Parole et Musique

Collège de France

Parole et Musique

Aux origines du dialogue humain

Colloque annuel 2008

Sous la direction de

Stanislas Dehaene et Christine Petit

Avec

Simha Arom, Anne Bargiacchi, Emmanuel Bigand, Jacques Bouveresse,
Roger Chartier, Ghislaine Dehaene-Lambertz, Michael Edwards,
Dan Gnansia, Claude Hagège, Martine Hausberger,
Régine Kolinsky, Christian Lorenzi, Helen Neville,
Pierre-Yves Oudeyer, Isabelle Peretz, Jean-Claude Risset,
Luigi Rizzi, Xavier Rodet, Peter Szendy, Monica Zilbovicius

Odile Jacob

Cet ouvrage s'inscrit dans le cadre de la collection
du Collège de France chez Odile Jacob.

Il est issu des travaux d'un colloque qui a eu lieu les 16 et 17 octobre 2008,
sous la responsabilité d'un comité scientifique
composé de Jean-Pierre Changeux, Roger Chartier, Antoine Compagnon,
Stanislas Dehaene, Pascal Dusapin, Christine Petit,
professeurs au Collège de France.
Il a reçu le soutien de la fondation Hugot du Collège de France.

La préparation de ce livre a été assurée
par Jean-Jacques Rosat, en collaboration
avec Patricia Llégou et Céline Vautrin.

15, rue Soufflot, 75005 Paris

www.odilejacob.fr

ISBN : 978-2-7381-2348-0
ISSN : 1265-9835

Préface

par Stanislas Dehaene et Christine Petit

Pour introduire notre sujet, un chiffre : quarante milliards de dollars. Tel est l'ordre de grandeur du marché annuel de la musique dans le monde. Chacun est prêt à dépenser des sommes importantes en appareils de haute-fidélité, en baladeurs, ou en disques dont la seule fonction n'est, après tout, que de dispenser quelques vibrations sonores. Ces vibrations ont cependant un pouvoir émotionnel considérable. Le pianiste Glenn Gould expliquait qu'il n'aurait pu imaginer sa vie sans une immersion totale dans la musique, et qu'il aurait été profondément malheureux au XIX^e^ siècle où les moments musicaux étaient rares et réservés à l'élite. Aujourd'hui, chacun aménage et transporte sur lui sa bibliothèque digitale, et chaque heure de trajet dans les transports en commun devient ainsi une heure de musique.

Robert Schumann écrivait : « La musique parle le langage général qui agite l'âme de façon libre et indéterminée. » Le puissant attrait qu'exerce la musique sur notre cerveau reste inexpliqué à ce jour. D'où provient le sens de l'harmonie, et pourquoi percevons-nous certains accords comme consonants et d'autres comme dissonants ? Pourquoi la gamme majeure nous paraît-elle vive et gaie, tandis que la gamme mineure nous semble effacée et introspective ? Ces messages musicaux varient-ils radicalement selon les cultures et les époques, ou bien font-ils appel à un petit jeu d'éléments invariants et universels, tels que l'octave, la quinte, le rythme binaire ou ternaire ? Ces éléments sont-ils propres à l'espèce humaine, ou bien se retrouvent-ils, au moins en germe, dans les comportements de communication d'autres espèces animales, tels que les chants des oiseaux ou les cris de haute fréquence qu'échangent les souris ? Notre cerveau a-t-il évolué pour la musique et, si oui, quelle en est la fonction primaire : communication sans langage, délimitation du groupe

social, renforcement des liens affectifs ? Ou bien la musique n'est-elle, comme le propose notre collègue le psychologue Steven Pinker, qu'une construction humaine récente et savante, un artifice culturel concocté dans le seul but de titiller les points les plus sensibles de nos facultés mentales ?

L'une des hypothèses les plus séduisantes est que la compétence musicale dérive de la faculté de langage propre à l'espèce humaine. Parole et musique partagent en effet de nombreux traits communs, dont le plus évident est une organisation hiérarchique par laquelle des éléments simples – notes ou phonèmes – se recombinent pour former, à plusieurs niveaux successifs, des structures de mots, de syntagmes et de phrases. Certes, l'origine du langage n'est pas moins disputée que celle de la musique, au point que cette question, comme on le sait, a été bannie officiellement des débats de la société de linguistique de Paris, dès l'année qui a suivi sa création en 1864. Cependant, des outils nouveaux, issus de l'imagerie cérébrale et de la génétique moléculaire, permettent d'en reprendre l'analyse. Il est donc particulièrement enrichissant de s'interroger sur les parallèles entre l'évolution de la musique et celle du langage parlé. Peut-on vraiment parler d'un langage musical ? Existe-t-il une forme de parenté entre les sonorités émises et traitées par l'un et l'autre système ? Ou bien, au contraire, peut-on affirmer avec Richard Wagner que « la musique commence là où s'arrête le pouvoir des mots » ? Au-delà des premières étapes du traitement acoustique, les messages linguistiques et musicaux empruntent des voies partiellement différentes dans notre cerveau. Cependant, la poésie, le chant et, tout particulièrement, l'opéra font s'entrecroiser langage et musique en une seule et même œuvre. Est-ce à dire qu'une seule et même grammaire universelle préside à l'organisation du langage parlé et de la musique ?

Le colloque *Aux origines du dialogue humain. Parole et musique*, qui s'est tenu les 16 et 17 octobre 2008 au Collège de France, entendait, sinon résoudre, du moins débattre de toutes ces questions en présence d'éminents spécialistes des différents champs disciplinaires impliqués : physiologie, neurosciences, psychologie, linguistique et anthropologie, mais également littérature, philosophie et création artistique, en particulier la musique et la poésie. Le présent livre en reprend les principales interventions.

Tout commence par… l'oreille, qui capte, mais aussi modifie et transforme les ondes sonores. Comment entendons-nous ? Cette question, Christine Petit la pose d'abord en physiologiste et généticienne, disséquant la structure de l'oreille interne où la cochlée, minuscule organe vibratoire, filtre les ondes sonores et les convertit en impulsions électri-

ques. Jacques Bouveresse reprend la question en philosophe et historien des sciences. Il rappelle les idées et les débats qui ont amené le grand physicien Hermann von Helmholtz à publier en 1862 sa magistrale *Théorie physiologique de la musique*, dans laquelle il explore les conséquences, pour la théorie musicale, de la décomposition spectrale des sons par l'oreille. Christian Lorenzi, enfin, démontre comment ces mécanismes, complétés par des traitements corticaux, contribuent à la perception de la parole en permettant la séparation des sources sonores et du bruit ambiant.

Au-delà du traitement perceptif initial, lorsqu'on s'interroge sur la capacité spécifique humaine de communiquer par la parole ou la musique, se pose immédiatement la question de l'arbitraire culturel. Comment les différentes cultures ont-elles stabilisé un code linguistique ou musical partagé par tous, qui permette le dialogue ? Pierre-Yves Oudeyer, informaticien, théorise et simule l'évolution d'un code phonologique par un phénomène d'« auto-organisation » : une population d'agents qui échangent des messages et s'imitent partiellement converge vers un code culturel partagé, stable, qui dépend à la fois des aléas de l'histoire et des attracteurs intrinsèques au système perceptif de chaque organisme.

Luigi Rizzi, linguiste, reprend cette analyse à un plus haut niveau, celui de la syntaxe des langues humaines. Toutes les langues de l'humanité, en dépit de leur apparente diversité, ne différeraient que par le choix d'un nombre limité de paramètres. Les principes linguistiques eux-mêmes seraient hautement invariants, et caractériseraient la compétence linguistique de l'espèce humaine. Dans le domaine musical, Isabelle Peretz, neuropsychologue de la musique, ne dit pas autre chose lorsqu'elle démontre l'existence d'amusies, des troubles sélectifs du développement qui peuvent affecter tel ou tel aspect de la perception musicale ou du chant. Chaque dimension de la musique (rythme, hauteur tonale, syntaxe…) ferait appel à des circuits cérébraux particuliers et susceptibles d'être sélectivement perturbés.

Cependant, à l'intérieur de l'espace des possibles, la diversité culturelle reste grande, à l'oral comme à l'écrit. Roger Chartier, historien de la lecture et du livre, analyse ce que la prosodie de la parole implique pour les systèmes de notation écrite : il a fallu, au fil des siècles, inventer des dispositifs de ponctuation tels que l'espace, la virgule ou le point d'interrogation afin de séparer les mots et d'en indiquer au lecteur le rythme et la respiration. À l'oral, voix et musique se mélangent fréquemment selon des modalités propres à chaque culture. Simha Arom, anthropologue, analyse une situation très étonnante : en Afrique subsaharienne : un code tambouriné permet de communiquer, entre des villages parfois distants de plusieurs kilomètres, par le biais d'un tambour à deux hauteurs tonales, de véritables phrases en partie stéréotypées et redondantes, mais toutefois véritables

foyers de communication à la frontière entre musique et langage. Xavier Rodet lui répond en disséquant, sous l'angle de l'informatique, ce qu'il y a de chanté et de parlé dans une voix humaine. Les logiciels de l'Ircam atteignent, dans ce domaine, des performances telles qu'il devient possible de modifier par ordinateur, par exemple, la ligne mélodique ou le caractère masculin ou féminin d'une voix.

Mais comment apprenons-nous la parole et la musique ? Ghislaine Dehaene-Lambertz, neuropédiatre et chercheur en sciences cognitives, étudie les circuits cérébraux de la parole et de la musique dès leur origine, chez le petit enfant de quelques mois, à l'aide de techniques innovantes d'imagerie cérébrale. Dès la naissance, les circuits du langage de l'hémisphère gauche sont organisés et prêts à apprendre le signal de parole, souvent en sélectionnant, parmi toutes les catégories linguistiques admissibles, celles qui sont utilisées dans la langue maternelle. Dès que l'enfant est âgé de quelques mois, parole et musique sont déjà partiellement séparés, respectivement dans les régions temporales supérieures des hémisphères gauche et droit. Pour Martine Hausberger, éthologue, l'apprentissage du langage présente de nombreux parallèles avec celui du chant chez l'oiseau. Les étourneaux apprennent différents dialectes selon leur région d'origine. Comme chez l'homme, les conditions sociales de l'apprentissage jouent un rôle déterminant : n'est bien appris que ce qui est enseigné par un « maître de chant » biologique, tandis que l'apprentissage purement associatif de sons délivrés par un haut-parleur donne des résultats beaucoup plus modestes, voire inexistants.

L'utilité de l'apprentissage du langage est évidente, mais qu'en est-il de la musique ? Helen Neville, psychologue et spécialiste de la neuroplasticité, utilise des études randomisées, semblables à celles que l'on mènerait pour tester un médicament, afin de mesurer l'impact de l'éducation musicale sur le développement cognitif et l'organisation cérébrale de l'enfant. La réponse qu'elle obtient est nette et importante : la musique, enseignée dans l'enfance, joue un rôle éminemment bénéfique, sans doute principalement lié à l'entraînement de l'attention. Monica Zilbovicius, neuropsychiatre, analyse une situation que l'on pourrait qualifier d'inverse – les troubles du langage et de la communication chez les enfants souffrant d'un syndrome autistique. Selon elle, une désorganisation bilatérale des régions auditives et communicatives du lobe temporal, visible dans différentes modalités d'imagerie cérébrale, existe chez toutes les populations d'enfants autistes, et pourrait rendre compte de leur communication déficiente.

Cependant, parole et musique sont bien plus que de simples instruments de communication. Michael Edwards nous restitue la pensée audible du poète et s'interroge : dans le domaine poétique, la « vie privée des mots » ne prend-elle pas le dessus sur la volonté authentique de com-

muniquer ? Sans doute la composition musicale est-elle tout aussi gouvernée par l'univers des possibles, l'ensemble des contraintes formelles propres à chaque époque, plus encore que par la nécessité d'exprimer à tout prix quelque message dont on voit mal le sens. Faisant appel à sa propre expérience, le compositeur et informaticien Jean-Claude Risset retrace les grandes étapes de l'informatique musicale et des joyaux sonores vraiment « inouïs » qu'elle permet aujourd'hui de synthétiser sur un coin de table. Chacun à leur manière, le linguiste Claude Hagège et le musicologue Peter Szendy étendent cette dialectique de l'inventivité et des contraintes dans la création artistique, l'un en s'intéressant à l'histoire de l'opéra, l'autre à celle, plus modeste mais tout aussi informative, d'une chanson populaire immortalisée par Mina et Dalida : *Paroles, paroles*.

L'émotion, plus encore que la communication, est au cœur de l'acte musical. Selon le psychologue cognitif Emmanuel Bigand, ses mécanismes psychologiques commencent à être mis en lumière. Le contexte de la phrase musicale passée induirait des attentes cognitives dont la résolution – soudaine ou différée, surprenante ou classique – déterminerait le pouvoir émotionnel de la composition musicale.

Si l'analyse scientifique jette ainsi une certaine lumière sur les mécanismes de la musique et de la parole, ni l'émotion musicale, ni la question des origines de la communication humaine, ne se laissent – pour l'instant ? – comprendre ou mettre en équations. Intime et personnel, l'envoûtant tourbillon de la musique et du chant doit s'apprécier sans discours. C'est pourquoi nous avions voulu que ce colloque de rentrée soit aussi l'occasion, tout simplement, d'entendre de la musique et de vibrer avec elle. Nous avions choisi, bien entendu, la voix chantée, lieu de rencontre naturel de la parole et de la musique. Ceux qui étaient présents se souviendront longtemps de la performance de la soprano Donatienne Michel-Dansac qui entreprit, en fin d'après-midi, après que notre cerveau eut été empli d'informations linguistiques, de nous « en-chanter » avec l'étonnante *Strette* d'Hector Parra et les *Quatorze récitations* de Georges Aperghis. Tous nos remerciements vont à l'Ircam (l'Institut de recherche et coordination acoustique/musique) qui avait accepté le pari d'installer, au Collège de France, quelques-uns de ses célèbres équipements d'informatique musicale afin de nous laisser entrevoir ce que deviendront, peut-être, la parole et la musique lorsque les machines entreront dans la partie.

Aucune parole ni aucun écrit ne saurait restituer ce que la musique nous apporte. Toutefois, les lecteurs pourront se référer au site Internet du Collège de France (www.college-de-france.fr) qui diffuse actuellement, en libre accès, les enregistrements audio et vidéo de ce colloque. Ultérieurement, ceux-ci seront consultables dans le cadre des archives du Collège de France.

I

ENTENDRE

Entendre : bases physiologiques de l'audition

par Christine Petit

L'intitulé de ce colloque, *Aux origines du dialogue humain. Parole et musique*, suggère l'enracinement du dialogue humain dans les vibrations de corps et d'objets, cordes vocales et instruments de musique, et leur perception. Le lien ainsi créé entre deux individus engendre une réponse de même nature, qui instaure la boucle du dialogue. Pour éclairer l'une des interrogations de ce colloque, le rapport entre les boucles de la parole et de la musique, ce chapitre introduit quelques notions élémentaires concernant le traitement des signaux acoustiques dans le système auditif et, tout particulièrement, des sons de parole et de musique. Il s'en tient à la rencontre du monde sonore et de la physiologie auditive chez l'homme. Les aspects plus intégrés de la perception auditive – son dialogue avec les autres sens ou avec l'émotion, qu'expriment intonation, prosodie et mélodie, l'entrée dans les silences, leur rupture, et ses relations avec la mémoire – ne sont pas discutés.

Tout objet vibrant est potentiellement audible, sous réserve que ces vibrations puissent gagner l'organe récepteur auditif en se propageant dans un milieu élastique (l'air pour les animaux terrestres), et que le système auditif soit sensible aux fréquences des vibrations émises. Peu d'espèces entendent : entendre est le privilège des vertébrés, à l'exception des plus primitifs comme la lamproie. Parce que l'audition est présente chez les requins, poissons à mâchoire, ce sens serait apparu chez les vertébrés il y a environ 400 millions d'années. D'abord dévolu à la perception des fréquences basses (sons graves), le système auditif est devenu, au cours de l'évolution des vertébrés, apte à traiter des fréquences de plus en plus élevées. Entendent aussi certains insectes, grillons, sauterelles, et d'autres arthropodes comme les crabes. Comme tous les autres sens, l'audition dans chaque espèce est adaptée à la niche écologique qu'elle occupe.

L'audition contribue à la reproduction, par le chant nuptial émis par certains insectes et oiseaux. Elle participe aux rapports de domination, par le chant dit de proclamation territoriale, par lequel certains oiseaux marquent l'espace qu'ils s'attribuent. Elle prend part à la survie des individus et des espèces par la détection des proies et des prédateurs. L'homme contemporain est rarement aux prises avec un animal agresseur. C'est plutôt à l'irruption d'un avertisseur sonore, quelle qu'en soit la nature, qu'il doit faire face. Il doit pouvoir le localiser et l'identifier dans des environnements de plus en plus bruités.

Il y a une trentaine d'années, Albert Bregman a développé le concept de « scène auditive reconstruite », repris dans un ouvrage publié en 1990[1]. Confronté à une scène sonore, le système auditif peut apporter réponse à une série d'interrogations : Combien êtes-vous ? Où êtes-vous ? Qui êtes-vous ? Et même, que faites-vous ? Voire, m'entendez-vous ? Les principes du traitement des signaux acoustiques qui président à cette analyse sont les mêmes dans les diverses espèces. La ségrégation d'événements environnementaux sonores en flux acoustiques et représentations perceptives distinctes, proposée par Bregman, repose sur des principes analogues à ceux qui opèrent dans le système visuel.

La singularité du sens auditif réside dans son lien avec la communication acoustique. Mode d'échange majeur dans les espèces entendantes, il l'est plus encore chez celles qui font l'apprentissage de vocalisations, parmi lesquelles les oiseaux dits chanteurs, dont beaucoup appartiennent à l'ordre des passereaux (corneilles, bruants, mésanges…). Cette aptitude est également présente chez les chauves-souris, les baleines et les dauphins, certains rats (rat-kangourou), certains éléphants (éléphant d'Afrique), et bien sûr l'homme. Les vocalisations apprises assurent la communication entre membres d'une même espèce. Les petits émettent tout d'abord des vocalisations innées. Puis, sous influence sociale, des vocalisations acquises vont progressivement se substituer aux précédentes. Ces apprentissages mettent en jeu la boucle audio-phonatoire. Ainsi, le nourrisson sourd profond émet des lallations, mais les sons qu'il émet ne s'organisent pas progressivement en mots, phrases, puis véritable langage parlé comme chez l'enfant entendant.

L'audition est un champ scientifique intrinsèquement pluridisciplinaire, auquel contribuent physiciens, psychoacousticiens, neurobiologistes et musiciens. Ce domaine est riche en énoncés de concepts, et les interfaces entre ces diverses disciplines sont de plus en plus fécondes. L'apport

1. Albert S. Bregman, *Auditory Scene Analysis. The Perceptual Organization of Sound*, Cambridge (Mass.), MIT Press, 1990.

des ingénieurs est aussi particulièrement important. Inventions et développements techniques se mêlent étroitement aux découvertes scientifiques. Ainsi, l'invention du téléphone, que l'on attribue à Alexandre Graham Bell (1876), est-elle à l'origine de la première prothèse auditive électroacoustique.

La démarche scientifique des physiciens se situe, pour l'essentiel, du côté de l'objet. Elle cherche à identifier les paramètres physiques des sources sonores qui sont extraits par le système auditif. De fait, la plupart des principes du fonctionnement de ce système, du moins dans les toutes premières étapes du traitement des signaux sonores, ont été découverts par les physiciens. Les psychoacousticiens se placent du côté du sujet et de la perception. Ils s'efforcent de comprendre comment le sujet attribue certaines qualités à l'objet sonore, et les différences qui existent entre réalité physique et perception sonore. Les neurobiologistes, quant à eux, étudient le fonctionnement du système auditif en termes d'activité individuelle ou collective de neurones ou de réseaux de neurones.

La première relation établie entre grandeur physique et perception sonore fut celle qui lie longueur des cordes vibrantes et consonance. Elle est généralement attribuée à Pythagore (environ 569-494 av. J.-C.). Toutefois, Helmholtz dans un ouvrage publié en 1862, *Die Lehre von den Tonempfindungen als physiologische Grundlage fur die Theorie der Musik*[2], curieusement traduit en français sous le titre *Théorie physiologique de la musique fondée sur l'étude des sensations auditives*, indique qu'il y a lieu de penser que ces connaissances sont bien antérieures. Deux cordes de même nature, soumises à une même tension, sont dans une consonance parfaite si leurs longueurs sont dans un rapport simple. Si ce rapport est de 1, elles vibrent à l'unisson. S'il est de 2, les deux sons émis sont séparés d'une octave ; s'il est de 2/3, d'une quinte ; de 3/4, d'une quarte… D'où l'idée selon laquelle la puissance émotionnelle associée à la perception de la consonance traduit l'expression d'un ordre mathématique de l'univers. L'abbé Marin Mersenne, un proche de Descartes, introduisit la notion de fréquence sonore. Elle représente le nombre de répétitions des vibrations (mouvements d'aller et de retour des corps vibrants) par seconde, avec pour unité de mesure, le hertz. Un son pur a une fréquence unique, qui est celle de l'onde sinusoïdale de pression qui le produit. L'amplitude de cette onde de pression correspond à l'intensité sonore. Vinrent ensuite les notions de fréquence fondamentale et d'harmoniques. La fréquence fondamentale est la fréquence la plus faible de la vibration stationnaire d'une

2. Voir Jacques Bouveresse, « Helmholtz et la théorie physiologique de la musique », dans le présent ouvrage, p. **NN-NN**.

corde. Les corps naturels, pour la plupart, ne vibrent pas à cette seule fréquence. Ils produisent également un mélange de sons purs dits « harmoniques », dont les fréquences sont des multiples entiers de la fréquence fondamentale. Le spectre fréquentiel d'un tel son « complexe » représente ses diverses composantes fréquentielles élémentaires et leurs amplitudes respectives. Le mouvement vibratoire produisant le son est aussi décrit par sa phase, qui indique la situation instantanée dans le cycle sonore. Les sons diffèrent aussi par leur timbre, une notion de nature perceptive, définie comme ce qui distingue deux sons de même hauteur, de même intensité et de même durée. Par exemple, le *la 440 Hz* sonne différemment selon qu'il est joué par une clarinette ou par un hautbois. Le timbre est conditionné par une variété de paramètres. Il dépend notamment de l'intensité relative des différents harmoniques du son considéré.

Les sons de notre environnement, ceux de la parole ou de la musique, sont complexes. Ils comportent de multiples harmoniques. Leur structure temporelle est, elle aussi, complexe. Cette dernière fait référence à une organisation en motifs qui se répètent régulièrement au cours du temps avec une fréquence propre, par exemple une modulation régulière en amplitude du spectre fréquentiel. Cette dualité spectrale et temporelle des sons, bien que dépourvue de signification en termes physiques, a été introduite parce qu'elle correspond à un traitement distinct de ces paramètres acoustiques dans le système auditif[3].

Le système auditif est un système acoustico-électrique ; il convertit les signaux acoustiques en signaux électriques. Chez l'homme, sa résolution fréquentielle est élevée pour les fréquences inférieures à 5 kHz qui caractérisent, pour l'essentiel, les sons de parole et de musique. Cette capacité à discriminer des fréquences proches est indispensable à l'écoute musicale. Elle atteint le 1/1 000 de la valeur fréquentielle pour les fréquences auxquelles le système est le plus sensible. Sa résolution temporelle est, quant à elle, exceptionnelle. Ces performances sont très supérieures à ce que nécessite la reconnaissance de la parole dans des environnements silencieux. En revanche, le système auditif peut extraire des signaux de parole émis dans une ambiance bruitée bien plus efficacement que n'importe quel algorithme que l'on sache produire aujourd'hui. Cette aptitude est cependant vulnérable. La difficulté à entendre dans le bruit est l'une des premières manifestations de la baisse de l'acuité auditive chez la personne vieillissante. De l'organe sensoriel, la cochlée, aux voies auditives afférentes, jusqu'au cortex auditif, plusieurs mécanismes contribuent

3. Christine Petit, « Des capteurs artificiels à la perception auditive », *in* Jean-Pierre Changeux (dir.), *L'Homme artificiel*, Paris, Odile Jacob, 2007, p. 211-222.

à extraire le message d'intérêt du bruit qui le parasite. Dès les premières étapes du traitement du signal auditif, c'est-à-dire dans la cochlée et les neurones auditifs primaires, des effets de masquage sont mis en œuvre. Le système nerveux efférent, beaucoup plus développé dans le système auditif que dans les autres systèmes sensoriels, augmente de façon importante la réponse à un stimulus acoustique en présence de bruit, par la modulation qu'il exerce sur les informations transmises dans les voies afférentes.

En 1754, le violoniste Giuseppe Tartini rapporta que si deux instruments jouent simultanément deux notes qui sont dans un rapport fréquentiel d'environ 1,2, des sons additionnels sont perçus par les auditeurs, dont les fréquences correspondent à la combinaison des fréquences jouées comme de leurs harmoniques. Ainsi, le système auditif n'offre pas une reproduction « haute-fidélité » des sons, et participe même à la création de notre monde sonore. Dans la cochlée, les messages sonores sont distordus. La distorsion porte sur le temps d'arrivée de l'onde sonore à la cellule sensorielle accordée à la fréquence de ce son. Ce temps est plus long pour les ondes de basse fréquence que pour celles de haute fréquence, en raison de l'organisation tonotopique de la cochlée (voir ci-dessous). La distorsion porte aussi sur l'intensité sonore. En effet, la cochlée amplifie les signaux acoustiques de manière non linéaire : plus faible est le son, plus il est amplifié dans la cochlée. Enfin, la cochlée distord la forme même de l'onde sonore et fait apparaître des produits fréquentiels absents du spectre acoustique initial. À côté de leur traduction perceptive, soulignée par l'observation de Tartini, ces distorsions ont aussi une traduction acoustique. En effet, l'oreille, spontanément ou en réponse à une stimulation sonore, émet elle-même des sons, que l'on nomme otoémissions acoustiques, et qui peuvent être détectés à l'aide d'un petit microphone placé dans le conduit auditif externe. En particulier, en réponse à deux sons purs simultanés de fréquences proches, la cochlée émet des produits acoustiques d'intermodulation, tels que les sons de Tartini, que l'on recueille dans les otoémissions acoustiques (produits de distorsion). Parce que la production de ces produits acoustiques de distorsion requiert l'activité de l'une des deux catégories de cellules sensorielles auditives (les cellules ciliées externes), presque toujours affectée dans les surdités précoces, ils sont mis à profit dans un test de dépistage de la surdité chez le nouveau-né. De surcroît, le système auditif a la propriété d'élaborer la perception de fréquences particulières sans même avoir produit l'onde correspondante. Il en est ainsi de la perception de la fréquence fondamentale à partir de ses seuls harmoniques. Enfin, ni la perception de hauteur (tonie) ni celle d'intensité (sonie) ne sont de simples représentations de la fréquence et de l'amplitude de l'onde sonore. Des multiples

facettes du système auditif, nous n'envisageons ici, dans le contexte de ce colloque, que celle de l'instrument de représentation de la réalité physique, puis celle de l'instrument de la conquête du dialogue.

Le système auditif, instrument de représentation de la réalité physique

Le système auditif comporte une partie dévolue à la transmission de l'onde sonore, l'oreille externe et moyenne. Parmi les fréquences auxquelles notre système auditif répond, de 20 Hz à 20 kHz, celles qui sont voisines de 3 500 Hz subissent une amplification d'environ 15 dB par résonance dans le conduit externe, tandis que celles qui sont inférieures à 500 Hz et supérieures à 5 000 Hz subissent au contraire une baisse de leur amplitude lors du passage des sons à travers l'oreille moyenne. Fait suite à cet appareil de transmission du son, le système sensoriel auditif périphérique, composé de la cochlée et du nerf auditif. Dans la cochlée, les cellules sensorielles auditives convertissent l'énergie associée à l'onde mécanique acoustique en signaux électriques transmis aux neurones du nerf auditif. Ces derniers codent l'information acoustique sous forme de potentiels d'action, et la transfèrent aux voies auditives centrales, qui comportent quatre relais jusqu'au cortex auditif[4]. Dans ces relais qui, à partir du second, reçoivent des signaux provenant des deux oreilles, sont extraits d'autres paramètres qui caractérisent le son, comme sa durée, ou la vitesse à laquelle il atteint son intensité maximale.

Les physiciens ont établi trois grands principes d'analyse des sons dans le système auditif : celui de l'analyse fréquentielle, celui de la localisation de la source sonore, enfin celui de la nécessité d'une amplification de la stimulation sonore au sein de la cochlée.

En 1843, Georg Simon Ohm[5] proposait que l'oreille se comporte comme un analyseur fréquentiel de Fourier. Vingt ans plus tôt, Joseph Fourier avait en effet énoncé le principe selon lequel toute fonction périodique peut être décomposée en une série de fonctions sinusoïdales élémentaires d'amplitudes et de phases appropriées. Pour Ohm, l'oreille devait ainsi décomposer les sons complexes en leurs fréquences élémentai-

4. Arthur N. Popper et Richard R. Fay (éd.), *The Mammalian Auditory Pathway : Neurophysiology*, Springer, 1992.
5. Georg Simon Ohm, « Über die Definition des Tones, nebst daran geknüpfter Theorie der Sirene und ähnlicher tonbildender Vorrichtungen », *Annalen der Physik und Chemie*, 59, p. 513-565 (1843).

res. De fait, dans la cochlée, une membrane tendue de sa base à son sommet se comporte comme un analyseur fréquentiel en raison de ses caractéristiques physiques, graduellement variables d'une extrémité à l'autre. Sous l'effet d'une stimulation sonore, cette membrane subit un déplacement maximal à un emplacement donné le long de son axe baso-apical. Cet emplacement est corrélé à la fréquence de l'onde sonore, de sorte qu'une représentation spatiale des fréquences sonores est créée le long de la cochlée : on parle de carte tonotopique. Parce que cette membrane est couplée mécaniquement aux cellules sensorielles, ces cellules sont elles aussi activées de façon optimale par une fréquence sonore donnée, qui dépend de leur emplacement. Les cellules sensorielles elles-mêmes participent à cette réponse fréquentielle. Toutes différentes les unes des autres par leurs propriétés morphologiques et biophysiques, elles affinent la réponse cochléaire par leur réponse préférentielle à une fréquence donnée. Elles sont, en quelque sorte, accordées en fréquence. La tonotopie est le principe fondamental d'organisation de l'ensemble du système sensoriel auditif. Établie au niveau cochléaire, on la retrouve dans tous les relais des voies auditives et au niveau du cortex auditif. La multiplicité de ces cartes tonotopiques, huit au moins dans le cortex auditif du macaque, indique que les autres paramètres des sons sont extraits, le long des voies et dans les aires corticales auditives, dans leur contexte fréquentiel.

En 1907, Lord Rayleigh[6] montra que la localisation des sources émettrices de sons de basse fréquence, inférieure à 1500 Hz, repose sur la détection de la différence des temps d'arrivée de l'onde sonore à l'une et à l'autre oreille. Chez l'homme, le délai temporel minimal décelable est d'environ 13 microsecondes. Chez certaines chauves-souris, il atteindrait le centième de microseconde. Quel est le support physiologique d'une telle résolution temporelle du traitement des messages sonores ? Un de ces éléments, bien établi, est la transformation de la carte tonotopique cochléaire en une représentation temporelle très précise. Pour les fréquences allant jusqu'à 4 à 5 kHz, les neurones auditifs répondent par des décharges de potentiels d'action strictement synchronisées avec une phase donnée de l'onde sonore. C'est à partir de cette information neuronale provenant de chacune des deux oreilles que des neurones situés dans le complexe olivaire supérieur, second relais du système auditif central, détectent le délai temporel interauriculaire, c'est-à-dire le décalage de phase que présente l'onde sonore lorsqu'elle parvient à l'une et à l'autre oreille. Une représentation spatiale de ces délais temporels, c'est-à-dire une

6. Lord Rayleigh (John William Strout), « On our perception of sound direction », *Philosophical. Magazine*, 13, p. 214-232 (1907).

autre carte, associée à une carte fréquentielle, se loge dans ce second relais central. On retrouve une carte similaire dans le cortex auditif primaire.

Comment rendre compte de la sensibilité considérable de l'audition, de son seuil de détection d'une énergie acoustique à peine dix fois supérieure à l'énergie du bruit thermique, alors que les cellules sensorielles de la cochlée baignent dans un milieu liquidien, qui amortit donc toute stimulation mécanique ? C'est la question que souleva Thomas Gold en 1948. Il postula alors l'existence d'un amplificateur mécanique actif au sein de la cochlée. Aujourd'hui l'existence d'une fonction amplificatrice cochléaire est bien établie. Elle est assurée par les cellules ciliées externes, qui ont la particularité de se contracter lorsqu'elles sont dépolarisées, une propriété connue sous le nom d'électromotilité.

Ces quelques exemples illustrent la curiosité que suscitent depuis longtemps, chez les physiciens, les performances du système auditif en sensibilité et en précision temporelle.

Le système auditif, instrument de la conquête du dialogue

Le système auditif se conçoit aussi comme un instrument de mise en forme des signaux sonores, qui peuvent ainsi prendre leur signification et s'inscrire dans la communication acoustique par le langage et la musique. Que sont les sons de parole et de musique ? Les modalités de leur perception sont-elles distinctes ?

La production des sons de parole par le tractus vocal permet d'en comprendre les caractéristiques. Le tractus vocal, de la glotte aux lèvres, comporte dix-sept points d'articulation. Par leur mobilité, ils modifient la géométrie des différentes cavités de ce tractus. Les cordes vocales, simples replis de la muqueuse du larynx, font vibrer l'air venu des poumons. Elles produisent des harmoniques, qui cheminent à travers les cavités du tractus vocal. Ces cavités se comportent comme une suite de résonateurs. Elles amplifient sélectivement la fréquence sonore qui est celle de leur résonance propre, elle-même conditionnée par leur géométrie. Tel harmonique, parmi l'ensemble des harmoniques d'un son de parole, sera ou non amplifié lors du passage à travers l'une des cavités. On appelle « formants » les pics de résonance des harmoniques de parole ainsi produits. Or le tractus vocal est animé d'une plasticité considérable. Des changements dynamiques de sa configuration surviennent à un rythme soutenu, qui modifient sa géométrie, parfois de façon très importante, comme par exemple lors des mouvements de la langue. Le tractus vocal est donc par

excellence l'instrument des contrastes fréquentiels rapides. La voix humaine peut produire jusqu'à huit cents phonèmes distincts, consonnes et voyelles, au rythme d'une dizaine par seconde, et qui présentent des variations internes de leur spectre fréquentiel, avec même des vides sonores pour certaines consonnes. Transition rapide d'un formant à un autre et montée subite en amplitude de l'un ou de l'autre sont les caractéristiques des sons de parole. Ces contrastes dynamiques excèdent, par leur amplitude et leur vitesse, ceux que peuvent produire les instruments de musique.

Les sons musicaux sont, quant à eux, catégorisés davantage par leur hauteur et par leurs intervalles de hauteur. Ceci renvoie à la notion de consonance/dissonance. Les sons distants d'une octave sont perçus comme parfaitement consonants : les fréquences de leurs harmoniques respectifs sont identiques. Si on considère maintenant deux sons à la quinte, de 220 Hz et de 330 Hz par exemple, un harmonique sur deux du son le plus aigu coïncide avec celles du son le plus grave. L'intervalle paraît juste. D'où l'idée selon laquelle la perception de consonance repose sur une fusion des harmoniques réalisée par le système auditif. Qu'il s'agisse de la parole ou de la musique, la sensation de hauteur sonore repose principalement sur la perception de la fréquence fondamentale. Or le regroupement des harmoniques, bien réel, permet la détection de la fréquence fondamentale.

En fait, des compositions sonores très différentes peuvent conduire à la perception d'une même hauteur sonore[7]. Soit un son harmonique, constitué de trois fréquences, 100, 200 et 300 Hz. Sa fréquence fondamentale est de 100 Hz, et il est perçu comme un son de 100 Hz. De brèves séquences de bruit, qui par définition comportent un grand nombre de fréquences différentes et qui se répètent toutes les 10 ms, soit avec une fréquence de 100 Hz, sont aussi perçues comme un son de 100 Hz, même si son timbre est différent de celui du son harmonique précédent. Soit un son de haute fréquence, dont l'amplitude est modulée avec une fréquence de 100 Hz. Il prend la couleur de la modulation et est perçu comme un son grave de fréquence 100 Hz. Soit, enfin, trois sons de fréquences 200, 300 et 400 Hz émis simultanément. Ce mélange va conduire à la perception d'un son unique de 100 Hz, c'est-à-dire un son dont la fréquence fondamentale est le plus grand dénominateur commun à ces trois sons, alors même que cette fréquence est absente du spectre fréquentiel du mélange sonore. Les échanges téléphoniques offrent une illus-

7. Christopher J. Plack, Andrew J. Oxenham, Richard R. Fay et Arthur N. Popper (éd.), *Pitch. Neural Coding and Perception*, Springer, 2005.

tration de la détection de la fréquence fondamentale manquante par le système auditif. Les fréquences fondamentales de la voix humaine, principalement masculine, ne sont pas transmises par le téléphone. Pourtant l'interlocuteur les perçoit. Si la voix ne lui paraît pas déformée, c'est parce que son système auditif « construit » la perception de cette fréquence fondamentale à partir du spectre des harmoniques qui lui parviennent. Seules les fréquences fondamentales manquantes inférieures à 800 Hz sont extraites. Leur détection est très robuste : on peut supprimer certains harmoniques, et elle persiste. Quelles fonctions attribuer à la détection de la fréquence fondamentale ? Quel avantage sélectif y serait associé ? Parce qu'elle est fondée sur le regroupement de tous les harmoniques d'une même fondamentale, cette performance du système auditif conduit à les identifier comme provenant d'une source sonore unique, et permet donc l'appréciation du nombre des sources émettrices : animaux agresseurs dans les milieux naturels, par exemple. Elle permet aussi l'appréciation du gabarit de l'émetteur, en raison de la relation qui existe entre fréquence fondamentale et longueur des cordes vocales. Autre intérêt, au quotidien : l'écoute simultanée de deux locuteurs, l'un masculin et l'autre féminin, très proches l'un de l'autre et situés à distance de l'auditeur. Si ce dernier ne parvient pas à ségréger leurs sons de parole respectifs en se fondant sur la localisation spatiale des sources sonores, il peut réussir à les écouter conjointement, par l'extraction des fréquences fondamentales de leurs sons de parole, différentes chez l'homme et chez la femme. L'extraction de la fréquence fondamentale manquante est une fonction ancestrale : on la trace jusqu'au poisson rouge.

L'extraction de la fréquence fondamentale repose donc soit sur le traitement des fréquences harmoniques des sons, soit sur celui de leur structure temporelle (voir ci-dessus). Le traitement fréquentiel des harmoniques suppose qu'ils puissent être distingués les uns des autres lorsqu'ils sont présents simultanément. Ceci nous renvoie à la question de la distinction de deux sons joués conjointement. Considérons deux sons de fréquences voisines, 200 et 230 Hz par exemple, joués successivement, l'auditeur les perçoit comme de hauteurs différentes. Joués ensemble, ils créent un son composite rugueux et d'amplitude fluctuante (phénomène de battement). Cette perception est mise à profit pour accorder les instruments de musique. Ainsi, dans un orchestre, les musiciens accordent-ils leurs instruments sur une note de référence, en ajustant progressivement la tension des cordes de leur instrument jusqu'à ce que ces sons désagréables disparaissent. Si maintenant on éloigne les deux fréquences l'une de l'autre, les sons initiaux joués conjointement deviennent distincts. Ceci définit la bande critique. Deux sons de fréquence voisine appartiennent à

une même bande critique si, émis simultanément, ils sont perçus comme un son rugueux, désagréable. Considérons maintenant les harmoniques d'un son : les premières, celles dont les fréquences sont les plus basses, appartiennent à des bandes critiques différentes, elles sont dites résolues et peuvent servir à l'extraction de la fréquence fondamentale. Au-delà de le 6e harmonique, deux harmoniques peuvent occuper la même bande critique. Dans ce cas, d'autres mécanismes doivent être mis en œuvre pour extraire la fréquence fondamentale. La structure temporelle du son pourrait être le fondement de l'extraction de la fondamentale à partir des seuls harmoniques non résolus. Quoi qu'il en soit, la notion de bande critique renvoie à celle de dissonance, puisque deux sons dont les bandes critiques sont les mêmes, ou se chevauchent largement, sont dissonants.

Suivons ces sons de parole et de musique jusqu'au cortex. La révolution qu'a introduite l'imagerie fonctionnelle cérébrale (imagerie par résonance magnétique nucléaire, tomographie par émission de positons) dans les neurosciences, en permettant de suivre l'activation des aires corticales au décours de la perception, a confirmé la latéralisation cérébrale de la perception de la parole et de la musique, établie depuis bien longtemps en se fondant sur l'effet des lésions du cortex observées chez certains patients. Chez la plupart des individus, l'hémisphère gauche est dévolu à la parole, l'hémisphère droit à la musique. L'imagerie fonctionnelle a authentifié l'existence d'une carte des hauteurs sonores, aux confins des aires auditives corticales primaires (gyrus de Heschl) et de leur ceinture, là où convergent les régions « basse fréquence » de plusieurs cartes tonotopiques. L'imagerie cérébrale permet de réinterroger beaucoup plus finement cette apparente différenciation des deux hémisphères pour le traitement des sons de parole et de musique dans leur étape finale. S'ancre-t-elle dans les signatures acoustiques distinctes des sons de parole et de musique évoquées plus haut ? Ou bien, la perception de la parole dépend-elle de mécanismes spéciaux exclusivement dédiés au traitement des sons de parole ? Des travaux récents indiquent qu'il existe des paramètres physiques des séquences sonores dont on peut prédire qu'ils activeront de manière asymétrique les cortex auditifs. Ainsi, il semble que les sons dont les variations temporelles sont très rapides, et ceux dont la fréquence est particulièrement élevée, sont traités de manière asymétrique. Chez la plupart des individus, la résolution temporelle appartiendrait préférentiellement à l'hémisphère gauche, et la résolution fréquentielle à l'hémisphère droit. Le rapide glissement des formants, en rapport avec l'articulation de l'appareil phonatoire, solliciterait préférentiellement l'hémisphère gauche. Au-delà, la composition linguistique des sons de parole, qui conduit à leur reconnaissance phonétique et à l'émergence du

sens, implique un contact avec les traces de mémoire et, quelles que soient les caractéristiques acoustiques des sons de parole, cette reconnaissance est traitée de manière asymétrique, généralement à gauche, dans les aires du langage.

Nous avons lié ici audition et langage, sous-entendant qu'il n'y a de langage qu'oral. Rappelons donc, pour finir, que la « parole gestuelle », dont les messages visuels sont traités dans les aires corticales du langage, est aussi un vecteur possible du dialogue, en particulier entre personnes non entendantes[8].

8. REMERCIEMENTS. Je remercie chaleureusement Jean-Pierre Hardelin pour la relecture de ce texte.

Helmholtz et la théorie physiologique de la musique

par JACQUES BOUVERESSE

L'acoustique et la théorie de la musique selon Helmholtz

Pour donner une idée du genre de bouleversement que les recherches de Helmholtz ont provoqué à la fois dans l'acoustique en général et l'acoustique musicale en particulier, dans la théorie musicale, et dans la musique elle-même, je commencerai par citer deux extraits du livre que Marcus Rieger a publié en 2006 sur cette question. Le premier a trait à la manière dont le travail de Helmholtz, après avoir suscité, pour commencer, des résistances très fortes, a fini par être accepté à peu près par tout le monde :

> L'histoire de la réception de *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik*[1] montre que dans l'espace d'à peine 40 ans le refus initial qu'un bon nombre de musiciens et de musicologues opposaient à l'œuvre de Helmholtz a disparu. Les conceptualisations, les définitions, les instruments et les projets sortis de son laboratoire ne sont pas seulement entrés dans les écrits de science musicale, mais sont apparus également dans le quotidien de musiciens et d'autres profanes en matière de sciences naturelles. Des écrits de science

1. Ce titre peut se traduire : « La théorie des sensations sonores comme fondement physiologique pour la théorie de la musique ». La traduction française du livre a été publiée cinq ans après la première édition allemande, sous le titre *Théorie physiologique de la musique fondée sur l'étude des sensations auditives*, traduit de l'allemand par M. G. Guéroult, Paris, Victor Masson et fils, 1868 ; réédité par Sceaux, Jacques Gabay, 1990 (désormais *TPM*).

> populaire, le téléphone et le phonographe ont contribué à ce résultat que, vers la fin du XIX[e] siècle, le son objectif et l'oreille objectivée ont reçu un accueil favorable également en dehors du laboratoire.
> Les voix sceptiques de musiciens et de musicologues, qui pour commencer déterminaient encore la réception de l'œuvre, se sont tues assez rapidement. Alors que les critiques de la première édition mettaient encore en doute la pertinence des *Tonempfindungen* pour la recherche musicale, les musicologues comme [Hugo] Riemann ou [Carl] Stumpf n'ont trouvé à contester que certains résultats particuliers de la recherche, sans mettre en question, ce faisant, le cadre de la science naturelle en tant que tel. Dans la pratique musicale également, on peut observer la manière dont la récusation des *Tonempfindungen* se transmue en un grand intérêt pour les connaissances qui relèvent des sciences de la nature. Alors que les musiciens dans les années 60 et 70 du XIX[e] siècle ne savaient pas encore faire grand-chose avec les phonautographes, les analyseurs de son, les sirènes et les machines à synthétiser, ceux-ci pour Varèse sont devenus déjà fondamentaux pour une nouvelle musique qui s'est établie au-delà du monde sonore de la tonalité majeur-mineur que Helmholtz avait fait reposer sur une base relevant des sciences de la nature[2].

Les idées et les recherches de Helmholtz ont effectivement exercé une influence importante sur certains musiciens qui connaissaient de près son livre sur les sensations sonores. On cite généralement, à ce propos, Leos Janacek, George Ives (le père de Charles Ives) et surtout Edgar Varese. La critique que Helmholtz a formulée contre le tempérament égal n'est pas non plus restée totalement sans effet dans le monde musical. Il n'est pas impossible que Paul Hindemith, bien qu'il ne se réfère pas à Helmholtz, s'en soit inspiré dans *Unterweisung im Tonsatz* (1937). Et, parmi les contemporains, il y a au moins un compositeur, Wolfgang von Schweinitz, qui a pris fait et cause pour le tempérament helmholtzien et écrit des œuvres dont certaines se réfèrent jusque dans leur titre à Helmholtz et sont conçues pour des instruments accordés de la façon qui lui semblait à la fois la plus naturelle et musicalement la plus satisfaisante.

Dans le deuxième extrait que je voudrais citer, il est question de la transformation radicale que Helmholtz a, selon l'auteur, fait subir à notre compréhension de la musique :

---

2. Marcus Rieger, *Helmholtz Musicus. Die Objektivierung der Musik im 19. Jahrhundert durch Helmholtz' Lehre von den Tonempfindungen*, Darmstadt, Wissenschaftliche Buchgesellschaft, 2006, p. 151-152.

> Dans les *Tonempfindungen*, Helmholtz formule une compréhension de la musique qui ne repose plus sur la tradition ou sur les expériences de musiciens, mais sur les mesures d'instruments de laboratoire objectivants. Pourquoi il ne pouvait pas, ce faisant, ne pas frapper d'un coup sur la tête tous ceux qui, soit pratiquement soit théoriquement, s'occupaient de musique, c'est ce que montrent ses définitions de concepts et de termes originairement musicaux : il dégrade le son en une vibration singulière périodique ; l'oreille musicale, dont l'éducation pour le musicien était encore une partie de son art, il la présente comme un appareil qui sent des excitations ; et l'activité qui consiste à faire de la musique est devenue chez lui une forme de production de sons, qui pouvait être exécutée aussi bien par l'acousticien dans le laboratoire avec des appareils pour la synthèse du son que par des musiciens dans la salle de concert. Et, pour finir, il explique à ses lecteurs que la consonance et la dissonance diffèrent non pas, comme on l'avait admis depuis Pythagore, qualitativement mais graduellement.
> […] Jusque dans les années 60 du XIX[e] siècle l'importance de l'acoustique musicale en dehors du laboratoire était marginale. Quand Helmholtz en 1863 s'est présenté à ses lecteurs avec sa conception de la musique scientifico-objective d'un type nouveau, ce qui dominait dans la science de la musique était encore l'histoire, l'esthétique et la théorie de la composition[3].

Je trouve personnellement très étrange cette façon de s'exprimer. Modifier la compréhension que nous avons des fondements de la musique en essayant de la rendre scientifique, ce qui est bien ce que Helmholtz a cherché à faire, n'est pas du tout la même chose que chercher à modifier et à rendre scientifique la compréhension que nous avons de la musique, si on entend par là la compréhension des œuvres musicales. Helmholtz lui-même était à la fois un trop grand scientifique et un trop bon musicien pour suggérer quoi que ce soit de ce genre, et il n'a sûrement jamais pensé qu'une compréhension scientifique des fondements de la musique était susceptible de modifier radicalement la relation que nous entretenons avec celle-ci et de se substituer plus ou moins au genre de connaissance et d'expérience qu'implique ce qu'on appelle ordinairement la compréhension de la musique. Il n'avait manifestement aucun doute sur le fait que cette dernière dépend et continuera à dépendre de façon essentielle de la formation musicale, de l'histoire, de la tradition et de la culture. Mais il y a malgré tout une chose qui n'est sûrement pas contestable dans ce que dit Marcus Rieger : une bonne partie de la musique qui se fait et s'écoute aujourd'hui n'aurait pas été concevable sans le genre de révolution que Helmholtz a provoqué à la fois dans la

3. *Ibid.* p. 2.

théorie et dans la pratique de la musique. Comme cela se passe pratiquement toujours en pareil cas, cela ne signifie pas nécessairement que les changements qui ont eu lieu correspondent à une évolution qu'il aurait pu souhaiter et approuver lui-même ; mais c'est évidemment une tout autre question.

Helmholtz a commencé ses recherches sur l'acoustique physiologique en 1855, à une époque où il n'en avait pas encore terminé avec la rédaction de son *Manuel d'optique physiologique* (*Handbuch der physiologischen Optik*), une œuvre monumentale dont la première partie a été publiée en 1856, la deuxième en 1860 et la troisième seulement en 1867. Il s'est lancé dans cette nouvelle entreprise avec l'ambition déclarée de réformer en profondeur l'acoustique physiologique et de faire pour la physiologie du sens auditif quelque chose de comparable à ce qu'il avait fait pour celle du sens visuel. Le livre qu'il a publié huit ans plus tard, en 1863, *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik,* constitue l'exposé synthétique et systématique des résultats auxquels il a abouti au cours de cette période et qu'il avait déjà présentés, au moins en partie, dans différents articles publiés antérieurement.

Il semble cependant qu'il ait été très vite en possession des idées principales qui sont développées dans le livre. Cela ressort clairement d'une lettre envoyée le 16 janvier 1857 à William Thomson (le futur Lord Kelvin), dans laquelle il écrit :

> Je publierai bientôt une deuxième partie de mes expériences acoustiques. […] La théorie de l'harmonie et de la disharmonie et des accords peut être dérivée complètement des recherches sur les battements des harmoniques et des sons résultants[4]. J'ai préparé un appareil pour étudier la qualité sonore. Grâce à l'hypothèse selon laquelle tout son d'une hauteur donnée est senti par une fibre nerveuse particulière qui est connectée à son extrémité à un pendule vibrant de fréquence correspondante – une hypothèse qui s'appuie sur des découvertes anatomiques récentes –, il apparaît que l'acoustique physiologique recevra bientôt un habillement mathématique exactement aussi rigoureux que l'optique[5].

4. Je traduis *combination tones* par « sons résultants », qui est l'expression utilisée par le traducteur français du livre de Helmholtz pour l'allemand *Combinationstöne* (littéralement, « sons de combinaison »).
5. Cité par Stephen Vogel, « Sensation of tone, perception of sound, and empiricism », in *Hermann von Helmholtz and the Foundations of Nineteenth-Century Science*, édition établie par David Cahan, Berkeley/Los Angeles/Londres, University of California Press, 1993, p. 267-268.

En d'autres termes, Helmholtz explique qu'il a trouvé le moyen de faire reposer toute la théorie de l'harmonie sur une base qui n'a plus de rien de spéculatif et est devenue enfin scientifique, qu'il est en train de se procurer les instruments nécessaires pour une étude mathématique et expérimentale des timbres – qui pourrait sembler à première vue impossible parce que, à la différence de la hauteur et de l'intensité, le timbre donne l'impression de représenter, dans le son, un élément purement qualitatif à peu près inanalysable –, et enfin qu'il est en mesure de proposer une hypothèse explicative susceptible de conférer à l'acoustique physiologique le statut de science à part entière et de justifier la place fondamentale qui doit lui être octroyée dans la théorie de la musique.

En ce qui concerne le timbre, le résultat essentiel auquel va aboutir Helmholtz constitue une précision importante apportée à l'idée communément admise à l'époque qu'alors que la hauteur du son dépend de la fréquence de la vibration et l'intensité de son amplitude, le timbre dépend de la forme de la vibration. On peut vérifier que le timbre dépend effectivement de la forme de la vibration, mais seulement dans la mesure où celle-ci détermine également le nombre et la nature des sons accessoires qui entrent dans la composition du son global. Des ondes de forme très différente peuvent, en fait, très bien donner le même timbre, à la condition que les vibrations de l'air qui arrivent à l'oreille présentent le même nombre de vibrations pendulaires de la même intensité. L'oreille ne distingue justement pas les diverses formes d'onde comme l'œil est capable de le faire pour les représentations des diverses formes de vibrations, quand celles-ci sont rendues visibles : les différences des timbres musicaux perçus ne dépendent que du nombre et de l'intensité des sons partiels que comporte le son et que l'oreille sépare par l'analyse du son, autrement dit, du mouvement vibratoire, en ses constituants élémentaires[6]. Sur ce point, l'oreille se comporte donc de façon fondamentalement différente de celle dont le fait l'œil.

Sur le premier problème qu'il mentionne dans le passage cité, Helmholtz, comme il l'explique dans une lettre du 15 juillet 1861 à son éditeur, estime qu'il est désormais possible de connaître les causes physiques et physiologiques de la différence entre l'harmonie et la disharmonie, et que le moment est venu d'abandonner une fois pour toutes les explications habituellement reçues, qui prétendent fournir une base scientifique à l'harmonie mais sont en réalité purement métaphysiques et se réduisent, à ses yeux, à peu près à une forme de verbiage creux. Les explications auxquelles il songe sont, bien entendu, celles qui sont héritées, sous une forme

6. *TPM*, p. 161 *sq*.

ou sous une autre, de la tradition pythagorico-platonicienne et qui soutiennent que l'harmonie musicale reflète quelque chose comme l'harmonie du monde ou du cosmos et que ce qui est expérimenté par l'âme dans la musique est précisément l'ordre et la beauté de l'univers, qui reposent en dernière analyse sur des rapports mathématiques d'une espèce particulièrement simple. On n'a malheureusement pas jusqu'à présent, estime Helmholtz, réussi à donner ne serait-ce qu'un commencement de réponse à la question de savoir ce que les accords musicaux ont à voir au juste avec les six premiers nombres entiers :

> Cette relation entre les nombres entiers et les consonances musicales a été considérée de tout temps comme un admirable et important mystère. Déjà les pythagoriciens la rangeaient dans leurs spéculations sur l'harmonie des sphères. Elle resta depuis lors tantôt la fin, tantôt le point de départ de suppositions singulières et hardies, fantastiques ou philosophiques, jusqu'aux temps modernes, où les savants adoptèrent, pour la plupart, l'opinion, déjà émise par Euler, que l'âme humaine trouvait un bien-être particulier dans les rapports simples, parce qu'elle pouvait plus facilement les embrasser et les saisir. Mais il restait à examiner, comment l'âme d'un auditeur entièrement étranger à la physique, et qui ne s'est peut-être jamais rendu compte que les sons proviennent de vibrations, peut arriver à reconnaître et à comparer les rapports des nombres de vibrations. Déterminer les phénomènes qui rendent sensible à l'oreille la différence entre les consonances et les dissonances, sera une des questions principales de la seconde partie de ce livre[7].

On ne peut donc plus se contenter de caractériser de façon scientifique la différence qui existe entre les consonances et les dissonances, il faut essayer également de répondre de façon scientifique, ou en tout cas plus scientifique, à la question de savoir de quelle manière la différence peut être sentie par l'oreille. Un philosophe comme Leibniz, par exemple, s'est déjà approché sensiblement de la question et même de ce qui constitue, aux yeux de Helmholtz, le principe de la réponse, quand il a remarqué que

> Les coups sur le tambour, le rythme et la cadence dans les danses et les autres mouvements de cette sorte qui s'effectuent selon la mesure et la règle tirent leur agrément de l'ordre, car tout ordre est bénéfique à l'âme, et un ordre régulier, bien qu'invisible, se rencontre également dans les coups ou les mouvements provoqués avec art des cordes, tuyaux ou cloches tremblants ou vibrants, et même de l'air, qui est mis par là dans une agitation régulière, laquelle produit alors encore en plus une résonance

7. *TPM*, p. 21.

> accordée avec elle (*einen mitstimmenden Widerschall*) en nous-mêmes par l'intermédiaire de l'oreille, d'après laquelle nos esprits vitaux sont agités eux aussi. C'est pourquoi la musique est si appropriée pour mouvoir les esprits, bien qu'au total un tel but principal ne soit pas suffisamment observé ni cherché[8].

Mais s'il est entendu que la consonance est agréable à l'esprit parce que l'ordre et la régularité le sont dans tous les cas, il reste évidemment encore à expliquer de façon plus précise en quoi consistent, en l'occurrence, l'ordre et la régularité concernés (et leurs contraires : le désordre et l'irrégularité) et par quel genre de processus s'effectue la perception (inconsciente) que, selon Leibniz, nous avons d'eux dans la musique.

À cette dernière question, « les musiciens, constate Helmholtz, aussi bien que les philosophes et les physiciens, se sont la plupart du temps bornés à répondre que l'âme humaine, par un mécanisme quelconque, inconnu de nous, avait la faculté d'apprécier les rapports numériques des vibrations sonores, et qu'elle éprouvait un plaisir particulier à trouver devant elle des rapports simples et facilement perceptibles[9] ». Il n'est pas contestable que des progrès importants ont été réalisés dans le traitement des questions qui relèvent de la psychologie et de l'esthétique. Mais il manque toujours ce qui constitue, aux yeux de Helmholtz, « le vrai point de départ, le principe fondamental, c'est-à-dire, la base scientifique des règles élémentaires pour la construction de la gamme, des accords, des modes, et généralement de tout ce qui est ordinairement compris dans ce qu'on appelle l'harmonie (*Generalbass*). Dans ces domaines élémentaires, nous avons affaire non seulement aux libres inventions de l'art, mais encore à une aveugle et inflexible loi de la nature, aux activités physiologiques de la sensation[10] ».

Helmholtz a le sentiment d'arriver, par conséquent, à un moment où le genre de réponse dont on s'était satisfait pendant longtemps ne peut plus être considéré comme acceptable, et il va s'efforcer de remplacer une apparence de fondement par ce qu'il croit être un fondement réel, sans toutefois se rendre compte clairement que le fondement nouveau ne présentera pas nécessairement les mêmes garanties de stabilité que l'ancien et, s'il consolide l'édifice, peut aussi, en un autre sens, le fragiliser et le rendre plus vulnérable et moins capable de résister au temps et au changement.

---

8. G. W. Leibniz, « Von der Weisheit », in *Gottfried Wilhelm Leibniz*, Auswahl und *Einleitung* von Friedrich Heer, Francfort-sur-le-Main et Hamburg, Fischer Bücherei, 1958, p. 204.
9. *TPM,* p. 2.
10. *TPM*, p. 3.

Comme l'explique Carl Dallhaus, dans son livre sur la théorie de la musique aux XVII^e^ et XVIII^e^ siècles :

> La règle de composition musicale selon laquelle la résolution d'une consonance imparfaite en une consonance parfaite est un but du déroulement musical – un point de repos auquel il tend – était, en tant que norme de *musica pratica* ou *poetica*, fondée dans des présuppositions ontologiques qui constituaient l'objet de la *musica theoretica* ou *speculativa*. L'idée, qui s'était transmise depuis longtemps, que la musique artificielle était de la mathématique sonnante (*tönende Mathematik*) énonce dans la *musica theoretica*, considérée comme une discipline servant de fondement à la *musica poetica*, que la simplicité mathématique de l'octave (*do-do'* = 2/1) par rapport à la sixte majeure plus compliquée (8/5) est la cause du phénomène, fondamental pour la technique de construction musicale du Moyen Âge tardif et des débuts des Temps modernes, d'une tendance […] contraignante de la sixte imparfaite à l'octave parfaite. Le principe porteur de la progression musicale dans la phrase polyphonique se révèle comme l'empreinte évidente d'une structure profonde mathématique latente. […] Tous les présupposés par lesquels avait été portée la *musica theoretica* comme prémisse ou implication de la *musica poetica* ont été détruits aux XVII^e^ et XVIII^e^ siècles, sans que, il est vrai, soit parvenu avec une clarté suffisante à la conscience des musiciens l'ébranlement du sol sur lequel ils continuaient toujours à croire qu'ils se tenaient. En premier lieu, la cadence, qui part d'une dominante pour arriver à un accord de tonique à trois notes, n'est plus mathématiquement justifiable comme différence entre le plus compliqué et le plus simple, parce que la structure de l'un des accords concorde avec celle de l'autre. En deuxième lieu, l'interprétation platonico-pythagoricienne du nombre comme cause ou principe actif des phénomènes musicaux a été abandonnée dans la philosophie des XVII^e^ et XVIII^e^ siècles, inspirée par la science de la nature moderne, tout comme la catégorie de la *causa finalis*, qui avait expliqué la tendance de l'imparfait à la perfection[11].

Helmholtz conteste vigoureusement les présupposés ontologiques sur lesquels est censée s'appuyer l'idée que la consonance constitue le but du déroulement musical. Mais il ne remet certainement pas en question l'idée elle-même. Les dissonances constituent dans la musique « soit un moyen de contraste pour renforcer l'impression des consonances, soit un moyen d'expression, et cela non seulement pour certains mouvements particuliers de l'âme ; elles servent aussi, d'une manière tout à fait géné-

11. Carl Dahlhaus, *Die Musiktheorie im 17. und 18. Jahrhundert*, Zweiter Teil : Deutschland [Geschichte der Musik, Bd. 11], Darmstadt, 1989, p. 131, cité par Marcus Rieger, *op. cit.*, p. 120, note 98.

rale, à renforcer l'expression de l'activité et de la marche en avant du mouvement musical, parce que l'oreille, tourmentée par les dissonances, aspire de nouveau au calme, au flux pur et régulier des sons formant consonance. Dans ce dernier sens, elles trouvent, surtout immédiatement avant la fin du morceau, à remplir un rôle d'une nature importante[12] ». Si le but n'est plus d'accéder à nouveau à la perfection, au sens ontologique du terme, après un passage effectué par l'imparfait, il s'agit néanmoins toujours de retrouver un état de calme et de repos, après une phase d'agitation dont la raison d'être principale réside dans le fait qu'elle exprime l'effort énergique qui est fait par l'âme pour revenir au premier. On peut dire, au total, que Helmholtz cherche à rebâtir une construction ébranlée dans ses fondations sur un sol qu'il croit plus ferme ; mais il est vrai également qu'il poursuit et achève à son insu un processus qui va conduire à quelque chose de bien différent de ce qui constituait pour lui la musique et qu'il cherchait encore à fonder.

Physiologie et psychologie des phénomènes sonores en général et de la musique en particulier

Puisque j'ai évoqué la question de savoir pourquoi la réponse à la question du fondement doit être cherchée, selon Helmholtz, dans la physiologie, plutôt que dans la psychologie, il me faut dire un mot de la façon surprenante dont a été traduit le titre de son livre dans la version française. Le titre allemand est, comme je l'ai indiqué, *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik* et il est devenu, dans la traduction française, *Théorie physiologique de la musique*. C'est d'autant plus étonnant que la traduction a été revue et approuvée par l'auteur, à la place duquel je n'aurais sûrement pas laissé passer un titre aussi contestable. Le titre français renverse, en effet, complètement la perspective adoptée par Helmholtz. Il ne se proposait, en réalité, nullement de construire une théorie physiologique de la musique ; il cherchait, ce qui est bien différent, à construire une théorie des sensations sonores susceptible de fournir un fondement approprié – lequel se trouve, en l'occurrence, être physiologique – à la théorie de la musique. Il n'est donc pas question chez lui d'une théorie physiologique de la musique mais seulement d'un fondement physiologique à apporter à une théo-

12. *Théorie physiologique de la musique*, p. 436

rie dont les concepts et les méthodes peuvent être bien différents de ceux de la physiologie.

Mais sur quoi repose au juste la position fondamentale et fondationnelle qui doit être reconnue ici à la physiologie ? Helmholtz donne une idée très claire de ce que doit être la réponse quand il écrit :

> Les relations entre la physiologie du sens de l'ouïe et la théorie de la musique sont particulièrement frappantes et claires, parce que les formes élémentaires de la mise en forme (*Gestaltung*) musicale dépendent de façon beaucoup plus pure de la nature et des particularités de nos sensations que ça n'est le cas avec les autres arts, dans lesquels l'espèce du matériau à utiliser et des objets à représenter fait sentir son influence de façon beaucoup plus forte[13].

Au début de la *Théorie physiologique de la musique*, Helmholtz souligne que « la musique se rattache à la sensation pure et simple par des liens bien plus étroits que les autres arts, qui ont affaire plutôt aux perceptions nous venant des sens, c'est-à-dire aux notions sur les objets extérieurs, que nous tirons des sensations par des procédés psychiques[14] ». Il est important de se souvenir ici que, pour Helmholtz, les perceptions sont des inférences inconscientes effectuées à partir de prémisses constituées par les sensations, qui sont elles aussi pour l'essentiel inconscientes, et que, pour lui, le trajet qui est effectué de la présence des sensations à la perception d'un objet ne relève pas de la physiologie mais de la psychologie. Or il se trouve que, dans la musique, la sensation cesse d'être traitée et évaluée en fonction de la contribution qu'elle est susceptible d'apporter à la connaissance de l'objet ; elle est considérée en elle-même et pour elle-même, ce qui, aux yeux de Helmholtz, constitue une différence significative par rapport à la façon dont les choses se passent, par exemple, dans la peinture :

> Dans la peinture, la couleur est le seul élément qui s'adresse immédiatement à la sensation, sans l'intermédiaire d'un acte de l'entendement. Dans la musique, au contraire, les sensations auditives sont précisément ce qui forme la matière de l'art ; nous ne transformons point ces sensations, au moins dans les limites où elles appartiennent à la musique, en symboles d'objets ou de phénomènes extérieurs. En d'autres termes, bien que, dans un concert, il nous arrive de distinguer certains sons produits, les uns par le violon, d'autres par la clarinette, la jouissance artistique ne réside pas

13. Helmholtz, cité dans Leo Königsberger, *Hermann von Helmholtz*, Braunschweig, Friedrich Vieweg und Sohn, 1902-1903, Bd. 2, p. 27
14. *TPM,* p. 3

> dans la représentation que nous nous faisons de l'existence matérielle du violon ou de la clarinette, mais dans la sensation des sons qui en émanent. […] En ce sens, il est évident que la musique a, avec la sensation proprement dite, des liens plus immédiats qu'aucun des autres arts ; il s'ensuit que la théorie des sensations auditives sera appelée à jouer, dans l'esthétique musicale, un rôle beaucoup plus essentiel que la théorie de l'éclairement ou de la perspective dans la peinture. Ces deux dernières sciences sont sans doute utiles au peintre pour atteindre autant que possible à une fidèle représentation de la nature, mais elles n'ont rien de commun avec l'effet artistique de l'œuvre. Dans la musique, au contraire, il ne s'agit pas d'arriver à la fidèle représentation de la nature ; les sons et les sensations correspondantes sont là pour eux-mêmes, et agissent tout à fait indépendamment de leur rapport avec un objet extérieur quelconque[15].

Nous avons ainsi obtenu la réponse à la question posée. L'étude des phénomènes qui se produisent dans les organes des sens comporte, selon Helmholtz, trois parties distinctes. La première s'occupe de la manière dont l'agent extérieur (la lumière pour l'œil, le son pour l'oreille) pénètre jusqu'aux nerfs. Cette partie peut être appelée la partie physique de l'étude physiologique – dans le cas du son, l'*acoustique physique*. La deuxième partie s'occupe des différentes espèces d'excitations nerveuses correspondant aux diverses sensations, ce qui constitue, pour le son, l'objet de l'*acoustique physiologique*. Enfin une troisième partie, qui correspond, dans le cas qui nous intéresse, à ce que l'on peut appeler l'*acoustique psychologique*, s'occupe de la manière dont les sensations sont transformées en représentations d'objets extérieurs. Or, comme on l'a vu, dans la musique on en reste normalement au deuxième stade et on ne va pas jusqu'au troisième, qui est celui de la perception et dont traite la psychologie.

Leibniz dit de la musique qu'elle est une sorte d'arithmétique cachée ou inconsciente à laquelle se livre une âme qui ne se rend pas compte qu'elle est en train de compter (« *Musica est exercitium arithmeticae occultum nescientis se numerare animi*[16] *»)* et il utilise ce fait comme un argument en faveur de l'existence de perceptions dont nous ne sommes pas conscients. La musique peut donc être comprise comme consistant dans la perception d'un ordre et d'une régularité cachés, et, si on la considère

15. *TPM,* p. 4.
16. G. W. Leibniz, « Lettre à Christian Golbach, 17 avril 1712 », *in* Leibniz, *Epistolae ad diversos*, publiées par Christian Korholt, Leipzig, Christian Breitkopf, 1734, Band 1, p. 240-241. Pour plus de détails sur cette question, voir Patrice Bailhache, « La musique, une pratique cachée de l'arithmétique ? », *in L'Actualité de Leibniz : les deux labyrinthes*, Décade de Cerisy-la-Salle (15-22 juin 1995), Stuttgart, Franz Steiner Verlag, 1999, p. 405-419.

de cette façon, on en conclura assez logiquement qu'elle est au fond le plus abstrait et le plus intellectuel de tous les arts. « Les plaisirs des sens qui approchent le plus des plaisirs de l'esprit <, et sont les plus purs et les plus sûrs>, sont, écrit Leibniz, ceux de la musique, et ceux de la symétrie, les uns des oreilles, les autres des yeux, car il est aisé de comprendre les raisons de l'harmonie ou de cette perfection qui nous y donne du plaisir. La seule chose qu'on y peut craindre, c'est d'y employer trop de temps[17]. » Ce qu'il faut dire, du point de vue de Helmholtz, semble être plutôt, au contraire, que la musique est le plus sensible de tous les arts, puisqu'on y a affaire uniquement à la sensation pure et simple, considérée indépendamment de la relation à un objet quelconque. Est-ce à dire que l'on peut laisser de côté la psychologie, qui n'a plus aucune fonction à remplir en l'occurrence, et que l'on peut également ignorer l'aspect intellectuel évident que semble comporter aussi le processus ? La réponse est dans les deux cas un non catégorique.

En ce qui concerne la première question, Helmholtz est bien le dernier à pouvoir être soupçonné de chercher à éliminer la psychologie au profit de la physiologie ou d'essayer de réduire la première à la seconde. Il a au contraire toujours défendu avec une vigueur particulière l'idée de l'autonomie et de la spécificité de la psychologie par rapport à la physiologie ; et cela lui a valu, du reste, d'être critiqué violemment par Hering, qui l'a accusé de chercher ainsi à défendre hypocritement une forme de dualisme et même de spiritualisme qui ne dit pas son nom. Ce qui est vrai, dans le cas de la musique, est que la psychologie n'intervient pas sous la forme de la transformation des sensations auditives en représentations d'objets ; mais cela ne l'empêche en aucune manière de continuer à jouer, sous une autre forme, un rôle essentiel. Rien ne montre plus clairement ce que peuvent être sur ce point les convictions de Helmholtz que le passage suivant, tiré de l'article « Sur les causes physiologiques de l'harmonie » :

> L'onde composée reste, en progressant, telle qu'elle est, et là où elle atteint l'oreille, personne ne peut voir en la regardant si elle est provenue sous cette forme d'un instrument musical, ou si elle s'est composée en cours de route à partir de deux ou plusieurs trains d'ondes.
> Que fait alors l'oreille ? L'analyse-t-elle ou l'appréhende-t-elle comme un tout ? – La réponse à cela peut, selon le sens qui est donné à la question, se révéler être différente, car nous devons distinguer ici deux choses, à savoir premièrement la *sensation* dans le nerf auditif, telle qu'elle se développe sans immixtion d'une activité intellectuelle, et la *représentation* que

17. G. W. Leibniz, *Textes inédits d'après les manuscrits de la Bibliothèque provinciale de Hanovre*, publiés et annotés par Gaston Grua, Paris, PUF, 1948, tome II, p. 580.

> nous nous formons en conséquence de cette sensation. Nous devons donc, en quelque sorte, distinguer l'oreille corporelle (*das leibliche Ohr*) du corps et l'oreille mentale (*das geistige Ohr*) de la faculté de représentation. L'oreille corporelle fait toujours exactement la même chose, ce que le mathématicien fait à l'aide du théorème de Fourier, et ce que fait le piano avec une masse sonore composée : elle analyse les formes d'onde qui ne correspondent pas déjà originairement, comme les sons du diapason, aux formes d'onde simples en une somme d'ondes simples, et ressent individuellement le son correspondant à chacune des ondes simples, que l'onde soit provenue originairement telle quelle de la source sonore ou se soit composée seulement en cours de route[18].

On peut dire, par conséquent, que l'oreille psychique, s'il est permis de l'appeler ainsi, conserve bel et bien toute sa place et son importance, qui sont déterminantes, à côté de l'oreille physiologique. Et la même chose est vraie du rôle de l'intellect, comparé à celui de la sensation proprement dite. L'harmonie sensible, dont le livre de Helmholtz a pour but de révéler les fondements physiologiques, n'est, précise-t-il, que le premier et le plus bas degré du beau musical :

> Ces phénomènes du son harmonieux purement sensible ne sont, il est vrai, que le degré le plus bas du beau musical. Pour la beauté supérieure, intellectuelle, de la musique l'harmonie et la disharmonie ne sont que des moyens, mais des moyens essentiels et puissants. Dans la disharmonie, le nerf auditif se sent tourmenté par les à-coups de sons incompatibles, il aspire à l'écoulement pur des sons dans l'harmonie, et y tend pour y demeurer dans un état d'apaisement. Ainsi, tous les deux propulsent et calment alternativement le flux des sons, dans le mouvement incorporel duquel l'esprit contemple une image du cours de ses représentations et états d'âme. De même que devant la mer agitée par la houle, il est fasciné ici par le mode du mouvement, qui se répète de façon rythmique et change pourtant constamment, et il le perpétue en lui. Mais, alors que là seules des forces naturelles règnent de façon aveugle et que dans l'état d'esprit du spectateur domine malgré tout finalement, pour cette raison, la sensation de réalité désertique, dans l'œuvre d'art musicale le mouvement suit les courants de l'âme excitée de l'artiste. Tantôt coulant doucement, tantôt bondissant avec grâce, tantôt agité fortement, sous l'emprise subite des sons naturels de la passion ou grâce à un travail considérable, le flux des sons transmet dans leur vivacité originaire des états d'âme insoupçonnés que l'artiste a surpris dans son âme à l'âme de l'auditeur, pour l'élever enfin à la paix de la beauté éternelle, dont seul un petit

18. Helmholtz, « Über die physiologischen Ursachen der musikalischen Harmonie » (1857), *in* Hermann von Helmholtz, *Vorträge und Reden*, fünfte Ausgabe, Friedrich Vieweg und Sohn, 1903, Band 1, p. 78-79.

> nombre de ses préférés qu'elle a élus ont été consacrés par la divinité comme les annonciateurs.
> Mais là se trouvent les limites de l'étude de la nature et elles m'ordonnent de m'arrêter[19].

Par conséquent, si on ne peut pas dire de la musique qu'elle reflète certaines des propriétés les plus profondes de la réalité extérieure, on peut, en revanche, tout à fait dire qu'elle constitue néanmoins l'image ou le reflet de quelque chose qui est d'une importance cruciale, au moins pour nous, à savoir les mouvements de l'âme. Il arrive même à Helmholtz de suggérer, de façon très métaphorique, que, dans la musique, les vibrations du son se trouvent en quelque sorte en accord avec les vibrations de l'âme. Tout dans la musique dépend, comme on l'a vu, initialement de la sensation, et la dernière chose à laquelle on pourrait s'attendre de la part de la sensation est qu'elle nous mette en contact avec certaines des caractéristiques les plus fondamentales du monde extérieur. Les sensations ne sont, en effet, en aucun cas des reproductions ou des images des objets dont elles proviennent. Elles ne constituent, selon Helmholtz, rien de plus que des signes arbitraires qui nous permettent de reconnaître, avec une sûreté suffisante, les identités et les différences entre les objets, sans toutefois être en mesure de nous renseigner de façon fiable sur la nature exacte de ceux-ci. Dans la musique, cette fonction, qui peut être qualifiée d'« utilitaire » et qui rend possible une adaptation convenable à la réalité extérieure, n'entre plus en ligne de compte. Mais c'est justement ce qui, d'une certaine façon, rend la sensation disponible pour une autre tâche, qui consiste à représenter les actions de l'âme. Dans la *Théorie physiologique de la musique*, Helmholtz cite, sur ce point, avec approbation Aristote :

> Aristote a déjà défini l'action de la musique d'une manière tout à fait semblable. Dans le vingt-neuvième problème[20], il pose la question suivante : « Pourquoi les rythmes et les mélodies, c'est-à-dire les sons, se prêtent-ils à exprimer les mouvements de l'âme, tandis qu'il n'en est pas de même des goûts, des couleurs et des parfums ? Serait-ce parce que ce sont des mouvements comme les gestes ? L'énergie particulière aux mélodies et aux rythmes provient d'une disposition de l'âme et agit sur elle. Les goûts et les couleurs, au contraire, n'y parviennent pas au même degré. » Aristote

19. *Ibid.*, p. 91 ; *cf.* aussi *TPM,* p. 330-331.
20. Il s'agit en réalité du vingt-septième problème du Livre XXIX [ΟΣΑ ΡΕΠΙΑΡΜΟΝΙΑΝ (Problèmes concernant l'harmonie)], qui est cité dans une traduction pour le moins un peu libre.

> dit encore, à la fin du vingt-septième problème : « Ces mouvements (ceux du rythme et de la mélodie) sont énergiques ; ce sont des phénomènes qui représentent l'état de l'âme [21]. »

Helmholtz est, il faut le rappeler, un ennemi farouche de toutes les théories qui postulent quelque chose comme une relation d'identité ou même simplement de proportionnalité qui serait donnée au départ entre la nature et l'esprit ; et cela suffit évidemment à rendre, à ses yeux, peu plausible et même peu compréhensible la théorie pythagoricienne. Mais il ne suffirait pas non plus de renoncer à attribuer à l'expérience musicale une fonction cognitive et une portée ontologique comme celles dont il est question dans ce genre de théorie et de se contenter de dire, comme le faisait encore Chladni, un des prédécesseurs immédiats de Helmholtz, qu'un intervalle est consonant quand les nombres de vibrations sont dans un rapport tellement simple que l'âme l'appréhende avec la plus grande facilité et en tire une sensation d'apaisement. En 1744, Sorge avait proposé, pour l'existence des sons résultants qu'il avait découverts quelques années plus tôt (en 1740), une explication qui illustre assez bien ce que Helmholtz veut dire quand il parle de verbiage creux :

> Comment se fait-il que dans l'accord d'une quinte 2-3 s'annonce et se fasse entendre encore en plus, en une subtile résonance concomitante, le troisième son, et à chaque fois une octave plus bas que le son grave de la quinte ? La nature joue ici son jeu charmant et montre que dans 2-3 le 1 manque encore, et qu'elle aimerait bien avoir aussi, en l'occurrence, ce son, afin que l'ordre 1-2-3, par exemple do_1 do_2 sol_2, soit parfait ; de là vient également qu'une quinte de 3 pieds rend le son si parfait et porte avec elle un troisième son qui est presque aussi fort qu'un tuyau bouché doux (*ein gelindes Gedackt*)[22].

---

21. *TPM,* p. 330-331 ; *cf.* Aristotle, *Problems*, Books 1-21, dans une traduction anglaise de W. S. Hett, Cambridge (Mass.) & Londres, Harvard University Press, 1936, Book XIX, Problem 27, p. 394-395. La traduction allemande de la dernière phrase citée d'Aristote (ou, en tout cas, du texte des *Problèmes*) est : « *Thaten aber sind Zeichen der Gemütsbestimmung* » (*Die Lehre von den Tonsempfindungen*, troisième édition revue, Braunschweig, Friedrich Vieweg und Sohn, 1870, p. 398). Ce que dit l'original est :
αἱ δὲ κινήσεις αὗται πρακτιαί εἰσιν, αἱ δὲ πράξεις ἤθους σημασία ἐστίν
Ce que le traducteur anglais a rendu par : « *But the movements with which we are dealing are connected with action, and actions are symptoms of moral character.* » (« Les mouvements auxquels nous avons affaire sont reliés à l'action, et les actions sont les symptômes du caractère moral. »)

22. Georg Andreas Sorge, *Anweisung zur Stimmung und Temperatur der Orgelwerke, als auch andrer Instrumente, sonderlich aber des Klaviers* (1744), cité dans Marcus Rieger, *op. cit.*, p. 112-113.

La situation que décrit Sorge est celle qui est représentée sur la deuxième mesure de la portée dans le tableau que Helmholtz donne des premiers sons résultants différentiels[23] des intervalles consonants usuels (figure 1).

INTERVALLES.	RAPPORTS DE VIBRATIONS.	DIFFÉRENCES	LE SON RÉSULTANT EST PLUS GRAVE QUE LE PLUS BAS des sons primaires D'UN INTERVALLE ÉGAL à
Octave.....	1 : 2	1	l'unisson.
Quinte.....	2 : 3	1	l'octave.
Quarte.....	3 : 4	1	la douzième.
Tierce majeure....	4 : 5	1	la 2e octave.
Tierce mineure....	5 : 6	1	la 2e octave, plus la tierce majeure.
Sixte majeure.....	3 : 5	2	la quinte.
Sixte mineure.....	5 : 8	3	la sixte majeure.

ou en notation usuelle, en figurant les sons primaires par des blanches et les sons résultants par des noires :

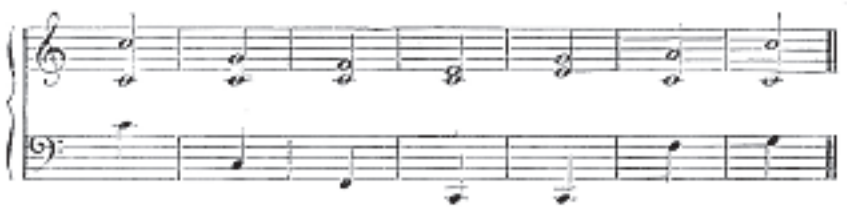


Figure 1 : Tableau donné par Helmholtz des premiers sons différentiels des intervalles consonants usuels[24].

Leibniz avait affirmé que, dans l'espèce d'arithmétique cachée que constitue la musique, l'âme n'a vraisemblablement pas besoin d'aller plus loin que 5. Tous les intervalles musicaux utilisés correspondent, en effet, à des rapports qui peuvent être exprimés à l'aide des seuls nombres 1, 2, 3 et 5 :

23. Il existe, en fait, deux espèces différentes de sons résultants : les sons *différentiels*, qui sont ceux dont parle Sorge, et qui présentent des vibrations dont le nombre est égal à la différence des nombres de vibrations des sons primaires, et les sons *additionnels*, qui ont été découverts par Helmholtz et dont le nombre de vibrations est égal à la somme des nombres de vibrations des sons primaires (*cf.* Helmholtz, *TPM,* p. 192).
24. *TPM,* p. 193.

> Nous, en musique, nous nes comptons pas au-delà de cinq, semblables à ces gens qui même en arithmétique ne progressaient pas au-delà du nombre 3, et chez lesquels se vérifierait ce que disent les Allemands à propos d'un homme simple : *Er kann nicht über drey zehlen.* Car nos intervalles usités sont tous issus de rapports composés à partir de rapports entre deux des nombres premiers 1, 2, 3, 5. Si nous étions dotés d'un peu plus de subtilité, nous pourrions aller jusqu'au nombre premier sept. Et je pense qu'il y a réellement des gens qui le font. C'est pourquoi les anciens ne récusaient pas non plus complètement le nombre 7. Mais on ne trouvera guère de gens qui aillent jusqu'aux nombres premiers qui sont ses successeurs les plus proches, 11 et 13[25].

Leibniz admettait, comme on le voit, qu'une plus grande finesse de nos sens pourrait nous permettre d'apprécier des proportions musicales différentes, et éventuellement d'aller même peut-être au-delà du nombre 7. Selon Euler, l'usage courant qui est fait d'ores et déjà de l'accord de septième de dominante dans la musique oblige à considérer que l'extension envisagée par Leibniz s'est bel et bien déjà produite : la composition musicale compte désormais au moins jusqu'à sept et l'oreille s'est habituée également à le faire.

> On soutient communément qu'on ne se sert dans la musique que des proportions composées de ces trois nombres premiers 2, 3 et 5, et le grand Leibniz a déjà remarqué que dans la musique on n'a pas encore appris à compter au-delà de 5, ce qui est aussi incontestablement vrai dans les instruments accordés selon les principes de l'harmonie. Mais, si ma conjecture a lieu, on peut dire que dans la composition on compte déjà jusqu'à 7 et que l'oreille y est déjà accoutumée ; c'est un nouveau genre de musique qu'on a commencé à mettre en usage et qui a été inconnu aux anciens. Dans ce genre d'accord, 4, 5, 6, 7 est la plus complète harmonie, puisqu'elle renferme les nombres 2, 3, 5 et 7 ; mais il est aussi plus compliqué que l'accord parfait dans le genre commun qui ne contient que les nombres 2, 3 et 5. Si c'est une perfection dans la composition, on tâchera peut-être de porter les instruments au même degré[26].

Les proportions auxquelles on recourt le plus normalement sont 1/2 (pour l'octave), 2/3 (pour la quinte), 3/4 (pour la quarte), 4/5 (pour la tierce majeure) et 5/6 (pour la tierce mineure). Pour la sixte mineure, il faudrait aller jusqu'à 5/8 (la proportion qui correspond à la sixte majeure est 6/10, autrement dit, 3/5). Tous les rapports concernés sont exprima-

25. G. W. Leibniz, « Lettre à Christian Goldbach », *op. cit.*, p. 240.
26. Cité par Patrice Bailhache, *op. cit.*, p. 419.

bles à l'aide des nombres entiers allant de 1 à 6. Si maintenant on voulait aller jusqu'à la septième mineure, qui intervient dans l'accord de septième de dominante, *sol – si – ré – fa*, il faudrait faire intervenir la proportion 4/7 (ou plus exactement 5/9, la proportion à laquelle correspond l'intervalle *sol-fa*). La facilité relative avec laquelle l'accord de septième de dominante a réussi à se faire accepter dans la musique et même à y jouer pour finir un rôle déterminant s'explique aisément si l'on tient compte du fait que la théorie de la consonance et de la dissonance que propose Helmholtz permet de le faire apparaître comme le moins dissonant de tous les accords dissonants, pour la raison que, de toutes les septièmes, la septième mineure est la plus proche de la septième naturelle, pour laquelle le rapport est de 7/4 et dont Hemholtz dit qu'elle « ne le cède pas en harmonie aux consonances[27] » :

> La prime et l'octave agissent considérablement sur les intervalles voisins, par les consonances qui s'y transforment en dissonances ; ainsi la seconde mineure *ut-ré*$_b$ et la septième majeure *ut-si*, différant chacune d'un demi-ton de la prime ou de l'octave, sont les dissonances les plus dures de nos gammes. La seconde majeure *ut-ré* et la septième mineure *ut-si*$_b$, différant d'un ton entier des intervalles en question, doivent encore être rangées parmi les dissonances, mais, en raison de l'écart plus grand des éléments dissonants, elles sont beaucoup plus douces que les précédentes. Dans les régions supérieures de la gamme surtout, leur dureté diminue beaucoup à cause du grand nombre de leurs battements. Comme la septième mineure doit sa dissonance au premier harmonique, plus faible que le son fondamental dans la plupart des timbres musicaux, elle présente une dissonance plus douce encore que celle de la seconde majeure, et doit être placée sur la limite des intervalles dissonants[28].

Le livre d'Euler auquel se réfère Helmholtz dans la *Théorie physiologique de la musique*, le *Tentamen novae theoriae musicae*, a été publié en 1739. En 1863, au moment où paraît le sien, Helmholtz constate que l'accord de septième de dominante, pour des raisons que sa théorie permet d'expliquer assez facilement, est devenu l'accord le plus important après l'accord de tonique :

> Aussi est-ce l'accord de septième de dominante qui, dans la musique moderne, joue le principal rôle après l'accord tonique. Il précise la tonalité, plus que le simple accord de dominante *sol – si – ré*, plus exactement

27. *TMP.* p. 441.
28. *TMP*, p. 243.

> que l'accord diminué *si – ré – fa*[29]. *Comme accord dissonant, il tend à se résoudre sur l'accord de tonique, ce que ne fait pas le simple accord de dominante. À cela, enfin, il faut encore ajouter que l'harmonie s'y trouve extraordinairement peu troublée, en sorte que c'est le plus doux des accords dissonants. Aussi, dans la musique moderne pourrions-nous à peine nous en passer. Il a été trouvé, à ce qu'il paraît, par Monteverdi, au commencement du dix-septième siècle*[30].

L'explication du rôle privilégié que joue cet accord peut être trouvée dans le fait que :

> L'accord de septième de dominante, *sol – si – ré – fa*, contient trois notes appartenant au son complexe *sol*, savoir *sol*, *si* et *ré*, tandis que la septième *fa* est la note dissonante. Il faut remarquer, cependant, que cette septième mineure *sol – fa* se rapproche déjà suffisamment du rapport 7/4, qui serait exprimé presque exactement par l'intervalle *sol – fa*, pour que la note *fa* puisse presque jouer le rôle du septième son partiel de *sol*. Ce son complexe serait plus exactement représenté par *sol – si – ré – fa*. Les chanteurs changent aussi très facilement le *fa* de l'accord de septième de dominante *fa*, soit qu'en général cette note descende sur le *mi*, soit que l'accord obtenu par cette modification ait plus de douceur[31].

Si l'on en croit Leibniz, « la raison de la consonance doit être cherchée dans la congruence des coups (*congruentia ictuum*)[32] » : quand deux cordes, par exemple, vibrent à l'octave l'une de l'autre, un coup sur deux de l'une des séries de coups coïncide avec chacun des coups de l'autre série ; dans la quinte, chaque troisième coup de l'une des séries coïncide avec chaque deuxième coup de l'autre série, etc. Cela exclut, en principe, non seulement les proportions sourdes, qui ne donnent pas lieu à des conjonctions régulières arithmétiquement exactes, mais également les cas dans lesquels la multitude des coups avant la conjonction est excessive, et ceux dans lesquels, du fait de la trop grande complexité du son, le nombre des comparaisons à effectuer est trop grand. Il faut donc, estime Leibniz, exclure de la constitution des consonances tous les nombres supérieurs à 8 et, parmi les nombres plus petits, le nombre 7, qui (pour nous et pour le moment en tout cas) semble représenter une limite difficile à atteindre et plus encore à dépasser. Néanmoins, même les propor-

---

29. Dans la notation adoptée par Helmholtz, les notes surlignées correspondent à des sons plus élevés de 81/80 (c'est-à-dire, d'un *comma*) que les notes utilisées normalement et les notes soulignées à des sons plus graves de la même quantité (*cf. TMP*, p. 366).
30. *TMP*, p. 454
31. *Ibid.*
32. Leibniz, « Lettre à Christian Goldbach », *op. cit.*, p. 241

tions sourdes, qui en elles-mêmes déplaisent à l'âme, sauf quand elles se rapprochent suffisamment de proportions rationnelles, peuvent, dans certaines conditions être trouvées plaisantes par accident : « Par accident, cependant, les dissonances plaisent de temps à autre, et elles sont employées utilement, et elles sont interposées parmi les suavités comme les ombres au sein de l'ordre et de la lumière, pour que nous soyons d'autant plus charmés ensuite par l'ordre[33]. » Comme on le verra, cette idée que les dissonances ont assurément leur place dans la musique mais ne peuvent être agréables que par accident et comme un moyen d'augmenter encore le plaisir suscité par la consonance, n'est pas remise en question par Helmholtz et est même, à ses yeux, confirmée entièrement, par sa théorie.

Pour lui, malgré tout, rien n'a encore été expliqué réellement jusqu'à présent. On ne sait toujours pas, en effet, sur quoi repose au juste la satisfaction particulière que l'âme ou plus exactement l'oreille, car c'est, bien entendu, avec elle que tout commence, éprouve à expérimenter des rapports arithmétiques privilégiés comme ceux dont il s'agit. Ce qui est insatisfaisant dans l'hypothèse d'Euler est, dit Helmholtz, qui prend soin de rendre un hommage appuyé à son prédécesseur mathématicien, qu'« elle ne dit pas comment fait l'âme, pour arriver à percevoir le rapport numérique de deux sons simultanés. Il restait donc, explique-t-il, à indiquer les moyens par lesquels les rapports des nombres de vibrations pénètrent dans la sensation, deviennent sensibles[34] ». Avant de parler du rapport que l'âme perçoit entre les nombres de vibrations, la moindre des choses serait évidemment de commencer par expliquer comment elle s'y prend pour calculer d'abord les nombres de vibrations eux-mêmes et déterminer leurs rapports.

La résolution de l'« énigme de Pythagore »

Bien entendu, le point de vue de Leibniz et d'Euler n'est pas le point de vue pythagoricien au sens strict, qui repose sur l'idée que « tout dans l'univers est nombre et harmonie » et que, dans la musique, ce que les sens perçoivent est quelque chose comme la musique de l'univers lui-même. Helmholtz ne se trompe pas sur ce point et dit d'Euler[35] qu'il a

33. Leibniz, *ibid*.
34. *TPM*, p. 298.
35. *TPM*, p. 295-296.

essayé de faire reposer sur des considérations psychologiques la relation qui avait été établie entre les consonances et les nombres entiers, ce qui correspond probablement à ce que les savants du siècle précédent, étant donné l'état des connaissances de l'époque et l'ignorance presque complète dans laquelle on était encore de ce qui se passe du point de vue physiologique, pouvaient faire de meilleur. Helmholtz suggère poliment qu'il s'agissait au fond simplement, pour lui, de compléter de façon appropriée ce qu'avaient dit Euler et avant lui tous ceux qui ont considéré la musique comme une sorte de « mathématique sonore ». Mais il est évidemment tout à fait conscient du fait que le changement qu'il propose est en réalité bien plus radical que cela. Pour résoudre l'énigme de Pythagore, il a fallu abandonner le mathématisme qui caractérise encore l'approche d'Euler et déplacer la question des mathématiques vers la physique et la physiologie :

> Des phénomènes physiologiques qui rendent sensible la différence entre la consonance et la dissonance, ou, suivant Euler, entre le rapport ordonné et non ordonné des sons, il résulte […], en définitive, une différence essentielle entre notre système d'explication et celui d'Euler. Selon ce dernier, l'âme doit percevoir directement les rapports rationnels des vibrations sonores ; selon nous, elle perçoit seulement un effet dû à ces rapports : la sensation continue ou intermittente des nerfs de l'audition. Le physicien sait que la sensation d'une consonance est continue, parce que les rapports numériques des vibrations sont rationnels ; mais un morceau de musique ne porte rien de semblable à la connaissance d'un auditeur étranger à la physique, et, pour le physicien lui-même, un accord n'est pas rendu plus harmonieux par cette vue plus exacte des choses. Il en est tout autrement de l'ordre qui règne dans le rythme. Avec quelque attention et sans instruction préliminaire, tout le monde peut apprécier qu'une ronde vaut exactement deux blanches ou quatre noires. Le rapport rationnel des vibrations de deux sons simultanés, au contraire, produit sur l'oreille une action particulière qui le distingue des rapports irrationnels, mais cette distinction, entre la consonance et la dissonance, repose sur des phénomènes physiques et non psychologiques[36].

En d'autres termes, la musique nous dit certainement quelque chose et quelque chose d'important ; et elle le fait à l'aide de sons dont les nombres de vibrations sont dans des rapports déterminés. Mais elle ne nous dit pas que ces rapports sont de telle ou telle nature, et en particulier qu'ils sont rationnels ou irrationnels. Ce que Helmholtz objecte à Euler

36. *TPM*, p. 299

est que l'oreille physiologique, qui se comporte comme un analyseur de sons, c'est-à-dire de mouvements vibratoires d'une certaine sorte, fait une différence entre deux espèces d'influence physique, résultant respectivement de l'action de rapports rationnels et de celle de rapports irrationnels entre des nombres de vibrations, grâce à un mécanisme sur la nature duquel on peut formuler une hypothèse raisonnable ; mais elle ne nous renseigne aucunement sur l'origine et la nature exactes de cette différence, et la psychologie, sans l'aide de la physiologie, est, bien entendu encore plus incapable de le faire.

La résolution de l'énigme de Pythagore ne pouvait donc être obtenue qu'en empruntant le chemin qui a été suivi dans la *Théorie physiologique de la musique* :

> La solution de l'énigme que Pythagore a posée il y a 2 500 ans à la science qui cherche les raisons des choses, à propos de la relation des consonances aux rapports des petits nombres entiers, a été à présent obtenue par le fait que l'oreille analyse les sons composés selon les lois de la vibration par influence en vibrations pendulaires et qu'elle n'appréhende comme son harmonieux qu'une excitation qui dure de façon régulière. Mais l'analyse en sons partiels s'effectue, si on exprime les choses en termes mathématiques, selon la loi qui a été formulée par Fourier, qui enseigne la manière dont toute grandeur constituée de façon quelconque qui varie de manière périodique peut être exprimée par une somme de grandeurs périodiques simples[37].
>
> L'origine des rapports rationnels donnés par Pythagore se trouve en dernière analyse, dans la loi de Fourier, qui, dans un certain sens, peut-être considérée comme le fondement de l'harmonie[38].

Cela étant, Helmholtz constate qu'Euler, en dépit des limitations et des insuffisantes de sa théorie, était déjà arrivé à la même hiérarchie des degrés de consonance que lui, à une différence près : il a attribué à tort aux accords mineurs le même degré de consonance qu'aux accords majeurs parce qu'il n'a pas tenu compte des sons résultants qui, dans leur cas, perturbent davantage l'harmonie. Cette influence perturbatrice plus sensible des sons résultants n'est assurément pas suffisante pour transformer l'accord mineur en une dissonance, mais elle a pour effet de lui conférer un caractère réellement spécial :

---

37. *Die Lehre von den Tonempfindungen*, 4. Aufl., 1877, p. 374 ; *cf. TPM*, p. 294.
38. *TPM*, p. 295.

> Aussi l'accord mineur présente-t-il quelque chose d'étrange, qui n'est pas assez prononcé pour détruire entièrement la sensation de la consonance, mais qui suffit cependant pour donner, à l'harmonie et à la signification musicale de cet accord, quelque chose de voilé, de vague, dont l'auditeur ne sait pas démêler la cause, parce que les faibles sons résultants qui le produisent sont couverts par d'autres sons plus forts, et ne peuvent être distingués que par une oreille exercée. C'est ce qui fait que les accords mineurs sont si propres à exprimer des sentiments vagues, sombres ou austères[39].

En dépit de l'opinion, qui semble prévaloir désormais, que la construction des accords mineurs est aussi logique et leur consonance aussi satisfaisante que celle des accords majeurs, Helmholtz pense que son point de vue a pour lui le sentiment des grands compositeurs :

> À mon avis, l'histoire de la musique, le développement lent et timide du mode mineur aux seizième et dix-septième siècles, l'emploi également timide chez Haendel, presque rare encore chez Mozart, de la cadence mineure, toutes ces circonstances réunies ne peuvent permettre de douter que le sentiment artistique des grands compositeurs ne soit d'accord avec mes conclusions[40].

Je me permettrai d'emprunter à Stephen Vogel la description condensée des progrès spectaculaires que Helmholtz, sur la base des présupposés et des principes dont j'ai essayé de donner une idée, a fait réaliser en quelques années à l'acoustique physiologique et du même coup, au moins pour ceux qui acceptent son point de vue, sur la question des fondements, à la théorie de la musique.

> Helmholtz a formulé quatre théories en acoustique physiologique : une théorie non linéaire des sons résultants ; une théorie des battements pour expliquer la consonance et la dissonance ; une théorie de la qualité du son et, comme application spéciale de celle-ci, des voyelles ; et, finalement, une théorie de l'audition fondée sur la résonance. Dans tous ces domaines de la physiologie acoustique, sa méthode de recherche était orientée vers la théorie. En outre […], il a traité les problèmes qui se posent en physiologie acoustique d'une façon mathématique ; cette prédilection pour l'analyse mathématique des problèmes physiologiques faisait partie de son programme plus vaste d'application des méthodes des sciences physiques à la physiologie. En même temps, Helmholtz a manifesté une aptitude à la

39. *TPM*, p. 277-278.
40. *TPM*, p. 397.

> conception de nouveaux instruments, par exemple d'appareils à diapason et de résonateurs sphériques. L'appréciation de l'utilisation de ces instruments […] est d'une importance centrale pour la compréhension de l'acoustique physiologique de Helmholtz, car ils matérialisaient exactement ses engagements théoriques. En résumé, la caractéristique essentielle de l'approche que Helmholtz a de l'étude des processus sensoriels était, par conséquent, son aptitude à intégrer des éléments mathématiques, théoriques et instrumentaux dans une structure complexe et néanmoins unifiée. Chaque élément dans la structure renforçait les autres. En conséquence, même une observation ou une hypothèse qui, prise isolément, restait faible, devenait plausible quand elle était connectée aux autres éléments dans le réseau plus vaste[41].

Un des aspects les plus déterminants de la révolution que Helmholtz a effectuée, et dont le moins que l'on puisse dire est qu'il n'a pas été facile à accepter, a consisté, comme je l'ai dit, dans le remplacement du fondement métaphysico-ontologique que la conception pythagoricienne avait essayé de fournir à la musique par un fondement de nature bien différente et d'une espèce autrement plus contingente, à savoir l'oreille physiologique elle-même. Ce point-là est, aux yeux de Helmholtz, tout à fait crucial. La possibilité de la musique repose entièrement sur une particularité de l'oreille que d'autres sens, comme par exemple la vue, ne possèdent pas :

> En premier lieu, le phénomène des à-coups ou des battements repose sur l'interférence du mouvement ondulatoire ; cela ne pouvait par conséquent advenir au son que parce qu'il est un mouvement ondulatoire. D'autre part, pour le constat des intervalles consonants, était requise la capacité qu'a l'oreille de sentir les harmoniques et d'analyser des systèmes ondulatoires composés en systèmes simples selon le théorème de Fourier. Que les harmoniques des sons musicalement utilisables soient au son fondamental dans le rapport des nombres entiers à un, et que les rapports de vibrations des intervalles harmoniques correspondent pour cette raison aux plus petits nombres entiers, est fondé entièrement dans le théorème de Fourier. À quel point est essentielle la particularité physiologique de l'oreille qui a été mentionnée, cela devient notamment clair quand nous comparons cette dernière avec l'œil. La lumière aussi est un mouvement ondulatoire d'un milieu particulier répandu à travers l'espace universel, l'éther luminifère, la lumière aussi manifeste le phénomène de l'interférence. La lumière aussi a des ondes de différentes durées de vibration, que l'œil perçoit comme des couleurs différentes, à savoir celle qui a la plus grande durée de vibration comme rouge ; suivent ensuite orange, jaune, vert, bleu, vio-

41. Stephan Vogel, *op. cit.*, p. 260.

> let, dont la durée de vibration est à peu près à moitié aussi grande que celle du rouge le plus extrême. Mais l'œil ne peut pas distinguer l'un de l'autre des systèmes d'ondes lumineuses composés, autrement dit, des couleurs composées ; il les sent dans une sensation simple qui ne peut être analysée, celle d'une couleur mixte. Il lui est par conséquent indifférent de savoir si dans la couleur mixte sont réunies des couleurs fondamentales ayant des rapports de vibrations simples ou non simples. Il n'a pas d'harmonie au sens auquel l'oreille en a une ; il n'a pas de musique[42].

L'émergence du principe de la tonalité et l'avenir de la musique

La troisième partie de la *Théorie physiologique de la musique* s'intitule « Affinités des sons, gammes et tonalités ». Helmholtz y entreprend une sorte de déduction du système de la tonalité, avec ses deux modes majeur et mineur, à partir des principes de l'acoustique physiologique tels qu'il a été en mesure de les établir. Sa théorie lui semble capable d'expliquer pourquoi l'histoire de la musique a conduit à l'émergence – et, pour finir, à la domination complète – du système de la tonalité sur tous les autres types de solution qui avaient été essayés antérieurement. Et il pense qu'inversement l'histoire de la théorie musicale et celle de la musique elle-même, qui sont tombées d'accord pour sélectionner et consacrer finalement ce genre de système, de préférence à tous les autres, constituent une confirmation du fait que les fondements qu'il a essayé de procurer à la musique, et qui ne pouvaient être, comme on l'a vu, que physiologiques, étaient réellement les bons. Comme le dit Marcus Rieger :

> La tonalité majeure-mineure des XVII^e^-XIX^e^ siècles a pu, selon l'opinion de Helmholtz, s'imposer avant tout parce qu'elle rend le mieux justice à la sensation naturelle de la relation fonctionnelle entre les sons ou les accords individuels. Dans les tonalités majeure et mineure, les relations des sons entre eux sont les plus étroites qu'elles puissent être ; ces deux modes correspondent par conséquent également le mieux à la sensation de la tonalité[43].

On peut donc dire, pour reprendre les expressions qu'utilise Rieger, que, dans la *Théorie physiologique de la musique*, la colonisation, qui a

42. Helmholtz, « Über die physiologischen Ursachen… », *op. cit.*, p. 90.
43. Marcus Rieger, *op. cit.*, p. 128.

finalement réussi, du système des concepts fondamentaux de la théorie de la musique, est complétée par une tentative de colonisation du passé de la musique lui-même. Le résultat final est, selon lui, que « si le lecteur suit la théorie des sensations sonores, alors l'histoire de la musique apparaît comme l'histoire d'une longue série d'erreurs sur les véritables fondements de la musique, qui a commencé avec Pythagore *Musicus* et qui ne s'achève qu'avec les *Sensations sonores* de Helmholtz[44] ». Même s'il serait sûrement plus exact, du point de vue de Helmholtz, de parler de « vérités partielles » que d'erreurs, et si l'on prend soin de préciser que l'on doit aux « erreurs » en question toute la considération qu'elles méritent en tant qu'étapes qui devaient être franchies sur le chemin de la découverte progressive de la vérité, la conception de l'histoire que cela implique ne peut évidemment manquer de susciter des résistances compréhensibles.

Le point crucial, dans l'argumentation de Helmholtz, est la façon dont la relation des sons d'un système peut être déduite de la relation à un son fondamental :

> La musique moderne établit une liaison intime, purement musicale, entre tous les sons d'une phrase en établissant entre eux et une certaine tonique un rapport d'affinité aussi net que possible pour l'oreille. Nous pouvons, avec Fétis, désigner sous le nom de *principe de la tonalité* cette domination de la tonique formant comme le lien de tous les sons de la phrase. Ce savant musicien remarque avec raison que la tonalité s'est développée à des degrés très divers, et suivant des voies différentes, dans les mélodies des diverses nations[45].

Pour caractériser la relation qui existe entre les sons au sein d'un système, Helmholtz se sert d'une notion de parenté entre les sons, qui dépend de la composition physique de ceux-ci ; et il affirme que c'est dans le système de la tonalité avec ses deux modes que les relations de parenté naturelle entre les sons sont les mieux représentées et respectées. Ce sont les harmoniques d'un son qui décident dans chaque cas du degré auquel deux sons sont reliés l'un à l'autre. Plus un son composé a d'harmoniques en commun avec le son fondamental de la tonique, plus les affinités des sons entre eux sont fortes.

---

44. *Ibid.*, p. 157.
45. *TPM*, p. 313. Helmholtz renvoie ici à François-Joseph Fétis, *Biographie universelle des musiciens et bibliographie générale de la musique*, Paris, 1834-1836.

> Nous appelons apparentés au premier degré des sons qui ont deux sons partiels identiques ; apparentés au deuxième degré des sons qui sont apparentés au premier degré avec le même troisième son[46].

Comme on le voit, il y a ici deux principes différents qui interviennent simultanément : d'une part, les affinités entre les sons, pris deux à deux, et, d'autre part, la relation de parenté avec le son fondamental et principal, autrement dit, la tonique. Or c'est dans la gamme majeure et à un moindre degré dans la gamme mineure que le système des affinités entre les sons utilisés, compris de cette manière, est le plus étroit :

> La relation la plus étroite et la plus simple de toutes se trouve dans la gamme majeure, parce que tous les sons d'un air majeur apparaissent comme harmoniques, soit de la tonique, soit de sa quinte supérieure ou inférieure. Par suite, toutes les affinités des sons se trouvent ramenées aux affinités les plus fortes et les plus étroites du système musical, c'est-à-dire à la relation de quinte[47].

Dans la façon dont Helmholtz se représente l'histoire de la musique, une sorte de sentiment de la tonalité et d'aspiration à la tonalité ont dû être présents depuis le début, et on reconnaît à certains signes qu'ils l'ont été effectivement. C'est ce que confirment notamment les renseignements dont nous disposons sur la façon dont les Grecs concevaient et pratiquaient la musique :

> Il résulte des faits que nous venons de citer – et c'est là ce qui présente un intérêt particulier pour l'objet de nos recherches – que les Grecs, chez lesquels notre gamme diatonique a pris naissance, n'étaient pas absolument dépourvus du sentiment de la tonalité au point de vue esthétique ; seulement, ce sentiment n'était pas encore aussi nettement prononcé que dans la musique moderne, et surtout, à ce qu'il semble, il ne jouait aucun rôle bien caractérisé dans les règles techniques de la construction mélodique. Aussi Aristote, qui s'occupe de la musique comme esthéticien, est-il le seul écrivain connu jusqu'ici qui dise quelques mots de la question ; les auteurs spéciaux gardent un silence absolu à cet égard. Malheureusement, les indications d'Aristote sont encore si vagues qu'elles laissent encore subsister un assez grand nombre de doutes[48].

---

46. *TPM*, p. 359-360 (Helmholtz, *Die Lehre von den Ton empfindungen*, *op. cit.*, p 423).
47. *TPM*, p. 384.
48. *TPM*, p. 316.

Tant qu'on en reste au stade de la musique homophonique, les sept modes mélodiques que Helmholtz redécouvre par la voie de la déduction logique et recense comme ayant été ceux des Grecs et de l'ancienne Église chrétienne apparaissent comme également justifiés[49]. Mais la situation change évidemment de façon déterminante quand, après l'époque de la musique homophonique et celle de la musique polyphonique, on en arrive à celle de la musique harmonique, au sens propre du terme.

La théorie musicale et l'esthétique

Helmholtz soutient que l'harmonie a des causes naturelles et qu'elles peuvent être identifiées avec précision, mais il ne propose pas pour autant une théorie qui pourrait être qualifiée de « naturaliste », au sens strict. Dans la *Théorie physiologique de la musique*, il souligne à plusieurs reprises que les explications que la science naturelle est en mesure de donner sur les causes de l'harmonie ne peuvent prétendre en aucun cas remplacer celles de l'esthétique. Mais il ne croit pas non plus que les principes et les règles du beau puissent reposer sur des conventions plus ou moins arbitraires. Autrement dit, il récuse aussi bien la conception conventionnaliste radicale que la conception naturaliste :

> Le système des gammes, des modes et de leur enchaînement harmonique ne repose pas sur des lois naturelles invariables, mais […] il est, au contraire, la conséquence de principes esthétiques qui ont varié avec le développement progressif de l'humanité, et qui varieront encore. Il ne s'ensuit pas que le choix des éléments de la technique musicale soit purement arbitraire, et qu'ils ne puissent pas se déduire d'une loi plus générale. Au contraire, les règles de chaque style artistique forment un tout bien coordonné, surtout quand le style dont il s'agit est parvenu à un riche et complet développement. […] La théorie musicale, notamment, où des activités physiologiques particulières de l'oreille, non immédiatement accessibles à l'observation consciente, jouent un grand rôle, offre aux investigations esthétiques un champ vaste et fécond, pour déterminer le caractère de nécessité des règles techniques qui président à chaque direction particulière prise par l'art dans ses développements successifs[50].

49. *TPM*, p. 364-365.
50. *TPM*, p. 306-307.

Helmholtz estime qu'il ne faut pas céder à la tentation d'ériger la nature et le naturel en instance suprême dont le verdict ne peut pas être contesté. C'est, selon lui, une tentation à laquelle Rameau n'a pas suffisamment résisté[51]. L'auteur de la *Théorie physiologique de la musique* sait évidemment mieux que personne à quel point ce qui est perçu et accepté comme naturel est dépendant de la culture et de l'histoire, et peut changer au cours de l'évolution. Il a fallu dans certains cas beaucoup de temps pour que des intervalles et des accords qui avaient été déconseillés ou prohibés d'abord comme dissonants en viennent à être reconnus et classés finalement comme des consonances : la tierce et la sixte, par exemple, n'étaient pas considérées comme des intervalles consonants pendant la période de la musique polyphonique du Moyen Âge et ont mis bien plus de temps que la quinte à conquérir leur place dans la musique. Mais, naturellement, il est possible de soutenir, et c'est ce que fait Helmholtz, que, maintenant que l'on a découvert la hiérarchie objective des degrés de consonance et de dissonance, les choses ne devraient en principe plus changer dans ce domaine. Ce qui pourrait éventuellement encore changer, bien sûr, est l'utilisation qui est faite dans la musique de la consonance et de la dissonance, et la place respective qui est accordée à chacune des deux. Mais c'est une autre question.

Il y a de nombreux exemples qui montrent, en outre, clairement que, là où on croit encore souvent être en train d'entendre la voix de la nature, ce que l'on entend est en réalité plutôt celle de la convention et de l'habitude. Un exemple assez typique de cela est, aux yeux de Helmholtz, la prohibition des suites d'octaves et des suites de quintes dans l'harmonie, dont on peut montrer, selon lui, que, contrairement à ce que l'on affirme la plupart du temps, elle n'a pas de fondement naturel réel et a été instaurée pour des raisons de nature différente[52]. Si les lois de la composition exigent que les suites de quintes, par exemple, soient évitées, « ce n'est pas parce qu'elles sonnent mal pour l'oreille ; c'est suffisamment prouvé par ce fait que presque tous les sons de la voix et de la plupart des instruments sont accompagnés de douzièmes, et que c'est sur cet accompagnement même que repose toute la structure de notre système musical. Par conséquent, tant que les quintes apparaissent comme éléments mécaniquement constitutifs du son, elles sont pleinement justifiées. C'est ce qui arrive pour les jeux de fourniture de l'orgue »[53]. La justification véritable de la prohibition des suites de quintes et des suites

51. *TPM*, p. 300.
52. *TPM*, p. 473-476.
53. *TPM*, p. 474.

d'octave réside dans le fait que, bien que l'importance de l'harmonie se soit accrue de façon considérable par rapport à celle de la mélodie dans la musique moderne, « la véritable perfection n'en consiste pas moins en ce que, dans la combinaison de plusieurs parties, chacune ait sa beauté propre, une marche facile à saisir ; le mouvement de l'ensemble restant pour l'auditeur d'une compréhension facile[54] ».

Comment est-il possible que l'histoire de la musique ait cherché à atteindre, et ait fini par atteindre presque complètement un but dont les agents qui ont été impliqués historiquement dans cette évolution n'avaient pas connaissance, puisqu'il n'a pu être reconnu qu'à peu près au moment où il a été atteint ? Pour expliquer cela, Helmholtz, dans le dernier chapitre de son livre, qui est intitulé « Points de contact avec l'esthétique », invoque une sorte de principe de la rationalité inconsciente qui est à l'œuvre dans l'œuvre et d'art et également dans le jugement de goût et dans la critique :

> En cherchant, par la critique, à comprendre la beauté d'une œuvre de ce genre, résultat auquel il nous est donné d'atteindre jusqu'à un certain point, nous montrons bien que nous supposons dans l'œuvre d'art un raisonnement inconscient que l'intelligence peut découvrir, mais dont la connaissance n'est indispensable ni à la création, ni au sentiment du beau. […] La principale difficulté, dans ce domaine, consiste à comprendre comment nous pouvons apprécier par intuition la conformité d'une œuvre aux lois de la raison, sans que nous ayons réellement conscience de ces dernières. Et cette intuition inconsciente des lois esthétiques n'est pas, dans l'action du Beau sur notre esprit, un accessoire qui peut être ou ne pas être ; il est évident, au contraire, qu'elle en est précisément le point capital, saillant[55].

Helmholtz insiste lui-même sur le fait que les lois naturelles de la perception auditive ne jouent ici que le rôle d'un matériau, dont le véritable créateur aussi bien des œuvres musicales que des systèmes musicaux qui ont existé historiquement – autrement dit finalement, si on considère les choses comme il propose de le faire, la raison elle-même – se sert pour réaliser ses fins :

> De même que les peuples ont été conduits, par la différence de leurs goûts, à construire avec les mêmes pierres des édifices de caractères très différents, de même aussi nous voyons, dans l'histoire de la musique, les

54. *TPM*, p. 473.
55. *TPM*, p. 480.

> mêmes propriétés de l'oreille humaine servir de base à des systèmes musicaux très divers. D'après cela, à mon avis, nous ne pouvons douter que non seulement la naissance des chefs-d'œuvre de la musique, mais même la construction de notre système de gammes, de tons, d'accords, en un mot tout ce qui rentre ordinairement dans la théorie de l'harmonie, ne soient une création du sentiment artistique, et, par conséquent, ne doivent être soumises aux lois de la beauté esthétique[56].

Comme on a pu le constater, pour un auteur que l'on qualifie fréquemment, de façon beaucoup trop simpliste à mon sens, de « mécaniste », Helmholtz, dans ses recherches sur la théorie de la musique, fait intervenir largement des considérations téléologiques. C'est évident dans sa conception de l'histoire de la musique. Mais ça l'est tout autant dans sa théorie de la consonance. Il est, comme je l'ai dit, parfaitement conscient du fait que la distinction entre ce qui peut être accepté comme consonant et ce qui ne le peut pas est susceptible de changer de façon importante, qu'elle l'a fait effectivement, et qu'elle pourrait en principe encore le faire. Qui plus est, ce sont précisément ses travaux qui ont contraint le monde musical à accepter l'idée, qui était au départ à peu près irrecevable pour lui, que la différence entre la consonance et la dissonance est une différence de degré, et non de nature, une conclusion qui s'impose à peu près immédiatement dès que l'on se rend compte que, quand deux sons résonnent simultanément, il faut tenir compte, pour évaluer le degré de consonance, non seulement des battements des deux sons fondamentaux, mais également de ceux de leurs harmoniques et de ceux des sons résultants qu'engendre leur combinaison, sans oublier le fait que les harmoniques, pris deux à deux, peuvent donner, eux aussi, des sons résultants.

Mais Helmholtz ne pouvait évidemment pas aller jusqu'à remettre en question le rapport de subordination qui était censé exister entre la dissonance et la consonance, et le principe téléologique qui veut que la dissonance ne puisse être admise qu'en vue de la consonance, considérée dans tous les cas comme le but réel :

> Les consonances ont, par elles-mêmes, le droit d'exister ; c'est d'après elles que se sont formées nos gammes modernes. Les dissonances, au contraire, ne sont admises que comme transitions entre des consonances. Elles n'ont aucun droit logique à l'existence, et, en les traversant, les parties restent par conséquent soumises, pour parcourir les degrés de la gamme, à la loi établie pour favoriser les consonances[57].

---

56. *TPM*, p. 478.
57. *TPM*, p. 436.

Le problème qui se pose est évidemment que, comme je l'ai déjà suggéré, le travail de Helmholtz et les résultats auxquels il parvient, achèvent, d'une certaine façon, d'ébranler des certitudes auxquelles il cherchait justement à procurer un fondement.

De la parole et du bruit
L'organisation auditive de l'identification de la parole

par Dan Gnansia et Christian Lorenzi

Chez les personnes entendantes, l'intelligibilité de la parole en présence de bruit de fond masquant est substantiellement améliorée dans les bruits fluctuants (modulés en amplitude et en fréquence) par rapport aux bruits stationnaires. Cette capacité de « démasquage de la parole » s'avère extrêmement utile en condition réelle d'écoute car elle assure une grande résistance face aux bruits environnants. Le démasquage semble résulter de notre capacité à « écouter » la parole dans les vallées spectro-temporelles du bruit masquant. Cette capacité implique non seulement une acuité temporelle et une résolution fréquentielle auditive normale, mais également une capacité à extraire et utiliser des indices acoustiques extrêmement rapides (dits de structure temporelle fine) grâce aux propriétés de synchronisation des neurones auditifs. Pris ensemble, ces résultats sur le démasquage militent en faveur d'une conception du système auditif comme système de démodulation.

Démasquage de la parole

Dans une étude psychoacoustique considérée aujourd'hui comme *princeps*, Miller et Licklider[1] ont démontré que notre capacité auditive à identifier des sons de parole – l'« intelligibilité de la parole » – est bien meilleure en présence de bruit masquant interrompu que de bruit masquant stationnaire. Cette amélioration – parfois considérable – révèle la

1. Miller et Licklider (1950).

grande résistance ou robustesse des processus auditifs de traitement des signaux de parole, qu'ils soient périphériques ou centraux. Ce phénomène dépend des paramètres de la stimulation, car elle varie en fonction du niveau sonore de présentation du masque (le rapport signal-sur-bruit), de la fréquence, et du rapport cyclique des interruptions, mais pas de la régularité des interruptions (et ainsi des attentes possibles des auditeurs).

La figure 1 résume certains résultats expérimentaux de cette étude, en présentant la quantité de « démasquage » (*masking release*), autrement dit la différence entre performance en bruit fluctuant et performance en bruit stationnaire, en fonction du rapport signal-sur-bruit à long terme, pour un bruit interrompu à 10 Hz avec un rapport cyclique de 50 %.

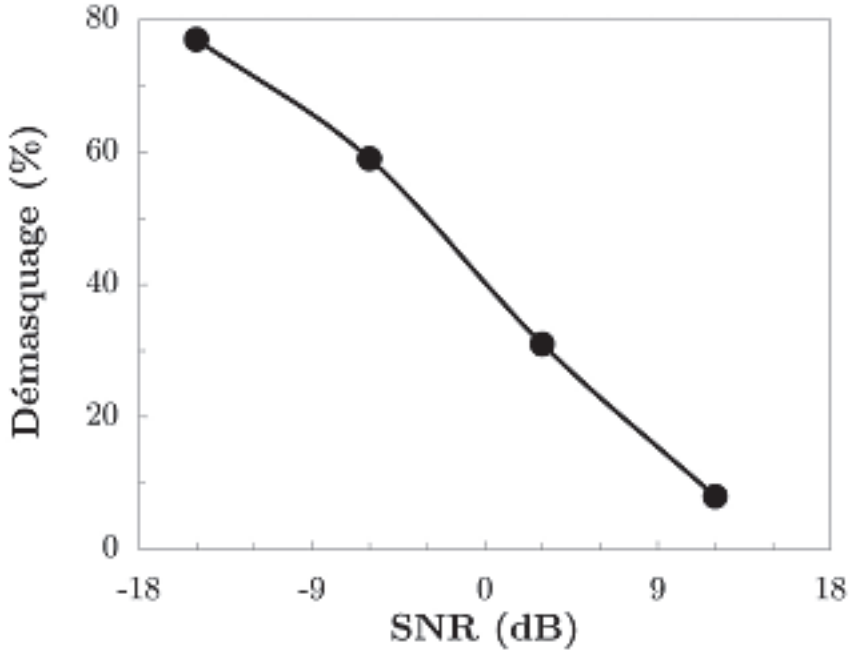


Figure 1 : *Démasquage de la parole en différence de pourcentage de réponses correctes en fonction du rapport signal-sur-bruit à long terme (SNR). D'après Miller et Licklider (1950).*

Quels sont les paramètres physiques du signal de parole et du bruit masquant contribuant à ce mécanisme remarquable et comment les étudier ? Les sons masquant non stationnaires de notre environnement sont généralement des brouhahas (*babble*) produits par la parole d'autres locuteurs (une situation quotidienne d'écoute). L'utilisation de bruit stationnaire que l'expérimentateur module pour y introduire des fluctuations temporelles ou spectrales répond à un besoin de contrôle expérimental

pour l'étude de la perception de la parole dans le bruit fluctuant. En effet, ce type de masque permet, d'une part, une étude quantitative précise du démasquage dans certaines conditions choisies, car les caractéristiques des « vallées » du bruit (ses *minima*) sont maîtrisées, et, d'autre part, une détermination des principales caractéristiques acoustiques et auditives rendant possible l'extraction d'un son de parole cible au sein du mélange parole/bruit. Nous allons voir par la suite que d'autres masques peuvent être utilisés, et notamment différents types de signaux de parole concurrents.

Effets paramétriques

EFFET DU NOMBRE DE LOCUTEURS INTERFÉRENTS

Afin de se rapprocher d'un contexte de masque écologique de type « Cocktail Party[2] », de nombreuses études sur l'intelligibilité de la parole ont été menées en utilisant des signaux de parole concurrents (*natural babble*) afin de masquer le signal de parole cible[3].

Une étude réalisée par Miller[4] a fourni les premiers résultats caractérisant l'effet du nombre de locuteurs interférents sur l'intelligibilité de la parole. La mesure (ou variable dépendante) utilisée dans cette étude est le SRT (*speech reception threshold*) qui correspond à la valeur (en dB) du rapport signal-sur-bruit nécessaire pour atteindre un score d'identification correcte de 50 % (ou 40 %, selon les études). Miller rapporte une diminution de 8 dB SRT entre 1 et 2 locuteurs interférents, et une autre diminution de 3 à 4 dB SRT entre 2 et 4, 6 et 8 locuteurs interférents. Cette étude révèle donc que l'effet de démasquage culmine pour un seul locuteur interférent, et diminue substantiellement avec le nombre de signaux de parole concurrents pour converger vers le niveau observé pour un bruit stationnaire.

Pourtant, l'intelligibilité de la parole est meilleure lorsque la cible est masquée par 4 à 8 locuteurs interférents qu'avec un masque comportant une infinité de locuteurs interférents[5]. Ce bruit quasi stationnaire, obtenu en additionnant une infinité de locuteurs interférents ayant le même spectre long terme que la parole, est appelé bruit, « *speech-shaped noise* » (SSN). Ces résultats mettent une nouvelle fois en lumière le mécanisme

2. Cherry (1953).
3. Pour une revue détaillée, voir Bronkhorst (2000).
4. Miller (1947).
5. Duquesnoy (1983) ; Festen et Plomp (1990).

de démasquage de la parole, car un nombre limité de locuteurs interférents correspond à un masque présentant des fluctuations temporelles et spectrales (et, de fait, des *minima* ou vallées) perceptivement saillantes. De la même façon, Danhauer et Lepper[6] rapportent de meilleurs scores pour l'intelligibilité lorsque la parole est masquée par 4 à 9 locuteurs que lorsqu'elle est masquée par un bruit stationnaire de même énergie.

L'utilisation de locuteurs interférents comme masque est certes écologique mais elle engage plusieurs mécanismes perceptifs de bas et haut niveau, dont des effets de masquage dits « informationnels » (par contraste avec le masquage énergétique ayant lieu au sein des bandes fréquentielles cochléaires). Ces effets informationnels sont liés pour partie au contenu sémantique du masque[7]. Afin de limiter au maximum l'influence de facteurs linguistiques de haut niveau (de type sémantique, par exemple), plusieurs études sur l'intelligibilité de la parole dans le bruit utilisent un masque fluctuant construit à partir d'un bruit SSN. Ce masque est modulé par l'enveloppe des signaux de parole interférents et appelé « *babble-modulated noise*[8] ». Ainsi, le contenu sémantique du masque modulé est fortement appauvri car ce dernier n'est que peu ou pas intelligible, mais ses fluctuations temporelles restent similaires à celles du signal de parole interférent d'origine.

La figure 2 présente les résultats de Simpson et Cooke[9] obtenus avec ce type de masque. Conformément aux résultats précédents, cette figure indique que le pourcentage d'identification de non-mots bisyllabiques (VCV, voyelle/consonne/voyelle) diminue substantiellement en fonction du nombre de locuteurs interférents pour un *natural babble* et pour un *babble-modulated noise*, pour converger vers le niveau obtenu avec un bruit stationnaire.

EFFET DU BRUIT MODULÉ EN AMPLITUDE

Pour les raisons citées plus haut, de nombreux travaux ont été menés en utilisant comme masque des bruits SSN modulés en amplitude. Le *babble-modulated noise* que nous venons de voir est un masque de ce type, mais il présente des *minima* temporels et spectraux sans régularité, et donc difficilement contrôlables. Ainsi, afin de mesurer quantitativement la durée et la profondeur de ces instants pendant lesquels l'énergie

6. Danhauer et Lepper (1979).
7. Rhebergen *et al.*, (2005).
8. Festen et Plomp (1990) ; Bronkhorst et Plomp (1992) ; Peters *et al.* (1998) ; Versfeld et Dreschler (2002) ; Cooke (2006).
9. Simpson et Cooke (2005).

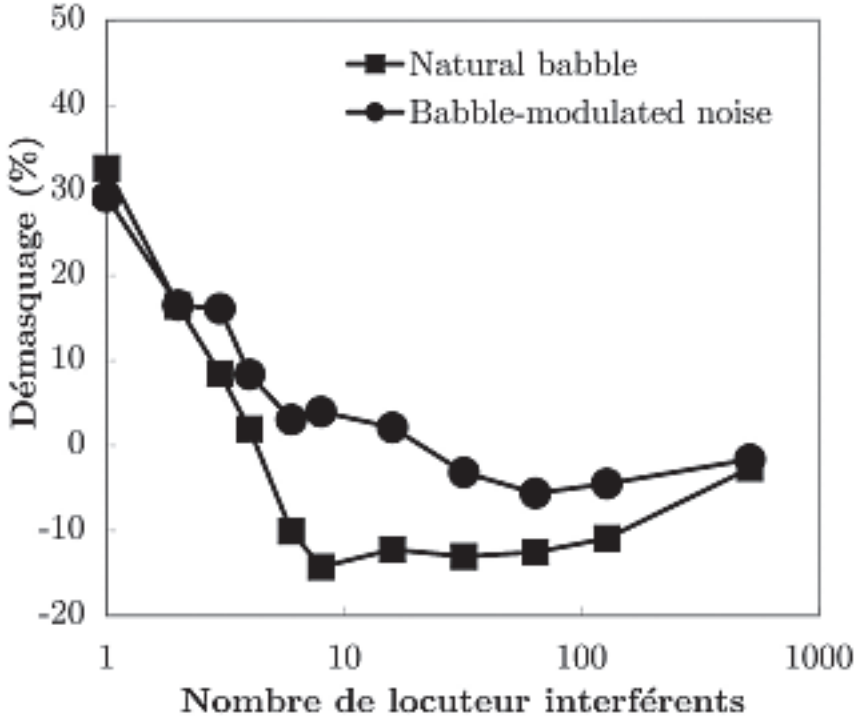


Figure 2 : *Niveau de démasquage (en pourcentage d'identification de non-mots bisyllabiques [VCV, voyelle/consonne/voyelle]) en fonction du nombre de locuteurs interférents (N) pour un natural babble (triangles, trait plein), et pour un babble-modulated noise (ronds, trait pointillé). Les résultats obtenus pour un speech-shaped noise stationnaire sont représentés par un cercle ouvert sur la droite du graphique. D'après Simpson et Cooke (2005).*

du bruit est moindre et le démasquage de la parole possible, des modulations simples et régulières telles que des modulations sinusoïdales ou carrées peuvent être appliquées. Le choix du bruit porteur est le plus souvent un bruit SSN car un tel bruit masque plus efficacement la parole sur l'ensemble du spectre audible qu'un bruit blanc[10].

Voyons maintenant les effets des paramètres physiques de la modulation sur le démasquage de la parole.

Effets de la fréquence de modulation

L'étude de Miller et Licklider présentée plus haut[11] nous fournit un premier aperçu de l'effet de la fréquence de modulation sur le démas-

10. Horii *et al.* (1970) ; Berry et Nerbonne (1972).
11. Miller et Licklider (1950).

quage de la parole. En effet, cette étude révèle que l'intelligibilité de phrases présentées dans un masque modulé de manière carrée est maximale pour des fréquences comprises entre 2 et 10 Hz environ, et varie selon le rapport signal-sur-bruit de présentation (pour _18 dB, intelligibilité maximum à 8 Hz ; pour 9 dB, intelligibilité maximum à 15 Hz), puis, pour les valeurs supérieures, le démasquage diminue jusqu'à être quasiment aboli pour des fréquences de modulation supérieures à 100 Hz. Pour les valeurs de fréquence de modulation les plus basses, les auteurs observent une diminution du démasquage, conduisant à une *fonction de démasquage d'allure passe-bande (sélective) centrée sur l'intervalle 8-16 Hz*.

Des ordres de grandeur similaires ont été constatés par Gustafsson et Arlinger[12] avec des stimuli de type phrase. En effet, ces auteurs relèvent un pic de démasquage de 45 points de pourcentage lorsque le masque est modulé sinusoïdalement à des fréquences comprises entre 10 et 20 Hz, puis une diminution progressive et une abolition du démasquage à partir de 100 Hz. En revanche, contrairement à Miller et Licklider, Gustafsson et Arlinger n'observent pas de réelle diminution du démasquage pour les plus basses valeurs de fréquence de modulation testées.

Avec des stimuli de type non-mots bisyllabiques (VCV, voyelle/consonne/voyelle), un maximum de démasquage de 35 points de pourcentage est obtenu pour un masque modulé sinusoïdalement à 8 Hz[13], avec une diminution pour les plus faibles et plus hautes valeurs de fréquence de modulation, accompagnée d'une quasi-abolition du démasquage à partir de 100 Hz. Ici, l'utilisation de stimuli VCV est avantageuse car elle permet de calculer la quantité d'information reçue pour certains traits phonétiques[14] comme le voisement, le lieu d'articulation et le mode d'articulation. Les *maxima* de démasquage observés pour ces trois traits phonétiques sont différents : le démasquage exprimé en termes de voisement reçu est maximum pour des fréquences de modulation allant de 2 à 64 Hz, le lieu d'articulation présente un pic à 32 Hz chez les sujets normo-entendants jeunes, et à 8 Hz chez les sujets normo-entendants âgés[15] ; enfin, le mode d'articulation présente un pic de démasquage pour une fréquence de modulation de 8 Hz. Les caractéristiques du démasquage (l'allure de la fonction de démasquage, par exemple, reliant le démasquage à la fréquence de modulation du masque) dépendent donc de la nature des informations phonétiques reçues. Ceci révèle l'existence

12. Gustafsson et Arlinger (1994).
13. Füllgrabe *et al.* (2006) ; Lorenzi *et al.*, (2006b).
14. Miller et Nicely (1955).
15. Füllgrabe *et al.* (2006) ; Lorenzi *et al.*, (2006b).

d'une analyse *à différentes échelles de temps* des informations acoustiques et phonétiques extraites au sein du bruit fluctuant.

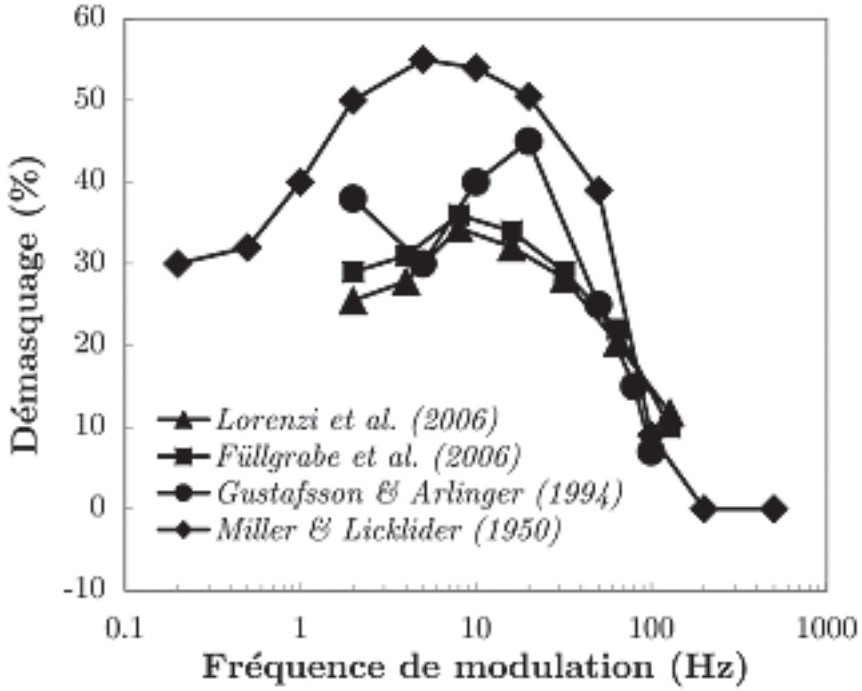


Figure 3 : *Démasquage de la parole exprimé en termes de différence de pourcentage de réponses correctes, en fonction de la fréquence de modulation du masque. D'après Lorenzi* et al. *(2006b ; stimuli : VCV ; triangles), Füllgrabe* et al. *(2006 ; stimuli : VCV ; carrés), Gustafsson et Arlinger (1994 ; stimuli : phrases ; cercles) et Miller et Licklider (1950 ; stimuli : mots ; losanges).*

La figure 3 résume les résultats de ces quatre études en présentant le démasquage observé (calculé à partir des pourcentages de réponses correctes) en fonction de la fréquence de modulation du masque. Un maximum de démasquage est observé pour des fréquences de modulation comprises entre 5 et 32 Hz environ. Ces études présentent toutefois des différences méthodologiques importantes qu'il convient de souligner. En effet, Miller et Licklider (1950) ont choisi de fixer le rapport signal-sur-bruit à travers les sujets (pour la figure 3, le rapport signal-sur-bruit à long terme est de -6 dB), alors que les autres études ajustent individuellement le rapport signal-sur-bruit afin que le sujet atteigne 30 % (Gustafsson et Arlinger, 1994) ou 50 % (Füllgrabe *et al.*, 2006 ; Lorenzi *et al.*, 2006b) de réponses correctes dans la condition de bruit stationnaire. De plus, Miller et

Licklider (1950) modulent le bruit à l'aide d'un signal carré, alors que Füllgrabe *et al.* (2006) et Lorenzi *et al.* (2006b) modulent le bruit à l'aide d'une sinusoïde, et Gustafsson et Arlinger (1994) avec un signal périodique complexe. Toutefois, la comparaison de ces différents résultats permet d'affirmer que le démasquage de la parole reste optimal pour des fréquences de modulation comprises entre 5 et 32 Hz, ce qui correspond à une « *fenêtre temporelle d'écoute dans les vallées du bruit* » comprise entre 16 et 100 ms, fenêtre variable selon le type d'indice acoustique/phonétique extrait ou traité par le sujet. De nombreuses autres études confirment ce résultat[16].

Effet de la profondeur de modulation

L'effet de la profondeur de modulation du masque (le contraste entre pics et creux d'énergie du masque) sur l'intelligibilité de la parole n'a fait l'objet que de peu d'études. Howard-Jones et Rosen (1993a) et Gnansia *et al.* (2008) utilisent ici un masque fluctuant à modulations carrées ou sinusoïdales, de fréquence de modulation fixe et égale à 10 ou 8 Hz, et dont la profondeur de modulation varie systématiquement entre 0 et 100 %. L'intelligibilité est mesurée avec des stimuli VCV, en fonction du niveau du bruit dans les vallées. Les auteurs remarquent que *plus le niveau de bruit est faible dans les minima (plus la vallée est profonde), meilleure est l'intelligibilité de la parole*. Il est à noter que les effets de démasquage sont observés dès une profondeur de modulation de 12,5 % (soit un rapport signal-sur-bruit local de –5 dB) et s'accentuent nettement à partir d'une profondeur de modulation de 50 % (soit un rapport signal-sur-bruit local de 0 dB)[17]. D'autres études rapportent également une meilleure intelligibilité aux profondeurs de modulation élevées[18].

Prises ensemble, ces études suggèrent que les sujets normo-entendants « entraperçoivent » le signal de parole au sein d'un bruit masquant si les vallées spectro-temporelles du bruit présentent certaines caractéristiques rapportées ci-dessus. Dans une étude de modélisation informatique de ce phénomène dit *glimpsing*, Cooke[19] observe que, pour simuler quantitativement les performances globales de démasquage de sujets normo-entendants, les vallées spectro-temporelles du bruit utilisées par le modèle afin d'extraire les informations partielles de parole et de procéder à une reconnaissance doi-

16. *Cf.* Licklider et Guttman (1957) ; Festen (1987) ; Kwon et Turner (2001) ; Dubno *et al.*, (2002) ; Buss *et al.* (2003) ; Dubno *et al.* (2003) ; Nelson *et al.*, (2003) ; Rhebergen *et al.* (2006).
17. Gnansia *et al.* (2008).
18. *Cf.* Gustafsson et Arlinger (1994) ; George *et al.* (2006).
19. Cooke (2003 ; 2006).

vent produire un rapport signal-sur-bruit local d'au moins –5 dB (pour des stimuli VCV). Cette étude n'a certes pas été menée à l'aide d'un bruit modulé par un signal périodique, mais cette valeur de –5 dB fournit une première estimation de la *profondeur minimale de la vallée de bruit* utilisée par le système auditif humain dans le démasquage. Ces résultats semblent compatibles avec les résultats obtenus par Gnansia *et al.* (2008).

Effet de la forme de la modulation

Les travaux cités ci-dessus utilisent des modulations d'amplitude de formes différentes. En effet, certains auteurs utilisent des modulations carrées[20], sinusoïdales[21], ou périodiques et complexes[22]. D'après les résultats précédents portant sur les effets de la profondeur de modulation, plus la vallée de bruit est importante (large et définie), plus la capacité d'écoute dans les vallées (*glimpsing*) est efficace, et ainsi, meilleure est l'intelligibilité (et donc le démasquage). La figure 3 indique effectivement que le plus fort effet de démasquage est observé pour l'étude de Miller et Licklider (1950), ces derniers utilisant un masque à modulations carrées, à savoir le masque dont les vallées sont les mieux définies et les plus longues.

Cet effet se vérifie également sur la figure 4 : Rhebergen *et al.* (2006) comparent ici les SRT pour l'identification de phrases dans un bruit fluctuant présentant des modulations carrées ou sinusoïdales, et ce, pour plusieurs fréquences de modulation. Les valeurs de fréquence de modulation les plus basses et les interruptions carrées engendrent une meilleure performance d'identification que la modulation sinusoïdale. Ces résultats suggèrent une nouvelle fois que l'intelligibilité est meilleure lorsque les vallées du bruit sont les mieux définies et les plus longues.

Les travaux de Füllgrabe *et al.* (2006) illustrent une autre manière de modifier la forme de la modulation du bruit et d'en étudier les effets sur le démasquage. Les auteurs comparent ici les effets de deux types de modulation du bruit masquant sur le démasquage de la parole. Les fluctuations du masque sont obtenues en appliquant une modulation sinusoïdale à l'amplitude du bruit dans un premier temps (modulation sinusoïdale dite « du premier ordre »), puis à la profondeur de modulation d'un bruit modulé sinusoïdalement en amplitude (modulation sinusoïdale dite « du second ordre ») dans un deuxième temps. Les résultats indiquent que le démasquage est indépendant de la fréquence de la variation de la pro-

20. Miller et Licklider (1950) ; Howard-Jones et Rosen (1993a) ; Dubno *et al.* (2002, 2003) ; Nelson *et al.* (2003) ; George *et al.* (2006) ; Rhebergen *et al.* (2006).
21. Takahashi et Bacon (1992) ; Füllgrabe *et al.* (2006) ; Lorenzi *et al.* (2006b) ; Rhebergen *et al.* (2006).
22. Gustafsson et Arlinger (1994) ; Füllgrabe *et al.* (2006) ; Rhebergen *et al.* (2006).

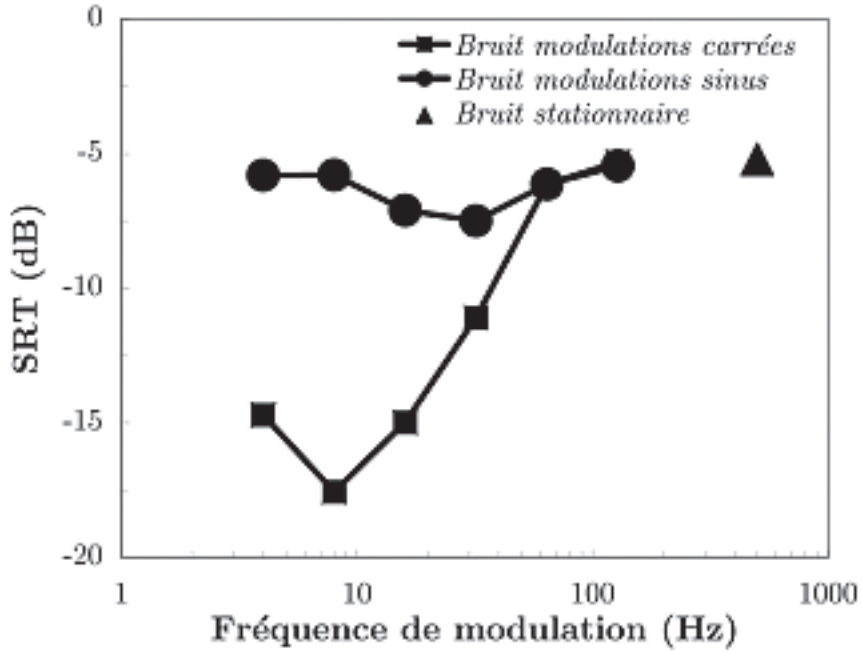


Figure 4 : *SRT en fonction de la fréquence de modulation du bruit masquant, pour un bruit modulé par un signal carré (carrés), une sinusoïde (ronds), ou sans modulation (triangle), pour une tache d'identification de phrases. D'après Rhebergen et al. (2006).*

fondeur, suggérant d'une part que la distribution temporelle des vallées du masque n'influence pas les résultats, et d'autre part que la profondeur et la durée des vallées du masque correspondent bien aux facteurs cruciaux qui contrôlent le phénomène de démasquage.

Autres effets

D'autres paramètres peuvent influencer le démasquage de la parole, comme par exemple le rapport cyclique (*duty cycle*). Plusieurs travaux confirment les effets de ce paramètre sur l'intelligibilité de la parole et le démasquage[23]. En effet, l'intelligibilité de la parole diminue à mesure que le rapport cyclique augmente. Ainsi, conformément aux résultats précédents, plus la vallée de bruit est conséquente, meilleure est l'intelligibilité dans le bruit, et donc le démasquage.

23. Nelson *et al.* (2003) ; Rhebergen *et al.* (2006).

Parmi les facteurs influençant la mesure de l'intelligibilité et du démasquage, le choix des stimuli et la méthode d'évaluation semblent également jouer un rôle critique. Ainsi, les auteurs recourent soit à l'identification de syllabes sans signification[24], soit à l'identification de mots[25] ou de phrases[26]. Ces stimuli diffèrent considérablement par leur complexité à plusieurs niveaux linguistiques (acoustique, phonétique, phonologique, morphologique, syntaxique et sémantique). Segmentation et/ou accès au lexique sont requis dans certains cas (la reconnaissance de mots ou phrases par exemple) alors que seule l'identification de segments phonétiques isolés est requise dans d'autres (la reconnaissance de logatomes sans signification par exemple). Par ailleurs, la mesure d'intelligibilité peut être réalisée à partir d'une tâche en choix forcé parmi un nombre limité d'alternatives (le sujet choisit un non-mot/un mot/une phrase au sein d'une liste fermée) ou non. Enfin, la métrique choisie pour la mesure de l'intelligibilité peut être différente selon l'étude : il peut s'agir du nombre de réponses correctes, mais aussi du nombre de mots clés correctement identifiés dans une phrase[27], ou encore du nombre de mots correctement identifiés dans une phrase[28]. Enfin, pour la mesure de l'intelligibilité dans le bruit et du démasquage de la parole, la grandeur choisie peut être soit un pourcentage de réponse correcte, soit un niveau de bruit pour lequel le sujet obtient 40 ou 50 % de réponses correctes (le SRT), soit une différence de pourcentage entre la performance obtenue avec un bruit fluctuant et celle obtenue avec un bruit stationnaire.

Ces variations importantes entre paradigmes expérimentaux limitent parfois fortement la possibilité de comparer les résultats des nombreuses études portant sur le démassquage, et expliquent pour partie certaines divergences notées quant aux effets des paramètres étudiés ci-dessus.

---

24. Howard-Jones et Rosen (1993a, b) ; Assmann et Summerfield (1994) ; Dubno *et al.* (2002, 2003) ; Buss *et al.* (2004) ; Cutler *et al.* (2004) ; Cooke (2006) ; Füllgrabe *et al.* (2006) ; Lorenzi *et al.* (2006b).
25. Miller et Licklider (1950) ; Carhart *et al.* (1975) ; Snell *et al.* (2002) ; Buss *et al.* (2003) ; Turner *et al.* (2004).
26. Plomp et Mimpen (1979) ; Duquesnoy (1983) ; Festen et Plomp (1990) ; Gustafsson et Arlinger (1994) ; Dorman *et al.* (1998) ; Nelson et Jin (2004) ; George *et al.* (2006).
27. *Cf.* Nelson et Jin (2004).
28. *Cf.* Dorman *et al.* (1998) ; George *et al.* (2006).

Mécanismes auditifs du démasquage

NATURE DES INDICES AUDITIFS

D'un point de vue acoustique, la parole est un signal présentant une structure spectrale et temporelle complexe. L'identification de ces formes complexes peut être conçue comme reposant d'une part sur la perception d'*indices auditifs spectraux* fournis par la décomposition fréquentielle réalisée par la membrane basilaire (aboutissant à une représentation spatiale ou tonotopique de ces indices) et d'autre part sur la perception d'*indices auditifs temporels* fournis par les distributions de décharges neurales des fibres afférentes du nerf auditif[29]. Les indices fréquentiels dits « de place d'excitation » (*i.e.* tonotopiques) peuvent être à la fois statiques (formants des voyelles) mais également dynamiques (transitions formantiques caractérisant l'articulation entre deux phonèmes). Les indices temporels correspondent, quant à eux, aux fluctuations de l'amplitude ou de la fréquence instantanée du signal de parole au sein de chaque bande de fréquence cochléaire (*i.e.*, pour chaque place d'excitation dans la cochlée). Ces dernières peuvent être divisées sommairement et – peut-être avec un certain degré d'arbitraire – en trois groupes en fonction de leurs corrélats acoustiques, perceptifs et linguistiques[30] (voir figure 5) : les fluctuations les plus lentes (entre 2 et 50 Hz environ) qui correspondent aux informations transmises par l'*enveloppe d'amplitude* du signal au sens strict (véhiculant notamment les informations sur l'intensité, la durée, le temps d'attaque et de chute du signal de parole, le tempo), et les fluctuations plus rapides (supérieures à 50 Hz) qui correspondent à la *périodicité* et à la *structure temporelle fine* du signal. Les fluctuations de périodicité comprises entre 50 et 800 Hz environ véhiculent en particulier l'information sur la fréquence fondamentale du signal (F0), ou fondamentale laryngée[31], mais aussi sur le voisement et la qualité de la voix[32], alors que les fluctuations de structure temporelle fine semblent véhiculer essentiellement des informations de timbre, de lieu d'articulation, de voisement et certaines informations sur le mode d'articulation (contraste nasales/non nasales)[33].

29. Pour une revue, voir Moore (2007).
30. Rosen (1992).
31. Terhardt (1972a, b).
32. Rosen (1992).
33. Rosen (1992).

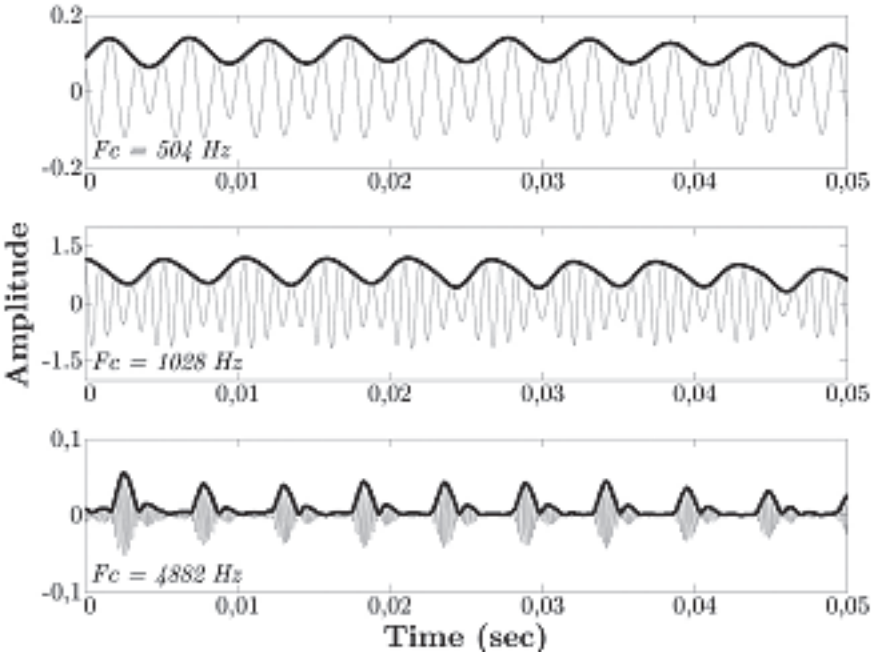


Figure 5 : *Signal d'onde en sortie de filtres cochléaires simulés centrés à 504, 1 028 et 4 882 Hz, répondant au son « a » de « aba » en Français. La ligne épaisse et la ligne fine montrent respectivement le résultat d'une décomposition des signaux en bande étroite en une enveloppe d'amplitude (modulateur de la porteuse) et une structure temporelle fine (porteuse).*

Ces indices temporels peuvent également, et toujours probablement avec un certain degré d'arbitraire[34], être scindés en deux catégories, résultant d'une décomposition du signal à bande étroite en chaque point de la membrane basilaire en une enveloppe temporelle et une structure temporelle fine. En première approche, du point de vue du signal, l'enveloppe temporelle et la structure temporelle fine peuvent être vues respectivement comme un modulateur et un signal porteur (le module et l'argument de la transformée de Hilbert du signal sonore en bande étroite) mais, du point de vue de la cochlée, comme les éléments du codage spatial et temporel. En effet, l'enveloppe temporelle correspondant aux fluctuations lentes et énergétiques du signal peut être considérée comme le principal élément du codage tonotopique le long de la membrane basilaire. La structure temporelle fine, suivant quant à elle les fluctuations rapides de la fréquence instantanée du signal, correspond aux informations temporelles proprement dites dans le nerf auditif, notamment grâce

34. *Cf.* Zeng *et al.* (2004) ; Gilbert et Lorenzi (2006).

aux phénomènes de *verrouillage de phase* des fibres du nerf auditif (décharges dans le nerf auditif synchronisées avec la structure temporelle fine jusqu'à environ 5 kHz[35]).

Ces fluctuations lentes et rapides d'enveloppe temporelle et de structure temporelle fine semblent jouer un rôle capital dans l'intelligibilité de la parole dans le silence. En effet, les indices d'enveloppe temporelle, même dégradés, suffisent pour une bonne intelligibilité[36]. Concernant l'intelligibilité dans le bruit et le démasquage de la parole, l'importance qualitative et quantitative de l'enveloppe temporelle (liée également à la résolution spectrale) et de la structure temporelle fine sera précisée par la suite. On peut cependant indiquer que, si l'enveloppe temporelle au sein d'un nombre limité de bandes de fréquence (4-8 bandes) suffit à percevoir un signal de parole dans le silence, la structure temporelle fine semble jouer un rôle majeur dans le bruit[37]. Ce point sera détaillé plus loin.

Écoute dans les vallées du bruit

Le démasquage est plus important pour un bruit masquant présentant des modulations sinusoïdales que pour des modulations reprenant l'enveloppe temporelle de la parole[38]. Nous avons vu qu'il augmente avec la profondeur de modulation du bruit modulé en amplitude[39]. Les fréquences de modulation optimales pour observer le phénomène se situent entre 8 et 25 Hz[40], ceci dépendant toutefois des signaux de paroles utilisés. De plus, le démasquage augmente lorsqu'on introduit des vallées spectrales dans un bruit stationnaire ou modulé temporellement en amplitude[41], particulièrement lorsque la profondeur de ces vallées spectrales augmente[42]. Par ailleurs, le démasquage est minimal pour un bruit masquant de type « locuteurs multiples » présentant une enveloppe temporelle et spectrale relativement plate, et se réduit encore si les portions les plus faibles de la parole présentée sont masquées par un bruit stationnaire[43].

Ces résultats démontrent que les sujets sont capables de tirer profit des *minima* – parfois extrêmement circonscrits sur le plan spectro-temporel – au sein du bruit afin de détecter des informations partielles de parole. Cette

35. *Cf.* Johnson (1980).
36. Shannon *et al.* (1995) ; Smith *et al.* (2002).
37. Nelson *et al.* (2003) ; Qin et Oxenham (2003) ; Hopkins *et al.* (2008).
38. Bacon *et al.* (1998).
39. Howard-Jones et Rosen (1993a) ; Gustafsson et Arlinger (1994).
40. Miller et Licklider (1950) ; Gustafsson et Arlinger (1994) ; Kwon et Turner (2001) ; Nelson *et al.* (2003) ; Füllgrabe *et al.* (2006).
41. Howard-Jones et Rosen (1993b) ; Buss *et al.* (2003) ; Buss *et al.* (2004).
42. Peters *et al.* (1998).
43. Eisenberg *et al.*, (1995) ; Bacon *et al.* (1998).

capacité est souvent appelée « écoute dans les vallées » (*glimpsing*). De plus, Howard-Jones et Rosen[44] ont démontré à l'aide de masques présentant des vallées spectrales et temporelles déphasées (asynchrones) que, dans une certaine mesure, les auditeurs normo-entendants peuvent regrouper des informations dans différentes régions fréquentielles et à différents instants pour démasquer au mieux la parole (capacité dénommée *uncomodulated glimpsing* par les auteurs).

De facto, tous ces mécanismes nécessitent un certain niveau de résolution temporelle (c'est-à-dire une capacité à suivre les fluctuations du masque afin d'extraire de l'information pendant les vallées) et spectrale (avoir accès à des portions du signal de parole qui ne sont pas ou peu masquées dans le domaine spectral). En rapport avec cette notion, on note une baisse du démasquage pour des fréquences de modulation du bruit masquant supérieures à 30 Hz (*i.e.*, des vallées de bruit de moins de 17 ms), cette baisse pouvant être mise sur le compte d'effets de masquage proactif, un facteur limitant la résolution temporelle auditive[45].

Toutefois, il est important de noter que la démonstration d'une stratégie d'écoute dans les vallées du bruit n'explique pas entièrement nos capacités de démasquage. En effet, une dernière question reste à élucider, à savoir : comment le système auditif s'y prend-il (*i.e.* quels indices auditifs sont utilisés par l'auditeur) pour savoir que des informations de parole se trouvent dans les vallées du bruit ? Des travaux récents réalisés grâce à des signaux de synthèse suggèrent fortement que les indices de structure temporelle fine (des disparités de hauteur fondamentale par exemple) jouent un rôle majeur dans ces distinctions entre parole et bruit au sein des vallées. En effet, quelle que soit la métrique utilisée ou la tâche, le démasquage est fortement dégradé voire aboli chez des auditeurs normo-entendants lorsque les signaux de parole et les bruits masquant sont dégradés par vocodeurs[46] au sein d'un nombre variable de bandes de fréquence (et particulièrement au sein des bandes de fréquence inférieures à 2 kHz environ) de façon à éliminer principalement les indices de structure temporelle fine[47].

L'effet d'une dégradation des informations de structure temporelle fine sur le démasquage est illustré en figure 6. Dans cette étude menée par Gnansia *et al.* (2008), les performances d'identification de consonnes au

44. Howard-Jones et Rosen (1993b).
45. Festen (1993) ; Dubno *et al.* (2002).
46. Dudley (1939, 1940).
47. Nelson *et al.*, (2003) ; Qin et Oxenham (2003) ; Fu et Nogaki (2005) ; Zeng *et al.* (2005) ; Füllgrabe *et al.* (2006) ; Lorenzi *et al.* (2006a) ; Gnansia *et al.* (2008) ; Hopkins *et al.* (2008).

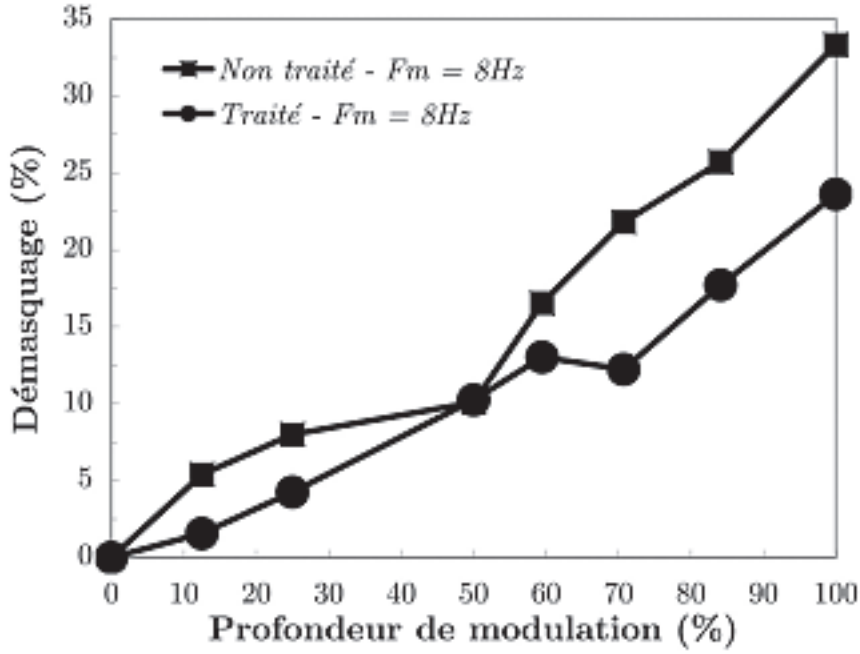


Figure 6 : *Démasquage de syllabes sans signification en fonction de la profondeur de modulation du bruit masquant. La fréquence de modulation du masque est fixée à 8 Hz. Les mesures sont réalisées sur les mélanges parole+bruit non traités (symboles pleins) et sur les mélanges traités par un vocodeur dégradant les indices de structure temporelle fine au sein de 32 bandes de fréquence (symboles ouverts). D'après Gnansia* et al. *(2008).*

sein de syllabes sans signification sont mesurées chez des auditeurs normo-entendants, en présence d'un bruit SSN stationnaire et d'un bruit SSN modulé sinusoïdalement en amplitude à 8 Hz. Ici, les performances d'identification sont mesurées à différentes profondeurs de modulation (de 0 à 100 %) du bruit SSN, avec un rapport signal-sur-bruit fixé individuellement afin de produire environ 50 % d'intelligibilité en condition de bruit stationnaire. Les signaux (parole et bruit) sont soit laissés intacts, soit traités par un vocodeur (*tone-excited vocoder*) utilisant 32 bandes fréquentielles d'analyse. Dans ce dernier cas, les signaux sont traités de façon à remplacer leur structure temporelle fine au sein de chacune des bandes par des sons purs à la fréquence centrale du filtre d'analyse. Conformément aux résultats discutés plus haut, les mesures indiquent que le démasquage augmente en fonction de la profondeur de modulation du bruit. De plus, ce démasquage est significativement réduit après dégradation des indices de structure temporelle fine, et ce, tout particulièrement aux profondeurs de modulation du bruit les plus élevées (> 50 %).

Conclusion

Ces derniers résultats suggèrent que nos capacités de démasquage de la parole – et plus précisément nos capacités d'écoute dans les vallées des bruits fluctuants – dépendent de façon critique des propriétés temporelles de décharge (*i.e.* des propriétés de synchronisation) des neurones auditifs. Plus généralement, l'étude du démasquage de la parole démontre l'importance (I) des indices acoustiques de modulation d'amplitude (enveloppe) et de fréquence (structure temporelle fine) au sein de la parole et des bruits masquant et (II) des mécanismes auditifs impliqués dans leur extraction. Ces résultats militent ainsi en faveur d'une conception de la parole comme signal « porteur » de modulations[48], et d'une conception du système auditif comme « système de démodulation »[49]. Au-delà du champ de la psychoacoustique, ces résultats devraient avoir des conséquences importantes dans les domaines de l'ingénierie de la parole (le développement de systèmes de reconnaissance automatique de la parole robustes par exemple), et de l'audiologie (nouveaux tests vocaux, dépistage des surdités neurosensorielles, évaluation des prothèses auditives et implants cochléaires)[50].

RÉFÉRENCES BIBLIOGRAPHIQUES

– Assmann P. F. et Summerfield Q. (1994), « The contribution of waveform interactions to the perception of concurrent vowels », *J. Acoust. Soc. Am.*, 95, p. 471-484.

– Bacon S. P., Opie J. M. et Montoya D. Y. (1998), « The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds », *J. Speech Lang. Hear. Res.*, 41, p. 549-563.

– Berry R. C. et Nerbonne G. P. (1972), « Comparison of the masking functions of speech-modulated and white noise », *J. Acoust. Soc. Am.*, 51, p. 121.

– Bronkhorst A. W. (2000), « The cocktail party phenomenon : A review of research on speech intelligibility in multiple-talker conditions », *Acustica*, 86, p. 117-128.

48. Dudley (1939, 1940).
49. Lorenzi *et al.* (2006a) ; Moore (2008).
50. *Remerciements*. Dan Gnansia est financé par une bourse CIFRE (ANRT/MXM-Neurelec). Certains travaux rapportés dans ce chapitre ont été soutenus par le GDR CNRS 2967 GRAEC.

– Bronkhorst A. W. et Plomp R. (1992), « Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing », *J. Acoust. Soc. Am.*, 92, p. 3132-3139.
– Buss E., Hall J. W. 3rd et Grose J. H. (2004), « Spectral integration of synchronous and asynchronous cues to consonant identification », *J. Acoust. Soc. Am.*, 115, p. 2278-2285.
– Buss E., Wall J. W. 3rd et Grose J. H. (2003), « Effect of amplitude modulation coherence for masked speech signals filtered into narrow bands », *J. Acoust. Soc. Am.*, 113, p. 462-467.
– Carhart R., Johnson C. et Goodman J. (1975), « Perceptual masking of spondees by combinations of talkers », *J. Acoust. Soc. Am.*, 58, S35.
– Cherry E. C. (1953), « Some experiments on the recognition of speech, with one and with two ears », *J. Acoust. Soc. Am.*, 25, p. 975-979.
– Cooke M. (2003), « Glimpsing speech », *J. Phonetics*, 31, p. 579-584.
– Cooke M. (2006), « A glimpsing model of speech perception in noise », *J. Acoust. Soc. Am.*, 119, p. 1562-1573.
– Cutler A., Weber A., Smits R. et Cooper N. (2004), « Patterns of English phoneme confusions by native and non-native listeners », *J. Acoust. Soc. Am.*, 116, p. 3668-3678.
– Danhauer J. L. et Leppler J. G. (1979), « Effects of four noise competitors on the California Consonant Test », *J. Speech. Hear. Disord.*, 44, p. 354-362.
– Dorman M. F., Loizou P. C., Fitzke J. et Tu Z. (1998), « The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6-20 channels », *J. Acoust. Soc. Am.*, 104, p. 3583-3585.
– Dubno J. R., Horwitz A. R. et Ahlstrom J. B. (2002), « Benefit of modulated maskers for speech recognition by younger and older adults with normal hearing », *J. Acoust. Soc. Am.*, 111, p. 2897-2907.
– Dubno J. R., Horwitz A. R. et Ahlstrom J. B. (2003), « Recovery from prior stimulation : masking of speech by interrupted noise for younger and older adults with normal hearing », *J. Acoust. Soc. Am.*, 113, p. 2084-2094.
– Dudley H. (1939), « The Vocoder », *Bell Labs. Rec.*, 17, p. 122-126.
– Dudley H. (1940), « The carrier nature of speech », *Bell System Tech. J.*, 19, p. 495-515.
– Duquesnoy A. J. (1983), « Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons », *J. Acoust. Soc. Am.*, 74, p. 739-743.
– Eisenberg L. S., Dirks D. D. et Bell T. S. (1995), « Speech recognition in amplitude-modulated noise of listeners with normal and listeners with impaired hearing », *J. Speech Hear. Res.*, 38, p. 222-233.
– Festen J. M. (1987), « Speech-perception threshold in a fluctuating background sound and its possible relation to temporal resolution », *in* M. E. H. Schouten (éd.), *The Psychophysics of Speech Perception*, Dordrecht, Martinus Nijhoff Publishers, 461-466.

– Festen J. M. (1993), « Contributions of comodulation masking release and temporal resolution to the speech-reception threshold masked by an interfering voice », *J. Acoust. Soc. Am.*, 94, p. 1295-1300.
– Festen J. M. et Plomp R. (1990), « Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing », *J. Acoust. Soc. Am.*, 88, p. 1725-1736.
– Fu Q. J. et Nogaki G. (2005), « Noise susceptibility of cochlear implant users : the role of spectral resolution and smearing », *J. Assoc. Res. Otolaryngol.*, 6, p. 19-27.
– Füllgrabe C., Berthommier F. et Lorenzi C. (2006), « Masking release for consonant features in temporally fluctuating background noise », *Hear. Res.*, 211, p. 74-84.
– George E. L., Festen J. M. et Houtgast T. (2006), « Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners », *J. Acoust. Soc. Am.*, 120, p. 2295-2311.
– Gilbert G. et Lorenzi C. (2006), « The ability of listeners to use recovered envelope cues from speech fine structure », *J. Acoust. Soc. Am.*, 119, p. 2438-2444.
– Gnansia D., Jourdes V. et Lorenzi C. (2008), « Effect of masker modulation depth on speech masking release », *Hear. Res.*, 239, p. 60-68.
– Gustafsson H. A. et Arlinger S. D. (1994), « Masking of speech by amplitude-modulated noise », *J. Acoust. Soc. Am.*, 95, p. 518-529.
– Hopkins K., Moore B. C. J. et Stone M. A. (2008), « Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech », *J. Acoust. Soc. Am.*, 123, p. 1140-1153.
– Horii Y., House A. S. et Hughes G. W. (1970), « Development and evaluation of a noise with speech-envelope characteristics », *J. Acoust. Soc. Am.*, 47, p. 75.
– Howard-Jones P. A. et Rosen S. (1993a), « The perception of speech in fluctuating noise », *Acustica*, 78, p. 258-272.
– Howard-Jones P. A. et Rosen S. (1993b), « Uncomodulated glimpsing in "checkerboard" noise », *J. Acoust. Soc. Am.*, 93, p. 2915-2922.
– Johnson D. H. (1980), « The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones », *J. Acoust. Soc. Am.*, 68, p. 1115-1122.
– Kwon B. J. et Turner C. W. (2001), « Consonant identification under maskers with sinusoidal modulation : masking release or modulation interference ? », *J. Acoust. Soc. Am.*, 110, p. 1130-1140.
– Licklider J. C. R. et Guttman N. (1957), « Masking of speech by line-spectrum interference », *J. Acoust. Soc. Am.*, 29, p. 287-296.
– Lorenzi C., Gilbert G., Carn H., Garnier S. et Moore B. C. J. (2006), « Speech perception problems of the hearing impaired reflect inability to use temporal fine structure », *Proc. Natl. Acad. Sci. USA*, 103, p. 18866-18869.
– Lorenzi C., Husson M., Ardoint M. et Debruille X. (2006), « Speech masking release in listeners with flat hearing loss : effects of masker fluctuation rate on identification scores and phonetic feature reception », *Int. J. Audiol.*, 45, p. 487-495.
– Miller G. A. (1947), « The masking of speech », *Psychol. Bull.*, 44, p. 105-129.

– Miller G. A. et Heise G. A. (1950), « The trill threshold », *J. Acoust. Soc. Am.*, 22, p. 637-638.
– Miller G. A. et Nicely P. E. (1955), « An analysis of perceptual confusions among some english consonants », *J. Acoust. Soc. Am.*, 27, p. 338-352.
– Moore B. C. (2008), « The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people », *J. Assoc. Res. Otolaryngol.*, 9, p. 399-406.
– Moore B. C. J. (2007), *Cochlear Hearing Loss : Physiological, Psychological and Technical Issues*, Chichester, Wiley.
– Nelson P. B. et Jin S. H. (2004), « Factors affecting speech understanding in gated interference : cochlear implant users and normal-hearing listeners », *J. Acoust. Soc. Am.*, 115, p. 2286-2294.
– Nelson P. B., Jin S. H., Carney A. E. et Nelson D. A. (2003), « Understanding speech in modulated interference : cochlear implant users and normal-hearing listeners », *J. Acoust. Soc. Am.*, 113, p. 961-968.
– Peters R. W., Moore B. C. J. et Baer T. (1998), « Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people », *J. Acoust. Soc. Am.*, 103, p. 577-587.
– Plomp R. et Mimpen A. M. (1979), « Improving the reliability of testing the speech reception threshold for sentences », *Audiology*, 18, p. 43-52.
– Qin M. K. et Oxenham A. J. (2003), « Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers », *J. Acoust. Soc. Am.*, 114, p. 446-454.
– Rhebergen K. S., Versfeld N. J. et Dreschler W. A. (2005), « Release from informational masking by time reversal of native and non-native interfering speech », *J. Acoust. Soc. Am.*, 118, p. 1274-1277.
– Rhebergen K. S., Versfeld N. J. et Dreschler W. A. (2006), « Extended speech intelligibility index for the prediction of the speech reception threshold in fluctuating noise », *J. Acoust. Soc. Am.*, 120, p. 3988-3997.
– Rosen S. (1992), « Temporal information in speech : acoustic, auditory and linguistic aspects », *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 336, p. 367-373.
– Shannon R. V., Zeng F. G., Kamath V., Wygonski J. et Ekelid M. (1995), « Speech recognition with primarily temporal cues », *Science*, 270, p. 303-304.
– Simpson S. A. et Cooke M. (2005), « Consonant identification in N-talker babble is a nonmonotonic function of N », *J. Acoust. Soc. Am.*, 118, p. 2775-2778.
– Smith Z. M., Delgutte B. et Oxenham A. J. (2002), « Chimaeric sounds reveal dichotomies in auditory perception », *Nature*, 416, p. 87-90.
– Snell K. B., Mapes F. M., Hickman E. D. et Frisina D. R. (2002), « Word recognition in competing babble and the effects of age, temporal processing, and absolute sensitivity », *J. Acoust. Soc. Am.*, 112, p. 720-727.
– Takahashi G. A. et Bacon S. P. (1992), « Modulation detection, modulation masking, and speech understanding in noise in the elderly », *J. Speech Hear. Res.*, 35, p. 1410-1421.
– Terhardt E. (1972), « Zur Tonhöhenwarhnemung von Klangen. I. Psychoakustiche Grundlagen », *Acustica*, 26, p. 173-186.

– Terhardt E. (1972), « Zur Tonhöhenwarhnemung von Klangen. II. Ein Funktionenschema », *Acustica*, 26, p. 187-199.
– Turner C. W., Gantz B. J., Vidal C., Behrens A. et Henry B. A. (2004), « Speech recognition in noise for cochlear implant listeners : benefits of residual acoustic hearing », *J. Acoust. Soc. Am.*, 115, p. 1729-1735.
– Versfeld N. J. et Dreschler W. A. (2002), « The relationship between the intelligibility of time-compressed speech and speech in noise in young and elderly listeners », *J. Acoust. Soc. Am.*, 111, p. 401-408.
– Zeng F. G., Nie K., Liu S., Stickney G., Del Rio E., Kong Y. Y. et Chen H. (2004), « On the dichotomy in auditory perception between temporal envelope and fine structure cues », *J. Acoust. Soc. Am.*, 116, p. 1351-1354.
– Zeng F. G., Nie K., Stickney G. S., Kong Y. Y., Vongphoe M., Bhargave A., Wei C. et Cao K. (2005), « Speech recognition with amplitude and frequency modulations », *Proc. Natl. Acad. Sci. USA*, 102, p. 2293-2298.

II

PARLER ET CHANTER

L'auto-organisation dans l'évolution de la parole

par Pierre-Yves Oudeyer

Les systèmes de vocalisation humains, véhicules physiques du langage, sont caractérisés par des formes et des propriétés structurales complexes. Ils sont combinatoires, basés sur la réutilisation systématique de phonèmes, et l'ensemble des répertoires de phonèmes des langues du monde est marqué à la fois par de fortes régularités statistiques, les universaux, et une grande diversité. En outre, ce sont des codes culturellement partagés par chaque communauté de locuteurs. Quelle est l'origine des formes de la parole ? Quels sont les mécanismes qui, au cours de la phylogenèse et de l'évolution culturelle, ont permis leur évolution ? Comment un code de la parole partagé peut-il se former dans une communauté d'individus ? Je vais m'intéresser dans ce chapitre à la manière dont les phénomènes d'auto-organisation et leurs interactions avec la sélection naturelle peuvent permettre d'éclairer ces trois questions.

La tendance qu'ont de nombreux systèmes physiques complexes à générer spontanément des formes nouvelles et organisées, comme les cristaux de glaces ou les spirales galactiques, est présente tout autant dans le monde inorganique que dans le monde vivant. Ainsi, l'explication de l'origine des formes du vivant ne peut reposer uniquement sur le principe de sélection naturelle, qui doit être complémenté par la compréhension des mécanismes de génération de formes nouvelles dans lesquels l'auto-organisation est centrale. Or, ceci s'applique aux formes sociales et culturelles du vivant, en particulier aux formes de la parole et du langage. Je commencerai par articuler de manière générale les relations entre auto-organisation, sélection naturelle et néodarwinisme pour la compréhension de la genèse des formes du vivant. J'instancierai ensuite ces relations dans le cadre des trois questions que j'ai énoncées ci-dessus. J'expliquerai alors pourquoi l'utilisation de simulations et de modèles informatiques est fon-

damentale pour faire progresser les théories qui y sont afférentes. Enfin, je présenterai un exemple d'expérimentation d'un modèle informatique qui montre que certains mécanismes simples de couplage sensori-moteur permettent de générer des systèmes de parole combinatoires, caractérisés par la dualité universaux/diversité, et partagés culturellement. Je conclurai par les scénarios évolutionnaires que cette expérimentation informatique vient complémenter ou renouveler.

Auto-organisation et évolution des formes du vivant

LA PHYSIQUE, CHAUDRON DE FORMES AUTO-ORGANISÉES

La nature regorge de formes et de motifs fascinants d'organisation, et en particulier dans sa partie inorganique. La silhouette des montagnes est la même que l'on regarde à l'échelle du rocher, du pic ou de la chaîne. Les dunes de sable s'alignent souvent en longues bandes parallèles. L'eau se cristallise en flocons symétriques et dentelés quand la température s'y prête. Et quand elle coule dans les rivières et tombe des cascades, apparaissent des tourbillons en forme de trompettes, tandis que les bulles se rassemblent en structures parfois polyhédrales. Les éclairs dessinent dans le ciel des ramifications à l'allure végétale. L'alternance de gel et de dégel sur les sols pierreux de la toundra laisse des empreintes polygonales sur le sol. La liste de ces formes rivalise de complexité avec bien des artefacts humains, comme on peut l'apprécier sur la figure 1. Et pourtant rien ni personne ne les a dessinées ou conçues. Pas même la sélection naturelle, le concepteur aveugle de Dawkins[1]. Quelle est donc leur origine ?

En fait, toutes ces structures organisées ont un point commun : elles sont le résultat macroscopique des interactions locales entre les nombreux composants du système dans lequel elles prennent forme. Leurs propriétés organisationnelles globales ne sont pas présentes au niveau local. En effet, la forme d'une molécule d'eau ainsi que ses propriétés physico-chimiques individuelles n'ont rien à voir avec celles des cristaux de glace, des tourbillons, ou encore des polyèdres de bulles. Les empreintes polygonales de la toundra ne correspondent pas à la forme des pierres qui les composent, et ont une organisation spatiale très différente de l'organisation temporelle du gel et du dégel. Voilà la marque de l'auto-organisation. Les phé-

1. Richard Dawkins, *The Blind Watchmaker*, New York, W. W. Norton & Company, 1986 (*L'Horloger aveugle*, Paris, Robert Laffont, 1989).

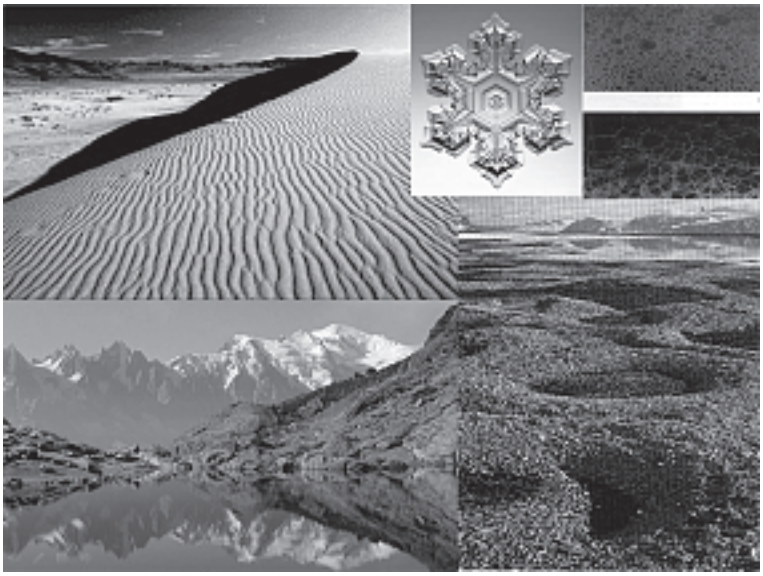


Figure 1 : *La nature regorge de formes et de motifs organisés sans qu'il y ait quelque part de plans qui aient servi à les construire : on dit qu'ils sont auto-organisés. Ici, des bandes parallèles qui courent sur les dunes, des bulles d'eau à la surface du liquide qu'on a agité, et les structures polyhédrales qui restent quand elles sèchent, un cristal de glace, des montagnes dont les formes sont les mêmes qu'on les regarde à l'échelle du rocher ou à l'échelle du pic.*

nomènes auto-organisés dans la nature caractérisent des systèmes physiques très variés, mais certaines propriétés typiques peuvent être identifiées. Non-linéarité, brisures de symétrie et « attracteurs » sont ainsi souvent présents. Par exemple, lorsqu'on chauffe par le dessous une fine couche d'huile s'étalant sur une surface plane, des courants de convection prenant des formes géométriques particulières (lignes ou polygones) s'auto-organisent et ces formes changent brutalement quand on passe certains seuils de température (voir figure 2). Entre ces seuils, au contraire, les formes restent globalement assez stables même si on les perturbe, constituant des attracteurs. Une autre propriété de nombreux systèmes auto-organisés est l'historicité, souvent associée à la sensibilité aux conditions initiales des systèmes chaotiques : l'attracteur dans lequel le système tombe, et qui contraint les formes produites par le système complexe, peut être très différent en fonction de petites variations des conditions initiales. C'est le cas par exemple de la magnétisation du fer : chacun des atomes d'une plaque de fer peut être vu comme une sorte d'aimant ayant plusieurs orientations possibles, laquelle est aléatoire si la température est élevée. Si elle passe en dessous d'un certain seuil, alors un phénomène d'auto-organisation se produit : tous les atomes adoptent spontanément

la même orientation magnétique. Cependant, cette orientation commune est quasiment imprédictible au départ, et de toutes petites variations aléatoires dans l'orientation initiale de quelques atomes peuvent faire se magnétiser la plaque dans une direction très différente. Or ces variations des conditions initiales sont typiquement liées à des événements contingents ayant interagi avec la plaque de métal : c'est pourquoi l'état final de la plaque dépend de son histoire en plus de ses propriétés physiques intrinsèques, d'où le terme d'historicité.

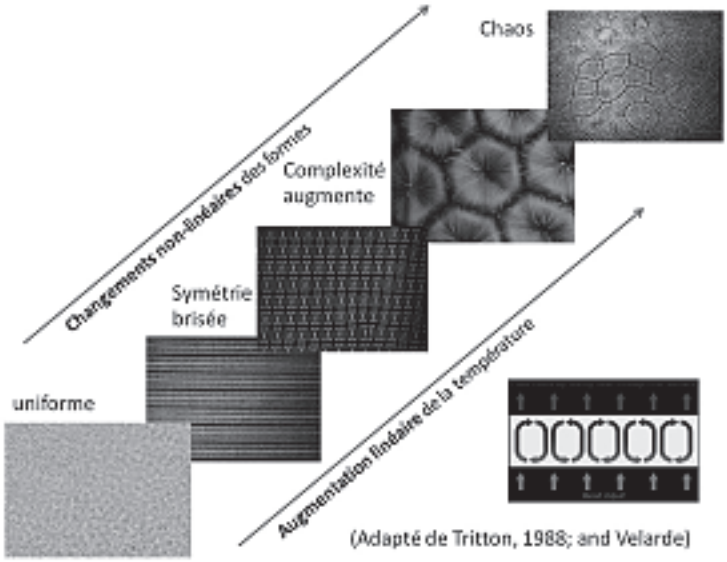


Figure 2 : *Cellules de Rayleigh-Bénard : si on chauffe par le dessous une fine couche d'huile s'étalant sur une surface plane, des courants de convections prenant des formes géométriques particulières (lignes, pavages hexagonaux) s'auto-organisent, et ces formes changent brutalement quand on passe certains seuils de températures. Ce type de non-linéarité caractérise de nombreux systèmes auto-organisés du monde inorganique comme du monde vivant*[2].

Le concept d'auto-organisation des systèmes complexes constitue la pierre de touche du changement paradigmatique que les sciences de la complexité ont opéré au XX^e^ siècle[3]. Depuis Newton, la bonne science se

2. Photos adaptées à partir de D. J. Tritton, *Physical Fluid Dynamics*, Oxford, Oxford University Press, 1988, et Manuel Velarde, Universidad Computense, Madrid.
3. William Ross Ashby, *An Introduction to Cybernetics*, Chapman & Hall, 1956 ; Grégoire Nicolis et Ilya Prigogine, *Self-Organization in Non-Equilibrium Systems : from Dissipative Structures to Order through Fluctuations*, New York, Wiley, 1977 ; Stuart Kauffman, *At Home in the Universe : the Search for Laws of Self-Organization and Complexity*, Oxford, Oxford University Press, 1996 ; Philip Ball, *The Self-Made Tapestry : Pattern Formation in Nature*, Oxford, Oxford University Press, 2001.

devait d'être réductionniste, et consistait à étudier les systèmes naturels en les décomposant en sous-systèmes plus simples. Par exemple, pour comprendre comment le corps humain fonctionnait, on se devait d'étudier d'un côté le cœur, de l'autre le système nerveux, et d'un autre côté encore par exemple le système limbique. D'ailleurs, on ne s'arrêtait pas là, et l'étude du système nerveux par exemple se devait d'être divisée en l'étude du cortex, du thalamus ou des innervations motrices périphériques. Cette méthode nous a évidemment permis d'apprendre une somme impressionnante de connaissances. Mais les chantres de la complexité l'ont battue en brèche. Leur credo : « le tout est plus que la somme des parties ».

L'IMPACT DE L'AUTO-ORGANISATION SUR L'ORIGINE DES FORMES DU VIVANT

Or les systèmes complexes, c'est-à-dire les systèmes composés de nombreux sous-systèmes en interaction, abondent dans la nature et ont la très forte tendance à s'auto-organiser. Les exemples de la partie précédente ont été volontairement choisis parmi les systèmes inorganiques pour montrer que la propriété d'auto-organisation peut caractériser des systèmes dont les mécanismes n'ont rien à voir avec celui de la sélection naturelle. Cependant, l'auto-organisation s'applique de la même manière aux systèmes vivants. C'est d'ailleurs un concept largement utilisé dans plusieurs champs de la biologie. Il est en particulier central aux théories qui expliquent les facultés des sociétés d'insectes à construire des nids ou des ruches, à chasser en groupe ou explorer de manière décentralisée et efficace les ressources de nourriture de l'environnement[4]. En biologie du développement, il sert aussi, par exemple, à expliquer la formation des patterns colorés sur la peau des animaux comme les papillons, les zèbres, les jaguars ou les coccinelles[5].

Il semble donc qu'il soit possible qu'il y ait dans les systèmes biologiques des mécanismes créateurs de formes et patterns qui soient orthogonaux à la sélection naturelle, et ceci grâce à leur propriété d'auto-organisation. Or la sélection naturelle, et les explications fonctionnalistes qui lui sont associées, constitue le cœur de la majorité des argumentations en biologie quand il s'agit d'expliquer la présence d'une structure, d'une forme ou d'un pattern dans un organisme. Quelle est donc l'articulation entre la théorie de la sélection naturelle et l'auto-organisation ?

---

4. Scott Camazine, Jean-Louis Deneubourg, Nigel R. Franks, James Sneyd, Guy Theraulaz et Éric Bonabeau, *Self-Organization in Biological Systems*, Princeton, Princeton University Press, 2002.
5. Philip Ball, *op. cit.*

Certains chercheurs ont proposé l'idée que l'auto-organisation remettait en question le rôle central de la sélection naturelle dans l'explication de l'évolution des organismes vivants. Waldrop explique :

> Les systèmes dynamiques complexes peuvent parfois passer spontanément d'un état de désordre à un état d'ordre ; est-ce une force motrice de l'évolution ? Avons-nous manqué quelque chose à propos de l'évolution – un principe clé qui a contrôlé le développement de la vie de manière différente de la sélection naturelle, des dérives génétiques, et de tous les autres mécanismes que les biologistes ont invoqué au cours des années ? Oui ! Et l'élément manquant est l'auto-organisation spontanée : la tendance qu'ont les systèmes dynamiques complexes à se placer dans des états ordonnés sans qu'il soit besoin d'aucune pression de sélection[6].

Cependant, ce n'est pas la position qui est prise dans cet article. Plutôt que de voir l'auto-organisation comme un concept qui minimise le rôle de la sélection naturelle en proposant des mécanismes créateurs de formes concurrents, il est plus exact de le voir comme, d'une part, correspondant à un niveau d'explication différent et surtout, d'autre part, comme décrivant des mécanismes qui décuplent la puissance de la sélection naturelle. Les mécanismes ayant la propriété d'auto-organisation sont donc complètement intégrables au mécanisme de la sélection naturelle pour l'explication de l'évolution des formes du vivant.

L'APPROCHE CLASSIQUE DU NÉODARWINISME

Pour le voir précisément, il faut d'abord rappeler en quoi consiste le mécanisme de la sélection naturelle selon le néodarwinisme. Il caractérise un système composé d'individus ayant chacun des traits, formes ou structures particuliers. Ensuite, les individus de ce système sont capables de se répliquer. Cette réplication doit parfois générer des individus qui ne sont pas les exactes copies de leurs ancêtres, mais des petites variations. Ce sont ces variations qui sont à la source de la diversité des individus. Enfin, chaque individu a la capacité de se répliquer plus ou moins facilement selon sa structure et l'environnement qui l'entoure. Cette réplication différentielle des individus donne lieu à une « sélection » de ceux qui sont le plus aptes à se répliquer. La combinaison du processus de variation et du processus de sélection fait qu'au cours des générations, les structures ou traits des individus qui les aident à se reproduire sont conservés et améliorés.

---

6. M. Waldrop (1990), « Spontaneous order, evolution, and life », *Science*, 247, p. 1543-1545.

Il est un point crucial sur lequel la théorie de la sélection naturelle reste neutre : c'est la manière dont sont générées les variations, et plus généralement la manière dont sont générés les individus, avec leurs formes, leurs traits et leurs structures. Certains arguments néodarwiniens considèrent les mécanismes de variations des formes comme secondaires par rapport aux avantages reproductifs de ces formes quand il s'agit d'expliquer leur évolution. Cette mise au second plan implique implicitement que la relation entre le niveau des gènes, considéré comme l'espace principal dans lequel s'opèrent les variations par mutations et réarrangements, et le niveau du phénotype, considéré comme une image isomorphe de l'espace des gènes, est simple et linéaire. Selon cette vision, l'exploration de l'espace des formes (qui déterminent, avec l'environnement, le degré d'efficacité de réplication des gènes) peut être comprise simplement en regardant la manière dont on se déplace dans l'espace des génomes. Or les mécanismes de mutation qui permettent justement ces déplacements sont de petite amplitude (la plupart des mutations n'affectent qu'une toute petite partie des génomes quand il y a réplication), et donc des variations aléatoires des gènes permettent d'explorer uniformément tout l'espace des génomes. Ce qui veut dire que, dans l'hypothèse où les deux espaces génotypiques et phénotypiques ont la même structure, alors l'espace des formes est exploré de manière quasi continue, par petites modifications successives des formes préexistantes. Pour l'apparition des formes de vie complexes, ce n'est heureusement pas le cas. En effet, si ce mécanisme de petites variations successives des formes était efficace pour le réglage fin des structures des organismes, il rendrait la recherche des formes aussi complexe que celle des organismes humains équivalente à la recherche d'une aiguille dans une botte de foin car les génomes sont de trop grande dimension[7].

L'AUTO-ORGANISATION CONTRAINT L'ESPACE DES FORMES À EXPLORER : TOUTES LES FORMES NE SONT PAS ÉGALEMENT FACILES À FAIRE ÉMERGER

C'est là que le concept d'auto-organisation vient à l'aide de ce mécanisme d'exploration naïf de l'espace des formes dans le cadre de la théorie néodarwinienne. En effet, la relation entre les gènes et les formes des organismes se caractérise par sa complexité et sa forte non-linéarité, qui s'expriment pendant le développement ontogénétique et épigénétique de chaque organisme. En fait, les organismes sont construits à partir d'une cellule souche qui contient un génome, et cette cellule souche peut être

7. Keefe et Szostak (2001).

vue comme un système dynamique paramétré par son génome et sous l'influence des perturbations imposées par l'environnement. Ce système dynamique constitue un système auto-organisé, qui a le même type de propriétés que les systèmes auto-organisés présentés dans la partie précédente. Le génome est un ensemble de paramètres analogues à la température des liquides dans les systèmes de Bénard, et l'environnement est l'analogue du bruit pour la magnétisation du fer (mais c'est un bruit évidemment très structuré et structurant !). Ainsi, le développement d'un organisme à partir de sa cellule souche partage un certain nombre de propriétés avec la formation auto-organisée des systèmes physiques : des formes, des structures et des patterns apparaissent au niveau global, et sont qualitativement différentes de ceux qui définissent le fonctionnement local, c'est-à-dire différents de ceux qui caractérisent la structure de la cellule souche et de son génome. Le pavage hexagonal qui peut apparaître à partir d'une simple différence de température dans un liquide homogène donne une idée de la manière dont une simple suite de nucléotides, entourée d'un système moléculaire qui les transforme automatiquement en protéines, peut générer un organisme bipède doté de deux yeux pour voir, d'oreilles, et d'un cerveau immensément complexe.

En outre, comme les systèmes de Bénard ou les plaques ferromagnétiques, les systèmes dynamiques définis par les génomes et les cellules qui les contiennent sont caractérisés par un paysage d'attracteurs : il y a de larges zones de l'espace des paramètres pour lesquelles le système dynamique adopte systématiquement un comportement dont la structure reste à peu près la même. Pour les systèmes de Bénard, il y a une plage de température qui permet de générer des bandes parallèles et qui est assez large pour qu'on puisse la trouver facilement. Pour les plaques ferromagnétiques, la zone de température dans laquelle le système arrive à trouver une cohérence magnétique globale est aussi très large. Ainsi, pour les organismes vivants, non seulement il est possible de générer des structures auto-organisées aux propriétés globales complexes mais, en plus, ces structures sont certainement générées par les génomes appartenant à de larges sous-espaces dans l'espace des génomes, qu'on appelle des bassins d'attraction. La structuration de l'espace en bassins d'attraction de ce genre de système dynamique permet donc ainsi de faire que l'exploration de l'espace des formes soit facilitée et ne s'apparente pas à la recherche d'une aiguille dans une botte de foin.

Comme dans les systèmes ferromagnétiques, le bruit (structuré) imposé par l'environnement sur le développement du système dynamique peut conduire à ce qu'il prenne des voies de développement différentes. Pour les morceaux de fer à basse température, cela correspondait à la

magnétisation dans un sens ou dans l'autre. Pour un organisme vivant, cela correspond à ses formes : c'est ainsi qu'il arrive que même des vrais jumeaux aient des différences morphologiques assez importantes. Ceci illustre aussi pourquoi la relation entre les gènes et les formes des organismes ne se caractérise pas seulement par sa complexité et sa non-linéarité, mais aussi par son non-déterminisme. Comme dans les systèmes de Bénard, où l'exploration de l'espace du paramètre de température peut parfois conduire à des changements rapides et conséquents du comportement du système (par exemple, le passage des bandes parallèles aux cellules carrées), l'exploration de l'espace des génomes peut aussi conduire parfois à des changements de formes rapides et conséquents. Cela correspond possiblement à de nombreuses observations de changements de formes très rapides dans l'évolution, comme en témoignent les fossiles qu'étudient les anthropologues, et qui sont à la base de la théorie des équilibres ponctués proposée par Eldredge et Gould en 1972.

Pour résumer, les propriétés d'auto-organisation du système dynamique composé par les cellules et leurs génomes apportent une structuration cruciale à l'espace des formes en le contraignant, ce qui rend la découverte de formes complexes et robustes beaucoup plus facile pour la sélection naturelle. D'une part, elles permettent à un génome de générer des formes complexes et très organisées sans qu'il soit besoin d'en spécifier précisément chaque détail dans le génome (de la même manière que les formes polygonales de Bénard ne sont pas spécifiées précisément, ou sous la forme d'un plan, dans les propriétés des molécules de liquide). D'autre part, elles organisent le paysage de ces formes possibles en bassins d'attraction à l'intérieur desquels celles-ci se ressemblent beaucoup (c'est là que se font les évolutions graduelles, avec des réglages fins des structures existantes), et entre lesquelles les formes peuvent différer substantiellement (c'est le passage de l'un à l'autre qui peut provoquer des inventions soudaines et puissantes dans l'évolution). Pour donner une image simple, l'auto-organisation fournit un catalogue de formes complexes réparties dans un paysage de vallées dans lesquelles et entre lesquelles la sélection naturelle se déplace et fait son choix : l'auto-organisation propose, la sélection naturelle dispose. Évidemment, ceci n'est qu'une image, car la sélection naturelle par son déplacement permet justement de faire apparaître de nouveaux mécanismes eux-mêmes auto-organisés, qui structurent l'espace des formes dans lequel elle se déplace ; donc la sélection naturelle participe à la formation de ces mécanismes qui l'aident à avoir un déplacement efficace dans l'espace des formes ; vice versa, le mécanisme de la sélection naturelle est certainement apparu dans l'histoire de la vie grâce aux comportements auto-organisés de systèmes qui étaient

encore complètement étrangers à la sélection naturelle ; la sélection naturelle et les mécanismes auto-organisés s'aident donc mutuellement dans une sorte de spirale qui permet à la complexité d'augmenter au cours de l'évolution.

La conséquence de cette intrication entre sélection naturelle et auto-organisation est que l'explication de l'origine et de l'évolution des formes du vivant nécessite au moins deux types d'argumentaires. Le premier, classique, est l'argumentaire fonctionnaliste néodarwiniste : il consiste à identifier le contexte écologique dans lequel un trait nouveau a pu apparaître et d'en articuler le bilan confrontant les coûts et les avantages reproductifs. Le second argumentaire est plus rarement utilisé, mais tout aussi essentiel : il s'agit d'identifier les mécanismes développementaux/épigénétiques, et les contraintes qui y sont associées, qui ont pu faciliter, ou au contraire rendre difficile, la genèse de ces traits nouveaux. Or la notion d'auto-organisation est centrale dans la manière dont ces mécanismes développementaux impactent la genèse des formes.

Auto-organisation et évolution des formes du langage et des langues

La question de savoir comment parole et langage sont venus à l'être humain et celle de savoir comment des langues nouvelles se forment et évoluent sont parmi les plus difficiles qui soient posées à la science. Alors qu'elles ont été écartées de la scène scientifique pendant la presque totalité du XX^{e} siècle, à la suite de la déclaration de la Société linguistique de Paris qui les avait bannies de sa constitution, ces questions sont revenues au centre des recherches de toute une communauté de scientifiques. Un consensus se dégage de la communauté des chercheurs qui aujourd'hui s'y attellent : la recherche doit être multidisciplinaire. En effet, c'est un puzzle aux ramifications immenses qui dépassent les compétences de chaque domaine de recherche pris indépendamment. C'est d'abord parce que les deux grandes questions, celle de l'origine du langage, d'une part, et celle de l'origine et de l'évolution des langues, d'autre part, doivent être décomposées en sous-questions elles-mêmes déjà fort complexes : Qu'est-ce que le langage ? Qu'est-ce qu'une langue ? Comment s'articulent entre eux les sons, les mots, les phrases, les représentations sémantiques ? Comment le cerveau représente-t-il et manipule-t-il ces sons, ces phrases, et les concepts que celles-ci véhiculent ? Comment apprend-on à parler ? Peut-on discerner l'inné de l'acquis ? À quoi sert le langage ? Quel est son rôle social ? Comment une langue se forme-t-elle et

change-t-elle au cours des générations successives de ses locuteurs ? Que sait-on de l'histoire de chaque langue ? Pourquoi le langage et les langues sont-ils tels qu'ils sont ? Pourquoi y a-t-il des tendances universelles et, en même temps, une grande diversité des structures linguistiques ? Quelle est l'influence du langage sur la perception et la conception du monde ? Que sait-on de l'histoire de la capacité de parler chez les humains ? Est-ce plutôt le résultat d'une évolution génétique (comme l'apparition des yeux) ou une invention culturelle (comme l'écriture) ? Est-ce une adaptation à un environnement changeant ? Une modification interne de l'individu qui a permis d'augmenter ses chances de reproduction ? Est-ce une exaptation, effet collatéral de changements qui n'étaient pas initialement reliés au comportement de communication ? Quels sont les prérequis évolutionnaires qui ont permis l'apparition de la capacité de parler ? Comment eux-mêmes sont-ils apparus ? Indépendamment ? Génétiquement ? Culturellement ?

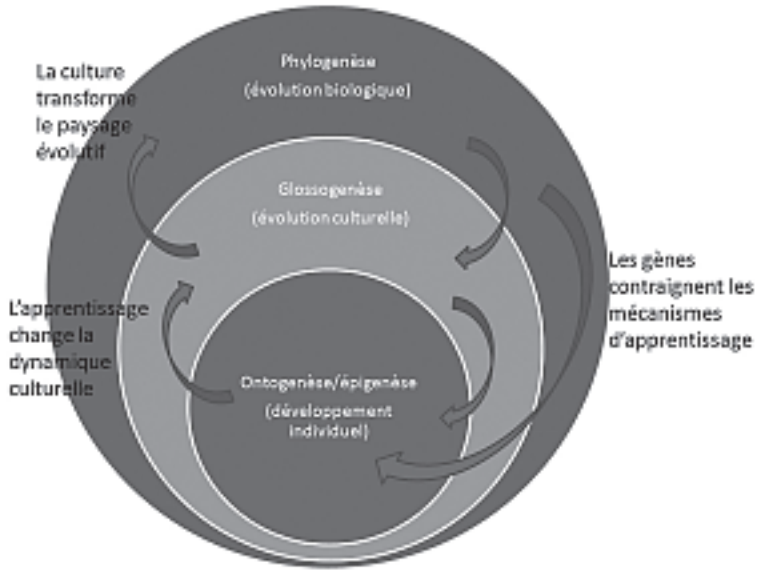


Figure 3 : *Les multiples échelles d'interactions à l'origine du langage.*

Face à la diversité de ces questions se dresse une diversité encore plus grande de disciplines et de méthodes. Les linguistes, même s'ils continuent à fournir des données cruciales sur l'histoire des langues ainsi que sur les tendances universelles de leurs structures, ne sont plus les seuls acteurs. La psychologie développementale, la psychologie cognitive et la neuropsychologie font des études comportementales de l'acquisition du

langage ainsi que des troubles du langage, souvent révélateurs des mécanismes cognitifs qui sont impliqués dans le traitement du langage. Les neurosciences, en particulier avec les dispositifs d'imagerie cérébrale qui permettent de visualiser quelles zones du cerveau sont actives quand on effectue une tâche donnée, essaient de trouver les corrélats neuronaux des comportements de parole, pour en découvrir l'organisation cérébrale. D'autres chercheurs étudient la physiologie de l'appareil vocal, pour essayer de comprendre la manière dont nous produisons des sons. La physiologie de l'oreille, capteur essentiel dans la chaîne de décodage de la parole (c'est la vision quand il s'agit de la langue des signes), est aussi au centre des recherches. Les archéologues examinent les fossiles et les artefacts qui sont les restes des premiers hommes ; ils tentent, d'une part, d'en déduire l'évolution morphologique des hommes (en particulier de son larynx) et, d'autre part, de se faire une idée des activités qu'ils pratiquaient (quels outils fabriquaient-ils ? comment les utilisaient-ils ? comment ces outils peuvent-ils nous renseigner sur le degré de développement cognitif ?). Les anthropologues vont à la rencontre des peuples isolés, et rendent compte des différences culturelles, en particulier celles liées aux langues et aux conceptions qu'elles véhiculent. Les primatologues essaient de rendre compte des capacités de communication de nos ancêtres les chimpanzés, et de les comparer aux nôtres. D'une part, les généticiens séquencent les génomes de l'homme et des espèces qui sont ses ancêtres potentiels, quand c'est possible, pour préciser leurs liens phylogénétiques, et, d'autre part, ils utilisent les informations génétiques des différents peuples de la planète pour aider à la reconstruction de l'histoire des langues, qui est souvent corrélée avec l'histoire des gènes des locuteurs qui les parlent.

Le langage implique donc une multitude de composantes qui interagissent de manière complexe sur plusieurs échelles de temps en parallèle : l'échelle ontogénétique, qui caractérise le développement de l'individu, l'échelle glosso-génétique ou culturelle, qui caractérise l'évolution des cultures, et l'échelle génétique, qui caractérise l'évolution des espèces (voir figure 3). En particulier, le langage met en œuvre des interactions complexes, à la fois physiques et fonctionnelles, entre de multiples circuits cérébraux, plusieurs organes, les individus qui en sont équipés et l'environnement dans lequel ils vivent. Or, comme on l'a vu dans la section précédente, s'il est fondamental d'étudier chacune de ces composantes indépendamment, afin de réduire la complexité du problème, il est aussi nécessaire d'en étudier les interactions. Un certain nombre de chercheurs ont en effet proposé l'idée selon laquelle de nombreuses propriétés du langage et des langues ne sont codées dans aucun de leurs composants, c'est-à-dire ni dans certaines structures cérébrales spécifiques, ni dans les pro-

priétés des appareils auditifs ou vocaux, ou ni même chez un individu considéré indépendamment, mais pourraient être des résultats auto-organisés de l'interaction complexe et dynamique de ces composants et de ses individus[8]. Or ces phénomènes d'auto-organisation sont souvent compliqués à comprendre, à prévoir intuitivement, et à formuler verbalement ; d'où le recours de plus en plus important à la modélisation mathématique et informatique comme nous allons le voir.

Modèles et simulations informatiques de l'évolution du langage

EXPÉRIMENTER LES SYSTÈMES COMPLEXES

L'une des manières les plus efficaces aujourd'hui pour développer notre compréhension de la dynamique des systèmes auto-organisés est l'utilisation des ordinateurs ou des robots. En effet, ils permettent d'élaborer des modèles opérationnels dont on connaît toutes les hypothèses, de les faire fonctionner, et d'en observer le comportement selon les valeurs des paramètres fixés dans le cadre de ces modèles. C'est pourquoi, en plus des linguistes, des psychologues, des anthropologues, des chercheurs en neurosciences, des généticiens et des physiologistes, les mathématiciens et les informaticiens/roboticiens ont désormais un rôle crucial dans cette recherche.

Un modèle opérationnel est un système qui définit formellement l'ensemble de ses présuppositions et surtout qui permet de calculer ses conséquences, c'est-à-dire de prouver qu'il mène à un ensemble de conclusions données. Il existe deux grands types de modèles opérationnels. Le premier, celui utilisé par les mathématiciens et certains biologistes théoriciens, consiste à abstraire du phénomène du langage un certain nombre de variables et leurs lois d'évolution sous la forme d'équations mathématiques. Cela ressemble le plus souvent à des systèmes d'équations différentielles couplées, et bénéficie du cadre de la théorie des systèmes dynamiques. Le second type, qui permet d'étudier certains phénomènes complexes qui se prêtent difficilement à la modélisation mathématique,

8. James R. Hurford, Michael Studdert-Kennedy et Chris Knight, *Approaches to the Evolution of Language : Social and Cognitive Bases*, Cambridge, Cambridge Universtity Press, 1998 ; B. Lindblom, P. MacNeilage et M. Studdert-Kennedy, « Self-organizing processes and the explanation of language universals », *in* Brian Butterworth, Bernard Comrie et Osten Dahl (éd.), *Explanations for Language Universals*, 1984, p. 181-203 ; Pierre-Yves Oudeyer, *Self-Organization in the Evolution of Speech*, Oxford, Oxford University Press, 2006.

est celui utilisé par les chercheurs en informatique : il consiste à construire des systèmes artificiels implantés dans des ordinateurs ou sur des robots. Ces systèmes artificiels sont composés de programmes qui, le plus souvent, prennent la forme d'agents artificiels dotés de cerveaux et de corps artificiels ; on pourra les appeler robots même s'ils évoluent dans des environnements virtuels. Ces robots sont alors mis en interaction dans un environnement artificiel ou réel, et on peut étudier leur dynamique. C'est ce qu'on appelle la « méthode de l'artificiel ». L'utilisation de machines computationnelles pour simuler et étudier les phénomènes naturels n'est d'ailleurs pas nouvelle : Lorenz a utilisé les premiers ordinateurs pour étudier le comportement de modèles climatologiques, Fermi pour simuler l'interaction entre des particules magnétisées, Turing pour imaginer comment les processus de morphogenèse pouvaient s'auto-organiser, von Neumann pour étudier l'auto-réplication.

Plus récemment et grâce à cette méthode, l'éthologie a fait un bon en avant pour la compréhension des comportements et des performances des insectes sociaux[9]. Des simulations informatiques de sociétés d'insectes, basées sur le concept d'agents informatiques modélisant chaque insecte individuellement – ce qu'on appelle parfois modèles individuo-centrés –, ont été construites. Cela a permis d'établir des caractéristiques suffisantes du comportement et des capacités des insectes pour observer la formation de structures collectives, comme la construction des nids chez les termites, la formation de collectifs de chasse ou de recherche de nourriture chez les fourmis, la formation des bancs de poissons, la thermorégulation dans les ruches des abeilles ou la formation de structures sociales chez les guêpes. De manière générale, ces simulations informatiques ont montré qu'il n'était souvent pas nécessaire que les insectes soient équipés de structures cognitives complexes pour que, collectivement, ils forment des structures complexes. Ces modèles informatiques sont mêmes parvenus à faire des prédictions qui ont été vérifiées par la suite sur le terrain.

Les physiciens utilisent aussi de plus en plus l'ordinateur pour construire des simulations de systèmes complexes qui leur permettent de développer leurs intuitions. En manipulant des automates cellulaires – sortes de grilles dont les cases peuvent être allumées ou éteintes et dont l'évolution dépend de l'état de leurs voisines selon des règles simples –, ils ont découvert comment, à partir de structures soit complètement aléatoires soit complètement ordonnées, des motifs complexes avec des symétries non triviales pouvaient se former. Les exemples sont fort divers : les cris-

9. E. Bonabeau, G. Theraulaz, J. L. Deneubourg, S. Aron et S. Camazine, (1997), « Self-organization in social insects », *Trends in Ecology and Evolution*, 12, p. 188-193.

taux de glace, les distributions des avalanches dans les tas de sables ou dans les montagnes, les dunes dans le désert, les formes des deltas fluviaux, la formation des galaxies ou celle des polyèdres de bulles au pied des cascades. Pour les physiciens, les automates cellulaires ne sont évidemment pas à proprement parler des modèles physiques des cristaux de glace ou des avalanches, mais ils ont joué un rôle de métaphore et d'analogie, qui a déclenché un renouvellement de la manière dont ils percevaient ces phénomènes.

INFORMATIQUE ET ORIGINES DU LANGAGE ET DES LANGUES

Il est également possible d'utiliser les ordinateurs et les simulations à base d'agents non seulement pour nous aider à comprendre les phénomènes qui caractérisent l'auto-organisation de la matière, des structures biologiques simples, ou des sociétés d'insectes, mais aussi pour l'étude des phénomènes qui caractérisent l'homme et ses sociétés. Le temps est venu de faire entrer l'ordinateur et les robots parmi les outils des sciences humaines. Ainsi, la construction de systèmes artificiels dans le cadre de la recherche sur les origines du langage et de l'évolution des langues bénéficie d'une popularité grandissante dans la communauté scientifique en tant qu'outil pour étudier les phénomènes du langage liés à l'interaction complexe de ses composants[10].

Il y a deux grands types d'utilisation de ces systèmes : 1) ils servent à évaluer la cohérence interne des théories verbales déjà proposées, en clarifiant toutes les hypothèses et en vérifiant qu'elles mènent bien aux conclusions proposées (et, souvent, on découvre des failles dans les présupposés ainsi que dans les conclusions, qui doivent être révisées) ; 2) ils servent à engendrer de nouvelles théories ou à explorer celles qui, souvent, apparaissent d'elles-mêmes quand on essaie tout simplement de construire un système artificiel qui reproduit les comportements de parole des humains.

Un certain nombre de résultats décisifs ont déjà été obtenus et ont permis d'ouvrir la voie à la résolution de questions jusque-là sans réponses : la génération décentralisée de conventions lexicales et sémantiques dans des communautés de robots[11], la formation de répertoires partagés de voyelles ou de syllabes dans des sociétés d'agents, avec des propriétés de régularités structurelles qui ressemblent beaucoup à celles des langues

10. L. Steels (1997), « The synthetic modeling of language origins », *Evolution of Communication*, 1 (1), p. 1-35.
11. L. Steels, art. cit. ; Frédéric Kaplan, *La Naissance d'une langue chez les robots*, Paris, Hermès, 2001.

humaines[12], la formation de conventions syntaxiques[13] ou les conditions dans lesquelles la compositionalité peut être sélectionnée[14].

Il est important de noter que, dans le cadre de la recherche sur les origines du langage, cette méthodologie de l'artificiel est avant tout une *méthodologie exploratoire*. Elle s'insère dans une logique scientifique d'abduction, c'est-à-dire une logique dans laquelle on cherche des prémisses qui peuvent mener à une conclusion donnée (au contraire de la déduction dans laquelle on cherche les conclusions auxquelles peuvent mener des prémisses données).

Le mot « modèle » a ici un sens différent de son acception traditionnelle. Selon cette dernière, modéliser consiste à observer un phénomène naturel, puis à essayer d'en abstraire les mécanismes et les variables fondamentales pour construire à partir d'elles un formalisme capable de prédire précisément la réalité. Dans le cas qui nous intéresse, il s'agit plutôt de s'interroger qualitativement sur les types de mécanismes que la nature a pu mettre en œuvre pour résoudre tel ou tel problème. En effet, le langage est un phénomène tellement complexe que la simple observation ne permet pas de *déduire* des mécanismes explicatifs. Au contraire, il est nécessaire d'avoir au préalable une bonne conceptualisation de l'espace des mécanismes et des hypothèses qui pourraient expliquer les phénomènes complexes du langage. Et c'est là le rôle des systèmes artificiels, ceux qu'on appelle parfois « modèles » : développer notre intuition sur les dynamiques de formation du langage et des langues, et ébaucher l'espace des hypothèses.

Il ne s'agit donc pas d'établir la liste des mécanismes responsables de l'origine de tel ou tel aspect du langage. L'objectif est plus modestement d'essayer de faire une liste des candidats possibles, de contraindre l'espace des hypothèses, en particulier en montrant des exemples de mécanismes qui sont suffisants et des exemples de mécanismes qui ne sont pas nécessaires.

---

12. Bart de Boer, *The Origins of Vowel Systems*, Oxford, Oxford University Press, 2001. P.-Y. Oudeyer, « Origins and learnability of syllable systems, a cultural evolutionary model », *in* Pierre Collet, Cyril Fonlupt, Jin-Kao Hao, Evelyne Lutton et Marc Schoenauer (éd.), *Artificial Evolution, 5th International Conference Proceedings 2001*, Berlin, Springer-Verlag, 2002, p. 143-155.

13. J. Batali, « Computational simulations of the emergence of grammar », *in* Hurford *et al.*, *op. cit.*

14. S. Kirby (2001), « Spontaneous evolution of linguistic structure. An iterated learning model of the emergence of regularity and irregularity », *IEEE Transactions on Evolutionary Computation*, 5 (2), p. 102-110.

Le code de la parole

Je vais maintenant illustrer ce travail de modélisation informatique de l'évolution du langage et des langues par la description d'une expérimentation qui se focalise sur le problème de l'origine de la parole, c'est-à-dire les systèmes de sons en tant que véhicules et supports physiques du langage (au même titre que peuvent l'être les gestes dans les langues des signes). L'objectif de cette expérimentation est de participer à la reconceptualisation de ce problème en explicitant, en évaluant et en ouvrant plusieurs hypothèses scientifiques.

DIGITALITÉ ET COMBINATORIALITÉ

Les humains ont un système de vocalisations complexe. Celles-ci sont digitales et combinatoriales, c'est-à-dire qu'elles sont construites à partir d'unités élémentaires, « sculptées » dans un continuum auditif et vocal, et systématiquement recombinées, puis réutilisées dans les vocalisations. Ces unités sont présentes à plusieurs niveaux (par exemple : les primitives motrices d'obstruction du flux de l'air dans le conduit vocal, qu'on appelle gestes ; les coordinations de gestes, que l'on appelle phonèmes et dont font partie les consonnes et les voyelles ; les syllabes). Alors que l'espace articulatoire est continu et permet potentiellement une infinité de gestes et de phonèmes, chaque langue discrétise cet espace à sa manière en utilisant un répertoire de gestes et de phonèmes à la fois petit et fini[15]. C'est pourquoi on appelle aussi parfois cette propriété le *codage phonémique*.

UNIVERSAUX ET DIVERSITÉ

En outre, malgré la grande diversité de ces unités dans les langues du monde, on y rencontre en même temps de fortes régularités. Par exemple, certains systèmes de voyelles sont beaucoup plus fréquents que d'autres, comme le système à cinq voyelles *e, i, o, a, u*. Il en va de même pour les consonnes. La manière dont les unités sont combinées est aussi très particulière : d'une part toutes les séquences de phonèmes ne sont pas autorisées dans une langue donnée, d'autre part l'ensemble des combinaisons de

15. M. Studdert-Kennedy et L. Goldstein (2003), « Launching language : the gestural origin of discrete infinity », *in* Morten Christiansen et Simon Kirby (éd.), *Language Evolution : the State of the Art*, p. 235-254, Oxford, Oxford University Press.

phonèmes est organisé en types génériques. Cette organisation en types génériques veut dire qu'on peut, par exemple, résumer les combinaisons de phonèmes autorisées en Japonais pour former des syllabes (« moras » plus exactement) par les types « CV/CVC/VC », où par exemple « CV » est un type qui désigne les syllabes composées de deux emplacements, avec dans le premier emplacement uniquement des phonèmes de la catégorie que l'on appelle « consonnes », alors que dans le second emplacement seuls les phonèmes de la catégorie « voyelles » sont autorisés.

PARTAGE CULTUREL

En outre, il faut remarquer que la parole est un code conventionnel. Alors qu'il y a des régularités statistiques au travers des langues humaines, chaque communauté linguistique possède sa propre manière de catégoriser les sons, et son propre répertoire de règles de combinaisons de ces sons. Par exemple, les Japonais n'entendent pas la différence entre le *r* de *read* et le *l* de *lead* en anglais. Comment alors une communauté linguistique en arrive-t-elle à former un code qui est partagé par tous ses membres, sans qu'il y ait de contrôle supervisé global ?

Depuis les travaux de De Boer ou de Kaplan[16], on sait comment un nouveau son ou un nouveau mot peut se propager et être accepté dans une population donnée. Mais ces mécanismes de négociation, encore appelés « dynamiques du consensus », font appel à la préexistence de conventions et d'interactions linguistiques. Les modèles associés concernent donc plutôt la formation et l'évolution des langues, mais ne proposent pas de solutions relatives à l'origine du langage. En effet, quand il n'y avait pas déjà de systèmes de communication conventionnels, comment sont apparues les premières conventions de la parole ?

C'est à cette dernière question en particulier que le modèle que je vais présenter s'intéresse. Elle est évidemment liée à celle de la formation des langues, car il s'agit de comprendre comment un code de la parole a pu être formé pour constituer la base des toutes premières langues. La différence principale entre les deux questions réside dans les propriétés qui doivent caractériser le mécanisme que l'on cherche. Pour la question de l'origine de la parole, on doit en particulier chercher un mécanisme explicatif qui ne présuppose ni l'existence de conventions linguistiques, ni l'existence de structures cognitives spécifiques au langage. Cela impliquerait en effet qu'on a affaire à des individus qui parlent déjà, et donc pour lesquels le langage est déjà apparu.

16. B. de Boer, *op. cit.* ; F. Kaplan, *op. cit.*

Auto-organisation et évolution de la parole

Il est donc naturel de se demander d'où vient cette organisation de la parole et comment un tel code conventionnel et partagé a pu se former dans une société d'agents qui ne disposaient pas déjà de conventions linguistiques. Comme je l'ai argumenté précédemment, deux types de réponses doivent être apportés. Le premier type est une réponse fonctionnelle : il établit la fonction des systèmes sonores, et montre que les systèmes qui ont l'organisation que nous avons décrite sont efficaces pour remplir cette fonction. Cela a par exemple été proposé par Liljencrantz et Lindblöm[17] qui ont montré que les régularités statistiques des répertoires de phonèmes peuvent être prédites en recherchant les systèmes de vocalisations les plus efficaces. Ce type de réponse est nécessaire, mais non suffisant : il ne permet pas d'expliquer *comment* l'évolution (génétique ou culturelle) pourrait avoir trouvé cette structure quasi optimale, ni *comment* une communauté linguistique fait le « choix » d'une solution particulière parmi les nombreuses solutions quasi optimales. En particulier, il se peut que la recherche darwinienne « naïve » avec des mutations aléatoires ne soit pas suffisamment efficace pour trouver des structures complexes comme celles de la parole : l'espace de recherche est trop grand.

C'est pourquoi un second type de réponse est nécessaire : il faut aussi trouver le moyen d'établir comment la sélection naturelle a pu « trouver » ces structures. On peut pour cela montrer que l'auto-organisation est susceptible, dans ce cas précis, de contraindre l'espace de recherche et d'aider la sélection naturelle. Il suffit de montrer qu'un système beaucoup plus simple que la structure que l'on cherche à expliquer s'auto-organise spontanément *en générant cette structure.*

Nous allons donc présenter maintenant un tel système et montrer comment des prémisses relativement simples d'un point de vue évolutionnaire peuvent conduire à la formation auto-organisée de codes de la parole.

---

17. J. Liljencrantz et B. Lindblom (1972), « Numerical simulation of vowel quality systems : the role of perceptual contrast », *Language*, 48, p. 839-862.

UNE EXPÉRIMENTATION INFORMATIQUE DE LA FORMATION DES STRUCTURES FONDAMENTALES DE LA PAROLE

Ce modèle informatique est individuo-centré : il consiste en la mise au point de robots virtuels, caractérisés par un modèle des appareils auditifs et moteurs, et par des réseaux de neurones artificiels qui connectent la modalité perceptuelle et la modalité motrice. Ces réseaux de neurones artificiels déterminent leur comportement, qui consiste principalement à effectuer du babillage vocal. Ce babillage, couplé avec les propriétés de plasticité des réseaux neuronaux, permet à ces robots d'apprendre les correspondances entre l'espace des perceptions auditives et l'espace des mouvements du conduit vocal. Enfin, ces robots sont placés ensemble dans un même environnement dans lequel ils peuvent non seulement entendre leur propre babillage, mais également celui de leurs voisins, qui les influence. Nous allons ainsi montrer que des propriétés émergentes caractérisant les vocalisations produites par les robots d'une même population se forment spontanément.

Plus techniquement[18], les agents disposent d'une oreille artificielle (dont les propriétés peuvent être modifiées pour étudier leur rôle spécifique, voir ci-dessous), capable de transformer un signal acoustique en impulsions nerveuses qui stimulent les neurones d'une carte de neurones artificiels perceptuels. Ils disposent aussi d'une carte de neurones moteurs dont l'activation produit des mouvements d'un modèle du conduit vocal, qui lui-même produit une onde acoustique (et dont le degré de réalisme peut également être modifié). Les cartes nerveuses (perceptuelle et motrice) sont totalement connectées entre elles. Au départ, les paramètres internes de tous les neurones, ainsi que les connexions entre les deux cartes, sont aléatoires. Pour produire une vocalisation, un robot active aléatoirement plusieurs neurones de sa carte motrice, dont les paramètres internes codent pour des configurations articulatoires à atteindre en séquence, ce qui produit une trajectoire articulatoire et, par le biais du modèle du conduit vocal, une trajectoire acoustique qui est perçue grâce au modèle de l'appareil auditif. Ceci définit leur comportement de babillage. C'est pourquoi, au départ de l'expérience, les robots produisent des vocalisations qui sont aléatoires dans l'espace vocal. Cependant, ces réseaux de neurones sont caractérisés par deux formes de plasticité : 1) les connections intermodales évoluent de telle manière que le robot apprend les correspondances entre les trajectoires auditives perçues et les comman-

18. Nous ne donnons ici qu'une description générale du système ; pour une description mathématique précise, voir P.-Y. Oudeyer, *Self-Organization…*, *op. cit.*

des motrices qui les génèrent quand il babille[19]; 2) Les neurones de chacune des cartes évoluent de telle manière qu'ils tendent à modéliser la distribution des sons que le robot entend[20] : les connections entre les deux cartes perceptuelles sont aussi telles que la distribution des sons codés par la carte perceptuelle (celle des sons perçus) reste à peu près la même que la distribution des sons codés par la carte motrice (celle des sons produits). Autrement dit, l'architecture nerveuse de l'agent est telle qu'il a tendance à produire la même distribution de sons que celle qu'il entend. Ainsi, si l'on fait écouter à un robot un flux de parole d'une langue donnée, son babillage va se régler/s'aligner sur la distribution des sons présents dans cette langue. Par exemple, si cette langue contient les voyelles [a, e, i] mais pas [o], le robot va se mettre à babiller en prononçant beaucoup plus souvent des [a, e, i] que des [o]. Ce comportement correspond à ce qui est observé chez les jeunes enfants, et parfois référé sous le terme de *phonological attunement*[21].

UN MÉCANISME UNIFIÉ POUR L'AUTO-ORGANISATION DE LA COMBINATORIALITÉ, DE LA DUALITÉ UNIVERSAUX/DIVERSITÉ ET DU PARTAGE CULTUREL

Ce type d'architecture a été utilisé fréquemment dans la littérature pour modéliser les phénomènes d'acquisition de la parole chez les enfants[22], dans des expériences dans lesquelles le système apprenait à prononcer les sons/syllabes d'une langue qu'on lui faisait entendre. Cependant, l'expérience qui est ici présentée est différente : on ne suppose pas qu'il existe au départ un système de parole déjà constitué. Au contraire, on va placer une population de robots babilleurs dans un environnement commun, de telle manière qu'ils vont percevoir à la fois leurs propres babillages et ceux de leurs voisins. Étant donné que les propriétés de plasticité de leurs cerveaux les font babiller en s'alignant sur les vocalisations

19. Les connections entre les deux cartes de neurones évoluent selon la loi de Hebb : celles qui relient des neurones qui sont souvent activés en même temps deviennent plus fortes, et celles qui relient des neurones dont l'activité n'est pas corrélée deviennent plus faibles. Ces connexions sont aléatoires au début, et, grâce au babillage des agents, elles s'organisent de telle manière que l'agent devient capable de trouver les commandes motrices correspondant à un son qu'il « entend ».

20. Les neurones s'adaptent aux stimuli par sensibilisation : leur dynamique est telle que, si un stimulus S est perçu, alors ils sont modifiés de telle manière que, si l'on présente le même stimulus S juste après, ils répondront encore plus.

21. Marilyn Vihman, *Phonological Development : the Origins of Language in the Child*, Cambridge (Mass), Blackwell, 1996.

22. T. Kohonen (1988), « The neural phonetic typewriter », *Computer*, 21 (3), p. 11-22 ; V. Sanguineti, R. Laboissière et D. J. Ostry (1998), « A dynamic biomechanical model for neural control of speech production », *Journal of the Acoustical Society of America*, 103 (3), p. 1615-1627.

qu'ils entendent autour d'eux, et comme statistiquement ils produisent tous des vocalisations aléatoires dans l'espace vocal, l'état initial est un état d'équilibre.

Seulement, si l'on fait fonctionner la simulation, on s'aperçoit que cet équilibre n'est pas stable. En effet, il y a du bruit – de la « stochasticité » – qui fait que, par hasard et de temps en temps, certains types de vocalisations seront prononcés plus souvent que d'autres. Or, le mécanisme de couplage décrit plus haut introduit une boucle de rétroaction positive : ces déviations de la moyenne sont amplifiées lorsqu'elles sont assez grandes, et la symétrie du système se brise. Les cartes de neurones s'auto-organisent alors en groupes concentrés de neurones, codant pour des configurations acoustiques et articulatoires très précises dans l'espace des vocalisations (voir figure 4). En bref, l'espace continu des vocalisations a été discrétisé. Les vocalisations que les agents produisent ne sont plus holistiques, mais digitales : elles sont systématiquement construites par la mise en séquence de quelques configurations clés, que l'on peut alors appeler phonèmes. On voit apparaître le codage phonémique/digital et combinatorial qu'on a décrit plus haut. En outre, le « code phonémique » qui apparaît est le même chez tous les agents d'une même simulation, alors qu'il est différent d'une simulation à l'autre. On observe donc la formation d'une « convention culturelle », qui peut être diverse d'un groupe à l'autre.

Plusieurs variantes de cette expérience peuvent être réalisées et permettent d'affiner ces conclusions. Tout d'abord, il est possible de réaliser cette expérience avec un seul agent qui s'écoute babiller. Dans ce cas, on observe également un phénomène de cristallisation de ses vocalisations : très vite il ne produit plus que des trajectoires vocales qui transitent toutes par quelques configurations articulatoires clés qui sont systématiquement réutilisées. On peut donc en déduire que l'apparition du codage phonémique, c'est-à-dire de la digitalité/combinatorialité, n'est pas le résultat des interactions sociales, mais des propriétés du couplage interne à chaque robot entre les modalités perceptuelles et motrices de la parole. Cependant, alors que les vocalisations de robots babillants isolés vont se cristalliser sur des systèmes de vocalisations différents, ces systèmes de vocalisations se synchronisent spontanément si les robots partagent le même environnement et sont capables de s'entendre les uns les autres : dans ce cas, ces systèmes sont tous quasiment les mêmes dans une même population.

Une seconde variante importante de cette expérience consiste à faire varier les propriétés morphophysiologiques des modèles des appareils auditifs et phonatoires afin de déterminer quel est l'impact de ces propriétés sur les systèmes qui se forment (ou qui ne se forment pas). En particulier, une propriété importante de ces appareils est la non-linéarité de la

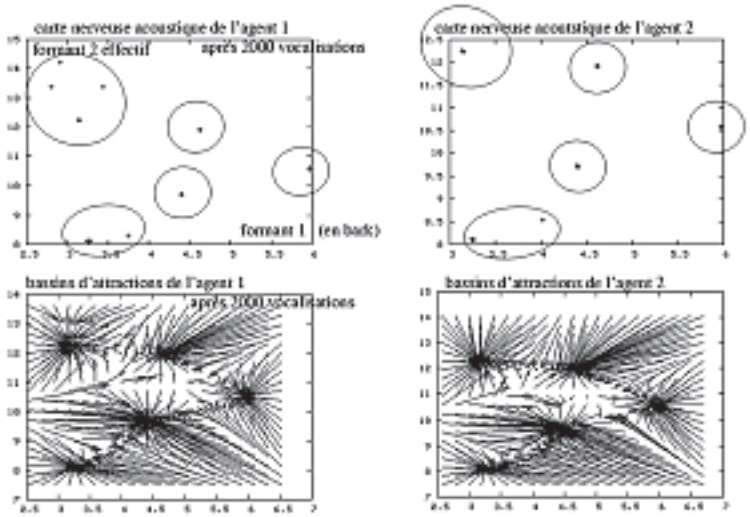


Figure 4 : *Très rapidement, la symétrie initiale des cartes neurales des robots se brise, et les neurones qui encodaient initialement des configurations vocales aléatoirement réparties dans l'espace vocal encodent maintenant un petit nombre de configurations qui sont systématiquement réutilisées par les agents quand ils babillent : l'espace vocal a été discrétisé. En outre, ces configurations élémentaires qui émergent sont les mêmes chez tous les agents d'une même population, mais différentes entre populations. On peut le voir sur cette figure qui représente la carte des neurones perceptuels de deux agents après 2 000 vocalisations (*en haut*), ainsi qu'une représentation de leur distribution (*en bas*). L'espace auditif est ici projeté sur le premier et le second formant effectif, exprimés en barks, ce qui permet de représenter les voyelles du système sonore auto-organisé dans cette simulation.*

fonction qui fait correspondre une onde acoustique et une perception auditive à des commandes motrices et à des configurations du conduit vocal. Le conduit vocal humain est en effet tel que, pour certaines configurations articulatoires, de petites variations produisent de petites variations du son produit et perçu. Pour d'autres configurations, de petites variations produisent de grandes variations du son produit. Cette propriété est centrale dans plusieurs théories qui proposent d'expliquer le codage phonémique de la parole, par exemple dans la théorie quantale de Stevens[23], ou dans le modèle DRM[24]. Or il est possible d'utiliser un

23. K. N. Stevens (1989), « On the quantal nature of speech », *Journal of Phonetics*, 17, p. 3-45.
24. M. Mrayati, R. Carre et B. Guerin (1988), « Distinctive regions and modes : a new theory of speech production », *Speech Communication*, 7, p. 257-286.

modèle des appareils acoustiques et phonatoires qui soit réaliste et qui inclue ces types de non-linéarités, mais il est aussi possible de construire un modèle non-réaliste volontairement linéaire afin de déterminer l'impact des non-linéarités. Ces expériences ont donc été menées. Tout d'abord, avec un modèle linéaire, on observe dans une population de robots babillants la cristallisation décrite précédemment : leurs vocalisations s'auto-organisent en un système combinatorial dans lequel quelques configurations articulatoires sont systématiquement réutilisées comme points clés des trajectoires vocales. Nous pouvons donc faire une première conclusion : ces simulations montrent que le codage phonémique peut apparaître spontanément sans linéarité dans les appareils phonatoires et auditifs. Cela n'implique pas que ces non-linéarités n'accélèrent pas la formation du codage phonémique, mais cela montre qu'elles ne sont pas nécessaires, contrairement à ce qui est proposé par la théorie quantale ou le modèle DRM.

Si, quand on utilise le modèle audiophonatoire linéaire, on fait de nombreuses simulations et que l'on s'intéresse à la distribution de ces configurations articulatoires clés (qu'on peut voir comme des sortes de phonèmes), on s'aperçoit qu'elles sont globalement positionnées à des endroits aléatoires et uniformément dans l'espace des configurations possibles. Cependant, quand on utilise un modèle réaliste, reproduisant en particulier les propriétés de production et de perception des voyelles des humains[25], on observe un phénomène supplémentaire. Outre le phénomène de cristallisation qui est le même qu'avec le modèle linéaire, les systèmes de vocalisations qui se forment sont caractérisés par des régularités statistiques qui ressemblent beaucoup aux systèmes humains. On peut par exemple faire des statistiques sur les voyelles qui apparaissent en tant que configurations articulatoires clés dans les systèmes émergents de cette expérience en faisant de nombreuses simulations. Les résultats, illustrés sur la figure 5, montrent que d'une part une certaine diversité de systèmes de voyelle apparaît, mais en même temps certains systèmes de voyelles apparaissent beaucoup plus fréquemment que d'autres. Il y a donc la même dualité universaux/diversité que celle que l'on observe dans les langues humaines, et cette simulation en propose une explication unifiée :

1) le système dynamique constitué par l'ensemble des agents babillants et par les couplages sensori-moteurs internes qui les caractérisent comporte un certain nombre d'attracteurs que sont les systèmes de

25. Pour une description précise du modèle basé sur les travaux de B. de Boer (*op. cit.*), voir P.-Y. Oudeyer, *Self-Organization…*, *op. cit.*

vocalisation combinatoriaux/avec un codage phonémique partagés par tous les membres d'une population ;

2) sous l'influence du bruit et des petites variations des conditions initiales, le système dynamique tombe dans un attracteur particulier, ce qui permet d'expliquer le « choix collectif décentralisé » de la population pour adopter un système plutôt qu'un autre ;

3) les non-linéarités des appareils auditifs et phonatoires introduisent des asymétries entre les attracteurs : certains d'entre eux ont un bassin d'attraction plus grand, en particulier ceux dans lesquels les phonèmes sont dans des zones ou de petites variations articulatoires provoquent de petites variations perceptuelles, ce qui a pour conséquence d'augmenter la probabilité que le système y « tombe ».

En outre, non seulement il y a une correspondance structurelle entre les simulations et la réalité, mais les systèmes de voyelles apparaissant le plus fréquemment dans les populations de robot sont à peu près les mêmes, et dans les mêmes proportions, que les systèmes les plus fréquents dans les langues du monde. Il y a donc aussi une certaine correspondance quantitative. On peut donc en conclure que les non-linéarités des appareils auditifs et phonatoires sont ici déterminantes pour expliquer pourquoi certains systèmes de phonèmes sont statistiquement plus fréquents que d'autres. Par contre, encore une fois, l'existence même des phonèmes, c'est-à-dire l'existence d'un système de vocalisations dans lequel des configurations articulatoires ou auditives invariantes sont systématiquement réutilisées, ne nécessite pas ces non-linéarités.

Il est cependant important de noter que, dans ces simulations, les architectures neurales sont caractérisées par plusieurs paramètres, et que toutes les valeurs de ces paramètres ne permettent pas d'obtenir ces résultats. Néanmoins, il se trouve qu'un seul de ces paramètres a une influence importante sur les résultats[26] : les neurones sont caractérisés par une sélectivité σ aux stimuli plus ou moins grande. Si cette sélectivité est trop petite, aucune cristallisation ne se passe, bien que les robots restent capables d'apprendre les relations entre l'espace auditif et l'espace moteur. Si elle est trop grande, alors le système se cristallise en un système dégénéré, dans lequel toutes les vocalisations sont exactement les mêmes et sont stationnaires : il n'y a qu'un phonème. Cependant, il y a une très grande plage de valeurs entre ces deux extrêmes, qui permet de faire apparaître une cristallisation dans laquelle un système combinatoire avec de multiples phonèmes se forme dans la population de robot babillants.

---

26. Voir P.-Y. Oudeyer, *Self-Organization…*, *op. cit.*

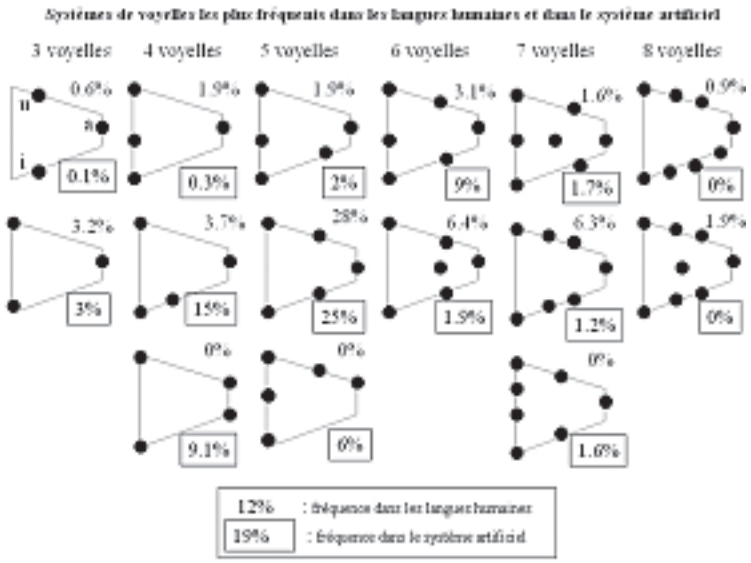


Figure 5 : *Comparaison entre la distribution des systèmes de voyelles auto-organisés dans le système artificiel et celle des systèmes de voyelles dans les langues humaines (d'après la base de données UPSID[27]). Les systèmes de voyelles sont représentés sur le triangle vocalique, dont la dimension horizontale correspond au premier formant, et la dimension verticale au second formant effectif. On observe que les systèmes les plus fréquents engendrés par les agents artificiels sont aussi les plus fréquents chez les humains, en particulier le système à 5 voyelles symétriques /a, e, i, o, u/ avec 25 % dans les systèmes artificiels et 28 % dans les langues humaines.*

VERS UNE VISION NOUVELLE DES SCÉNARIOS ÉVOLUTIONNAIRES DES ORIGINES DE LA PAROLE

Ces remarques ont une importante conséquence si l'on utilise ce modèle pour imaginer des scénarios évolutionnaires ayant pu mener à la formation des premiers systèmes de vocalisations partageant des propriétés fondamentales des systèmes de parole de l'humain contemporain. En effet, elles impliquent d'abord que beaucoup de variations des paramètres de cette architecture neurale permettent de faire apparaître des systèmes de paroles combinatoriaux partagés par toute une population. Ensuite, elles impliquent qu'avec une telle architecture neurale, un tel système de parole peut apparaître sans supposer d'autres propriétés particulières des appareils auditifs et phonatoires que celle de pouvoir babiller initialement avec une certaine variété de sons. En particulier, aucune non-linéarité n'est néces-

27. Ian Maddieson, *Patterns of Sounds*, Cambridge, Cambridge University Press, 1984.

saire. Enfin, l'architecture neurale elle-même est relativement simple : elle met en œuvre des unités neurales dont les propriétés intrinsèques et les propriétés de plasticité sont très classiques et correspondent, au niveau fonctionnel, à la manière dont fonctionnent la plupart des unités neurales des cerveaux des mammifères[28]. La spécificité de cette architecture est le fait que les cartes neurales auditives sont fortement et directement connectées aux cartes motrices, et que ces connexions sont plastiques. Or cette architecture et ces connexions apparaissent comme étant les éléments essentiels, au départ, non pas de la parole mais tout simplement de la capacité à apprendre à imiter des sons, ce que nous appelons ici l'imitation vocale adaptative[29]. Ce qui nous amène au scénario évolutionnaire suivant pour conceptualiser l'origine des systèmes de vocalisations phonémiques et partagés par les membres d'une communauté.

1) L'imitation vocale adaptative est présente chez de nombreux animaux[30] qui disposent de systèmes de vocalisations partagés et appris, mais chez qui le langage n'existe pas. Les éthologues ont d'ailleurs établi de nombreux avantages reproductifs potentiels caractérisant la capacité d'imitation vocale adaptative dans une communauté d'individus (comme par exemple le fait que cela permet de marquer l'appartenance au groupe). Il est donc raisonnable de penser qu'avant de savoir parler, les humains ont pu disposer de la capacité à s'imiter vocalement et que cette capacité d'imitation est apparue avant le langage.

2) Être capable d'imitation vocale adaptative, ainsi que la plupart des avantages reproductifs identifiés par les éthologues pour expliquer la présence de cette capacité chez certains animaux, n'implique pas et ne nécessite pas un système de vocalisation combinatoire et codé phonémiquement. La zone de paramètres pour lesquels le paramètre de sélectivité σ est petit permet d'ailleurs aux robots d'apprendre très bien les correspondances perceptuo-motrices vocales sans pour autant générer de système phonémique.

3) Si maintenant l'on se place dans un contexte écologique dans lequel la présence d'un système de parole combinatorial apporterait un avantage reproductif à ses possesseurs, alors les expériences que nous avons décrites permettent de dire qu'un simple changement de la valeur du paramètre de sélectivité σ des cartes de neurones auditives et motrices permettrait de faire apparaître spontanément des systèmes de vocalisa-

28. *Ibid.*
29. Et qui correspond à la terminologie anglaise : *adaptive vocal mimicry*.
30. Charles T. Snowdown et Martine Hausberger, *Social Influences on Vocal Development*, Cambridge, Cambridge University Press, 1997 ; Marc D. Hauser, *The Evolution of Communication*, MIT Press (Bradford Books), 1997.

tions qui partagent plusieurs des propriétés fondamentales des systèmes de parole humains contemporains, au point même de pouvoir approximativement en prédire les systèmes de voyelles quand on utilise un modèle de la production et de la perception des voyelles chez l'humain. Cela permet alors de comprendre que ce qui fut un grand pas dans l'origine du langage, c'est-à-dire la formation de systèmes de vocalisations combinatoriales, a pu être la conséquence d'un petit changement biologique grâce aux propriétés auto-organisatrices de la matière neurale et de ses couplages multimodaux.

Ce scénario dans lequel les systèmes de parole phonémiques auraient été sélectionnés grâce aux avantages reproductifs qu'ils auraient pu procurer – sélection rendue possible par la relative facilité de générer ces systèmes à partir des bases biologiques de l'imitation vocale adaptative – n'est cependant pas le seul que ce modèle peut appuyer. En effet, j'ai expliqué plus haut que, dans la plage de valeurs du paramètre σ qui permet à des systèmes phonémiques de se former, la capacité d'imitation vocale est intacte et aussi performante. Or, à performances d'imitation égales, le passage du paramètre σ entre cette zone de paramètre et la zone de petite sélectivité ne présente pas de coût métabolique *a priori*. Cela implique que, dans un contexte écologique dans lequel ces structures neurales sont apparues sous l'effet d'une pression sélective pour l'imitation vocale adaptative, des mutations/variations neutres ont pu se produire et générer spontanément des systèmes de parole phonémique sans pression de sélection linguistique. Une observation rend ce scénario particulièrement stimulant : parmi les espèces animales capables d'apprendre à faire des imitations vocales, et pour lesquels il existe des systèmes de vocalisations partagés culturellement mais qui n'ont pas de langage, un certain nombre produisent des vocalisations ou des chants structurés autour d'unités de base systématiquement réutilisées. C'est le cas par exemple des canaris ou des diamants mandarins chez les oiseaux[31], ou des baleines à bosses[32]. La fonction de cette structuration quasi phonémique est encore très mal conceptualisée en éthologie. Or, le modèle que j'ai présenté, parce qu'il est neutre vis-à-vis de beaucoup de propriétés des appareils auditifs et vocaux, et parce que l'architecture neurale qu'il suppose correspond à l'équipement minimal pour l'imitation vocale adaptative, peut s'appliquer également à la formation des chants chez ces animaux. Dans ce cas, il fournit une hypothèse que l'incertitude sur la fonction du codage combi-

31. E. A. Brenowitz et M. D. Beecher (2005), « Song learning in birds : diversity and plasticity, opportunities and challenges », *Trends in Neuroscience*, 28 (3), p. 127-132.
32. P. Tyack (1981), « Interactions between singing hawaian humpback whales and conspecifics nearby », *Behavioral Ecology and Sociobiology*, 8 (2), p. 105-116.

natoire de ces chants vient renforcer : une combinatoire d'unités vocales pourrait s'être formée spontanément comme un effet collatéral de l'équipement biologique pour l'imitation vocale adaptative. Il est donc également raisonnable d'imaginer que cela aurait pu se produire chez l'être humain : les systèmes combinatoires de parole auraient été recrutés seulement par la suite pour réaliser leur fonction linguistique. Dans ce cas, plusieurs des propriétés fondamentales des systèmes de parole de l'humain contemporain seraient des exaptations[33].

Conclusion

Grâce à la construction et à l'utilisation d'un modèle informatique, j'ai montré comment une architecture de couplage sensori-moteur audio-phonatoire relativement simple permettait, dans une dynamique auto-organisatrice, la formation spontanée de systèmes de vocalisations combinatoires et codés phonémiquement, partagés par tous les individus d'une même communauté, et caractérisés par la dualité universaux/diversité. La première contribution de ce travail est qu'il permet pour la première fois de proposer une explication unifiée de ces trois phénomènes.

En outre, cette architecture de couplage multimodal correspond au matériel neuronal minimal nécessaire à l'imitation vocale adaptative, et ne comporte aucun élément biologique spécifique de la parole humaine. Étant donné que le phénomène de cristallisation du système se réalise pour une large partie de l'espace des paramètres de ce modèle, cela montre que l'innovation biologique permettant de passer de systèmes de vocalisations inarticulés à des systèmes caractérisés par plusieurs des propriétés fondamentales de la parole de l'humain contemporain a pu être modeste. Il ne semble en effet pas nécessaire que des structures neuronales encodant *a priori* et spécifiquement l'organisation phonémique ainsi que les régularités typiques de la parole soient générées de manière innée pour permettre la formation et l'apprentissage de tels codes de la parole. C'est la seconde contribution de ce travail : il permet de comprendre comment les propriétés d'auto-organisation de structures neuronales simples ont pu contraindre l'espace des formes biologiques de la parole et permettre qu'elles soient générées puis sélectionnées lors de la phylogenèse.

---

33. Ce terme a été introduit par S. J. Gould et E. S. Vrba (1982), « Exaptation. A missing term in the science of form », *Paleobiology*, 8 (1), p. 4-15.

Ces nouvelles hypothèses n'auraient probablement pas pu être identifiées sans l'utilisation de simulations informatiques, parce que les dynamiques qu'elles impliquent sont complexes et difficiles à prévoir par un travail uniquement verbal. Cela illustre l'importance potentielle que peuvent avoir ces nouveaux outils méthodologiques pour les sciences humaines et pour les sciences de la nature. Cependant, ces modèles informatiques font abstraction de nombreux mécanismes biologiques et comportementaux, et donc consistent avant tout en un travail théorique de réflexion sur l'espace des hypothèses : une fois cet espace reconceptualisé et la cohérence interne des hypothèses évaluée par des simulations, tout le travail de validation et d'ancrage de ces hypothèses dans la réalité des observations biologiques reste à faire. Ainsi, la troisième contribution de ce travail est, plus que l'élaboration de nouvelles hypothèses spécifiques, la mise en place d'un cadre de travail et d'outils qui permettent de développer de nouvelles intuitions et de nouveaux concepts pour penser les origines et l'évolution de la parole[34].

34. REMERCIEMENTS. Ce travail a été réalisé en grande partie au Sony Computer Science Laboratory, à Paris, et il a bénéficié du soutien de Luc Steels.

Comment formaliser la diversité des langues ?

par LUIGI RIZZI

Le thème de la diversité des langues ne peut être abordé que dans le contexte du thème complémentaire de l'uniformité des langues, et de l'unité du langage humain. Ces deux aspects se présupposent mutuellement et ne sont jamais totalement dissociables, quoique le choix d'une perspective particulière doive mettre l'accent sur l'un ou sur l'autre. Du point de vue des circonstances pratiques de la vie quotidienne, l'aspect saillant est sans doute la diversité : c'est la diversité des langues qui peut affecter très directement nos vies dans des situations concrètes d'interaction humaine, en permettant ou en empêchant la communication et le dialogue, et en créant l'appartenance ou l'exclusion du groupe. Par contre, la recherche formelle sur le langage, qui réfléchit sur la structure des langues dans le vaste ensemble des systèmes symboliques et communicatifs possibles, souligne l'uniformité profonde des langues : elle met en évidence l'existence d'un système très structuré d'universaux linguistiques, et essaye d'identifier les limites de la variation linguistique.

En effet, les deux aspects de l'invariance et de la variation des langues sont toujours coprésents, et il faut les aborder ensemble. Dans ce travail je voudrais d'abord mentionner quelques éléments invariants du langage, en particulier en relation à ses aspects combinatoires, et illustrer certaines procédures formelles qui sont à la base du caractère illimité et « créatif » du langage. Ceci nous permettra d'aborder le thème de la variation, et d'illustrer les modèles « paramétriques » qui ont eu une diffusion considérable dans le dernier quart de siècle. On évoquera plusieurs types de preuves empiriques qui peuvent être utilisés pour tester ces modèles, et on regardera de près quelques prédictions générées par des hypothèses analytiques précises : la microcomparaison et la diachronie, la macrocomparaison et la typologie, l'acquisition du langage. Je voudrais terminer en

évoquant la pertinence potentielle des études d'imagerie cérébrale pour la question des règles linguistiques possibles et impossibles.

La « créativité » et la combinatoire illimitée des langues humaines

Par « créativité » dans l'usage normal de la langue, on entend le fait que nous sommes constamment confrontés à des phrases nouvelles : des séquences de mots que nous n'avons jamais rencontrées dans notre expérience linguistique précédente et que nous n'avons néanmoins aucune difficulté à les comprendre ; et nous produisons constamment des phrases que nous n'avons jamais entendues. En effet, nos capacités linguistiques nous donnent la maîtrise d'un ensemble potentiellement infini de phrases : on ne peut pas identifier la phrase française la plus longue ! Comment caractériser une telle familiarité avec ce qui est constamment nouveau et l'infinité potentielle des objets linguistiques que nous pouvons construire ?

Ces questions, centrales pour la linguistique moderne, ont des racines anciennes. Dans un passage très célèbre du *Discours de la méthode*, Descartes a identifié dans cette capacité de produire des arrangements de mots nouveaux et appropriés au contexte la propriété critique permettant de distinguer l'homme de la machine : les hommes « les plus hébétés » en sont capables, mais cette capacité reste inatteignable pour les automates les plus perfectionnés. Une telle capacité dépend de plusieurs facteurs, mais elle est certainement liée aux propriétés combinatoires du langage humain. Plus ou moins en même temps, Galilée avait mis en relief, de manière indirecte, ces propriétés combinatoires dans son *Dialogo sopra i due massimi sistemi del mondo*. Il est question, dans le passage pertinent du *Dialogo*, d'identifier la plus grande invention de l'humanité. Selon l'un des participants au dialogue, c'est l'invention de l'écriture alphabétique qui crée la possibilité de communiquer « *i… più reconditi pensieri* » (les pensées les plus cachées) à travers les distances temporelle et spatiale grâce aux arrangements d'une vingtaine de petits caractères sur une feuille. Galilée est frappé par l'extrême simplicité d'une invention si puissante. Il s'agit bien d'une invention, mais une invention fondée sur une découverte aussi extraordinaire : la découverte de la combinatoire sur plusieurs niveaux hiérarchisés qui régit le fonctionnement des langues naturelles. Les phonèmes, repris de manière plus ou moins fiable par les graphèmes de l'écriture alphabétique, se combinent en entités plus grandes,

les morphèmes, puis en mots ; et les mots se combinent dans la syntaxe pour donner lieu à l'infinité potentielle des phrases de la langue, capables d'exprimer « *i… più reconditi pensieri* ».

Le caractère central de la syntaxe dans l'organisation des langues naturelles est reconnu par Ferdinand de Saussure : nous lisons dans le *Cours de linguistique générale* que la syntaxe fait bien partie du système de la langue, non seulement dans les expressions figées, mais dans toutes les constructions qui peuvent être ramenées à des « patrons réguliers ». Néanmoins, il faudra attendre encore un demi-siècle, jusqu'au milieu des années 1950, avant de pouvoir disposer d'un mécanisme précis capable d'exprimer le caractère régulier et infini de la syntaxe.

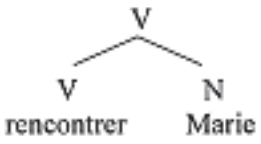
L'hypothèse d'un tel mécanisme a été la première contribution formelle importante de Chomsky : les règles syntaxiques sont récursives, en ce qu'elles peuvent se réappliquer indéfiniment à leur propre résultat. Par exemple, une phrase peut contenir une expression nominale, qui peut contenir une phrase, qui peut contenir une expression nominale, et ainsi de suite. La récursivité est le secret de l'infinitude de la syntaxe.

Les linguistes ont beaucoup discuté, dans le dernier demi-siècle, sur la nature des mécanismes syntaxiques récursifs. Dans les versions récentes du programme minimaliste, le mécanisme récursif optimal est aussi le plus simple et le plus général qu'on puisse concevoir. C'est l'opération appelée « fusion » (*merge* en anglais) :

(1)

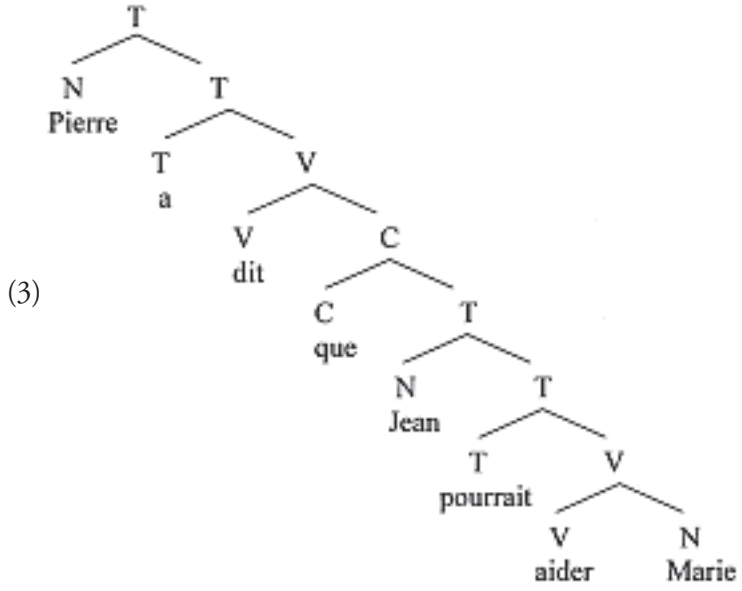
La fusion dit simplement : prends deux éléments A et B (par exemple deux mots du lexique), mets-les ensemble pour former un élément complexe [A B], qui aura à son tour une étiquette, héritée de A ou bien de B, et qui pourra à son tour être soumis à la fusion avec un autre élément, et ainsi de suite, indéfiniment. L'assignation de l'étiquette à la structure ainsi créée est déterminée algorithmiquement. Quand deux éléments sont soumis à la fusion, l'un des deux sélectionne l'autre : le sélectionneur est la « tête » de la construction, l'élément dont l'étiquette est assignée à la structure. Par exemple, un verbe transitif sélectionne un nom, son objet direct, et la structure créée est un syntagme de nature verbale, un syntagme verbal, dont la tête est le verbe :

(2)



De cette manière on peut construire, pas à pas, des structures indéfiniment complexes, dont on peut exprimer la structure par des représentations arborescentes comme la suivante :

(3)



Invariance et variation, principes et paramètres

On arrive ainsi à la question de la variation. Toutes les langues sont fondées sur le mécanisme de la fusion : on ne connaît aucune langue qui ne l'utiliserait pas et se limiterait à des expressions atomiques, à des mots isolés. Toutes les langues comprennent des phrases à structure hiérarchisée, où l'organisation hiérarchique des mots est déterminée par la séquence des applications successives de la fusion. Mais ce qui varie est l'ordre des éléments : dans certaines langues, comme le français, le verbe précède son complément dans l'ordre fondamental des mots ; dans d'autres langues, comme le japonais, c'est le contraire. Les langues du monde se laissent ainsi classifier en langues VO et langues OV. Voici quelques exemples[1] :

1. Les langues varient aussi considérablement par rapport à la rigidité de l'ordre des mots. Il ne s'agit pas d'un simple paramètre binaire (« ordre libre ou pas »), mais plutôt du jeu de plusieurs paramètres plus fins, permettant plusieurs degrés de liberté. Même dans le cas des langues qui manifestent une très grande liberté dans les ordres possibles, il paraît toujours justifié d'identifier un ordre fondamental, qui peut être bouleversé par des processus de mouvement, un type d'opération qui sera discuté par la suite.

(4)

	[V O]		[O V]
Francais:	Jean [aime Marie] 'Jean loves Marie'	Latin :	Tullius [Flaviam amat] 'Tullius Flavia loves'
Chicewa:	Njuchi [zi-na-wa-lum-a nlenje] 'bees bit hunters'	Japonais:	John-ga [Mary-o butta] 'John Mary hit'
Thaï:	khăw [sīi aahăan] 'he buy food'	Navajo:	Ashkii [at'ééd yiyiiltsa] 'Boy girl saw'

Il faut donc admettre que l'opération de fusion permet un paramètre de variation concernant l'ordre des éléments :

(5)

Ce simple point de choix, en apparence banal, a des conséquences très systématiques sur la structure des phrases. Dans les langues qui choisissent l'option (la plus simple) de fixer le paramètre une seule fois pour tous les types de têtes (la majorité des langues, selon le travail fondamental de Greenberg[2], quoique une fixation « incohérente », c'est-à-dire différente pour différents types de tête, soit possible), le choix de l'une ou de l'autre valeur de (5) détermine des différences majeures dans l'ordre des mots. Considérons deux langues « cohérentes » à tête initiale ou finale, comme l'anglais et le japonais : deux phrases complexes comme les suivantes, (6)a-b, reçoivent les structures arborescentes de (7) :

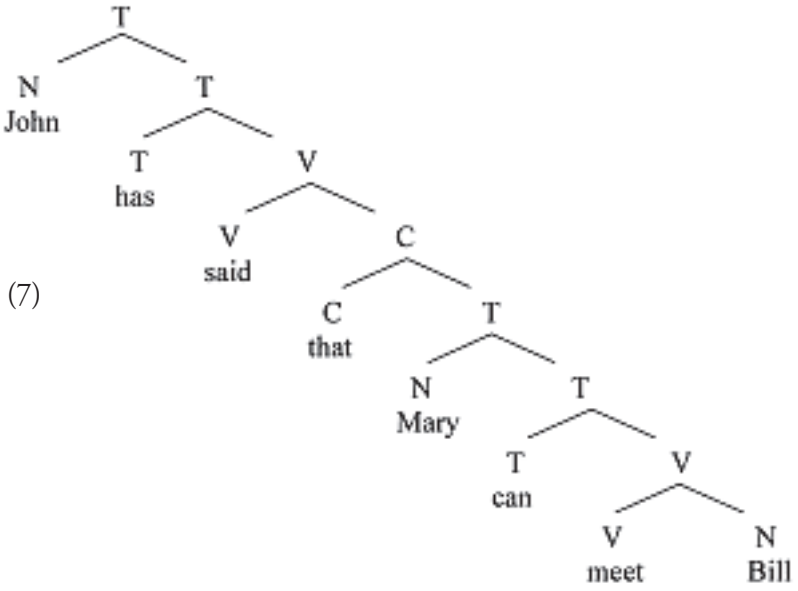
(6)

a John has said [that Mary can meet Bill]

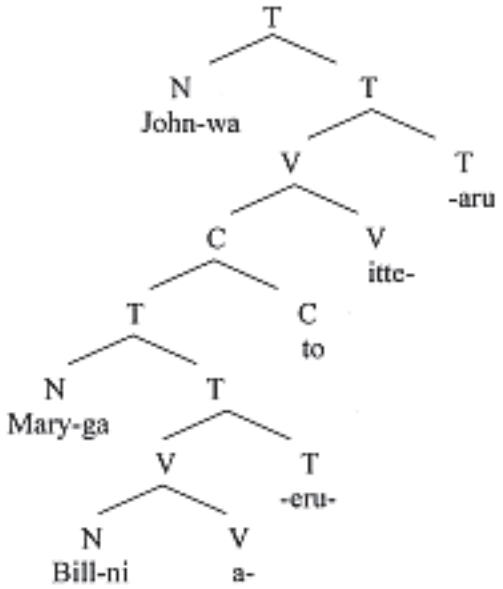
b John-wa [Mary-ga Bill-ni a -eru- to] itte-aru.
'John-Top [Mary-Nom Bill-Dat meet – can - that] said - has'

2. Greenberg (1963).

(7)



(8)



L'arbre japonais est presque entièrement l'image miroir de l'arbre anglais (pas entièrement parce que certaines propriétés d'ordre demeurent constantes entre les deux langues, l'ordre sujet-prédicat par exemple).

Essayons de généraliser la perspective à partir de ces exemples. L'analyse détaillée des langues révèle certaines propriétés invariantes, les universaux linguistiques, et certaines propriétés qui varient d'une langue à

l'autre. Il faut donc qu'une théorie générale des langues et du langage exprime ces deux aspects. Il y a environ un quart de siècle, une approche formelle de l'invariance et de la variation a été proposée, qui devait profondément influencer la linguistique comparative des années à venir. Si on appelle « grammaire universelle » la théorie générale du langage humain, on peut concevoir sa structure interne comme étant constituée de deux sortes d'entités :

– les principes, qui expriment les propriétés universelles ;

– et les paramètres, les « points de choix », qui expriment la variation possible.

Par exemple, nous avons vu que, en ce qui concerne la constitution des structures, il faut postuler un principe général, le mécanisme de fusion, qui crée des structures hiérarchiques, et un paramètre d'ordre. Le défi que ce cadre « des principes et des paramètres » a tenté de relever était de caractériser l'ensemble de la variation syntaxique[3] en termes d'un système de points de choix généralement binaires au sein de la structure invariante des principes.

Cette méthode analytique a constitué un langage formel très bien adapté à la comparaison des langues. En effet, les études de syntaxe comparative ont connu une période d'expansion sans précédent[4]. Dans cette conception, la grammaire d'une langue particulière est constituée par le système général, la grammaire universelle, avec les paramètres fixés d'une certaine manière : les grammaires du français, du chinois, du maori, etc., expriment la grammaire universelle sous certains ensembles distincts de valeurs paramétriques. Le but fondamental du travail comparatif est donc d'identifier les paramètres irréductibles à la base de la variation. Parallèlement, l'acquisition d'une grammaire peut être conçue comme une opération de fixation paramétrique : l'enfant détermine sur la base de son expérience, à chaque bifurcation sous-tendue par un paramètre, quel chemin a pris la langue à laquelle il est exposé. Sur cette base, il a été possible de réorienter l'étude de l'acquisition de la syntaxe, désormais conçue comme l'identification de la séquence temporelle de la fixation des paramètres, dans le développement de l'enfant[5].

---

3. Mais pas seulement : on a proposé des modèles de paramétrisation phonologique, morphologique, sémantique, etc.

4. Chomsky (1981), Rizzi (1982), Kayne (1983) et, pour une synthèse rétrospective, Baker (2001).

5. Hyams (1986) et Rizzi (2000), introduction (pour une synthèse).

Le mouvement

Si la fusion est le mécanisme essentiel pour la constitution des structures, les computations syntaxiques impliquent une autre opération fondamentale : le mouvement. De façon très générale, le mouvement peut être caractérisé comme le fait que les expressions linguistiques sont souvent prononcées dans des positions distinctes des positions où elles sont interprétées. Par exemple, dans une phrase comme la suivante,

(9) C'est ton livre que Jean a dit … que je devrais lire ___

l'expression nominale *ton livre* doit être interprétée comme l'objet direct du verbe *lire*, mais elle est prononcée dans une position qui peut être indéfiniment éloignée du verbe. Une manière classique de comprendre cet état de choses consiste à admettre que, dans une représentation abstraite, chaque expression occupe la position où elle est interprétée (ici la position d'objet de *lire*, indiquée par l'espace ___), et est déplacée à une autre position afin de satisfaire d'autres propriétés formelles ou interprétatives (en [9], le déplacement de *ton livre* est un moyen pour focaliser l'élément dans la construction clivée).

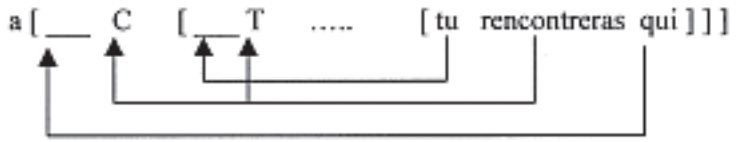
Le mouvement est une opération cruciale et constamment appliquée dans la syntaxe des langues naturelles : il est normal que tous les éléments d'une représentation abstraite se déplacent, vidant ainsi des portions importantes de la proposition, et donnant lieu parfois à un bouleversement total de l'ordre des constituants. Ainsi, l'interrogative (10b) est dérivée d'une représentation de base comme (10a) par une série de mouvements, schématiquement indiqués par les flèches en (11a), qui ont l'effet de vider complètement le syntagme verbal, au sein duquel la structure argumentaire de la phrase (qui fait quoi à qui) est établie, donnant lieu à la représentation dérivée (11b) :

(10)

a Tu rencontreras quelqu'un
S V O

b Qui rencontreras-tu?
O V S

(11)



b [Qui rencontreras [tu ___ … [____ ____ ____]]]

Les opérations de mouvement en jeu dans la dérivation d'une interrogative ne sont pas toutes du même genre. Il y a une typologie des mouvements possibles. Dans ce qui suit, je voudrais illustrer les conséquences du mouvement en prenant en considération un cas relativement moins spectaculaire que le mouvement « à longue distance » en jeu en (9), mais d'une grande importance pour éclaircir l'interaction entre syntaxe et morphologie : le mouvement du verbe dans l'espace fonctionnel de la proposition.

Le mouvement du verbe : implications diachroniques et microcomparatives

Les phrases des langues naturelles, en tout cas les phrases principales, présentent un certain événement en le situant sur l'axe temporel par rapport au moment de l'énonciation. Un élément grammatical, le temps verbal, fournit cette information. Or, le marqueur de temps est, dans certaines langues, un élément indépendant, un mot autonome, qui typiquement apparaît entre le sujet et le prédicat dans les langues VO. C'est le cas dans l'expression du futur en anglais par le modal *will* (*you will meet Mary*). Cette manière d'exprimer le temps est généralisée à tous les temps dans les langues créoles par exemple (les exemples [12] suivants sont tirés du jamaïquain[6], le marqueur du présent, le temps non marqué, étant zéro) :

(12)

a Im nuo dat
'He PRES knows that'

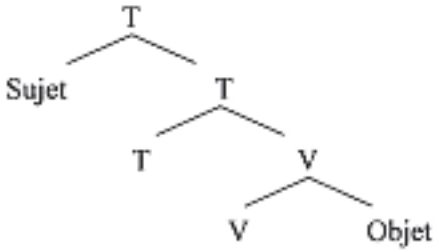
b Im en nuo dat
'He PAST know that'

c Im wi nuo dat
'He FUT know that'

6. Durrleman (2008).

Il paraît donc naturel d'admettre que la forme de la phrase exprime très directement l'articulation indiquée en (12), avec l'élément fonctionnel T qui joue le rôle de tête de la phrase (l'hypothèse d'un tel statut pour T nous permet de comprendre immédiatement le fait que dans les langues à tête finale, comme le japonais, le marqueur du temps se trouve en position finale, par exemple en [8][7]) :

(13)



Que se passe-t-il dans les langues où le temps n'est pas exprimé par un mot indépendant, mais par un affixe attaché au verbe, comme en français ?

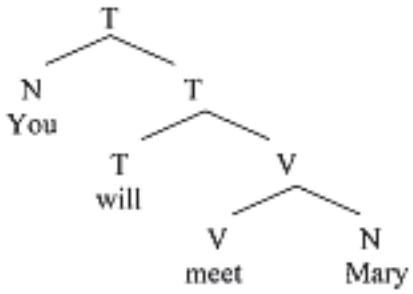
(14) Tu rencontr-eras Marie

Les études linguistiques ont suivi ici, comme dans beaucoup d'autres cas, une intuition d'uniformité. Peut-être la structure de la phrase est-elle toujours la même à travers les langues, comme indiqué en (13) (*modulo* l'ordre des mots). On a donc, par exemple, la même configuration de phrase en anglais et en français ([15] et [16]). Ce qui varie n'est pas la structure hiérarchique créée par la fusion, mais la nature morphologique de l'élément qui encode le temps. Si c'est un mot indépendant, comme dans les langues créoles, ou bien dans le cas du futur anglais exprimé par *will*, rien ne se passe dans la syntaxe, et la configuration de base est prononcée sans variation, comme en (15) ; mais si le temps est exprimé par un affixe, comme dans la structure française (16), la configuration n'est pas morphologiquement bien formée parce que *-eras* n'est pas un mot indépendant ; quelque chose doit se passer pour aligner la syntaxe et la morphologie : la racine verbale se déplace donc pour s'accorder avec la morphologie temporelle, ce qui forme le mot complexe bien formé *rencontr-eras*.

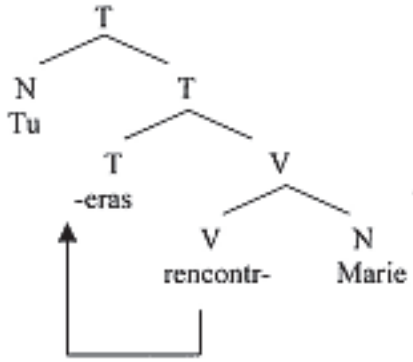
---

7. Des représentations comme (13), (8), etc. sont simplifiées à bien des égards : les structures fonctionnelles des phrases et des syntagmes sont beaucoup plus riches et articulées ; par exemple, la structure de la phrase implique aussi des marqueurs d'aspect, de mode, de voix, etc. En effet, un programme de recherche très actif dans le contexte des études syntaxiques actuelles a précisément pour but de construire une « cartographie » détaillée et réaliste des structures syntaxiques, qui rende justice à la complexité des configurations syntaxiques (Belletti [2004] ; Cinque [1999], [2002] ; Rizzi [1997], [2004]). Dans notre présentation nous nous contenterons de structures simplifiées comme (13), sans entrer dans les détails cartographiques.

(15)



(16)



Ce mouvement du verbe à la flexion temporelle (un cas particulier du mouvement d'une tête à la position de tête immédiatement supérieure) a des conséquences morphologiques – la formation d'un mot complexe – mais aussi des conséquences proprement syntaxiques : le mouvement du verbe a pour effet de le déplacer au-delà de certains adverbes comme *souvent*, qui s'interpolent donc entre le verbe fléchi et l'objet direct dans l'ordre superficiel en français.

(17)

a You will [meet Mary]

b Tu -eras [rencontr- Marie]

(18)

a You often [meet Mary]

b Tu rencontr-es souvent [___ Marie]

Emonds et Pollock[8] ont proposé d'expliquer ainsi une différence saillante entre l'anglais contemporain et le français. Tandis qu'en anglais les adverbes précèdent la séquence VO, comme en (17)a et (18)a, en français ils s'interposent systématiquement entre le verbe et l'objet, comme en (17)b et (18)b. Cette différence peut être élégamment expliquée en admettant une structure de la phrase identique dans les deux langues, une position identique des adverbes, et une seule différence paramétrique indépendante : le mouvement du verbe avant la flexion temporelle en français, mais non en anglais contemporain.

Ce n'est probablement pas par hasard si les choses sont ainsi. Un regard sur la conjugaison verbale dans les deux langues montre que le verbe français est plus richement fléchi que le verbe anglais, avec une spécification d'accord avec le sujet sur plusieurs temps verbaux (*part-ons, part-ir-ons, part-i-ons*), en contraste avec l'extrême pauvreté morphologique de la flexion verbale anglaise, laquelle inclut uniquement la marque résiduelle de l'accord de troisième personne du singulier au présent *-s*. Cette considération a alimenté l'hypothèse suivante :

(19) Hypothèse : une morphologie verbale « riche » sur T déclenche le mouvement de V.

Les linguistes ont mis beaucoup d'énergie à essayer de préciser autant que possible la notion de morphologie « riche » qui semble jouer un rôle en syntaxe[9], ainsi qu'à définir la direction du rapport de causalité : est-ce la morphologie riche qui « cause » l'attraction du verbe ? Ou bien la morphologie se limite-t-elle à enregistrer et à signaler le mouvement du verbe ? Je n'aborderai pas ce débat ici, et je me limiterai à garder la discussion sur le plan intuitif. Le point que j'aimerais aborder est plutôt qu'une hypothèse comme (19) peut être soumise à vérification empirique sur la base des données de la diachronie linguistique et comparative.

8. Emonds (1978) et Pollock (1989).
9. Vikner (1997), Roberts (1993), Bobalijk (1996).

Vérifications diachroniques et microcomparatives

On sait que les langues changent dans le temps, et le changement diachronique peut typiquement affecter la morphologie : par exemple, les paradigmes flexionnels des verbes peuvent se simplifier. L'hypothèse (19) fait donc la prédiction qu'un appauvrissement morphologique doit entraîner un changement syntaxique, la perte du mouvement du verbe. Cette prédiction a été validée par les données de l'histoire de l'anglais. Comme Roberts l'a fait remarquer[10], l'anglais jusqu'au début du XVII[e] siècle admettait le mouvement du verbe. Ceci est montré, entre autres, par l'ordre du verbe lexical avant le marqueur de négation *not*, un ordre qu'on trouve encore systématiquement en anglais shakespearien :

(20)
a I know *not*…
b go *not* to Wittenberg
c I speak *not* to him (Hamlet)

Parallèlement, à ce stade de l'histoire de l'anglais, le paradigme verbal était plus riche morphologiquement ; voilà un paradigme verbal de l'époque :

(21)

Is: cast	Ip: cast(-e)
IIs: cast-est	IIp: cast(-e)
IIIs: cast-eth	IIIp: cast(-e)

L'appauvrissement du paradigme morphologique a déterminé un changement syntaxique, l'impossibilité de déplacer le verbe lexical avant la flexion temporelle, comme on s'y attend sur la base de l'hypothèse (19).

La connexion entre la richesse de la morphologie verbale et le mouvement du verbe est aussi soutenue par des considérations comparatives. L'islandais diffère systématiquement du scandinave continental en ce qu'il a préservé une remarquable richesse de l'expression morphologique de l'accord, comme les paradigmes suivants le montrent :

(22) Islandais (heyra 'hear')
présent: heyr-i, heyr-ir, heyr-ir, heyr-um, heir-ið, heyr-a
prétérit: heyr-ði, heyr-ði-r, heyr-ði, heyr-ðu-m, heyr-ðu-ð, heyr-ðu

10. Roberts (1993).

Les langues scandinaves continentales, par contre, ont été soumises à un processus de simplification morphologique encore plus radical que l'anglais. Une seule forme verbale est utilisée pour toutes les personnes et nombres au présent et au prétérit, par exemple, en danois :

(23)

Danois (høre 'hear')
présent: hør-er
prétérit: hør-te

Parallèlement, l'islandais implique le mouvement du verbe à T, comme le montre le fait que le verbe lexical précède le marqueur de négation, tandis que, dans les langues scandinaves continentales, le verbe, typiquement, ne se déplace pas vers T et reste dans une position plus basse dans l'arbre que la négation[11] :

(24)

… að	hann	keypti	ekki	[___	bokina]	(Islandais)
' that	he	bought	not		the book'	

(25)

… at	han	ikke	[kobte	bogen]	(Danois)
' that	he	not	bought	the book'	

Il n'est pas surprenant de constater, à ce point, que la corrélation comparative a une contrepartie diachronique : le suédois du XVII[e] siècle montrait et la richesse du paradigme verbal et le mouvement du V vers T. Parallèlement, la variété scandinave parlée aux îles Féroé, qui se situent approximativement à mi-chemin entre l'Islande et le continent, est instable, avec une simplification partielle du paradigme verbal, et une variation dialectale concernant le mouvement du verbe – des caractéristiques qu'on peut vraisemblablement attendre d'un système en transition[12].

Réfléchissons un instant sur la méthodologie suivie ici. On a comparé des langues ou des variétés dialectales très proches, ou bien des couches diachroniques relativement rapprochées d'une même langue. C'est la dimension qui a été appelée microcomparative. Comme Richard Kayne

11. Dans les exemples suivants, il faut regarder l'ordre des mots dans les phrases subordonnées parce que l'ordre des principales est déterminé par un autre phénomène indépendant, le « V-2 », commun à toutes les langues germaniques sauf l'anglais, qui déplace le verbe fléchi en deuxième position, et qui « masque » les effets du mouvement du verbe à T. Pour tester les propriétés du mouvement du verbe vers T il faut donc regarder les phrases subordonnées.
12. Jonas (1996).

l'a fait remarquer[13], la microcomparaison est la meilleure approximation d'une expérience contrôlée en syntaxe comparative : en comparant des systèmes très proches, qui ont eu relativement peu de temps pour se différencier d'un antécédent commun, les linguistes ont de meilleures chances d'identifier les vrais paramètres, les propriétés irréductibles qui sont à la base de la variation syntaxique. La comparaison directe de langues très éloignées, qui se sont différenciées au cours de plusieurs millénaires, court en effet le risque de se trouver confrontée à une situation trop opaque, où les interactions entre valeurs paramétriques différentes sont trop complexes pour permettre aux facteurs primitifs de variation d'émerger avec clarté. Pour ces raisons, la recherche comparative a pris résolument la direction des études microcomparatives. L'un des résultats les plus importants de ce choix de perspective a été le développement de la dialectologie inspirée des études théoriques de la matrice paramétrique. Un grand nombre de recherches ont été consacrées à la nouvelle dialectologie syntaxique des langues romanes[14] et germaniques[15].

S'il est vrai que la recherche comparative a besoin du microscope de la microcomparaison, il ne faudrait pas pour autant perdre de vue la dimension de la variation à une plus grande échelle. Dans certains domaines privilégiés, la macrocomparaison est également possible, et fondamentale. C'est le cas, par exemple, des réinterprétations, théoriquement motivées, des découvertes empiriques des études typologiques que Guglielmo Cinque a menées récemment[16].

Le grand angle de la *macrocomparaison* est un complément indispensable du zoom de la *microcomparaison* : la première nous offre une approximation des limites quant à la variation possible, une information certes encore grossière dans laquelle l'erreur de détail (liée à la granularité grossière de l'analyse d'un grand nombre de langues) est fort possible, mais susceptible d'être affinée et corrigée au fur et à mesure ; la seconde nous procure les conditions optimales pour l'identification des paramètres irréductibles[17].

---

13. Kayne (2000), introduction.
14. Benincà et Munaro (à paraître), Manzini et Savoia (2005), Poletto (2000), Kayne (2000).
15. Haegeman (1992).
16. Voir en particulier Cinque (2005) et son analyse, au moyen des outils fins de la syntaxe formelle, de l'Universel 20 de Greenberg (1963), ainsi que les travaux de Marc Baker sur les langues polysynthétiques (Baker [1996], [2001]), et sur le caractère universel des catégories lexicales de nom, verbe et adjectif (Baker, 2003).
17. La capacité de l'approche typologique d'isoler des universaux fiables est controversée à cause de l'exiguïté de l'échantillon des langues prises en compte (Evans et Levinson

Si la recherche comparative produit les données empiriques fondamentales qui nourrissent la réflexion sur l'invariance et la variation linguistique, d'autres domaines de recherche sur le langage peuvent aussi donner des apports décisifs. Je voudrais considérer ici certaines contributions à la problématique des « règles linguistiques possibles et impossibles », qui peuvent venir de l'étude de l'acquisition de la langue maternelle, et de l'imagerie cérébrale appliquée au langage.

La variation et l'acquisition de la langue

L'acquisition de la langue maternelle est un phénomène familier, constamment sous nos yeux : qu'y a-t-il de plus naturel et attendu que le fait qu'un enfant apprenne à parler ? Néanmoins, la rapidité avec laquelle le petit enfant acquiert la langue exige une explication : comment un enfant en bas âge peut-il maîtriser un système extraordinairement complexe comme une langue naturelle ? Ce genre de considérations nous amène à postuler une prédisposition, un don biologique propre à notre espèce pour l'apprentissage d'une langue naturelle, « *an instinctive tendency to speak* », selon l'expression de Charles Darwin. L'étude de l'acquisition conduit donc à s'interroger sur la nature de cette tendance instinc-

---

[2008]) et (ce qui est à mon avis beaucoup plus important) du fait que beaucoup de langues dans l'échantillon ne sont analysées que de manière extrêmement superficielle. Néanmoins, les résultats obtenus dans les travaux cités dans le texte me semblent justifier plus d'optimisme quant à la perfectibilité de la méthode. L'article cité de Evans et Levinson attaque de manière radicale les approches universalistes, proposant pour l'essentiel un retour à la position structuraliste selon laquelle les langues peuvent varier sans limites (« *without limit and in unpredictable ways* » [Joos <1957>]). Cette conclusion me semble surestimer fortement la variation, sur la base d'analyses douteuses de la structure en constituants, de la récursivité, des principes régissant le mouvement et le liage, etc. En particulier, cette approche ne semble pas tenir compte des universaux architecturaux, concernant par exemple l'articulation en cascade des niveaux hiérarchisés de la combinatoire, et bien d'autres traits (*design features*) qui ne sont pas logiquement impliqués par le concept de langage, mais que nous retrouvons dans les langues naturelles : on peut facilement imaginer des langages artificiels qui auraient d'autres traits, qui « sauteraient » par exemple le niveau de la morphologie, ou celui de la syntaxe, ou qui auraient des règles non dépendantes de la structure (voir *infra* p. **NNN**), ou des modalités différentes d'application du mouvement (mouvement uniquement vers le fond de l'arbre, plutôt que vers le sommet, comme dans les langues naturelles), etc. La question fondamentale que la linguistique générale est censée aborder est précisément de déterminer quelles sont les dimensions de variation des langues naturelles, et pourquoi nous observons empiriquement précisément ces propriétés, et non pas d'autres traits et propriétés imaginables.

tive, et donc à se poser la question des fondements biologiques du langage. Je ne pourrai pas aborder ici ces aspects dans toute leur complexité, mais je voudrais me concentrer sur la pertinence de l'étude de l'acquisition pour la question de la variation des langues.

Quand les enfants commencent à produire les premières combinaisons de mots (peu avant 2 ans), puis dans la phase d'« explosion » de la syntaxe (dans le courant de la troisième année de la vie), ils ne parlent pas exactement comme les adultes : les productions enfantines ont une série de caractéristiques convergentes avec la langue adulte cible, mais elles diffèrent aussi systématiquement des phrases adultes ; en d'autres termes, les enfants produisent certaines « fautes » systématiques. Mais l'observation de ce que l'enfant ne fait pas est au moins aussi intéressante que celle de ce qu'il fait : il y a des types de fautes parfaitement imaginables que l'enfant ne commet pourtant pas. L'analyse des fautes systématiquement attestées et non attestées nous offre une fenêtre précieuse pour comprendre le processus d'acquisition. D'une part, l'enfant produit des structures qu'il n'entend pas : il ne se limite donc pas à enregistrer en mémoire certaines structures ; il explore certaines options qui vont au-delà de son expérience, et qui doivent donc refléter quelque nécessité interne au système enfantin. D'autre part, l'enfant semble exclure *a priori* certaines options. Ceci nous donne une image précise de l'espace grammatical exploré par l'enfant. Une observation importante de l'étude de l'acquisition est que cet espace semble correspondre à l'espace grammatical couvert par les différentes langues adultes, l'espace de la grammaire universelle. En d'autres termes, l'enfant explore, pendant quelque temps, des options grammaticales qui ne sont pas choisies dans sa langue cible, mais qu'on trouve dans d'autres langues adultes. Les productions enfantines nous donnent donc des indices sur les explorations grammaticales de l'enfant à l'intérieur de l'espace de la grammaire universelle.

Donnons un exemple concret. Les langues humaines disposent d'un certain nombre de procès d'effacement (ou de non-prononciation) du sujet de la phrase. Par exemple, le sujet peut être systématiquement omis dans les langues à sujet nul, comme l'italien ou l'espagnol (___ *parlo italiano* alterne avec *io parlo italiano*). Or on a constaté que la possibilité d'omission sélective du sujet semble être une constante du système enfantin : l'enfant entre 2 et 3 ans peut omettre systématiquement et sélectivement le sujet même si sa langue cible ne possède aucun des mécanismes d'omission du sujet. Ce phénomène est solidement attesté dans l'acquisition du français, de l'anglais, de l'allemand, du néerlandais, des langues scandinaves continentales et insulaires, etc. – c'est-à-dire des langues adultes qui n'admettent pas l'omission.

On a également observé que l'omission enfantine du sujet est structurellement sélective. Elle a lieu, fréquemment et sur une période de plusieurs mois, quand le sujet est l'élément initial, comme dans les exemples en (26) et (27), tirés de corpus de productions spontanées d'enfants anglophones et francophones ; mais non pas quand le sujet est précédé par un autre élément, par exemple un pronom interrogatif antéposé, comme en (28), un type de structure qui n'est presque pas attesté dans les corpus de productions spontanées enfantines dans l'acquisition de langues comme le français et l'anglais :

(26)
a ___ was a green one (Eve, 1;10)
b ___ falled in the briefcase (Eve 1;10)

(27)
a ___ a tout tout tout mangé (Augustin 2,0)
b ___ est où? (Augustin 2;6)

(28)
a * Où ___ est? (OK : Où il est ?)
b * Where ___ is ? (OK : where is it ?)

Une omission sélective du sujet en position initiale n'est pas une propriété du français ou de l'anglais adulte, mais c'est une propriété qu'on retrouve dans les registres familiers de plusieurs autres langues naturelles : allemand familier, portugais brésilien, variétés dialectales arabes, romanes, etc.[18]. Par exemple, Shlonsky et De Crousaz ont observé dans le patois franco-provençal parlé à Gruyère (Suisse), l'articulation suivante[19] :

(29)
a I travayè pra
'il/elle travaille beaucoup'

a' ___ travayè pra

b Portyè i travayè?
'Pourquoi il/elle travaille?'

b' * Portyè ___ travayè ?

c I travayè kan?
'il/elle travaille quand?'

c' ___ travayè kan ?

18. Rizzi (2005).
19. Shlonsky et De Crousaz (2000).

Le pronom sujet peut être omis, quand il est en position initiale, dans les déclaratives comme (29a), et dans les questions *in situ*, sans mouvement du pronom interrogatif, comme (29c) ; mais le sujet ne peut pas être omis dans une question comme (29b), où la position initiale est occupée par l'élément interrogatif qui correspond à *pourquoi*.

Cette option d'omission initiale, qui caractérise plusieurs systèmes adultes, se retrouve très généralement dans les systèmes enfantins, même quand la langue cible ne l'adopte pas. L'enfant explore donc une véritable option de la grammaire universelle, et il l'abandonne seulement plus tard, sur la base de la pression de l'expérience, si les données auxquelles il a accès ne lui permettent pas de confirmer cette option[20].

On peut identifier ici une sorte d'analogue syntaxique d'une modalité d'apprentissage bien documentée et décrite dans le domaine de la phonologie : l'« apprentissage par l'oubli sélectif ». Le bébé est initialement sensible à toutes les distinctions phonétiques utilisées par les langues du monde pour distinguer les mots (par exemple la distinction /l/ - /r/, qui permet de distinguer des mots comme *lire* et *rire* en français), mais à la fin de la première année de la vie il n'est plus sensible qu'aux distinctions utilisées dans la langue à laquelle il est exposé : il semble avoir « oublié » les autres distinctions phonétiques (par exemple, l'enfant exposé au japonais perd la distinction /l/-/r/, qui n'est pas utilisée par sa langue)[21]. L'exploration de l'espace phonétique, suivie par la sélection des distinctions conformes à l'expérience, semble donc avoir un analogue dans le cas de la syntaxe : certaines options non conformes à celles de la langue cible, mais attestées dans d'autres langues, sont explorées par l'enfant pendant quelque temps et ensuite abandonnées. L'apprentissage prend donc la forme d'une sélection entre des ressources disponibles initialement, sur la base de l'expérience.

Ces considérations portent sur des questions fondamentales concernant la nature de l'apprentissage, en ce qu'elles apportent un soutien empirique aux modèles « sélectifs » comportant une restriction progressive des ressources par rapport à l'inventaire initial[22]. Elles éclairent également la question de la variation des langues. L'apprenant explore certaines régions de l'espace grammatical concevable, mais pas d'autres. Par exem-

---

20. Il resterait encore à comprendre quelle pression interne induit l'enfant à explorer une option compatible avec la grammaire universelle, mais qui n'est pas validée par l'expérience. Un facteur vraisemblablement en jeu est la nécessité de minimiser la séquence prononcée afin de faciliter la tâche à un système de production qui n'a pas atteint la maturité (Rizzi, 2005).
21. Voir Werker *et al.* (1981) pour la découverte de cet effet avec d'autres distinctions, et Piattelli-Palmarini (1989), Mehler et Dupoux (1991) pour une discussion générale.
22. Changeux (2002).

ple, une omission généralisée et non sélective de toutes les formes pronominales, indépendamment du contexte syntaxique, n'est pas prise en considération par l'enfant, et, parallèlement, elle n'est pas attestée à travers les langues adultes (des cas apparents d'omission généralisée du pronom, par exemple en chinois et dans d'autres langues de l'Extrême-Orient, montrent, dans le cadre d'une analyse plus raffinée, des formes importantes de sensibilité à des contraintes syntaxiques fines)[23]. L'étude des « fautes » attestées et non attestées ouvre donc une nouvelle fenêtre sur la variabilité linguistique et, par conséquent, directement sur la genèse de la grammaire chez l'enfant.

Règles possibles et impossibles et l'imagerie cérébrale

Reformulons la question des limites sur la variation de manière légèrement différente : les langues humaines utilisent seulement un petit sous-ensemble, une petite zone continue, au sein de la vaste classe des règles formelles concevables. Seules les règles qui peuvent être ramenées aux opérations fondamentales de la fusion et du mouvement, et les paramètres qui peuvent y être associés, sont utilisées par les langues. Une sorte de métaprincipe, qui découle de ce qui précède, est ce qu'on a appelé « la dépendance de la structure » : les règles des langues humaines ne manipulent jamais les mots de façon structuralement arbitraire, mais toujours en fonction de leur organisation hiérarchique dans les structures créées par la fusion. Par exemple :

(30)

- Dans aucune langue humaine on n'obtient l'interrogative en organisant les mots en image miroir de l'ordre de la déclarative (c.à.d. avec une transformation déclarative → interrogative qui donnerait *la voiture part → part voiture la* ?).
- Dans aucune langue humaine on n'obtient une phrase impérative en permutant les mots pair et impairs (c.à.d. *tu ouvres la porte → ouvres tu porte la*).
- Dans aucune langue on n'obtient la négative en plaçant le marqueur négatif dans une position fixe ene terme de nombre de mots, par ex. en position de quatrième mot (c.à.d. où les phrases *la fille achète le livre, Marie achète le livre, la fille de Marie achète le livre* auraient comme contreparties négatives *la fille achète PAS le livre, Marie achète le PAS livre, la fille de PAS Marie achète le livre*).

De telles règles sont formellement simples : on pourrait facilement programmer un langage artificiel spécifié de cette manière, et nous pouvons les apprendre sans difficulté particulière, en dehors de l'apprentis-

23. Huang (1984).

sage linguistique, dans le contexte d'un jeu de société par exemple. Mais aucune langue humaine ne les utilise. Nous devons donc nous poser la question de la nature et de l'origine de cette exclusion.

La distinction entre règle linguistique possible et règle formellement concevable mais linguistiquement impossible est-elle un pur accident historique ? *A priori*, on pourrait imaginer qu'une hypothétique « langue humaine primordiale » se soit organisée de cette manière pour des raisons accidentelles, et que certaines propriétés, dont la dépendance de la structure, se soient simplement préservées dans toutes les langues issues de la première : un simple cas de transmission culturelle. C'est une possibilité logique : les universaux linguistiques pourraient tous avoir une origine liée simplement à l'histoire « externe » des langues, sans qu'il soit nécessaire de postuler un mécanisme interne pour les expliquer.

Néanmoins, l'observation des modalités de l'acquisition chez l'enfant suggère que ce genre d'approche ne peut pas être correct. On a vu que l'enfant explore certaines possibilités qui ne se conforment pas immédiatement aux données qu'il entend, et il en écarte *a priori* d'autres (par exemple, les règles indépendantes de la structure) ; il semble donc suivre une logique et des contraintes internes dans sa construction du système des connaissances linguistiques. Ces considérations rendent plausible une autre possibilité : peut-être notre cerveau est-il prédisposé à apprendre et à utiliser pour le langage seulement un petit sous-ensemble de règles, celles qui dépendent de la structure et qui sont en dernière analyse susceptibles d'être ramenées aux opérations élémentaires de la fusion et du mouvement. Y aurait-il donc des circuits neuronaux dédiés à l'utilisation des règles linguistiquement possibles ? Les nouvelles techniques de neuro-imagerie nous offrent des moyens pour commencer à aborder ce genre de question.

Dans certains travaux récents, des équipes de linguistes, de neuropsychologues et de spécialistes de l'imagerie[24] ont obtenu des résultats prometteurs sur la question. Ces auteurs ont appris à des sujets locuteurs monolingues de l'allemand des langues inventées qui étaient « hybrides » par rapport à la distinction entre règles possibles et impossibles. Ces langues utilisaient des lexiques de langues réelles (de l'italien et du japonais) et mélangeaient des règles linguistiques possibles (comme l'omission du sujet, certaines propriétés d'ordre des mots et d'accord grammatical), et des règles impossibles (la formation d'interrogatives et de négatives non

24. Tettamanti, Alkadhi, Moro, Perani, Kollias et Weniger (2002) ; Musso, Moro, Glauche, Rijntjes, Reichenbach, Buechel et Weiller (2003) ; les expériences sont reprises et commentées dans Moro (2008).

dépendantes de la structure, comme en [30]). Ils ont ensuite observé, en IRM fonctionnelle, les configurations d'activation cérébrale lorsque les sujets apprenaient les règles linguistiquement possibles ou impossibles.

Les auteurs ont observé dans l'aire de Broca une augmentation progressive de l'activation dans le traitement de phrases qui impliquaient des règles possibles, et une diminution de l'activation dans le traitement de phrases qui impliquaient des règles impossibles, corrélée à l'amélioration progressive de la performance des sujets, indiquant la maîtrise progressive des deux types de règles. La conclusion que les auteurs en ont tirée est que l'aire de Broca fait partie d'un circuit recruté pour le traitement des règles sensibles à la structure, s'appliquant sur des configurations hiérarchiques du type créé par la fusion, et en respectant la structure ; tandis qu'elle n'est pas spécifiquement recrutée pour des règles dont la computation est indépendante de la structure. Ces dernières règles peuvent être apprises par nos capacités générales d'apprentissage, que nous pouvons utiliser pour apprendre toutes sortes de régularités en dehors du langage, mais elles ne sollicitent pas les structures cérébrales sélectivement dédiées au langage naturel.

Ces observations apportent une nouvelle contribution à un débat classique. Elles ouvrent aussi des questions ultérieures : est-ce que les circuits activés dans l'usage des « règles possibles » sont effectivement dédiés aux computations purement linguistiques ? Ou bien sont-ils recrutés dans toutes les computations mentales qui impliquent des objets organisés hiérarchiquement, par exemple dans les computations arithmétiques[25] ? Ou encore, dans la constitution des circuits neuronaux en jeu dans certaines capacités spécifiques de notre espèce, quel est le rôle respectif des mécanismes sélectifs classiques et du « recyclage neuronal[26] », la réutilisation des circuits préexistants pour des innovations culturelles ? Ces questions, et bien d'autres, seront sans doute abordées par les études sur le langage et le cerveau dans un futur proche. Ce qui rend fascinante cette étape de la recherche, c'est le fait que des communautés scientifiques distinctes de linguistes, psycholinguistes, neuropsychologues, spécialistes de l'imagerie ont commencé à interagir à un certain niveau de profondeur, jusqu'à dessiner ensemble des expériences tenant compte des expertises des uns et des autres. On peut attendre de cet effort commun des progrès décisifs sur les questions cruciales des fondements biologiques du langage.

25. Dehaene, 1997
26. Dehaene et Cohen (2007)

Conclusion

La question de la diversité des langues ne peut être abordée de manière scientifiquement intéressante que dans le contexte de la question complémentaire de l'uniformité du langage. C'est seulement dans le cadre d'une théorie générale du langage que la question de la diversité des langues acquiert une dimension descriptive précise, et peut être ramenée à un système de principes explicatifs. La linguistique théorique a, d'une part, constitué des modèles précis des computations linguistiques communes à toutes les langues, réduites à certaines opérations récursives de base. D'autre part, les études formelles ont élaboré des modèles de la variation linguistique par le biais de la notion de paramètre.

Une théorie précise de l'invariance et de la variation peut être soumise à plusieurs types de contrôles et de validations empiriques. Les études microcomparatives – y compris le cas méthodologiquement optimal de la dialectologie formelle, ainsi que la description et l'analyse du changement linguistique – créent les meilleures conditions pour l'identification des paramètres (les points de choix ultimes et irréductibles). Les études macrocomparatives, avec l'identification de certaines généralisations empiriques de base par les études typologiques, et la possibilité de réinterprétation des généralisations en termes théoriquement plus profonds, nous offrent une image d'ensemble de la variation linguistique, qu'il convient de relier aux études beaucoup plus raffinées sur le plan descriptif et théorique de la perspective microparamétrique. L'étude de l'acquisition du langage nous permet de voir les questions de l'invariance et de la variation sous un angle particulier et privilégié : celui de la genèse des compétences grammaticales chez l'enfant, des hypothèses qu'il explore et de celles qu'il écarte *a priori*. Les techniques récentes de la neuro-imagerie nous permettent d'enrichir les recherches de la neurolinguistique classique en approfondissant l'étude des circuits neuronaux impliqués dans les règles linguistiques possibles et impossibles, ouvrant ainsi une nouvelle dimension à la recherche sur les bases biologiques de l'invariance et de la variation à travers les langues. Le langage est un objet extraordinairement complexe, à la frontière entre nature et culture, bases biologiques et histoire. Nous ne pouvons espérer arriver à une compréhension scientifiquement satisfaisante de ses mécanismes que par les efforts coordonnés et intégrés de plusieurs disciplines, de la description comparative et diachronique des langues à la modélisation formelle, de l'expérimentation psy-

cholinguistique sur l'acquisition et le comportement linguistique adulte aux études par imagerie des circuits neuronaux impliqués dans le langage.

RÉFÉRENCES

– Baker M. (1996), *The Polysynthesis Parameter*, Oxford/New York, Oxford University Press.
– Baker M. (2001), *The Atoms of Language*, New York, Basic Books.
– Baker M. (2003), *Lexical Categories. Verbs, Nouns and Adjectifs*, Cambridge, Cambridge University Press.
– Belletti A (éd.) (2004), *Structures and Beyond. The Cartography of Syntactic Structures,* vol. 3, Oxford University Press.
– Benincà P. et Munaro N. (éd.) (à paraître), *Mapping the Left Periphery. The Cartography of Syntactic Structures,* vol. 5, New York, Oxford University Press.
– Bobaljik J. (1995), *Morphosyntax. The Syntax of Verbal Inflection*, PhD dissertation, MIT.
– Changeux J.-P. (2002), *L'Homme de vérité*, Paris, Odile Jacob.
– Chomsky N. (1957), *Syntactic Structures*, La Haye, Mouton.
– Chomsky N. (1981), *Lectures on Government and Binding*, Dordrecht, Foris Publications.
– Chomsky N. (1995), *The Minimalist Program*, Cambridge (Mass.), MIT Press.
– Cinque G. (1999), *Adverbs and Functional Heads*, New York, Oxford University Press.
– Cinque G. (éd.) (2002), *The Structure of DP and IP. The Cartography of Syntactic Structures,* vol. 1, New York, Oxford University Press.
– Darwin C. ([1871] 1981), *The Descent of Man*, Princeton, Princeton University Press.
– De Crousaz I. et Shlonsky U. (2003), « The distribution of a subject clitic pronoun in a Franco-Provençal dialect », *Linguistic Inquiry*, 34, p. 413-442.
– Dehaene S. (1997), *La Bosse des maths*, Paris, Odile Jacob.
– Dehaene S. et Cohen L. (2007), « Cultural recycling of cortical maps », *Neuron*, 56, p. 384-398.
– Descartes R. ([1637] 1951), *Discours de la méthode*, Paris, Union générale d'éditions.
– Emonds J. (1978), « The verbal complex V'-V in French », *Linguistic Inquiry*, 9, p. 151-175.
– Evans N. et Levinson S. (2008), « The myth of language universals : Language diversity and its important for cognitive science », ms., Nijmegen, Max Planck Institute.
– Galilei G ([1630] 1970), *Dialogo sopra I due massimi sistemi del mondo*, Turin, Einaudi.
– Greenberg J. (1963), « Some universals of grammar with particular reference to the order of meaningful elements », *in* Greenberg J. (éd.), *Universals of Language*, Cambridge (Mass.), MIT Press, p. 73-113.

– Grodzinsky Y. (2000), « The neurology of syntax : Language use without Broca's area », *Behavioral and Brain Sciences*, 23, p. 1-71.
– Haegeman L. (1992), *Theory and Description in Generative Syntax*, Cambridge, Cambridge University Press.
– Hauser M., Chomsky N. et Fitch T. (2002), « The faculty of language : What is it, who has it, and how did it evolve ? », *Science*, 298, 2002.
– Hyams N. (1986), *Language Acquisition and the Theory of Parameters*, Dordrecht, Reidel.
– Jonas D. (1996), *Clause Structure and Verb Syntax in Scandinavian and English,* PhD dissertation, Harvard University.
– Huang C.-T. J. (1984), « On the distribution and reference of empty pronouns », *Linguistic Inquiry*, 15, p. 531-574.
– Joos M. (1957), *Readings in Linguistics I.*, New York, American Council of Learned Society.
– Kayne R. (1983), *Connectedness and Binary Branching*, Dordrecht, Foris Publications.
– Kayne R. (1994), *The Antisymmetry of Syntax*, Cambridge (Mass.), MIT Press.
– Kayne R. (2000), *Parameters and Universals*, Oxford/New York, Oxford University Press.
– Mancini M. R. et Savoia L. (2005), *I dialetti italiani e romanzi*, Alessandria, Edizioni dell'Orso.
– Mehler J. et Dupoux E. (1990), *Naître humain*, Paris, Odile Jacob.
– Moro A. (2008), *The Boundaries of Babel*, Cambridge, MIT Press.
– Musso M., Moro A., Glauche V., Rijntjes M., Reichenbach J., Buechel C. Weiller C., (2003), « Broca's area and the language instinct », *Nature Neuroscience*, vol. 6, p. 774-781.
– Piattelli-Palmarini M. (1989), « Evolution, selection and cognition : from "learning" to parameter setting in biology and the study of language », *Cognition*, 31, p. 1-44.
– Poletto C. (2000), *The Higher Functional Field. Evidence from Northern Italian Dialects,* New York, Oxford University Press.
– Pollock J.-Y. (1989), « Verb movement, universal grammar, and the structure of IP », *Linguistic Inquiry*, 20, p. 365-424.
– Rizzi L. (1982), *Issues in Italian Syntax*, Dordrecht, Foris Publications.
– Rizzi L. (1990), *Relativized Minimality*, Cambridge (Mass.), MIT Press.
– Rizzi L. (1997), « The fine structure of the left periphery », *in* Haegeman L. (éd.), *Elements of Grammar*, Dordrecht, Kluwer, p. 281-338.
– Rizzi L. (2000), *Comparative Syntax and Language Acquisition*, Londres, Routledge.
– Rizzi L. (2004), *The Structure of CP and IP. The Cartography of Syntactic Structures,* vol. 2, New York, Oxford University Press.
– Rizzi L. (2005), « Grammatically-based target-inconsistencies in child language », *in* Deen K. U., Nomura, J., Schulz B. et Schwartz B. D. (éd.) (2006), *The Proceedings of the Inaugural Conference on Generative Approaches to Language Acqui-*

sition. North America (GALANA), UCONN/MIT Working Papers in Linguistics, Cambridge (Mass.), MIT Press.
– Roberts, I. (1993), *Verbs and Diachronic Syntax*, Dordrecht, Kluwer.
– Saussure F. de ([1916] 1985), *Cours de linguistique générale*, Paris, Payot.
– Tettamanti M., Alkadhi H., Moro A., Perani D., Kollias S. et Weniger D. (2002), « Neural correlates for the acquisition of natural language syntax », *NeuroImage*, 17, p. 700-709.
– Vikner S. (1997), « V to I and inflection for person in all tenses », *in* Haegeman L. (éd.), *The New Comparative Syntax,* Longman, p. 189-213.
– Werker J. F., Gilbert J. H. V., Humphreys G. W. et Tees R. C. (1981), « Developmental aspects of cross-language speech perception », *Child Development*, 52, p. 349-355.

Paroles et musique dans le chant : Échec du dialogue ?

par Isabelle Peretz et Régine Kolinsky

Du fado aux chromosomes

RÉGINE KOLINSKY – Écoutons ce disque de *fado bailado*. Les musiciens du groupe portugais Rão Kyão ont eu l'idée de remplacer la voix humaine, souvent déchirante, du chanteur (*fadista*), par le timbre tout aussi envoûtant du saxophone. N'est-ce pas une illustration frappante de la similitude entre langage et musique ?

ISABELLE PERETZ – C'est vraiment très beau… Mais le fait que la voix et un instrument de musique puissent éveiller en nous les mêmes sentiments, les mêmes émotions, ne prouve rien quant aux liens entre langage et musique. Il est à la mode de s'interroger sur ce que ces deux systèmes partagent ; certains trouvent de nombreuses similitudes, et en arrivent à conclure que les deux systèmes recouvrent les mêmes fonctions. Pourtant, il est tout aussi important de tenter de comprendre les divergences entre langage et musique[1] car ces différences ont des implications pour l'étude de la musique en général, et pour celle de ses origines en particulier.

R. K. – La recherche sur les origines animales du comportement musical, bien que très prometteuse, n'en est qu'à ses balbutiements[2]. Par contre, il est intéressant de nous pencher sur les études génétiques.

I. P. – Exactement, parce que si la musique et le langage impliquent vraiment des fonctions similaires, ayant des origines similaires, un point de départ serait par exemple d'étudier les gènes qui ont déjà été identifiés

1. *Cf.* Peretz et Morais (1989) ; Peretz (2006).
2. Peretz, (2006).

comme impliqués dans la parole, et d'examiner s'ils le sont aussi dans la musicalité. Partons donc du langage et du gène *FOXP2*. L'idée que ce gène puisse être impliqué dans la parole a émergé lors de l'étude d'une famille anglaise d'origine pakistanaise, dont la moitié des membres, à travers trois générations, souffre d'un trouble de la parole et du langage[3]. Dans cette famille, environ la moitié des enfants des personnes affectées par ce trouble en souffre aussi, alors qu'aucun enfant des personnes non affectées n'est lui-même atteint. Ce trouble héréditaire a pu être rattaché à une anomalie siégeant dans un petit segment du chromosome 7[4]. La découverte, par hasard, d'un jeune garçon non apparenté à la famille précédente, et qui présentait un déficit de parole similaire, a permis d'en identifier la cause. Une mutation d'un gène de cette région du chromosome 7, *FOXP2*, le rend non fonctionnel chez ces individus[5]. Le gène *FOXP2* semble ainsi jouer un rôle dans le développement des réseaux cérébraux qui sous-tendent la parole et le langage[6].

R. K. – Mais n'oublions pas que le trouble présent dans la famille que nous venons d'évoquer n'est pas spécifique du langage, et porte aussi sur les praxies bucco-faciales notamment[7]. Au-delà de problèmes linguistiques plus généraux, ces personnes semblent en effet incapables de réaliser les mouvements fins de la langue et des lèvres qui sont nécessaires à la production d'une parole intelligible, à tel point que même les membres non atteints de leur famille ne les comprennent pas…

I. P. – C'est justement la question : on peut se demander si la mutation du gène *FOXP2* affecte aussi leurs aptitudes vocales dans le chant, par exemple. Et c'est le cas ! Alcock et ses collaborateurs[8] ont examiné neuf membres de cette famille et ont montré qu'ils avaient également des anomalies de la production et de la perception du rythme.

R. K. – Musique et parole pourraient donc quand même avoir des origines communes. N'est-ce pas un argument qui s'oppose au point de vue modulaire, selon lequel il y aurait des processus spécifiquement dédiés au traitement de la musique ?

I. P. – Pas forcément. Nous avons montré qu'en musique il y a indépendance entre la perception et la production du rythme, d'une part, et des variations mélodiques, d'autre part, c'est-à-dire des variations musicales fondées sur la hauteur des notes. À la suite d'une lésion cérébrale, une

3. Hurst *et al.* (1990).
4. Fisher *et al.* (1998) ; Hurst *et al.* (1990).
5. Lai *et al.* (2001).
6. Marcus et Fisher (2003).
7. Vargha-Khadem *et al.* (1995).
8. Alcock *et al.* (2000).

patiente, que nous avons examinée ensemble en Belgique, et un autre cas similaire, que j'ai rencontré au Canada, étaient tous deux très affectés tant dans la production que dans la perception des variations mélodiques, alors que pour le rythme, ces patients ne rencontraient aucune difficulté[9]. La situation inverse a aussi été observée, puisque certains individus, à la suite d'une lésion, présentent un trouble qui concerne le rythme mais pas la mélodie[10]. Or, et c'est le point crucial, les membres de la famille évoquée précédemment, même s'ils étaient affectés dans la production (et la perception) du rythme, étaient tout aussi performants que des individus sains dans celles de la mélodie[11].

R. K. – Bien sûr. D'ailleurs, pour en revenir aux troubles congénitaux, les travaux de l'équipe que tu diriges sur les cas d'amusie congénitale suggèrent aussi que les aptitudes musicales fondées sur la hauteur des notes sont guidées par des facteurs génétiques distincts. En effet, contrairement aux membres de la famille dont il a déjà été question, ces personnes qui, toute leur vie durant, présentent d'importantes difficultés à apprécier la musique et à s'engager dans des activités musicales, souffrent en réalité de troubles de la perception et de la production qui ne portent que sur les variations mélodiques, et pas sur le rythme[12].

I. P. – Nous avons effectivement montré que les personnes qui souffrent d'amusie congénitale présentent d'importantes difficultés dans toutes les tâches qui reposent sur l'organisation séquentielle de la hauteur des notes, mais qu'elles n'ont pas nécessairement de problème avec les intervalles temporels[13]. Leur trouble est en revanche manifeste lorsqu'on leur demande de détecter une « fausse note » dans une mélodie par ailleurs conventionnelle[14]. Ce trouble de perception de la hauteur des sons musicaux (en anglais, *pitch*) est lui aussi héréditaire[15]. Et les personnes identifiées jusqu'à présent comme souffrant d'amusie congénitale ne présentent pas de trouble de la parole…

R. K. – Donc, les données dont nous disposons suggèrent qu'il y aurait deux facteurs innés à l'origine des aptitudes musicales, l'un en rapport avec les caractéristiques temporelles de la musique (et dans lequel le gène *FOXP2* est vraisemblablement impliqué), et l'autre en rapport avec la hauteur des notes (pour lequel les gènes impliqués restent à déterminer).

---

9. Peretz *et al.* (1994) ; Peretz et Kolinsky (1993).
10. *Cf.* Di Pietro *et al.* (2004).
11. Alcock *et al.* (2000a).
12. *Cf.* Ayotte *et al.* (2002a).
13. Hyde et Peretz (2004).
14. Ayotte *et al.* (2002).
15. Peretz *et al.* (2007).

I. P. – Oui. Et donc, comme nous venons de l'illustrer, la comparaison entre musique et parole est précieuse car elle constitue un point de départ vers l'identification des facteurs génétiques qui contribuent aux capacités potentiellement partagées par ces deux domaines, ou au contraire, qui sous-tendent uniquement les aptitudes musicales. Les facteurs génétiques mis en jeu uniquement dans les aptitudes musicales agissent vraisemblablement sur l'aptitude à percevoir la hauteur des notes. Ils pourraient toutefois ne pas être vraiment spécifiques de la musique, et être aussi impliqués par exemple dans la prosodie du langage, la « musique du langage ».

R. K. – Ce que tu viens de dire soulève deux questions fondamentales : celle de savoir ce qu'il faut comparer entre musique et langage, et celle de déterminer comment nous pouvons évaluer la spécificité de chacun de ces domaines.

I. P. – Cette question de la « modularité » de traitement a souvent été discutée en ce qui concerne la perception[16]. Il serait intéressant d'examiner l'autre volet et d'estimer l'adéquation d'une position modulaire par rapport à l'action, c'est-à-dire la production de la parole et du chant. Étonnamment, cet aspect est plus rarement évoqué !

R. K. – Mais la notion contemporaine de modularité a elle aussi évolué. De nos jours, elle ne recouvre plus nécessairement le même concept que celui qui avait été proposé par Fodor en 1983 dans son livre *La Modularité de l'esprit*[17]. Avant d'évoquer cette question, changeons de disque ; les *Suites pour violoncelle seul*, de Bach me paraissent appropriées pour la suite de la discussion.

Prélude : modularité ou spécificité

I. P. – Depuis le livre de Fodor, le concept de modularité a effectivement fort évolué. Ce concept a malheureusement alimenté plusieurs débats qui sont restés sans réponse, tant en ce qui concerne le langage[18] que d'autres domaines comme celui du traitement de l'information liée à

16. *Cf.* Peretz (2001) ; Peretz et Coltheart (2003) ; Justus et Hutsler (2005) ; McDermott et Hauser (2005).
17. Jerry Fodor, *The Modularity of Mind. Essay on Faculty Psychology*, MIT Press, 1983 ; *La Modularité de l'esprit. Essai sur la psychologie des facultés,* Paris, Éditions de Minuit, 1986.
18. *Cf.* Liberman et Whalen (2000).

la perception visuelle des visages d'autrui[19]. Et ce concept a aussi « contaminé » les études sur la musique. Il est dès lors important, lorsqu'on s'intéresse à des questions comme celle de la spécialisation du traitement de l'information, la « spécificité de domaine », la localisation cérébrale, ou le caractère inné de certaines aptitudes, de distinguer soigneusement et de clarifier des notions qui sont souvent confondues[20], car elles étaient interconnectées dans la proposition originelle de Fodor. De nos jours, de toutes les caractéristiques de la modularité qui ont été discutées par Fodor, la spécificité au domaine reste la plus importante[21]. En effet, une opération spécifique à un domaine constitue un mécanisme distinct qui traite d'un aspect particulier de l'entrée sensorielle, soit exclusivement, soit beaucoup plus efficacement que n'importe quel autre mécanisme.

R. K. – En d'autres termes, ce qui individualise un module, c'est sa spécialisation fonctionnelle[22]. Actuellement, la plupart des scientifiques seraient sans doute d'accord pour considérer que l'esprit humain comporte des systèmes fonctionnels distincts, par exemple un pour la perception, et un autre pour le contrôle de la motricité. La question est de savoir si, à l'intérieur de ces grands systèmes, il existe des spécialisations fonctionnelles, dans le cas qui nous occupe, pour la musique et pour la parole.

I. P. – Cette description reste cependant trop classique : elle ne considère que la possibilité d'une spécialisation fonctionnelle en relation avec une faculté mentale considérée dans sa globalité, comme la musique ou le langage. Or, la spécialisation fonctionnelle pourrait concerner des processus beaucoup plus élémentaires. En d'autres termes, il n'y a aucune raison de ne pas considérer la possibilité d'une spécificité à l'échelle des composantes de chaque grand domaine envisagé[23]. Un « domaine » peut être aussi large et général que l'analyse de la scène auditive, ou aussi étroit et spécifique que l'encodage tonal de la hauteur des notes. Ces deux sous-systèmes réalisent des opérations hautement spécialisées et, en ce sens, sont spécifiques à un domaine. Tous deux traitent d'un aspect particulier de la musique et le font exclusivement ou plus efficacement que n'importe quel autre mécanisme. Cependant, l'analyse de la scène auditive est censée intervenir pour tous les types de son[24], alors que l'encodage tonal de la hauteur des notes n'intervient, lui, que pour la musique.

19. *Cf.* Gauthier et Curby (2005).
20. *Cf.* Peretz (2006).
21. *Cf.* Peretz et Coltheart (2003).
22. Barrett et Kursban (2006).
23. Coltheart (1999).
24. Bregman (1990).

R. K. – Donc, la spécificité au domaine n'implique pas nécessairement la spécificité pour la musique ou pour le langage. Ce point de vue a comme conséquence que l'hypothèse de spécificité pour la musique doit être examinée au niveau de chaque sous-système ou composante de traitement.

I. P. – C'est exact. Si l'on considère le chant, la question qui nous intéresse est de savoir dans quelle mesure le traitement de la musique repose sur des mécanismes distincts de ceux impliqués dans le traitement du langage, ou si certaines composantes sont partagées. Il reste possible que chanter n'implique aucune composante spécifique à la musique… La musique pourrait en fait agir comme un parasite de la parole, par exemple en faisant intervenir dans le chant les mécanismes responsables du traitement de l'intonation linguistique. On pourrait imaginer que la musique soit au langage ce que des masques artistiques sont au système de reconnaissance des visages. Et nous pourrions même étendre l'argument, en disant que la musique doit son efficacité au fait qu'elle « cannibalise » nos dispositions naturelles pour la parole, en en exagérant des caractéristiques particulières comme l'intonation et la tonalité affective, qui sont si efficaces pour renforcer les interactions entre les êtres humains, et donc pour la cohésion sociale. On pourrait dès lors considérer que le « domaine propre » du module du langage a été envahi[25]. Et, pour continuer dans cette voie, nous pourrions spéculer sur le fait que la musique se serait établie dans toutes les cultures car elle est particulièrement efficace pour coopter un ou plusieurs modules évolués. L'ancrage multiple dans divers modules pourrait même contribuer à l'ubiquité et au pouvoir de la musique.

R. K. – Cette situation de « parasitage » est plus plausible dans le cas des masques, ou d'une autre invention strictement culturelle, que dans des systèmes qui, tous deux, ont une base biologique bien ancrée, ce qui semble être le cas de la musique et du langage. De plus, même certaines inventions culturelles pourraient mener au développement, au cours de la vie, de systèmes de traitement hautement spécialisés, de type modulaire. Ce processus ontogénétique de « recyclage neuronal[26] » correspond à ce qui est appelé une « exaptation » en biologie, lors de l'adaptation phylogénétique[27]. Or une exaptation est précisément une adaptation sélective dans laquelle la fonction actuellement remplie n'est pas celle qui était remplie initialement. Dans le cas du cerveau, une aire cérébrale particulière est ainsi réutilisée pour une fonction non sélectionnée à l'origine, qui tire parti de certaines composantes ou caractéristiques du ou des systèmes

25. Sperber et Hirschfeld (2004).
26. Dehaene (2007).
27. Gould et Vrba (1982).

préexistants. Cette idée a déjà été défendue dans le domaine de l'apprentissage de la lecture, qui est un acquis culturel mais mène à des processus hautement spécialisés[28].

I. P. – Oui, la question peut être reformulée en termes de « recyclage neuronal ». Nous pouvons nous demander si la musique n'est pas une façon très agréable de recycler les circuits neuronaux impliqués dans le langage. C'est à cette notion de parasitage que fait référence Pinker[29] lorsqu'il parle de la musique comme d'une pâtisserie auditive. Cette notion de parasitage peut être testée. Il y a plusieurs façons de tester la spécificité (ou la modularité) de la production et de la perception de la musique. En ce qui concerne la production de la parole et du chant, nous devrions envisager quatre types d'arguments :

(i) les dissociations entre langage et musique qui ont été observées en neuropsychologie, chez des individus porteurs d'une lésion cérébrale ;

(ii) les profils d'activation d'aires cérébrales communes pour la musique et le langage observés dans les études d'imagerie cérébrale fonctionnelle du cerveau d'individus sains ;

(iii) les effets d'interférence observés chez les sujets sains soumis à une stimulation magnétique transcrânienne ;

(iv) les effets de transfert entre domaines, c'est-à-dire entre les aptitudes musicales et celles liées à la parole.

Par souci d'objectivité, il nous faudra aussi examiner une autre approche théorique, celle qui considère que les ressources neurales nécessaires au traitement de l'information pourraient être partagées[30]. Mais commençons par passer en revue les résultats des tests de la spécificité de domaine pour la production du chant et de la parole.

Sarabande : tests de la spécificité

I. P. – Avant d'examiner les divers tests de spécificité, il faut rappeler que le chant n'est pas, contrairement à ce que beaucoup de personnes croient, une activité réservée à une élite. Tout le monde, ou presque, peut chanter. Des chanteurs occasionnels peuvent même atteindre le niveau d'aptitude des chanteurs professionnels[31].

---

28. Morais et Kolinsky (2002 ; Dehaene (2007).
29. Pinker (1997).
30. Patel, (2003, 2008, sous presses).
31. Dalla Bella *et al.* (2007).

R. K. – L'idée que le chant constitue une disposition naturelle de l'être humain est d'ailleurs beaucoup plus cohérente avec le fait que le chant est universel, puisqu'il se retrouve dans toutes les cultures. De plus, chanter est une activité de groupe. Par sa nature participative, qui exige une coordination de l'action, le chant est aussi une expérience très agréable. C'est pourquoi chanter est une aptitude humaine fondamentale, dont on pense qu'elle renforce et même stimule la cohésion du groupe[32]. L'importance sociale du chant est bien illustrée par le fait que toutes les mères, dans toutes les cultures, chantent pour leurs enfants. Par ailleurs, le chant apparaît précocement et « spontanément » au cours du développement de l'enfant.

I. P. – Les premières chansons sont en effet produites vers la fin de la première année de vie, et, dès 18 mois, les enfants commencent à produire des chansons reconnaissables[33]. Cette expertise initiale se retrouve dans le chant produit ou reproduit par des adultes, dont les caractéristiques sont remarquablement constantes, tant chez un individu donné[34] qu'entre personnes différentes[35], en termes de hauteur des notes de départ et de tempo. Donc, la population adulte semble posséder les capacités de base nécessaires au chant populaire.

R. K. – Étant donné son universalité, son émergence précoce et spontanée, sa constance et sa fonction sociale, le chant constitue l'une des sources les plus riches d'information sur la nature et les origines du comportement musical. De plus, les chansons présentent une combinaison unique de paroles et de musique, ce qui en fait un matériel d'étude exceptionnel pour analyser les relations fonctionnelles entre musique et langage. En effet, bien que les paroles et la mélodie des chansons reposent sur des codes différents, et puissent même être composées par des personnes différentes, elles sont souvent (sinon toujours) entendues et produites conjointement. Les chansons constituent ainsi une alliance naturelle entre la musique et la parole, un stimulus hautement « écologique ». L'étude du chant constitue une nouvelle approche importante pour la compréhension de la cognition musicale, puisque le chant est lié au langage et semble, par ailleurs, guidé par des processus largement inconscients[36].

I. P. – C'est l'une des raisons pour lesquelles l'examen des conséquences des lésions cérébrales sur le chant est particulièrement intéressant.

32. Wallin *et al.*, (2003).
33. *Cf.* Ostwald (1973 ; pour une revue, voir Dowling (1999).
34. Bergeson et Trehub (2002) ; Halpern (1989).
35. Levitin (1994) ; Levitin et Cook (1996).
36. Loui *et al.* (2008).

LES DISSOCIATIONS NEUROPSYCHOLOGIQUES

R. K. – Avant d'examiner les conséquences des lésions cérébrales sur le chant, il faut rappeler qu'un module ou une opération spécifique d'un domaine n'est pas nécessairement distinct ou dissociable à l'échelle neuronale. Il est possible que le substrat neural d'un module musical soit entrelacé avec les réseaux dédiés aux modules de parole. Dans ce cas, bien que ces modules soient fonctionnellement distincts, un accident cérébral ne pourrait pas juste affecter le module musical et épargner les modules de parole. En revanche, si ces modules impliquaient des aires cérébrales distinctes du cerveau, il devrait être possible d'observer des dissociations neuropsychologiques. Celles-ci constituent évidemment un argument très fort en faveur de l'existence de modules distincts.

I. P. – Les données dont nous disposons actuellement sont compatibles avec l'existence de modules séparables, au niveau neural, pour la production de parole et de mélodie. Nous avons montré que des patients peuvent perdre leur aptitude à chanter les mélodies de chansons familières tout en continuant à être parfaitement capables de réciter les textes de ces chansons et à parler normalement, avec une prosodie correcte[37]. Ce type de dissociation est révélateur à plus d'un égard. D'abord, cela montre bien que les paroles et la musique d'une chanson sont représentées indépendamment dans la mémoire. Sinon, on comprendrait mal comment ces patients peuvent encore réciter les paroles (ou simplement les reconnaître), alors qu'ils ne retrouvent pas la mélodie. Ensuite, ce maintien général de la parole, qu'elle soit chantée ou parlée, suggère qu'il n'y a rien de spécial dans la parole chantée. Nous y reviendrons.

R. K. – Mais cet argument ne suffit pas, car une objection fréquente est que la plupart des gens ne sont que des dilettantes en musique, dont ils sont amateurs mais pas experts, alors qu'ils sont surentraînés, et donc hautement experts, en langage. En conséquence, la musique, plus fragile car moins bien ancrée dans leur système cognitif, pourrait souffrir davantage que la parole d'une lésion cérébrale.

I. P. – Oui, je te l'accorde, les compétences musicales semblent plus « fragiles » chez le chanteur occasionnel. Mais le niveau d'expertise ne constitue pourtant pas un contre-argument sérieux. Premièrement, les déficits constatés ne sont pas limités aux personnes dépourvues d'éducation musicale. Schön et ses collaborateurs[38] ont rapporté le cas d'un chanteur d'opéra qui n'était plus capable de produire des intervalles musicaux,

37. *Cf.* Peretz *et al.* (1994) ; Peretz *et al.* (1997) ; Murayama *et al.* (2004).
38. Schön *et al.* (2004).

mais qui continuait à parler normalement, avec une expression et une intonation correctes. Deuxièmement, la situation inverse est très fréquente. La plupart des aphasiques non fluents peuvent encore chanter la mélodie correctement[39]. De fait, ces patients restent capables de chanter des mélodies familières et d'apprendre de nouvelles mélodies, alors qu'ils n'arrivent plus à produire la parole de manière intelligible, que ce soit en chantant ou en parlant[40]. Ces résultats indiquent que la production de parole, chantée ou parlée, est sous-tendue par le même système linguistique de sortie (qui dans le cas présent est affecté), et que cette voie articulatoire est distincte de la voie musicale (ici épargnée).

R. K. – Nous observons donc ici une dissociation dans les deux sens : certains patients parlent mais ne chantent plus les mélodies correctement, tandis que d'autres ne parlent plus mais conservent la capacité de fredonner l'air.

I. P. – Une telle dissociation se retrouve d'ailleurs dans les troubles du développement. Les enfants présentant des troubles spécifiques du langage chantent bien, mais ne parviennent pas à parler correctement[41]. Inversement, les personnes souffrant d'amusie congénitale ne sont pas capables de chanter correctement, mais parlent tout à fait normalement[42].

R. K. – En somme, la spécificité de domaine de la perception du langage et de la musique s'étend à leur production.

I. P. – Ces cas neuropsychopathologiques constituent l'argument le plus convaincant en faveur d'une modularité sous-jacente pour la parole et la musique. La dissociation implique l'existence de systèmes anatomiquement et fonctionnellement distincts pour la musique et la parole : un système de production peut fonctionner relativement indépendamment de l'autre, qui peut être sélectivement atteint.

R. K. – Pourtant, certains sceptiques ont suggéré que les dissociations constatées ne seraient pas nécessairement concluantes : de telles dissociations peuvent être simulées dans un réseau de neurones artificiel, alors que celui-ci représente un système unitaire. En d'autres termes, la « lésion » de certains systèmes de connexions peut engendrer des dissociations dans les deux sens en l'absence d'une séparation claire de fonctions, ou de l'existence de modules[43].

---

39. Hébert *et al.* (2003) ; Peretz *et al.* (2004) ; Schlaug *et al.* (2008) ; Warren *et al.* (2003) ; Wilson *et al.* (2006).
40. Hébert *et al.* (2003) ; Peretz *et al.* (2004) ; Racette *et al.* (2006).
41. *Cf.* El Mogharbel *et al.* (2005-2006).
42. *Cf.* Ayotte *et al.* (2002).
43. *Cf.* Plaut (1995).

I. P. – Cela reste théorique. Il n'y a jusqu'à présent aucune explication unitaire qui puisse rendre compte de manière plausible des profils de perte et de maintien des aptitudes musicales que nous venons de discuter. Donc, les données neuropsychologiques révèlent l'existence d'au moins un module distinct pour la musique et la parole. Pour la musique, ce module distinct pourrait être lié à la production de la hauteur des notes. En effet, comme nous l'avons déjà suggéré, il n'est pas nécessaire de postuler que toutes les composantes qui contribuent à la production du chant sont spécifiques de la musique. Une seule composante, si elle est endommagée ou absente, pourrait rendre compte de toutes les manifestations de spécificité pour la musique. Or tous les individus atteints d'amusie congénitale que nous avons étudiés semblent souffrir d'un dysfonctionnement de la discrimination des intervalles mélodiques[44] et, en conséquence, chantent faux. De plus, tous les cas d'amusie acquise qui souffrent d'un trouble de la reconnaissance ou de la production musicale suite à une lésion cérébrale[45] sont systématiquement affectés dans leur perception de la hauteur des notes, mais pas dans celle du rythme.

R. K. – En principe, une atteinte de la dimension temporelle, en particulier de la perception du rythme, devrait aussi affecter les autres performances musicales, tant le rythme semble être l'essence même de la musique.

I. P. – Comme nous l'avons déjà vu, des troubles de la perception du rythme peuvent survenir indépendamment des difficultés de perception mélodique[46], ce qui renforce l'idée de la séparabilité fonctionnelle du traitement du rythme et de la hauteur des sons. Il reste à déterminer dans quelle mesure ces troubles de la perception rythmique affectent exclusivement les aptitudes musicales. Inversement, la préservation des aptitudes rythmiques pourrait expliquer pourquoi chanter en chœur à l'unisson, en se synchronisant avec quelqu'un d'autre, favorise l'intelligibilité de la parole chez les aphasiques non fluents, alors que parler à l'unisson n'est d'aucune aide[47]. Les facteurs rythmiques pourraient aussi varier en fonction de la langue parlée. Par exemple, l'articulation des mots dans le chant semble être plus souvent préservée chez les aphasiques anglophones[48] que chez les francophones. Ceci pourrait être dû au fait que les accents de la musique coïncident avec l'accentuation des paroles en anglais. Il serait nécessaire de faire plus de recherches sur la dimension

44. Peretz (2008).
45. Peretz (2006).
46. Di Pietro *et al.* (2004) ; Alcock *et al.* (2000a, 2000b).
47. Racette *et al.* (2006).
48. *Cf.* Schlaug *et al.* (2008).

temporelle, à la fois dans la production de parole et dans celle du chant, en particulier du chant choral.

R. K. – Mais nous pouvons déjà conclure que les preuves disponibles montrent que la capacité musicale est le produit d'un ensemble de modules fonctionnellement distincts…

I. P. – … et que, jusqu'à présent, seules les aptitudes liées à la perception de la hauteur des notes semblent spécifiques de la musique. Il reste bien sûr à examiner la spécificité musicale de nombreuses autres composantes[49]. Néanmoins, nous pouvons affirmer que les données expérimentales disponibles, qui concernent surtout des processus liés à la perception de la hauteur des notes, s'opposent à l'idée selon laquelle la capacité musicale aurait simplement « envahi » les modules du langage.

R. K. – Il n'en reste pas moins que nous ne pouvons pas nous contenter de données expérimentales neuropsychologiques. Nous devrions expliquer aussi les profils d'activation d'aires cérébrales commune entre musique et langage qui ont été rapportés dans les études d'imagerie fonctionnelle du cerveau sain.

ACTIVATIONS COMMUNES EN NEURO-IMAGERIE FONCTIONNELLE

I. P. – En effet, le traitement de la musique, sans doute davantage encore que celui du langage, recrute un grand nombre de régions cérébrales, qui sont localisées tant dans l'hémisphère gauche que dans l'hémisphère droit, avec une asymétrie en faveur du côté droit plus prononcée pour le traitement de la mélodie[50]. Il n'est donc pas surprenant que la neuro-imagerie fonctionnelle du cerveau normal révèle un recouvrement significatif des aires d'activation entre la musique et le langage. Ceci est aussi le cas des sept études de neuro-imagerie dans lesquelles le chant et la parole (produits ouvertement ou non) ont été comparés[51]. Un tel recouvrement était prévisible, non seulement parce que parole et musique recrutent des réseaux largement distribués de régions cérébrales, mais aussi parce qu'elles impliquent de nombreux systèmes de traitement dont certains pourraient être partagés. Le nombre de réseaux impliqués est particulièrement important dans les tâches de production, puisque les systèmes de sortie impliquent aussi les systèmes perceptifs de contrôle (ou *monitoring*) auditif.

---

49. *Cf.* Peretz et Coltheart (2003).
50. Peretz et Zatorre (2005).
51. Callan *et al.* (2006) ; Jeffries *et al.* (2003) ; Hickok *et al.* (2003) ; Özdemir *et al.* (2006) ; Saito *et al.* (2006) ; Brown *et al* (2006) ; Koelsch *et al.* (2008).

R. K. – En fait, l'identification de profils d'activation distincts pour le chant et pour la parole serait plus instructive que la mise en évidence de profils d'activation communs.

I. P. – Oui, et toutes les études publiées, à l'exception d'une seule[52], rapportent des aires d'activation distinctes pour la parole et le chant.

R. K. – Il faut cependant être prudent dans l'interprétation de ces activations distinctes. L'augmentation d'activité observée dans une région cérébrale donnée pourrait parfois refléter l'augmentation de la difficulté de la tâche plutôt que révéler un corrélat neural distinct. Ainsi, Özdemir et ses collaborateurs[53] ont utilisé une tâche d'imitation vocale de mots bisyllabiques parlés ou chantés (par exemple « *money* » chanté sur une tierce mineure). Ces mots étaient prononcés à un débit anormalement lent (une syllabe par seconde) et donc étaient plus similaires au chant qu'à la parole (dont le taux peut monter jusqu'à dix consonnes et voyelles par seconde[54]). Les aires d'activation communes à toutes les tâches incluaient le gyrus inférieur pré- et postcentral, le gyrus temporal supérieur (GTS) et le sillon temporal supérieur des deux hémisphères. Le chant s'accompagnait d'une activation supplémentaire dans le GTS droit et dans le cortex sensori-moteur primaire. De plus, chanter les mots provoquait davantage d'activation que fredonner (chanter sans les paroles) dans le GTS droit, l'operculum et le gyrus frontal inférieur. Ce résultat fut interprété comme pouvant refléter l'existence d'une voie distincte pour les mots chantés, une voie qui pourrait être utilisée dans le chant par les aphasiques non fluents dont nous avons déjà parlé. L'hypothèse est certes séduisante, mais en réalité ces observations en imagerie fonctionnelle pourraient simplement indiquer que chanter les mots est une tâche plus difficile que seulement les parler, ou que seulement fredonner[55].

I. P. – Mais d'autres études sont beaucoup plus convaincantes. En utilisant des chansons connues, plutôt qu'un intervalle chanté lentement, comme l'avait fait Özdemir[56], Saito et ses collaborateurs[57] ont mis en évidence un réseau neural distinct pour le chant. Ils ont comparé le chant et la récitation du texte de la chanson, que les sujets devaient produire soit seuls soit à l'unisson, en se synchronisant avec une voix enregistrée. Chanter seul ou à l'unisson activait des régions cérébrales qui n'étaient pas activées lors de la récitation des paroles, à savoir le gyrus frontal infé-

52. Koelsch *et al.* (2008).
53. Özdemir *et al.* (2006).
54. *Cf.* Liberman (1991).
55. Racette et Peretz (2007).
56. Özdemir *et al.* (2006).
57. Saito *et al.* (2006).

rieur droit, le cortex prémoteur droit et l'insula antérieure droite. En se fondant sur une logique de soustraction, on peut considérer ces aires cérébrales comme étant liées à la production de la mélodie. Il est cependant dommage que ces chercheurs n'aient pas testé également la production de la mélodie seule (fredonnement) en imagerie. Il est intéressant de remarquer que le chant synchrone (à l'unisson) activait davantage que la parole synchronisée la partie antérieure gauche du lobe pariétal inférieur, le *planum temporale* postérieur droit, le *planum polare* droit et l'insula médiane droite. Ces aires corticales pourraient offrir une base neurale à l'observation clinique selon laquelle l'intelligibilité de la parole des aphasiques non fluents est améliorée lors du chant synchrone, comme nous l'avons déjà mentionné[58].

R. K. – Par ailleurs, il est remarquable qu'il y ait aussi plus d'activation lors du chant que lors de la production de parole dans les régions cérébrales qui sont impliquées dans le circuit de la récompense, comme le noyau accumbens[59]. Ce résultat, qui suggère une composante émotionnelle plus importante dans le chant, est cohérent avec le fait que chanter, plus que parler, est vécu par chacun comme une expérience agréable. Et c'est encore plus le cas des patients aphasiques, pour qui chanter est souvent le seul mode d'expression vocale qui soit préservé[60].

I. P. – Les études de neuro-imagerie fonctionnelle peuvent en effet fournir des hypothèses intéressantes sur les similitudes et différences entre langage et musique, en particulier lorsqu'on les combine avec les études de lésions cérébrales, comme nous venons de l'illustrer avec ces déficits du langage. Toutefois, les données de la neuro-imagerie sont beaucoup moins informatives que celles de la neuropsychologie, car les recouvrements ou associations observés n'offrent pas la même puissance d'inférence. Seules les dissociations neuropsychologiques permettent d'identifier quelle composante de traitement est essentielle à une fonction, alors que les études d'activation révèlent seulement celles qui participent ou sont associées à ce traitement, sans être pour autant nécessairement essentielles. De plus, les aires cérébrales activées sont relativement vastes, et pourraient donc chacune abriter plus d'un réseau de traitement de l'information. Une meilleure résolution spatiale pourrait en fait révéler des sous-régions distinctes.

R. K. – Donc les résultats de la neuro-imagerie, pris isolément, sont difficiles à interpréter. Néanmoins, si toutes les tentatives de montrer la

58. Racette *et al.* (2006).
59. *Cf.* Callan *et al.* (2006).
60. Racette *et al.* (2006).

séparabilité neurale échouaient, nous serions sans doute conduits à remettre en cause la séparabilité du traitement musical par rapport au traitement linguistique. Heureusement, nous pouvons utiliser, dans cette recherche, d'autres outils, plus appropriés. La stimulation magnétique transcrânienne, dont nous allons parler maintenant, est l'un de ces outils parmi les plus prometteurs.

LES EFFETS D'INTERFÉRENCE PAR STIMULATION MAGNÉTIQUE TRANSCRÂNIENNE

I. P. – La stimulation magnétique transcrânienne est maintenant largement utilisée dans les neurosciences cognitives car c'est la meilleure méthode dont nous disposons. Elle permet de produire une interférence temporaire avec un processus cérébral. Contrairement aux études de neuro-imagerie fonctionnelle, cette méthode ne se contente pas de montrer qu'une réponse neurale donnée est associée au comportement d'intérêt, mais permet, en interférant avec le déroulement de celui-ci, de vérifier quels sont les processus neuraux essentiels à l'activité en cours.

R. K. – En ce sens, la stimulation magnétique transcrânienne est un équivalent transitoire (car l'interférence sur le traitement de l'information est temporaire, réversible) et expérimental (puisque la localisation de la « lésion » est choisie et peut être manipulée) des études neuropsychologiques de patients porteurs de lésions cérébrales. Par rapport à l'étude des lésions, cette méthode offre encore un troisième avantage : l'interférence est locale, et survient dans un cerveau par ailleurs normal, sans comorbidité due à l'accident cérébral, ni risque de contamination des données par des stratégies compensatoires que le patient pourrait avoir développées pour pallier son déficit.

I. P. – Lorsque la stimulation magnétique transcrânienne est appliquée de manière à inhiber le fonctionnement du cortex frontal gauche chez des sujets droitiers sains, elle engendre un arrêt de la parole, alors que la même stimulation appliquée à la région homologue droite n'interfère ni avec la parole, ni avec le chant[61]. Arriver à interférer avec le chant semble d'ailleurs très difficile, quel que soit le côté de stimulation[62].

R. K. – L'inhibition provoquée par la stimulation magnétique transcrânienne pourrait même aider certains patients à récupérer mieux ou plus vite de leurs troubles. Ceci peut paraître contre-intuitif, mais s'appliquerait notamment aux aphasiques non fluents. L'imagerie cérébrale fonctionnelle a en effet montré que le cerveau des aphasiques présentait,

61. Epstein *et al.* (1999) ; Stewart *et al.* (2001).
62. Epstein *et al.* (1999) ; Walsh (communication personnelle).

dans l'hémisphère droit, une hyperactivation des régions homologues aux aires périsylviennes gauches responsables de la parole[63]. Il est possible que cette suractivité reflète une tentative inadéquate de leur cerveau de s'adapter à ces nouvelles conditions pathologiques. L'idée est donc d'appliquer la stimulation magnétique transcrânienne de manière à diminuer l'activité de ces régions de l'hémisphère droit. Et cela semble faciliter la performance des aphasiques non fluents dans des tâches linguistiques comme la nomination d'images[64], avec des effets qui perdurent de deux à huit mois après la stimulation. On pourrait aussi tester les performances musicales de ces patients, avec l'idée qu'il pourrait exister une variation des performances musicales concomitante à l'amélioration linguistique ?

I. P. – C'est une excellente idée. Mais nous avons déjà des résultats intéressants issus d'études qui ont utilisé la stimulation magnétique transcrânienne sur un mode « facilitateur ». Lorsque cette technique est ainsi appliquée aux régions du cortex moteur correspondant à la main, les activités de parole et de chant modifient l'amplitude des potentiels évoqués moteurs induits par la stimulation magnétique transcrânienne[65]. Les potentiels de la main droite sont amplifiés par la production de parole, alors que ceux de la main gauche le sont par la production de chant et le fredonnement, en comparaison avec l'articulation de syllabes sans signification. Cette approche offre un argument supplémentaire en faveur de l'existence de mécanismes différemment latéralisés qui sous-tendent la planification et l'exécution vocale de parole et de musique.

R. K. – Les effets d'interférence et de facilitation entre les paroles et la mélodie dans le chant peuvent aussi être étudiés dans le chant « normal », sans aucune stimulation.

I. P. – De tels effets ont déjà été observés dans une tâche d'apprentissage de chansons, tant chez les musiciens que chez les « non musiciens », c'est-à-dire les sujets sans éducation musicale formelle[66]. Produire à la fois les paroles et la mélodie était plus difficile que réciter le texte ou chanter la mélodie sur /*la*/. Produire à la fois les paroles et la mélodie d'une nouvelle chanson semble donc être une tâche double, dans laquelle la mélodie et le texte entrent en compétition pour des ressources attentionnelles ou mnésiques qui seraient, elles, générales et limitées. Nous avons interprété cette interférence entre le traitement de la musique et celui du langage comme une preuve en faveur de l'intervention de mécanismes distincts.

63. Belin *et al.* (1996) ; Naeser *et al.* (2004) ; Rosen *et al.* (2000).
64. Martin *et al.* (2007) ; Naeser *et al.* (2005a, 2005b).
65. Sparing *et al.* (2007) ; Lo *et al.* (2003).
66. Racette et Peretz (2007).

R. K. – Voilà qui doit paraître étrange à des « modularistes radicaux »… En effet, si les composantes de traitement de la musique et du langage étaient complètement modulaires au sens où l'entend Fodor, le traitement d'un domaine, par exemple la musique, devrait être « encapsulé », et donc imperméable au traitement de la parole, qui aurait lieu en parallèle. Or, comme nous venons de l'illustrer, les traitements de la parole et de la mélodie interagissent l'une avec l'autre dans la production du chant.

I. P. – Mais ceci n'implique pas que la mélodie et les paroles soient traitées par un centre commun d'opérations. L'observation d'une interférence (ou d'une facilitation) ne remet pas en cause la modularité des traitements, mais seulement leur « encapsulation ». L'utilisation d'informations provenant de sources diverses est ce que nous attendons d'un système de traitement efficace, et s'applique en particulier au chant, qui par définition est un stimulus multidimensionnel. Mais l'intégration entre les informations ne s'oppose pas à l'idée d'une utilisation spécialisée de l'information par des systèmes de traitement dédiés à la musique ou à la parole.

R. K. – Voyons maintenant si cette hypothèse de spécialisation de traitement pourrait être réfutée sur la base d'un autre type d'argument : l'occurrence d'effets de transfert entre domaines, c'est-à-dire entre les aptitudes musicales et de parole.

EFFETS DE TRANSFERT ENTRE DOMAINES

I. P. – La recherche récente a examiné les effets de transfert entre les performances musicales et linguistiques en supposant que ce transfert est sous-tendu par des mécanismes partagés[67]. Cependant, nous comprenons encore mal ces effets. C'est le cas, notamment, des effets de l'expertise ou de l'entraînement musical sur le langage[68]. Néanmoins, de nombreux chercheurs spéculent sur la nature des associations observées. Ainsi, Patel[69] propose que l'expertise musicale améliore l'encodage sensoriel, ce qui à son tour serait bénéfique pour la perception de la parole.

R. K. – En principe, cette proposition pourrait être applicable à la production de la parole. Par exemple, on pourrait prédire que les musiciens sont plus aptes à apprendre à parler une seconde langue que les non musiciens. Une telle association a d'ailleurs été récemment rapportée par Slevc et Miyake[70] chez des Japonais arrivés tardivement aux

67. *Cf.* Patel (sous presses).
68. Schellenberg (2006).
69. Patel (sous presses).
70. Slevc et Miyake (2006).

États-Unis. Ceux qui présentaient un haut degré d'aptitude musicale percevaient et prononçaient mieux l'anglais que leurs pairs musicalement moins talentueux.

I. P. – Inversement, on s'attendrait à ce que des locuteurs de langues tonales soient plus « musiciens » que des locuteurs de langues non tonales. Par exemple, dans ces cultures, l'amusie devrait être quasi inexistante. Nous testons actuellement cette prédiction à Hong Kong (en collaboration avec Patrick Wong). À ma connaissance, la possibilité d'un transfert entre domaines, du langage vers la musique, n'a jamais été examinée.

R. K. – Or nous pourrions non seulement nous intéresser davantage aux locuteurs de langues tonales, mais aussi étudier les effets de l'expertise linguistique sur le traitement de la musique, que ce soit une expertise relativement informelle, comme celle qui consiste à maîtriser plusieurs langues, ou une expertise beaucoup plus fine et formelle, comme celle des experts en phonétique. La seule fois que des experts phonéticiens ont été examinés dans le cadre des effets de transfert, ce fut uniquement pour montrer qu'ils bénéficient d'une expertise musicale dans l'analyse de fines variations prosodiques[71].

I. P. – De manière plus générale, les études actuelles sur les effets de transfert entre domaines se heurtent à différents obstacles[72]. Premièrement, l'aptitude musicale, le fait de prendre des leçons de musique, et le fait d'être musicien sont des concepts liés mais non identiques. L'aptitude réfère à l'inné (hors apprentissage), le fait de suivre des leçons de musique implique l'apprentissage. Être musicien est vraisemblablement la conséquence à la fois d'une aptitude personnelle et de l'entraînement, combinés à d'autres facteurs. Le nombre d'années de leçons de musique prédit des capacités cognitives – y compris linguistiques – chez les enfants et les adultes[73], alors que la comparaison entre musiciens et non musiciens a mené à des résultats nuls ou incohérents[74]. De la même manière, on ne peut pas rendre compte des effets de l'entraînement musical lorsqu'on étudie l'aptitude[75], car cet entraînement améliore la performance aux tests d'aptitude musicale.

R. K. – En d'autres termes, les associations qui ont été observées pourraient tout aussi bien avoir une origine génétique qu'être la conséquence de l'apprentissage musical…

---

71. Dancovicová *et al.* (2007).
72. Schellenberg et Peretz (2008).
73. Schellenberg (2006).
74. Helmbold *et al.* (2005).
75. *Cf.* Slevc et Miyake (2006).

I. P. – Un deuxième problème, lié au premier, concerne la nature et la spécificité des associations entre l'expérience musicale et la cognition. La discussion de « liens spéciaux » entre musique et langage[76] est trompeuse, car les associations entre la pratique musicale et les aptitudes cognitives sont bien plus générales, s'étendant à la mémoire de travail ainsi qu'aux aptitudes mathématiques et à la reconnaissance spatiale.

R. K. – Suivre des leçons de musique pourrait constituer une expérience d'apprentissage qui améliore ce qu'on appelle les « fonctions exécutives », c'est-à-dire les processus de planification, de mémoire de travail, de contrôle et d'attention, qui favorisent un comportement flexible et adapté au contexte. Si c'était le cas, on s'attendrait évidemment à observer une amélioration des résultats à une multitude de tests cognitifs. Nos résultats récents soutiennent cette hypothèse : les musiciens sont plus sensibles que les non musiciens à de fines variations prosodiques[77], mais restent parfaitement capables d'ignorer ces variations si la tâche l'exige[78]. Ils ont donc une meilleure capacité d'attention sélective que les non musiciens, du moins dans le domaine auditif.

I. P. – De plus, les inférences sur les relations de causalité qui sont proposées dans les études de transfert entre domaines ne sont pas fondées. Bien que certaines expériences suggèrent que le fait de prendre des leçons de musique mène à un transfert cognitif[79], identifier la nature de l'association entre la musique et le langage requiert des études supplémentaires, dans lesquelles on utiliserait des conditions témoins appropriées, ainsi qu'une assignation aléatoire des sujets aux différents groupes étudiés.

R. K. – En somme, les associations qui ont été observées entre musique et langage, comme celle rapportée par Slevc et Miyake[80], pourraient n'être que le produit des fonctions exécutives, d'influences attentionnelles ou cortico-fugales[81] générales. Et il semble très difficile de préciser la nature exacte de la relation entre ces domaines, même en utilisant des témoins et des protocoles expérimentaux appropriés. Il n'en reste pas moins que ces études ont des implications cliniques et éducatives importantes, pour l'apprentissage des langues comme pour la rééducation du langage chez les patients porteurs de lésions cérébrales.

---

76. *Cf.* Patel (sous presses) ; Slevc et Miyake (2006).
77. Kolinsky *et al.* (sous presses) ; voir aussi Marques *et al.* (2007).
78. Kolinsky *et al.* (sous presses).
79. Schellenberg (2004).
80. Slevc et Miyake (2006).
81. Wong *et al.* (2007).

Discussion en contrepoint

I. P. – Il est clair aussi que si nous voulons progresser dans la connaissance des relations entre perception de la musique et de la parole, nous devons continuer les recherches, plutôt que rester sur des positions théoriques rigides.

R. K. – À ce propos, il faut citer une approche théorique alternative, qui considère que les ressources neurales nécessaires aux traitements des informations pourraient être partagées. En effet, Patel[82] propose que la spécificité de domaine ne s'applique qu'aux représentations, aux bases de données qui constituent notre connaissance. Les opérations qui ont lieu sur ces représentations seraient, elles, partagées ou générales, œuvrant à travers les domaines. Patel se réfère à ces opérations comme étant des « ressources neurales partagées ». En d'autres termes, son cadre théorique fait la distinction entre spécificité représentationnelle et spécificité de traitement. Dans la théorie modulaire, la spécificité de domaine s'applique à la fois à l'opération et à sa représentation.

I. P. – Il devrait être possible, en principe, de tester empiriquement ces deux points de vue. De la même manière qu'un programme Excel peut être utilisé avec des nombres ou des noms mais est indépendant de ces codes, nous devrions pouvoir dissocier une composante de traitement de sa base de connaissance et tester sa spécificité. Par exemple, l'acquisition de la connaissance tonale utilise des principes généraux, notamment en extrayant les régularités statistiques de l'environnement[83]. Bien que l'encodage tonal de la hauteur des notes soit spécifique à la musique, il pourrait être bâti sur « la sensibilité des auditeurs à la distribution des hauteurs de notes, [qui est] une manifestation de stratégies perceptives générales d'exploitation des régularités du monde physique[84] ». Donc, l'entrée et la sortie du traitement statistique peuvent être spécifiques d'un domaine, alors que le mécanisme d'apprentissage ne l'est pas[85]. Une fois acquis, le fonctionnement du système – dans le cas présent : celui de l'encodage tonal de la hauteur – peut être modulaire, en encodant exclu-

82. Patel (2003, 2008, sous presses).
83. Krumhansl (1990) ; Tillmann *et al.* (2000).
84. Oram et Cuddy (1995), p. 114.
85. Peretz (2006) ; Saffran et Thiessen (2006).

sivement et automatiquement la hauteur des notes en termes de gammes musicales.

R. K. – Un raisonnement similaire peut être appliqué à l'analyse de la scène auditive et au groupement auditif. Le fait que ces deux composantes de traitement organisent les stimuli selon les principes généraux de la *Gestalt* (comme la proximité en hauteur, dans le cas des sons, ou spatiale, dans le cas des stimuli visuels) ne signifie pas que leur fonctionnement soit général et sous-tendu par un seul système de traitement. Il serait d'ailleurs très surprenant que l'analyse de la scène auditive soit prise en charge par le même système que celui qui s'occupe de l'analyse de la scène visuelle. Il est bien plus vraisemblable que les codes d'entrée (visuel et auditif) ajustent ces mécanismes à leurs propres nécessités de traitement. Donc, le code d'entrée peut transformer des mécanismes généraux en mécanismes hautement spécialisés. Par ailleurs, l'existence de microsystèmes multiples et hautement spécialisés, même s'ils fonctionnent d'une manière très similaire, est plus vraisemblable, car la « modularisation » est plus efficace[86].

I. P. – Il est ainsi possible que la spécificité de domaine émerge de l'opération d'un mécanisme général, ou de ressources neurales partagées, comme le propose Patel. En pratique, il sera toutefois fort difficile de démontrer cette hypothèse, puisque les mécanismes généraux ou « partagés » sont susceptibles de se « modulariser » avec l'expérience[87]. Les études du développement normal et pathologique pourraient nous aider à départager l'état final, modularisé, de l'état initial. Les troubles du développement pourraient fournir des pistes sérieuses, notamment en nous montrant dans quelle mesure il y a (ou non) cooccurrence de troubles de l'acquisition de la musique et du langage (ou d'autres sphères de la cognition, comme la cognition spatiale).

R. K. – L'étude des animaux serait précieuse elle aussi. En effet, le point de vue que nous avons évoqué pourrait renforcer l'idée que la spécificité de domaine ne dépend que de très peu de composantes de traitement, qui se détacheraient du fond cognitif commun, partagé entre les divers domaines de la connaissance. Ces composantes clés pourraient correspondre à des adaptations spécifiquement humaines, alors que le système cognitif commun pourrait être aussi partagé avec les animaux.

I. P. – Il reste beaucoup de questions à résoudre quant aux relations entre musique et langage. Même si nous disposons déjà de données montrant que les aptitudes musicales reposent, en partie, sur des mécanismes

86. Marr (1982).
87. Saffran et Thiessen (2006).

cérébraux spécialisés, la parole et le chant impliquent de nombreuses composantes de traitement de l'information. Chacune peut être considérée comme modulaire parce qu'elle réalise un traitement hautement spécifique ; mais certaines pourraient aussi être spécifiques à la musique, c'est-à-dire uniquement impliquées dans le traitement de la musique. Les recherches futures devraient aussi s'intéresser au détail des fonctions que ces mécanismes sous-tendent, et pas seulement à leur spécificité. Comme nous l'avons dit, les études sur le développement pourraient être critiques dans ce débat. Les nouveaux outils neuroscientifiques, comme la stimulation magnétique transcrânienne et l'imagerie optique, pourraient eux aussi nous aider à estimer si des mécanismes cérébraux distincts sous-tendent l'acquisition des différents domaines de connaissance, non seulement chez les adultes, mais aussi chez les jeunes enfants[88].

R. K. – Cette discussion illustre en tout cas à quel point la notion de modularité reste importante dans la recherche contemporaine.

I. P. – L'hypothèse de modularité est instructive au niveau empirique, en nous faisant rechercher des spécialisations. De plus, elle fournit des candidats plausibles de mécanismes évolués de traitement de l'information, et donc aussi de mécanismes génétiquement déterminés. Pour conclure, on peut dire que le concept moderne de modularité « offre un cadre conceptuel utile dans lequel des débats constructifs portant sur les systèmes cognitifs peuvent continuer à être structurés[89] »[90].

RÉFÉRENCES BIBLIOGRAPHIQUES

– Alcock K. J., Passingham R. E., Watkins A. J. et Vargha-Khadem F. (2000a), « Pitch and timing abilities in inherited speech and language impairment », *Brain and Language*, 75, p. 34-46.

– Alcock K. J., Wade D., Anslow P. et Passingham R. E. (2000b), « Pitch and timing abilities in adult left-hemisphere-dysphasic and right-hemisphere subjects », *Brain and Language*, 75, p. 47-65.

– Ayotte J., Peretz I. et Hyde K. (2002), « Congenital amusia : A group study of adults afflicted with a music-specific disorder », *Brain*, 125, p. 238-251.

– Barrett H. C. et Kurzban R. (2006), « Modularity in cognition : Framing the debate », *Psychological Review*, 113, p. 628-647.

88. *Cf.* Pena *et al.* (2003).

89. Barrett et Kurzaban (2006), p. 644.

90. *Remerciements*. La préparation de ce chapitre a bénéficié de l'appui du conseil de recherche en sciences naturelles et en génie du Canada, des instituts de recherche en santé du Canada et d'une chaire de recherche du Canada, ainsi que d'un fonds belge du FRFC (2.4633.06, représentations mentales de la musique et du langage dans le chant et nature de leurs interactions). Le deuxième auteur est maître de recherche du fonds de la recherche scientifique – FNRS.

– Belin P., Van Eeckhout P., Zilbovicious M., Remy P., Francois C. et Guillaume S. *et al.* (1996), « Recovery from nonfluent aphasia after melodic intonation therapy : A PET study », *Neurology*, 47, p. 1504-1511.
– Bergeson T. R. et Trehub S. E. (2002), « Absolute pitch and tempo in mother's sons to infants », *Psychological Science*, 13, p. 72-75.
– Bregman A. (1990), *Auditory Scene Analysis. The perceptual organization of sound*, Londres, MIT press.
– Brown S., Martinez M. J. et Parsons L. M. (2006), « Music and language side by side in the brain : A PET study of the generation of melodies and sentences », *European Journal of Neuroscience*, 23, p. 2791-2803.
– Callan D. E., Tsytsarev V., Hanakawa T., Callan A. M., Katsuhara M., Fukuyama H. *et al.* (2006), « Song and speech : Brain regions involved with perception and covert production », *Neuroimage*, 31, p. 1327-1342.
– Coltheart M. (1999), « Modularity and cognition », *Trends in Cognitive Science*, 3, p. 115-120.
– Dalla Bella S., Giguère J.-F. et Peretz I. (2007), « Singing proficiency in the general population », *Journal of Acoustical Society of America*, 121, p. 1182-1189.
– Dancovicová, J., House, J., Crooks, A. et Jones, K. (2007), « The relationship between musical skills, music training, and intonation analysis skills », *Language and Speech*, 50, p. 177-225.
– Dehaene S. (2007), *Les Neurones de la lecture,* Paris, Odile Jacob.
– Di Pietro M., Laganaro M., Leeman B. et Schnider A. (2004), « Receptive amusia : Temporal auditory processing deficit in a professional musician following a left temporo-parietal lesion », *Neuropsychologia*, 42, p. 868-877.
– Dowling W. J. (1999), « The development of music perception and cognition », *in* Deutsch D. (éd.), *The Psychology of Music*, San Diego, Academic Press, 2nde éd., p. 603-625.
– El Mogharbel C., Sommer G., Deutsch W., Wenglorz M. et Laufs I. (2005-2006), « The vocal development of a girl who sings but does not speak », *Musicae Scientiae*, p. 235-258.
– Epstein C. M., Meador K. J., Loring D. W., Wright R. J., Weisman J. D., Sheppard S. *et al.* (1999), « Localization and characterization of speech arrest during transcranial magnetic stimulation », *Clinical Neurophysiology*, 110, p. 1073-1079.
– Fisher S. E., Vargha-Khadem F., Watkins K. E., Monaco A. P. et Pembrey M. E. (1998), « Localisation of a gene implicated in a severe speech and language disorder », *Nature Genetics*, 18, p. 168-170.
– Fodor J. (1983), *The Modularity of Mind*, Cambridge (Mass.), MIT Press ; trad. fr. *La Modularité de l'esprit. Essai sur la psychologie des facultés,* Paris, Éditions de Minuit, 1986.
– Gauthier I. et Curby K. M. (2005), « A perceptual traffic jam on highway N170 », *Psychological Science*, 14, p. 30-32.
– Gould S. J. et Vrba E. S. (1982), « Exaptation : A missing term in the science of form », *Paleobiology*, 8, p. 4-15.

– Halpern A. R. (1989), « Memory for the absolute pitch of familiar songs », *Memory and Cognition*, 17, p. 572-581.
– Hébert S., Racette A., Gagnon L. et Peretz I. (2003), « Revisiting the dissociation between singing and speaking in expressive aphasia », *Brain*, 126, p. 1838-1850.
– Helmbold N., Rammsayer T. et Altenmüller E. (2005), « Differences in primary mental abilities between musicians and nonmusicians », *Journal of Individual Differences*, 26, p. 74-85.
– Hickok G., Buchsbaum B., Humphries C. et Muftuler T. (2003), « Auditory-motor interaction revealed by fMRI : Speech, music, and working memory in area SPT », *Journal of Cognitive Neuroscience*, 15 (5), p. 673-682.
– Hurst J. A., Baraister M., Auger E., Graham F. et Norell S. (1990), « An extended family with dominantly inherited speech disorder », *Developmental Medecine and Child Neurology*, 32, p. 352-355.
– Hyde K. L. et Peretz I. (2004), « Brains that are out-of-tune but in-time », *Psychological Science*, 15, p. 356-360.
– Jeffries K. J., Fritz J. B. et Braun A. R. (2003), « Words in melody : An H2 15O PET study of brain activation during singing and speaking », *NeuroReport*, 14, p. 749-754.
– Justus T. et Hutsler J. J. (2005), « Fundamental issues in the evolutionary psychology of music : Assessing innateness and domain-specificity », *Music Perception*, 23, p. 1-27.
– Koelsch S., Schulze K., Sammler D., Fritz T., Müller K. et Gruber O. (2008), « Functional architecture of verbal and tonal working memory : An fMRI study », *Human Brain Mapping*.
– Kolinsky R., Cuvelier H., Goetry V., Peretz I. et Morais J. (sous presses), « Music training facilitates lexical stress processing », *Music Perception*.
– Krumhansl C. L. (1990), *Cognitive Foundations of Musical Pitch*, New York, Oxford University Press.
– Lai C. S., Fisher S. E., Hurst J. A., Vargha-Khadem F. et Monaco A. P. (2001), « A forkhead-domain gene is mutated in a severe speech and language disorder », *Nature*, 413, p. 519-523.
– Levitin D. J. (1994), « Absolute memory for musical pitch : evidence from the production of learned melodies », *Perception and Psychophysics*, 56, p. 414-423.
– Levitin D. J. et Cook P. R. (1996), « Memory for musical tempo : additional evidence that auditory memory is absolute », *Perception et Psychophysics*, 58, p. 927-935.
– Liberman A. M. (1991), « Reading is hard just because listening is easy », *in* Euler C. von (éd.), *Wenner-Gren International Symposium Series : Brain and Reading*, Hampshire, Macmillan.
– Liberman A. M. et Whalen D. H. (2000), « On the relation of speech to language », *Trends in Cognitive Sciences*, 4, p. 187-196.
– Lo Y., Fook-Chong S., Lau D. P. et Tan E. K. (2003), « Cortical excitability changes associated with musical tasks : A transcranial magnetic stimulation study in humans », *Neuroscience Letters*, 252, p. 85-88.

– Loui P., Guenther F., Mathys C. et Schlaug G. (2008), « Action-perception mismatch in tone-deafness », *Current Biology*, 18, R331-R332.
– Marcus G. F. et Fisher S. E. (2003), « FOXP2 in focus : What can genes tell us about speech and language ? », *Trends in Cognitive Sciences*, 7, p. 257-262.
– Marques C., Moreno S., Castro S.-L. et Besson M. (2007), « Musicians detect pitch violation in a foreign language better than nonmusicians : Behavioral and electrophysiological evidence », *Journal of Cognitive Neuroscience*, 19, p. 1453-1463
– Marr D. (1982), *Vision*, W. H. Freeman.
– Martin P. I., Naeser M. A., Ho M., Doron K. W., Kurland J., Kaplan J. *et al.* (2007), « Overt naming fMRI pre- and post-TMS : Two nonfluent aphasia patients, with and without improved naming post-TMS », *Brain and Language*, 103, p. 248-249.
– McDermott J. et Hauser M. (2005), « The origins of music : Innateness, uniqueness, and evolution », *Music Perception*, 23, p. 29-59.
– Murayama J., Kashiwagi T., Kashiwagi A. et Mimura M. (2004), « Impaired pitch production and preserved rhythm production in a right brain-damaged patient with amusia », *Brain and Cognition*, 56, p. 36-42.
– Morais J. et Kolinsky R. (2002), « L'esprit "lettré" et l'esprit humain universel », *in* Dupoux E. (éd.), *Les Langages du cerveau : autour de Jacques Mehler*, Paris, Odile Jacob.
– Naeser M. A., Martin P. I., Nicholas M., Baker E. H., Seekins H., Helm-Estabrooks N. *et al.* (2005), « Improved naming after TMS treatments in a chronic, global aphasia patient – case report », *Neurocase*, 11, p. 182-193.
– Naeser M. A., Martin P. I., Nicholas M., Baker E. H., Seekins H., Kobayashi M. *et al.* (2005), « Improved picture naming in chronic aphasia after TMS to part of right Broca's area : an open-protocol study », *Brain and Language*, 93, p. 95-105.
– Naeser M. A., Martin P. I., Baker E. H., Hodge S. M., Sczerzenie S. E., Nicholas M. *et al.* (2004), « Overt propositional speech in chronic nonfluent aphasia studied with the dynamic susceptibility contrast fMRI method », *NeuroImage*, 22, p. 29-41.
– Oram N. et Cuddy L. (1995), « Responsiveness of Western adults to pitch-distributional information in melodic sequences », *Psychological Research*, 57, p. 103-118.
– Ostwald P. F. (1973), « Musical behavior in early childhood. Developmental medicine and child », *Neurology*, 15, p. 367-375.
– Özdemir E., Norton A. et Schlaug G. (2006), « Shared and distinct neural correlates of singing and speaking », *NeuroImage*, 33, p. 628-635.
– Patel A. (2003), « Language, music, syntax and the brain », *Nature Neuroscience*, 6, p. 674-681.
– Patel A. (2008), *Music, Language, and the Brain*, Oxford, Oxford University Press.
– Patel A. D. (sous presses), « Language, music, and the brain : A resource-sharing framework », *in* Rebuschat P., Rohrmeier M., Hawkins J. et Cross I. (éd.) *Language and Music as Cognitive Systems*, Oxford, Oxford University Press.

– Pena M., Maki A., Kovacic D., Dehaene-Lambertz G., Koizumi H., Bouquet F. *et al.* (2003), « Sounds and silence : An optical topography study of language recognition at birth », *Proceedings of the National Academy of Sciences USA*, 100, p. 11702-11705.
– Peretz I. (2001), « Music perception and recognition », *in* Rapp B. (éd.), *The Handbook of Cognitive Neuropsychology*, Hove, Psychology Press, p. 519-540.
– Peretz I. (2006), « The nature of music from a biological perspective », *Cognition*, 100, p. 1-32.
– Peretz I. (2008), « Musical disorders : From behavior to genes. Current directions », *Psychological Science*, 17, p. 329-333.
– Peretz I., Belleville S. et Fontaine F.-S. (1997), « Dissociations entre musique et langage après atteinte cérébrale : un nouveau cas d'amusie sans aphasie », *Revue canadienne de psychologie expérimentale*, 51, p. 354-367.
– Peretz I. et Coltheart M. (2003), « Modularity of music processing », *Nature Neuroscience*, 6, p. 688-691.
– Peretz I., Cummings S. et Dubé M.-P. (2007), « The genetics of congenital amusia (or tone-deafness) : A family aggregation study », *American Journal of Human Genetics*, 81, p. 582-588.
– Peretz I, Gagnon L, Hébert S. et Macoir J. (2004), « Singing in the brain : Insights from cognitive neuropsychology », *Music Perception*, 21, p. 373-390.
– Peretz I. et Kolinsky R. (1993), « Boundaries of separability between melody and rhythm in music discrimination : A neuropsychological perspective », *Quarterly Journal of Experimental Psychology*, 46 A, p. 301-325.
– Peretz I., Kolinsky R., Tramo M., Labrecque R., Hublet C., Demeurisse G. *et al.* (1994), « Functional dissociations following bilateral lesions of auditory cortex », *Brain*, 117, p. 1283-1301.
– Peretz I. et Morais J. (1989), « Music and modularity », *Contemporary Music Review*, 4, p. 277-291.
– Peretz I. et Zatorre R. J. (2005), « Brain organization for music processing », *Annual Review of Psychology*, 56, p. 89-114.
– Pinker S. (1997), *How the Mind Works*, New York, Norton ; trad. fr. *Comment fonctionne l'esprit*, Paris, Odile Jacob, 2000.
– Plaut D. C. (1995), « Double dissociation without modularity : Evidence from connectionist neuropsychology », *Journal of Clinical and Experimental Neuropsychology*, 17, p. 291-321.
– Racette A., Bard C. et Peretz I. (2006), « Making non-fluent aphasics speak : Sing along ! », *Brain*, 129, p. 2571-2584.
– Racette A. et Peretz I. (2007), « Learning lyrics : To sing or not to sing ? », *Memory et Cognition*, 35, p. 242-253.
– Rosen H. J., Petersen S. E., Linenweber M. R., Snyder A. Z., White D. A., Chapman L. *et al.* (2000), « Neural correlates of recovery from aphasia after damage to left inferior frontal cortex », *Neurology*, 55, p. 1883-1894.
– Saffran J. R. et Thiessen E. D. (2006), « Domain-general learning capacities », *in* Hoff E. et Shatz M. (éd.), *Handbook of Language Development*, Cambridge, Blackwell, p. 68-86.

– Saito Y., Ishii K., Yagi K., Tatsumi I. et Mizusawa H. (2006), « Cerebral networks for spontaneous and synchronized singing and speaking », *NeuroReport*, 17, p. 1893-1897.
– Schellenberg E. G. (2004), « Music lessons enhance IQ », *Psychological Science*, 15, p. 511-514.
– Schellenberg E. G. (2006), « Exposure to music : The truth about the consequences », *in* McPherson G. E. (éd.), *The Child as Musician : A Handbook of Musical Development*, Oxford (G.-B.), Oxford University Press, p. 111-134.
– Schlaug G., Marchina S. et Norton A. (2008), « From singing to speaking : Why singing may lead to recovery of expressive language function in patients with Broca's aphasia », *Music Perception*, 25, p. 315-323.
– Schön D., Lorber B., Spacal M. et Semenza C. (2004), « A selective deficit in the production of exact musical intervals following right-hemisphere damage », *Cognitive Neuropsychology*, 21, p. 773-784.
– Slevc L. R. et Miyake A. (2006), « Individual differences in second language proficiency : Does musical ability matter ? », *Psychological Science*, 17, p. 675-681.
– Sparing R., Meister I. G., Wienemann M., Buelte D., Staedtgen M. et Boroojerdi B. (2007), « Task-dependent modulation of functional connectivity between hand motor cortices and neuronal networks underlying language and music : A transcranial magnetic stimulation study in humans », *European Journal of Neuroscience*, 25, p. 319-323.
– Sperber D. et Hirschfeld L. A. (2004), « The cognitive foundations of cultural stability and diversity », *Trends in Cognitive Sciences*, 8, p. 40-47.
– Stewart L., Walsh V., Frith U. et Rothwell J. (2001), « Transcranial magnetic stimulation produces speech arrest but not song arrest », *Annals of the New York Academy of Sciences*, 930, p. 433-435
– Tillmann B., Bharucha J. et Bigand E. (2000), « Implicit learning of tonality : A self-organizing approach », *Psychological Review*, 107, p. 885-913.
– Vargha-Khadem F. Watkins K., Alcock K., Fletcher P. et Passingham R. (1995), « Praxic and nonverbal cognitive deficits in a large family with a genetically transmitted speech and language disorder », *Proceedings of the National Academy of Sciences USA*, 92, p. 930-933.
– Wallin N., Merker B. et Brown S. (éd.) (2000), *The Origins of Music,* Cambridge (Mass.), MIT Press.
– Warren J. D., Warren J. E., Fox N. C. et Warrington E. K. (2003), « Nothing to say, something to sing : Primary progressive dynamic aphasia », *Neurocase*, 9, p. 140-155.
– Wilson S., Parsons K. et Reutens D. (2006), « Preserved singing in aphasia : A case study of the efficacy of melodic intonation therapy », *Music Perception*, 24, p. 23-36.
– Wong P. C. M., Skoe E., Russo N. M., Dees T. et Kraus N. (2007), « Musical experience shapes human brainstem encoding of linguistic pitch patterns », *Nature Neuroscience*, 10, p. 420-422.

III

L'INVENTION DE NOUVEAUX MODES DE COMMUNICATION

Capter la parole vive

par Roger Chartier

« *Eduardum occidere nolite timere bonum est.* » Tels sont les mots écrits sur le billet remis par Mortimer à Lightborne, lorsqu'il l'envoie auprès du roi Édouard II, emprisonné au château de Berkeley. Six mots, mais qu'ordonnent-ils ? Si celui qui les lira marque une pause après les quatre premiers, il devra assassiner le roi : « Ne craignez pas de tuer Édouard, cela est bien ». Mais s'il découpe la phrase en deux parties égales, plaçant la pause après *nolite* et non pas *timere*, l'ordre est tout autre et le monarque aura la vie sauve : « Ne tuez pas Édouard, il est bien de craindre. » De la manière de ponctuer la « *scriptio continua* » de la sentence latine dépend donc la vie ou la mort d'un souverain, qui sera imputée, non pas à celui qui a écrit le billet et, de fait, ordonné le meurtre, mais à ceux qui l'auront reçu[1].

La ponctuation des textes n'a pas toujours, fort heureusement, un aussi dramatique enjeu. Mais, toujours, elle construit la signification en guidant l'œil – ou la voix. C'est sur celle-ci que je voudrais porter mon attention dans ce colloque voué à la relation entre parole et musique et, dans ce cas, à la « musicalité » des textes sans partition ou sans mélodie.

1. Christopher Marlowe, *The Troublesome Raigne and Lamentable Death of Edward the Second* (1598), *in The Complete Works of Christopher Marlowe*, édité par Fredson Bowers, vol. IIn, Cambridge, Cambridge University Press, 1973, p. 86. Tr. fr. Marlowe, *Edouard II*, traduction de C. Pons, Paris, Aubier/Éditions Montaigne, 1964, p. 321 : « Cette lettre écrite par un de nos amis / Contient sa mort, avec l'ordre de lui sauver la vie : / *Eduardum occidere nolite timere, bonum est*, / Ne craignez pas de tuer le Roi, il est bon qu'il meure. / Mais lisez-la ainsi, et c'est un autre sens : / *Eduardum occidere nolite,* / *timere bonum est*, / Ne tuez pas le Roi : il est bon de craindre le pire. / Sans ponctuation elle partira telle qu'elle est, / De sorte qu'à sa mort, si par hasard on la découvre, / Matrevis et les autres en portent la responsabilité, / Et que nous soyons déchargés, qui avons ordonné le meurtre. »

Ma communication, attentive à la ponctuation de l'écrit, sera donc un peu décalée par rapport à la basse continue du colloque, mais peut-être trouvera-t-elle sa justification dans la remarque faite par Yves Bonnefoy dans un texte intitulé « Les deux points, c'est un peu, en prose la poésie ». Il y distingue, en effet, deux systèmes de ponctuation :

> La ponctuation qui dégage les articulations d'un texte, c'est celle que réclame la syntaxe, je suppose ; et qui tend ainsi à coïncider avec les structures de la pensée ? Tandis que celle qui aiderait la lecture serait là plutôt pour comprendre les besoins de la voix, ou mettre en évidence des rythmes, des sons : en somme, non pour penser mais pour séduire[2] ?

S'il conclut décidément en faveur de la première (« Je m'en tiens pour ma part – au moins c'est ce que j'espère – à la ponctuation qui suit les contours de la réflexion »), c'est à la seconde que m'attacherai.

Aux XVI^e^ et XVII^e^ siècles, toutes les réformations de l'orthographe proposées en France ou en Angleterre visent à approcher la perfection ou, plutôt, la moindre imperfection de la langue castillane dans laquelle, comme l'écrit Antonio de Nebrija dans sa *Grammaire de la langue castillane* publiée *l'annus mirabilis* de 1492, celle de l'achèvement de la Reconquista et de la découverte d'un monde nouveau, « *tenemos de escrivir como pronunciamos : i proninciar como escrivimos* » – « il faut écrire comme l'on prononce, et prononcer comme l'on écrit[3] ». Dans toutes les langues européennes, obtenir une telle coïncidence entre la diction et la graphie n'est pas chose aisée. Une première possibilité, contraire aux usages, serait de prononcer toutes les lettres des mots, comme on le fait en latin. C'est cette manière bizarre et pédante de prononcer l'anglais qui, dans *Love's Labour's Lost* (*Peines d'amour perdues*) de Shakespeare, fait le ridicule du maître d'école Holopherne, qui stigmatise ainsi les manières de dire de Don Adriano de Armado : « J'ai horreur de ces bourreaux de l'orthographe qui vous prononcent "*dout*" *sine* "b", alors qu'il devrait dire "*doubt*" ; "*det*", alors qu'il devrait prononcer "*debt*" – "d, e, b, t", pas "d, e, t". […] C'est abhominable – qu'il prononcerait "abominable"[4]. »

Une autre solution, moins extravagante, consiste à transformer l'écriture même des mots pour les ajuster à la façon dont ils sont prononcés. Les ouvrages qui proposent une telle réforme indiquent clairement

---

2. Yves Bonnefoy, *La Petite Phrase et la Longue Phrase*, La Tilv Éditeur, 1994, p. 15-22.
3. Elio Antonio de Nebrija, *Gramática Castellana*, Introducción y notas : Miguel Angel Esparza, Ramón Sarmiento, Madrid, Fundación Antonio de Nebrija, 1992, p. 158-159.
4. William Shakespeare, *Peines d'amour perdues (Love's Labour's Lost)*, *in* William Shakespeare, *Œuvres complètes*, *Comédies,* t. I, édition bilingue, sous la direction de Michel Grivelet et Gilles Monsarrat, Paris, Robert Laffont, 2000, p. 568-569.

que le but visé, bien plus que la réduction de la diversité des graphies, est l'identité entre le dire et l'écriture. Le traité publié par John Hart en 1596 est ainsi intitulé *An orthographie, conteyning the due order howe to write thimage of mannes voice* (*Orthographe, contenant l'ordre adéquat pour écrire l'image de la voix*), et celui de William Bullokar en 1580 était tout aussi explicite : *Booke at large, for the Amendment of Orthographie for English speech*[5].

En France, les réformes qui entendent imposer une « écriture orale », selon l'expression de Nina Catach, entièrement commandée par les manières de dire, vont au-delà de la transformation des graphies. Ronsard, par exemple, propose dans son *Abbrégé de l'art poëtique françois* de supprimer « toute ortographie superflue » (c'est-à-dire toutes les lettres qui ne se prononcent pas), de transformer la graphie des mots afin de la rapprocher de la façon dont ils sont dits (ainsi « roze », « kalité », « Franse », « langaje », etc. – ce qui rendra inutiles le *q* et le *c*) et d'introduire des lettres doubles, à l'imitation du *ll* ou du *ñ* espagnol, pour fixer une prononciation plus exacte des mots « orgueilleux » ou « Monseigneur »[6].

La pratique des imprimeurs ne suivra pas ces propositions radicales. En revanche, elle introduira une innovation décisive pour une plus forte adéquation entre manières de dire et formes d'inscription des textes, à savoir : la fixation de la longueur des pauses. Le texte fondamental est ici celui de l'imprimeur (et auteur) Étienne Dolet, intitulé *La Punctuation de la langue françoise*. Il définit en 1540 les nouvelles conventions typographiques qui doivent distinguer, selon la durée des silences et la position dans la phrase, le « point à queue ou virgule », le « comma » (ou point virgule), « lequel se met en sentence suspendue et non du tout finie », et le « point rond » (ou point final) qui « se met toujours à la fin de la sentence[7] ». Les dictionnaires de langue de la fin du XVII[e] siècle enregistrent, tout ensemble, l'efficacité du système proposé par Dolet (enrichi des deux points qui indiquent une pause d'une durée intermédiaire entre le comma et le point final) et, déjà, la distance prise entre la voix lectrice et la

5. Jonathan Goldberg, *Writing Matter. From the Hands of the English Renaissance*, Stanford, Stanford University Press, 1989, et Jeffrey Masten, « Pressing subjects or The secret lives of Shakespeare's compositors », *in* Jeffrey Masten, Peter Stallybrass et Nancy Vickers (éd.), *Language Machines. Technologies of Literary and Cultural Production*, New York et Londres, Routledge, 1997, p. 75-107.

6. Ronsard, *Abbrégé de l'art poëtique françois* (1565), *in* Ronsard, *Œuvres complètes*, édition établie par Gustave Cohen, Paris, Gallimard, « Bibliothèque de la Pléiade », 1966, II, p. 995-1009.

7. Le texte de Dolet est reproduit en hors-texte dans l'ouvrage fondamental de Nina Catach, *L'Orthographe française à l'époque de la Renaissance (auteurs, imprimeurs, ateliers d'imprimerie)*, Genève, Librairie Droz, 1968.

ponctuation, considérée désormais, selon le terme du dictionnaire de Furetière, comme une « observation grammaticale » qui marque les divisions du discours.

Ainsi équipé pour indiquer les durées variables des pauses, le système de la ponctuation des textes ne l'est pas pour marquer les différences d'intensité ou de hauteur. De là, le détournement de la signification de certains signes utilisés pour signaler au lecteur les phrases ou les mots qu'il lui faut accentuer. Ainsi, pour Ronsard, le point d'exclamation. Dans l'avis qu'il adresse au lecteur dans les préliminaires aux quatre premiers livres de la *Franciade*, il indique :

> Je te supliray seulement d'une chose, Lecteur : de vouloir bien prononcer mes vers et accomoder ta voix à leur passion, et non comme quelques-uns les lisent, plutost à la façon d'une missive, ou de quelques lettres Royaux, que d'un Poëme bien prononcé ; et te suplie encore derechef, où tu verras cette marque *!* vouloir un peu eslever ta voix pour donner grace à ce que tu liras[8].

Il en va de même avec le point d'interrogation pour Racine. Comme l'a montré Georges Forestier, sa présence inattendue peut indiquer, exceptionnellement, un signe d'intonation, comme dans ce vers de *La Thébaïde* :

> Parlez, parlez, ma Fille **?**

Inversement, et plus fréquemment, l'absence de point d'interrogation à la fin de phrases interrogatives signale que la voix doit rester égale, sans montée d'intensité – ainsi dans cet autre vers dans la première édition de La Thébaïde :

> Ma Fille, avez-vous su l'excès de nos misères[9]**.**

Une autre pratique est celle qui dote d'une lettre capitale les mots qui doivent être accentués ou détachés. Elle est codifiée par les traités qui décrivent l'art de l'imprimerie, ainsi les *Mechanick Exercices on the Whole Art of Printing* de Joseph Moxon, publiés en 1683-1684, qui imposent l'emploi des majuscules pour des mots en cours de phrase et qui ne sont

8. Ronsard, « Au lecteur » *in Les Quatre Premiers Livres de la Franciade* (1572), *in* Ronsard, *Œuvres complètes*, *op. cit.*, II, p. 1009-1013.
9. Georges Forestier, « Lire Racine », *in* Racine, *Œuvres complètes*, t. I, *Théâtre-Poésie*, édition par Georges Forestier, Paris, Gallimard, « Bibliothèque de la Pléiade », 1999, p. LIX-LXVIII.

pas des noms propres lorsqu'ils doivent être l'objet d'une « *emphasis* »[10]. Un exemple frappant d'emploi de majuscules d'intensité est cité par Georges Forestier avec ce vers de *Bajazet* dit par Atalide, et maintenu dans toutes les éditions de la tragédie :

J'ai cédé mon **A**mant, **T**u t'étonnes du reste[11].

De l'usage musical de la ponctuation des pauses et de l'emploi des capitales comme marques d'intensité, *Les Caractères* de La Bruyère sont un magnifique exemple. En retournant à la ponctuation de l'édition de 1696, qui est la dernière que La Bruyère a pu revoir, et en débarrassant le texte d'une ponctuation anachronique, lourde et grammaticale, Louis Van Delft a pu, dans sa propre édition, restituer l'oralité de la composition comme de la lecture des *Caractères*[12]. La Bruyère privilégie l'usage de la virgule, traitée comme un soupir, refuse les guillemets et, surtout, traite chaque « remarque » comme une phrase musicale unique, qui alterne les séquences rapides et agitées, rythmées par les césures, avec des périodes plus longues, sans ponctuation. Cette composition, où la ponctuation est distribuée en fonction du souffle, est une claire invitation à lire le texte à haute voix, pour soi-même ou pour d'autres. Et, de fait, on sait que Jean-Marie Villégier a fait des *Caractères* la matière de lectures publiques. D'autre part, les majuscules placées au début de nombreux mots dans le cours même du texte témoignent pour l'acuité de ce que l'on pourrait appeler la « conscience typographique » de La Bruyère, qui joue souvent avec les effets, visuels ou sémantiques, produits par les formes données au texte. Par exemple, en mettant en italique ou en enserrant par des virgules les propos rapportés ou les langages en usage qui sont la cible même de l'ironie et de la critique, ou bien en utilisant les majuscules comme « un coefficient de dignité » selon l'expression de Louis Van Delft.

Mais peut-on supposer que tous les auteurs ont été aussi attentifs que Ronsard ou La Bruyère à la ponctuation de leurs œuvres ? Ce n'est pas ce qu'indiquent les textes anciens qui, écrits par des hommes de l'art typographiques, insistent sur le rôle décisif des compositeurs et des correcteurs des imprimeries. Dans l'Espagne du Siècle d'Or, la « *apuntuación* » est une de leurs tâches essentielles lors de la préparation ou de la

10. Joseph Moxon, *Mechanick Exercises on the Whole Art of Printing (1683-4)*, édité par Herbert Davis et Harry Carter, Oxford et Londres, Oxford University Press, 1958, p. 216-217.
11. Georges Forestier, « Lire Racine », *in* Racine, *Œuvres complètes, op. cit.,* p. LXI, note 4.
12. Louis Van Delft, « Principes d'édition », *in* La Bruyère, *Les Caractères*, présentation et notes de Louis Van Delft, Paris, Imprimerie nationale, 1998, p. 45-57.

composition du texte, tout comme la colocation des accents et celle des parenthèses. En 1619, Gonzalo de Ayala, qui était lui-même correcteur d'imprimerie, indique que le correcteur « doit connaître la grammaire, l'orthographe, les étymologies, la ponctuation, la disposition des accents ». En 1675, Melchor de Cabrera, un avocat du Conseil du Roi qui défend les exemptions fiscales des imprimeurs madrilènes, souligne que le compositeur doit savoir « placer les points d'interrogation et d'exclamation et les parenthèses ; parce que souvent l'intention des écrivains est rendue confuse du fait de l'absence de ces éléments, nécessaires, et importants pour l'intelligibilité et la compréhension de ce qui est écrit ou imprimé, parce que si l'un ou l'autre fait défaut, le sens est changé, inversé et transformé ». Quelques années plus tard, pour Alonso Víctor de Paredes, le correcteur doit « comprendre l'intention de l'Auteur dans ce qu'il fait imprimer, non seulement pour introduire la ponctuation adéquate, mais aussi pour voir s'il n'a pas commis quelques négligences, afin de l'en avertir[13] ». Les formes et les dispositions du texte imprimé ne dépendent donc pas de l'auteur, qui délègue à celui qui prépare la copie ou à ceux qui composent les pages les décisions quant à la ponctuation, l'accentuation et les graphies.

Dans ce partage de la responsabilité quant à la ponctuation, chaque tradition de la critique textuelle a privilégié l'un ou l'autre des acteurs engagés dans le processus de composition et de publication des textes entre le XV^e^ et le XVIII^e^ siècle, à l'âge de ce que l'on peut appeler l'ancien régime typographique. Pour la bibliographie matérielle, les choix graphiques et orthographiques sont le fait des compositeurs. Les ouvriers typographes des ateliers anciens n'avaient pas tous la même manière d'orthographier les mots ou de marquer la ponctuation. De là, la récurrence régulière des mêmes graphies ou des mêmes usages des signes de ponctuation dans les différents cahiers d'un même ouvrage en fonction des préférences et des habitudes du compositeur qui a en composé les pages. C'est pourquoi les « *spelling analysis* », avec l'étude de la récurrence des caractères endommagés ou des ornements, ont permis d'attribuer la composition de telle ou telle page à tel ou tel compositeur et, ainsi, de reconstituer le processus même de fabrication du livre, soit *seriatim* (c'est-à-dire en suivant l'ordre du texte), soit par formes (c'est-à-dire en composant à la suite

13. Melchor de Cabrera Nuñez de Guzman, *Discurso legal, histórico y político en prueba del origen, progressos, utilidad y excellencias del Arte de la Imprenta* ; *y de que se le deben (y a sus Artifices) todas las Honras, Exempciones, Inmunidades, Franquezas y Privilegios del Arte Liberal, por ser, como es, Arte de las Artes*, Madrid, 1675 ; Alonso Víctor de Paredes, *Institución y Origen del Arte de la Imprenta y Reglas generales para los componedores*, édition de Jaime Moll, Madrid, Calembur, 2002.

toutes les pages assemblées dans un même châssis de bois, appelé « forme », et imprimées sur le même côté d'une feuille d'imprimerie – par exemple, pour un in-quarto les pages 2, 3, 6 et 7 –, ce qui permet de commencer l'impression d'une feuille alors même que toutes les pages d'un même cahier n'ont pas encore été composées mais ce qui suppose, aussi, le calibrage préalable et exact de la copie manuscrite)[14].

Dans la perspective des « *compositor studies* », fondée sur l'examen méticuleux de la matérialité des ouvrages imprimés et des modalités d'inscription des textes sur la page, la ponctuation est considérée, à l'instar des variations graphiques et orthographiques, comme le résultat, non des volontés de l'auteur du texte, mais des habitudes ou, parfois, des obligations, si le calibrage a été mal fait, des ouvriers qui l'ont composé pour qu'il devienne un livre imprimé. Comme l'écrit joliment Alonso Víctor de Paredes, « *no son Angeles los que cuentan* » – ce ne sont pas des Anges qui font le calibrage. Si la division de la copie a été maladroite, la composition des dernières pages d'un même cahier exige des ajustements qui peuvent aller, comme il le dit avec réprobation, jusqu'à « l'emploi des procédés laids et qui ne sont pas permis », entendons des ajouts ou des suppressions de mots ou de phrases qui ne doivent rien à la volonté de l'auteur, mais tout aux embarras des compositeurs, qui peuvent aussi jouer sur la mise en page, la taille des caractères ou la ponctuation, allégée pour économiser les espaces blancs ou alourdie pour étirer le texte.

En un temps caractérisé par une grande « plasticité phonétique, orthographique et sémantique », particulièrement dans le cas de la langue anglaise[15], la marge de décision laissée aux typographes est très large et certains auteurs, dans le contrat signé avec l'imprimeur, délèguent explicitement au jugement des compositeurs la ponctuation de leur ouvrage[16]. Moxon, dans ses *Mechanick Exercises*, insiste sur la contribution fondamentale des compositeurs non seulement à la production du livre mais à celle du texte :

> L'ambition d'un bon Compositeur doit être de rendre intelligible au Lecteur le sens voulu par l'auteur et faire que son travail soit gracieux pour

14. À titre d'exemples, *cf.* D. F. McKenzie, « Compositor B's Role in the "Merchant of Venice" Q2 (1619) », *Studies in Bibliography*, vol. 12, 1959, p. 75-89 ; ou Charlton Hinman, *The Printing and Proof-Reading of the First Folio of Shakespeare*, Oxford, Clarendon Press, 1963.
15. Margreta de Grazia et Peter Stallybrass, « The Materiality of the Shakespearean Text », *Shakespeare Quartely*, Vol. 44, No 3, 1993, p. 255-283 (citation p. 266).
16. Un exemple dans James Binns, « STC Latin Book : Evidence for printing-house practice », *The Library*, Fifth Series, vol. XXXII, n° 1, 1997, p. 1-27 : n° 10 : M. A. de Dominis, *De republica ecclesiastica*, 1617.

> l'œil et plaisant pour la lecture. Si la copie est écrite dans une langue qu'il comprend, il doit la lire avec attention, de façon à entrer dans le sens voulu par l'auteur et, en conséquence, il doit considérer comment organiser son travail le mieux possible, tant pour la page de titre que pour le corps du livre : c'est-à-dire comment distribuer les paragraphes, la ponctuation, les coupures de lignes, les italiques, etc., de la manière la mieux accordée avec le génie de l'auteur et aussi la capacité du Lecteur[17].

Moxon oppose ainsi un processus de publication fondé sur une collaboration éclairée aux plaintes fréquentes des auteurs déplorant que leur œuvre ait été déformée par l'ignorance ou la négligence des ouvriers typographes.

Dans une autre perspective, plus philologique, l'essentiel est ailleurs : dans la préparation du manuscrit pour la composition telle qu'elle est opérée par les « correcteurs » qui ajoutent capitales, accents et ponctuation, qui normalisent l'orthographe, qui fixent les conventions graphiques et qui, souvent, sont en charge de la correction des épreuves. C'est pourquoi, pour Moxon, les décisions prises par le compositeur sont toujours soumises en dernière instance au contrôle du correcteur qui « examine les épreuves, et considère la ponctuation, les italiques, les majuscules, ainsi que toute erreur que celui-ci a pu commettre par méprise ou par manque de jugement[18] ». S'ils restent le résultat d'un travail d'atelier, les choix quant à la ponctuation ne sont plus ici assignés seulement ou principalement aux compositeurs, mais aux humanistes (clercs, gradués des universités, maîtres d'école) employés par les libraires et les imprimeurs pour assurer la plus grande correction possible de leurs éditions. Paolo Trovato a rappelé combien il était important pour les éditeurs du Cinquecento d'insister sur la « correction » de leurs éditions, affirmée sur les pages de titre par l'expression « *con ogni diligenza corretto*[19] ». *D'où le rôle décisif des « correcteurs » dont les interventions se déploient à plusieurs moments du processus d'édition : de la préparation du manuscrit à la correction des épreuves, des corrections en cours de tirage, à partir de la révision des feuilles déjà imprimées, à l'établissement des errata*, en leurs diverse formes – les corrections à la plume sur les exemplaires imprimées, les feuillets d'*errata* ajoutés à la fin du livre ou les invitations faites au lecteur pour qu'il corrige lui-même son propre exemplaire[20]. À

17. Joseph Moxon, *Mechanick Exercises*, *op. cit.,* p. 211-212.
18. *Ibid.,* p. 247.
19. Paolo Trovato, *« Con ogni diligenza corretto ». La stampa e le revisioni dei testi letterari italiani (1470-1570)*, Bologne, Il Mulino, 1991.
20. Des exemples de ces invitations au lecteur, accompagnées d'une liste d'errata ou adressées à son propre jugement, dans l'article de James Binns, art. cit., n° 32, 33, 35 et 36.

chacune de ces étapes, la ponctuation du texte peut-être corrigée, transformée ou enrichie.

Aux XVI^e^ et XVII^e^ siècles, les textes soumis ainsi à la ponctuation des « correcteurs », intervenant comme « *copy editor* » ou « *proofreader* », appartiennent à différents répertoires : les textes classiques, grecs ou latins[21] ; les œuvres en langue vulgaire qui ont eu une circulation manuscrite et auxquelles l'imprimerie impose ses propres normes de présentation du texte et, dans certains cas comme celui des éditions italiennes, une normalisation graphique et linguistique[22] ; enfin, les manuscrits des contemporains dont la fort médiocre lisibilité irritait fort les correcteurs. Jérôme Hornschuch dans son *Orthotypographia. Instruction utile et nécessaire pour ceux qui vont corriger les livres imprimés*, de 1608, vilipende les auteurs qui remettent aux imprimeurs des manuscrits qu'ils ont rédigés avec négligence ou qu'ils ont fait copier par des scribes peu soigneux. Le travail des compositeurs et celui des correcteurs s'en trouvent profondément affectés et c'est pourquoi, déclare Hornschuch, « je voudrais donc, non tant au nom des correcteurs, qu'à celui des imprimeurs, admoneste et supplier avec insistance tous ceux qui feront un jour publier quelque chose, de le présenter de telle façon qu'il ne soit pas nécessaire de demander dans l'atelier de l'imprimeur avec l'esclave de la comédie : "Les poules ont-elles aussi des mains[23] ?" » – allusion à une réplique de la comédie de Plaute *Pseudolus*. Renversant les rôles tels qu'ils sont ordinairement distribués, Hornschuch demande à l'auteur de prendre un soin tout particulier de la ponctuation :

> De plus, ce qui est presque le plus important de tout, qu'il mette une ponctuation exacte. Car, chaque jour, de nombreuses erreurs sont commises par beaucoup de gens dans ce domaine. Et en poésie, rien n'est plus fâcheux et plus blâmable que le nombre de gens qui omettent la ponctuation, ce qui nous rappelle peut-être un je ne sais quoi de la corneille d'Ésope. À coup sûr, une bonne ponctuation apporte une grande élégance au texte et plus que toute autre chose permet une bonne compréhension du sujet, alors que ne pas s'en soucier semble être le fait d'un esprit dissolu[24].

---

21. Anthony Graftton, « Printer's correctors and the publication of classical text », *in* Anthony Grafton, *Bring Out Your Dead. The Past as Revelation*, Cambridge (Mass.) et Londres, Harvard University Press, 2001, p. 141-155.
22. Brian Richardson, *Print Culture in Renaissance Italy. The Editor and the Vernacular 1470-1600*, Cambridge, Cambridge University Press, 1994.
23. Jérôme Hornschuch, *Orthotypographia. Instruction utile et nécessaire pour ceux qui vont corriger des livres imprimés et conseils à ceux qui vont les publier (1608),* traduction du latin de Susan Baddeley avec une introduction et des notes de Jean-François Gilmont, Paris, Édition des Cendres, 1997, p. 87.
24. *Ibid.,* p. 89.

Pourtant, au Siècle d'Or, les manuscrits des auteurs n'étaient jamais utilisés par les typographes qui composaient avec les caractères mobiles les pages du livre à venir. La copie qu'ils utilisaient était le texte mis au propre par un scribe professionnel qui avait été envoyé au Conseil du Roi pour recevoir les approbations des censeurs, puis la permission d'imprimer et le privilège du roi. Rendu à l'auteur, c'est ce manuscrit qui était remis au libraire-éditeur, puis au maître imprimeur et à ses ouvriers. Un premier écart sépare donc le texte tel que l'a rédigé l'écrivain (que Francisco Rico désigne comme le « *borrador* », le manuscrit raturé) de la « *copia en limpio* » ou « original », mis en forme par un copiste qui lui impose des normes tout à fait absentes des manuscrits d'auteur[25]. Comme on le sait, ceux-ci n'observent aucune régularité graphique et ignorent quasiment la ponctuation, tandis que les « originaux » (qui, de fait, ne le sont pas) devaient assurer une meilleure lisibilité du texte soumis à l'examen des censeurs.

Dans son dictionnaire, Furetière propose deux exemples d'emplois pour le terme « ponctuation » :

> Ce Correcteur d'Imprimerie entend fort bien la ponctuation

et

> L'exactitude de cet Autheur va jusques là qu'il prend soin des points et des virgules.

Si le premier exemple assigne tout à fait normalement la ponctuation à la compétence technique propre aux correcteurs employés par les imprimeurs, le second, implicitement, renvoie au désintérêt ordinaire des auteurs pour la ponctuation, mais il signale aussi qu'il est des auteurs attentifs à la ponctuation de leurs textes, comme le montrent les exemples de Ronsard et La Bruyère. Soit un autre cas, celui de Molière. Est-il possible de trouver trace de l'« exactitude » qu'évoque Furetière dans les éditions imprimées de ses œuvres ? Il serait très risqué de lui attribuer trop directement les choix de ponctuation tels que les donnent les éditions originales de ses pièces puisque, comme Jeanne Veyrin-Forrer l'a montré pour l'édition de 1660 des *Précieuses ridicules*, ils varient selon les différentes feuilles, voire les différentes formes, au gré des préférences des compositeurs ou des correcteurs[26]. Pourtant, les écarts de ponctuation qui

25. Francisco Rico, *El texto del « Quijote ». Preliminares a una ecdótica del Siglo de Oro*, Barcelone, Ediciones Destino, 2006.
26. Jeanne Veyrin-Forrer, « À la recherche des "Précieuses" », *in* Jeanne Veyrin-Forrer, *La Lettre et le Texte. Trente années de recherches sur l'histoire du livre*, Paris, Collection de l'École normale supérieure de jeunes filles, 1987, p. 338-366.

existent entre les premières éditions des pièces, publiées peu de temps après leurs premières représentations parisiennes, et les éditions postérieures permettent de reconstruire, sinon les intentions de l'auteur, du moins les destinations attendues du texte imprimé.

On sait les réticences de Molière devant la publication imprimée de ses pièces. Avant les *Précieuses ridicules* et la nécessité de devancer la publication du texte par Somaize et Ribou, faite à partir d'une copie dérobée et sous le couvert d'un privilège obtenu par surprise, jamais Molière n'avait livré l'une de ses comédies à l'impression. Il y avait à cela des raisons financières – puisque, une fois publiée, une pièce peut être jouée par n'importe quelle troupe – mais aussi des raisons esthétiques. Pour Molière, en effet, l'effet du texte de théâtre tient tout entier dans l'« action », c'est-à-dire dans la représentation. L'adresse au lecteur qui ouvre l'édition de l'*Amour médecin*, représenté à Versailles puis sur le théâtre du Palais-Royal en 1665 et publié l'année suivante, souligne l'écart entre le spectacle et la lecture :

> Il n'est pas nécessaire de vous avertir qu'il y a beaucoup de choses qui dépendent de l'action : on sait bien que les comédies ne sont faites que pour être jouées ; et je ne conseille de lire celle-ci qu'aux personnes qui ont des yeux pour découvrir dans la lecture tout le jeu du théâtre[27].

Pourquoi, alors, ne pas penser que la ponctuation est l'un des supports possibles (avec l'illustration des frontispices et les didascalies) pour que soit restitué dans le texte imprimé et dans sa lecture quelque chose de l'« action » et de la parole du théâtre ?

Comparée systématiquement à celle adoptée dans les éditions postérieures (non seulement au XIX^e^ siècle mais aussi dès les XVII^e^ et XVIII^e^ siècles), la ponctuation des premières éditions des pièces de Molière atteste clairement son lien à l'oralité, soit qu'elle destine le texte imprimé à une lecture à haute voix ou à une récitation, soit qu'elle permette au lecteur qui lira en silence de reconstruire, intérieurement, les temps et les pauses du jeu des acteurs. Comme l'a montré Gaston Hill, le passage d'une ponctuation à l'autre est loin d'être sans effets sur le sens même des œuvres[28]. D'une part, les ponctuations premières, toujours plus nombreuses, caractérisent différemment les personnages – ainsi la virgule présente

27. Molière, « Au lecteur », *L'Amour médecin*, *in* Molière, *Œuvres complètes*, t. II, textes établis par Georges Couton, Gallimard, « Bibliothèque de la Pléiade », 1971, p. 95.
28. Gaston H. Hill, « Ponctuation et dramaturgie chez Molière », *in* Roger Laufer, *La Bibliographie matérielle* (présentée par Roger Laufer, table ronde du CNRS organisée par Jacques Petit), Paris, Éditions du CNRS, 1983, p. 125-141.

dans l'édition de 1669 et disparue ensuite après le premier mot (« Gros ») dans ce vers plus que célèbre de Tartuffe :

> Gros, et gras, le teint frais, et la bouche vermeille[29].

ou encore la multiplication des virgules et des capitales pour distinguer les manières de dire du maître de philosophie de celles du maître de danse dans *Le Bourgeois Gentilhomme*[30]. *D'autre part, les ponctuations des éditions originales donnent des pauses qui permettent les jeux de scène (ou leur reconstitution imaginée). Par exemple, dans la scène des portraits du Misanthrope*[31], l'édition de 1667 contient six virgules de plus que les éditions modernes, ce qui permet à Célimène de détacher les mots, de prendre des temps, de multiplier les mimiques. Enfin, ces ponctuations originelles mettent en évidence des mots chargés d'une signification particulière. Alors que les deux derniers vers de Tartuffe ne comportent aucune virgule dans les éditions modernes, il n'en va pas ainsi dans l'édition de 1669 :

> Et par un doux hymen, couronner en Valère,
> La flame d'un Amant généreux, & sincère.

Le dernier mot de la pièce, « sincère » est ainsi clairement désigné comme l'antonyme celui qui figure au titre, *Le Tartuffe ou l'Imposteur*. Cette ponctuation abondante, qui indique des pauses plus nombreuses et, généralement, plus longues que celles retenues ensuite, enseigne au lecteur comment il doit dire (ou lire) les vers et faire ressortir un certain nombre de mots, généralement dotés de capitales d'imprimerie, elles aussi supprimées dans les éditions postérieures. Quel que soit le responsable de cette ponctuation (Molière, un copiste, un correcteur, les compositeurs), elle indique une forte relation avec l'oralité, celle de la représentation du théâtre ou celle de la lecture de la pièce.

Dans l'Angleterre des XVI^e^ et XVII^e^ siècles, nombreux sont les jeux poétiques ou dramatiques avec la ponctuation dont les variations transforment ou inversent le sens d'un texte sans en changer un seul mot. Les créations poétiques s'emparent de la formule en proposant plusieurs significations d'un même vers s'il est lu en suivant une ponctuation ou une autre[32]. Il en va de même sur la scène. Au dernier acte de *A Midsummer*

29. Acte I, scène 4, vers 233.
30. Acte II, scène 3.
31. Acte II, scène 4, vers 586-594
32. Un exemple de « *punctuation poem* » est donné par Malcom Parkes, *Pause and Effect. An Introduction to the History of Punctuation in the West*, Berkeley et Los Angeles, University of California Press, 1993, p. 210-211.

Night's Dream (*Le Songe d'une nuit d'été*), Quince ouvre la représentation de Pyrame et Thisbé, donnée par les artisans d'Athènes à la cour de Thésée, par un prologue qui se veut une *captatio benevolentiae*. Mais il coupe les phrases par des pauses mal placées et le texte énonce le contraire de son intention, ce qui fait dire à Thésée « *This fellow does not stand upon points* » et à Lysandre « *he knows not the stop* ». Dans le Folio de 1623, les imprimeurs ont respecté la comique maladresse de Quince en introduisant des points, ou « *stops* », là où il ne le faudrait pas.

> If we offend, it is with our good will.
> That you should think, we come not to offend,
> But with good will. To show our simple skill,
> That is the true beginning of our end[33].

ce que l'on pourrait traduire comme

> Si nous vous déplaisons, c'est notre intention.
> Ne pensez pas que nous ne voulons pas le faire, car c'est notre intention.
> De vous montrer notre simple savoir-faire,
> Tel est de notre fin le vrai commencement.

La ponctuation attendue et correcte de ces mêmes vers leur donne un sens tout à fait opposé sans qu'un seul mot soit changé – ce qui est impossible à rendre en traduction :

> If we offend, it is with our good will
> That you should think, we come not to offend.
> But with good will to show our simple skill :
> That is the true beginning of our end.

soit :

> Si nous vous déplaisons, c'est notre intention
> Que vous ne pensiez pas que nous voulons le faire.
> Car c'est notre intention de vous montrer notre simple savoir-faire :
> Tel est de notre fin le vrai commencement.

Mortimer le fourbe et Quince le maladroit rappellent ainsi que la ponctuation construit la signification et qu'elle entretient un lien puissant

33. William Shakespeare, *A Midsommer Nights Dreame*, *in The First Folio of Shakespeare*, 1623, préparé par Doug Moston, *Facsimile*, New York et Londres, Applause, 1995, p. 145.

avec les manières de dire. Faut-il dès lors souscrire à la thèse de William Nelson selon laquelle, à la fin du XVII^e siècle, une ponctuation grammaticale et syntaxique, éloignée de la voix, remplacerait une ponctuation d'oralité qui indiquait pauses et intonations[34] ? Ou bien doit-on considérer avec Malcolm Parkes qu'à partir de la Renaissance l'essentiel est, pour une même époque, voire dans un même texte, l'oscillation entre une ponctuation rhétorique, qui marque la structure des périodes, et une autre, syntaxique, qui identifie les articulations logiques du discours[35] ? Et peut-on supposer que tous les acteurs auxquels la ponctuation d'un texte ancien peut être attribuée, aux différents moments de sa trajectoire, ont partagé les mêmes normes et les mêmes attentes ? Ce sont là les questions qui peuvent servir de toile de fond à une interrogation sur les retours à une ponctuation d'oralité au XVIII^e siècle, illustrés par l'introduction dans la langue castillane, à l'initiative de la Real Academia, des points d'interrogation et d'exclamation inversés, ou par la volonté de Benjamin Franklin de construire le nouvel espace public sur la parole vive des orateurs, enseignée au collège et reproduite grâce aux dispositifs typographiques mobilisant les capitales, les italiques et, du moins le souhaitait-il, les signes de ponctuation inversés de l'espagnol qui indiquent, d'emblée, comment doit être posée la voix. Ces questions peuvent aussi, je crois, permettre de mieux comprendre la nostalgie pour une oralité perdue qui tourmente nos sociétés.

34. William Nelson, « From "Listen Lordings" to "Dear Reader" », *University of Toronto Quarterly. A Canadian Journal of the Humanities*, vol. XLVI, n° 2, 1976-1977, p. 110-124.
35. Malcolm Parkes, *Pause and Effect*, *op. cit.*, p. 5.

Entre parole et musique : les langages tambourinés d'Afrique subsaharienne

par SIMHA AROM

Les langages tambourinés constituent un mode de communication très répandu en Afrique subsaharienne. Ils sont régis par des lois relevant à la fois de la langue et de la musique puisque, pour transmettre des messages *linguistiques*, on a recours à des *instruments de musique*.

Ce mode de communication est largement exploité grâce à une caractéristique commune à de nombreuses langues africaines – celle d'être des langues à tons. Dans les langues à tons africaines, chaque voyelle est nécessairement affectée d'une hauteur : tout changement du ton d'une voyelle peut modifier le sens du mot au sein duquel elle figure. Ainsi, dans la langue ngbaka (République centrafricaine), qui a recours à trois hauteurs tonales – haut, représenté par un accent aigu ; bas, par un accent grave ; moyen, par un trait horizontal – et à deux tons dits modulés, figurés par un accent grave suivi d'un accent aigu et inversement, on relève :

kà	bientôt
kā	plaie, blessure
ká	saleté, ordure
kāā	seulement
kàà	encore, toujours
kàáà	très longtemps
mà	soulever, porter
mā	orage, pluie
má	comment
màā	je (accompli verbal)
màá	à moi, mien
lɛ̀	enfant
lɛ̄	essayer
lɛ́	pêche, pêchage

Dans ces langues, les tons ont une fonction distinctive, au même titre que les voyelles et les consonnes ; « à eux seuls, ils permettent non seulement des oppositions lexicales mais aussi morphologiques[1] ». En effet, dans de nombreuses langues à tons africaines, l'essentiel du système de conjugaison repose sur des alternances tonales. Les tons de la langue n'ont pas de hauteur fixe ; leur pertinence tient à leur opposition, non à la grandeur des intervalles qui les séparent.

> C'est dans le cadre d'un paradigme tonal que tel ton se caractérise comme haut, par rapport à tel autre qui se caractérise comme bas. C'est ce qui explique que, dans la chaîne [parlée], un ton haut puisse être réalisé plus bas qu'un ton bas précédent[2].

Pour expliciter les différents aspects du fonctionnement du langage tambouriné, je m'appuierai sur celui des Banda-Linda de Centrafrique, branche de l'ethnie banda, établie au centre du pays ; leur dialecte, le linda, est parlé par 27 000 personnes environ.

L'étude d'un tel sujet – par définition interdisciplinaire, puisque situé à l'intersection entre langage et musique – ne peut être menée à bien qu'en étroite collaboration entre chercheurs des deux disciplines que sont la linguistique et l'ethnomusicologie.

Ce qui suit est le résultat d'un travail effectué en 1973 dans le cadre du LACITO du CNRS avec ma collègue linguiste France Cloarec-Heiss, spécialiste de la langue banda-linda, dans la minuscule sous-préfecture d'Ippy[3].

Le linda ne connaît ni opposition entre voyelles longues et voyelles brèves, ni marque accentuelle. En revanche, il présente trois hauteurs tonales – haut, moyen, bas – et deux tons modulés, réalisés par un glissement rapide du ton bas au ton haut et inversement (dans le schéma ci-dessous, V désigne une voyelle).

ton haut	V́	ton haut-bas	V̌
ton bas	V̀	ton bas-haut	V̂
ton moyen	V̄		

Rappelons que la hauteur d'un ton donné n'est pas absolue, mais relative.

La transmission des messages s'effectue au moyen de deux tambours de bois, `l"eng"a`, de taille différente. Leurs flancs convexes sont d'épaisseur inégale – ce qui permet sur chacun d'eux la production de deux sons de hauteur différente. Le plus grand est appelé `èy"K.l"eng"a`, "mère-

1. Cloarec-Heiss (1997), p. 137.
2. Thomas, Bouquiaux, Cloarec-Heiss (1976), p. 110.
3. Arom et Cloarec-Heiss (1976).

lenga", le plus petit àk"O.l"eng"a, "époux-lenga". Les tons bas et moyen sont réalisés de part et d'autre de la fente longitudinale du tambour èy"K.l"eng"a, le ton haut sur l'un des flancs du àk"O.l"eng"a. La frappe est réalisée au moyen de deux mailloches dont l'extrémité est entourée de bandelettes de caoutchouc sauvage.

Figure 1 : *Un tambourinaire*

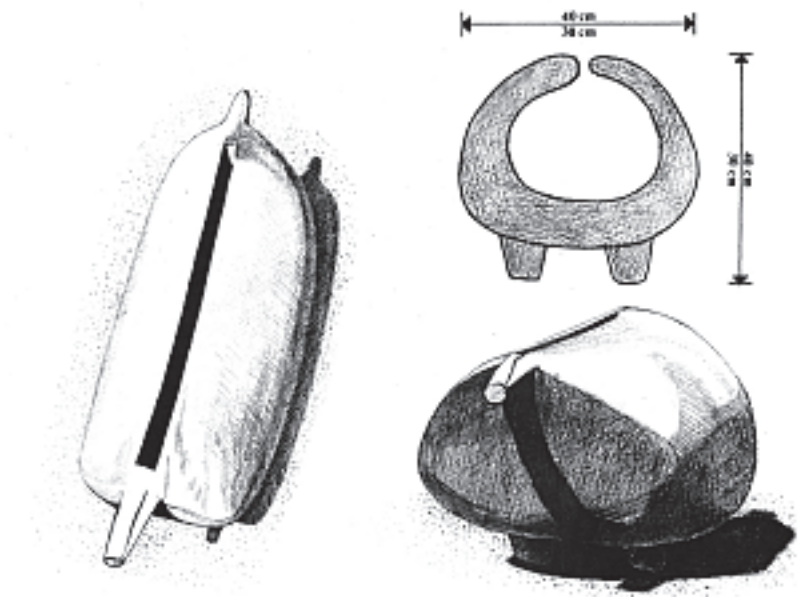


Figure 2 : *Tambours*

La durée des messages tambourinés est de cinq minutes environ. Leur portée, qui dépend de l'environnement et des conditions météorologiques, peut couvrir une douzaine de kilomètres. Les messages peuvent être répercutés dans un rayon infiniment plus vaste au moyen de relais. Toutefois, leur diffusion est tributaire des limites dialectales. Pour accroître leur rayon d'efficacité, on les transmet de préférence la nuit, à l'heure la plus favorable, juste avant le lever du soleil.

Traditionnellement, chaque village possède une paire de `l"eng"a` dont le chef est le dépositaire.

Tout comme les autres langages tambourinés de cette région, celui des Linda ne conserve de la langue que les hauteurs tonales et le rythme de l'élocution.

> On peut se demander comment, après une réduction aussi importante du matériel linguistique, la communication reste encore possible et comment sont résolus les problèmes posés par l'homophonie tonale. On va voir que ce sont les conditions sociolinguistiques d'utilisation de ce système et la structure des messages qui, en permettant la mise en œuvre d'opérations cognitives de reconstruction du sens, compensent la perte d'information entraînée par cette réduction[4].

Il s'agit d'un mode de communication *unilatéral* ; le but des informations ainsi diffusées est de susciter la venue d'une personne ou d'un groupe de personnes vers la source d'émission, mais non d'obtenir une réponse.

Fonctions du langage tambouriné

En théorie, un tel système permet d'émettre tout énoncé ou succession d'énoncés possible dans la langue. Félix Éboué n'a pas tort lorsqu'il affirme qu'au moyen du langage tambouriné, les Banda peuvent « exprimer absolument toutes les idées qu'il est possible d'émettre avec la langue parlée[5] ». Toutefois, en raison de l'importante homophonie tonale entre des mots différents, la probabilité de compréhension serait extrêmement faible. Bref, si l'on peut tout dire, on ne peut pas tout comprendre…

4. Cloarec-Heiss (1997), p. 137.
5. Éboué (1933), p. 80.

Le langage tambouriné accentue donc la *dissymétrie* entre l'émetteur et le récepteur : ce qui n'est pas ambigu pour le premier risque souvent de l'être pour le second.

Dans la réalité, le recours au langage tambouriné répond avant tout à une fonction sociale : aussi est-il limité à une quinzaine de situations qui correspondent chacune soit à un événement – naissance, décès, retrait de deuil, consécration de nouveaux tambours, fête d'investiture –, à des accidents (chasse, incendie, personne perdue en brousse), ou encore à des circonstances apparues avec la colonisation, telles l'arrivée d'un administrateur, d'une équipe sanitaire, le marché au coton ou la collecte de l'impôt. Toutefois, la nature du corpus recueilli montre que le système permet d'engendrer des messages « inédits ».

Méthodologie

Notre objectif était de mettre au jour la systématique qui régit le langage tambouriné des Banda-Linda, de vérifier dans quelle mesure les destinataires des messages parviennent à en saisir le contenu, et de comprendre les processus cognitifs de leur reconstitution. Dans cette perspective, l'enquête se devait de faire intervenir – outre le tambourinaire-*émetteur*, Michel Waanga, un villageois – Simon Ngoadukuzu, faisant office de *décodeur*. C'est pourquoi la totalité des messages a été recueillie en deux versions successives, l'une, dans des conditions normales d'émission, c'est-à-dire en *continu*, l'autre en discontinu ou *alterné*. Dans la version alternée, chaque énoncé tambouriné était suivi d'une pause, durant laquelle le décodeur – *qui ignorait la teneur du message* – le décryptait aussitôt en banda-linda.

Cette procédure était destinée à confirmer le parallélisme du profil mélodique des schèmes tonals des messages tambourinés avec celui du langage articulé, et à vérifier le degré et la vitesse de compréhension du destinataire. Il faut savoir que, en milieu traditionnel, le décodage des messages tambourinés n'est pas affaire de spécialiste : tout locuteur banda-linda est à même de les comprendre. La rapidité et la précision avec laquelle les phrases successives ont été décodées attestent, de façon spectaculaire, l'efficacité de ce type de communication.

Les messages ne sont pas figés. La comparaison, pour chacun d'eux, des versions continue et alternée met en évidence de nombreuses variantes qui concernent notamment l'ordre de succession de leurs segments mélodico-rythmiques. À quelques exceptions près, leur contenu linguistique

est conforme à la langue : les messages sont constitués d'énoncés complets dont la syntaxe correspond à celle de la langue parlée. L'observation a mis en évidence une correspondance terme à terme entre les trois hauteurs frappées par les tambours et celles qui affectent chacune des voyelles des énoncés correspondants dans la langue. De même, la comparaison des messages tambourinés avec leur décodage en banda-linda montre une adéquation parfaite de leur articulation rythmique.

Structure des messages

À l'audition, un message tambouriné se présente comme une longue succession de coups frappés à une cadence très rapide sur trois hauteurs fixes, formant des motifs "musicaux" séparés par des pauses ; certains motifs sont instantanément identifiables en raison de leur récurrence sporadique, attestant par là un fort taux de redondance.

Le phénomène de récurrence met en évidence l'existence de deux classes paradigmatiques. Par *paradigme*, il faut entendre ici un ensemble d'énoncés, ou de parties d'énoncés dont la notion centrale est définie par le sens. L'analyse de la totalité du corpus a fait apparaître qu'*il n'y a pas d'énoncé qui ne trouve sa place dans l'une ou l'autre de ces classes*.

La première comprend les énoncés communs à tous les messages – formules d'adresse, de convocation, d'incitation à l'écoute, d'authentification et formules de fin ; la seconde regroupe l'inventaire des formules spécifiques contenues dans l'ensemble des messages.

> *Formules d'adresse* : Eh, les cultivateurs / Eh, les chefs de village / Eh, les invités !

Lorsqu'un message concerne un seul destinataire, son identification est assurée par un procédé de surdétermination. L'émetteur frappera alors : « Untel, fils d'Untel, des enfants du lignage Untel. »

> *Formules de convocation* : Rassemblez-vous / Rassemblez-vous tous / Rassemblez vous vite / Venez vite vers moi / Venez vers moi vite vite vite / Courez vers moi / Venez vers moi en foule.
> *Formules d'incitation* : Avez-vous compris ? /Avez-vous tout compris ? / Avez-vous tout bien compris ? / Ne faites pas les idiots / Vous êtes en train d'hésiter.
> *Formules d'authentification* : C'est la vérité que je vous dis / C'est la vérité que je vous dis là / C'est la vérité que je suis en train de vous dire / Je ne me trompe pas / C'est à vous que je m'adresse.
> *Formule de fin de message* : C'est la dernière fois que je vous parle.

La seconde classe paradigmatique regroupe les formules propres à chaque message.

Le message lançant l'invitation à la cérémonie d'investiture d'un chef traditionnel servira d'exemple pour illustrer et rendre concret ce qui vient d'être dit. Voici successivement, traduites en français, la version continue, puis la version alternée. La segmentation du texte, tel qu'il est présenté, reproduit la segmentation sonore du message. Les unités ainsi définies sont numérotées par ordre d'occurrence :

1. Eh ! les invités !
2. C'est de la fête que je vous parle.
3. C'est d'une fête que je vous parle,
4. et je dis :
5. On donne l'investiture.
6. Et je dis :
7. Rassemblez-vous tous !
8. C'est de la fête du tambour que je vous parle.
9. Et je dis :
10. Ne faites pas les idiots !
11. Venez vite vers moi !
12. C'est d'une fête que je vous parle.
13. C'est de la fête de Wayewo que je vous parle.
14. Venez vers moi pour la fête de Wayewo dont je suis en train de vous parler.
15. Rassemblez-vous et venez vite vers moi !
16. Rassemblez-vous et venez vite vers moi !
17. C'est de la fête d'investiture que je vous parle.
18. C'est de la fête d'investiture que je vous parle.
19. Venez vers moi pour les réjouissances de la fête.
20. C'est de l'investiture de Wayewo que je vous parle.
21. Je ne me suis pas trompé.
22. C'est bien de vous qu'il s'agit.
23. Rassemblez-vous tous et venez vers moi, vite, vite, vite !
24. C'est de la fête que je vous parle.
25. Rassemblez-vous et venez vite vers moi !
26. C'est de l'investiture que je vous parle.
27. Rassemblez-vous et venez vite vers moi !
28. Vite, venez vite vers moi, venez vers moi !
29. Rassemblez-vous tous.
30. Et venez vers moi pour la fête de l'investiture de Wayewo du lignage Ngora.
31. Rassemblez-vous et venez vite vers moi !
32. Rassemblez-vous et venez vite vers moi !
33. C'est la vérité que je suis en train de vous dire.

34. Vous êtes en train d'hésiter.
35. Vous êtes en train d'hésiter.
36. Je ne me suis pas trompé.
37. Rassemblez-vous et venez vite vers moi !
38. Venez vite vers moi !
39. Venez vite vers moi !
40. C'est de la fête d'investiture de Wayewo du lignage Ngora,
41. C'est de la fête d'investiture de Wayewo du lignage Ngora que je vous parle.
42. Avez-vous compris ?
43. C'est à vous que je m'adresse !
44. Rassemblez-vous et venez vite vers moi !
45. C'est des réjouissances de la fête que je vous parle.
46. C'est de la fête d'investiture de Wayewo du lignage Ngora.
47. C'est de la fête d'investiture de Wayewo du lignage Ngora que je vous parle.
48. Rassemblez-vous et venez vite vers moi, venez vers moi !
49. Je ne me suis pas trompé.
50. Venez vite vers moi !
51. Venez, venez vers moi !
52. C'est de la fête d'investiture de Wayewo du lignage Ngora que je vous parle.
53. Rassemblez-vous et venez vers moi, vite, vite, vite !
54. C'est la dernière fois que je vous parle.

Dans le tableau qui suit, sous chacun des sept premiers énoncés, transcrits en banda-linda, figure :

a) sa traduction mot à mot et son analyse morphologique,
b) sa traduction en français.

1. wá ātāyī.mòlō
Eh/pl. + invités (Formule d'appel)
Eh ! Les invités !

2. īpī sándákà sándákà nà dá[1] mā ná-pà ndá-nà kō-?ē kó
histoire|fête x 2|la/acc. + est + (que) # je/acc. + progr.|dit/au sujet de|le/à|vous/là
C'est de la fête que je vous parle.

3. īpī sándákà sándákà dá mā ná-pà ndá-nà kō-?ē kó
histoire|fête x 2/acc. + est + (que) # je/acc. + progr.|dit/au sujet de|le/à|vous/là
C'est de la fête que je vous parle.

4. á pà dá yē
et + je/dis/que
Et je dis :

5. ànjē zá màdāyà màdāyà[2]
on/acc.+donne/médaille x 2
On donne l'investiture.

6. á pà dá yē
et + je/dis/que
Et je dis :

7. yɛ̄ ngbɔ̀ɔ̀ᵌ tɨ́-ʔɛ̄ gbɔ́lɔ́ gbɔ́lɔ́ gbɔ́lɔ́
// vous/*inj.* + rassemble/vous-mêmes/tous x 3 //
Rassemblez-vous tous !

La comparaison de la version continue avec celle alternée (ci-dessous), où chaque énoncé était suivi aussitôt par son décodage en banda-linda, met en évidence – par-delà les variations quant à la forme et à l'ordre de succession des énoncés – que leur contenu ne change guère.

1. uà ɜ̄mɔ̀kɔ̄nɟì
// Eh|*pl.* + chef de village //
Eh ! les invités !

2. sándákà mɔ̀dɜ̄yɛ̀ mɔ̀dɜ̄yɛ̀ dɔ́ mɜ̄ sɔ́-tɨ́-pà ndɨ́-nɔ̀
// fête|médaille x 2/*acc.* + est + (que) # je/*acc.* + *progr.*|*dit.*| dit|au sujet de|
C'est au sujet de la fête d'investiture que je vous parle. 1e //

3. ávórɨ́ ɔ̄ɪpɛ̄ dɔ́ mɜ̄ sá-pà kɛ̄
// vrai|histoire/ *acc.* + est + (que) # je/*acc.* + *progr.*|dit/là //
C'est la vérité que je vous dis là.

4. sándákà mɔ̀dɜ̄yɛ̀ mɔ̀dɜ̄yɛ̀ nɔ́ wàyɛ̄wɔ̀ nálɔ́ gɔ̄gɔ̄ nɔ́ ɜ̄nɛ́ʃɛ̄ ngòrà ngòrà
fête|médaille x 2|de|Wayewo/dans|village|de|enfants|| Ngora x 2/
C'est de la fête d'investiture de Wayewo du lignage Ngora

dɔ́ mɜ̄ nɨ́-pà ndɨ́-nɔ̀ kɜ̄-ʔɛ̄ kɔ́
/*acc.* + est + (que) # je/*acc.* + *prog.*|dit/au sujet de|le/à|vous/là #
que je vous parle.

5. á pà dɔ́-yɛ̄
et + je/dis/que
Et je dis :

6. yɛ̄ ngbùrù gù gà-nɨ́ (ndɔ́) sándákà (sándákà) mɔ̀dɜ̄yɛ̀ mɔ̀dɜ̄yɛ̀
// vous/*inj.* + rassemble|*succ.* + vient/vers|moi/(pour)/fête x 2|médaille x 2|
Rassemblez-vous et venez vers moi pour la fête d'investiture

nɨ́ wàyɛ̄wɔ̀ málɔ́ gɔ̄gɔ̄ nɨ́ ɜ̄nɛ́ʃɛ̄ ngòrà ngòrà
de|Wayewo/dans|village|de|enfants|| Ngora x 2/
de Wayewo du lignage Ngora

dɔ́ mɜ̄ nɨ́-pà ndɨ́-nɔ̀ kɜ̄-ʔɛ̄
/*acc.* + est + (que) # je/*acc.* + *progr.*|dit/au sujet de|le/à|vous //
dont je vous parle.

7. mɜ̄ fórɨ́ tɨ́-mɜ̄ fófórɔ̀ nɛ̄
// je/*rév.* + trompe|/moi-même/|*réf.* + trompe/pas *N.E.* //
Je ne me suis pas trompé.

8. yɛ̄ jí gbàrá gbàrá nɨ́
// vous/*acc.* + comprend/distinctement x 2/tout + *(interrog.* //
Avez-vous bien tout compris?

9. yɛ̄ ngbùrùᵌ gù gà-mɜ̄ yófffffa
// vous/*inj.* + rassemble|*succ.* + vient/vers|moi/tous #
Rassemblez-vous et venez tous vers moi!

10. ɔ gûgū gɜ̌-mɜ̄
et + vous/*dur.* + vient|*succ.* + vient +/vers|moi //
Venez vers moi !

1. Eh ! les chefs de village !
2. C'est au sujet de la fête d'investiture que je vous parle.
3. C'est la vérité que je vous dis là.

4. C'est de la fête d'investiture de Wayewo du lignage Ngora que je vous parle.
5. Et je dis :
6. Rassemblez-vous et venez vers moi pour la fête d'investiture de Wayewo du lignage Ngora dont je vous parle.
7. Je ne me suis pas trompé.
8. Avez-vous tout bien compris ?
9. Rassemblez-vous et venez tous vers moi !
10. Et venez vers moi !
11. C'est pour la fête d'investiture de Wayewo du lignage Ngora que je vous appelle.
12. Rassemblez-vous et venez vers moi !
13. Rassemblez-vous et venez vers moi !
14. C'est la vérité que je vous dis là.
15. Je ne me suis pas trompé.
16. Rassemblez-vous et venez vite vers moi !
17. Eh ! mes pères !
18. Eh ! mes mères !
19. Rassemblez-vous et venez vers moi !
20. C'est pour la fête d'investiture de Wayewo du lignage Ngora que je vous appelle.
21. Rassemblez-vous et venez vers moi pour les réjouissances !
22. C'est de la fête d'investiture que je parle.
23. Avez-vous compris ?
24. Rassemblez-vous et venez vers moi pour les réjouissances !
25. C'est au sujet de l'investiture de Wayewo que je vous appelle.
26. Rassemblez-vous et venez vers moi !
27. Courez vers moi !
28. Venez vers moi !
29. Vite, vite, vite !
30. Venez vers moi !
31. Tous !
32. C'est la vérité que je vous dis.
33. Avez-vous compris ?
34. Avez-vous compris ?
35. Ne faites pas les idiots !
36. Venez vers moi !
37. C'est des réjouissances de la fête que je vous parle.
38. C'est pour l'investiture de Wayewo que je vous appelle.
39. Venez vers moi !
40. C'est la fête d'investiture.
41. Avez-vous compris ?
42. Rassemblez-vous tous !
43. Rassemblez-vous tous !

44. Venez vers moi en foule !
45. Venez vers moi en foule !
46. C'est pour la fête d'investiture de Wayewo du lignage Ngora que je vous appelle.
47. Ne faites pas les idiots !
48. C'est la dernière fois que je vous parle.

L'observation des deux versions de ce message atteste la coexistence des deux catégories mentionnées plus haut :

– l'une d'ordre général, qui peut apparaître dans tout message, comme : « C'est la vérité que je vous dis là » ;

– l'autre, qui porte une information spécifique : « C'est au sujet de la fête d'investiture que je vous parle ».

On le voit, l'information spécifique est toujours enchâssée dans les formules générales, dont la fonction est d'attirer et de maintenir l'attention des destinataires, et d'indiquer la fin du message. L'ordre d'occurrence des différentes formules – qu'elles soient générales ou spécifiques – n'est pas codifié, mais laissé au choix du tambourinaire. Cela signifie que la syntaxe des messages est libre.

Dans la plupart des cas, formules générales et formules spécifiques, formules d'adresse et de conclusion constituent les éléments nécessaires et suffisants à l'élaboration des messages. Les formules générales servent de cadre à l'information spécifique ; elles ont une fonction démarcative qui seule permet d'isoler le sens propre au message. Celui-ci s'ordonne toujours autour des nominaux, qui portent la charge sémantique la plus grande ; c'est pourquoi leur frappe est systématiquement redoublée.

Décodage

La stratégie du décodage repose sur la capacité du destinataire à reconstruire le sens à partir d'éléments dont une partie seulement lui est connue. Cette connaissance lui est fournie par la situation extralinguistique et par la structure des messages[6].

Situation extralinguistique. Les destinataires localisent la source sonore et situent, dans l'espace, l'origine du message – ce qui exclut d'emblée une grande partie des destinataires potentiels.

6. Cloarec-Heiss (1997), p. 139.

Structure du message. La plus grande part du corps du message est constituée de formules générales, dont le contour mélodique est connu : la « mélodie » fait directement sens pour l'auditeur. Ces formules sont figées *sémantiquement*, mais non *formellement*. Elles se résument à une dizaine, dont chacune admet différentes réalisations au sein d'un même message, ce qui produit un taux considérable de redondance.

Du point de vue cognitif, il est remarquable que l'identification de ces formules ne soit pas entravée par les nombreuses variantes auxquelles elles se prêtent. Comme le note F. Cloarec-Heiss,

> l'identification du paradigme auquel une formule appartient tient au fait que chaque variante comporte une suite *typique minimale* – une *signature* – qui en constitue le *prototype* et dont la seule présence garantit l'interprétation correcte. Ainsi, la formule « Venez vers moi » peut présenter, parmi d'autres, l'une des formes suivantes[7] :

Formule paradigmatique "Venez vers moi"

Le décodage fait appel à deux procédures distinctes, selon que le segment de message concerné est une formule générale ou spécifique. Pour les formules générales, c'est par le contour mélodique que l'auditeur accède directement au sens.

L'inventaire des formules spécifiques montre que la quantité d'information propre à chaque message est très réduite. Au sein d'un même message, on ne compte guère plus de trois – parfois seulement deux – formules *sémantiquement distinctes*. Ces unités sont balisées avant et après par les deux formules générales suivantes : « Et je dis que » et « C'est de cela que je vous parle ». À partir de ces repères, l'auditeur peut localiser le segment sur lequel il doit concentrer son attention.

7. *Ibid.*, p. 142.

À l'inverse des formules générales, les formules *spécifiques* ne sont pas décodables sans le recours à la transposition dans la langue ; leur compréhension est cependant facilitée par leur fort taux de récurrence.

Le processus cognitif de reconstitution du message linguistique met en œuvre deux opérations : la *segmentation* et la *sélection par recombinaison*.

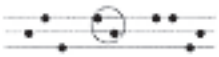
La segmentation. Le rythme, par sa fonction démarcative, garantit l'efficacité du décodage linguistique. Grâce à la segmentation, il assure la localisation et l'identification des segments à interpréter – mots, syntagmes ou parties d'énoncés. C'est le contour mélodique des segments rythmiques qui est décodé globalement par un processus de sélection des possibles et de recombinaison.

La sélection par recombinaison. Les segments étant identifiés, l'auditeur décode globalement un ensemble de suites tonales. Au sein d'un segment donné, l'ambiguïté est d'autant plus réduite que le nombre de combinaisons tonales possibles est grand. Deux facteurs font varier le nombre de combinaisons : l'un, paradigmatique, correspond au nombre de tons pertinents dans la langue ; l'autre, syntagmatique, varie avec le nombre de syllabes qui composent le segment. L'accroissement du nombre de syllabes sur l'axe syntagmatique restreint, pour chaque suite émise, le nombre des combinaisons possibles. En effet, plus grand est le nombre de syllabes au sein d'un segment mélodico-rythmique, plus réduite est la part d'indétermination[8].

Le phénomène de recombinaison est illustré par deux formules quasi homonymes :

Le processus de rétro-identification

mā fórɔ̀ tá-mā fófórɔ̀ nē
Je ne me trompe pas.



yē ngbórɔ̀ ándà ngbóngbórɔ̀ nē
Ne faites pas les idiots !

Cet exemple met en évidence *le caractère non linéaire de la perception des formules et le processus de rétro-identification par recombinaison*. En effet, les deux formules ne diffèrent que par les quatrième et cinquième tons. C'est la présence d'un schème H-M [Haut-Moyen] dans le premier cas, et B-B [Bas-Bas] dans le second qui permet à l'auditeur d'interpréter non seule-

8. *Ibid.*, p. 145.

> ment les deux tons comme deux mots différents, respectivement "moi-même" et "idiot", mais également, *rétroactivement, les deuxième et troisième tons qui sont identiques dans les deux formules*, par le jeu des incompatibilités sémantiques comme "tromper" dans un cas et "faire" dans l'autre[9].

Il y a donc décryptage par *rétro-identification* et *recombinaison* des éléments ou, si l'on préfère, par élimination des possibles. Ce processus suppose un stockage mémoriel jusqu'à un effet de seuil où la compréhension se produit par recombinaison. C'est donc bien l'ensemble du contour mélodique qui est décodé globalement.

Le processus de décodage des formules *spécifiques* est symétrique de celui de l'encodage : l'auditeur reconstruit l'énoncé en redéployant mentalement dans la langue l'équivalent des tons qu'il entend.

Parole et musique en fusion

Chez les Banda-Linda, les tambours à fente `l"eng"a` sont avant tout des instruments de musique dont le rôle consiste à maintenir l'assise métrique de la danse et à en assurer le soubassement rythmique. Utilisés en paire, ils requièrent la présence de deux tambourinaires ; à chacun d'eux est assigné l'un des instruments.

L'instrument le plus aigu a pour fonction de maintenir la synchronisation des pas des danseurs : son rôle se limite à l'exécution ininterrompue, sous forme d'un *ostinato* à peine varié, de la charpente rythmique spécifique à chacune des danses. Le tambour le plus grave est l'apanage du maître-tambourinaire, lequel est généralement un virtuose. À lui aussi est dévolue pour chaque danse une figure rythmique particulière ; mais il lui est permis de s'en affranchir pour se livrer à des improvisations libres de toute contrainte, cependant que le tambour le plus aigu assure l'assise métrique de la danse. Il est d'usage que, dans les séquences improvisées, le maître-tambourinaire alterne improvisations et paroles adressées aux danseurs et aux spectateurs, qui contiennent diverses invectives, plaisanteries – le plus souvent grivoises – et insultes. Ce qui a pour effet de susciter les rires de l'assistance. C'est à ces moments, lorsque des énoncés linguistiques, dont le rythme n'est pas soumis à une contrainte métrique, s'intègrent dans un ordonnancement musical rigoureusement mesuré, que musique et langage fusionnent réellement.

---

9. *Ibid.*, p. 147.

Figure 3 : *Tambourinaires*

Le langage tambouriné banda-linda se présente comme un système fonctionnel, synchroniquement clos mais diachroniquement ouvert ; il a recours à un nombre défini de paradigmes dont chacun peut donner lieu à des réalisations différentes. On peut donc le considérer comme l'équivalent d'un système phonologique. Aussi, les langages tambourinés de la plupart des langues à tons africaines qui en font usage reposent, pour l'essentiel, sur des principes semblables.

RÉFÉRENCES BIBLIOGRAPHIQUES

– Alexandre P. (1969), « Langages tambourinés, une écriture sonore ? », *Semiotica*, 1 (3), p. 272-281.
– Ames D. *et al.* (1971), « Taaken sàmàarii : A drum language of hausa youth », *Africa*, 41 (1), p. 12-31.
– Armstrong R. G. (1955), « Talking instruments in West Africa », *Explorations*, 4, p. 140-153.
– Arom S. (2006), « Language and music in fusion : The drum language of the Banda-Linda (Central African Republic) », *in* E. Camara de Landa et S. Garcia (éd.), *Approaches to African Musics*, Valladolid, Universidad de Valladolid, Centro Buendia, p. 17-34.
– Arom S. et Cloarec-Heiss F. (1976), « Le langage tambouriné des Banda-Linda (R.C.A.) », *in* L. Bouquiaux (éd.), *Théories et méthodes en linguistique africaine*, Paris, Selaf, p. 113-169.
– Béart C. (1953), « Contribution à l'étude des langages tambourinés, sifflés, musicaux », *Notes africaines de l'IFAN*, 57, p. 11-14.
– Beier U. (1954), « The talking drums of the Yoruba », *African Music*, 1, p. 29-31.
– Betz R. (1898), « Die Trommelsprache der Duala », *Mitteilungen von Forschungsreisenden und Gelehrten aus den deutschen Schutzgebieten*, 11, p. 1-86.
– Burssens A. (1939), « Le Luba, langue à intonation et le tambour-signal », *Proceedings of the Third International Congress of Phonetic Sciences*, p. 503-507.
– Carrington J. F. (1949a), *A Comparative Study of Some Central African Gong Languages*, Bruxelles, Falk G. Van Campenhout.
– Carrington J. F. (1949b), *Talking Drums of Africa*, Londres, The Carey Kingsgate Press.
– Clarke R. T. (1934), « The drum language of the Tumba People », *American Journal of Sociology*, 40, p. 34-48.
– Cloarec-Heiss F. (1997), « Langage naturel, langage tambouriné : un encodage économique (Banda-Linda de Centrafrique) », *in* C. Fuchs et S. Robert (éd.), *Diversité des langues et représentations cognitives*, Paris-Gap, Ophyris, p. 136-149.
– Denett R. E. (1909), *At the Back of the Black Man's Mind*, Londres, Macmillan.
– Éboué F. (1933), *Les Peuples de l'Oubangui-Chari. Essai d'ethnographie, de linguistique et d'économie sociale*, Paris, Comité de l'Afrique centrale.
– Éboué F. (1935), « La clef musicale des langages tambourinés et sifflés », *Comité de l'Afrique occidentale française*, 18, p. 353-360.
– Guillemin L. (1948), « Le tambour d'appel des Ewondo », *Études camerounaises (IFAN)*, 21-22.
– Graf W. (1950), « Einige Bemerkungen zur Schlitztrommel-Verständigung in Neuguinea », *Anthropos*, 45, p. 861-868.
– Heepe M. (1920), « Die Trommelsprache der Jaunde in Kamerun », *Zeitschrift für Kolonialsprachen*, 10, p. 43-60.
– Heinitz W. (1943), « Probleme der afrikanischen Trommelsprache », *Beiträge zur Kolonialforschung*, 4, p. 69-100.

– Hermann E. (1943), « Schallsignalsprachen in Melanesien und Afrika », *Nachrichten von* der *Akademie* der *Wissenschaften in Göttingen, Philologisch-Historische Klasse*, 5, p. 127-186.
– Herzog G. (1945), « Drum-signalling in a West African Tribe », *Word*, 1, p. 217-238.
– Hulstaert G. (1935), « De telefon der Nkundo (Belgische Kongo) », *Anthropos*, 30, p. 655-668.
– Junod H. A. (1927), *The Life of a South African Tribe*, Londres, Macmillan.
– Labouret L. (1923), « Langage tambouriné et sifflé », *Bulletin du Comité de l'Afrique occidentale française*, 6, p. 120-158.
– Meinhof C. (1894), « Die Geheimsprachen Afrikas », *Globus*, 66, p. 117-119.
– Nekes H. (1912), « Trommelsprache und Fernruf bei den Jaunde und Duala in Südkamerun », *Mitteilungen des Seminars für Orientalische Sprachen*, 15 (3), p. 69-83.
– Nketia J. H. K. (1963), *Drumming in the Akan Communities of Ghana*, Édimbourg, Thomas Nelson & Sons.
– Peters C. (1891), *New Light on Dark Africa*, Londres, Ward Lock.
– Rattray R. S. (1923), *Ashanti*, Oxford, Clarendon Press.
– Rialland A. (1974), « Les langages instrumentaux sifflés ou criés en Afrique », *La Linguistique*, 10 (2), p. 105-121.
– Rouget G. 1964, « Tons de la langue en Gun (Dahomey) et tons du tambour », *Revue de musicologie*, 10 (1), p. 3-29.
– Schaeffner A. (1951), *Les Kissi, une société noire et ses instruments de musique*, Paris, Hermann.
– Schneider M. (1952), « Zur Trommelsprache der Duala », *Anthropos*, 47, p. 235-243.
– Sebeok T. A. et Umiker-Sebeok D. J. (éd.) (1976), *Speech Surrogates : Drum and Whistle Systems*, La Haye et Paris, Mouton, 2 vol.
– Stern T. (1957), « Drum and whistle languages : An analysis of speech surrogates », *American Anthropologist*, **59**, p. 487-506.
– Thilenius G., Meinhof, C. et W. Heinitz (1916), « Die Trommelsprache in Afrika und in der Südsee », *Vox*, 26, p. 179-208.
– Thomas J. M. C., Bouquiaux L., Cloarec-Heiss F. (éd.) (1976), *Initiation à la phonétique. Phonétique articulaire et phonétique distinctive,* Paris, PUF.
– Umiker D. J. (1974), « Speech surrogates : Drum and whistle systems », *in* T. Sebeok (éd.), *Current Trends in Linguistics. Linguistics and Adjacent Arts and Sciences*, Paris et La Haye, Mouton, vol. 12, p. 497-536.
– Van Avermaet E. (1945), « Les tons en kiluba-samba et le tambour-téléphone », *Aequatoria*, 8 (1), p. 1-12.
– Verbeken A. (1920), « Le tambour-téléphone chez les indigènes de l'Afrique centrale », *Congo*, 1, p. 253-284.
– Westermann D. H. (1907), « Zeichensprache des Ewevolkes in Deutsch-Togo », *Mitteilungen des Seminars für Orientalische Sprachen*, 10 (3), p. 1-14.
– Wilson W. A. A. (1963), « Talking drums in Guinea », *Estudios, Ensaios e Documentos*, 3, p. 201-219.
– Witte P. A. (1910), « Zur Trommelsprache bei den Ewe Leuten », *Anthropos*, 5, p. 50-53.

Transformation et synthèse de la voix parlée et de la voix chantée

par Xavier Rodet[1]

La mission principale de l'Institut de recherche et coordination acoustique/musique (Ircam) est la création musicale et la création artistique en général, ce qui inclut notamment les arts du spectacle comme le théâtre ou le film. Cet institut possède une longue expérience dans l'analyse et la synthèse des sons, et en particulier de la parole. En effet, de nombreux compositeurs contemporains portent un vif intérêt à la voix, chantée mais aussi parlée. Ils considèrent la voix non seulement comme un matériau musical qui peut entrer, d'une façon ou d'une autre, dans leurs compositions, mais aussi pour sa structure, depuis les niveaux acoustiques et phonétiques jusqu'aux niveaux linguistiques les plus élevés.

Dans ce contexte, l'équipe « Analyse-synthèse » des sons de l'Ircam a développé depuis plusieurs années un savoir-faire, des études et des outils, en particulier informatiques, concernant l'analyse, le traitement et la synthèse de la voix et de la parole. Ces moyens sont d'abord utilisés pour la création musicale à l'Ircam. Ils ont été employés, par exemple, pour des pièces récentes de Jean-Baptiste Barrière, Joshua Fineberg, Stefano Gervasoni ou Jonathan Harvey. Mais ces moyens trouvent également des applications dans le multimédia en général. En effet, alors que les images de synthèse ont envahi de nombreux médias, dessins animés, jeux vidéo et films notamment, la voix reste aujourd'hui le parent pauvre en la matière : elle est, la plupart du temps, simplement enregistrée par des acteurs, souvent synchronisée de façon « manuelle » avec le mouvement des personnages et n'utilise presque aucune technique de synthèse, sauf à de rares

1. Avec Grégory Beller, Niels Bogaards, Gilles Degottex, Snorre Farner, Pierre Lanchantin, Nicolas Obin, Axel Roebel, Christophe Veaux et Fernando Villavicencio.

exceptions. Cependant, les méthodes et outils développés par l'Ircam ont permis de créer la voix du castrat dans le film *Farinelli* de Gérard Corbiaud, d'améliorer la prononciation anglaise de Gérard Depardieu pour le film *Vatel* de Roland Joffé, ou de transformer une voix de femme en voix d'homme pour le film *Tirésia* de Bertrand Bonello, et, inversement, une voix d'homme en voix de femme pour le film *Les Amours d'Astrée et de Céladon* d'Éric Rohmer (2007). Bien d'autres applications sont expérimentées pour les jeux vidéos, le dessin animé, les avatars, etc.

Les principaux sujets sur la voix traités à l'Ircam sont la constitution de corpus oraux, l'analyse de ces corpus, la synthèse à partir du texte, la transformation du type et de la nature de la voix, la conversion d'identité de la voix, l'étude de la transformation d'expressivité de la voix, la séparation de la source glottique de l'influence du conduit vocal, et la modélisation de la prosodie dans différents modes de discours. Ces divers travaux ont pour but premier de fournir de nouveaux moyens aux compositeurs et artistes travaillant à, ou avec, l'Ircam. Pour cela, l'institut collabore avec de nombreux centres de recherches et mène des projets de recherche dans les cadres institutionnels français (agence nationale de la recherche, CNRS), européens ou autres. Enfin l'Ircam valorise ses compétences, connaissances et résultats vers d'autres instituts et vers l'industrie.

Dans la première section de ce chapitre, j'exposerai la problématique et les moyens nécessaires à la gestion de corpus de parole enregistrée. En effet, les méthodes scientifiques et techniques d'étude de l'oral s'appuient de plus en plus sur l'analyse statistique de grands corpus enregistrés, qui requièrent donc un outil spécifique. Dans la deuxième section, j'exposerai cet outil : la plate-forme logicielle IrcamCorpusTools, développée par l'équipe « Analyse-synthèse » des sons de l'Ircam pour la gestion de corpus de parole enregistrée. Dans la troisième section seront présentés les principaux logiciels développés par l'équipe pour l'analyse, la synthèse et la transformation de voix (SuperVP et AudioSculpt). Enfin, je conclurai en décrivant quelques applications d'analyse, de synthèse et de transformation de voix, aussi bien dans le domaine de la recherche que dans la création musicale ou le multimédia.

Gestion de corpus de parole enregistrée

LES MÉTHODES À BASE DE CORPUS

Les méthodes à base de corpus sont désormais très largement répandues en traitement de la parole et en traitement du langage pour le développement de modèles théoriques et d'applications technologiques. Que ce soit pour vérifier des heuristiques, découvrir des tendances ou modéliser des données, l'introduction de traitements calculatoires et/ou statistiques fondés sur les données des corpus a multiplié les possibilités et permis des avancées considérables dans les technologies de la parole et du langage. La reconnaissance et la synthèse de parole en sont des exemples pour le traitement automatique de la parole. De même, l'utilisation de corpus annotés (annotations d'ordre phonétique, prosodique, des phénomènes paraverbaux et des disfluences, par exemple du corpus LeaP[2], mais aussi d'ordre syntaxique et discursif) intéresse la recherche en linguistique aussi bien qu'en traitement de la parole. Toutefois, cette complémentarité n'est possible que par la mise en commun des corpus. C'est pourquoi les questions de représentation et de gestion des données des corpus sont centrales. Les corpus oraux sont constitués de deux types principaux de ressources : les signaux temporels et les annotations. Les signaux temporels sont les enregistrements audio, vidéo et/ou physiologiques, ainsi que leurs descriptions (fréquence fondamentale, spectrogramme, etc.). Les annotations sont la transcription textuelle ainsi que toutes les notations ajoutées manuellement ou automatiquement qui permettent de caractériser d'un point de vue linguistique le signal acoustique (transcription phonétique, catégories grammaticales, structure du discours, etc.). Les différents niveaux d'annotations possèdent généralement des relations hiérarchiques et/ou séquentielles et sont synchronisés temporellement sur le signal acoustique. Les outils de gestion des corpus recouvrent tout un ensemble de fonctionnalités, allant de la création et de la synchronisation des ressources aux requêtes (pouvant porter autant sur les annotations que sur les signaux temporels), en passant par le stockage et l'accès aux données. La plupart des systèmes de gestion de corpus existants ont été développés pour des corpus spécifiques et sont difficilement adaptables et extensibles[3].

2. Learning Prosody Project : http://leap.lili.uni-bielefeld.de
3. Oostdijk (2000).

Des efforts ont été faits pour faciliter l'échange de données par la conversion de formats[4] ou pour dégager une représentation formelle pouvant servir d'interface commune entre les divers outils et les données[5]. Cette notion d'interface entre les méthodes et les données est à la base de la plate-forme IrcamCorpusTools présentée dans ce chapitre. Cette plate-forme utilise l'environnement de programmation Matlab afin d'être facilement extensible. Elle permet notamment la synchronisation d'informations provenant de différentes sources (vidéo, audio, symbolique, etc.) ainsi que la gestion de nombreux formats (XML, AVI, WAV, SDIF[6], etc.). Elle est munie d'un langage de requête prenant en compte les relations hiérarchiques multiples, les relations séquentielles et les contraintes acoustiques. Elle permet ainsi l'analyse contextuelle de variables acoustiques (prosodie, enveloppe spectrale) en fonction de variables linguistiques (mots, groupe de sens, syntaxe). Elle est employée pour la synthèse de la parole par sélection d'unités, les analyses prosodiques et phonétiques contextuelles, la modélisation de l'expressivité, et pour exploiter divers corpus de parole en français et en d'autres langues.

SYSTÈMES DE GESTION ET DE CRÉATION DE CORPUS DE PAROLE

Depuis l'essor de la linguistique de corpus[7], de nombreux corpus annotés ont été exploités par le traitement automatique des langues, dont des corpus oraux comme ceux qui sont recensés par LDC[8]. L'automatisation de ce traitement nécessite de traiter une grande quantité de métadonnées linguistiques. Aussi, de nombreux systèmes de gestion de larges corpus sont aujourd'hui disponibles pour cette communauté[9]. Dans le domaine du traitement automatique de la parole, le corpus TIMIT fut le premier corpus annoté à être largement diffusé. Une tendance actuelle est l'utilisation de corpus multimodaux avec l'intégration de données visuelles, ce qui accroît encore la diversité des formats à gérer. Permettre à une communauté de chercheurs de partager et d'exploiter de tels corpus ne pose pas simplement la question de la gestion des formats, mais aussi celles de la représentation des données, du partage des outils de génération, d'accès et d'exploitation, et du langage de requête associé.

4. Gut *et al.* (2004).
5. Bird *et al.* (2000).
6. http://sdif.sourceforge.net/
7. Chafe (1992).
8. *Linguistic Data Consortium* : http://www.ldc.upenn.edu/Catalog/
9. Cunningham *et al.* (2002).

MODÈLES DE REPRÉSENTATION DES DONNÉES

Un modèle de représentation des données doit pouvoir capturer les caractéristiques importantes de celles-ci et les rendre facilement accessibles aux méthodes utilisées pour leur traitement. Ce modèle constitue en fait une hypothèse sous-jacente sur la nature des données et sur leur structure. Il doit donc être aussi général que possible afin de pouvoir représenter différents types de structures phonologiques et de permettre une grande variété de requêtes sur ses structures. Les modèles principalement utilisés pour le traitement automatique du langage sont des structures hiérarchiques comme celles du Penn Treebank[10], qui peuvent être alignées temporellement dans le cas des corpus oraux. Certains systèmes, comme Festival[11] ou EMU[12], vont au-delà de ces modèles en arbre unique et supportent des hiérarchies multiples, c'est-à-dire qu'un élément peut avoir des parents dans deux hiérarchies distinctes sans que ces éléments parents soient reliés entre eux. Ces représentations sont particulièrement adaptées pour les requêtes multiniveaux sur les données du corpus. D'autres approches, telles que celle de Bird et Liberman[13] ou de Müller[14], se concentrent sur des représentations des données qui facilitent la manipulation et le partage des corpus multiniveaux. Il s'agit généralement de représentations « à plat » des données qui donnent uniquement la structure temporelle : les relations hiérarchiques y sont représentées implicitement par la relation d'inclusion entre les marques temporelles. Enfin, Gut expose une méthode et des spécifications minimales permettant de convertir entre elles les différentes représentations des données utilisées par les corpus[15].

PARTAGE DES DONNÉES

Afin de pouvoir partager les corpus, comme dans le cas du projet PFC[16], des efforts de standardisation ont été entrepris à différents niveaux. Un premier niveau de standardisation consiste à établir des conventions sur les formats de fichiers et les métadonnées décrivant leur contenu. Ainsi, le format XML[17] s'est de plus en plus imposé comme le

10. *Penn Treebank* : http://www.cis.upenn.edu/treebank/home.html/
11. Taylor *et al.* (2001).
12. Cassidy et Harrington (2001).
13. Bird et Liberman (2001).
14. Müller (2005).
15. Gut *et al.* (2004).
16. PFC : Phonologie du français contemporain : http://www.projet-pfc.net/ ; *cf.* Durand et Tarrier (2006).
17. XML : eXtensible Markup Language : http://www.w3.org/XML/

format d'échange des annotations. Cette solution permet la compréhension des données par tous les utilisateurs, tout en leur permettant de créer de nouveaux types de données selon leurs besoins. Un second niveau consiste à standardiser le processus de génération des données elles-mêmes. Cela conduit par exemple à des recommandations comme celles de la Text Encoding Initiative[18] pour les annotations des corpus oraux. Certains projets, tel CHILDES pour l'analyse des situations de dialogues[19], proposent à la fois des normes de transcription et les outils conçus pour analyser les fichiers transcrits selon ces normes.

PARTAGE DES OUTILS

Des efforts ont également été entrepris pour créer des outils libres adaptés aux annotations des ressources audio et/ou vidéo des corpus comme Transcriber[20] ou ELAN du projet DOBES[21]. Toujours pour l'annotation, des outils de visualisation et d'analyse acoustique sont disponibles et largement utilisés, comme WaveSurfer[22] ou Praat[23]. Ces logiciels permettent l'analyse, la visualisation/annotation, la transformation et la synthèse de la parole. Mais ils sont limités, soit par un format propriétaire pour les données, soit par le langage de requêtes, soit pour la gestion des données.

LANGAGES DE REQUÊTE

Pour être exploitable par une large communauté d'utilisateurs, un corpus doit être muni d'un langage de requête qui soit à la fois simple et suffisamment expressif pour formuler des requêtes variées[24]. On peut distinguer deux grandes familles de systèmes utilisés pour stocker et rechercher de l'information structurée : les bases de données et les langages de balises de textes comme le XML[25]. Les langages de requête comme XSLT/XPath sont naturellement adaptés à la formulation des contraintes d'ordre hiérarchique, mais la syntaxe des requêtes se complique lorsqu'il s'agit d'exprimer des contraintes séquentielles. Les systèmes fondés sur le XML offrent une « extensibilité » limitée car ils nécessitent une recherche linéaire dans le sys-

18. Text Encoding Initiative : http://www.tei-c.org/
19. MacWhinney (2000).
20. Transcriber : http://trans.sourceforge.net/en/presentation.php ; *cf.* Barras *et al.* (1998).
21. DOBES : documentation sur les langues rares : http://www.mpi.nl/DOBES/
22. WaveSurfer : http://www.speech.kth.se/wavesurfer/ ; *cf.* Sjölander et Beskow (2000).
23. Praat : http://www.fon.hum.uva.nl/praat/ ; *cf.* Boersma (2001).
24. Lai et Bird (2004).
25. Gut *et al.* (2004), Cassidy et Harrington (2001).

tème de fichiers[26]. À l'inverse, les systèmes de bases de données sont capables de stocker de très grandes quantités d'informations et d'effectuer des requêtes relativement rapides sur celles-ci. Cependant, le modèle relationnel étant par nature moins adapté à la représentation des contraintes hiérarchiques et séquentielles que le XML, une requête donnée en XML se traduit de manière beaucoup plus complexe en SQL[27]. Si des langages intermédiaires plus simples comme LQL ont été proposés[28], les requêtes les plus complexes ne sont pas toujours formulables selon cette approche.

EXPLOITATION DES DONNÉES

Une fonctionnalité essentielle des plate-formes de gestion de corpus est la possibilité d'interfacer les données (éventuellement après filtrage par des requêtes) avec des outils de modélisation. Ainsi, alors que certains environnements de développement linguistique permettent de construire, de tester et de gérer des descriptions formalisées[29], d'autres se sont tournés vers les traitements statistiques[30]. L'apprentissage automatique pour les tâches de classification, de régression et d'estimation de densités de probabilités est aujourd'hui largement employé. Qu'elles soient déterministes ou probabilistes, ces méthodes nécessitent des accès directs aux données et à leurs descriptions. C'est pourquoi certains systèmes de gestion de corpus tentent de faciliter la communication entre leurs données et les machines d'apprentissage et d'inférence de règles, comme c'est le cas pour le projet EMU et le projet R[31].

IrcamCorpusTools : une plate-forme complète de gestion de corpus

Comme nous venons de le voir, si certains outils comme Praat apportent des solutions partielles permettant l'exploitation des corpus, peu de systèmes proposent une solution complète allant de la génération des données jusqu'aux requêtes sur celles-ci. Lorsque de tels systèmes existent, ils ont été le plus souvent conçus au départ pour une application

26. Cassidy et Harrington (2001).
27. Structured Query Language (Langage structuré de requête) ; cf. http://www.sql.org/
28. Nakov *et al.* (2005) ; sur LQL, cf. http://biotext.berkeley.edu/lql/
29. Bilhaut et Widlöcher (2006).
30. Cassidy et Harrington (2001).
31. *R Project* : http://www.r-project.org/

spécifique comme la synthèse de parole[32] ou l'observation de pathologies, comme c'est le cas pour le projet CSL (Computerized Speech Lab). Cela implique des limitations intrinsèques sur le type de données, sur leur représentation et donc sur leur capacité à être partagées. Ainsi le chercheur à la frontière des traitements automatiques du langage et de la parole est-il, pour le moment, contraint d'utiliser une batterie d'outils dédiés et fondés sur plusieurs langages de programmation, ce qui l'oblige à effectuer de nombreuses conversions de formats et interdit toute automatisation complète d'un processus.

LA PLATE-FORME IRCAMCORPUSTOOLS

Pour répondre aux besoins spécifiques de la parole, de son traitement et de l'analyse de corpus, la plate-forme IrcamCorpusTools (ICT) a été développée et est utilisée dans une grande variété d'applications. Elle s'inscrit à l'intersection de deux domaines de recherches complémentaires : la recherche linguistique et le développement de technologies vocales. Nous la présentons dans cette section en commençant par une vue générale du système et de son architecture. Puis nous présentons deux spécificités de la plate-forme : son langage de requête, qui prend simultanément en compte des contraintes d'ordre linguistique et des contraintes sur les signaux ; et le principe d'autodescription des données et des outils, qui permet de répondre à certaines des problématiques concernant les systèmes de gestion et de création de corpus de parole.

ARCHITECTURE DE LA PLATE-FORME

Afin de répondre à différentes demandes de recherche et de développement industriel, l'architecture d'origine[33] s'est naturellement orientée vers une solution extensible, modulaire et partagée par plusieurs utilisateurs et développeurs[34]. Cette mutualisation des outils et des données implique une certaine modularité tout en maintenant des contraintes de standardisation qui assurent la cohérence du système. La solution choisie repose sur le principe d'autodescription des données, et des outils permettant de définir une interface commune entre ces objets. Une vue générale de l'architecture de IrcamCorpusTools (ICT) est offerte par la figure 1. Elle fait apparaître la couche d'interface que nous introduisons entre les données et les outils, et qui est constituée par notre environnement Matlab. Cette architecture à trois niveaux est semblable à celle proposée

32. Taylor *et al.* (2001).
33. Beller et Marty (2006).
34. Veaux *et al.* (2008).

pour le système ATLAS[35]. Elle permet à différentes applications externes ou internes de manipuler et d'échanger entre elles des informations sur les données du corpus. Les différents éléments composant IrcamCorpusTools sont des instances (objets) de classes qui forment le cœur de la plate-forme. Ces classes sont représentées dans la figure 2 et nous les présentons maintenant en détail.

DESCRIPTEURS

L'activité de parole est intrinsèquement multimodale. La coexistence du texte, de la voix et de gestes (faciaux, articulatoires, etc.) génère une forte hétérogénéité des données relatives à la parole. Le système doit être capable de gérer ces données de différentes natures. Voici les types de données gérées par IrcamCorpusTools.

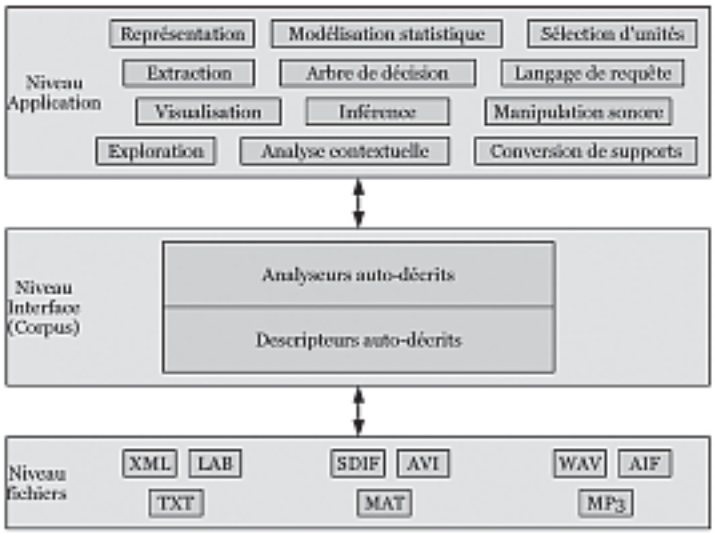


Figure 1 : *Vue d'ensemble de la plate-forme IrcamCorpusTools.*

Informations de type signal

Les signaux correspondent soit aux enregistrements provenant d'un microphone ou d'autres instruments de mesure (EGG, fMRI, ultrasons ou autres), soit à des résultats d'analyse de ces enregistrements. Ils peuvent être unidimensionnels ou multidimensionnels. Parmi les signaux les plus courants figurent ceux qui sont relatifs à la prosodie, comme la fréquence fondamentale f0, l'énergie, le débit de parole, le degré d'articula-

35. Bird et Liberman (2001).

tion mesuré à partir des formants (fréquence, amplitude, largeur de bande), et la qualité vocale (coefficient de relaxation, modèle LF, mesure du voisement), mais aussi ceux qui sont relatifs à l'enveloppe spectrale, donnés par différents estimateurs (FFT, MFCC, TrueEnvelope, LPC) et représentables sous la forme de coefficients autorégressifs (AR), de paires de lignes spectrales (LSF), de pôles, ou d'aires de sections du conduit vocal (LAR). Enfin, cette liste non exhaustive peut être augmentée de signaux issus d'autres modalités comme c'est le cas par exemple pour la mesure de l'aire glottique par caméra ultrarapide (voir, dans les exemples d'application d'analyse acoustique, l'étude de la qualité vocale).

Informations de type métadonnée

Ces informations peuvent, par exemple, servir à spécifier un contexte d'enregistrement (lieu, date, locuteur, consigne donnée, expressivité, genre de discours, etc.). Elles comprennent les transcriptions textuelles *a priori* (parole lue) ou *a posteriori* (parole spontanée). Elles permettent de définir n'importe quelle information sous la forme de mots/symboles ou de séquence de mots.

Informations de type annotation

Ces informations sont de nature textuelle et possèdent, de plus, un temps de début et un temps de fin, permettant d'attribuer une information de type linguistique à une portion de signal. Cette sorte de donnée est cruciale pour une plate-forme de gestion de corpus de parole, puisqu'elle permet de faire le lien entre les signaux et les catégories linguistiques, entre la physique (flux de parole continu) et le symbolique (unités de sens discrètes). Elles constituent souvent des dictionnaires clos, comme c'est le cas pour les phonèmes d'une langue ou pour d'autres étiquettes phonologiques (onset, nucleus, coda, etc.). Parmi ces informations, les segmentations phonétiques sont les plus courantes. Les annotations syntaxiques, de phénomènes prosodiques ou de mots, sont autant d'étiquettes qui peuvent être placées manuellement et/ou automatiquement. Elles définissent alors des segments, aussi appelés *unités*, dont la durée est variable : senone, semiphone, phone, diphone, triphone, syllabe, groupe accentuel, mot, groupe prosodique, phrase, paragraphe, discours, etc.

Informations de type statistique

Sur l'horizon temporel de chacune des unités, les signaux continus peuvent être modélisés par des valeurs statistiques. Ces valeurs, décrivant le comportement d'un signal sur cette unité, sont appelées *valeurs caractéristiques* : moyenne arithmétique et géométrique, variance, intervalle de

variation, maximum, minimum, moment d'ordre N, valeur médiane, centre de gravité, pente, courbure, etc.

UNITÉS

Les unités sont les objets permettant de relier les données entre elles. Elles sont définies pour chaque niveau d'annotation et regroupent les données symboliques ou acoustiques sur la base de la segmentation temporelle associée à ce niveau d'annotation. Les unités sont reliées entre elles par des relations de type séquentiel et/ou hiérarchique. Les relations hiérarchiques sont représentées sous la forme d'arbres (« phrase->mots->syllabes->phones », par exemple) dont les nœuds correspondent chacun à une unité. Afin de représenter des relations hiérarchiques multiples, une liste d'arbres est utilisée à la manière de Festival[36]. Les unités du niveau « phone » sont par exemple dans une relation de parenté avec celles du niveau « syllabe » et avec celles du niveau « mot » ; en revanche, les syllabes et les mots n'ont pas de relation de parenté entre eux (parce que, à cause des liaisons, certaines syllabes ne peuvent être liées de façon unique). Ces arbres permettent de propager les marques temporelles au sein d'une hiérarchie d'unités à partir d'un seul niveau d'annotation synchronisé avec le signal de parole (typiquement le niveau d'annotation issu de la segmentation phonétique). Inversement, à partir d'annotations indépendamment alignées, on peut construire les différentes hiérarchies entre unités, en se basant sur l'intersection des marques temporelles. Cela permet notamment de maintenir la cohérence des diverses données relatives aux unités, tout en autorisant des interventions manuelles à tous les niveaux. À l'inverse des relations hiérarchiques, les relations séquentielles entre unités ne sont définies qu'au sein d'un même niveau d'annotation.

FICHIERS

Nous avons choisi de stocker les différents descripteurs indépendamment les uns des autres afin de faciliter la mise à jour et l'échange des données du corpus[37]. Ces fichiers reposent sur plusieurs supports dont les formats les plus répandus sont :

– LAB, XML, ASCII, TextGRID, pour les données de type métadonnée et annotation ;

– SDIF, AVI, WAV, AIFF, AU, MP3, MIDI, pour les données de type signal ;

36. Taylor *et al.* (2001).
37. Müller (2005).

– MAT (Matlab), pour les données de type relation et statistique.

En revanche, les unités et leurs relations sont stockées dans un fichier unique. Une fonction permet de reconstruire les unités et leurs relations lorsqu'un descripteur (symbolique ou acoustique) a été modifié.

ANALYSEURS

Les analyseurs regroupent toutes les méthodes de génération ou de conversion des données. On peut les enchaîner si nécessaire, par exemple si on veut obtenir la moyenne de la fréquence fondamentale sur le groupe prosodique avoisinant une syllabe[38]. Certaines de ces méthodes sont *internes* au logiciel, d'autres utilisent des logiciels externes qui peuvent être exécutés par appel depuis IrcamCorpusTools. Grâce à l'interface du système de fichiers, les données engendrées par un tel logiciel sont automatiquement rendues accessibles au sein de notre environnement. D'un point de vue utilisateur, le caractère interne/externe ne fait aucune différence. Dans l'exemple cité précédemment, l'utilisateur peut remplacer un estimateur interne de la fréquence fondamentale (à titre d'exemple, par celui de Praat, de WaveSurfer ou de SuperVP[39]) sans avoir à changer d'environnement.

CORPUS

Un corpus peut être représenté comme un ensemble d'énoncés. Chacun de ces énoncés contient un ensemble d'analyses. Chacune de ces analyses comporte un ou plusieurs descripteurs. Par exemple, l'analyse « audio » comporte le descripteur « forme d'onde » qui n'est autre que le signal acoustique de la phrase enregistrée. Ces analyses sont donc synchronisées au niveau de la phrase dans un corpus. Mais une synchronisation plus fine existe aussi grâce à l'ajout d'unités décrites par l'analyse « segmentation ». Les objets « corpus » sont des interfaces avec le système de fichiers. Lorsqu'un analyseur est appliqué à un corpus, celui-ci fait appel à des fichiers d'entrée et de sortie. Il stocke ainsi toute création d'un fichier, qu'il relie à une configuration particulière de l'analyseur et des paramètres qui l'ont généré, et y attache les objets descripteurs. L'objet « corpus » est lui-même stocké dans un fichier XML, à la racine du système de fichier, ce qui permet à plusieurs personnes d'ajouter ou de supprimer des données dans un corpus sans que cela entraîne de conflit. En effet, l'objet « Corpus » conserve au fur et à mesure l'historique

38. Voir l'exemple donné *infra*, p. **NNN**.
39. Bogaards *et al.* (2004).

des opérations effectuées sur un corpus et lui confère donc un accès multi-utilisateur.

LANGAGE DE REQUÊTE

Certains outils de requête XML (Xpath, Xquery, NXT search) présentent une syntaxe complexe. Dans IrcamCorpusTools, nous privilégions l'expressivité du langage de requête. Une requête élémentaire est ainsi constituée :

1) du niveau dans lequel on effectue la recherche d'unité ;
2) d'une relation séquentielle par rapport à l'unité recherchée ;
3) d'une relation hiérarchique par rapport à l'unité recherchée ;
4) d'une condition à tester sur les données numériques associées aux unités.

Ces requêtes sont rapides car elles ne s'appliquent qu'aux données préalablement stockées en mémoire vive. De plus, elles peuvent être composées à volonté, afin de faire des recherches complexes entre les multiples niveaux d'unités.

PRINCIPE D'AUTODESCRIPTION

L'expressivité du langage de requêtes provient de la possibilité de mélanger des contraintes sur des données de types différents. Cela est rendu possible par le principe d'autodescription sur lequel repose IrcamCorpusTools. Chaque instance d'une classe (corpus, fichier, analyseur ou descripteur) est accompagnée de métadonnées décrivant son type, sa provenance, comment y accéder et comment la représenter. Cela permet une compréhension et une exploitation immédiates de tous les objets par tous les utilisateurs, mais aussi par le système lui-même. À l'instar du caractère interne/externe des analyseurs, l'hétérogénéité des données est invisible pour l'utilisateur, qui ne possède qu'un seul lexique restreint de commandes, avec lesquelles il peut rapidement se familiariser. Aucune donnée ne se « perd », car l'objet « corpus » garde une trace des différentes opérations réalisées sur lui et donc, des différentes analyses ayant engendré ses données. Cela permet notamment de conserver un historique de l'accès aux données. En effet, on peut toujours accéder à d'anciennes informations, même si la méthode d'accès à celles-ci a changé entre-temps. Enfin, n'importe quel utilisateur peut comprendre les données des autres et utiliser leurs analyseurs sur ses corpus, sans avoir à changer d'environnement. En résumé, le principe d'autodescription d'IrcamCorpusTools lui assure la pérennité des données, lui fournit un langage de requête expressif et lui confère la possibilité de mutualiser les données, les fichiers, les corpus et, surtout, les analyseurs. La mise en commun des outils est un

facteur déterminant pour le développement des recherches en TAL et en TAP, car leur complexité s'accroît rapidement.

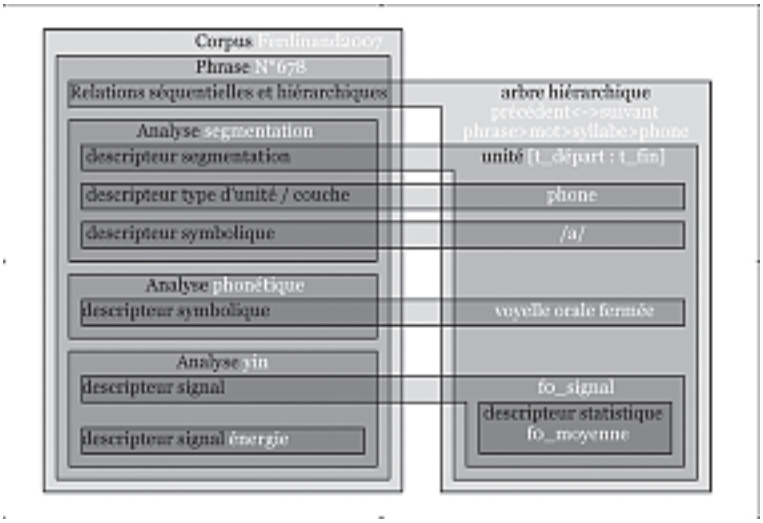


Figure 2 : *Exemple d'utilisation : une instance particulière.*

CRÉATION ET ANALYSE DE CORPUS

Conception de corpus

Si l'approche qui consiste à soumettre des hypothèses théoriques à l'épreuve de grands corpus oraux est de plus en plus répandue, c'est parce que la taille de ces corpus leur permet d'être considérés comme exhaustifs (sous certaines hypothèses)[40]. Pour le reste, l'approche traditionnelle consiste à créer des corpus en vue de valider certaines hypothèses théoriques prises en compte lors de la conception de ces corpus. Il en va de même pour la conception d'un synthétiseur de parole qui débute par une phase de conception de corpus, afin de minimiser les traitements ultérieurs. Nous avons élaboré un ensemble d'outils dans le dessein de sélectionner des ensembles de phrases respectant certaines contraintes linguistiques. Ces ensembles sont extraits de larges corpus textuels, par exemple Corpatext[41] de plus de 37 millions de mots. L'extraction est motivée par différentes recherches de couvertures maximales sous contraintes. Pour la synthèse TTS, l'ensemble des phrases retenues doit présenter le meilleur compromis entre une taille minimale et une couverture maximale des phonèmes par rapport à des contextes donnés (phonétique, lexical, syn-

40. Habert (2000).
41. Corpatext : http://www.lexique.org/public/corpatext.php

taxique, etc.). Ici, une couverture maximale pourra être interprétée comme ayant au moins un candidat pour chaque contexte ou bien comme ayant une distribution statistique des candidats reflétant une distribution naturelle (comme la distribution sur tout le corpus textuel, par exemple).

Décodage acoustico-phonétique

Pour permettre des études en linguistique de corpus, il est nécessaire qu'un certain nombre d'étapes soient automatisées. Dans le cadre de la synthèse de parole, de la modélisation prosodique et expressive, le décodage acoustico-phonétique est une étape essentielle en amont d'une chaîne de traitements linguistiques permettant de représenter la structure de la parole. Cette étape permet la segmentation d'un signal de parole en ses unités linguistiques minimales. Celles-ci sont ensuite regroupées en des unités linguistiques de dimensions supérieures (syllabes, groupes accentuels, groupes prosodiques). Une fois la conception du corpus réalisée (parole de laboratoire ou parole spontanée par exemple), les enregistrements sont automatiquement segmentés en phones à l'aide de l'analyseur IrcamAlign[42]. Ce dernier prend en entrée le signal de parole, sa transcription textuelle, ainsi qu'un dictionnaire constitué de modèles statistiques paramétriques (Hidden Markov Models, HMM[43]) de chacun des phones en contexte, appris sur le corpus multilocuteur BREF80[44]. À partir de la transcription textuelle et du dictionnaire, un modèle statistique de la phrase est constitué en prenant en compte les différentes variantes de prononciation. La meilleure séquence de phones peut alors être sélectionnée, puis alignée sur le signal de parole. Finalement, afin de détecter les erreurs éventuelles et de simplifier une phase de correction manuelle, un indice de confiance est associé automatiquement à chacun des phones segmentés.

Création des unités

Le système IrcamCorpusTools offre une grande modularité dans l'étape de spécification des unités, ce qui permet d'envisager un large champ d'application possible en étude de la parole. Il est ainsi possible de définir arbitrairement une structure de parole (tant au niveau des unités utilisées que de leurs attributs associés) à partir de considérations particulières au domaine d'étude considéré. Cette propriété se révèle nécessaire

42. Lanchantin *et al.* (2008).
43. Rabiner (1989).
44. Lamel *et al.* (1991).

dans l'étude des phénomènes rattachés à la parole, que ce soit pour définir des structures de la parole à partir de théories phonologiques spécifiques au sein d'une langue, pour représenter la variabilité des structures observées entre les langues, ou bien pour définir des niveaux d'analyses supplémentaires pour des domaines d'études spécifiques (acquisition du langage, pathologie, etc.). À partir de la segmentation phonétique présentée précédemment, la représentation de la structure phonologique segmentale et suprasegmentale de la parole dans IrcamCorpusTools est décrite avec les éléments suivants : le phonème et ses attributs phonologiques, la structure syllabique (onset/rhyme, nucleus/coda), la syllabe et ses attributs phonologiques, le groupe accentuel, le groupe prosodique et le discours.

Analyse, synthèse et transformation de voix dans SuperVP et AudioSculpt

L'analyse, la synthèse et la transformation de voix requièrent finalement des logiciels dits de traitement du signal adaptés aux signaux sonores, musicaux, et spécialement au signal de la voix. L'équipe « Analyse-synthèse » des sons de l'Ircam développe en particulier les logiciels SuperVP et AudioSculpt pour la musique mais aussi pour la voix. Le logiciel SuperVP est un « moteur » de traitement du signal qui peut être utilisé seul ou appelé par d'autres logiciels comme AudioSculpt, Xspect ou Max/MSP.

LE LOGICIEL SUPERVP

Le logiciel SuperVP[45] (pour Super Vocoder de Phase[46]) est une implémentation très évoluée d'un vocodeur de phase généralisé, qui est utilisable soit en mode autonome (en ligne de commande) soit comme bibliothèque. Il est appelé comme moteur de calcul pour les traitements et les analyses dans les logiciels AudioSculpt, Xspect ou Max/MSP, et dans un grand nombre de projets de recherche et de production musicale. Il a été perfectionné sur de nombreux points et, en particulier, pour améliorer la qualité des transformations de la parole. Il est aussi utilisé ou distribué sous licence par plusieurs sociétés commerciales.

45. Bogaards *et al.* (2004), Roebel (2003).
46. http://forumnet.Ircam.fr/

SuperVP permet de faire une grande quantité d'analyses, de traitements et de synthèses. Il inclut plusieurs méthodes d'estimation d'enveloppe spectrale (AutoRegressive ou LPC, Cepstre, Cepstre Discret, TrueEnvelope[47], formants). Il offre diverses méthodes d'analyse telles que celle des partiels harmoniques et inharmoniques ou de la fréquence fondamentale (F0), plusieurs techniques de détection des transitoires (il permet de les traiter spécifiquement), d'estimation sinusoïdes/bruit et voisé/non voisé. SuperVP permet de faire toutes sortes de filtrages, par bandes, par surfaces, par enveloppe spectrale, par phase, etc. Divers traitements sont possibles comme le « déplacement de fréquences » (*frequency shift*), la suppression de bruits aléatoires et tonaux (*denoiser*), la transposition, la dilatation/contraction, etc.

SuperVP offre des méthodes de synthèse, telles que la synthèse source-filtre et la synthèse croisée généralisée, qui sont particulièrement adaptées à la transformation de voix, par exemple par application de l'enveloppe spectrale d'amplitude et/ou de phase d'une voix sur une autre. En particulier, SuperVP inclut les traitements les mieux adaptés pour la parole, notamment des filtrages temporels (LPC) et fréquentiels (par FFT), la transposition et la modification de durée avec conservation de l'enveloppe spectrale et de la forme d'onde (méthode dite *shape invariant*[48]). Il permet également d'utiliser le résultat d'une analyse de voisement non seulement en temps mais même en fréquence dans les transformations. Tous ces traitements spéciaux sont indispensables pour préserver la qualité de la voix. Les entrées et sorties d'analyse de SuperVP sont dans le format standard SDIF qui facilite leur gestion, leur maintenance et leur compatibilité avec les autres logiciels.

LE LOGICIEL AUDIOSCULPT

AudioSculpt[49] est une application graphique interactive d'analyse et de traitement musical des signaux sonores qui utilise essentiellement SuperVP comme moteur de traitement du signal. Ce logiciel permet l'étude très détaillée d'un son, de son spectre, de sa forme d'onde, de sa fréquence fondamentale et de son contenu en « partiels ». Toutes les analyses (comme le filtrage, la dilatation/contraction du temps et la suppression du bruit) peuvent être éditées, stockées (notamment en format SDIF) et employées pour guider le traitement dans l'application, ou peuvent servir d'entrée spectrale pour des environnements de composition.

47. Roebel et Rodet (2005).
48. Quatieri et McAulay (1992).
49. http://forumnet.ircam.fr ; *cf.* Bogaards *et al.* (2004).

Au cœur du logiciel se trouve une représentation très flexible du « sonagramme » du son suivant les diverses méthodes d'estimation spectrale de SuperVP. Une fois que le sonagramme a été obtenu, des filtres ou des traitements peuvent être dessinés directement dessus. AudioSculpt permet ainsi de « sculpter » un son de manière visuelle.

AudioSculpt comporte une classe unique de filtres spectraux de gain, appelés les filtres de « surface », qui permettent l'amplification ou l'atténuation d'une région arbitraire du plan temps-fréquence. D'autres traitements, qui peuvent être appliqués à une section ou à la totalité du son, incluent : passe-bande/rejection, transposition, dilatation, *écoute interactive en vitesse lente sans transposition* (et même « gel » spectral), suppression de bruit, écrêtage, filtrage *break-point* et par formants, etc.

Tous les traitements s'accompagnent d'un objet graphique sur le sonagramme aussi bien que d'un objet dans le « séquenceur de traitements », qui est un des outils les plus intéressants d'AudioSculpt. Ces objets peuvent être déplacés, modifiés, copiés, collés ou répliqués en temps et en fréquence. Le concept de séquenceur de traitements est très utile, et il permet aussi de se concentrer sur certains traitements et de les essayer seuls en débranchant momentanément les autres. Le traitement final complet sera ainsi réglé de façon optimale.

Outre la musique, il est aussi largement utilisé dans d'autres applications comme le *design* sonore, la postproduction cinéma et vidéo, la recherche et le développement scientifique ou la musicologie[50].

Exemples d'applications d'analyse, de transformation et de synthèse de voix

ANALYSE ACOUSTIQUE : ÉTUDE DE LA QUALITÉ VOCALE

En parallèle et indépendamment de la structure spécifiée du corpus, il est possible d'associer des analyses acoustiques produites par des analyseurs. Dans le cas de l'analyse et la synthèse de l'expressivité, l'une des caractéristiques prosodiques importantes est la qualité vocale. C'est pourquoi nous cherchons, par inversion du conduit vocal, à estimer une source d'excitation du conduit vocal. De nombreuses méthodes existent déjà[51], mais la difficulté réside dans leurs validations. En effet, il n'existe

50. La documentation est accessible à l'adresse suivante : http://support.Ircam.fr/forum-ol-doc/audiosculpt.

aucun moyen de mesurer *in vivo* cette source d'excitation. Cependant, à l'aide de la vidéo-endoscopie à haute vitesse, il est possible de filmer la glotte à un taux de 4 000 images/secondes. Sur cette suite d'images, la variation temporelle de l'aire de la glotte est calculée automatiquement[52]. La courbe de variation de l'aire permet alors l'estimation d'un débit glottique[53], qui est donc comparé à la source d'excitation estimée qui lui est fortement corrélée[54]. Dans ce contexte, il est donc indispensable de pouvoir manipuler des informations tant visuelles qu'acoustiques s'exprimant dans une ou plusieurs dimensions. La plate-forme IrcamCorpusTools permet une visualisation synchronisée de ces différentes informations et facilite ainsi l'interprétation des données multimodales. L'ensemble des paramètres glottiques estimés est alors calculé sur le corpus par un analyseur acoustique, et accessible par le langage de requête. Les étapes de construction d'un corpus et la spécification de ses différents niveaux d'analyse pertinents d'un point de vue linguistique, et l'estimation de divers signaux relatifs à la parole, permettent un grand nombre d'études linguistiques et/ou statistiques sur des régularités au sein de ces corpus.

CARACTÉRISATION ACOUSTIQUE DU PHÉNOMÈNE DE PROÉMINENCE

La proéminence est un phénomène prosodique majeur pour l'analyse et la modélisation de la prosodie[55]. L'analyse de ce phénomène peut se conduire en trois temps : dans une première étape, des outils statistiques sont utilisés pour permettre l'émergence des corrélats acoustiques de la proéminence et permettre leur détection automatique ; dans une deuxième étape, les proéminences détectées automatiquement sont utilisées pour faire émerger un ensemble de formes de la proéminence ; enfin, ces formes prosodiques sont étudiées par des linguistes pour réaliser une correspondance forme/fonction. Nous nous arrêterons ici seulement sur la première étape de cette étude : l'émergence automatique des corrélats acoustiques de la proéminence et sa détection automatique. Une modélisation statistique des corrélats acoustiques de la proéminence est rendue possible grâce à la complémentarité des possibilités d'IrcamCorpusTools et des méthodes statistiques implémentées en Matlab[56].

À titre d'exemple, voici les étapes menant à une caractérisation acoustique de la proéminence reposant sur la hauteur f0 moyenne des

51. Vincent *et al.* (2005), Henrich (2001).
52. Degottex *et al.* (2008a).
53. Maeda (1982).
54. Degottex *et al.* (2008b).
55. Rosenberg et Hirschberg (2007), Avanzi *et al.* (2008).
56. Obin *et al.* (2008c).

unités « syllabe », relativisée par rapport à la hauteur f0 moyenne des syllabes adjacentes ou par rapport à la hauteur f0 moyenne du « groupe prosodique » parent. Ainsi, nous pouvons accéder aux unités syllabes de la phrase 678 du corpus Ferdinand2007, ainsi qu'à leurs f0 moyennes comme précédemment[57] :

```
» syls = loadfeatures(corpus, 678, _syllabe_) ;
» f0_mean_syl = mean(segment(f0, syls)) ;
```

Grâce aux relations entre unités, on accède au groupe prosodique parent de chaque syllabe et à leurs f0 moyennes respectives :

```
» prosos = getparent(corpus, syls, _prosodic_);
» f0_mean_proso = mean(segment(f0, prosos));
```

Enfin, des valeurs relatives sont déterminées pour chaque syllabe, en divisant leurs hauteurs moyennes par la hauteur moyenne de leur groupe prosodique parent respectif (la fonction gv() extrait les valeurs des objets) :

```
» f0_mean_syl_rel = gv(f0_mean_syl) / gv(f0_mean_proso);
```

Cette étape montre comment les relations hiérarchiques entre les différentes unités de la phrase permettent à une unité d'un niveau donné d'hériter des données associées de ses « parents » ou d'agréger les données associées à ses « enfants ». On peut utiliser ces procédés d'analyse pour tout type de signaux et sur un corpus entier (ou sur la réunion de plusieurs corpus). Dans l'étude présentée, nous avons utilisé ces procédés pour générer une description du signal de parole pour toutes les syllabes.

Cette description comprend :

a) plusieurs corrélats acoustiques (fréquence fondamentale, durée, intensité, information spectrale et information spectrale perceptive) ;

b) plusieurs mesures sur ces corrélats au niveau de la syllabe (valeur moyenne, valeur maximale, etc.) ;

c) plusieurs empans de relativisation mis en œuvre pour relativiser la valeur observée sur une syllabe donnée, en fonction des valeurs observées dans son contexte (aucun contexte, syllabes adjacentes, groupe accentuel précédent, groupe prosodique précédent). Vis-à-vis d'une annotation de la proéminence manuelle ou découlant automatiquement de cette description, il est alors possible de filtrer les syllabes présentant une proéminence grâce à la requête :

57. Voir *supra*, p. NNN.

» syl_pro = getunits(corpus, _syllabe_, {_prominence_,_is_,_P_}) ;

On récupère, de la même façon, les syllabes ne présentant pas de proéminence. La figure 3 montre cette distinction binaire dans un espace réduit de la description dont les axes sont la durée et la f0 moyenne relativisée.

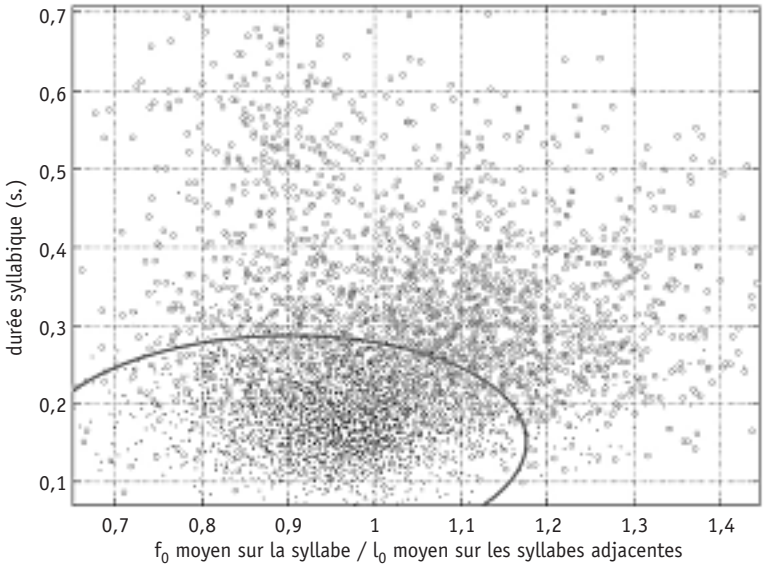


Figure 3 : *Distribution des syllabes proéminentes (cercles clairs) et non proéminentes (points sombres) dans un espace prosodique choisi. Le trait plein gras représente leur courbe de séparation quadratique.*

Les corrélats acoustiques de la proéminence sont appris grâce à des outils statistiques (machines d'apprentissage) de Matlab, dont le but est de déterminer l'ensemble ordonné des corrélats acoustiques observés sur la syllabe, qui permet la meilleure discrimination entre les syllabes proéminentes et les syllabes non proéminentes. La puissance du langage de requête d'IrcamCorpusTools a ainsi permis la caractérisation et la modélisation de la proéminence sur un corpus de voix parlée monolocuteur[58]. Grâce à la facilité d'intégration d'analyseurs externes, cette méthode a été confrontée à d'autres sur des corpus de parole spontanée[59]. Enfin, elle a permis la mise en place d'une méthode de caractérisation automatique des

58. Obin *et al.* (2008c).
59. Obin *et al.* (2008a).

genres de discours[60] : en plus de la simple lecture, le genre de discours d'interview, de discours politique, de dialogue, d'aide vocale, etc.

TRANSFORMATION DE VOIX

Sur la base des outils présentés dans les paragraphes précédents, diverses applications de traitement de voix parlée et de voix chantée ont été développées par l'Ircam. Nous exposerons ci-après la transformation de type et de nature (par exemple, transformation entre voix d'homme et voix de femme, ou chuchotement, voix rauque, etc.), la transformation de l'expressivité (joie, colère, doute, ironie, etc.) et la conversion d'identité (transformer la voix d'un locuteur pour qu'elle paraisse avoir été émise par un autre locuteur).

Transformation de type et de nature

Deux des caractéristiques les plus importantes d'une parole enregistrée sont la hauteur (ou *pitch*, assimilable à la fréquence fondamentale) et le timbre. La hauteur est liée à la longueur et à l'épaisseur des plis vocaux qui distinguent notamment les voix masculines graves des voix aiguës de femme et d'enfant. Mais une voix de femme relativement grave et une voix masculine légère, dont les hauteurs seraient semblables, diffèrent en général par leur *timbre*, ou leur *couleur*, par exemple une voix plus claire ou plus sombre, plus ou moins nasale, tendue, enrouée, avec plus ou moins de souffle, etc. Après avoir étudié les caractéristiques de différents types et natures de voix, nous avons développé la bibliothèque logicielle IrcamVoiceTrans[61]. IrcamVoiceTrans est fondée sur la bibliothèque SuperVP et permet de changer le type de voix entre les hommes, femmes, enfants, personnes âgées, etc., en temps réel. D'autres modifications de la voix originale peuvent donner l'impression d'une voix de fumeur ou d'une voix rauque, chuchotée, etc., ou même excitée ou ennuyée[62]. Un exemple typique d'application est une vidéo où les voix des quatre personnages (homme, vieille femme, adolescent et enfant) sont dérivées, par transformation d'âge et de genre, de la voix d'une seule locutrice (jeune femme)[63]. Cette technologie peut aussi être employée pour modifier la voix produite par notre système de synthèse à partir du texte IrcamTTS (voir ci-dessous). Les applications comprennent la musique, les installations, le théâtre, le spectacle vivant en général, et le multimédia, jeu

60. Obin *et al.* (2008b).
61. Farner *et al.* (2009).
62. Farner *et al.* (2008).
63. http://recherche.ircam.fr/equipes/analyse-synthese/demos.html

vidéo, animation, vidéo et films. Par exemple, pour le film *Tirésia* de B. Bonello nous avons transformé la voix de femme de l'héroïne en voix d'homme, et pour le film *Les Amours d'Astrée et de Céladon* d'Éric Rohmer, la voix d'homme de Céladon en voix de femme. Pour le film *Vatel* de Roland Joffé, c'est la prononciation anglaise de l'acteur Gérard Depardieu qui a été corrigée, par modification de l'accent tonique ou de certaines voyelles et consonnes. Certaines de ces transformations sont déjà opérationnelles en temps réel aussi dans l'environnement Max/MSP (objets SuperVP~) et dans l'application SuperVP-TRaX, distribués dans le Forum des logiciels de l'Ircam[64].

Modélisation et transformation contextuelles de l'expressivité

On entend par *expressivité* les aspects, essentiels dans la communication humaine, transportés par la parole comme l'émotion (joie, colère, peur, etc., chacune d'elles pouvant être « introvertie » ou « extravertie »), les affects, les intentions (ironie, doute, etc.), aussi bien que des notions comme surprise positive ou négative, discrétion, excitation ou confusion. Alors que les logiciels de synthèse à partir du texte (*text-to-speech* ou TTS[65]) ne fournissent que des phrases ayant une expressivité neutre, souvent un style de lecture, la plupart des applications, en particulier artistiques, nécessitent au contraire que la voix soit expressive. De la même façon, pour des applications en jeux vidéo, animation et film, il est indispensable de pouvoir modifier la façon de prononcer tel ou tel segment ou phrase, ou d'obtenir des effets qui ne correspondent pas à une voix habituelle, comme le fait un acteur. Celui-ci peut typiquement, sur une portion de phrase, modifier sa voix pour obtenir quantité d'effets comme un certain sous-entendu, une voix de « Mickey Mouse », une voix grinçante, une sorte de rire ou de gémissement, etc.

L'analyse et la synthèse de l'expressivité dans la parole sont donc un nouvel enjeu pour la communauté de la parole. Elles permettent de rendre les systèmes de reconnaissance vocale plus robustes et d'accroître le registre des synthétiseurs de la parole à partir du texte. De plus, elles sont un outil pour les psychologues qui étudient les émotions et les éventuelles pathologies qui y sont liées. Notre approche est avant tout motivée par le désir de modifier, à l'instar d'un acteur, l'expressivité d'une phrase parlée, qu'elle soit synthétisée ou bien naturelle.

Dans un projet récent, nous avons développé des solutions permettant de produire automatiquement ces modifications. Toutes ces transfor-

64. http://forumnet.ircam.fr
65. Voir *supra*, p. **NNN**.

mations doivent être d'une très haute qualité permettant d'étendre leur usage à des domaines d'application, restés jusque-là à l'écart, comme le film d'animation et le doublage de film.

Nous avons donc enregistré des acteurs exprimant un même texte avec différentes expressivités et avec différents niveaux d'intensité expressive. Ces corpus servent à l'établissement de modèles de jeux d'acteur. Ces modèles sont dépendants de variables contextuelles et linguistiques[66] comme le phonème ou le degré de proéminence. Par exemple, les variations acoustiques liées à l'expressivité dépendent du degré de proéminence (particulièrement pour le débit de parole)[67]. Et la transformation du degré d'articulation nécessite la connaissance du contexte phonétique[68]. La capacité d'IrcamCorpusTools de gérer différents types de données y est donc pleinement exploitée. Les variables linguistiques sont utilisées par un *réseau bayésien* pour estimer des densités de probabilités conditionnelles des variables acoustiques relatives aux cinq dimensions de la prosodie (hauteur, débit de parole, intensité, degré d'articulation et qualité vocale[69]). La comparaison de ces densités conduit à différents facteurs de transformation utilisés par le logiciel SuperVP pour modifier l'expressivité d'une phrase neutre. Des exemples sonores sont disponibles sur le Web[70].

Conversion d'identité

Comme en ce qui concerne l'expressivité[71], la plupart des applications, en particulier artistiques, nécessitent de donner à divers personnages, ou à des avatars, des voix bien différenciées et spécifiques du personnage. Il s'agit de produire, par un logiciel, une transformation artificielle de la voix naturelle d'un locuteur pour obtenir la voix d'un personnage différent (apprise sur l'enregistrement d'une voix réelle), ce qui est appelé *conversion d'identité*. En plus des applications artistiques, du doublage et du jeu vidéo, cette technologie trouve des applications dans d'autres domaines comme les avatars et les agents conversationnels. En effet, la méthode de doublage utilisée traditionnellement repose toujours sur le simple enregistrement d'acteurs. Dans les dessins animés comme dans les jeux vidéo, ou même dans le doublage de films, l'utilisation de techniques de modification de voix, au contraire, permet de créer les voix de plusieurs personnages avec l'enregistrement d'un seul acteur. Ainsi,

66. Beller et Rodet (2007).
67. Beller *et al.* (2006), Beller (2007b).
68. Beller (2007a), Beller *et al.* (2008).
69. Pfitzinger (2006).
70. http://recherche.Ircam.fr/equipes/analyse-synthese/beller
71. Voir *supra*, p. **NNN**.

l'Ircam peut proposer l'ensemble des technologies requises pour apporter une solution complète à l'utilisation de la voix dans les applications multimédias en général. Dans le domaine du film, plusieurs applications sont possibles. La première est une modification de voix lorsque l'acteur ne peut pas le faire lui-même. Cette technique a été employée, entre autres, dans le film *Vatel* de Roland Joffé pour améliorer la prononciation de l'anglais de Gérard Depardieu. Il serait même possible de modifier, voir de créer par synthèse à partir du texte, telle ou telle réplique d'un acteur sans avoir à le faire revenir en studio pour un enregistrement. Une autre application est la modification du timbre d'un acteur pour des contraintes liées au rôle. Cette technique a été testée pour le rôle de Klaus Maria Brandauer dans le film *Vercingétorix*. Enfin, la synthèse de la voix pourrait permettre de faire entendre la voix spécifique d'une personne non disponible (acteur non disponible après le tournage par exemple) ou disparue comme cela a été demandé à l'Ircam pour la voix du général de Gaulle ou celle du poète Jean Cocteau. Enfin, dans d'autres domaines comme le théâtre et la musique, des metteurs en scène et des compositeurs souhaitent pouvoir utiliser dans leurs créations des traitements et des synthèses de voix. Dans le cas du théâtre, il s'agirait, par exemple, de générer pendant les répétitions ou le spectacle une voix en temps réel qui servirait d'*alter ego* de l'acteur ou qui démultiplierait la voix de l'acteur. Ou encore de transformer en temps réel la voix d'un acteur en celle d'un autre personnage.

La conversion d'identité d'une voix source en une voix cible consiste à modifier le *timbre*, la hauteur (fréquence fondamentale ou *pitch*) et la durée de chaque phonème prononcé[72]. La méthodologie de conversion d'identité consiste d'abord à apprendre une fonction de transformation à partir d'enregistrements des locuteurs source et cible (par exemple, par la technique dite *mélange de gaussiennes*). Cette transformation peut être ensuite appliquée à la voix du locuteur source au moyen d'une technique d'analyse-synthèse comme SuperVP[73].

SYNTHÈSE DE LA PAROLE À PARTIR DU TEXTE

Le langage de requête de IrcamCorpusTools[74] peut être utilisé de manière manuelle par une succession de lignes de commandes comme dans l'exemple précédent. Mais il peut aussi être invoqué de manière automatique à différents niveaux, de manière à construire des arbres de

72. Villaviciencio *et al.* (2006).
73. Voir supra, p. **NNN**.
74. Voir supra, p. **NNN**.

données par de multiples décisions successives. Un synthétiseur de parole à partir du texte (*text-to-speech* ou TTS) a été construit de cette façon. Les procédés employés dans les systèmes TTS à base de corpus sont aujourd'hui bien connus[75], mais nous présentons ce système, IrcamTTS, car c'est une application du langage de requête. Après une phase d'apprentissage automatique impliquant de nombreuses requêtes sur les données symboliques de plusieurs niveaux (phones, diphones, syllabes, mots, groupes prosodiques, etc.), un arbre de décision est construit de manière à fournir en ses feuilles de nombreux sous-ensembles d'unités du niveau « diphone », qui peuvent d'ailleurs appartenir à des corpus différents. Chacune de ces feuilles correspond à des sous-ensembles d'unités acoustiquement homogènes. À l'étape de synthèse, ces sous-ensembles sont accessibles *via* une succession de requêtes construites à partir du texte à synthétiser. Il en résulte des sous-ensembles d'unités candidates. On sélectionne, parmi ces sous-ensembles, les unités qui minimisent une distance de concaténation, grâce à la programmation dynamique[76]. Cette distance est, elle aussi, apprise automatiquement, et permet de favoriser le naturel des transitions au niveau segmental, mais aussi au niveau suprasegmental. Comme le langage de requête d'IrcamCorpusTools permet de stipuler des contraintes acoustiques, celles-ci peuvent être définies et ajoutées, de manière à influencer la prosodie finale de la phrase de synthèse. Ces contraintes prosodiques peuvent, par exemple, être fournies par un modèle de la proéminence ou par un modèle de l'expressivité[77]. Dans la requête, il est également possible de bannir certaines unités, afin que l'algorithme de sélection en choisisse d'autres parmi les sous-ensembles candidats possibles : ainsi l'utilisateur, ou le programme appelant la requête, peut modifier ou améliorer la synthèse résultante.

VOIX CHANTÉE

L'Ircam a travaillé depuis longtemps sur la synthèse de la voix chantée. Ainsi, l'air de la Reine de la nuit de l'opéra de Mozart *La Flûte enchantée*[78] a entièrement été synthétisé par le logiciel Chant avec la technique dite *formes d'ondes formantiques*.

---

75. Hunt et Black (1996).
76. Viterbi (1967).
77. Voir respectivement supra p. **NNN** et **NNN**.
78. On peut l'entendre à l'adresse : http://recherche.ircam.fr/equipes/analyse-synthese/reine.html

De même, l'Ircam a développé une méthode de synthèse temps-réel[79] de chœurs parlés et chantés[80] pour la plate-forme logicielle Max/MSP, utilisée notamment dans l'opéra *K* de Philippe Manoury.

Enfin, pour le film *Farinelli* de Gérard Corbiaud, la voix du célèbre castrat a été créée par ordinateur par l'équipe « Analyse-synthèse » de l'Ircam[81]. Une soprane et un haute-contre ont été enregistrés, et leur voix montées note à note suivant les parties qu'ils pouvaient chanter au mieux. Ces voix ont été transformées, d'une part pour les rendre égales (et supprimer l'impression de passage d'une voix d'homme à une voix de femme à chaque transition de chanteur), d'autre part pour obtenir la réalisation artistique d'une voix de castrat exceptionnelle, sujet essentiel du film[82].

Transformation de voix parlée en voix chantée

Cette transformation a été conçue notamment pour l'œuvre *Lolita* de Joshua Fineberg (créée en 2006), fondée sur le livre éponyme de Vladimir Nabokov. Le travail est conçu comme un opéra, mais qui se produirait dans l'esprit du narrateur. Toutes les voix chantées entendues par l'assistance sont le résultat de traitements par ordinateur de la voix parlée du narrateur, graduellement transformée en voix de femme[83]. De cette façon, le morceau est un opéra véritablement imaginaire : c'est l'opéra imaginé dans l'esprit du narrateur.

La technique utilisée est une modification *shape invariant* par le logiciel SuperVP qui permet de grandes transpositions et de longs étirements de la voix parlée[84]. L'objectif final est un système automatique de transformation en temps réel de voix parlée en voix chantée, où les syllabes sont automatiquement assignées aux notes correspondantes de la composition. Dans son état actuel, le système n'est pas encore capable de réaliser l'alignement automatique des syllabes parlées et des notes. Pour cela, nous devons encore adapter à cette fin le système d'alignement (présenté p. **NNN**), comme nous l'avons fait pour le chant *fado* à l'occasion de l'œuvre *Com que voz* de Stefano Gervasoni.

Les notes de la composition exigent une transposition de la voix parlée en voix chantée, jusqu'à trois octaves au-dessus et une octave au-dessous. De même, pour donner aux voyelles de la voix parlée les durées des notes de la composition, il faut les étirer d'un facteur qui peut attein-

79. Elle est accessible dans le club d'utilisateurs de l'Ircam : http://forumnet.ircam.fr/
80. Schnell *et al.* (2000).
81. Depalle *et al.* (1995).
82. Depalle *et al.* (1994).
83. Roebel et Fineberg (2007).
84. Roebel et Rodet (2005).

dre la valeur huit ou dix. Ces deux transformations sont effectuées avec une grande qualité par le logiciel SuperVP. Enfin l'enveloppe spectrale du narrateur est transformée en celle d'une chanteuse.

Analyse de voix chantée

L'œuvre *Com que voz* de Stefano Gervasoni confronte le chant portugais traditionnel *fado* au chant en écriture contemporaine d'un ténor. Des enregistrements de la chanteuse de *fado* Christina Branco ont aussi été traités à l'Ircam. Pour cela le logiciel IrcamAlign[85] a été adapté à la phonétique de la langue portugaise et à la voix chantée. Ainsi les syllabes des enregistrements ont pu être segmentées et ont été utilisées comme matériau musical par le compositeur.

Conclusion

Dans ce chapitre, nous avons présenté des études et des développements, en particulier informatiques, d'analyse, de traitement et de synthèse de la voix et de la parole. En premier lieu, nous avons montré comment l'utilisation de grands corpus permet d'extraire des informations essentielles sur la structure de la parole et son contenu linguistique et acoustique. Gérer de tels corpus nécessite des outils particuliers. Dans l'équipe « Analyse-synthèse des sons », nous avons développé un tel outil, qui ouvre des possibilités remarquables. IrcamCorpusTools, une plate-forme extensible pour la création, la gestion et l'exploitation des corpus de parole, permet d'interfacer facilement des données hétérogènes avec des analyseurs internes ou externes, en utilisant le principe d'autodescription des données et des analyseurs. En outre, l'autodescription des données garantit leur pérennité, favorise l'introduction de nouveaux types et leur confère une plus grande visibilité. De même, l'autodescription des analyseurs assure l'extensibilité de la plate-forme ainsi que sa modularité et la mutualisation des corpus. La plate-forme IrcamCorpusTools est capable de gérer les relations hiérarchiques multiples et séquentielles entre des unités. Un langage de requête simple et expressif donne un accès immédiat aux données de ces unités. Ces fonctionnalités appliquées à différents corpus de parole (parole contrôlée et parole spontanée pour des études de la prosodie et/ou de l'expressivité) intéressent directement les

85. Voir supra, p. **NNN**.

recherches à la frontière entre le traitement automatique des langues et le traitement automatique de la parole. En guise d'exemple, un processus de synthèse de parole expressive à partir du texte peut être entièrement réalisé, depuis sa genèse jusqu'au résultat sonore. L'intégration de différentes exploitations rassemblées au sein d'une même plate-forme illustre les avantages de l'interopérabilité. C'est pourquoi nous avons le projet de distribuer publiquement IrcamCorpusTools à des communautés de chercheurs.

Pour le traitement du signal sonore lui-même, deux outils ont été présentés, SuperVP et AudioSculpt. Ils permettent de sculpter littéralement le son, comme un plasticien hors temps réel, ou même dans le temps réel de la parole, ce qui est essentiel sur scène notamment mais aussi pour la rapidité des mises au point.

Enfin nous avons montré comment ces recherches et logiciels sont utilisés dans la création musicale, artistique en général, et dans les multimédias où de tels outils sont de plus en plus demandés et mis en œuvre[86].

RÉFÉRENCES BIBLIOGRAPHIQUES

– Avanzi M., Lacheret-Dujour A. et Victorri B. (2008), « ANALOR. A tool for semi-automatic annotation of French prosodic structure », *Speech Prosody*.
– Barras C., Geoffrois E., Wu Z. et Liberman M. (1998), « Transcriber : A free tool for segmenting, labelingand transcribing speech », *LREC*, p. 1373-1376.
– Beller G. (2007a), « Influence de l'expressivité sur le degré d'articulation », *RJCP*, Rencontres jeunes chercheurs de la parole.
– Beller G. (2007b), « Transformation de la parole dépendante de l'expressivité et du texte », Journée des sciences de la parole.
– Beller G., Marty A. (2006), « Talkapillar : outil d'analyse de corpus oraux », *RJC-ED268*, Paris-III-Sorbonne-nouvelle.
– Beller G., Obin N. et Rodet X. (2008), « Articulation degree as a prosodic dimension of expressive speech », *Speech Prosody*, Campinas.
– Beller G. et Rodet X. (2007), « Content-based transformation of the expressivity in speech », *ICPhS*, Saarbruecken.
– Beller G., Schwarz D., Hueber T. et Rodet X. (2006), « Speech rates in French expressive speech », *Speech Prosody*, *SproSig*, ISCA, Dresde.
– Bilhaut F. et Widlöcher A. (2006), « LinguaStream : An integrated environment for computational linguistics experimentation », Trente, Italie.

---

86. REMERCIEMENTS. Le développement d'IrcamCorpusTools et des outils et logiciels décrits dans cet article, sont partiellement financés par le projet RIAM VIVOS (VIVOS : http://www.vivos.fr) sur la création de voix expressives pour des applications multimédias, par le projet ANR RHAPSODIE (RHAPSODIE : http://rhapsodie.risc.cnrs.fr) sur l'élaboration de corpus prosodiques de référence en français parlé et par le projet RIAM AFFECTIVE AVATARS, nouvelle génération d'avatars expressifs pilotés par la voix.

– Bird S., Day D., Garofolo J., Henderson J., Laprun C. et Liberman M. (2000), « ATLAS : A flexible and extensible architecture for linguistic annotation », *arXiv*.
– Bird S. et Liberman M. (2001), « A formal framework for linguistic annotation », *Speech Commun.*, vol. 33, n° 1-2, p. 23-60.
– Boersma P. (2001), « Praat, a system for doing phonetics by computer », *Glot international*, vol. 5-9, p. 341-345.
– Bogaards N., Roebel A. et Rodet X. (2004), « Sound analysis and processing with AudioSculpt 2 », *International Computer Music Conference* (ICMC), Miami (E.-U.).
– Cassidy S. et Harrington J. (2001), « Multi-level annotation in the Emu speech database management system », *Speech Communication*, vol. 33, p. 1-2, Elsevier Science Publishers B. V., Amsterdam, p. 61-77.
– Chafe W. (1992), « The importance of corpus linguistics to understanding the nature of language », *in* J. Svartvik (éd.), *Directions in Corpus Linguistics, Proceedings of Nobel Symposium 82*, Berlin-New York, Mouton de Gruyter, p. 79-97.
– Cunningham H., Maynard D., Bontcheva K. et Tablan V. (2002), « GATE : A framework and graphical development environment for robust NLP tools and applications », *Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics*.
– De Cheveigné A. et Kawahara H. (2002), « YIN, a fundamental frequency estimator for speech and music », *JASA*, vol. 111, p. 1917-1930.
– Degottex G., Bianco E. et Rodet X. (2008a), « Measure of glottal area on high-speed videoendoscopy », *Production Workshop : Instrumentation-based Approach*, Paris.
– Degottex G., Bianco E. et Rodet X. (2008b), « Usual to particular phonatory situations studied with highspeed videoendoscopy », *International Conference on Voice Physiology and Biomechanics*.
– Depalle_P., Garcia G. et Rodet X. (1994), *A Virtual Castrato, International Computer Music Conference* (ICMC), Arrhus.
– P. Depalle, G. Garcia et X. Rodet (1995), « Reconstruction of a castrato voice : Farinelli's voice », *IEEE WASPAA*, Mohonk Mountain House (NY).
– Durand J. et Tarrier J.-M. (2006), « PFC, corpus et systèmes de transcription », *Cahiers de grammaire*, vol. 30, p. 139-158.
– Farner S., Rodet X. et Ach L. (2008), *Voice Transformation and Speech Synthesis for Video Games*, Paris Game Developers Conference, Paris.
– Farner S., Roebel A. et Rodet X. (2009), *Natural Transformation of Type and Nature of the Voice for Extending Vocal Repertoire in High-Fidelity Applications*, soumis à l'AES.
– Gussenhoven C. et Jacobs H. (2004), *Understanding Phonology*, Arnold, 2005.
– Gut U., Milde J.-T., Voormann H. et Heid U. (2004), « Querying annotated speech corpora », *Speech Prosody*, Nara, Japan.
– Habert B. (2000), « Des corpus représentatifs : de quoi, pour quoi, comment ? », *Linguistique sur corpus*, Perpignan, p. 11-58.
– Henrich N. (2001), *Étude de la source glottique en voix parlée et chantée*, PhD thesis, université Paris-VI.

– Hunt A. J. et Black A. W. (1996), « Unit selection in a concatenative speech synthesis system using a large speech database », *ICASSP*, IEEE Computer Society, Washington (DC), p. 373-376.
– Lai C. et Bird S. (2004), « Querying and updating treebanks : A critical survey and requirements analysis ».
– Lamel L., Gauvain J.-L. et Eskénazi M. (1991), « Bref, a large vocabulary spoken corpus for French », *EuroSpeech*.
– Lanchantin P., Morris A., Rodet X. et Veaux C. (2008), « Automatic phoneme segmentation with relaxed textual constraints », *Proc. of LREC*, Marrakech.
– Laroche J. et Dolson M. (1999), « New phase-vocoder techniques for real-time pitch shifting, chorusing, harmonizing and other exotic audio modifications », *Journal of the AES*, vol. 47, n° 11, p. 928-936.
– Quatieri T. F. et McAulay R. J. (1992), « Shape invariant time-scale and pitch modification of speech », *IEEE Transactions on Signal Processing*, 40 (3), p. 497-510.
– MacWhinney B. (2000), *The CHILDES Project : Tools for Analyzing Talk*, troisième édition, vol. I : *Transcription Format and Programs*, Lawrence Erlbaum Associates, Mahwah (NJ)
– Maeda S. (1982), « A digital simulation method of the vocal-tract system », *Speech Communication*.
– Müller C. (2005), « A flexible stand-off data model with query language for multi-level annotation », *ACL*, p. 109-112.
– Nakov P., Schwartz A., Wolf B. et Hearst M. (2005), « Supporting annotation layers for natural language processing », *ACL*, p. 65-68.
– Obin N., Goldman J., Avanzi M. et Lacheret-Dujour A. (2008a), « Comparaison de 3 outils de détection automatique de proéminence en français parlé », Journées d'études de la parole, Avignon.
– Obin N., Lacheret-Dujour A., Veaux C., Rodet X. et Simon A.-C. (2008b), « A method for automatic and dynamic estimation of discourse genre typology with prosodic features », soumis à *Interspeech*, Brisbane, Australia.
– Obin N., Rodet X. et Lacheret-Dujour A. (2008c), « French prominence : a probabilistic framework », *Proc. of ICCASP*, Las Vegas (NV), P. 3993-3996.
– Oostdijk N. (2000), « The spoken Dutch corpus : Overview and first evaluation », *LREC*, 887-893.
– Pfitzinger H. (2006), « Five dimensions of prosody : Intensity, intonation, timing, voice quality, and degree of reduction », *in* Hoffmann, H., Mixdorff, R. (éd.), *Speech Prosody*, n° 40, Abstract Book, Dresde, p. 6-9.
– Rabiner L. (1989), « A tutorial on hidden Markov Models and selected applications in speech recognition », *IEEE*, vol. 77, 2, p. 257-286.
– Roebel A. (2003), « Transient detection and preservation in the phase vocoder », *International Computer Music Conference (ICMC)*, Singapour, p. 247-250.
– Roebel A. et Rodet X. (2005), « Efficient spectral envelope estimation and its application to pitch shifting and envelope preservation », *Proc. of the 8th Int. Conf. on Digital Audio Effects (DAFx05)*, p. 30-35.
– Roebel A. et Fineberg J. (2007), « Speech to chant transformation with the phase vocoder », *Interspeech*, Anvers.

– Roebel A., Villaviciencio F. et Rodet X. (2007), « On cepstral and all-pole based spectral envelope modeling with unknown model order », *Pattern Recognition Letters, Special issue on Advances in Pattern Recognition for Speech and Audio Processing*, accepté pour publication.
– Rodet X., Escribe J. et Durigon S. (2004), « Improving score to audio alignment : Percussion alignment and precise onset estimation », *Proc Int. Conf on Computer Music* (*ICMC*), p. 450-453.
– Rosenberg A. et Hirschberg J. (2007), « Detecting pitch accent using pitch-corrected energy-based predictors », *Interspeech*, Anvers, p. 2777-2780.
– Schnell N., Peeters G., Lemouton S., Manoury, P. et Rodet X. (2000), *Synthesizing a Choir in Real-time Using Pitch Synchronous Overlap Add (PSOLA)*, ICMC : International Computer Music Conference, Berlin.
– Sjölander K. et Beskow J. (2000), « WaveSurfer. An open source speech tool », International Conference on Spoken Language Processing.
– Taylor P., Black A. W. et Caley R. (2001), « Heterogeneous relation graphs as a mechanism for representing linguistic tnformation », *Speech Communication*, vol. 3, p. 153-174.
– Veaux C., Beller G. et Rodet X. (2008), « IrcamCorpusTools : An extensible platform for speech corpora exploitation », *LREC*, Marrakech, Maroc.
– Villavicencio F., Roebel A. et Rodet X. (2006), « Improving LPC spectral envelope extraction of voiced speech by true envelope stimation », *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, vol. I, p. 869-872.
– Villavicencio, F., Roebel, A. et Rodet, X. (2007), « All-pole spectral envelope modelling with order selection for harmonic signals », *IEEE International Conference on Acoustics, Speech and Signal processing* (*ICASSP'07*), 1, I-49 I-52.
– Villavicencio, F., Roebel, A. et Rodet, X. (2008), « Extending efficient spectral envelope modeling to mel-frequency based representation », *IEEE International Conference on Acoustics, Speech, and Signal processing* (*ICASSP'08*), p. 1625-1628.
– Villavicencio F., Roebel A. et Rodet X. (2009), « Applying improved spectral modeling for high-quality voice conversion », soumis à *ICASSP*.
– Vincent D., Rosec O. et Chonavel T. (2005), « Estimation of LF glottal source parameters based on an ARX model », *Interspeech*.
– Viterbi A. J. (1967), « Error bounds for convolutional codes and an asymptotically optimum decoding algorithm », *IEEE TIT*, vol. 13 (2), p. 260-269.

IV

PLASTICITÉ ET ÉDUCATION

L'apprentissage du chant chez les oiseaux : l'importance des influences sociales

par MARTINE HAUSBERGER[1]

La communication vocale : un mode d'échange privilégié dans le règne animal

La communication peut prendre différentes formes. Ainsi des actes non intentionnels peuvent résulter en un transfert d'informations : le bruit de masticage d'une proie ou les traces laissées dans la neige peuvent informer un prédateur de sa présence. Mais d'autres actions ont évolué spécifiquement dans un but de communication.

Là encore, cependant, des différences sont observables. Beaucoup de signaux sont fixés et liés à une fonction particulière : le cri de menace d'un félin, l'alarme du geai, voire les cris des nouveau-nés ont une forme relativement stéréotypée et pas fondamentalement modifiée par l'émetteur. Ils ne requièrent pas d'apprentissage. De tels signaux sont largement répandus, caractérisés par leur fixité et leur lien à une fonction donnée (menace, alarme ou détresse par exemple[2]). Certains groupes animaux possèdent essentiellement de tels signaux (beaucoup de mammifères terrestres, oiseaux non chanteurs, certains amphibiens). Pour d'autres, apparaissent, en plus, des signaux vocaux flexibles et acquis par l'expérience, en particulier sociale. Sont ainsi concernés les mammifères marins (dauphins, baleines, orques…), certains autres mammifères (chauves-souris, éléphants) et surtout les oiseaux chanteurs, dont cela constitue une caractéristique majeure. De façon intéressante, cette notion même de flexibilité

1. Avec Laurence Henry, Hugo Cousillas, Jean-Pierre Richard, Françoise Joubaud et Isabelle George.
2. Oller et Griebel (2008).

sépare la communication vocale des primates non humains du langage humain. Si les primates non humains font preuve d'importantes capacités d'ajustement au contexte de leurs signaux vocaux, voire d'éléments dits « sémantiques[3] », ils montrent peu ou pas d'aptitude à modifier leurs signaux en fonction de l'expérience. Les seuls exemples montrent des variations fines de modulation de cris préexistants[4].

Un grand pas vers l'évolution du langage a certainement été cette libération de la fixité des signaux (de forme et de fonction) et l'ouverture à l'expérience. Hommes, cétacés et oiseaux chanteurs (mais aussi les perroquets et les colibris) partagent cette capacité à « construire » leur communication vocale au cours du développement sous l'importante « pression » des influences sociales. Parmi les animaux, les oiseaux chanteurs ont été les mieux étudiés, et la diversité de leurs modes de vie amène à de fascinantes questions : qui apprend quoi, de qui, pourquoi et comment ? Ils offrent des modèles d'observation et d'expérimentation privilégiés où peuvent se rejoindre des questions fonctionnelles (pourquoi chanter ainsi ?), et plus « mécanistiques » (comment le cerveau traite-t-il l'information ?). Nous allons donc poursuivre en nous focalisant sur ce groupe privilégié qui a fasciné l'homme de tout temps.

L'oiseau et son chant

Les oiseaux ont toujours eu une place privilégiée auprès de l'homme, signes d'élévation et de lien avec le divin selon Platon. Leur chant, particulièrement, fascine : chez les Bambaras, c'est à l'oiseau que se rattache le don de la parole[5] ; pour le Coran, le langage des oiseaux est le langage des hommes[6]. Enfin, le chant de certaines espèces servait de présage, leur valant leur nom actuel d'oscines, du latin « *oscen, inis*, nom [qui] signifie : tout oiseau dont le chant servait de présage[7] ».

Ces premiers récits montrent que les Anciens observaient leur comportement avec attention, remarquant que « nul oiseau ne chante quand il souffre » (Platon), que « le merle chante en été, bégaie en hiver, et est

3. Zuberbühler (2000).
4. Elowson et Snowdon (1994) ; Snowdon et Elowson (1999) ; Lemasson, Zuberbühler et Hausberger (2005).
5. Calame-Griaule (1965).
6. Gheerbrant et Chevalier (1997).
7. Gaffiot (1934).

muet vers le solstice d'été » ou encore que « chaque rossignol a plusieurs airs, et ces airs ne sont pas les mêmes pour tous ; chacun a les siens » (Pline l'Ancien). Toutes ces observations sont confirmées par les études modernes : le chant semble être bien associé à un contexte où l'animal ne souffre pas, au sens propre du terme, où il est perché, actif[8] ; la plupart des espèces montrent bien des rythmes annuels où le chant est concentré entre mars et juin, période de reproduction ; et enfin différentes études expérimentales ont révélé que le chant était porteur d'information sur l'identité individuelle de l'émetteur, comme par exemple le rouge-gorge[9]. Outre la reconnaissance spécifique et individuelle, le chant permet aux oiseaux selon leur système social, de défendre un territoire, d'attirer un partenaire sexuel ou d'assurer la cohésion sociale[10]. L'autre particularité fascinante est certainement la capacité des oiseaux à apprendre leur chant d'un autre : « les rossignols plus jeunes étudient et reçoivent la leçon qu'ils doivent apprendre ; l'élève écoute avec attention et il répète » (Pline l'Ancien), mais aussi pour certains à imiter la voix humaine (par exemple, l'étourneau, le mainate, l'oiseau-lyre, la perruche et le perroquet[11]). Les performances bien connues du célèbre Alex (perroquet gris du Gabon, oiseau non chanteur mais ouvert à l'apprentissage[12]) ont révélé l'aptitude de tels oiseaux à répondre vocalement et de façon appropriée à des questions comme la matière, la couleur, la forme ou encore la taille d'un objet : apprentissage du son mais également de la pertinence du contexte. Comment est-ce possible ? Par l'existence chez ces oiseaux d'un organe phonatoire, la syrinx, correspondant à un renforcement de la trachée, à la jonction des bronches et sur lequel viennent s'attacher plusieurs paires de muscles. Les premiers anneaux bronchiques droits et gauches sont modifiés en deux membranes tympaniformes internes pouvant vibrer de façon indépendante. La production du chant résulte de l'activité des muscles respiratoires abdominaux et thoraciques, qui font se mouvoir l'air dans la syrinx. Deux sources de sons existent, une à la base de chaque bronche, amenant à une synchronie ou à la production de deux sons différents[13]. La production du chant requiert donc une coordination motrice très précise, harmonisant respiration, ouverture du bec et contraction des muscles appropriés.

8. Hartshorne (1973).
9. Brémond (1968).
10. Catchpole et Slater (1995) ; Snowdon et Hausberger (1997).
11. Feare (1984) ; Thorpe (1961) ; Pepperberg (1997).
12. Pepperberg (1997).
13. *Cf.* Hausberger *et al.* (2002).

Il est considéré maintenant que les oscines correspondent à des oiseaux possédant au moins quatre paires de muscles en commande de l'organe phonatoire. L'apprentissage du chant va donc requérir de l'oiseau l'apprentissage de la maîtrise de ces différents éléments. Ceci, comme l'indiquait Pline l'Ancien, demande un modèle, généralement un adulte de même espèce ou, lorsqu'il s'agit d'imitation humaine, d'un humain agissant comme substitut social pour un oiseau élevé seul[14]. Dans tous les cas, le jeune oiseau est dépendant de la présence des autres, et ne pourra développer le chant de l'espèce en absence d'un congénère.

Vers un modèle d'apprentissage

Les premières expériences prouvant un apprentissage datent du XVIII[e] siècle, où le baron von Pernau, un ornithologue autrichien amateur, a réalisé une série d'expérimentations en volières, en particulier sur les pinsons. Néanmoins, à cette époque et pendant longtemps encore, la précision de cet apprentissage et ses étapes ne pouvaient être clairement définies. En effet, comment rendre compte d'impressions sonores ? La préoccupation de représenter visuellement le son a été une problématique constante, que seul a résolu l'avènement du sonagraphe dans les années 1960 en permettant la représentation de chants sur un diagramme à deux axes correspondant à la fréquence et au temps, et donc d'obtenir une trace mesurable et comparable[15]. C'est alors qu'ont pu se développer des travaux sur l'apprentissage du chant qui allaient prendre une importance exponentielle dans ces dernières décennies.

Thorpe (dès 1958) puis Nottebohm (en 1968), ont travaillé sur le développement du chant chez les pinsons des arbres. Ils ont montré que des jeunes oisillons, prélevés au nid à quelques jours, élevés « à la main » par des humains et donc privés de contact auditif avec des adultes de leur espèce, développent seulement un chant pauvre ne ressemblant pas au chant typique du pinson. Des contacts sociaux entre jeunes pinsons permettent un développement plus poussé, mais là encore atypique. Seules des interactions avec les adultes permettent un développement normal : une phase de préchant (gazouillis très compliqué doux et décousu), suivie d'une phase de chant plastique (chant plus structuré, ressemblant au chant de l'adulte mais encore

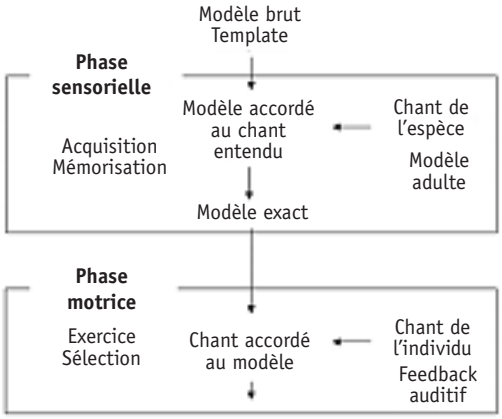
14. Thorpe (1961) ; Pepperberg (1997) ; West, Stroud et King (1983).
15. *Cf.* Hausberger *et al.* (2002).

variable), puis d'une cristallisation du chant spécifique définitif. Le cas le plus extrême est observé lorsque l'oisillon est assourdi alors qu'il n'a que quelques jours. Le fait que l'oisillon ne puisse pas entendre ses propres vocalisations (feed-back auditif) aboutit à la production de vocalisations totalement déstructurées. Le jeune doit comparer ses propres émissions à celles d'un adulte pour aboutir à une copie fidèle du modèle. Ces observations indiquent donc que les jeunes oiseaux doivent s'entendre et entendre un ou des adultes de leur espèce pour pouvoir produire à leur tour des chants typiques.

Dans les années 1970, Peter Marler a réalisé des séries d'expériences qui ont insisté sur deux notions importantes : la notion de période sensible et la notion de *template*. Ces expériences, réalisées selon un protocole très strict, consistaient à placer des jeunes oiseaux, prélevés au nid et élevés à la main (des bruants à tête couronnée), seuls dans une chambre sourde (pièce isolée phonétiquement) et à diffuser via un haut-parleur des chants de l'espèce à différents âges de l'oiseau. Marler a ainsi montré que ces oiseaux doivent entendre le chant de l'adulte dans une fenêtre très courte (la période sensible) comprise entre 10 et 50 jours de vie, faute de quoi ils n'apprennent pas. De même, la diffusion via des haut-parleurs de chants d'autres espèces à de jeunes bruants a le même effet que l'isolement total. Cependant, la diffusion de ces mêmes chants couplés à des chants spécifiques induit le développement de chants typiques de l'espèce. Cette expérience souligne donc que le jeune filtre les informations pertinentes dans son environnement sonore. Marler définit le *template* comme un modèle auditif inné qui permet au jeune de sélectionner les chants de ses congénères et de rejeter les chants des autres espèces[16].

Ces résultats ont permis de proposer un modèle d'apprentissage du chant (figure 1) selon lequel son développement se réalise en deux étapes : une étape de mémorisation, au cours de laquelle le jeune oiseau entend et mémorise les chants des adultes, et une étape motrice au cours de laquelle les jeunes oiseaux produisent des vocalisations. Les oiseaux s'entraînent donc à chanter jusqu'à ce que leurs productions s'accordent au modèle mémorisé plus tôt. Le feed-back auditif intervient dans cette phase. Chez certaines espèces, ces deux étapes peuvent être décalées dans le temps (ex : le canari), alors que, chez d'autres, elles se chevauchent (ex : le diamant mandarin). Deux catégories d'espèces ont été définies : les oiseaux qui n'apprennent qu'au cours d'une période limitée et pour qui le chant acquis pendant la période sensible reste inchangé (*age-limited learners*), et ceux qui maintiennent la capacité de modifier leur chant tout au long de leur vie (*open-ended learners*). Nous allons voir cependant que la situation n'est pas aussi caricaturale.

16. Marler (1970).



d'après Catchpole & Slater, 1995
(Thorpe, 1958 ; Konishi, 1965 ; Marler, 1970)

Figure 1 : *Un modèle d'apprentissage du chant. L'apprentissage du chant se réalise au cours de deux phases : une phase de mémorisation du chant des adultes et une phase motrice au cours de laquelle le jeune oiseau s'entraîne à chanter jusqu'à ce que ses propres vocalisations s'accordent au modèle mémorisé plus tôt.*

Du modèle théorique à la « réalité » : l'impact des influences sociales

Le modèle théorique basé sur les expérimentations en conditions contrôlées de Marler mettait l'accent sur les nombreux parallèles existant entre le développement du langage chez l'homme et le développement du chant chez les oiseaux : existence d'une période sensible, nécessité d'une exposition aux signaux spécifiques, besoin de s'exercer à la production (préchant/babillage) et d'entendre ses propres productions (feed-back auditif) et absence de changements après la fin de la période sensible. Or Baptista et Petrinovitch[17] ont remis en cause ce paradigme, se basant sur les conditions de vie habituelles d'un jeune oiseau qui est en contact direct avec un « congénère modèle ». Ils ont donc repris les expériences de Marler en utilisant des modèles sociaux vivants comme modèles (et non des haut-parleurs). Dans ce dispositif expérimental, les jeunes oiseaux pouvaient donc voir, entendre et interagir directement avec l'adulte. Ces

17. Baptista et Petrinovitch (1984, 1986).

chercheurs ont montré que, dans ce cas, l'apprentissage est possible au-delà de la période sensible et que même des chants d'une autre espèce peuvent être appris si le modèle vivant appartient à une autre espèce. Les influences sociales peuvent donc permettre de lever des inhibitions et d'aboutir à des apprentissages hors période sensible ou des apprentissages hétérospécifiques (*cf.* les imitations humaines précitées).

Baptista et Gaunt[18] ont montré que les influences sociales sur l'apprentissage du chant peuvent prendre des formes diverses. Pour certaines espèces, comme le grimpereau par exemple, le contact social est totalement nécessaire. Ces espèces sont qualifiées d'*obligate social learners*. Pour d'autres espèces (canari), le contact n'est pas nécessaire mais favorise un meilleur apprentissage. Ce sont les *facultative social learners*. Par ailleurs, le modèle choisi par le jeune oiseau peut-être différent. Ainsi le père nourricier peut être choisi comme tuteur, comme chez le bouvreuil où un jeune élevé par un canari produit plus tard un chant de canari. De même, les diamants mandarins copient 60 % de leur chant sur leur père en condition naturelle[19]. D'autres espèces apprennent le chant plus tard, à l'âge adulte ; ils pourront alors copier leurs voisins de nid ou partenaires de groupe. Dans certains cas apparaissent alors des signaux communs, qui définissent le groupe ou la localité (dialectes et « microdialectes »). Bien que, chez la plupart des espèces des régions tempérées, seul le mâle chante, au moins pendant la reproduction, chez d'autres, les femelles sont également actives et amenées à apprendre le chant. Chez les mainates, l'apprentissage se fait selon des lignées mâles et femelles respectivement. Ainsi, selon les espèces, les oiseaux peuvent apprendre tôt, tard ou tout au long de leur vie, copier leur parent, un voisin ou un partenaire social, choisir de quel sexe apprendre… Une diversité qui rend l'approche générale complexe mais qui néanmoins fait ressortir des éléments communs, et en particulier l'importance des influences sociales.

Influences sociales et apprentissage du chant : aspects généraux

Un certain nombre de principes généraux émergent des différentes études[20] :

18. Baptista et Gaunt (1997).
19. Zann (1997).
20. Snowdon et Hausberger (1997).

– L'apprentissage ne consiste pas seulement à apprendre à produire mais aussi à utiliser les sons de façon appropriée. Des oscines vachers élevés dans un environnement sans adulte produisent des chants normaux mais les utilisent de façon non appropriée, ce qui induit des réactions agressives lorsqu'ils sont replacés avec des adultes.

– Les influences sociales peuvent, comme indiqué précédemment, retarder la période d'apprentissage. C'est le cas chez le bruant à couronne blanche, le bruant indigo et bien d'autres espèces encore.

– Le chant appris peut servir de « marqueur social » : la grande majorité des apprentissages se fait entre individus familiers, le plus souvent avec un lien social marqué. Même chez les espèces territoriales, les voisins, qui partagent des chants similaires, se montrent peu agressifs entre eux. L'apprentissage du chant, par la copie précise faite du modèle, permet l'émergence de groupes d'individus présentant des chants plus similaires que ceux d'autres individus de l'espèce. Ce phénomène fait que le chant est porteur de l'identité de groupe, voire de la population (dialectes).

– La nature des influences sociales est complexe. La compréhension des mécanismes impliqués a été un des grands défis des dernières années et nous sommes bien loin encore de comprendre tous les rouages de ce mode de transmission. Lors d'une interaction, le partenaire social est une source de focus attentionnel, de stimulation multimodale et de renforcement. Chacun de ces aspects peut être impliqué. Parfois la simple proximité suffit à exprimer une affinité et à induire un apprentissage. L'attention du modèle et de l'apprenant est une caractéristique commune et majeure de l'apprentissage vocal. Enfin, l'apprenant est un acteur de son développement, faisant activement le choix du modèle qu'il va copier et du moment où il va imiter.

Il reste néanmoins encore beaucoup à comprendre de ces mécanismes. Comment les influences sociales contribuent-elles à la mise en place de systèmes de traitement des signaux ? Comment peuvent-elles lever des inhibitions comme celle de la fenêtre développementale qu'est la période sensible. Quels en sont les éléments majeurs : vision, audition, multimodalité ? Seules des études intégratives combinant analyses fines du comportement et approches expérimentales du développement et de la perception peuvent permettre d'appréhender ces aspects. C'est ce que nos études sur les étourneaux sansonnets apportent.

Une espèce privilégiée : l'étourneau sansonnet

UN IMITATEUR RECONNU

L'étourneau sansonnet (*Sturnus vulgaris*) se révèle être un modèle particulièrement pertinent pour l'étude des influences sociales sur le développement de la communication vocale. Cette espèce d'oiseau chanteur présente une vie sociale riche et focalise depuis plusieurs centaines d'années l'intérêt de savants et d'artistes. Pline l'Ancien avait déjà perçu les extraordinaires capacités d'apprentissage et d'imitation de l'oiseau. Il écrivait ainsi :

> Les jeunes césars, Britannicus et Néron avaient un étourneau apprenant à parler grec et latin, étudiant chaque jour, prononçant incessamment de nouvelles paroles… On instruit les oiseaux dans un lieu retiré et aucune voix ne se fait entendre ; le maître, assis à côté, répète fréquemment ce qu'il veut graver dans leur mémoire, et leur donne des aliments qui les flattent.

Plus tard, Schubert et Mozart ont eux aussi élevé un étourneau et popularisé cet oiseau par leur musique. L'une des caractéristiques de cette espèce, relativement facile à élever en captivité, est de maintenir des capacités d'apprentissage tout au long de sa vie[21].

L'étourneau occupe une très vaste aire de répartition et est maintenant présent sur tous les continents. Originaire de l'Eurasie, il a colonisé l'Amérique du Nord, le sud de l'Afrique et le sud de l'Australie, où l'espèce a été introduite il y a deux siècles. Sur tous ces sites, les oiseaux vivent en groupes de taille variable, le jour, pour se reproduire (quelques couples dans de petites colonies de reproduction) et s'alimenter (quelques dizaines d'individus sur des sites alimentaires proches des colonies), ou bien la nuit pour dormir (plusieurs millions d'oiseaux). Le chant est présent dans tous ces aspects de la vie sociale tout au long de l'année.

Chaque étourneau possède un répertoire de chant, sorte de catalogue d'éléments vocaux différents en structure et servant souvent des fonctions différentes. Trois classes de chant peuvent ainsi être décrites. Les deux premières classes correspondent à des éléments brefs, discontinus et relativement puissants (sifflements) fréquemment utilisés lors d'interac-

21. West *et al.* (1990).

tions vocales. Certains d'entre eux (classe I) sont partagés par tous les individus mâles de l'espèce et présentent la même gamme de variations dans toutes les populations étudiées. Ces sifflements sont nommés « thèmes universels ». D'autres sifflements (classe II) sont caractéristiques d'un individu ou d'une toute petite communauté d'individus (2 ou 3) appartenant à une unité sociale stable. Ces sifflements de classe II sont produits par les mâles et par les femelles. La troisième classe de chant concerne un chant long, continu (une à deux minutes), peu sonore (le gazouillis), produit à la fois par les mâles et les femelles. Ce chant, de par sa structure, ne permet pas d'interactions vocales et est constitué d'une succession de motifs répétés, caractéristiques d'un individu ou partagés par une unité sociale stable.

CHANT ET VIE SOCIALE

Les études de terrain, menées sur plus de 400 individus appartenant à 10 populations en Europe et en Australie, indiquent que les 5 thèmes de sifflements de classe I servent de support à un système complexe de dialecte. Les oiseaux partageant une même variante dialectale d'un thème donné sont dans des zones dialectales bien délimitées. Les zones dialectales ainsi mises en évidence ont une surface différente d'un thème à l'autre (de quelques m_ à plusieurs centaines de km_) ce qui souligne la complexité du système de partage de chant. Ces partages de sifflements entre oiseaux vivant sur une même zone constituent de véritables « labels sociaux[22] ». En effet, lorsque les oiseaux quittent les colonies de reproduction pour rejoindre les sites nocturnes, ils ont tendance à se regrouper entre individus partageant le même dialecte.

Les études réalisées en captivité sur plusieurs groupes d'individus ont révélé un lien fort entre affinité et partage de chant. Nous avons pu montrer qu'en dehors de la reproduction, les oiseaux présentent des affinités sociales fortes avec des individus de même sexe qu'eux. Les femelles forment des paires stables et les mâles des petits groupes plus lâches. Dans tous les cas, la ressemblance vocale (sifflements de classe II et motifs de gazouillis de classe III) reflète les affinités sociales et des lignées sexuelles d'apprentissage du chant s'observent donc chez l'étourneau[23].

Un apprentissage du chant marqué par les influences sociales

Les étourneaux font partie des *facultative social learners* : ils peuvent apprendre des sons diffusés via un haut-parleur mais ils le font moins

22. Hausberger *et al.* (2008).
23. Hausberger *et al.* (1995).

bien que s'ils sont directement confrontés à un modèle vivant. Même dans les cas mentionnés plus haut d'imitation de la voix humaine, seul un contact hétérospécifique intense (l'homme devenant un « substitut social ») permet atteindre ces niveaux d'apprentissage « exceptionnels[24] ». L'importance du contact direct est donc indéniable.

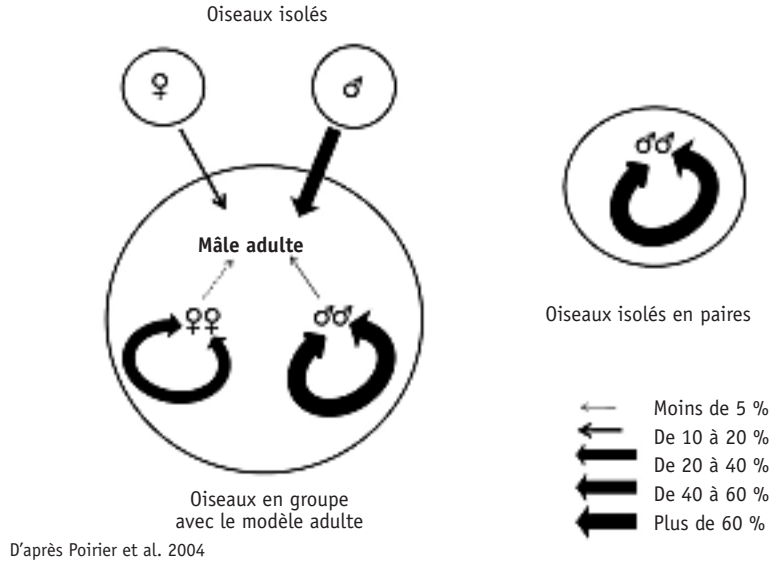


Figure 2 : *Partage de gazouillis observé entre d'une part, de jeunes oiseaux élevés dans des situations sociales différentes, et d'autre part leur modèle adulte. Le taux de partage de chant est représenté par la taille des flèches. Les oiseaux élevés isolément ont plus copié le modèle adulte que les oiseaux élevés en contact direct avec lui. Les oiseaux élevés socialement forment des groupes d'oiseaux de même sexe et se ressemblent vocalement. Les oiseaux élevés en paires ont totalement négligé l'information auditive venant de l'adulte et partagent une part très importante de leurs vocalisations avec leur partenaire social.*

Dans une série d'expériences, nous avons tenté de faire émerger la part de l'influence sociale et de l'influence auditive. Ainsi, si l'on place pendant un an de jeunes étourneaux (précédemment pris au nid et élevés à la main), soit en situation sociale normale (adultes + autres jeunes), soit seuls, soit en paires de jeunes inexpérimentés, et que tous ont accès à la même information auditive, on peut constater l'énorme impact de l'expérience sociale précoce à l'âge adulte. Seuls les jeunes placés directement avec des adultes ont un chant correctement structuré, suivis des animaux

24. West et King (1990).

placés en isolement. Ceux-ci, en l'absence d'autre stimulation, ont clairement appris quelques-uns des chants diffusés. Au contraire, ceux élevés en paires ont développé une sorte de « dialecte » propre et ne présentent aucune structuration de chant d'adulte. L'hypothèse est que l'expérience sociale focalise l'attention ; dans ce dernier cas, elle la focalise sur le partenaire peu expérimenté mais présent, plutôt que sur la source sonore (non « sociale »). De façon intéressante, même les oiseaux élevés en groupe avec adultes ont négligé une partie de l'information : ils ont tendance à former des groupes de jeunes de même sexe, se ressemblant vocalement (l'un improvise et les autres copient) et se distanciant de l'adulte présent socialement mais pas partenaire privilégié[25] (Figure 2).

Dans une série d'expériences complémentaires[26], nous avons pu montrer que la seule situation où un jeune étourneau copie pleinement un adulte est un ratio de 1/1. Un contact unique avec l'adulte semble essentiel pour le développement du jeune. Plus la proportion de jeunes augmente, plus l'information issue de l'adulte semble négligée, les jeunes développant alors une sorte de « vocabulaire à eux » et non commun avec l'adulte. Même les femelles se sont montrées sensibles à ce ratio, étant particulièrement perturbées dans des situations triadiques (2 adultes/ 1 jeune). Il semble donc bien que, chez cette espèce, le contact direct avec l'adulte mais aussi la focalisation de son attention sur ce modèle – une situation rencontrée lorsque le jeune adulte s'installe dans une colonie – soient des prérequis pour un apprentissage correct.

Il apparaît ainsi qu'une séparation physique, et même une séparation sociale, peuvent altérer l'apprentissage : une incapacité à produire ou bien l'absence d'attention envers le modèle peuvent-elles expliquer cette production altérée ? À force de ne pas écouter, le jeune animal ne finit-il pas par ne pas « entendre » ?

De l'écoute à l'apprentissage : les bases cérébrales de la perception du chant

Le chant d'oiseau et son apprentissage ont suscité un grand intérêt dans les dernières décennies, constituant ainsi un modèle privilégié de neuroéthologie. Les travaux pionniers de Nottebohm et son équipe, ame-

25. Poirier *et al.* (2004).
26. Bertin *et al.* (2007), sous presses.

nant à la première découverte d'une neurogenèse chez l'adulte[27], ont également posé les bases de ce qui est communément appelé le « système de chant ». Par des études successives basées sur la neuroanatomie et sur des lésions cérébrales, cette équipe a montré l'existence d'un ensemble de noyaux cérébraux impliqués dans la production et la perception du chant (Figure 3).

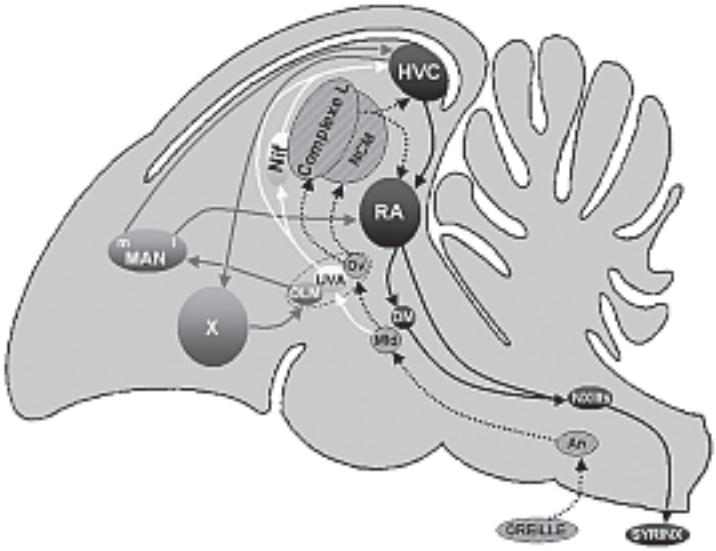


Figure 3 : *Schéma d'une vue latérale gauche de l'encéphale d'oiseau représentant les voies auditives (en hachuré), motrices (en noir et blanc) et d'apprentissage du chant (en gris). La voie auditive qui prend naissance dans l'oreille aboutit, après trois relais synaptiques dans le Complexe L (homologue du cortex auditif primaire). Les voies de la commande vocale partant du HVC aboutissent dans la syrinx (analogue du larynx), La voie d'apprentissage qui part du HVC est composée de circuits reliant localement des noyaux télencéphaliques et thalamiques.*

Sur le plan perceptuel, la zone auditive majeure, le complexe L, correspond à l'aire auditive primaire des mammifères. Des travaux antérieurs avaient pu montrer qu'elle présente de la même façon une tonotopie – c'est-à-dire que les fréquences y sont traitées selon un gradient spatial[28] –, des sélectivités envers des paramètres clés de chants de l'espèce[29], et une ouver-

27. Nottebohm (1981).
28. Leppelsack (1974).
29. Hausberger *et al.* (2000).

ture à l'expérience : les oiseaux dépourvus d'expérience auditive avec des chants d'adultes pendant la première année de développement ne présentent pas de tonotopie ni de sélectivité envers des chants de l'espèce[30], parallèle vocal à ce qui avait été démontré dans le système visuel[31].

De façon remarquable, la privation sociale (jeunes élevés seuls ou en paires, sans contact direct avec un adulte, *cf.* plus haut) peut avoir à elle seule le même effet qu'une privation sensorielle : les jeunes sans contact avec un adulte présentent une aire auditive ayant les mêmes caractéristiques que celles de jeunes oiseaux n'ayant jamais entendu un chant d'adulte.

Même les jeunes élevés avec des adultes (plusieurs jeunes avec un adulte) présentent des caractéristiques différentes d'un animal sauvage, avec en particulier moins de spécialisation neuronale envers les chants de l'espèce. Nous avons alors émis l'hypothèse que la préférence sociale active des jeunes envers leurs pairs de même âge pouvait avoir diminué leur attention envers le chant adulte et, par conséquent, par manque d'attention sélective envers lui, amené à une perte de sélectivité envers les chants de l'espèce[32]. Nous nous sommes alors intéressés à de jeunes femelles avec, comme seuls tuteurs, des mâles adultes. Nous savions que les femelles, hors de la période de reproduction, tendent à rester entre elles et n'apprennent pas des mâles. En plaçant de jeunes femelles avec des mâles, nous assurions une ségrégation sociale : les jeunes femelles tendaient à rester entre elles et n'apprenaient pas le chant du mâle adulte présent. Comme prévu, les jeunes femelles sont restées entre elles et n'ont que très peu appris du mâle. De plus, leur aire auditive apparaît aussi peu spécialisée, et conforme à celle d'animaux élevés en séparation physique de tout adulte[33].

Conclusion

Ce résultat remarquable permet de constater qu'une simple ségrégation sociale peut suffire à créer une perturbation du développement d'une aire sensorielle primaire, ouvrant de nouveaux champs d'investigation insoupçonnés. Comment une stimulation sociale peut-elle à elle seule

30. Cousillas *et al.* (2004).
31. Hubel et Wiesel (1963).
32. Puel, Bonfils et Pujol (1988).
33. Cousillas *et al.* (2008).

modifier la mise en place des caractéristiques d'une aire sensorielle primaire ?

La récente découverte de neurones auditifs sensibles à des images diffusées en même temps que le son pourrait amener à des éléments de compréhension et souligner l'importance du caractère multimodal d'un partenaire social[34]. Attention sélective et multimodalité pourraient bien être les clés de la compréhension de cette transmission du chant.

L'attention sélective, basée sur la stimulation intermodale et socialement pertinente, pourrait être une clé de compréhension du lien entre « cerveau social » et « cerveau vocal » chez l'oiseau et chez l'homme[35]. Ces éléments soulignent ainsi un parallèle de plus entre l'apprentissage du chant chez l'oiseau et le développement du langage humain[36].

RÉFÉRENCES BIBLIOGRAPHIQUES

– Baptista L. F. et Gaunt S. L. L. (1997), « Social interaction and vocal development in birds », *in* C. T. Snowdon et M. Hausberger (éd.), *Social Influences on Vocal Development*, Cambridge (G.-B.), Cambridge University Press, p. 23-40.

– Baptista L. F. et Petrinovivh L. (1984), « Social interaction, sensitive phases and the song template hypothesis in the white-crowned sparrow », *Animal Behaviour,* 32*,* p. 172-181.

– Baptista L. F. et Petrinovitch L. (1986) « Song development in the white-crowned sparrow ; social factors and sex differences », *Animal Behaviour,* 35*,* p. 1359-1371.

– Bertin A., Hausberger M., Henry L. et Richard-Yris M. (2007), « Adult and peer influences on starling song development », *Developmental Psychobiology*, 49, p. 362-374.

– Bertin A., Henry L., Richard-Yris M.-A. et Hausberger M. (sous presses). « Adult/ratio influences song acquisition in female European starlings (*Sturnus vulgaris*) », *Journal of Comparative Psychology*.

---

34. George *et al.* (soumis).
35. Kuhl (2003).
36. Nous remercions C. Aubry, S. Barbu, A. Bertin, C. Lunel, C. Petton, C. Poirier et F. Rousseau pour l'aide apportée à ces études.

Nos recherches ont été financées par les programmes suivants : European Training Program (Fondation pour la Science) (1991-1992). « Whistled song of the starling : ethological and neurophysiological aspects of perception ». — Programme Cognisciences (1991-1993). « Émission, perception et prise de signification des signaux acoustiques chez l'oiseau : rôle de l'apprentissage ». — ACI du Ministère « Cognitique » (1999-2002). « Perception, plasticité et vie sociale : perturbations et récupération du traitement de l'information auditive. Perspective comparative : animal / humain ». — Programme CNRS OHLL « Origine de l'Homme du Langage et des Langues » (2000-2005). « Vie sociale et communication : Un lien clé dans l'évolution du langage ? Une analyse comparative chez l'homme et l'animal ».

– Brémond J.-C. (1968), « Recherche sur la sémantique et les éléments vecteurs d'information dans les signaux acoustiques du rouge-gorge (*Erithacus rubecula*) », *Terre et Vie,* p. 109-220.
– Calame-Griaule G. (1965), *Ethnologie et langage. La parole chez les Dogons*, Paris, Gallimard.
– Catchpole C. K. et Slater P. J. B. (1995), « Bird song : biological themes and variations », Cambridge (G.-B.), Cambridge University Press.
– Cousillas H., Richard J.-P., Mathelier M., Henry L., George I. et Hausberger M. (2004), « Experience-dependent neuronal specialization and functional organization in the field L complex of European starlings », *European Journal of Neuroscience,* 19*,* p. 3343-3352.
– Cousillas H., George I., Henry L., Richard J.-P. et Hausberger M. (2008), « Linking social and vocal brains : Could social segregation prevent a proper development of a central auditory area in a female songbird ? », *PLoS ONE,* 3 (5), e2203.
– Elowson A. M. et Snowdon C. T. (1994), « Pygmy marmosets, *Cebuella pygmaea*, modify vocal structure in response to changed social environment », *Animal Behaviour,* 47*,* p. 1267-1277.
– Feare C. (1984), *The Starling*, New York, Oxford University Press.
– Gaffiot F. (1934), *Dictionnaire latin-français*, Paris, Hachette.
– Gheerbrant A. et Chevalier J. (1997), *Dictionnaire des symboles. Mythes, rêves, coutumes, gestes, formes, figures, couleurs, nombres*, Paris, Robert Laffont.
– Hartshorne C. (1973), *Born to Sing. An Interpretation and World Survey of Bird Song.* Bloomington, Indiana University Press.
– Hausberger M., Bigot E. et Clergeau P. (2008), « Dialect use in large assemblies : A study in European starling roosts », *Journal of Avian Biology*, 39*,* p. 672-682.
– Hausberger M., Cousillas H., George I. et Henry L. (2002), « Les bases neurobiologiques du chant des oiseaux », *Pour la science, « La communication animale »*, 34*,* p. 68-74.
– Hausberger M., Leppelsack E., Richard J.-P. et Leppelsack H. J. (2000), « Neuronal bases of categorization in starling song », *Behavioural Brain Research*, 114*,* p. 89-95.
– Hausberger M., Richard-Yris M.A., Henry L., Lepage L. et Schmidt I. (1995), « Song sharing reflects the social organization in a captive group of European starlings (*Sturnus vulgaris*) », *Journal of Comparative Psychology,* 109*,* p. 222-241.
– Hubel D. H. et Wiesel T. N. (1963), « Receptive fields of cells in striate cortex of very young, visually inexperienced kittens », *Journal of Neurophysiology,* 26*,* p. 994-1002.
– Kuhl P. K. (2003), « Human speech and birdsong : Communication and the social brain », *Proceedings of the National Academy of Sciences of the USA,* 100*,* p. 9645-9646.
– Lemasson A., Zuberbühler K. et Hausberger M. (2005), « Socially meaningful vocal plasticity in Campbell's monkeys », *Journal of Comparative Psychology,* 119, p. 220-229.

– Leppelsack H. J. (1974), « Funktionelle eigenschaften der höhrbahn im feld L des neostriatum caudale des staren (*Sturnus vulgaris*) », *Journal of Comparative Physiology,* 88, p. 271-320.
– Marler P. (1970), « Bird song and speech development : Could there be parallels ? », *American Scientist,* 58, p. 669-673.
– Nottebohm F. (1968), « Auditory experience and song development in the chaffinch *Fringilla coeleb* », *Ibis,* 110, p. 549-568.
– Nottebohm F. (1981), « A brain for all seasons : Cyclical anatomical changes in song control nuclei of the canary brain », *Science,* 214, p. 1368-1370.
– Oller D. K. et Griebel U. (2008), « Contextual flexibility in infant vocal development and the earliest steps in the evolution of language », *in* D. K. Oller et U. Griebel (éd.), *Evolution of Communicative Flexibility : Complexity, Creativity, and Adaptability in Human and Animal Communication*, Cambridge (Mass.), The MIT Press, p. 141-168.
– Pepperberg I. M. (1997), « Social influences on the acquisition of human-based codes in parrots and nonhuman primates », *in* C. T. Snowdon et M. Hausberger (éd.), *Social Influences on Vocal Development*, Cambridge, Cambridge University Press, p. 157-177.
– Poirier C., Henry L., Mathelier M., Lumineau S., Cousillas H. et Hausberger M. (2004), « Direct social contacts override auditory information in the song-learning process in starlings (*Sturnus vulgaris*) », *Journal of Comparative Psychology,* 118, p. 179-193.
– Puel J.-L., Bonfils P. et Pujol R. (1988), « Selective attention modifies the active micromechanical properties of the cochlea », *Brain Research,* 447, p. 380-383.
– Snowdon C. T. et Elowson A. M. (1999), « Pygmy marmosets modify call structure when paired », *Ethology,* 105, p. 893-904.
– Snowdon C. T. et Hausberger M. (éd.) (1997), *Social Influences on Vocal Development*, Cambridge (G.-B.), Cambridge University Press.
– Thorpe W. H. (1958), « The learning of song patterns by birds, with especial reference to the song of the chaffinch *Fringilla coelebs* », *Ibis*, 100, p. 535-570.
– Thorpe W. H. (1961), *Bird-Song. The Biology of Vocal Communication and Expression in birds*, Cambridge (G.-B.), Cambridge University Press.
– West M. J., Stroud A. N. et King A. P. (1983), « Mimicry of the human voice by European starlings (*Sturnus vulgaris*) : the role of social interaction », *Wilson Bulletin,* 95, p. 635-640.
– West M. J. et King A. P. (1990), « Mozart's starling », *American Scientist,* 78, p. 106-114.
– Zann R. (1997), « Vocal learning in wild and domesticated zebra finches : Signature cues for kin recognition or epiphenomena ? », *in* C. T. Snowdon et M. Hausberger (éd.), *Social Influences on Vocal Development*, Cambridge (G.-B.), Cambridge University Press, p. 85-97.
– Zuberbühler K. (2000), « Referential labelling in Diana monkeys », *Animal Behaviour,* 59, p. 917-927.

À l'origine du langage chez le nourrisson

par Ghislaine Dehaene-Lambertz

Tous ceux qui ont lutté pour apprendre une seconde langue ne peuvent manquer de s'émerveiller de la facilité avec laquelle les enfants apprennent leur langue maternelle. Alors que vers 3 ans, ils courent encore avec maladresse, ne maîtrisent pas le vélo, ils tiennent déjà de petits discours qui émerveillent leurs parents. À cet âge, ils ont compris que la parole sert à transmettre de l'information, et que ce signal acoustique continu peut être décomposé en briques élémentaires qui peuvent se recombiner de multiples façons pour créer de nouveaux messages. Ils exploitent cette propriété combinatoire avec efficacité, produisant des phrases qu'ils n'ont jamais entendues comme « ils sontaient partis ». De telles phrases, bien que non grammaticales, prouvent que les enfants de cet âge ne sont pas de simples perroquets mais ont la capacité d'analyser ce qu'ils entendent, qu'ils ont découvert les règles particulières qui permettent de former mots et phrases dans leur langue maternelle, et qu'ils peuvent judicieusement exploiter ces règles pour créer leurs propres messages (ici, prendre la forme du verbe au présent et ajouter la terminaison « aient » pour exprimer le passé, ce qui est plus logique que d'utiliser la forme irrégulière « étaient »). Comment les enfants arrivent-ils si vite et si efficacement à apprendre leur langue maternelle ? Les recherches en développement et en linguistique se sont surtout intéressées ces dernières années à caractériser le stimulus, c'est-à-dire à définir les propriétés formelles du langage et les algorithmes d'apprentissage que l'on peut supposer présents chez l'enfant pour lui permettre d'apprendre de telles propriétés.

Les progrès en imagerie cérébrale permettent désormais d'examiner « l'apprenti » lui-même et de s'interroger sur une organisation particulière du cerveau humain, notamment dans les régions périsylviennes gauches, qui faciliterait cet apprentissage. En effet, depuis la publication initiale de

Broca[1], en 1861, rapportant l'aphasie de M. Leborgne à la lésion qu'il présentait dans la région frontale gauche, de nombreuses études en neuropsychologie et en neuro-imagerie ont confirmé la relation privilégiée entre langage et régions périsylviennes gauches chez la plupart des adultes. Mais pourquoi le traitement du langage repose-t-il sur ces régions particulières ? Possèdent-elles des propriétés singulières qui pourraient expliquer l'émergence du langage dans notre espèce ? Pourquoi sont-elles situées dans l'hémisphère gauche chez la majorité des humains, quelles que soient la langue et la culture ? Ce biais en faveur de l'hémisphère gauche témoigne-t-il d'une évolution génétique modifiant l'organisation de la région périsylvienne de cet hémisphère et favorisant ce mode de communication, ou n'est-il que la conséquence de l'exposition prolongée et de la manipulation experte d'un stimulus auditif aux caractéristiques très particulières ? Pour tenter de répondre à ces questions et évaluer ce qui fait la spécificité du langage dans notre espèce, l'étude du nourrisson avant une exposition intense à la parole est indispensable.

D'où partent les nouveau-nés ?

Bien que la production verbale des enfants ne devienne conséquente qu'à la fin de la première année de vie, les nourrissons présentent dès la naissance des capacités perceptives sophistiquées, qui se modifient rapidement sous l'influence de la langue de leur entourage. Les nouveau-nés peuvent ainsi discriminer n'importe quel couple de langues dès lors qu'elles appartiennent à des familles rythmiques différentes[2]. Ils reconnaissent la voix de leur mère[3], mais même quand ils ne connaissent pas le locuteur, ils préfèrent écouter leur langue maternelle[4]. À 2 mois, ils s'orientent plus vite vers un haut-parleur s'il diffuse leur langue maternelle que s'il diffuse une langue étrangère[5]. Vers 4 mois, ils deviennent capables de discriminer deux langues proches appartenant à la même classe rythmique, comme le catalan et l'espagnol[6], ou l'anglais américain et l'anglais britannique[7]. Sur le plan segmental, ils discriminent les pho-

1. Broca (1861).
2. Mehler *et al.* (1988) ; Nazzi, Bertoncini et Mehler (1998).
3. Mehler, Bertoncini, Barrière et Jassik-Gerschenfeld (1978).
4. Mehler *et al.* (1988) ; Moon, Cooper et Fifer (1993).
5. Dehaene-Lambertz et Houston (1998).
6. Bosch et Sebastian-Galles (1997).
7. Nazzi, Jusczyk et Johnson (2000).

nèmes de façon catégorielle, même ceux n'appartenant pas à leur langue maternelle[8]. Ils négligent aisément les variations acoustiques non pertinentes, comme les différences de voix ou les différences d'intonation, pour extraire le segment phonétique[9]. Ceci n'est évidemment pas dû au fait qu'ils ne sont pas capables de percevoir des différences de voix, puisqu'ils reconnaissent la voix de leur mère dès les premiers jours de vie, et discriminent deux voix inconnues[10].

Ces capacités initiales sont rapidement modifiées par la langue de l'entourage et, à la fin de la première année de vie, les nourrissons ont acquis le répertoire phonétique de leur langue et les règles qui gouvernent les combinaisons de phonèmes dans les mots de cette langue[11]. Leur babillage est marqué par la langue maternelle, rendant facilement identifiables les petits Français, Cantonnais ou Arabes de 8 mois[12]. Vers la même époque, ils sont capables d'extraire des mots de la parole continue, et associent les mots les plus fréquents avec des objets. Ceci n'est pas une performance banale, puisque moins de 7 % de la parole adressée au nourrisson est constituée de mots isolés[13], et que les langues orales manquent de frontières de mots systématiques et évidentes comme le blanc qui sépare les mots écrits. De plus, les mêmes mots peuvent avoir des réalisations acoustiques très différentes en fonction des idiosyncrasies du tract vocal du locuteur, de ses émotions, de son débit, du bruit dans l'environnement, etc. Malgré toutes ces difficultés, vers un an (et sans doute avant), les nourrissons ont compris que le bruit que les autres humains produisent avec leur bouche transmet de l'information, grâce à la combinaison de mots qui doivent être décodés. D'un point de vue développemental, ce succès doit être comparé avec leur incapacité à marcher seul.

Quel est le rôle de l'apprentissage prénatal ?

Les performances du nouveau-né ne reflètent pas vraiment l'état initial du cerveau humain avant toute exposition à la parole, puisque l'audition se développe pendant le dernier trimestre de la grossesse. Comment

8. Werker et Tees (1984).
9. Kuhl (1983).
10. Dehaene-Lambertz (2000).
11. Jusczyk, Luce et Charles-Luce (1994) ; Werker et Tees (1984).
12. Boysson-Bardies, Sagart et Durand (1984).
13. Van De Weijer (1998).

alors séparer ce qui est dû à une exposition depuis plusieurs semaines de ce qui pourrait être dû à un déterminisme génétique favorisant le développement du langage dans l'espèce humaine ?

Effectivement, la voix de la mère est très largement audible au-dessus des bruits de l'environnement utérin (flux artériel dans l'artère placentaire, bruits du cœur de la mère, bruits intestinaux, etc.) grâce à la transmission directe des vibrations de la voix maternelle à l'oreille du fœtus. Les bruits extérieurs, dont la voix du père, sont eux plus distants et considérablement atténués par les tissus maternels et le liquide amniotique[14], et donc peu audibles par le fœtus. L'univers auditif du fœtus est donc centré sur la voix de la mère. Les modulations de la voix, liées aux émotions, sont même directement associées pour le fœtus aux modifications hormonales et du rythme cardiaque maternel, qu'il peut directement ressentir. Si la seule exposition prénatale expliquait les compétences néonatales, on s'attendrait à un tout autre comportement du nouveau-né. Premièrement, la prépondérance d'une voix unique, celle de la mère, dans l'environnement auditif du fœtus devrait favoriser une représentation précise des productions maternelles et des difficultés à généraliser à d'autres locuteurs. Deuxièmement, la perception de certains contrastes phonétiques, comme la place d'articulation, est très dégradée dans le bruit. À partir d'enregistrements chez la brebis gravide, Griffiths et ses collaborateurs[15] ont montré que l'intelligibilité de ces phonèmes était médiocre dans les conditions d'écoute fœtale. Pourtant, les nouveau-nés n'ont pas de difficulté pour discriminer des phonèmes qui diffèrent sur la place d'articulation comme /pa/ et /ta/[16]. Les nourrissons sont enfin capables de discriminer des contrastes non présents dans leur environnement linguistique[17]. Ainsi des nourrissons anglophones discriminent des contrastes hindi, que leurs parents ne perçoivent ni ne produisent.

Bien sûr, les possibilités d'apprentissage ne démarrent pas brutalement à la naissance, et les études chez les nouveau-nés montrent que ceux-ci ont mémorisé des caractéristiques de leur environnement auditif pendant la vie fœtale. Ils reconnaissent la voix de leur mère[18], préfèrent écouter leur langue maternelle[19], et réagissent particulièrement à une histoire que la mère a lue de façon répétée pendant les dernières semaines de

14. Busnel et Granier-Deferre (1983) ; Querleu, Renard, Versyp, Paris-Delrue, et Crépin (1988).
15. Griffiths *et al.* (1994).
16. Dehaene-Lambertz et Pena (2001).
17. Werker et Tees (1984).
18. DeCasper et Fifer (1980).
19. Mehler *et al.* (1988).

grossesse[20]. Néanmoins, leurs performances linguistiques à la naissance excèdent ce qu'ils auraient pu apprendre par la seule exposition à la parole *in utero*, et suggèrent une organisation particulière du cerveau humain.

Le cerveau humain est asymétrique

Quelle pourrait être cette organisation ? Pour découvrir quelques pistes, revenons brièvement sur les particularités structurales et fonctionnelles des régions linguistiques chez l'adulte. Chez l'adulte, les régions impliquées dans le traitement de la parole occupent les berges de la scissure de Sylvius, ce sillon profond qui sépare régions temporales et frontales à l'avant, temporales et pariétales à l'arrière. Les caractéristiques anatomiques de ces régions ne sont pas symétriques. Généralement, la vallée sylvienne est plus longue et le *planum temporale* plus étendu à gauche[21]. La substance blanche sous le gyrus de Heschl, région auditive primaire, occupe un plus grand volume à gauche[22]. Moins fréquemment la région frontale gauche est également plus large que la droite[23]. À l'échelle microscopique, les cellules pyramidales sont plus grandes dans le cortex auditif gauche[24] et sont associées à des fibres plus myélinisées[25]. La largeur des colonnes de neurones et la distance entre ces colonnes sont plus importantes dans le lobe temporal supérieur gauche[26]. Toutes ces caractéristiques structurales permettraient à l'hémisphère gauche de coder efficacement les transitions acoustiques rapides et complexes qui caractérisent la parole[27]. La différence entre /ba/ et /da/ par exemple ne porte que sur les 40 premières millisecondes du signal.

Pour expliquer l'avantage gauche dans le traitement de la parole chez l'adulte, deux hypothèses s'opposent. La première postule une asymétrie dans le traitement auditif. À gauche se font les traitements demandant une grande précision temporelle, et à droite, ceux demandant une bonne représentation spectrale ; donc la parole est traitée par l'hémisphère gauche car beaucoup des contrastes phonémiques (mais pas tous) nécessi-

20. DeCasper et Spence (1986).
21. Geschwind et Levitsky, 1968.
22. Penhune, Zatorre, MacDonald et Evans (1996).
23. Knaus (sous presses).
24. Hutsler (2003).
25. Anderson, Southernet et Powers (1999).
26. Seldon (1981).
27. Boemio, Fromm, Braun et Poeppel (2005) ; Zatorre et Belin (2001).

tent une analyse temporelle fine pour être perçus. La seconde hypothèse insiste sur le fait que c'est le caractère linguistique ou non des stimuli qui détermine la latéralisation de leur traitement, et non les propriétés acoustiques de la parole.

L'hypothèse selon laquelle les propriétés structurales des régions temporales gauches sont déterminantes pour expliquer que le traitement de la parole s'effectue à gauche est d'autant plus attirante que dans les groupes d'enfants dyslexiques ou dysphasiques, qui ont plus de difficultés à discriminer les phonèmes contenant des variations rapides, comme /ba/ et /da/, les asymétries comme celle du *planum temporale* sont moins importantes[28] que dans les groupes d'enfants témoins. Mais bien que ces caractéristiques structurales semblent en effet adaptées au traitement d'un stimulus variant rapidement comme la parole, l'hypothèse d'une simple relation causale entre leur présence et l'émergence du langage est sans doute simpliste, pour les raisons suivantes. Il convient d'abord de rappeler que les humains ont un entraînement intense avec la parole, ce qui rend difficile de savoir si les asymétries observées chez l'adulte sont la cause ou la conséquence du traitement privilégié de la parole dans l'hémisphère gauche. Ensuite, les réponses neurales de ces régions dépendent plus de la valeur linguistique des stimuli que de leurs caractéristiques acoustiques. Par exemple, dans deux études utilisant l'imagerie par résonance magnétique (IRM) fonctionnelle, l'asymétrie des activations observées dans la région temporale postérieure et la région pariétale inférieure n'étaient observées que quand les stimuli étaient pertinents pour les sujets sur un plan linguistique, mais s'atténuaient lorsque les sujets considéraient les stimuli comme des sifflements[29] ou lorsqu'ils n'appartenaient pas à leur langue maternelle[30]. Enfin et surtout, les langues de signes, qui reposent sur des indices spatiaux et non sur des indices temporels rapides comme les langues orales, activent les mêmes régions périsylviennes gauches que les langues orales[31].

28. Plante (1991).
29. Dehaene-Lambertz, Dupoux et Gout (2000).
30. Jacquemot, Pallier, LeBihan, Dehaene et Dupoux (2003).
31. Damasio, Bellugi, Poizner, et Gilder (1986) ; Sakai, Tatsuno, Suzuki, Kimura et Ichida (2005).

Asymétries cérébrales chez l'animal

Les asymétries hémisphériques sont-elles si rares dans les autres espèces qu'il faille relier l'émergence du langage à l'émergence d'une asymétrie hémisphérique ? Même si les asymétries à l'échelle macroscopique constatées chez certains animaux sont assurément plus discrètes que chez l'homme, les grands singes ont un *planum temporale* plus étendu à gauche[32]. Les différences cytoarchitectoniques sont également moins importantes que chez l'humain[33] et la latéralisation de l'organisation des minicolonnes observées dans le cortex temporal humain n'a été retrouvée ni chez le singe rhésus ni chez le chimpanzé[34].

On retrouve également chez l'animal des asymétries fonctionnelles. Des réponses électrophysiologiques asymétriques ont été enregistrées dans le thalamus de cochons d'Inde en réponse à des sons complexes, comme la syllabe /da/, alors que les réponses sont symétriques pour des sons simples[35]. Des singes rhésus sauvages[36], des otaries[37], des souris[38] et des aigles harpies[39] orientent leur oreille droite, et donc favorisent un traitement par l'hémisphère gauche, vers un haut-parleur jouant des vocalisations de leur propre espèce. Des macaques perdent la capacité à discriminer deux formes de leurs vocalisations après une lésion temporale supérieure gauche mais pas après une lésion similaire à droite[40]. Cette asymétrie fonctionnelle a été récemment confirmée par une étude en tomographie par émission de positons (TEP) montrant des activations du pôle temporal gauche quand des singes macaques écoutent des vocalisations de leur propre espèce[41]. Comme chez les humains, cette latéralisation semble plus liée à la pertinence de ces signaux dans la communication qu'aux caractéristiques acoustiques du signal[42]. Cette latéralisation semble se développer

32. Cantalupo, Pilcher et Hopkins (2003) ; Gannon, Holloway, Broadfield et Braun (1998).
33. Buxhoeveden, Switala, Litaker, Roy et Casanova (2001).
34. Buxhoeveden *et al.* (2001).
35. King, Nicol, McGee et Kraus (1999).
36. Hauser et Andersson (1994).
37. Boye, Gunturkun et Vauclair (2005).
38. Ehret (1987).
39. Palleroni et Hauser (2003).
40. Heffner et Heffner (1984).
41. Poremba *et al.* (2004).
42. Ehret (1987) ; Petersen *et al.* (1984).

progressivement, car elle n'est pas observée chez les bébés rhésus[43] ou otaries[44]. Chez les aigles harpies, c'est l'expérience active de la chasse qui modifie le biais initial d'orientation de l'oreille gauche vers le cri de la proie en un biais inverse[45]. Bien que la latéralisation des réponses auditives apparaisse donc comme un phénomène moins inhabituel que nous le pensions autrefois, les raisons de cette latéralisation, et surtout les parts respectives de l'analyse acoustique des sons et de la valeur sémantique des stimuli dans cette latéralisation sont encore mal connues, même chez l'animal.

Les animaux et la parole

Si les asymétries structurales ne sont pas l'apanage de l'espèce humaine, y a-t-il néanmoins une spécificité humaine du traitement de la parole ? Une des caractéristiques de la perception phonétique chez l'humain est qu'elle est catégorielle, et est normalisée à travers différentes productions. Ceci est vrai chez l'adulte et chez le nourrisson. Des propriétés similaires ont été décrites chez l'animal. Le singe macaque, le chinchilla, et même certains oiseaux comme la caille[46] peuvent être entraînés à discriminer des syllabes de façon catégorielle et sont capables de généraliser cet apprentissage à de nouveaux exemplaires jamais entendus. Néanmoins, en plus de différences évidentes entre espèces dans la durée nécessaire pour entraîner les animaux à ces discriminations, ces derniers n'utilisent pas forcément les mêmes indices que les humains (tout au moins que des adultes humains) pour réussir la tâche[47]. Par exemple, les singes, mais pas les humains, ont plus de difficultés à discriminer /b-d/ quand ces consonnes sont suivies des voyelles /i/ et /e/ que lorsqu'elles sont suivies des voyelles /a/ et /u/[48]. Les performances des singes peuvent être expliquées par des mécanismes auditifs généraux analysant la direction du deuxième formant. Cette direction est clairement différente pour /ba/ et /da/ mais cet indice est beaucoup plus ambigu dans /bi-di/ à cause de la coarticulation avec la voyelle /i/. Une différence similaire entre un

43. Hauser et Andersson (1994).
44. Boye *et al.* (2005).
45. Palleroni et Hauser (2003).
46. Kuhl et Padden (1983) ; Kuhl et Miller (1975) ; Kluender, Diehl, et Killeen (1987).
47. Kuhl (1991) ; Sinnott (2005).
48. Sinnott et Gilmore (2004).

codage auditif général et un codage spécifique au traitement linguistique peut être mise en évidence chez l'être humain quand on utilise des stimuli synthétiques, analogues sinusoïdaux de la parole. Ces stimuli peuvent être perçus soit comme des bruits électroniques, soit comme de la parole. Si l'on utilise un continuum de sons variant de /ba/ à /da/, la localisation de la frontière de catégorie n'est pas identique mais diffère d'un pas quand les stimuli sont perçus comme des bruits et quand ils sont perçus comme de la parole[49]. De plus, lorsque ces sons sont perçus comme de la parole, on observe une augmentation significative de l'activation dans la partie postérieure du sillon temporal supérieur et dans le gyrus supramarginal[50]. Ces résultats suggèrent que chez l'humain, au moins dans le cerveau adulte, les représentations phonétiques sont distinctes des représentations acoustiques.

Plus récemment, des études comparatives ont examiné l'autre capacité linguistique observée précocement chez le nourrisson, la capacité à reconnaître que des phrases proviennent de langues différentes. Des rats[51] entraînés avec des phrases en japonais et en hollandais, naturelles ou synthétisées pour ne conserver que l'information prosodique, réussissent à généraliser cet apprentissage à de nouvelles phrases. Néanmoins, contrairement au nouveau-né humain, ils n'y arrivent pas lorsque les nouvelles phrases sont produites par de nouveaux locuteurs jamais entendus auparavant. Les singes tamarins n'ont pas besoin d'entraînement, et discriminent spontanément ces deux langues humaines même lorsque les phrases sont produites par des locuteurs variés et inconnus[52]. Ces capacités à discriminer des langues humaines sont néanmoins limitées aux langues appartenant à des familles rythmiques différentes, par exemple hollandais et japonais, polonais et japonais. Les tamarins ne peuvent pas discriminer l'anglais et le hollandais, deux langues rythmiquement proches. Les deux espèces, comme les humains d'ailleurs, échouent également quand les phrases sont jouées à l'envers. Le traitement différentiel des informations rythmiques de la parole est donc possible chez d'autres mammifères, et ne se produit que sur des phrases qui respectent le déroulement temporel naturel de la parole humaine. Les rats sont moins performants que les tamarins pour extraire les invariants de la langue lorsque plusieurs locuteurs, avec leurs particularités, produisent les phrases, mais les tamarins eux, se rapprochent de ce qui est observé chez les nouveau-nés humains.

---

49. Serniclaes, Sprenger-Charolles, Carre et Demonet (2001).
50. Dehaene-Lambertz *et al.* (2005).
51. Toro, Trobalon et Sebastian-Galles (2005).
52. Ramus, Hauser, Miller, Morris et Mehler (2000) ; Tincoff *et al.* (2005).

Où se situent les nourrissons ?

Les capacités des nourrissons ne semblent donc pas si différentes initialement de celles des autres mammifères, mais l'évolution rapide de ces capacités en fonction de la langue maternelle dès les premières semaines de vie différencie le petit humain des autres animaux. Ces différentes trajectoires de développement sont-elles uniquement dues à l'exposition du nourrisson humain à la parole, ou les propriétés structurales et fonctionnelles du cerveau humain permettent-elles à l'enfant de tirer parti de cet environnement particulier ?

Asymétries structurales chez le nourrisson

L'asymétrie hémisphérique est évidente chez le fœtus humain car la plupart des sillons apparaissent d'abord à droite. Le sillon frontal supérieur, le gyrus temporal supérieur et le gyrus de Heschl sont détectables une à deux semaines plus tôt à droite qu'à gauche[53]. Cette asymétrie du développement de la gyration n'est pas rapportée chez le fœtus macaque[54]. À la naissance, la scissure sylvienne est plus longue à gauche, et est associée à un *planum temporale* plus étendu à gauche tandis que la surface du sillon temporal supérieur est plus importante à droite, toutes caractéristiques similaires à celles observées chez l'adulte humain[55]. Les études de jumeaux montrent une influence génétique forte sur le développement de ces régions[56], avec peu d'effet de la stimulation auditive, tout au moins à l'échelle macroscopique. Le *planum temporale* est en effet plus étendu et le gyrus de Heschl plus volumineux à gauche chez des adultes sourds qui ne se distinguent pas des normo-entendants sur ces aspects[57]. Dans l'espèce humaine, les hémisphères droit et gauche ont donc un développement extrêmement différencié dès la période fœtale, ce qui suggère un

53. Chi, Dooling et Gilles (1977) ; Dubois, Benders, Cachia *et al.* (2008).
54. Fukunishi *et al.* (2006).
55. Sowell *et al.* (2002) ; Chi *et al.* (1977) ; Witelson et Pallie (1973) ; Dubois, Benders, Borradori-Tolsa *et al.* (2008).
56. Thompson *et al.* (2001).
57. Emmorey, Allen, Bruss, Schenker et Damasio (2003).

déterminisme génétique. Effectivement, on observe que certains gènes ont une expression asymétrique, particulièrement dans l'espèce humaine[58]. *LMO4* par exemple, est systématiquement plus exprimé à droite qu'à gauche chez l'humain, alors que la différence en faveur d'un côté n'est pas systématique dans les populations de souris. Ce gène s'exprime à un stade précoce du développement fœtal, entre 12 et 14 semaines, une période critique dans la régionalisation corticale. D'autres gènes ont une expression asymétrique dans le cerveau fœtal, et présentent des différences de profil d'expression entre l'adulte humain et le chimpanzé. Leur profil d'expression pourrait avoir été l'objet d'une pression sélective récente, mais l'identification de ces gènes n'est pas terminée[59].

Asymétries fonctionnelles chez le nourrisson

Bien que les asymétries structurales soient particulièrement présentes dans les régions temporales supérieures, elles pourraient ne pas être reliées au développement du langage dans l'espèce humaine, mais simplement à un développement globalement asymétrique des deux hémisphères. Deux résultats suggèrent néanmoins que ces caractéristiques structurales sont liées au développement des asymétries fonctionnelles de l'adulte. L'imagerie du tenseur de diffusion permet d'obtenir des images des faisceaux de substance blanche. Comme les indices mesurés avec ce type d'imagerie sont sensibles à l'organisation du faisceau, par exemple à sa compacité et à sa myélinisation, il devient possible de suivre *in vivo* la maturation de ces faisceaux de substance blanche, qui transportent l'information aux centres de traitement. Une étude chez 23 nourrissons dans les premières semaines de vie a montré que la maturation est plus avancée à gauche pour deux faisceaux particulièrement intéressants car reliés à deux fonctions très latéralisées chez l'humain : le faisceau arqué, qui relie régions de compréhension et régions de production du langage, et le faisceau cortico-spinal, principal faisceau de la motricité volontaire[60]. Les asymétries dans ces faisceaux ne sont pas la conséquence d'une utilisation asymétrique puisqu'à l'âge testé, les nourrissons ont des capacités linguistiques et motrices limitées. Ces asymétries préexistent à la production du langage oral et à celle de gestes volontaires.

58. Sun, Collura, Ruvolo et Walsh (2006).
59. Sun *et al.* (2006).
60. Dubois, Hertz-Pannier *et al.* (2008).

Un deuxième élément vient des études d'imagerie fonctionnelle. Pendant la première année de vie, il n'y a pas de différence au repos dans le débit sanguin cérébral droit et gauche, même dans les régions linguistiques (régions frontale inférieure, temporale supérieure et temporo-pariétale plurimodale)[61]. En revanche, les activations observées en IRM ou par l'électroencéphalographie sont plus importantes à gauche en réponse à des stimuli auditifs. Des passages de 20 secondes de parole, ont été présentés normalement ou à l'envers à des nouveau-nés et à des nourrissons de 3 mois. Bien que la parole jouée à l'envers partage certaines caractéristiques avec la parole à l'endroit (présence de transitions rapides, information phonétique persistante dans le cas de phonèmes symétriques comme les voyelles), elle viole les règles universelles de l'organisation prosodique et les nourrissons ne savent plus discriminer deux langues lorsque les phrases sont jouées à l'envers[62]. Aux deux âges, les analyses statistiques montrent une asymétrie de l'activation en faveur de la gauche au niveau du *planum temporale* mais ne peuvent démontrer que cette asymétrie est plus importante pour la parole normale (à l'endroit) que pour la parole impossible (à l'envers)[63]. C'est donc la présence de variations temporelles rapides dans les deux types de stimuli qui pourrait être la cause de l'activation préférentielle de l'hémisphère gauche. Mais le faible nombre de participants dans ces deux études ne peut écarter que l'absence de différence entre les deux conditions soit éventuellement liée à un manque de puissance statistique car l'activation dans la région temporale postérieure gauche n'était significative que pour la parole à l'endroit, la parole à l'envers activant des régions plus antérieures et ventrales.

L'amplitude des potentiels évoqués auditifs enregistrés au-dessus de l'hémisphère gauche est également plus importante qu'au-dessus de l'hémisphère droit chez les nourrissons, mais cette différence d'amplitude est observée pour des syllabes et également pour des tons non verbaux[64]. Néanmoins, dans une étude récente portant sur la capacité des nourrissons de 2 mois à percevoir la relation entre un mouvement articulatoire vu et une voyelle entendue peu de temps après, ainsi que la relation entre le sexe du visage et celui de la voix, nous avons observé que le traitement linguistique et le traitement de l'identité de la personne avaient un biais hémisphérique différent. Le traitement linguistique était dévolu à l'hémisphère gauche, tandis que le traitement du genre de l'individu était

61. Chiron *et al.* (1997).
62. Mehler *et al.* (1988).
63. Pena *et al.* (2003) ; Dehaene-Lambertz, Dehaene et Hertz-Pannier (2002).
64. Dehaene-Lambertz (2000).

significativement plus latéralisé à droite[65]. Ces résultats ont donc mis en évidence, pour la première fois, l'existence d'une distinction fonctionnelle entre les deux hémisphères dès les premières semaines de vie.

En résumé, il existe des différences dans la maturation de l'hémisphère gauche par rapport à celle de l'hémisphère droit chez l'être humain, en particulier dans la partie postérieure du lobe temporal supérieur. Ces différences à l'échelle macroscopique s'accompagnent de différences dans l'organisation des faisceaux impliqués dans la manualité et le langage : le faisceau cortico-spinal et le faisceau arqué. Finalement, un avantage fonctionnel en faveur de l'hémisphère gauche est présent très précocement. Ce biais semble concerner l'ensemble des stimuli auditifs, même si les processus linguistiques semblent plus constamment et significativement plus latéralisés du côté gauche que les autres stimuli. Néanmoins, en cas de lésion, les régions contro-latérales périsylviennes droites peuvent assumer un traitement linguistique de bonne qualité (par exemple, de discrimination phonétique, et même un langage réceptif et expressif normal[66]). Plusieurs études ont ainsi souligné que le développement du langage est le plus souvent dans les limites normales chez des enfants ayant des lésions précoces, quel que soit le côté de la lésion[67]. Ces résultats montrent que la préférence hémisphérique gauche dans le traitement du langage n'est pas une contrainte stricte. La plasticité du cerveau permet, en présence d'une lésion précoce, une réorganisation le plus souvent suffisante du point de vue fonctionnel, même si on ignore quelles auraient été les performances de ces enfants en l'absence de la lésion. Notons que la récupération post-lésionnelle a souvent été utilisée comme argument en faveur d'une équipotentialité initiale des deux hémisphères, mais les résultats obtenus par l'imagerie cérébrale chez l'enfant normal vont clairement à l'encontre de cette hypothèse même si la latéralisation se consolide pendant le développement et l'acquisition de compétences linguistiques plus élaborées[68],

65. Bristow *et al.* (2009).
66. Dehaene-Lambertz (2004) ; Hertz-Pannier *et al.* (2002).
67. Bates et Roe (2001).
68. Holland *et al.* (2001).

Une continuité fonctionnelle entre le nourrisson et l'adulte : la perception phonétique

La latéralisation fonctionnelle n'est pas la seule caractéristique commune aux nourrissons et aux adultes. D'autres propriétés caractéristiques du réseau linguistique adulte sont déjà retrouvées chez le nourrisson dès les premières semaines de vie. Les potentiels électriques évoqués ont été utilisés pour comprendre comment des sons brefs sont traités par le cerveau. En soustrayant la réponse provoquée par un son précédé par lui-même ou par un autre son proche (par exemple da da da *da versus* ba ba ba *da*), on observe une réponse de discrimination dont la latence et la topographie dépendent des caractéristiques qui ont changé. Quand le changement dans une série de syllabes répétées concerne la voix ou l'identité des phonèmes, les réponses de discrimination ont une topographie et une latence différentes, ce qui suggère que différents réseaux fonctionnent en parallèle pour coder les différentes caractéristiques des sons (hauteur, intensité, durée, etc.) chez le nourrisson comme chez l'adulte[69]. Un de ces réseaux possède « des propriétés phonétiques », comme la normalisation à travers différents locuteurs et la perception catégorielle[70]. La modélisation dipolaire des régions actives lors de la réponse de discrimination d'un changement phonétique suggère l'implication de régions plus postérieures et dorsales que lors de la perception d'un changement acoustique similaire. Ce recul du dipôle est compatible avec l'implication des régions temporales postérieures et pariétales observée chez l'adulte uniquement lors d'un changement phonétique[71], mais non lors d'un changement acoustique équivalent.

Un réseau temporal hiérarchiquement organisé

Grâce à l'utilisation de l'IRM fonctionnelle chez les nourrissons, nous avons pu étudier la réponse cérébrale à des stimuli auditifs plus longs que des syllabes, et nous rapprocher des conditions naturelles

69. Bristow *et al.* (sous presse) ; Dehaene-Lambertz (2000).
70. Dehaene-Lambertz et Pena (2001) ; Dehaene-Lambertz et Baillet (1998).
71. Dehaene-Lambertz *et al.* (2005) ; Jacquemot *et al.* (2003).

d'écoute de la parole. La parole, à l'endroit et à l'envers, active l'ensemble des régions temporales supérieures, du gyrus de Heschl au pôle temporal[72] (voir figure 1). Le gyrus angulaire et le précuneus sont plus activés par la parole à l'endroit que par la parole à l'envers. Ces régions, qui sont impliquées dans le stockage lexical chez l'adulte, jouent donc sans doute un rôle similaire dans le stockage des contours prosodiques utilisés par le nourrisson pour reconnaître les phrases de sa langue maternelle, évidemment absents de la parole à l'envers. Finalement, des activations dans la région dorso-latérale droite sont observées chez les nourrissons éveillés qui écoutent leur langue maternelle. Chez l'adulte, cette région est impliquée dans les réseaux de récupération en mémoire. Les nourrissons réveillés, mais pas ceux qui dormaient pendant l'étude, mobilisent donc de façon active leur mémoire pour reconnaître, dans ce qu'ils écoutent, ce qui est leur langue maternelle. Cette première étude chez des enfants aussi jeunes a permis de montrer que, bien que les nourrissons soient moins compétents que des adultes pour traiter la parole, les activations retrouvées étaient proches dans les deux populations. Nous n'avons observé ni une activation diffuse de tout le cerveau, ni une activation limitée aux régions les plus matures à cet âge, mais bien l'activation d'un réseau organisé combinant des régions distinctes aux propriétés différentes. L'observation de l'activation de régions frontales a d'ailleurs été une surprise. En effet, du fait de l'immaturité notoire du lobe frontal et de sa maturation très lente jusqu'à la puberté, l'opinion commune était que ces régions n'intervenaient pas dans la cognition du nourrisson.

Dans une deuxième étude, nous avons fragmenté la zone périsylvienne activée en sous-régions fonctionnellement distinctes selon le délai de l'activation et la sensibilité à la répétition de la phrase[73]. En réponse à une simple phrase, le délai de l'activation augmente au fur et à mesure que l'on se déplace des régions auditives primaires vers la partie postérieure du gyrus temporal supérieur, d'une part, et vers le pôle temporal et la région frontale inférieure (aire de Broca), d'autre part. La différence observée entre les régions les plus lentement et les plus rapidement activées est de plusieurs secondes (figure 1), et ne peut donc pas refléter seulement des délais en rapport avec la transmission synaptique. Ce gradient d'activation pourrait être le résultat des différentes opérations cognitives permettant l'intégration des sons de parole dans des unités de plus en plus grandes, et sans doute plus abstraites, qui requièrent une durée plus longue de traitement et une activité plus soutenue. Cette architecture hiérar-

72. Dehaene-Lambertz *et al.* (2002).
73. Dehaene-Lambertz *et al.* (2006).

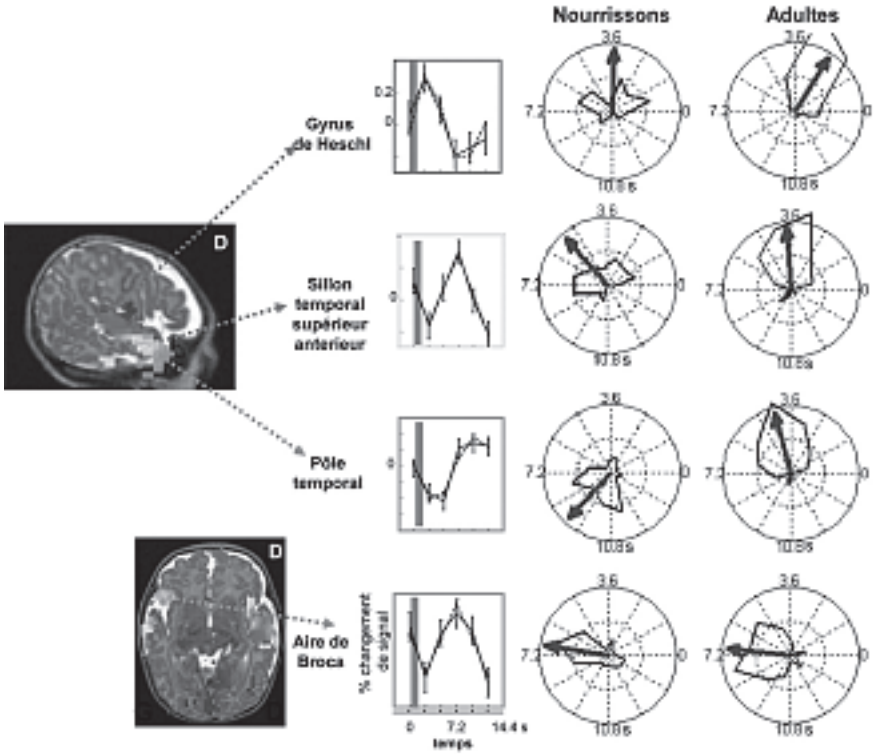


Figure 1 : *Décalage de la phase de la réponse dépendante du taux sanguin de dioxygène (en anglais,* blood-oxygen-level-dependent response *ou* BOLD response*) le long du sillon temporal supérieur, en IRM fonctionnelle. La phase de la réponse provoquée par l'audition d'une seule courte phrase a été mesurée chez des nourrissons de 3 mois et chez des adultes. On constate un décalage de phase croissant au fur et à mesure de la progression vers le pôle temporal, tandis que les réponses les plus rapides sont enregistrées dans le gyrus de Heschl. Bien que l'intervalle des phases soit plus large chez le nourrisson que chez l'adulte, un décalage similaire est observé dans les deux populations entre le gyrus de Heschl et les régions temporales les plus antérieures, comme le montre la projection des réponses mesurées dans les différentes régions d'intérêt sur un cercle figurant les 14,4 secondes qui s'écoulent entre le début des phrases successives. La région de Broca qui présente une réponse lente dans les deux populations est sans doute impliquée dans la mémorisation à court terme de la phrase entendue (Dehaene-Lambertz* et al.*, 2006a). La présence de ce gradient d'activation chez le très jeune enfant, avant le stade du babillage, et sa correspondance avec l'organisation corticale des projections anatomiques chez les autres primates (Kaas et Hackett, 2000 ; Pandya et Yeterian, 1990), suggèrent que cette organisation reflète une prédisposition innée, qui pourrait faciliter la découverte par l'enfant de la structure emboîtée du langage.*

chique du lobe temporal humain présente des analogies avec celle du cerveau des singes[74]. Il est donc possible que le réseau linguistique humain ait « recyclé » un système hiérarchique auditif préexistant chez les primates[75]. Une telle organisation d'unités emboîtées, avec des fenêtres d'intégration de plus en plus longues, fournirait ainsi au nourrisson un outil adapté de segmentation du flux de parole en ses unités prosodiques. En effet, les théories de l'apprentissage du langage se heurtent au problème de l'absence d'indices acoustiques non ambigus qui signaleraient comment découper la phrase. Sur un segment non limité dans le temps, le nombre d'analyses combinatoires que devrait théoriquement faire le nourrisson pour découvrir les régularités de sa langue devient rapidement énorme. Pour résoudre ce problème, certains auteurs[76] ont proposé que le nourrisson, du fait de sa mémoire et de ses capacités d'attention limitées, ne retiendrait et n'analyserait que de petits segments du discours. Une série de processeurs avec des fenêtres d'intégration de durées limitées est une autre solution qui permettrait une première segmentation. Cette segmentation initiale pourrait s'affiner progressivement, et s'adapter aux unités prosodiques propres à la langue maternelle grâce à l'analyse des variations acoustiques fréquemment rencontrées à ses bordures. À l'intérieur de ces unités bornées dans le temps, l'amélioration des capacités de mémoire auditive à long terme, qui serait, d'après Fritz et ses collaborateurs[77], une des différences essentielles entre l'humain et les autres primates, permettrait au nourrisson d'analyser et de mémoriser facilement les briques élémentaires que sont les syllabes. En effet, plusieurs études effectuées chez l'adulte ont montré que l'adjonction de frontières minimales permettait d'améliorer considérablement l'analyse de la structure de la parole présentée[78].

Enfin, nous avons observé des activations dans la région frontale inférieure gauche lorsque les nourrissons étaient engagés dans une tâche de mémorisation à court terme[79]. La parole est un stimulus mutimodal, qui est bien sûr auditif, mais aussi visuel, lorsque le nourrisson regarde sa mère qui parle, et moteur, lorsque lui-même s'essaye à vocaliser. La région frontale inférieure est une région de convergence des informations visuelles, auditives et motrices, ainsi qu'une région de planification. Elle

74. Kaas et Hackett (2000) ; Pandya et Yeterian (1990).
75. Dehaene et Cohen (2007).
76. Newport (1990).
77. Fritz *et al.* (2005).
78. Buiatti, Peña et Dehaene-Lambertz (2009) ; Pena, Bonatti, Nespor et Mehler (2002).
79. Bristow *et al.* (2009) ; Dehaene-Lambertz *et al.* (2006).

pourrait ainsi permettre au nourrisson d'utiliser ces informations redondantes pour conforter ses représentations perceptives et renforcer son apprentissage des séquences motrices complexes nécessaires à la production de la parole, en les confrontant aux modèles perceptifs.

Conclusion

Dès les premières semaines de vie, le cerveau humain présente des capacités adaptées au traitement de la parole, comme la normalisation à travers différents locuteurs, la perception catégorielle des phonèmes, la sensibilité aux propriétés prosodiques et rythmiques de la parole, une mémoire verbale à court terme et une mémoire à long terme efficaces. Ces capacités reposent sur des circuits cérébraux très semblables à ceux de l'adulte, impliquant les régions périsylviennes gauches. Les quelques semaines d'exposition à un environnement auditif, et notamment à la parole, pendant la vie intra-utérine et les premières semaines de vie postnatale, ne sont pas suffisantes pour expliquer l'organisation complexe de ces régions. Les similarités importantes constatées entre les nourrissons, fonctionnellement immatures, et les adultes, matures et « compétents », suggèrent que la part du déterminisme génétique pour favoriser le traitement de la parole dès le plus jeune âge est considérable. Cette aptitude pourrait provenir d'un recyclage des processus auditifs qui se sont développés chez les autres mammifères (par exemple, les discontinuités perceptives le long de certaines dimensions acoustiques sont réexploitées en tant que support des contrastes phonétiques, la sensibilité aux caractéristiques rythmiques des sons et une organisation hiérarchique dans le lobe temporal déjà présentes chez le singe permettraient de segmenter le signal continu de parole en unités plus facilement traitables), mais ne se limite pas à ceux-ci. La complexité du réseau activé par la parole chez le nourrisson qui inclue non seulement les régions temporales mais aussi les régions pariétales et frontales, ainsi que le développement du faisceau arqué reliant ces trois régions[80] suggèrent que des connexions à longue distance précocement efficaces entre régions impliquées dans la mémoire et dans l'organisation hiérarchique des patterns moteurs pourraient être le changement crucial favorisant l'apprentissage du langage dans notre espèce, en permettant la manipulation combinatoire des unités linguisti-

80. Rilling, Glasser *et al.* (2008).

ques. Enfin, n'oublions pas que la parole survient toujours pour le nourrisson dans un contexte d'échanges sociaux. Nous n'avons pas abordé ici l'importance de ces échanges et des émotions qu'ils créent, qui facilitent certainement l'apprentissage comme le montre notre étude récente où la voix de la mère activait fortement non seulement les régions émotionnelles mais amplifiait les réponses dans les régions temporales postérieures du réseau linguistique[81].

Alors que les études du développement du langage se sont classiquement focalisées sur les propriétés du stimulus de parole, l'analyse fonctionnelle des régions cérébrales que permet l'imagerie cérébrale pourrait aider à redéfinir quels sont les paramètres cruciaux pour l'acquisition du langage. Une telle analyse ouvre de nouvelles perspectives pour étudier les anomalies précoces du langage et de la communication chez l'enfant.

RÉFÉRENCES BIBLIOGRAPHIQUES

– Anderson B., Southern B. D. et Powers R. E. (1999), « Anatomic asymmetries of the posterior superior temporal lobes: A postmortem study », *Neuropsychiatry Neuropsychol. Behav. Neurol.,* 12 (4), 247-254.
– Bates E. et Roe K. (2001), « Language development in children with unilateral brain injury », *in* C. Nelson et M. Luciana (éd.), *Handbook of Developmental Cognitive Neurocsience*, Cambridge (Mass.), MIT Press, p. 281-307.
– Boemio A., Fromm S., Braun A. et Poeppel D. (2005), « Hierarchical and asymmetric temporal sensitivity in human auditory cortices », *Nat. Neurosci.,* 8 (3), p. 389-395.
– Bosch L. et Sebastian-Galles N. (1997), « Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments », *Cognition,* 65 (1), p. 33-69.
– Boye M., Gunturkun O. et Vauclair J. (2005) « Right ear advantage for conspecific calls in adults and subadults, but not infants, California sea lions (Zalophus californianus) : Hemispheric specialization for communication ? », *Eur. J. Neurosci.,* 21 (6), p. 1727-1732.
– Boysson-Bardies B., Sagart L. et Durand C. (1984), « Discernible differences in the babbling of infants according to target language », *Journal of Child Language,* 11, p. 1-15.
– Bristow D., Dehaene-Lambertz G., Mattout J., Soares C., Gliga T., Baillet S. *et al.* (2009), « Hearing faces : Crossmodal representations of speech in two-month-old infants », *Journal of Cognitive Neuroscience,* 21, p. 905-21.
– Broca P. (1861), « Remarques sur le siège de la faculté du langage articulé suivie d'une observation d'aphémie », *Bulletin de la Société anatomique de Paris,* 6, p. 330.

81. Dehaene-Lambertz, Montavont *et al.* (soumis).

– Buiatti M., Peña M. et Dehaene-Lambertz G. (2009), « Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses », *NeuroImage,* 44, p. 509-519.
– Busnel M.-C. et Granier-Deferre C. (1983), « And what of fetal audition ? », *in* A. A. Oliveirio et M. Zapelle (éd.), *The Behaviour of Human Infants*, New York, Plenum Press, p. 93-126.
– Buxhoeveden D. P., Switala A. E., Litaker M., Roy E. et Casanova M. F. (2001), « Lateralization of minicolumns in human planum temporale is absent in nonhuman primate cortex », *Brain Behav. Evol.,* 57 (6), p. 349-358.
– Cantalupo C., Pilcher D. L. et Hopkins W. D. (2003), « Are planum temporale and sylvian fissure asymmetries directly related ? A MRI study in great apes », *Neuropsychologia,* 41 (14), p. 1975-1981.
– Chi J. G., Dooling E. C. et Gilles F. H. (1977), « Gyral development of the human brain », *Annals of Neurology,* 1, p. 86-93.
– Chiron C., Jambaque I., Nabbout R., Lounes R., Syrota A. et Dulac O. (1997), « The right brain hemisphere is dominant in human infants », *Brain,* 120, p. 1057-1065.
– Damasio A., Bellugi U., Poizner H. et Gilder J. V. (1986), « Sign language aphasia during left-hemisphere Amytal injection », *Nature,* 322 (24 juillet), p. 363-365.
– DeCasper A. J. et Fifer W. P. (1980), « Of human bonding : Newborns prefer their mother's voices », *Science,* 208, p. 1174-1176.
– DeCasper A. J. et Spence M. J. (1986), « Prenatal maternal speech influences newborn's perception of speech sounds », *Infant Behavior and Development,* 9, p. 133-150.
– Dehaene-Lambertz G. (2000), « Cerebral specialization for speech and non-speech stimuli in infants », *Journal of Cognitive Neuroscience,* 12 (3), p. 449-460.
– Dehaene-Lambertz G. et Baillet S. (1998), « A phonological representation in the infant brain », *NeuroReport,* 9, p. 1885-1888.
– Dehaene-Lambertz G., Dehaene S. et Hertz-Pannier L. (2002), « Functional neuroimaging of speech perception in infants », *Science,* 298, p. 2013-2015.
– Dehaene-Lambertz G., Dupoux E. et Gout A. (2000), « Electrophysiological correlates of phonological processing : a cross-linguistic study », *Journal of Cognitive Neuroscience,* 12 (4), p. 635-647.
– Dehaene-Lambertz G., Hertz-Pannier L., Dubois J., Meriaux S., Roche A., Sigman M. *et al.* (2006), « Functional organization of perisylvian activation during presentation of sentences in preverbal infants », *Proc. Natl. Acad. Sci. USA,* 103 (38), p. 14240-14245.
– Dehaene-Lambertz G. et Houston D. (1998), « Faster orientation latencies toward native language in two-month-old infants », *Language and Speech,* 41, p. 21-43.
– Dehaene-Lambertz G., Montavont A., Jobert A., Allirol L., Dubois J., Hertz-Pannier L., et Dehaene S. (soumis), « Language or music, mother or Mozart ? Structural and environmental influences on infants' language networks' », *Brain & Language*.

– Dehaene-Lambertz G., Pallier C., Serniklaes W., Sprenger-Charolles L., Jobert A. et Dehaene S. (2005), « Neural correlates of switching from auditory to speech perception », *NeuroImage*, 24, p. 21-33.
– Dehaene-Lambertz G. et Pena M. (2001), « Electrophysiological evidence for automatic phonetic processing in neonates », *NeuroReport,* 12, p. 3155-3158.
– Dehaene-Lambertz G., Pena M., Christophe A., Charolais A. et Landrieu P. (2004). « Phoneme discrimination in a neonate with a left sylvian infarct », *Brain & Language,* 88, p. 26-38.
– Dehaene S. et Cohen L. (2007), « Cultural recycling of cortical maps » *Neuron,* 56 (2), p. 384-398.
– Dubois J., Benders M., Borradori-Tolsa C., Cachia A., Lazeyras F., Ha-Vinh Leuchter R. *et al.* (2008), « Primary cortical folding in the human newborn : An early marker of later functional development », *Brain,* 131 (Pt 8), p. 2028-2041.
– Dubois J., Benders M., Cachia A., Lazeyras F., Ha-Vinh Leuchter R., Sizonenko S. V. *et al.* (2008). « Mapping the early cortical folding process in the preterm newborn brain », *Cerebral Cortex,* 18 (6), p. 1444-1454.
– Dubois J., Hertz-Pannier L., Cachia A., Mangin J. F., Le Bihan D. et Dehaene-Lambertz G. (2008), « Structural asymmetries in the infant language and sensori-motor networks », *Cereb. Cortex*.
– Ehret G. (1987), « Left hemisphere advantage in the mouse brain for recognizing ultrasonic communication calls », *Nature,* 325, p. 249-251.
– Emmorey K., Allen J. S., Bruss J., Schenker N. et Damasio H. (2003), « A morphometric analysis of auditory brain regions in congenitally deaf adults », *Proc. Natl. Acad. Sci. USA,* 100 (17), p. 10049-10054.
– Fritz J., Mishkin M. et Saunders R. C. (2005), « In search of an auditory engram », *Proc. Natl. Acad. Sci. USA,* 102 (26), p. 9359-9364.
– Fukunishi K., Sawada K., Kashima M., Sakata-Haga H., Fukuzaki K. et Fukui Y. (2006), « Development of cerebral sulci and gyri in fetuses of cynomolgus monkeys (Macaca fascicularis) », *Anat. Embryol. (Berl),* 211 (6), p. 757-764.
– Gannon P. J., Holloway R. L., Broadfield D. C. et Braun A. R. (1998), « Asymmetry of chimpanzee planum temporale : Humanlike pattern of Wernicke's brain language area homolog », *Science,* 279 (9 janvier), p. 220-222.
– Geschwind N., et Levitsky W. (1968), « Human brain : Left-right asymmetries in temporal speech region », *Science,* 161, p. 186-187.
– Griffiths S. K., Brown W. S. Jr., Gerhardt K. J., Abrams R. M. et Morris R. J. (1994), « The perception of speech sounds recorded within the uterus of a pregnant sheep », *J. Acoust. Soc. Am.,* 96 (4), p. 2055-2063.
– Hauser M. D. et Andersson K. (1994), « Left hemisphere dominance for processing vocalizations in adult, but not infant, rhesus monkeys : Field experiments », *Proc. Natl. Acad. Sci. USA,* 91 (9), p. 3946-3948.
– Heffner H. E. et Heffner R. S. (1984), « Temporal lobe lesions and perception of species-specific vocalizations by macaques », *Science,* 226, p. 75-76.
– Hertz-Pannier L., Chiron C., Jambaque I., Renaux-Kieffer V., Van de Moortele P. F., Delalande O. *et al.* (2002), « Late plasticity for language in a child's non-

dominant hemisphere : a pre- and post-surgery fMRI study », *Brain,* 125 (Pt 2), p. 361-372.

– Holland S. K., Plante E., Weber Byars A., Strawsburg R. H., Schmithorst V. J. et Ball W. S. Jr. (2001), « Normal fMRI brain activation patterns in children performing a verb generation task », *NeuroImage,* 14 (4), p. 837-843.

– Hutsler J. J. (2003), « The specialized structure of human language cortex : Pyramidal cell size asymmetries within auditory and language-associated regions of the temporal lobes », *Brain Lang.,* 86 (2), p. 226-242.

– Jacquemot C., Pallier C., LeBihan D., Dehaene S. et Dupoux E. (2003), « Phonological grammar shapes the auditory cortex : A functional magnetic resonance imaging study », *Journal of Neuroscience,* 23, p. 9541-9546.

– Jusczyk P. W., Luce P. A. et Charles-Luce J. (1994), « Infants' sensitivity to phonotactic patterns in the native language », *Journal of Memory and Language,* 33, p. 630-645.

– Kaas J. H. et Hackett T. A. (2000), « Subdivisions of auditory cortex and processing streams in primates », *Proc. Natl. Acad. Sci. USA,* 97 (22), p. 11793-11799.

– King C., Nicol T., McGee T. et Kraus N. (1999), « Thalamic asymmetry is related to acoustic signal complexity », *Neurosci. Lett.,* 267 (2), p. 89-92.

– Kluender K. R., Diehl R. L., et Killeen P. R. (1987), « Japanese Quail can learn phonetic categories », *Science,* 237, p. 1195-1197.

– Kuhl P. K. (1983), « Perception of auditory equivalence classes for speech in early infancy », *Infant Behavior and Development,* 6, p. 263-285.

– Kuhl P. K. (1991), « Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not », *Percept Psychophys,* 50 (2), p. 93-107.

– Kuhl P. K. et Miller J. D. (1975), « Speech perception by the chinchilla : voiced-voiceless distinction in alveolar plosive consonants », *Science,* 190 (4209), p. 69-72.

– Kuhl P. K. et Padden D. M. (1983), « Enhanced discriminability at the phonetic boundaries for the place feature in macaques », *Journal of the Acoustical Society of America,* 73 (3), p. 1003-1010.

– Mehler J., Bertoncini J., Barrière M. et Jassik-Gerschenfeld D. (1978), « Infant recognition of mother's voice », *Perception*, 7, p. 491-497.

– Mehler J., Jusczyk P., Lambertz G., Halsted N., Bertoncini J. et Amiel-Tison C. (1988), « A precursor of language acquisition in young infants », *Cognition,* 29, p. 143-178.

– Moon C., Cooper R. P. et Fifer W. (1993), « Two-day-olds prefer their native language », *Infant Behavior and Development,* 16, p. 495-500.

– Nazzi T., Bertoncini J. et Mehler J. (1998), « Language discrimination by newborns : Toward an understanding of the role of rhythm », *J. Exp. Psychol. Hum. Percept. Perform.,* 24 (3), p. 756-766.

– Nazzi T., Jusczyk P. W. et Johnson E. K. (2000), « Language discrimination by English-learning 5-month-olds : Effects of rhythm and familiarity », *Journal of Memory and Language,* 43, p. 1-19.

– Newport E. L. (1990), « Maturational constraints on language learning », *Cognitive Science,* 14, p. 11-28.
– Palleroni A. et Hauser M. (2003), « Experience-dependent plasticity for auditory processing in a raptor », *Science,* 299 (5610), p. 1195.
– Pandya D. N. et Yeterian E. H. (1990), « Architecture and connections of cerebral cortex : Implications for brain evolution and function », *in* A. B. Scheibel et A. F. Wechsler (éd.), *Neurobiology of Higher Cognitive Function*, New York, Guilford Press, p. 53-83.
– Pena M., Bonatti L. L., Nespor M. et Mehler J. (2002), « Signal-driven computations in speech processing », *Science,* 298 (5593), p. 604-607.
– Pena M., Maki A., Kovacic D., Dehaene-Lambertz G., Koizumi H., Bouquet F. *et al.* (2003), « Sounds and silence : An optical topography study of language recognition at birth », *Proc. Natl. Acad. Sci. USA,* 100 (20), p. 11702-11705.
– Penhune V. B., Zatorre R. J., MacDonald J. D. et Evans A. C. (1996), « Interhemispheric anatomical differences in human primary auditory cortex : Probabilistic mapping and volume measurement from magnetic resonance scans », *Cereb. Cortex,* 6 (5), p. 661-672.
– Petersen M. R., Beecher M. D., Zoloth S. R., Green S., Marler P. R., Moody D. B. *et al.* (1984), « Neural lateralization of vocalizations by Japanese macaques : Communicative significance is more important than acoustic structure », *Behav. Neurosci.,* 98 (5), p. 779-790.
– Plante E. (1991), « MRI finding in the parents and siblings of specifically language-impaired boys », *Brain & Language,* 41, p. 67-80.
– Poremba A., Malloy M., Saunders R. C., Carson R. E., Herscovitch P. et Mishkin M. (2004), « Species-specific calls evoke asymmetric activity in the monkey's temporal poles », *Nature,* 427 (6973), p. 448-451.
– Querleu D., Renard X., Versyp F., Paris-Delrue L. et Crépin G. (1988). « Fetal Hearing », *European Journal of Obstetrics and Gynecology and Reproductive Biology,* 29, p. 191-212.
– Ramus F., Hauser M. D., Miller C., Morris D. et Mehler J. (2000), « Language discrimination by human newborns and by cotton-top tamarin monkeys », *Science,* 288, p. 349-351.
– Rilling J. K., Glasser M. F., Preuss T. M., Ma X., Zhao T., Hu X., *et al.* (2008), « The evolution of the arcuate fasciculus revealed with comparative DTI », *Nat. Neurosci.*, 11, p. 426-428.
– Sakai K. L., Tatsuno Y., Suzuki K., Kimura H. et Ichida Y. (2005), « Sign and speech : Amodal commonality in left hemisphere dominance for comprehension of sentences », *Brain,* 128 (Pt 6), p. 1407-1417.
– Seldon H. L. (1981), « Structure of human auditory cortex. II. Axon distributions and morphological correlates of speech perception », *Brain Res.,* 229 (2), p. 295-310.
– Serniclaes W., Sprenger-Charolles L., Carre R. et Demonet J. F. (2001), « Perceptual discrimination of speech sounds in developmental dyslexia », *J. Speech Lang. Hear Res.,* 44 (2), p. 384-399.

– Sinnott J. M. (2005), « Recent developments in animal speech perception », *Recent Results in Developmental Acoustics,* 2, p. 1-11.
– Sinnott J. M., et Gilmore C. S. (2004), « Perception of place-of-articulation information in natural speech by monkeys versus humans », *Percept Psychophys,* 66 (8), p. 1341-1350.
– Sowell E. R., Thompson P. M., Rex D., Kornsand D., Tessner K. D., Jernigan T. L. *et al.* (2002), « Mapping sulcal pattern asymmetry and local cortical surface gray matter distribution in vivo : Maturation in perisylvian cortices », *Cerebral Cortex,* 12 (1), p. 17-26.
– Sun T., Collura R. V., Ruvolo M. et Walsh C. A. (2006), « Genomic and evolutionary analyses of asymmetrically expressed genes in human fetal left and right cerebral cortex », *Cereb Cortex,* 16 (1), i18-25.
– Thompson P. M., Cannon T. D., Narr K. L., Van Erp T., Poutanen V. P., Huttunen M. *et al.* (2001), « Genetic influences on brain structure », *Nat. Neurosci.,* 4 (12), p. 1253-1258.
– Tincoff R., Hauser M., Tsao F., Spaepen G., Ramus F. et Mehler J. (2005), « The role of speech rhythm in language discrimination : Further tests with a non-human primate », *Dev. Sci.,* 8 (1), p. 26-35.
– Toro J. M., Trobalon J. B. et Sebastian-Galles N. (2005), « Effects of backward speech and speaker variability in language discrimination by rats », *J. Exp. Psychol. Anim. Behav. Process,* 31 (1), p. 95-100.
– Van De Weijer J. (1998), *Language Input for Word Discovery*, Nijmegen, Wageningen, Ponsen et Loijen.
– Werker J. F. et Tees R. C. (1984), « Cross-language speech perception : Evidence for perceptual reorganization during the first year of life », *Infant Behavior and Development,* 7, p. 49-63.
– Witelson S. F., et Pallie W. (1973), « Left hemisphere specialization for language in the newborn : Neuroanatomical evidence for asymmetry », *Brain,* 96, p. 641-646.
– Zatorre R. J., et Belin P. (2001), « Spectral and temporal processing in human auditory cortex », *Cerebral Cortex,* 11 (10), p. 946-953.

Comment la pratique de la musique améliore-t-elle les aptitudes cognitives ?

par HELEN NEVILLE[1]

La valeur intrinsèque de l'éducation à la pratique d'un instrument et de l'expérience du sens de la musique est universellement reconnue. En partie motivé par la diminution de la part budgétaire allouée à l'éducation artistique et musicale à l'école, un nombre croissant de recherches vise à fournir des arguments expérimentaux en faveur des effets bénéfiques potentiels de la pratique musicale sur le développement cognitif et scolaire de l'enfant. La grande majorité de ces études, qui consistent à comparer les aptitudes cognitives des sujets, rapporte que les sujets musiciens, comparativement aux non musiciens, présentent de meilleures capacités verbales, visuo-spatiales, numériques et intellectuelles[2]. De fait, ces études interprètent de telles corrélations comme une preuve empirique de l'influence *causale* de la pratique de la musique sur l'amélioration des capacités cognitives. Cependant, une interprétation tout aussi plausible pourrait substituer les causes aux effets, en stipulant que les sujets dotés de meilleures dispositions cognitives seraient plus enclins à faire l'effort d'apprendre la musique.

En effet, apprendre la musique sollicite fortement l'attention, nécessite la manipulation de signes et de relations abstraites, et requiert une intelligence fluide (autrement dit un excellent contrôle exécutif). Par conséquent, il semble fort probable que ces corrélations positives entre

1. Et (par ordre alphabétique) : Annika Andersson, Olivia Bagdade, Ted Bell, Jeff Currin, Jessica Fanning, Linda Heidenreich, Scott Klein, Brittni Lauinger, Eric Pakulak, David Paulsen, Laura Sabourin, Courtney Stevens, Stephanie Sundborg et Yoshiko Yamada (Laboratoire de développement cérébral, Université de l'Oregon). Nous remercions les familles et écoles du programme Head Start pour leur participation, ainsi que la Fondation Dana.
2. Pour une revue, *cf.* Schellenberg (2006), Norton *et al.* (2005).

musique et cognition résultent simplement du fait que les sujets doués de meilleures capacités cognitives choisissent d'apprendre la musique.

Il est également probable que l'apprentissage de la musique façonne et développe les ressources cognitives. Si l'on veut évaluer cette hypothèse, il est nécessaire d'allouer à des sujets pris au hasard soit un entraînement musical, soit un entraînement non musical, voire pas d'entraînement du tout. Or non seulement très peu d'études ont adopté une telle approche, mais, de surcroît, ces mêmes études se sont généralement limitées à certaines capacités cognitives. Ainsi, par exemple, Rauscher a pu constater que les cours particuliers de piano améliorent le traitement spatial et spatio-temporel du jeune enfant[3] ; Gardiner et ses collaborateurs[4] ont observé que des enfants de 6 ans ayant reçu une éducation en musique et en arts graphiques progressent plus vite dans les tests standardisés de lecture et d'arithmétique, que des enfants ayant reçu une éducation classique. Enfin, Schellenberg[5] a trouvé que des enfants de 6 ans qui suivent des cours de musique (chorale ou clavier) en petit groupe font davantage de progrès dans tous les sous-tests verbaux et non verbaux de l'échelle d'intelligence de Wechsler, que ceux qui prennent des cours de théâtre ou pas de cours du tout.

S'ils étaient répliqués, ces résultats pourraient d'une part suggérer que l'éducation musicale intervient de manière *causale* dans l'amélioration des capacités cognitives, et d'autre part soulever la question de savoir *comment* cette dernière peut produire de tels effets.

Dans la mesure où les améliorations rapportées jusque-là ne semblent pas restreintes à une capacité cognitive en particulier mais au contraire se manifester à travers un ensemble assez divers d'aptitudes, l'éducation musicale pourrait très probablement affecter des mécanismes impliqués dans l'amplification du traitement de l'information, tout à fait indépendamment du domaine considéré. À cet égard, un bon mécanisme candidat serait l'*attention*.

Dans la recherche présentée ici, nous envisageons l'hypothèse selon laquelle l'éducation musicale contribuerait directement à l'amélioration de diverses capacités cognitives et moyennant un entraînement sollicitant les fonctions attentionnelles. Dans la première partie, nous résumons brièvement ce qui est à présent connu au sujet de l'architecture, du développement, de la plasticité, de la vulnérabilité et de l'entraînement de l'attention.

---

3. Rauscher *et al.* (1997, 2002).
4. Gardiner *et al* (1996).
5. Schellenberg (2004).

Ce qu'on sait aujourd'hui sur l'attention

L'ARCHITECTURE DE L'ATTENTION

Les recherches menées ces dernières décennies en sciences et neurosciences cognitives convergent peu à peu vers une compréhension globale des différents composants de l'attention[6]. Bien que différents modèles ou groupes de recherche soient en désaccord sur des questions de détails conceptuels ou terminologiques, tous reconnaissent la pertinence d'une distinction fondamentale entre un niveau basique d'alerte/éveil, et un niveau d'amplification sélective de signaux spécifiques en vue d'un traitement transitoire ou soutenu des stimuli sélectionnés.

La sélection attentionnelle inclut aussi bien les processus de rehaussement de signaux sélectionnés (amplification du signal) que la suppression, en présence de « non-signaux » saillants, des informations non pertinentes (ou suppression des distracteurs). La suppression des distracteurs constitue une étape précoce de la sélection attentionnelle, et elle est également considérée comme un mécanisme de contrôle exécutif ou inhibiteur. Le contrôle exécutif, quant à lui, joue un rôle important aussi bien dans la suppression des réponses les plus immédiates que dans les engagements/désengagements très rapides de l'attention d'une cible vers une autre et dans l'allocation des ressources attentionnelles entre différentes tâches.

LE DÉVELOPPEMENT DE L'ATTENTION

Plusieurs études attestent la pertinence du concept d'attention pour rendre compte du développement cognitif de l'enfant, et en particulier du moment à partir duquel il est apte à aller à l'école[7]. Les études du développement de l'attention décrivent un processus de maturation très lent, y compris pour certains aspects de l'attention qui sont présents dès la petite enfance sous une forme encore imparfaite. Ainsi, tandis que les mécanismes d'alerte sont clairement présents dès la petite enfance, la capacité à maintenir ces mêmes mécanismes de manière volontaire et soutenue est, quant à elle, soumise à un développement lent qui s'étend

6. Driver *et al.* (2001) ; Raz et Buhle (2006) ; Shipp (2004).
7. Blair (2002), Early Child Care Research Network (2003), Posner et Rothbart (2000).

jusqu'à l'âge adulte[8]. Tandis que la sélection attentionnelle transitoire, en réponse à un stimulus exogène (*overt attention*), peut parvenir à maturation durant les dix premières années de la vie, le développement des mécanismes de sélection endogène (*covert attention*) se poursuit au moins jusqu'à l'âge de 30 ans[9].

Dans une revue des études du développement de l'attention sélective où figurent à la fois des données comportementales et de potentiels évoqués, Ridderinkhof et van der Stelt[10] ont développé l'idée selon laquelle les capacités à sélectionner un stimulus parmi d'autres, et à traiter préférentiellement l'information plus pertinente, sont pour l'essentiel présentes chez le très jeune enfant, mais que la vitesse et l'efficacité des systèmes qui contribuent à ces mêmes capacités, ou comportements, sont en revanche sujettes à s'améliorer au cours du développement. En vue de tester ces hypothèses plus directement, nous avons adapté le paradigme de potentiels évoqués (ERP) employé par Hink et Hilliard afin de rendre l'écoute dichotique plus attrayante à de jeunes enfants âgés de 3 à 8 ans[11]. Les auditeurs devaient sélectivement prêter attention à l'une des deux histoires qu'on leur présentait simultanément, et qui différaient quant à leur source spatiale (gauche/droite), à la voix du narrateur (masculine/féminine) et à leur contenu.

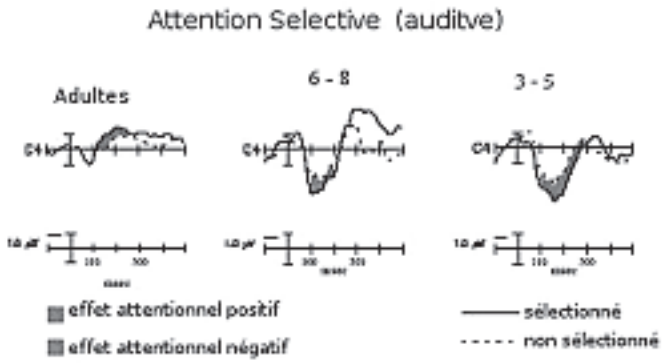


Figure 1

Comme on peut le voir ci-dessus, la forme des ERP en réponse à ces stimuli auditifs diffère fortement en fonction de l'âge. Néanmoins, on a

8. Gomes *et al.* (2000) ; Rueda *et al.* (2004).
9. Rueda *et al.* (2004) ; Schul *et al.* (2003).
10. Ridderinkhof et Van Der Stelt (2000).
11. Hink et Hillyard (1976) ; Coch *et al.* (2005) ; Sanders *et al.* (2006).

pu observer chez les enfants de 3 ans une amplification des ERP aux stimuli « sondes » (autrement dit un effet de l'attention sélective), et ce, avec une latence de réponse (100 ms) similaire à celle des adultes (*cf.* figure 1). Cette découverte suggère que, dès 3 ans, les enfants peuvent sélectivement prêter attention à une information auditive lorsqu'on leur donne des indices attentionnels, et que la nature et le déroulement temporel de ces effets sur le traitement des informations auditives sont similaires à ceux observés chez les sujets adultes.

PLASTICITÉ ET VULNÉRABILITÉ DE L'ATTENTION

Nos études précédentes, tant en comportement qu'en potentiels évoqués et en neuro-imagerie, vont dans le sens d'une plasticité considérable des systèmes neuraux qui sous-tendent l'attention sélective. Un exemple illustrant ce propos est celui des sourds, chez qui l'attention visuelle est considérablement plus performante que celle des entendants[12]. De même, on a pu observer chez des sujets adultes aveugles de naissance une augmentation similaire des facultés attentionnelles auditives[13]. Toutefois, il semblerait qu'il y ait une limite dans le développement, au-delà de laquelle ces modifications des aptitudes attentionnelles ne peuvent advenir. Nous avons effectivement pu observer que des sujets dont la cécité visuelle avait été tardive ne présentaient pas de telles compensations[14]. Ces résultats, qui illustrent la plasticité de certains mécanismes de l'attention sélective chez des sujets atteints de cécité congénitale (visuelle ou auditive), suggèrent que ces mécanismes attentionnels, à l'instar d'autres systèmes qui présentent un haut degré de neuroplasticité[15], pourraient être soumis à un développement assez lent et ainsi s'avérer particulièrement vulnérables au cours de celui-ci. Récemment, avec le même paradigme de potentiels évoqués mentionné plus haut, nous avons observé des déficits attentionnels chez des populations à risque, dont des enfants souffrant de déficiences spécifiques du langage (SLI), ou issus de milieux socio-économiques défavorisés (SES)[16].

L'entraînement de l'attention

Depuis la fin des années 1980, la recherche en réhabilitation cognitive mesure l'efficacité de certains protocoles de rééducation des fonctions

12. Bavelier *et al.* (2000, 2001) ; Neville et Lawson (1987a, 1987b, 1987c).
13. Röder *et al.* (1999, 2003).
14. Fieger *et al.* (2006).
15. Bavelier et Neville (2002).
16. Stevens *et al.* (2006) ; Lauinger *et al.* (2006).

attentionnelles dans diverses populations adultes, notamment des sujets atteints de traumatismes crâniens, traités pour une tumeur cérébrale ou à la suite d'un accident vasculaire cérébral[17]. Beaucoup de ces études rapportent des améliorations des fonctions attentionnelles (attention soutenue) et exécutives[18]. Toutefois, ces différents chercheurs se sont souvent penchés sur des fonctions attentionnelles différentes, afin de concevoir des protocoles d'entraînement adaptés aux déficits spécifiques présentés par les patients – ce qui rend l'interprétation des résultats de ces travaux assez délicate. À cet égard, une méta-analyse récente de cette littérature appelle à un recours plus rigoureux à des groupes de contrôle[19]. Des études effectuées très récemment auprès de sujets adultes normaux ont également observé des effets prononcés de l'entraînement par des jeux vidéo sur l'ensemble des fonctions attentionnelles[20]. Enfin, quelques articles rapportent des tentatives d'intégrer un entraînement des fonctions attentionnelles et de la mémoire de travail dans le traitement des enfants souffrant de troubles hyperactifs et déficits de l'attention (ADHD)[21].

Plus récemment, Posner et collègues ont mesuré l'impact de l'entraînement attentionnel chez des enfants d'âge préscolaire, issus de milieux socio-économiques élevés et sujets à un développement cognitif normal[22]. Les activités qui leur étaient proposées étaient adaptatives (au sens où la charge attentionnelle augmentait progressivement), informatisées et fondées sur une étude qui avait mis en évidence un gain de performance significatif chez des primates non humains[23]. Dans l'étude menée par Posner, bien que l'entraînement n'ait duré que cinq jours, le groupe expérimental présenta néanmoins des progrès significativement plus importants que le groupe de contrôle aux tests de contrôle exécutif et de QI non verbal.

En somme, ces recherches tendent à montrer que les fonctions attentionnelles sont sollicitées par toutes les aptitudes cognitives, et nécessaires aux bonnes performances scolaires. De plus, les processus attentionnels présentent un haut degré de plasticité et montrent à la fois une capacité de compensation (à la suite d'une privation sensorielle précoce), et une certaine vulnérabilité chez des populations à risque, dont les enfants sujets à des troubles développementaux ou issus des couches

17. Par exemple, Sohlberg et Mateer (1987, 2001, 2003) ; Niemann *et al.* (1990).
18. Ethier *et al.* (1989) ; Finlayson *et al.* (1987) ; Gray et Robertson (1989).
19. Park et Ingles (2001).
20. Green et Bavelier (2003).
21. Kerns *et al.* (1999), Klingberg *et al.* (2002).
22. Rueda *et al.* (2005).
23. Rumbaugh et Washburn (1995).

socio-économiques les plus basses[24]. Plusieurs études rigoureuses menées auprès d'enfants et adultes, toutes origines confondues, suggèrent que l'attention peut être significativement améliorée moyennant un entraînement adapté. Au vu de ces résultats, l'objectif de la recherche qui est présentée ici est de déterminer si une éducation musicale dispensée auprès d'enfants d'âge préscolaire est à même de produire des améliorations significatives tant des capacités cognitives que des performances scolaires, et qui soient comparables aux effets obtenus par un entraînement des fonctions attentionnelles.

Éducation musicale et attention : une expérience

NOS HYPOTHÈSES

Nous souhaitions étudier l'hypothèse selon laquelle des enfants d'âge préscolaire, après avoir suivi, durant huit semaines et chaque jour d'école, un entraînement soit musical, soit attentionnel et d'une durée de 40 minutes, pouvaient présenter des améliorations dans leurs aptitudes linguistiques, prélittéraires, visuo-spatiales, numériques et intellectuelles non verbales. De plus, nous souhaitions savoir si ces améliorations pouvaient s'avérer plus importantes que celles qu'on peut observer dans des groupes de contrôle, que ceux-ci reçoivent leur entraînement au sein de petits effectifs ou dans de grandes classes.

Chaque enfant fut donc inclus dans l'un des quatre groupes de manière aléatoire, et ces groupes furent appariés pour un certain nombre de variables connues pour être importantes dans les performances cognitives. Les effets des interventions furent évalués sur la base de mesures fiables et validées par des expérimentateurs qui ignoraient à quel groupe appartenaient les enfants.

LE DÉROULEMENT DE L'EXPÉRIENCE

Les groupes étaient bien homogènes au départ. Tous les participants étaient issus des couches socio-économiques basses, âgés de 3 à 5 ans, droitiers, monolingues, et sans trouble neurologique ou comportemental. Les enfants furent recrutés dans les sections préscolaires *Head Start* locales. *Head Start* est un programme préscolaire financé par l'état fédéral et

24. Stevens *et al.* (sous presses)

destiné aux enfants de familles à très faibles revenus. Un mois avant et après les interventions, les tests suivants furent administrés :

1. Évaluation clinique des acquis linguistiques préscolaires fondamentaux, 2e édition ;
2. Échelles d'intelligence Stanford-Binet, 5e édition (SB-5) ;
3. Test de vocabulaire Peabody Picture, 3e édition (PPVT-III) ;
4. Identification de lettres ;
5. Mesure développementale des capacités numériques.

Durant huit semaines, le groupe « musical » suivit un entraînement en petits groupes (ratio de 5 élèves pour 2 encadrants), et axé sur des activités musicales telles que l'écoute musicale, le mouvement en musique, le chant, et la pratique d'un instrument. Chaque cours durait 40 minutes et était dispensé pendant le temps scolaire à une fréquence de 4 jours par semaine. Pour déterminer dans quelle mesure les effets observés étaient spécifiques à l'entraînement musical, nous avions ajouté plusieurs groupes de contrôle. Ces groupes de contrôle étaient les suivants : (1) entraînement dispensé en plus grande classe (ratio de 18 élèves pour 2 encadrants) ; (2) élèves suivant les activités classiques du programme Head Start, mais en petits groupes (ratio de 5 élèves pour 2 encadrants), et (3) entraînement des fonctions attentionnelles en petit groupe (concentration, conscience des détails, etc.).

Tous les groupes de contrôle (excepté celui en grande classe) étaient encadrés par les mêmes instructeurs et pendant la même durée que celui recevant l'entraînement musical. Si les élèves des groupes de contrôle présentaient des profils d'amélioration différents de ceux du groupe expérimental, notre protocole expérimental permettait d'en déduire que l'entraînement musical conduit à des effets spécifiques. Si les enfants des autres groupes présentaient un profil d'effets similaire à ceux observés dans le groupe musical, nous pouvions inférer les mécanismes stimulés par la pratique musicale. De fait, notre hypothèse était que l'apprentissage musical peut être équivalent à un entraînement des fonctions attentionnelles, et qu'ainsi les effets des deux interventions peuvent être similaires.

LES GROUPES DE CONTRÔLE EN GRANDE CLASSE

Les enfants qui avaient suivi le programme *Head Start* classique (N = 19, avec un ratio de 18 élèves pour 2 encadrants) présentèrent des progrès similaires dans les tests CELF de compréhension et production linguistiques ($p < .01$), et de conscience phonologique ($p < .03$). Cependant aucun progrès significatif ne fut observé dans les autres batteries de tests.

Données comportementales

Groupes	Compréhension linguistique	Aptitudes prélittéraires	Traitement numérique	Cognition spatiale	QI non verbal
Programme classique HS en grande classe (19)	*	*			
Entraînement musical (26)	*	*	*	**	*
Entraînement attentionnel (23)	*	**	**	*	*
Programme classique HS en petit groupe (20)	*	*	*	*	*
Formation parentale (14)	**	**	*	*	**

Signifiance
* intra groupe avant/après
** inter groupes

Figure 2

L'ENTRAÎNEMENT MUSICAL

Après avoir suivi le programme d'entraînement musical, les enfants (N = 26) présentèrent des progrès dans les tests (CELF) de compréhension ($p < .01$) et production linguistiques ($p < .001$) (*cf.* figure 2). Ils firent également des progrès significatifs dans les tests d'intelligence de Wechsler en assemblage d'objets ($p < .01$), un résultat analogue à ceux de Schellenberg. Par ailleurs, le groupe musical fit des progrès significatifs aux tests que nous avions développés en vue de mesurer les aptitudes numériques ($p < .007$), en particulier ceux de comptage verbal et d'estimation de quantités continues. Enfin, ce même groupe progressa de manière significative aux tests de QI non verbal de Stanford-Binet ($p < .03$), qui comprenaient le sous-test de QI Stanford-Binet de raisonnement fluide et quantitatif ($p < .03$) ainsi que celui de « raisonnement critique » ($p < .01$) ? Dans ce dernier sous-test, par exemple, l'enfant doit déceler ce qui ne va pas dans une image où l'on voit deux personnages au soleil dont les ombres ne sont pas orientées dans la même direction.

L'ENTRAÎNEMENT DE L'ATTENTION

Les enfants ayant suivi ce type d'entraînement (N = 23) présentèrent des progrès comparables à ceux ayant suivi l'entraînement musical, et ce, dans des domaines analogues : compréhension ($p < .01$) et production ($p < .004$) linguistiques ; conscience phonologique ($p < .01$) (*cf.* figure 2) ; en cognition visuelle (ils s'améliorèrent au test d'assemblage

d'objets, $p < .007$) ; en aptitudes numériques ($p < .001$) ; enfin aux sous-tests de Stanford-Binet relatifs au raisonnement fluide et quantitatif, à la mémoire de travail visuo-spatiale, et au raisonnement critique ($p < .01$, pour chacun).

Ces résultats sont compatibles avec l'hypothèse selon laquelle l'effet bénéfique de la pratique musicale sur les aptitudes cognitives serait en partie dû au fait que la musique sollicite et développe les fonctions attentionnelles.

LE RÔLE DES PARENTS

Nous avons récemment complété ces travaux par des données supplémentaires, obtenues auprès de 27 enfants. Chez les groupes ayant suivi un entraînement musical ou attentionnel, nous continuons d'observer des progrès considérables dans les tests d'aptitudes intellectuelles non verbales, numériques, ainsi que relatives au traitement de l'espace – améliorations au demeurant non observées au sein du groupe de contrôle qui suivait le programme Head Start classique en grande classe, tandis que les enfants entraînés en petites classes présentaient désormais des effets d'apprentissage importants dans les mêmes aptitudes cognitives.

Ce dernier point suggère que la majorité des effets que nous avions supposés être dus à la pratique musicale et/ou à l'entraînement des fonctions attentionnelles peuvent advenir en petit groupe, lorsque les adultes consacrent plus de temps et d'attention à chaque enfant. Le rôle central de la tutelle de l'enfant par l'adulte est également corroboré par les résultats d'une étude menée en parallèle, où seuls les parents recevaient une formation en vue d'améliorer leurs propres attitudes parentales. À l'issue de celle-ci, la conduite des enfants s'améliora, et l'on put observer des progrès importants dans chacune des mesures rapportées ici[25]. Ces changements sont tout à fait significatifs, y compris dans les comparaisons intergroupes.

Pour finir, nous avons mené en 2008 une intervention « mixte », qui comprenait, pour les enfants, un entraînement par semaine en petit groupe, à la fois en musique et de l'attention, ainsi qu'une formation hebdomadaire destinée à leurs parents. À ce jour, les données suggèrent que cette approche pourrait être la plus efficace parmi toutes celles que nous avons essayées. Après huit semaines, les améliorations des aptitudes linguistiques, prélittéraires, numériques et intellectuelles des enfants se sont avérées significativement plus importantes que toutes celles observées dans les groupes de contrôle.

---

25. *Cf.* figure 2 ; et Fanning *et al.*, (2007).

Conclusion

En conclusion, l'étude de l'impact de l'éducation musicale sur l'amélioration des aptitudes cognitives nous a permis de découvrir les effets croisés de l'attention et de la tutelle parentale sur les bénéfices cognitifs de la musique.

(Traduction : Sarah Kouhou)

RÉFÉRENCES BIBLIOGRAPHIQUES

– Bavelier D., Brozinsky C., Tomann A., Mitchell T., Neville H. et Liu G. (2001), « Impact of early deafness and early exposure to sign language on the cerebral organization for motion processing », *Journal of Neuroscience,* 21 (22), p. 8931-8942.
– Bavelier D. et Neville H. J. (2002), « Cross-modal plasticity : Where and how ? », *Nature Reviews Neuroscience,* 3, p. 443-452.
– Bavelier D., Tomann A., Hutton C., Mitchell T., Liu G., Corina D. *et al.* (2000), « Visual attention to the periphery is enhanced in congenitally deaf individuals », *Journal of Neuroscience,* 20 (17), p. 1-6.
– Blair C. (2002), « School readiness : Integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry », *American Psychologist,* 57 (2), p. 111-127.
– Clay Marie M. (1993), *An Observation Survey of Early Literacy Achievement*, Auckland, Heinemann.
– Coch D., Sanders L. D. et Neville H. J. (2005), « An event-related potential study of selective auditory attention in children and adults », *Journal of Cognitive Neuroscience*, 17 (4), p. 605-622.
– Driver J., Davis G., Russell C., Turatto M. et Freeman E. (2001), « Segmentation, attention and phenomenal visual objects », *Cognition,* 80, p. 61-95.
– Dunn L.M. et Dunn L.M. (1997), *Peabody Picture Vocabulary Test*, Circle Pines (Minnesota), American Guidance Services (3e éd.).
– Ethier M., Braun C. et Baribeau J. M. C. (1989), « Computer-dispensed cognitive-perceptual training of closed head injury patients after spontaneous recovery. Study 1 : Speeded tasks », *Canadian Journal of Rehabilitation,* 2, p. 223-233.
– Fieger A., Röder B., Teder-Sälejärvi W., Hillyard S.A. et Neville H. J. (2006), « Auditory spatial tuning in late onset blind humans », *Journal of Cognitive Neuroscience,* 18 (2), p. 149-157.
– Finlayson M. A. J., Alfano D. P. et Sullivan J. F. (1987), « A neuropsychological approach to cognitive remediation : microcomputer applications », *Canadian Psychology,* 28, p. 180-190.

– Gardiner M. F., Fox A., Knowles F. et Jeffrey D. (1996), « Learning improved by arts training », *Nature,* 381, p. 284.
– Ginsburg H. P. et Baroody A. J. (2003), *The Test of Early Mathematics Ability*, Austin (Texas), PRO-ED (3e éd.).
– Gomes H., Molholm S., Christodoulou C., Ritter W. et Cowan N. (2000), « The development of auditory attention in children », *Frontiers in Bioscience,* 5, p. 108-120.
– Gray J. M. et Robertson I. H. (1989), « Remediation of attentional difficulties following brain injury : three experimental single case studies », *Brain Injury,* 3, p. 163-170.
– Green C. S. et Bavelier D. (2003), « Action video game modifies visual selective attention », *Nature,* 423, p. 534-537.
– Hink R. F. et Hillyard S. A. (1976), « Auditory evoked potentials during selective listening to dichotic speech messages », *Perception & Psychophysics,* 20, p. 236-242.
– Kerns K. A., Eso K. et Thomson J. (1999), « Investigation of a direct intervention for improving attention in young children with ADHD », *Developmental Neuropsychology,* 16 (2), p. 273-295.
– Klingberg T., Forssberg H. et Westerberg H. (2002), « Training of working memory in children with ADHD », *Journal of Clinical and Experimental Neuropsychology,* 24 (6), p. 781-791.
– Lauinger B., Sanders L., Stevens C. et Neville H. (2006), *An ERP Study of Selective Auditory Attention and Socioeconomic Status in Young Children*, préseté à la Cognitive Neuroscience Society, San Francisco.
– National Institute Of Child Health and Human Development, Early Child Care Research Network (2003), « Do children's attention processes mediate the link between family predictors and school readiness ? », *Developmental Psychology*, 39, p. 581-593.
– Neville H. J. et Lawson D. (1987a), « Attention to central and peripheral visual space in a movement detection task : an event-related potential and behavioral study. I. Normal hearing adults », *Brain Research,* 405, p. 253-267.
– Neville H. J. et Lawson D. (1987b), « Attention to central and peripheral visual space in a movement detection task : an event-related potential and behavioral study. II. Congenitally deaf adults », *Brain Research,* 405, p. 268-283.
– Neville H. J. et Lawson D. (1987c), « Attention to central and peripheral visual space in a movement detection task an event-related potential and behavioral study. III. Separate effects of auditory deprivation and acquisition of a visual language », *Brain Research,* 405 (2), p. 284-294.
– Niemann H., Ruff R. M. et Baser C. A. (1990), « Computer-assisted attention retraining in head-injured individuals : A controlled efficacy study of an outpatient program », *Journal of Consulting and Clinical Psychology,* 58 (6), p. 811-817.
– Norton A., Winner E., Cronin K., Overy K., Lee D. J. et Schlaug G. (2005), « Are there pre-existing neural, cognitive, or motoric markers for musical ability ? », *Brain and Cognition,* 59, p. 124-134.

– Park N. W. et Ingles J. (2001), « Effectiveness of attention rehabilitation after an acquired brain injury : a meta-analysis », *Neuropsychology,* 15 (2), p. 199-210.
– Posner M. I. et Rothbart M. K. (2000), « Developing mechanisms of self-regulation », *Development and Psychopathology,* 12, p. 427-441.
– Rauscher F. H. (2002), « Mozart and the mind : Factual and fictional effects of musical enrichment », *in* J. Aronson (éd.), *Improving Academic Achievement : Impact of Psychological Factors on Education*, San Diego, Academic Press, p. 267-278.
– Rauscher F. H., Shaw G. L., Levine L. J., Wright E. L., Dennis W. R. et Newcomb R. L. (1997), « Music training causes long-term enhancement of preschool children's spatial-temporal reasoning », *Neurological Research,* 19, p. 2-8.
– Raz A. et Buhle J. (2006), « Typologies of attention networks », *Nature Reviews Neuroscience,* 7, p. 367-379.
– Ridderinkhof K. R. et Van Der Stelt O. (2000), « Attention and selection in the growing child : views derived from developmental psychophysiology », *Biological Psychology*, 54, p. 55-106.
– Röder B. et Neville H. (2003), « Developmental plasticity », *in* J. Grafman et I. Robertson (éd.), *Plasticity and Rehabilitation. Handbook of Neuropsychology*, vol. 9, Amsterdam, Elsevier Science.
– Röder B., Teder-Sälejärvi W., Sterr A., Rösler F., Hillyard S. A. et Neville H. J. (1999), « Improved auditory spatial tuning in blind humans », *Nature,* 400 (6740), p. 162-166.
– Roid G. H. (2003), *Stanford-Binet Intelligence Scales*, Itasca (Ill.), Riverside Publishing (5e édition).
– Rueda M. R., Fan J., Mccandliss B. D., Halparin J. D., Gruber D. B., Lercari L. P. *et al.* (2004), « Development of attentional networks in childhood », *Neuropsychologia,* 42, p. 1029-1040.
– Rueda M. R., Rothbart M., Mccandliss B., Saccamanno L. et Posner M. (2005), « Training, maturation, and genetic influences on the development of executive attention », *Proceedings of the National Academy of Science USA,* 102, p. 14931-14936.
– Rumbaugh D. M. et Washburn D. A. (1995), *Attention, Memory and Executive Function*, Baltimore, Brookes.
– Sanders L. D., Stevens C., Coch D. et Neville H. J. (2006), « Selective auditory attention in 3- to 5-year-old children : An event-related potential study », *Neuropsychologia,* 44 (11), p. 2126-2138.
– Schellenberg E. G. (2004), « Music lessons enhance IQ », *Psychological Science*, 15, p. 511-514.
– Schellenberg E. G. (2006), « Long-term positive associations between music lessons and IQ », *Journal of Educational Psychology,* 98 (2), p. 457-468.
– Schul R., Townsend J. et Stiles J. (2003), « The development of attentional orienting during the school-age years », *Developmental Science,* 6 (3), p. 262-272.
– Shipp S. (2004), « The brain circuitry of attention », *Trends in Cognitive Sciences,* 8 (5), p. 223-230.

– Sohlberg M. M., Avery J., Kennedy M., Ylvisaker M., Coelho C., Turkstra L. *et al*. (2003), « Practice guidelines for direct attention training », *Journal of Medical Speech-Language Pathology,* 11 (3), p. 19-39.
– Sohlberg M. M. et Mateer C. A. (1987), « Effectiveness of an attention-training program », *Journal of Clinical and Experimental Neuropsychology,* 9, p. 117-130.
– Sohlberg M. M., Mclaughlin K. A., Pavese A., Heidrich A. et Posner M. (2001), « Evaluation of attention process training and brain injury education in persons with acquired brain injury », *Journal of Clinical and Experimental Neuropsychology,* 22, p. 656-676.
– Stevens C., Sanders L. et Neville H. (2006), « Neurophysiological evidence for selective auditory attention deficits in children with specific language impairment », *Brain Research,* 1111 (1), p. 142-152.
– Stevens C., Lauinger B. et Neville H. (sous presses), « Differences in the neural mechanisms of selective attention in children from different socioeconomic backgrounds », *Developmental Science*.
– Wiig E. H., Secord W. A. et Semel E. (2004), *Clinical Evaluation of Language Fundamentals*, San Antonio (Texas), The Psychological Corporation Harcourt Assessment, Inc (2e édition).

Les raisons de l'autisme

par Anne Bargiacchi et Monica Zilbovicius

Introduction

L'autisme est un trouble précoce, global et sévère du développement social de l'enfant, qui altère l'ensemble de ses capacités à établir des rapports avec d'autres individus. La présentation clinique et la sévérité de ce trouble sont variables, et l'on utilise aujourd'hui le terme « troubles du spectre autistique » pour illustrer cette variabilité clinique. Le symptôme le plus caractéristique de l'autisme est le déficit des interactions sociales, qui est associé à un déficit de la communication verbale et non verbale, et à des comportements restreints et stéréotypés. Les symptômes doivent apparaître avant l'âge de 3 ans. En France, l'autisme atteint environ 180 000 personnes, et la prévalence serait de 6 enfants pour 1 000. Il s'agit d'un handicap social majeur. Grâce aux études en imagerie cérébrale effectuées ces dernières années, nous connaissons mieux les circuits neuronaux impliqués dans l'autisme. L'imagerie cérébrale a aussi permis de mieux comprendre les régions cérébrales impliquées dans l'interaction sociale normale, en particulier les circuits impliqués dans la perception sociale (perception de mouvements biologiques tels que le regard, les mouvements des visages, des mains, du corps), et les réseaux qui sous-tendent la « théorie de l'esprit », deux fonctions cognitives qui sont anormales dans l'autisme. Ainsi, les études sur la perception sociale ont montré que la région du sillon temporal supérieur est essentielle dans le traitement des stimuli sociaux lors de l'interaction avec autrui. En parallèle, les études d'imagerie menées dans notre laboratoire nous ont conduites à formuler l'hypothèse que des anomalies du sillon temporal supérieur sont impliquées dans la pathogénie de l'autisme. En effet, les anomalies anatomiques et fonctionnelles qui ont été décrites dans cette région pour-

raient refléter une des premières étapes dans la cascade des anomalies survenues au cours du développement cérébral, et qui sous-tendent l'autisme.

Dans ce chapitre, nous allons nous concentrer sur le rôle du sillon temporal supérieur, tout d'abord au cours de l'interaction sociale normale, puis dans le cadre des anomalies de l'interaction sociale qui caractérisent l'autisme. Nous allons rappeler les données récentes sur la contribution du sillon temporal supérieur et des autres régions du « cerveau social » dans la cognition sociale, puis décrire les données d'imagerie cérébrale impliquant le sillon temporal supérieur dans l'autisme.

Sillon temporal supérieur et cognition sociale

Lorsqu'on s'intéresse aux origines du dialogue humain, il est beaucoup question d'« entendre ». Dans cet exposé sur « les raisons de l'autisme », nous allons cependant surtout évoquer le « voir », car un véritable dialogue existe dans le cerveau entre ces deux modalités sensorielles. Commençons par une expérience : si l'on enregistre, à l'aide d'une caméra à rayonnement infrarouge, la direction de votre regard (en anglais, *eye tracking*) pendant que vous regardez un film, on constatera que lors d'une scène d'interaction sociale, votre regard est le plus souvent fixé sur les yeux des personnages. Si à votre place se trouvait une personne atteinte d'autisme, le tracé obtenu montrerait au contraire un regard éloigné des yeux des personnages, traduisant une attention visuelle limitée à des régions de la scène peu informatives sur ce qui est en train de se passer entre les personnages. Pour tenter d'expliquer le résultat de cette expérience, nous allons partir d'un développement cérébral « normal ». Avec l'évolution de la complexité des relations sociales, notre cerveau s'est aussi organisé d'une façon de plus en complexe. L'un des fondements d'un comportement dit « social » est, pour les animaux, la capacité à reconnaître leurs pairs, et pour les humains, la capacité à reconnaître un individu dans un contexte social. Chez la plupart des mammifères non primates, cette reconnaissance se fait essentiellement par l'odorat. En revanche, chez les primates, la reconnaissance des visages et des voix est prépondérante : « l'autre » est vu ou entendu, puis reconnu. Notre cerveau possède en outre une capacité particulière de « cognition sociale », c'est-à-dire qu'il est capable d'un traitement de l'information qui aboutit à la perception des dispositions et intentions d'autrui[1]. Cette capacité fait en réalité intervenir un ensemble

1. Brothers L., Ring B. et Kling A. (1990), « Response of neurons in the macaque amygdala to complex social stimuli », *Behav. Brain Res.*, 41 (3), p. 199-213.

de régions cérébrales, appelé « cerveau social », qui comporte la région préfrontale, le sillon temporal supérieur, la région temporale interne périamygdalienne, et le gyrus fusiforme[2]. De nombreux travaux ont mis en évidence l'implication du sillon temporal supérieur dans la perception du regard et, de manière plus générale, dans la perception des mouvements : mouvements du regard, du visage, de la bouche, des mains, du corps[3]. La perception du regard est une des composantes essentielles de l'interaction sociale normale. Par ailleurs, Allison et ses collaborateurs ont souligné le rôle primordial du sillon temporal supérieur parmi les structures cérébrales sous-tendant la cognition sociale, en montrant que l'analyse initiale des indices sociaux visuels se produit dans le sillon temporal supérieur[4]. Ainsi, lors d'une interaction avec autrui, nous percevons à chaque instant des mouvements subtils, qui nous permettent d'obtenir des informations sur son « état mental ». Depuis, de nouveaux résultats ont été publiés, qui ont élargi nos connaissances sur le rôle du sillon temporal supérieur dans la cognition sociale. Ces résultats proviennent d'analyses en imagerie fonctionnelle, d'abord en utilisant des stimuli simples, puis avec des stimuli sociaux beaucoup plus complexes et de nature écologique[5]. Ces différentes études ont montré une implication du sillon temporal supérieur dans :

a) la perception sociale visuelle : il existe une activation du sillon temporal supérieur lors de la perception des mouvements du regard, de la main, du corps et de la bouche[6] ;

b) la perception de la voix humaine[7] ;

c) des aspects plus complexes de la cognition sociale, comme la compréhension et l'interprétation des intentions de l'autre (théorie de l'esprit)[8].

Le sillon temporal supérieur semble donc constituer un carrefour des voies de perception visuelle et de la voix humaine. Tout comme la perception des mouvements fournit des indices sociaux indispensables à l'interaction sociale, la voix humaine est probablement l'indice le plus

2. *Ibid.*
3. Allison T., Puce A. et McCarthy G. (2000), « Social perception from visual cues : role of the STS region », *Trends Cogn. Sci.*, 4 (7), p. 267-278.
4. *Ibid.*
5. Adolphs R. (2003), « Cognitive neuroscience of human social behaviour », *Nat. Rev. Neurosci.*, 4 (3), p. 165-78.
6. Allison, Puce et McCarthy, *op. cit.*
7. Belin P. *et al.* (2000), « Voice-selective areas in human auditory cortex », *Nature*, 403 (6767), p. 309-312.
8. Pelphrey K. A., Morris J. P. et McCarthy G. (2004), « Grasping the intentions of others : the perceived intentionality of an action influences activity in the superior temporal sulcus during social perception », *J. Cogn. Neurosci.*, 16 (10), p. 1706-1716.

important de notre environnement sonore lors d'une interaction avec autrui. Non seulement elle porte la parole, qui fait de l'homme une espèce unique, mais elle est aussi un « visage auditif », riche en informations concernant l'identité et l'état affectif de l'interlocuteur. Les réseaux cérébraux incluant le sillon temporal supérieur contribuent donc à analyser, d'une part, le regard et les autres mouvements, et, d'autre part, les indices contenus dans la voix humaine, dans la mesure où ces perceptions participent de façon significative à l'interaction sociale.

L'autisme, la cognition sociale et le cerveau social

Revenons maintenant à l'intitulé de ce chapitre : « Les raisons de l'autisme ». Le témoignage de parents d'enfants autistes ressemble souvent à ceci :

> Paul a maintenant 7 ans. Dès l'âge de 1 an, nous nous sommes inquiétés. Il ne ressemblait pas à son frère aîné. Il ne jouait pas avec ses mains, il ne gazouillait pas, il ne semblait pas intéressé par sa mère, son père, son frère… Lorsqu'on le prenait dans nos bras, il ne réagissait pas beaucoup. Il ne souriait pas non plus. On avait l'impression qu'il n'entendait pas. Paul était très mignon, et l'on ne pouvait pas imaginer qu'il était malade. Il ne présentait aucun signe corporel. Notre inquiétude n'a cessé de grandir. Paul était de plus en plus bizarre. Il ne parlait toujours pas. Il faisait des mouvements étranges avec ses mains, il se balançait sans arrêt. Et parfois, il se mettait en colère. Il ne jouait pas. Nous avons posé des questions au pédiatre, qui semblait aussi être un peu désemparé face à Paul : « les enfants ne se développent pas tous de la même façon… Les garçons sont parfois un peu lents… ». Le problème s'est aggravé à l'école maternelle. Selon la maîtresse, Paul était réellement différent des autres enfants. Nous sommes alors allés consulter un psychiatre pour enfants, qui a diagnostiqué l'autisme de Paul. Pourquoi Paul est-il autiste ? Il a pourtant été aimé, câliné et entouré comme son frère aîné.

Pour tenter de comprendre les mécanismes cérébraux qui sous-tendent l'autisme, nous effectuons des recherches en imagerie cérébrale chez les enfants atteints de cette maladie. Ces recherches utilisent trois techniques : la tomographie par émission de positons (TEP) au repos, qui permet de mesurer l'activité synaptique au repos, l'imagerie par résonance magnétique (IRM) « anatomique », qui révèle certains aspects de la struc-

ture du cerveau, et des méthodes d'activation du cerveau, accompagnées d'une évaluation de l'activité cérébrale par TEP ou IRM.

Nous avons pu ainsi montrer qu'il existe une diminution du flux sanguin dans les régions temporales supérieures, ce qui reflète une baisse de l'activité synaptique au repos, chez des enfants autistes comparés à des enfants non autistes appariés pour l'âge et le retard mental[9]. Ce résultat a rapidement été confirmé par une étude japonaise[10]. Nous avons également montré, en utilisant une échelle qui permet de mesurer la sévérité de l'autisme, que plus l'autisme est sévère, plus le débit sanguin cérébral dans le sillon temporal supérieur est diminué[11]. Ces résultats ont fourni la première preuve solide d'un dysfonctionnement de la région temporale supérieure chez des enfants autistes d'âge scolaire. Or je vous ai dit précédemment que cette région du cerveau est impliquée dans la perception et la cognition sociales.

Notre équipe travaille maintenant sur la possibilité de détecter cette anomalie du fonctionnement cérébral au repos chez chaque individu, par des méthodes d'analyse multivariée. Actuellement nous obtenons un taux de classement réussi (classification de l'enfant comme étant atteint d'autisme ou pas) de 88 %, grâce à l'étude plus particulière de deux régions du cerveau, dont le sillon temporal supérieur. Il est possible d'observer pour chacun des enfants atteints d'autisme la diminution du débit sanguin cérébral. L'endroit où cette diminution est maximale est variable pour chaque enfant, mais lorsque l'on regarde l'ensemble des points, ceux-ci donnent l'impression d'être distribués le long du sillon temporal supérieur[12].

Nous avons par ailleurs étudié des enfants atteints d'autisme en IRM anatomique, et avons trouvé une diminution de l'épaisseur de la substance grise dans ces régions temporales supérieures[13]. Une autre équipe a montré l'existence d'une différence significative dans l'aspect de sillons corticaux, principalement sillons frontal et temporal, chez les enfants autistes[14]. En utilisant une mesure directe de l'épaisseur du cortex pour examiner l'intégrité de la matière grise et explorer le substrat anato-

9. Zilbovicius, M. *et al.* (2000), « Temporal lobe dysfunction in childhood autism : a PET study. Positron emission tomography », *Am. J. Psychiatry*, 2000, 157 (12), p. 1988-1993.
10. Ohnishi, T. *et al.* (2000), « Abnormal regional cerebral blood flow in childhood autism », *Brain*, 123 (Pt 9), p. 1838-1844.
11. Gendry Meresse I. *et al.* (2005), « Autism severity and temporal lobe functional abnormalities », *Ann. Neurol.*, 58 (3), p. 466-469.
12. *Ibid.*
13. Boddaert N. *et al.* (2004), « Superior temporal sulcus anatomical abnormalities in childhood autism : a voxel-based morphometry MRI study », *NeuroImage*, 23 (1), p. 364-369.
14. Levitt J. G. *et al.* (2003), « Cortical sulcal maps in autism », *Cereb. Cortex*, 13 (7), p. 728-735.

mique des symptômes autistiques, Hadjikhani et ses collaborateurs ont trouvé, dans le groupe de sujets atteints d'autisme, une diminution locale de la matière grise dans le gyrus frontal inférieur, le lobule pariétal inférieur et les sillons temporaux supérieurs[15]. En outre, l'amincissement du cortex dans ces régions, qui sont toutes impliquées dans la cognition sociale, a été corrélé avec la sévérité des symptômes autistiques. Nous avons, quant à nous, mis en évidence l'existence d'une corrélation entre l'aspect (forme, profondeur…) du sillon temporal supérieur et les symptômes autistiques[16] (figure 1).

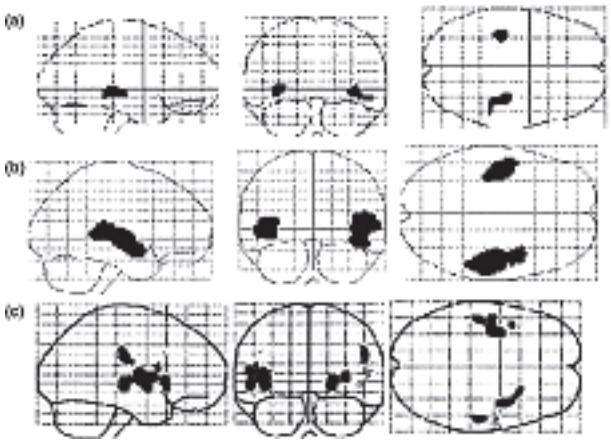


Figure 1 : *Les résultats obtenus en IRM anatomique et en imagerie cérébrale fonctionnelle au repos ont mis en évidence des anomalies du sillon temporal supérieur chez les enfants atteints d'un trouble du spectre autistique. Sur ces images du cerveau dans un plan sagittal, frontal, ou horizontal (*de gauche à droite*), les régions en noir sont celles qui diffèrent significativement quant à l'épaisseur de matière grise (mesurée par l'IRM)* ***(a)***[17] *le débit sanguin cérébral au repos (mesuré en TEP)* ***(b)***[18]*, ou en tomographie d'émission monophotonique (*single photon emission computed tomography *ou SPECT)* ***(c)***[19] *chez les enfants atteints d'un trouble du spectre autistique, par comparaison avec des enfants non atteints. Le même logiciel d'analyse des images a été utilisé dans ces trois études (*statistical parametric mapping*). Les régions en noir incluent le sillon temporal supérieur.*

15. Hadjikhani N. *et al.* (2006), « Anatomical differences in the mirror neuron system and social cognition network in autism », *Cereb. Cortex*, 16 (9), p. 1276-1282.
16. Cachia A. et *al*, « Autism severity and anatomical abnormalities of the superior temporal sulcus », en préparation.
17. Boddaert N. *et al.*, *op. cit.*
18. Zilbovicius M. *et al.*, *op. cit.*
19. Ohnishi T. *et al.* (2000), « Abnormal regional cerebral blood flow in childhood autism », *Brain*, 123 (Pt 9), p. 1838-1844.

Nous avons mentionné, au début de ce chapitre, l'importance de la perception de la voix dans les interactions sociales. Nous avons fait écouter des séries de sons produits par une voix humaine (sons « voix ») et des séries de sons « non-voix » à des sujets autistes et à des sujets non autistes, étudiés par IRM fonctionnelle. Chez les sujets autistes, nous n'avons pas observé d'activation de la région temporale supérieure au moment de l'écoute des sons « voix », contrairement à ce qui était obtenu chez les sujets témoins[20]. Chez les sujets atteints d'autisme, il n'y avait aucune différence, en termes d'activité cérébrale enregistrée par cette technique, entre le fait d'entendre un stimulus « voix » et celui d'entendre un stimulus « non-voix ». Ce résultat concorde avec les témoignages de parents d'enfants autistes, qui rapportent régulièrement le fait que leur enfant ne répond pas lorsqu'ils l'appellent, alors qu'il réagit à d'autres sons de l'environnement.

D'autres études ont permis de mettre en évidence des anomalies d'activation de l'aire cérébrale spécialisée dans la perception des visages (en anglais, *face fusiform area*). Cette aire semble être moins activée chez les sujets autistes[21], mais cette hypoactivation pourrait être liée à une différence dans la façon de regarder les visages, plutôt qu'à un dysfonctionnement primaire de cette région[22]. Par ailleurs, les sujets autistes ont des déficits dans la perception du regard, dans le « décodage » de l'information au cours du contact visuel, et des difficultés d'accès à l'information pour en déduire l'état mental de l'autre[23]. « Je n'avais aucune idée que d'autres personnes communiquaient par le biais de subtils mouvements de l'œil, jusqu'à ce que je l'aie lu dans un magazine, il y a cinq ans », dit un des adultes autistes. Il est souvent question de l'incapacité des sujets autistes à comprendre l'état émotionnel d'autrui. L'enregistrement de la

20. Gervais H. *et al.* (2004), « Abnormal cortical voice processing in autism », *Nat. Neurosci.*, 7 (8), p. 801-802.
21. Baron-Cohen S. *et al.* (1999), « Social intelligence in the normal and autistic brain : An fMRI study », *Eur. J. Neurosci.*, 11 (6), p. 1891-1898. Critchley H. D. *et al.* (2000), « The functional neuroanatomy of social behaviour : Changes in cerebral blood flow when people with autistic disorder process facial expressions », *Brain*, 123 (Pt 11), p. 2203-2212. Hubl D. *et al.* (2003), « Functional imbalance of visual pathways indicates alternative face processing strategies in autism », *Neurology*, 61 (9), p. 1232-1237. Pierce K. *et al.* (2001), « Face processing occurs outside the fusiform "face area" in autism : Evidence from functional MRI », *Brain*, 124 (Pt 10), p. 2059-2073. Schultz R. T. *et al.* (2000), « Abnormal ventral temporal cortical activity during face discrimination among individuals with autism and Asperger syndrome », *Arch. Gen. Psychiatry*, 57 (4), p. 331-340.
22. Hadjikhani N. *et al.* (2004), « Activation of the fusiform gyrus when individuals with autism spectrum disorder view faces », *NeuroImage*, 22 (3), p. 1141-1150.
23. Pelphrey K. A., Morris J. P. et McCarthy G. (2005), « Neural basis of eye gaze processing deficits in autism », *Brain*, 128 (Pt 5), 1038-1048.

direction du regard par la technique, que j'ai déjà mentionnée au début de ce chapitre, met en évidence, chez les personnes atteintes d'autisme, une perception visuelle anormale des interactions sociales[24]. Si on demande à des personnes non autistes de décrire l'état émotionnel de quelqu'un à partir d'une photographie, et que l'on enregistre alors la direction de leur regard, on s'aperçoit qu'il existe une façon reproductible de regarder le visage photographié pour fournir la réponse : le regard se porte sur les yeux et la bouche, et le tracé enregistré ressemble alors à un « triangle à l'envers[25] ». Les tracés obtenus chez des autistes à qui on demande d'effectuer la même tâche, permettent de mieux comprendre pourquoi il leur est difficile de donner la réponse correcte : leur regard ne se porte que rarement sur les yeux, et est plutôt dirigé vers le bas du visage photographié.

Un autre aspect très étudié dans l'autisme par l'équipe que dirige Utah Frith à Londres, concerne la « théorie de l'esprit », c'est-à-dire la capacité à se représenter la pensée de l'autre[26]. Un test simple, que l'on peut faire avec des enfants (qui donnent la bonne réponse en général dès l'âge de 4 ans), est de montrer une image d'un bonhomme entouré de quatre friandises : les yeux du bonhomme sont tournés vers les bonbons m&m's et il sourit. La question posée à l'enfant est : « Quelle friandise préfère ce bonhomme ? — Les m&m's. — Pourquoi ? — Parce qu'il les regarde. » Quand on pose ces questions à des enfants atteints d'autisme, ils ne savent pas dire ce que le bonhomme préfère, comme si la donnée « information véhiculée par le regard » était mal ou pas traitée par leur cerveau[27]. Par ailleurs, dans une récente série d'études, Pelphrey et ses collaborateurs ont étudié la modulation du degré d'activation du sillon temporal supérieur par le contexte social, et ont montré des anomalies chez les sujets atteints d'autisme[28]. Par exemple, ils ont étudié l'activation du sillon temporal supérieur lors de la visualisation d'un regard congruent ou non congruent par rapport à une cible visuelle, ou lorsque le regard traduit l'intention d'engager ou de se retirer d'une interaction sociale. Ils ont montré que le sillon temporal supérieur est sensible au contexte social

24. Klin A. *et al.* (2002), « Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism », *Arch. Gen. Psychiatry*, 59 (9), p. 809-816.
25. Pelphrey K. A. *et al.* (2002), « Visual scanning of faces in autism », *J. Autism Dev. Disord.*, 32 (4), p. 249-261.
26. Frith C. D. et Frith U., (1999), « Interacting minds – a biological basis », *Science*, 286 (5445), p. 1692-1695. Frith U., Morton J. et Leslie A. M. (1991), « The cognitive basis of a biological disorder : autism », *Trends Neurosci.*, 14 (10), p. 433-438.
27. Baron-Cohen S. *et al.* (1999), *op. cit.*
28. Pelphrey, Morris et McCarthy (2004), *op. cit.*

dans lequel se produit un changement du regard, c'est-à-dire si le regard est perçu comme compatible ou non avec l'attente par rapport à l'intention de la personne effectuant le mouvement des yeux[29].

Ceci suggère un rôle pour le sillon temporal supérieur dans la représentation de l'action intentionnelle, et non pas seulement dans la perception des mouvements. Cette intentionnalité peut être étudiée en imagerie. Une étude, faite à Londres par Castelli et ses collaborateurs, utilise comme stimuli deux triangles en mouvement[30]. Dans une première animation, les mouvements des triangles sont aléatoires ; dans la seconde, les mouvements conduisent à imaginer une histoire entre les triangles (ils dansent, se disputent, « se font un câlin »…). Cette animation suscite la plupart du temps des descriptions en termes d'états mentaux attribués aux triangles par les participants. Dans ce cas, un réseau cérébral particulier est activé chez le sujet sain : d'une part le cortex visuel, et en particulier la région V5, qui traite les mouvements dans le cortex visuel, mais également le cortex préfrontal, le sillon temporal supérieur et une région temporo-basale périamygdalienne. Chez les sujets atteints d'autisme, l'activation de la région V5 est également retrouvée (le mouvement est bien perçu), ainsi qu'une faible activation du cortex préfrontal, mais aucune activation temporale n'est observée. Du point de vue comportemental, les sujets atteints d'autisme ont également moins tendance à raconter une histoire à partir de ce qu'ils ont vu. Cette même équipe a montré que la connexion fonctionnelle entre la région visuelle et la région temporale supérieure est réduite chez les personnes atteintes d'autisme par rapport aux personnes non autistes[31].

Ces arguments nous font penser que des anomalies précoces du sillon temporal supérieur pourraient être impliquées dans le mécanisme des troubles sociaux et de communication dans l'autisme, en considérant que des anomalies de réseaux, plutôt que des anomalies d'une seule région cérébrale, sous-tendraient l'autisme[32]. Le sillon temporal supérieur est en effet très lié à d'autres régions du cerveau social, telles que la zone de reconnaissance des visages, le cortex orbito-frontal et l'amygdale (figure 2). Toutes ces régions semblent anormalement activées chez des individus autistes confrontés à des tâches de cognition sociale. Nous

29. Pelphrey K. A. *et al.* (2003), « Brain activation evoked by perception of gaze shifts : the influence of context », *Neuropsychologia*, 41 (2), p. 156-170.
30. Castelli F. *et al.* (2002), « Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes », *Brain*, 125 (Pt 8), p. 1839-1849.
31. *Ibid.*
32. Zilbovicius M. *et al.* (2006), « Autism, the superior temporal sulcus and social perception », *Trends Neurosci.*, 29 (7), p. 359-366.

étudions actuellement les connexions anatomiques qui sous-tendent ce réseau.

Figure 2 : *Le sillon temporal supérieur et ses connexions.*

(a) *Le sillon temporal supérieur est une région associative multimodale fortement connectée aux régions frontale, pariétale, limbique, auditive et visuelle*[33].

(b) *Castelli et ses collaborateurs*[34] *ont étudié la connectivité entre les régions cérébrales activées lors de la visualisation de triangles animés, qui sollicitait la mentalisation. Ils ont montré que, alors que les connexions de la région du sillon temporal supérieur avec les régions frontale et limbique étaient normales, celles du cortex occipital visuel extrastrié avec la région du sillon temporal supérieur étaient significativement réduites chez les sujets atteints d'un trouble du spectre autistique (flèche en pointillé). Ceci suggère que les difficultés de ces sujets relatives à la « théorie de l'esprit » testée par ces* ***animations*** *pourraient être liées à un défaut de transmission, du cortex occipital vers le sillon temporal supérieur, d'une information importante sur les mouvements des triangles.*

Par ailleurs, l'équipe de généticiens dirigée par Thomas Bourgeron a identifié, dans plusieurs familles comportant des individus atteints de troubles du spectre autistique, des mutations dans certains gènes codant pour des protéines qui semblent impliquées dans la formation ou la maturation des connexions synaptiques entre les neuro-

33. Basé sur les études anatomiques de Seltzer B. et Pandya D. N. (1978), « Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey », *Brain Res.*, 149 (1), p. 1-24 ; adapté de P. Gloor (1997).
34. Castelli F. *et al.* (2002), *op. cit.*

nes, et qui pourraient donc contribuer à la constitution de réseaux neuronaux intracérébraux[35]. De plus, l'étude comportementale de souris porteuses d'une mutation dans l'un de ces gènes a mis en évidence des anomalies des vocalisations sociales[36], qu'il est tentant de rapprocher des anomalies de la perception de la voix humaine observées chez les personnes autistes. Une des pistes de recherche pour les prochaines années est de tenter d'identifier le rôle de certaines variations génétiques (plus ou moins fréquentes) dans le développement structural et fonctionnel du cerveau humain. L'imagerie fonctionnelle du cerveau occupera sans aucun doute une place importante dans cette démarche.

Les études familiales et de jumeaux suggèrent que la part génétique dans le déterminisme de l'autisme est très importante[37]. Par ailleurs, nous savons que notre capacité à percevoir, reconnaître, interagir et comprendre autrui est déterminée non seulement par nos gènes (l'inné), mais aussi par nos expériences et notre environnement (l'acquis). Ainsi, les enfants naîtraient avec une capacité innée à reconnaître l'autre, à s'intéresser à lui et à se mettre en relation avec lui. Ceci a été constaté par différentes études chez le nouveau-né. Le nouveau-né a une préférence visuelle pour le visage humain. Il tourne la tête ou les yeux dans la direction d'un son. Il est particulièrement sensible à la voix humaine, notamment à celle de sa mère. Dès le premier jour, le nouveau-né bouge selon des rythmes précis, coordonnés à la voix humaine. Il réagit aux stimuli extérieurs, et va rapidement être capable d'imiter autrui, en particulier sa mère, lorsqu'elle ouvre ou ferme les yeux, sourit… Mais son cerveau est encore très immature. Au cours des premières années de la vie, des modifications importantes se produisent, aussi bien dans la structure que dans le fonctionnement du cerveau. Ces modifications sont très dépendantes de l'expérience. Ainsi, cette attirance innée pour autrui se transforme en une réelle capacité du cerveau à reconnaître et à interagir avec l'entourage. Des régions précises du cerveau se spécialisent dans la reconnaissance des visages, de la voix, des gestes, du regard, et des connexions s'établissent entre les différentes régions cérébrales.

---

35. Persico A. M. et Bourgeron, T. (2006), « Searching for ways out of the autism maze : genetic, epigenetic and environmental clues », *Trends Neurosci.*, 29 (7), p. 349-358.
36. Jamain S. *et al.* (2008), « Reduced social interaction and ultrasonic communication in a mouse model of monogenic heritable autism », *Proc. Natl. Acad. Sci. USA*, 105 (5), p. 1710-1715.
37. Persico et Bourgeron, *op. cit.*

L'un des résultats les plus récents concernant l'autisme provient de l'équipe dirigée par Ami Klin, et de ses expériences d'*eye-tracking*[38]. Ces chercheurs ont pu montrer, chez des enfants de 15 mois, que, lors d'une interaction avec une personne maternante, le regard de l'enfant non-autiste se porte essentiellement sur les yeux, tandis qu'un enfant du même âge qui sera plus tard reconnu comme autiste ne regarde pas les yeux mais le bas du visage. Ce résultat suggère que des troubles subtils du comportement peuvent être identifiés très précocement chez les enfants atteints d'autisme. Ainsi, il se pourrait qu'il existe, à l'origine de la trajectoire développementale qui conduit à l'autisme, une anomalie du cerveau qui empêche l'enfant de développer une expertise du monde social, et qui, au contraire, peut le conduire à s'intéresser plus particulièrement au « monde physique » (des objets en rotation, les chiffres…). Une des hypothèses est que cette anomalie initiale siège dans la région du sillon temporal supérieur, et qu'elle conduit à des connexions anormales entre les principales régions du cerveau social.

Il reste beaucoup à découvrir sur le dysfonctionnement du cerveau de Paul, l'enfant autiste, mais nous comprenons déjà un peu mieux qu'auparavant « l'origine » de ses difficultés relationnelles, en particulier grâce aux résultats récents issus de l'imagerie fonctionnelle du cerveau. Paul ne dirige pas son regard vers autrui probablement parce que l'information véhiculée par le regard de l'autre n'est pas traitée correctement dans son cerveau, si bien que pour Paul, ce regard a perdu toute signification particulière. Paul ne distingue pas non plus une voix d'un autre son, et ne prête donc pas une attention particulière aux personnes qui lui parlent. Plus généralement, Paul n'intègre probablement pas correctement les différentes informations sensorielles qui lui parviennent de ses congénères. Il ne joue pas avec les autres enfants parce qu'il est incapable d'analyser leurs gestes, leurs mimiques, leurs mots, leurs intentions. Le cerveau de Paul ne s'est pas construit de façon à nous comprendre, de sorte que Paul est incapable d'interagir « normalement » avec nous, et paraît s'isoler.

38. Jones W., Carr K. et Klin A. (2008), « Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism spectrum disorder », *Arch. Gen. Psychiatry*, 65 (8), p. 946-954.

V

MUSIQUE DU LANGAGE ET LANGAGE DE LA MUSIQUE

Poésie et musique : la pensée audible

par Michael Edwards

1

Je saisis l'occasion de notre colloque pour penser la poésie et la musique sous l'angle particulier du « dialogue humain ». Cet angle peut paraître curieux et peu prometteur, car ni un poème ni une œuvre musicale ne s'orientent à l'origine vers autrui, et la recherche du poète et du musicien, le travail que chacun accomplit, ne visent pas en premier lieu le rapport à l'autre. Cette approche est pourtant utile. Elle rappelle, d'abord, que la poésie est en effet une parole, qu'elle se lit à haute voix, que même écrite elle est faite de sons, qu'elle vibre dans la bouche et dans l'oreille, et qu'en poursuivant ses aspirations les plus élevées, elle ne quitte pas ces zones de la vie du corps. En poésie, la parole habite le corps entier et le rassemble dans la dynamique du souffle et du rythme, comme elle habite et concentre l'esprit, et comme elle laisse deviner la présence de l'inconscient, par l'obscurité de sa genèse et par certaines impulsions de son écriture. C'est par sa capacité d'élaborer un monde sonore et, ce faisant, de réorganiser le corps et toute l'âme du corps, que la poésie côtoie la musique.

Sous la perspective de notre dialogue les uns avec les autres, il est clair aussi que la musique et la poésie, dans la mesure où elles ont « quelque chose à dire », compliquent démesurément l'affaire. En parlant, nous arrivons tant bien que mal à nous comprendre (quoique aucun langage ne parvienne à transmettre le vécu de celui qui parle ni la façon dont il éprouve ses paroles), alors que le musicien livre à ses auditeurs un espace-temps sonore éloigné des usages de l'échange, et que le poète évite, dès les premiers sons, les premières cadences, le premier morceau de syntaxe modifiée, d'aller droit au but. Dans les deux cas, ce n'est ni par malice ni

uniquement parce que le poète et le musicien ont d'autres projets en tête, mais parce qu'ils cherchent à *dire* bien plus, de façon différente, et avec un surcroît de difficulté qui donne à ce qui est dit une autre épaisseur et finalement (pour anticiper) une plus grande joie.

Cette complication vient d'abord de ce que le musicien pense en musicien – de toute évidence, puisque dans l'acte créateur il pénètre, si je l'ai bien compris, dans un univers formel tremblant de virtualités à la fois intellectuelles et sensuelles qu'une pensée non verbale va découvrir et ordonner – et de ce que le poète pense en poète, vérité un peu moins manifeste, du fait qu'il utilise des mots et qu'il dit souvent des choses qui, sorties de leur contexte, pourraient se dire hors poésie : « Cueillez votre jeunesse », « La nature est un temple », « La chair est triste, hélas ! et j'ai lu tous les livres ». Il suffit de songer au vers, cependant, qui, en passant au vers suivant, interrompt, en un sens, ce que l'on était en train de dire, pour comprendre que la pensée du poète au travail diffère volontairement et allègrement des façons de penser qui peuvent être les siennes dans les situations courantes, et qu'en créant continuellement une tension, une attention, une attente, un bref silence empli de bruits à venir, il creuse, comme le mot « vers » l'indique, le sillon de la langue, comme il creuse, par son inventivité, le futur.

La complication vient aussi de ce que le poète écoute non seulement le son des mots, mais encore, comme on le sait, le murmure des connotations, des allusions et de la longue mémoire des mots, qui peut ressusciter les moments innombrables de l'histoire d'un peuple et rendre présentes les œuvres littéraires du passé. Plus que d'autres genres d'écrivains, me semble-t-il, et certainement d'une manière plus réfléchie, le poète est conscient qu'une langue fourmille de pensées et de souvenirs. En se déplaçant avec bonheur dans la vie privée des mots, parmi les sons qui les animent et toutes sortes de sous-entendus qui constituent leur profondeur, il cherche ce qu'il veut dire par le seul moyen qui lui permette de le trouver, par l'*indirection* (pour employer le mot bien utile de Polonius) qui le lui révèle peu à peu et, ce faisant, l'enrichit prodigieusement. Pour le lecteur, le poème paraîtra un *autrement dit*, une allégorie selon le sens originel du mot, une manière autre de dire réinventée à chaque fois qu'un poème vient à être, et dont la poésie a pour tâche de signaler sans cesse l'existence. La musique aussi est un *autrement dit*. Elle peut paraître un monde à part, un univers de sons minutieusement analysés et maîtrisés où l'on aimerait se réfugier en fuyant les difficultés du moi et les maladies du monde : une disposition de rapports abstraits qui ne toucheraient que l'oreille. Il suffit, cependant, de réfléchir sur l'effet que produit en nous-mêmes la musique la moins « figurative », pour ainsi dire, la plus tournée

vers elle-même, vers des questions de technique et de forme à résoudre (en prenant au hasard parmi les exemples évidents : l'*Art de la fugue* de Bach ou les *24 préludes et fugues* de Chostakovitch), pour sentir à quel point la musique nous concerne, combien ses calculs remuent les émotions, éveillent l'imagination, suggèrent de nouvelles configurations pour l'expérience humaine et même pour notre lecture du monde phénoménal. Avec toutes les apparences d'un domaine *sui generis*, la musique aussi est une allégorie, une façon de dire autrement, dont l'écoute philosophique aide à mieux comprendre l'*autrement dit* de la poésie.

La complication vient également de la difficulté des choses que « disent » la musique et la poésie. Même simple, la musique initie, dès ses premières notes, ses premiers accords, à une manière d'être où tout ce que nous reconnaissons d'humain n'a d'existence que par sa participation à ces mathématiques voluptueuses, et la plupart du temps, on le sait, elle découvre des états de conscience, des subtilités d'émotion et de pensée, au-delà du connu et impossibles à formuler sans elle. En poésie, dire « je » signifie déjà pour le poète se dédoubler, rencontrer un bizarre personnage suspendu entre la première personne et la troisième, et se trouver dans un monde où tous les repères sont à créer. (Ce « je » n'est pas moins étonnant lorsqu'on avait l'impression de simplement s'exprimer : on peut se troubler dans ce cas comme on se trouble en regardant sa photo.) Pour le lecteur, les premiers mots, si clairs soient-ils, sont un seuil à franchir, vers une région où tout ce qui est nommé, si même il s'agit d'objets familiers ou d'actes ordinaires, devient inhabituel dans la lumière neuve diffusée par un langage lui-même renouvelé. On a l'impression d'être attiré par-delà le monde spirituel exploré, et l'on comprend que les moyens usuels de communiquer par le langage n'auraient pas suffi pour la créer. La poésie interrompt le dialogue, les conversations incessantes que nous avons avec les autres et avec nous-mêmes, parce qu'elle pressent l'avènement de ce qui le dépasse, et qui sera peut-être, à la longue, une certaine joie ou un certain sens de l'exil.

2

Une fois formées, poésie et musique sont, chacune de son côté, une pensée audible. La réponse de l'auditeur qui a saisi le sens, la direction, de l'œuvre, est proprement : « Je vous entends. » Mais dans cette interruption du cours des choses, cette complication de l'entretien des gens, qu'effectuent exactement la poésie et la musique ? La poésie commence

– pour moi en tant que poète, mais aussi en tant que lecteur des autres poètes, chez qui je crois trouver (en m'aveuglant peut-être par une incapacité de reconnaître la vraie altérité de ce qui m'est étranger) une pensée semblable exprimée autrement et vécue de manière différente – à la fois dans l'émerveillement devant ce qui nous habite et nous entoure et l'élan vers l'éloge, dans le sentiment que nous n'avons cependant pas un rapport authentique avec nous-mêmes, autrui, et ce monde qui nous captive, les trois personnes du verbe étant sérieusement viciées, et dans le pressentiment du possible du moi comme du réel qui le soutient. La poésie est la recherche de ce qui est, et de ce qui pourrait être, la quête de soi n'obligeant pas le poète à devenir, selon une voie ouverte par un certain romantisme et prolongée par la psychanalyse entre autres, l'astronaute de l'espace du dedans, la quête du réel n'impliquant pas la poésie raréfiée, issue en partie d'une réaction contre le romantisme, qui exclut du réel poétique toute la réalité, devenue « prosaïque », de l'histoire, de la société, de la politique, de l'approfondissement du bien et du mal. La poésie est surtout la poursuite de ce que Rimbaud discernait selon le génie de l'adolescence : la vraie vie, que l'on devine, mais qui semble, la plupart du temps, hors d'atteinte.

Dans cette recherche, la poésie ne produit pas un savoir mais une connaissance. Le savoir, qui opère à distance, permet d'apercevoir *ce qu'il y a* et de comprendre pourquoi et comment. La connaissance, qui est transitive, nous met en relation avec ce qui nous sollicite ; elle a besoin, pour s'effectuer, d'un langage envahi par le superflu, par tout ce qui appartient en propre au langage, mais qui ne semble pas nécessaire pour véhiculer un sens : par ces sons, rythmes, silences, insinuations, dont j'ai déjà parlé. Pour Wagner, dans sa *Lettre sur la musique*, le rythme et ce qu'il appelle « l'ornement (presque musical) de la rime » servent à saisir et à gouverner les sentiments. On pourrait dire plus largement, hors du contexte romantique et opératique, que tout le luxe apparent ou caché de la parole poétique (dont beaucoup d'éléments rapprochent en effet la poésie de l'art musical) a pour rôle essentiel de nous mettre en présence de nous-mêmes, de l'autre et du monde. Tous les détails, disons, techniques portent l'assez grave responsabilité, dans le dialogue entre poète et lecteur, de l'orientation de l'être, dans ses efforts pour vivre au diapason d'un réel enfin trouvé. D'où le danger d'une écriture malavisée ou autosuffisante. « La vraie éloquence se moque de l'éloquence », dit Pascal, et la fausse éloquence se moque de la vie.

Mais comment tous ces jeux du langage, de la rhétorique, de la prosodie, œuvrent-ils ? Je remarque, non seulement en me lisant moi-même, mais en lisant toutes sortes de poèmes de poètes très divers, que la parole

poétique change ce qu'elle cherche à connaître. En respectant le plus délicatement possible l'en-soi de l'être ou de la chose, en s'abstenant de le dénaturer par les caprices navrants que la fantaisie est fertile à concevoir dans l'acte poétique, elle vient néanmoins à le présenter sous un autre jour, selon un rythme, par exemple, qui lui donne une nouvelle allure, ou une métaphore qui le situe dans un rapport inattendu mais juste. En modifiant, non la chose mais la perception que nous en avons, elle en découvre le possible, comme si la seule façon de connaître vraiment quelque chose était de déceler ce qu'il pourrait être – comme si l'être même d'une personne, d'une présence de la nature, d'un objet quelconque, était précisément ce qu'il est en puissance. Dans ce nouveau rapport créé avec l'être du réel, la parole poétique modifie aussi, quelque peu, l'autre terme du rapport, le moi du poète ou du lecteur qui s'y aventure.

Sous l'angle du dialogue, ce pouvoir heuristique, cette clairvoyance de la poésie, nous rappelle que dans tout vrai contact les personnes qui se rencontrent sont changées. Sous l'angle de la musique, les virtualités du langage, qui sont libérées par l'exercice irraisonné puis raisonnable de la poésie, rappellent le matériau varié et autonome à la disposition du compositeur. La « musique » du langage en poésie, les sons rendus audibles, le rythme qui entre dans le corps, les jeux de mots, de grammaire et de syntaxe, constituent, du reste, une musique pour l'esprit aussi complexe, aussi détaillée et aussi attirante à sa manière que celle du musicien. Et tout cet effort, tout ce travail heureux mais pénible aussi, rappelle, par la nécessité dont il est revêtu, que nous sommes en effet étrangers à nous-mêmes, que le réel que nous cherchons garde ses distances et se révèle terriblement imparfait, et que le dialogue humain n'est pas seulement une affaire de société et de savoir, mais que nous sommes embarqués, que nous avons été plongés, comme le lecteur d'Homère selon Horace, *in medias res*, au milieu d'une histoire turbulente de vie et de mort, et que nous nous entretenons d'une solution possible. Si « l'harmonie des sphères » peut encore éveiller en nous des résonances, c'est sans doute, comme pour Shakespeare, en figurant simplement une consonance de l'être, un concert des phénomènes, qui pourraient exister mais dont nous sommes (définitivement ou provisoirement) exclus. La « musicalité de tout » que Mallarmé croyait percevoir s'explique comme une réinterprétation symboliste de ce chant des corps célestes, mais, si attrayante qu'elle soit, elle nous leurre en ignorant le désordre du monde où nous nous trouvons (pauvreté, violence, deuil, mort, injustice…). La poésie cherche malgré tout, malgré le malheur du Tout, non pas à couvrir l'univers de ses convictions, mais à faire apparaître le possible de ce qu'elle touche, le futur présent en nous et dans le monde ambiant. Après les idées presti-

gieuses de l'harmonie des sphères et du grand théâtre du monde, le poète recherche, dans le monde d'ici, la poésie du réel.

3

Ce qui soulève par un autre biais la question du sens d'un poème, qui devient urgente, du reste, dans le contexte du dialogue. Nous savons qu'un poème n'est pas une façon élaborée de dire ce qu'on aurait pu dire en prose, qu'il ne signifie pas par les quelques phrases qui tenteraient de le résumer, mais cette évidence ne répond pas à la question, n'indique pas ce que le poème accomplit. J'ai appelé un poème, comme un morceau de musique, une pensée audible ; mais il faut préciser, dans les deux cas, que l'on entend moins une pensée formée par l'œuvre finie que l'acte de penser qui produit l'œuvre. Il s'agit, en effet, moins d'une pensée que d'un penser, ou, comme on peut le dire avec clarté selon l'empirisme de l'anglais, moins d'un *thought* que d'un *thinking*, où *think-ing* suppose l'activité de la personne dans le temps. Et ne concluons pas que le poète, comme le lecteur, s'intéresse aux processus de la pensée, à la manière de dire, aux dépens de ce qui est dit, dans une écoute distinguant l'esthétique de tout le reste et séparant soigneusement le beau du bon et du vrai. Un poème est pour le poète un acte, qui est soumis au jugement comme tous les autres actes de sa vie. Il implique l'exercice de tout son être et, vers par vers, des choix ontologiques et existentiels, des décisions quant à la nature du monde abordé et des mises en rapport supposées justes avec les présences du réel. Pour le lecteur un poème est un événement, le passage de quelque chose dans un mouvement de pensée et d'émotion, dans un bruissement de sensations, dans un ébranlement de la langue. Afin de l'entendre, lui aussi a besoin de toutes ses facultés, que le poème appelle et aiguise, y compris les moins « poétiques » selon une certaine conception désinfectée de la poésie, celles qui permettent de reconnaître la souveraineté de ce qui est et les façons d'être et d'agir qu'il exige. Si je peux encore citer l'anglais (je suppose qu'il est profitable aux Français de connaître les moyens distinctifs de la langue anglaise, comme aux Anglais de connaître ceux de la langue française) : la formule « *this poem has meant a lot to me* » – que l'on traduirait par : « ce poème m'a particulièrement touché, ou a eu de l'importance, une certaine résonance, pour moi », mais qui dit littéralement : « ce poème a signifié beaucoup pour moi » – institue un lien entre le sens, le *meaning*, d'un poème, que l'on cherche à comprendre, et sa pénétration de l'être du lecteur, son *meaning*

profond, la manière dont il touche et modifie émotions, pensées, volonté même. Du côté du poète – dans ce dialogue où le lecteur *répond* à l'auteur –, la parole où il s'engage est le geste du poème entier. Ce qui est dit est toujours problématique : la voix du poème n'est jamais celle du poète hors poème (même s'il a l'intention de « s'exprimer », il devient, il change, dans l'effort poétique pour le faire), et le poète peut choisir de parler obliquement ou même d'affirmer, dans un contexte (ironie, mise en scène) soigneusement contrôlé, le contraire de ce qu'il pense. Ces autres formes d'*indirection* sont rendues nécessaires par le fait que le sens du monde et notre propre sens ne sont pas donnés. Le poète s'en tient, non pas nécessairement à ce qu'il dit, mais à ce que dit le poème, grâce à l'acte qu'il accomplit et l'événement qu'il constitue.

Si la poésie et la musique cherchent à compliquer, à rendre plus complexes, le *dire* et le rapport à autrui, le poème assumé comme acte et comme événement rend le comportement du poète responsable. Concevoir ainsi le poème permet aussi de revenir de façon sérieuse à la question de la vérité, à condition de se rappeler que la vérité en poésie n'est pas une affaire de formules et que la vérité qui lui convient est celle du vécu. Il oblige surtout à examiner à nouveau la forme. Parlant, dans *Éclats 2002*, de la gestation de sa musique, de ce qui « peut déclencher une réflexion, permettre de découvrir une solution », Pierre Boulez évoque « le vol d'un oiseau ou l'ombre d'une feuille […], une scène de théâtre, un dialogue, un extrait de film ». Un poète qui remarquerait, pour le départ d'un mouvement de l'imagination, un vol d'oiseau ou une ombre de feuille, resterait probablement proche de ces choses vues. Elles paraîtraient dans son poème comme la source et la demeure d'une recherche de connaissance, d'une recréation du sensible par la vertu de la parole poétique ainsi provoquée, d'une modification de notre perception des choses à partir, par exemple, de ce que l'on est en présence, dans les deux cas, de deux êtres vivants mais aussi, et avant tout, de leurs prolongements, *l'ombre* d'une feuille et *le vol* d'un oiseau ayant chacun, de manière différente, sa part de mystère. Si le poète est saisi par la poésie du réel, le musicien, sans du tout y être insensible, est attiré aussitôt par une forme. Un vol d'oiseau et une ombre de feuille suggèrent des lignes, des contours ; une scène de théâtre, des mouvements dans l'espace (je suppose) et des rapports de force ; un dialogue, des différences d'intonation et de longueur de phrases ; un extrait de film, des plans divers, un ajustement de couleurs, ou de noirs et de blancs, une gestion de vitesses. Le musicien vise la forme à créer, soit directement par une idée, soit à partir d'un phénomène qui le requiert ou, comme le dit aussi Boulez, du matériau qu'il veut entendre (c'est alors « le matériau qui suggère une utilisation formelle »). Les phénomènes qui captent son attention ne sont pas spéciale-

ment liés à un art du son, étant des lignes, des corrélations, des structures, qui pourraient aiguillonner également le peintre, le sculpteur, l'architecte, le chorégraphe, le mathématicien, le poète : tous les fous de la forme.

Mais c'est surtout la musique qui concentre l'attention sur cette folie salutaire. Le désir de forme, chez toutes sortes d'artistes et toutes sortes de scientifiques, est, à bien y penser, surprenant. Il n'est pas compris dans les théories de l'œuvre d'art qui privilégient, soit l'imitation (le rapport au monde), soit l'expression (le rapport à l'artiste), soit la destination (le rapport au lecteur, à l'auditeur, au spectateur). Il se peut, cependant, qu'il constitue le premier mobile de l'art. C'est comme si nous sentions un manque, une absence d'ordre, le corps-esprit que nous sommes cherchant dans la création ou l'expérience d'une œuvre une harmonie qu'il sait ne pas atteindre dans sa vie au jour le jour. La forme serait, au fond et quoi qu'il se passe dans la conscience de l'artiste, l'intuition d'un mieux-être pour la vie : pour le moi et pour le monde qui le porte. La forme « *means a lot to us* », nous touche profondément, non pas comme une enveloppe, comme une affaire uniquement de proportions et de relations à contempler, mais, à l'instar de l'aspiration à la forme dans les plantes, comme la vie de l'œuvre, le cœur de son être-là. D'où probablement la raison essentielle de ce désir de musique chez le poète. Elle ne nous détache pas non plus de la vie ordinaire, ou ne devrait pas le faire. C'est à celui qui rencontre l'œuvre d'y chercher des liens avec le reste de son expérience, et à l'artiste d'en faire autant dans l'œuvre qu'il crée, peut-être à la manière de Boulez, qui écrit dans le même passage : « Plus je compose, plus je tiens compte de l'accident et du quotidien ». Il dit aussi : « La création est […] un processus organique », la comparaison se faisant, non avec des formes végétales comme chez Ruskin, mais avec un fleuve (rien de plus naturel, en effet, que l'artifice), et même ceci, qui ne peut manquer de plaire à un Anglais : la création, « c'est un empirisme très dirigé, qui tient compte des obstacles, qui se développe selon une direction précise. C'est un empirisme organique ». L'œuvre est en rapport, dans le meilleur des cas, avec la vie organisée et avec l'expérience des hommes.

4

La forme vient de l'intérieur (ou ne vient pas) pour constituer l'être du poème, de l'œuvre musicale, et c'est par cet ensemble, ce tout, que nous passons le seuil d'une nouvelle connaissance. J'hésite à parler de la musique : auditeur inexpert pour qui elle est simplement une présence

permanente et indispensable, je suis obligé de deviner à partir de mon expérience de la création poétique. Je sens néanmoins que la musique aussi est une connaissance – mais il vaudrait mieux qu'un musicien le dise –, une manière, à chaque fois nouvelle et différente, de penser, de sentir l'émotion, d'habiter le corps : une manière d'être insolite. Si je peux citer mon livre *De l'émerveillement*, la musique « est une intelligence singulière, un corps singulier, et un rapport singulier entre l'intelligence et le corps ». Elle n'est pas un monde à part, mais une autre langue, qui nous parle autrement de tout ce qui nous concerne. On évoque, un peu approximativement, le « langage » de l'architecture, du dessin, du ballet, mais le langage de la musique représente vraiment, comme celui des mathématiques, une langue désirée, qui n'est pas édénique et qui ne résout pas nos problèmes, mais qui cherche au-delà du connu et du malheureux, et qui nous fait la promesse que la réalité meilleure qu'elle nomme existe. Contrairement aux mathématiques, la musique est aussi une langue qui nous parle à l'oreille.

C'est la forme du poème, de l'œuvre musicale, en tant qu'unité multiple, qui nous change ainsi et qui change notre perception du réel, selon l'*anaktisis*, ou recréation, qui me semble, plus que la *mimèsis*, la raison d'être de tous les arts. Et cette multiplicité ne devient une, comme j'ai commencé à le dire, que si la forme semble se créer à mesure que l'œuvre avance, pour finalement l'habiter entièrement, sans être imposée du dehors, cette impression d'un corps-âme animé de l'intérieur pouvant naître autant dans les structures fixes, comme le sonnet ou la sonate, que dans, par exemple, le vers libre ou la polytonalité détendue. C'est la forme, finalement, qui compose l'acte de l'œuvre musicale ou poétique, l'événement que celle-ci déroule.

Cette complication audiblement exagérée du dialogue, afin de trouver des choses plus essentielles à se dire et d'approfondir l'idée même de communication, rapproche la poésie et la musique et explique encore le désir de musique en poésie. La poésie est le langage qui tend vers la musique, non pas à cause d'une quelconque prééminence de la musique comme dans une certaine pensée du XIX^e^ siècle (Schopenhauer, Pater), mais en vue de dépasser et d'enrichir le sens par le son. On parle du va-et-vient en poésie entre le son et le sens, et de la tentative angoissée et jamais satisfaite de les concilier ; et il est vrai que cette oscillation et cet acharnement entrent souvent dans le travail du poète. Ce qui importe davantage, cependant, ce sont les sens supplémentaires, et finalement intégraux, créés par les sons, et surtout l'autre sorte de sens véhiculée, ou plutôt incarnée, par le langage en sage délire. La musique proprement dite de la poésie, les sons et les rythmes agissent, pour simplifier, sur le

corps, et ce qui est musical par analogie, comme la structure, ou l'agencement des significations annexes, ou le contrepoint des images et des émotions, sur l'esprit.

La poésie et la musique sont des recherches, et ce qu'elles trouvent et communiquent est avant tout, me semble-t-il, une certaine joie. Celle-ci ne vient pas du sujet du poème, qui est souvent affligeant, ni des émotions « représentées » par une œuvre musicale, qui peuvent être déplaisantes, mais de la forme, qui est en elle-même une joie, et de cette entrevision, ou *entraudition*, d'un mieux-être imaginé du moi et du monde auquel il s'ouvre. Le « plaisir » dont on a si longtemps parlé en poétique n'est ni une voie vers autre chose (habituellement, l'« instruction »), ni un simple hédonisme, mais l'objectif salutaire et révélateur de l'art.

Je note aussi, cependant, que dans le cas de la poésie, on garde le contact avec la vie ordinaire, qui se déroule parmi les mots. Quels que soient la langue dorée à laquelle il aspire, le ravissement verbal auquel il parvient, le poète traîne avec lui le limon du langage et le souvenir des besoins habituels dont celui-ci est le serviteur. Il n'abandonne pas tout à fait la pauvreté au jour le jour d'une race qui parle. Voilà une des choses qui rappellent au lecteur que, sorti de l'œuvre, où il a pu se croire dans un lieu paradisiaque, il revient dans le monde, éclairé, mais toujours présent.

Musique et parole
De l'acoustique au numérique

par Jean-Claude Risset

L'œil et l'oreille

Parole et musique se transmettent par l'entremise des ondes acoustiques. Le son est le médium privilégié de la communication humaine : qu'on songe aux rituels ancestraux de diverses civilisations, aux conversations, aux débats démocratiques, aux ventes à la criée, au téléphone portable et au baladeur mp3. Cela peut surprendre : s'il y a conflit entre vu et entendu, l'œil tend à prendre le pas sur l'oreille. Voir, c'est croire.

L'image fascine et perturbe : la vision n'est pas l'entendement. Pythagore et, plus près de nous, François Bayle recommandent l'écoute *acousmatique*, celle où il n'y a rien à voir. La vue met à distance, l'oreille rapproche. Dès avant la naissance, le son nous baigne : l'audition nous relie au monde environnant. Grâce à la propagation des sons, l'audition est une espèce de toucher à distance, doté d'une sensibilité exquise. Aussi incroyable que cela puisse paraître, nous pouvons entendre les vibrations d'une membrane même lorsque l'amplitude du mouvement est inférieure aux dimensions d'un atome d'hydrogène. Et les vibrations sonores font le tour de la tête : l'audition joue un rôle essentiel d'alerte, et l'évolution l'a optimisée pour qu'elle fournisse sur le monde environnant des informations élaborées utiles à la survie. À l'écoute, on fait inconsciemment une enquête sur l'origine du signal : on essaie d'y démêler plusieurs sources, leurs natures, leurs directions, leur distance. *Quoi ?* qu'est-ce qui produit ce bruit ? quel est son mode vibratoire (percussion, frottement, souffle…) ? *Où* ? dans quelle direction ? à quelle distance ? par-devant, par-derrière ? la source sonore est-elle puissante et lointaine ? ou douce et proche ?

Selon Leroi-Gourhan, l'usage des outils et de la parole est spécifique de l'espèce humaine. Les instruments de musique sont des outils qui prolongent le corps ; la voix émane du corps lui-même. Les sons transmettent des messages importants entre les animaux, mais on ne peut pour eux parler de langage articulé par la phonation.

Parole, musique

Parole et musique comportent des structures hiérarchiques. Luigi Rizzi indique qu'on peut formaliser la diversité des langues. Il faut le signaler : des dizaines d'années avant Noam Chomsky, le musicologue viennois Herbert Schenker avait proposé un schéma de grammaire générative pour rendre compte de la diversité des musiques tonales. Et, comme l'a montré Johan Sundberg, il existe des codes spécifiques de l'expressivité aussi bien pour la parole que pour la musique : ils apportent dans les deux cas des informations similaires, soulignant les frontières structurales, facilitant la distinction des catégories perceptives, et accentuant telle ou telle portion du message.

Hugues Dufourt distingue deux archétypes de sons musicaux : la *voix* – les sons qu'il faut conduire du début à la fin (c'est le cas pour les sons des cordes frottées ou des cuivres) – et la *percussion* – les sons qu'on lance et qu'on laisse vivre et mourir (comme ceux des cloches des carillons, mais aussi des instruments à clavier ou à cordes pincées). Pour la parole, on peut songer à une analogie : les voyelles et certaines consonnes. Mais, comme le remarque Claude Hagège, consonnes et voyelles sont en succession, jamais en simultanéité.

La parole enrôle l'audition et ses capacités de différenciation, mais elle tire surtout remarquablement parti de nos possibilités phonatoires. Comme l'ont proposé René Carré, Björn Lindblom et Peter MacNeilage, certaines particularités du conduit vocal humain facilitent l'émission de signaux vocaliques différenciés. À partir d'un système de différences phonétiques, les langues parlées ont développé des codes plus ou moins arbitraires qui leur sont spécifiques ; mais un message parlé peut être traduit d'une langue dans un autre. La parole est le véhicule majeur de la communication humaine. Dans les cités démocratiques grecques de l'Antiquité, chaque citoyen devait pouvoir suivre les débats sur l'agora, la place publique. La radio porte la parole à distance. Elle a multiplié le pouvoir galvanisant des discours de Hitler. Avec le téléphone, la communication parlée devient planétaire.

La musique est un jeu sonore, un luxe de l'écoute : « plaisir délicieux d'une occupation inutile » selon Henri de Régnier, une jouissance

exploitant de façon gratuite nos capacités sensorielles, motrices et cérébrales. « La joie sans les mots », disait saint Augustin de la jubilation musicale. On peut traduire un texte, mais la signification de la musique réside dans sa forme même : si l'on ne connaît pas le dialecte de celui qui parle, on peut entendre sa parole comme musique. Les disques ont contribué à rendre plus proches de nous des musiques lointaines dans le temps et dans l'espace. Dans sa diversité, la musique apparaît comme un langage universel qui s'appuie sur un fonds commun d'intersubjectivité.

Précédence ou coopération ?

Les amateurs d'opéra au XVIII^e^ siècle se rangeaient dans deux camps. *Prima la parola*, disaient les partisans de Gluck, pour qui la musique n'était qu'un accompagnement qui devait se plier aux exigences du texte. Non, la musique est reine, répondaient les partisans de Piccinni : *prima la musica*. Aux origines du dialogue humain : parole ou musique ?

Dans nos sociétés, la parole est indispensable, alors que la musique n'est qu'un agrément : si nécessité fait loi, l'antériorité de la parole paraît plus plausible. Cependant, l'utilité pratique et l'avantage évolutif de l'apparition du langage dans les sociétés humaines archaïques de chasseurs-cueilleurs ou d'agriculteurs ne sont pas si évidents qu'on pourrait le penser : certains, comme Jean-Louis Dessalles, ont avancé l'hypothèse que la parole originaire a dû jouer d'abord un rôle d'ordre politique, permettant aux meneurs d'homme de convaincre et d'entraîner. La musique est apparue très tôt dans les sociétés humaines. André Jolivet voyait dans les musiques originaires « l'expression magique et incantatoire de la religiosité des groupements humains » ; « religiosité » : du verbe *religare*, « relier ». Selon François-Bernard Mâche, le sentiment esthétique est déjà présent dans le chant des oiseaux, au-delà de la transmission de signaux de ralliement, de reconnaissance, de défense du territoire ou d'appel sexuel.

Plutôt que spéculer sur l'antériorité, on peut observer que musique et parole sont souvent associées – dans l'opéra occidental, mais aussi dans les cantillations rituelles – et qu'il peut être difficile de les disjoindre. Je vais évoquer diverses formes d'alliance entre la musique et la parole[1].

1. Je les ai illustrées par une vingtaine d'exemples sonores lors du colloque dont ce livre est issu.

Instruments « voisés »

Certains instruments de musique très anciens tiraient parti de la voix ou des organes vocaux.

Depuis les temps préhistoriques, les aborigènes australiens jouent du didjeridoo : un long tube creux qu'ils font vibrer par une phonation rythmée à une extrémité du tube. Le résultat sonore est très particulier : il combine certains aspects de la voix avec des caractéristiques imposées par le tube (on peut vérifier qu'il est à peu près impossible de siffler un glissando dans un tuyau). Hybridation qui annonce nos « synthèses croisées », cette fusion du vivant et de l'inanimé avait une signification rituelle.

Dans l'arc à bouche africain, le conduit vocal se comporte comme un filtre résonant pour les sons pincés de la corde. La guimbarde relève du même principe : le spectre étendu de la tige est altéré par les résonances – les *formants* vocaliques –, qui évoluent au rythme de la phonation.

Résonances vocales

Dans la parole, l'articulation modifie le son émis par les cordes vocales. La phonation peut aussi être utilisée musicalement. Les chanteurs peuvent apprendre à maîtriser les résonances de leur conduit vocal de façon à filtrer sélectivement certains harmoniques aigus du spectre de l'onde complexe issu des cordes vocales : alors ces harmoniques ne fusionnent plus avec le fondamental et donnent l'impression d'un chant « diphonique ». C'est ainsi que procèdent les moines bouddhistes tibétains pratiquant leur cantillation. La scission de la voix en sons multiples est également pratiquée avec une extrême précision en Mongolie et dans certaines régions de Sibérie : les chanteurs produisent des mélodies d'harmoniques planant au-dessus d'un bourdon ou de mélodies graves. Cette technique délicate a été reprise par des chanteurs contemporains comme Linda Vickerman, David Hykes, Frank Royon Le Mée, Irène Jarsky, Tran Quang Hai.

Les langues sifflées, mentionnées par Simha Arom, permettent aux bergers des vallées montagneuses de divers continents de dialoguer à plus

d'un kilomètre de distance. Ces langues peuvent transmettre des messages variés en mimant la structure phonétique. Elles ont été décrites par André Classe et René-Guy Busnel, et tout récemment par Julien Meyer dans une étude remarquable.

Cantillation et notation

La musique rituelle est multiple et diverse. Les offices religieux sont souvent accompagnés de cantillations sur des textes sacrés. Au Tibet, la musique aussi bien que le texte des chants bouddhiques ont fait très tôt l'objet de notations.

En Occident, la notation s'est développée dans le chant grégorien : elle a d'abord joué un rôle mnémonique, mais elle a aussi aidé à l'apparition de la polyphonie, spécialement à Notre-Dame de Paris au XII^e^ siècle avec Léonin et Pérotin. Depuis lors, le développement de la polyphonie caractérise la musique européenne.

En retour, la polyphonie a forcé la notation graphique à se décanter et s'enrichir. « L'œil écoute », écrit Claudel : sur une représentation graphique, la symétrie apparaît manifestement, et il devient naturel d'associer à une courbe mélodique la courbe symétrique par rapport à un axe horizontal ou vertical. Cela a probablement suggéré les transformations du contrepoint – renversement et récurrence –, dont l'usage élaboré ne semble pas exister dans les musiques de tradition orale. La parole n'a pas d'équivalent du contrepoint, où l'on peut fusionner ou distinguer les voix.

Pour un historien comme Geoffroy Hindley, c'est l'exemple de la notation musicale qui explique que les coordonnées cartésiennes sont nées en Occident plutôt qu'ailleurs ; l'explosion scientifique en a découlé, avec la mise en équation des lois de Newton, la prévision des trajectoires balistiques et la notion de déterminisme.

Dans les motets – musiques sur des mots –, le chant suit la structure de la phrase. Au XIV^e^ et au XV^e^ siècle, les chants polyphoniques pouvaient atteindre une grande complexité : jusqu'à trente-six voix indépendantes dans le *Deo gratias* d'Ockeghem.

L'opéra

En 1600, Peri et Caccini présentent le drame lyrique *Euridice* au mariage d'Henri IV : on considère qu'il s'agit du premier opéra, précédant de peu *Orfeo*, chef-d'œuvre expressif de Monteverdi. L'opéra allie au théâtre musique vocale et instrumentale. Les « récitatifs » permettent de mieux comprendre le texte ; la musique reprend son hégémonie dans les arias. Sundberg a montré que les voyelles aiguës fortissimo des chanteuses ne peuvent guère suggérer d'autres voyelles que « a » : plus la hauteur monte, plus la chanteuse ouvre la bouche pour augmenter l'intensité en accordant la cavité vocale au fondamental de la note chantée. Il a également révélé que le style de chant propre à l'opéra – « dans le masque », en abaissant le larynx – avait une fonction ergonomique : il crée un nouveau formant aidant la voix des chanteurs à émerger du tutti de l'orchestre.

Je renvoie ici au texte, publié dans ce même ouvrage, où Claude Hagège survole avec brio l'histoire glorieuse de l'opéra, de Gluck et Mozart à Verdi, Bizet, Wagner et au-delà.

Modernité musicale et vocale

De même que les recherches acoustiques de Mersenne et Sauveur avaient marqué la théorie musicale de Rameau, les recherches acoustiques d'Ohm et Helmholtz, évoquées dans ce volume par Jacques Bouveresse, ont laissé indirectement leur trace sur les méthodes de composition. Elles ont en particulier favorisé la séparation de « paramètres » des notes de musique : plus tard, la technique sérielle traitera ces paramètres séparément. La crise de la musique tonale au début du XX^{e} siècle fait le constat de l'usure du langage tonal, de plus en plus battu en brèche par les innovations expressives de Gesualdo à Chopin, Liszt et Wagner. Cette crise est bien sûr à rapprocher d'autres ruptures : cubisme, abstraction et expressionnisme en peinture, quanta et relativité en physique, évolution des espèces, décadence de l'Empire austro-hongrois, prise en compte de l'inconscient.

La technique vocale, va à l'occasion, subvertir les contraintes traditionnelles du chant occidental : ainsi Schoenberg a-t-il recours au

Sprechgesang – parlé-chanté, intermédiaire entre parole et chant – dans son « mélodrame » *Pierrot lunaire* (1913), sur des textes d'Albert Giraud traduits en allemand.

Après 1940, les jazzmen du be-bop (après 1940) ont inventé le scat, un mode de chant délié et quasi instrumental, illustré notamment par Ella Fitzgerald, Sarah Vaughan, Anita O'Day, Dizzy Gillespie. Les musiciens d'« avant-garde » (une expression d'origine militaire qui a ses détracteurs) ont exploré de nouvelles techniques vocales. La *Sequenza* pour voix de Luciano Berio, les *Récitations* de Georges Aperghis mêlent rire, soupirs, phrases et vocalises rapides. Mauricio Kagel introduit un « théâtre musical » porteur de critique sociale. Dans les *Maulwerke* de Dieter Schnebel, les *Aventures* de György Ligeti, dans mes propres *Dérives*, les chanteurs parlent des langues imaginées et pratiquent des « effets de bouche ».

Les nouveautés technologiques sont aussi un ingrédient important de l'innovation vocale, même si l'acoustique y tient la plus grande part. Dans *Ecuatorial* (1934), Varèse, pour la première fois, sonorise un chanteur : après 1945, tous les groupes pop l'ont suivi. Le style vocal de Jimi Hendrix est inséparable de la distorsion imposée électriquement à la guitare électrique. Le rap est apparu après 1970, avec le Sugar Hill Gang : un texte intelligible y est rythmé et musicalisé d'une façon très particulière, préservant soigneusement l'intelligibilité. Le texte porte le plus souvent une protestation sociale : on peut parler ici d'une musique de ralliement.

Cela nous amène à quitter le domaine acoustique pour passer au son électrique.

Son électrique et synthèse des sons

Vers 1875, l'apparition de l'enregistrement sonore et du téléphone a bouleversé nos rapports avec le son. Jusqu'à cette date, tous les sons étaient des perturbations audibles, mais impalpables et éphémères, produites par des vibrations mécaniques, à l'exception du tonnerre et du canon. L'enregistrement reproduit les sons en l'absence de leur cause mécanique initiale. Et les « transducteurs » à l'œuvre dans le téléphone (microphone, haut-parleur) traduisent les ondes acoustiques en vibrations électriques et vice versa, conférant au traitement du son les ressources des technologies électriques.

La « révolution électrique » (Hugues Dufourt) a eu les conséquences que l'on sait. L'usage de l'électricité a longtemps été confiné à la diffusion des sons, par la radio, le téléphone et l'enregistrement. Cependant, dès

avant 1900, Thaddeus Cahill avait construit aux États-Unis un ensemble de dynamos à sons, le Dynamophone : on pouvait en jouer comme d'un instrument… mais il pesait trois cents tonnes. Aussi Cahill avait-il proposé des abonnements pour recevoir la musique du Dynamophone par téléphone, d'où son autre nom de Telharmonium. Les échos de cette réalisation étonnante mais éphémère ont mis en marche l'imagination d'Edgard Varèse : dès 1917, celui-ci n'a cessé d'appeler de ses vœux l'extension du vocabulaire de sons musicaux par la construction de machines pour *produire* des sons à partir de vibrations électriques, et pas seulement pour reproduire des sons d'origine acoustique. Après 1950, Varèse pourra tirer parti des ressources des musiques électroacoustiques avec *Déserts* et *Poème électronique* ; et il s'est beaucoup intéressé à la mise en œuvre par Max Mathews de la synthèse des sons par ordinateur.

Le son numérique (en anglais *digital*) bénéficie à la fois des ressources de l'ordinateur et du codage d'une fonction continue en nombres « discrets » (discontinus). L'ordinateur peut être programmé, ce qui donne lieu à une immense variété d'usages – en particulier pour analyser, produire et transformer le son.

L'ordinateur a produit des sons musicaux avant de parler. En 1957, Mathews a mis en œuvre aux Bell Telephone Laboratories la première conversion de sons en nombres – l'enregistrement numérique, utilisé dans les disques compacts – et la première synthèse de sons musicaux. En 1962, collaborant avec John Kelly, il a fait pour la première fois chanter par un ordinateur en synthétisant *Daisy, a bicycle built for two* – une chanson très populaire aux États-Unis. L'ordinateur, qui avait aussi produit l'accompagnement, avait un accent électrique marqué, mais sa parole était déjà intelligible. La synthèse imitant la voix parlée a été produite par une simulation du conduit vocal, qui est animé par l'articulation. Stanley Kubrick s'est souvenu cette réalisation remarquable dans son film *2001 : l'Odyssée de l'espace*, où l'ordinateur retombant en enfance se souvient de sa première chanson.

Les possibilités du son numérique sont en principe illimitées, mais il faut les conquérir. L'ordinateur permet de construire un son « sur plans », mais c'est l'effet auditif qui importe, que ce soit pour la parole ou pour la musique. De nombreux travaux visent à améliorer l'intelligibilité et la qualité de la voix de synthèse ; Xavier Rodet parle dans ce même volume de ce domaine qu'il a fait grandement progresser.

J'ai moi-même participé avec Max Mathews et John Chowning aux premières explorations du son musical numérique. La synthèse est exigeante : il faut tout spécifier à l'ordinateur, sans négliger aucun élément contribuant à la vie et à l'identité du son. Et les sons acoustiques « natu-

rels » sont plus complexes que l'on ne pensait : ainsi les descriptions des caractéristiques des instruments de musique qu'on pouvait trouver dans les manuels d'acoustique sont insuffisantes à produire un simulacre convaincant de ces instruments.

Il a donc fallu constituer un savoir-faire sonore pour tirer efficacement parti des ressources potentielles de la synthèse. Les données fournies à l'ordinateur spécifient précisément toutes les caractéristiques du son voulu : elles permettent donc de répéter sa synthèse à l'identique, même quarante ans plus tard. En écoutant le résultat, on fait l'expérience de la relation « psychoacoustique » entre la structure physique du son et son effet sensible, ce qui donne des indications nouvelles sur les mécanismes auditifs et leurs particularités.

La méthode de l'*analyse par synthèse* permet de distinguer dans la description complexe ce qui est pertinent pour l'oreille et ce qui ne l'est pas. J'ai pu ainsi donner une caractérisation concise des sons cuivrés, qui résistaient à l'imitation : le spectre de fréquence s'enrichit lorsque l'intensité augmente, même lors de l'attaque qui est très brève ; l'oreille reconnaît cette caractéristique. On dénomme « effet de chœur » la qualité particulière d'un ensemble de chanteurs – ou d'instruments – à l'unisson. Les voix à l'unisson ne sont pas rigoureusement identiques. On peut simuler l'effet de chœur *by brute force*, en ajoutant de nombreuses voix de fréquences légèrement différentes ; mais il est possible de le simuler plus économiquement en imposant à une onde unique une modulation aléatoire appropriée d'amplitude et de fréquence, que l'oreille pourra interpréter comme un effet de chœur. Je l'ai fait dans mon œuvre de synthèse *Little Boy* (1968) pour donner l'impression d'un chœur lointain parcourant une spirale glissant sans fin vers le grave.

Illusions acoustiques, vérités de la perception

Une telle descente est paradoxale : la fréquence d'un son ne peut diminuer indéfiniment – par exemple de trois octaves par minute – sans que le son sorte au bout de quelques minutes du domaine audible. On peut parler d'une illusion acoustique : à l'écoute, on a l'impression que la fréquence diminue ; la hauteur sonore perçue diminue sans conteste. Mais c'est une illusion cognitive d'assimiler la hauteur – un percept subjectif issu d'une expérience d'écoute – à la fréquence, paramètre physique objectif.

La synthèse numérique permet de construire des sons très spéciaux, donnant lieu à des illusions acoustiques : mouvements illusoires des sources sonores (John Chowning), gamme chromatique perpétuelle (Roger Shepard), accélérations perpétuelles (Kenneth Knowlton). J'ai, quant à moi, produit diverses illusions de hauteur : un son ne cessant de descendre et aboutissant à une hauteur plus aiguë, un son qui paraît baisser lorsqu'on double ses fréquences (par exemple en doublant la vitesse de lecture du magnétophone). J'ai aussi produit des illusions analogues pour le rythme. La hauteur perçue ne se réduit pas à une mesure de fréquence, ni le rythme à un comptage chronométrique.

Les particularités de l'audition permettent de « comprimer » les enregistrements sonores par un facteur allant jusqu'à plusieurs dizaines en tenant compte des effets de masquage de certaines composantes de fréquences par d'autres : le codage mp3 permet de faire tenir sur un CD-Rom d'une capacité de 700 Mo le contenu de dizaines de CD, avec une perte de qualité le plus souvent minime. Il s'agit d'un codage non linéaire : il est malcommode d'effectuer des transformations simples sur des sons codés en mp3.

Purkinje, parlant des illusions d'optique, a écrit : « Les illusions, erreurs des sens, sont des vérités de la perception ». Les mécanismes de la perception auditive sont hautement idiosyncrasiques, mais ils n'ont rien d'arbitraire : leur complexité se comprend dans la perspective de l'analyse de scènes auditives, une notion introduite par Albert Bregman et mentionnée par Christine Petit. Comme je l'ai dit au début, l'évolution des espèces a optimisé l'audition pour tirer parti au mieux des données des sens pour extraire des informations utiles sur le monde extérieur. Ainsi peut-on distinguer un son fort et lointain d'un son doux et proche, même si dans les deux cas l'oreille reçoit le même nombre de décibels : on ne sait pas faire faire cela par un programme.

Vocoder et traitement de la voix

Le vocoder est un dispositif électronique qui analyse puis resynthétise la parole en séparant plus ou moins bien une excitation – la vibration des cordes vocales – et une réponse – les résonances du conduit vocal, sans cesse modifiées par la phonation. Les premiers vocoders construits, depuis celui d'Homer Dudley en 1939, étaient analogiques. Dans les années 1960, on pouvait acquérir, auprès de la firme anglaise EMS, qui fabriquait des synthétiseurs électroniques (comme le Synthi 100 ou

l'AKS), un vocoder conçu pour l'usage musical par Peter Zinovieff, son directeur. Avec cet appareil, on peut remplacer le son des cordes vocales par une excitation montant du grave à l'aigu : une voix de basse profonde se mue graduellement en voix féminine, puis la parole devient suraiguë et inintelligible, ce qui est normal si la fréquence de l'excitation est supérieure à la fréquence des formants caractérisant les voyelles. On peut aujourd'hui tirer parti de synthétiseurs virtuels, mis en œuvre sur des ordinateurs individuels, comme celui réalisé par Xavier Rodet à l'Ircam.

Dans les années 1980 à Marseille, dans le Laboratoire de mécanique et d'acoustique (LMA) du CNRS, Daniel Arfib, Richard Kronland-Martinet et Pierre Dutilleux ont réalisé des analyses-synthèses d'un autre type, utilisant la méthode des grains de Gabor (*circa* 1946) et la méthode des ondelettes, qui venait d'être introduite par Jean Morlet et Alex Grossmann. La resynthèse est parfaitement fidèle, mais on peut, après analyse, exercer des modifications singulières. Ainsi, Frédéric Boyer a analysé une voix en utilisant douze ondelettes par octave, puis il l'a reconstituée de façon défective, en n'utilisant que ceux des degrés chromatiques qui donnent lieu à un accord donné : il a pu pour ainsi dire imprimer dans les cordes vocales divers accords : accord parfait majeur, accord parfait mineur, septième de dominante, septième diminuée. En écartant les grains de Gabor et en les raccordant avec soin, Arfib a pu étirer une phrase d'un facteur 100 dans le temps sans modifier la hauteur ni perdre l'intelligibilité. J'en ai tiré parti dans mes œuvres *Attracteurs étranges* et *Invisible*.

Synthèse de la voix chantée

Il n'est pas facile de synthétiser une voix d'une qualité « naturelle » : le mode de production contraint le signal d'une façon très spécifique ; nous y sommes familiers depuis avant notre naissance, et nous débusquons immédiatement les anomalies. Cependant, dans les années 1970, Mike McNabb et John Chowning ont pu suggérer très simplement l'apparition d'une voix chantée. Partant d'un son périodique ayant le spectre d'une voyelle chantée, ils lui ont appliqué graduellement un vibrato – une modulation de fréquence – soigneusement dosé et légèrement irrégulier : ce léger tremblement provoque la mutation du timbre anonyme de départ en un son qui impose son identité comme celle d'une voix chantée, suggérant l'apparition invisible d'une chanteuse ou d'un chanteur « virtuels ». Chowning fait ainsi surgir des voix féminines ou

masculines dans son œuvre *Phone**(1981). J'ai reproduit cette mutation dans mon œuvre pour soprano et ordinateur *L'autre face*, sur un poème de Roger Kowalski : j'y mets en scène une rencontre entre la soprano Irène Jarsky et un double illusoire, une voix de synthèse qui n'est la voix de personne – *présence-absence*.

Johan Sundberg s'est attaché à mieux comprendre le chant. Dans les années 1970, il a construit à cet effet un synthétiseur spécial, *Mussems*, et il a eu recours à l'analyse par synthèse pour débusquer nombre des caractéristiques de la voix chantée. Il a pu alors produire une imitation extrêmement convaincante des vocalises d'un chanteur. Jon Appleton a utilisé *Mussems* pour son œuvre de synthèse *Mussems Sangs* (1976). En 1979, à la faveur de résidences à l'IRCAM, Sunberg a inspiré plusieurs travaux sur la voix chantée. Xavier Rodet, Gerald Bennett et Jean-Baptiste Barrière ont développé le programme *Chant*, bien adapté à la production de sons tenu, et amplifié en un logiciel compositionnel, *Chant-Formes*. Bennett l'utilise dans plusieurs œuvres, comme *Un madrigal gentile* et *Rainstick* – dans cette dernière, les sons ténus du bâton de pluie s'agrègent soudainement en une voix insolite. John Chowning, quant à lui, a enrichi sa méthode de synthèse FM (*frequency modulation*) pour des imitations très réalistes de voix chantée. Il a extrapolé au-delà des possibilités vocaliques humaines : vers le grave, il a synthétisé une mélodie *basso profundissimo* que seul un géant pourrait chanter. Il a aussi réalisé des interpolations entre des timbres différents, métamorphosant par exemple une cloche en quatuor vocal.

Au cours de ses recherches sur l'imitation de la voix chantée, Chowning a élucidé une énigme : comment l'oreille fait-elle pour distinguer deux sons simultanés à l'unisson, ou deux sons tels que le fondamental de l'un ait la même fréquence qu'un harmonique de l'autre, alors que leurs composantes coïncident ? Superposant deux sons périodiques aux spectres de voix chantée dont les fréquences sont telles que toutes leurs composantes coïncident, il produit un seul son, une entité sonore unique : mais cette entité se scinde en deux « voix » lorsqu'il impose à ces deux sons des « micromodulations » différentes, notamment des vibratos qui leur confèrent une qualité quasi vocale. L'oreille dépiste alors une incohérence vibratoire entre deux ensembles de composantes dont chacun vibre de façon cohérente, et elle les traite comme deux entités sonores séparées, issues de deux sources virtuelles différentes. L'analyse par synthèse de Chowning a révélé ici un mécanisme très important de la cognition auditive.

Chowning a illustré de façon convaincante cette compréhension nouvelle de l'analyse auditive du simultané dans son œuvre de synthèse *Phone* : un son de cloche se transforme en un accord de sopranos, puis

des basses profondes surgissent d'un magma indémêlable. Il y a là une possibilité neuve et prometteuse de faire émerger à volonté certaines figures d'un fond indifférencié.

Je me suis borné ci-dessus à citer quelques cas où la musique d'aujourd'hui tire parti de la voix et de la parole de façon innovante.

Apprendre à parler/apprendre la musique

Les enfants apprennent à parler : peu apprennent la musique.

Parole et musique n'impliquent pas le même fonctionnement cérébral (*cf.* dans ce même ouvrage le texte d'Isabelle Peretz). Lorsqu'on parle ou qu'on écoute la parole, l'activité principale se situe dans l'hémisphère gauche, correspondant à la main droite et à l'oreille droite et appelé hémisphère dominant. Les choses sont plus compliquées pour la musique. Les aspects rythmiques font intervenir l'hémisphère gauche, alors que les accords sont plutôt perçus par l'hémisphère droit. Les expériences de Bever et Chiarello ont montré que la perception des mélodies musicales migre de l'hémisphère droit à l'hémisphère gauche lors de l'apprentissage du solfège : la référence aux degrés d'une gamme rapproche la musique d'un langage.

Récemment, Gottfried Schlaug et ses collègues de Harvard ont apporté des données anatomiques significatives concernant le cerveau et la musique. Ils ont montré, en particulier, que la pratique musicale chez l'enfant jeune augmente le volume de la partie antérieure du corps calleux, lequel est responsable de la communication entre les deux hémisphères cérébraux. Les mêmes auteurs remarquent que le corps calleux est proportionnellement plus important chez la femme que chez l'homme. Outre la coordination motrice entre les deux mains, qui impose la communication entre les deux hémisphères, on peut à bon droit spéculer que la musique favorise l'association de la précision et de l'affectivité, et pour ainsi dire de l'esprit de géométrie et de l'esprit de finesse.

On le voit, la pratique de la musique engage profondément les activités cérébrales aussi bien que motrices, et sa pratique peut être bénéfique, au-delà des problèmes pathologiques[2]. La musique est formatrice. C'est une activité complète, qui fait concourir perception, motricité, intelligence et sensibilité. Juste après la seconde guerre mondiale, une

2. *Cf.* Thaut (2005) ; Macintosh et *al.* (2006).

expérience pédagogique à grande échelle a été menée en Hongrie sous la direction du compositeur Zoltan Kodaly pour évaluer les bienfaits pédagogiques de la pratique musicale. Les élèves des lycées ont été répartis suivant deux groupes : l'un de ces groupes suivait l'enseignement habituel, l'autre un programme comportant un enseignement de musique actif – théorique et pratique – et très renforcé. Les résultats de ce dernier groupe ont été bien meilleurs en musique – rien d'étonnant – mais aussi pour les activités verbales, les mathématiques, les matières littéraires et même l'éducation physique. Plus récemment, les recherches approfondies de Glen Schellenberg à l'Université de Toronto (2005) montrent qu'une pratique musicale suivie exerce sur les performances cognitives des enfants des effets bénéfiques qu'on ne peut attribuer au milieu social ou à l'éducation des parents.

Outre le plaisir qu'elle procure, la musique peut stimuler les tâches cognitives, et elle donne accès à une autre forme de connaissance. La parole comme la musique, dans leurs formes les plus hautes, transcendent le quotidien, nous faisant sortir de nous-mêmes et peut-être entrevoir ce qui nous dépasse.

RÉFÉRENCES BIBLIOGRAPHIQUES

– T. G. Bever et R. J. Chiarello (1974), « Cerebral dominance in musicians and non-musicians », *Science*, 185, p. 537-539.

– B. Bossis (2005), *La Voix et la Machine*, Presses universitaires de Rennes.

– A. Bregman (1990), *Auditory Scene Analysis : The perceptual organization of sound*, Cambridge (Mass.), MIT Press.

– R. Carré, B. Lindblom et P. MacNeilage (1994), « Acoustic contrast and the origin of the human vowel space », *J. Acoust. Soc. Amer.*, 95, S2924.

– R.-G. Busnel et A. Classe (1976), *Whistled Languages*, Springer Verlag.

– J.-L. Dessalles (2000), *Aux origines du langage*, Paris, Hermès Sciences.

– B. Lechevalier, H. Platel, F. Eustache (2007), *Le Cerveau musicien*, Bruxelles, De Boeck.

– D. J. Lee, Y. Chen et G. Schlaug (2003), « Corpus callosum : Musician and gender effects », *NeuroReport*, 14, p. 205-209.

– F. Lerdahl et R. Jackendoff (1983), *A Generative Theory of Tonal Music*, Cambridge (Mass.), MIT Press.

– G. C. Mcintosh, D. A. Peterson et M. H. Thaut (2006), « Verbal learning with a musical template increases neuronal synchronization and improves verbal memory in patients with multiple sclerosis », *Neurorehabilitation et Neural Repair*, 20, p. 129.

– *Music perception*, 25, n° 4 (avril 2008), numéro spécial : *Music and Neurological Disorders*.

– J.-J. Nattiez (2004), *Musiques. Une encyclopédie pour le XXIe siècle*, t. 2, *Les Savoirs musicaux*. (*Cf.* J.-C. Risset, « Le timbre », p. 135-161 ; I. Peretz, « Le cerveau musical », p. 294-320).
– A. Pascal-Leone (2001), « The brain that plays music and is changed by it », *Annales of the New York Academy of Sciences*, 930, p. 315-239.
– *Portraits polychromes*, Paris, INA : n° 2, Jean-Claude Risset (2001-2008) ; n° 7, John Chowning (2005) ; n° 11, Max Mathews (2007).
– J. Sundberg (1987), *The Science of the Singing Voice*, DeKalb (Ill.), Northern Illinois University Press.
– E. G. Schellenberg (2005), « Music and cognitive Abilities », *Current Directions in Psychological Science*, 14, n° 6, p. 317-320.
– G. Schlaug, L. Jäncke, Y. Huang et H. Steinmetz (1995), « In vivo evidence of structural brain asymmetry in musicians », *Science*, 267, p. 699-671.
– G. Schlaug, L. Jäncke, Y. Huang, J. F. Staiger et H. Steinmetz (1995), « Increased corpus callosum size in musicians », *Neuropsychologia*, 33, p. 1047-1055.
– G. Schlaug, S. Marchina et A. Norton (2008), « From singing to speaking : why singing may lead to recovery of expressive language function in patients with Broca's aphasia (Melodic Intonation Therapy – MIT) », *Music Perception*, 25, p. 315-324
– H. Steinmetz, J. F. Staiger, G. Schlaug, Y. Huang Y, et L. Jäncke (1995), « Corpus callosum and brain volume in women and men », *NeuroReport*, 6, p. 1002-1004.
– J. Sundberg, L. Nord et R. Carlson (1991), *Music, Language, Speech and Brain*, Wenner-Gren International Symposium Series, vol. 59, McMillan Press. (*Cf.* J.-C. Risset, « Speech and music combined : an overview », p. 368-379 ; J. Sundberg, « Summary », p. 441-446.)
– M. H. Thaut (2005), « The future of music in therapy », *Annals of the New York Academy of Sciences*, 1060, p. 303-308.

Parole-chant : l'opéra

par Claude Hagège

Il n'est pas exclu que la musique et la langue aient la même origine, comme le pensait Jean-Jacques Rousseau. Il reste que les systèmes linguistique et musical sont très dissemblables. On ne retrouve pas en musique, comme on l'observe dans la langue, d'opposition entre lexèmes, ou mots pleins relevant du lexique, et morphèmes, ou mots-outils relevant de la grammaire. On ne peut pas, non plus, soutenir qu'il y ait une équivalence rigoureuse entre le spectre acoustique d'une note de musique et les formants d'un son de langue. Et surtout, cet aspect de la « grammaire » musicale qu'est le contrepoint, c'est-à-dire la superposition des lignes mélodiques, s'oppose à la syntaxe des langues d'une manière radicale.

En effet, la musique est capable non seulement d'inscrire les notes dans le flux du temps qui s'écoule, mais aussi d'associer les lignes mélodiques *dans le même moment*, en contrepoint, à savoir selon l'axe des simultanéités, alors que les langues humaines, dans leur totalité, déploient leur discours uniquement selon l'axe des successivités. Certes, nous entendons souvent, dans les langues, des tons, c'est-à-dire, par exemple, des mélodies montantes, descendantes, à l'unisson sur divers registres, ou mixtes, se superposer aux voyelles. Cela peut être illustré par les quatre tons du chinois mandarin : à l'unisson haut, montant, descendant-montant et descendant, ce qui donne, respectivement, *mā* « maman », *má* « chanvre », *mǎ* « cheval » et *mà* « insulter ». Mais ce phénomène de superposition ne s'observe qu'en phonétique, et de plus il ne concerne que les langues dites à tons, certes très nombreuses, mais qui ne couvrent qu'une partie du monde, notamment l'Asie du Sud-Est (chinois, thaï, vietnamien, miao-yao, birman, etc.), de nombreuses îles mélanésiennes, dont la Nouvelle-Calédonie, la plus grande partie de l'Afrique, où sont à tons quasiment toutes les langues, dont celles de la famille bantoue, enfin le nord, le cen-

tre et le sud du continent américain, où sont tonales de nombreuses langues amérindiennes.

Dans le domaine de la syntaxe, en revanche, bien qu'il y ait une différence entre :

— les langues à ordre des mots S(ujet) + V(erbe) + C(omplément), par exemple les langues indo-européennes d'Europe,

— les langues où la séquence est SCV (ex. hongrois, turc, mongol, amharique, japonais, coréen, etc.),

— celles où elle est VSC (arabe et hébreu classiques, etc.),

— celles qui présentent VCS (malgache, etc.),

— d'autres types de séquences encore,

le discours est toujours et invariablement soumis à la succession irréversible des mots, qui n'est autre que celle-là même du temps. *Il ne peut pas y avoir de contrepoint linguistique comme il y a un contrepoint musical.* Ou du moins, il ne peut y en avoir que lorsque beaucoup de personnes parlent en même temps. Mais alors sommes-nous en langue, sommes-nous en musique ? Nous ne sommes nulle part, il ne s'agit pas, dans ce cas, de communication, mais seulement de bruit, équivalant à un assourdissant silence.

Si, à présent, nous examinons la morphologie, c'est-à-dire la structure des unités minimales, en l'occurrence les mots pour ce qui est de la langue et les sons pour ce qui est de la musique, la différence est tout aussi radicale. Un mot d'une langue est une unité, soit indécomposable, soit analysable en sous-unités, comme le sont les mots dérivés et les mots composés. Ces unités ont deux faces, une sémantique et une phonique, dont chacune se déconstruit en constituants plus petits. Car un mot peut s'analyser en micro-unités de sens : la notion de « cheval », par exemple, est analysable en « équidé » + « mâle ». Un mot s'analyse également, quant à son autre face, la face sonore, en micro-unités de sonorité, soit, pour cheval, [š] + [ə] + [v] + [a] +[l]. Ces micro-unités sont elles-mêmes, à leur tour, des faisceaux de traits, dits distinctifs, soit, pour [š], chuintant + prépalatal + sourd.

Qu'est-ce, maintenant, qu'un son d'une partition musicale ? C'est un phénomène définissable par cinq propriétés : la durée, le rythme, la hauteur, l'intensité et le timbre. Un son musical, en effet, dure pendant une certaine portion de temps, correspondant, dans la notation musicale de l'Occident, à une ronde pour la durée la plus longue et à une quadruple croche pour la plus courte. Mais les sons sont également agencés entre eux selon des rythmes, qui peuvent être binaires, ternaires, quaternaires, etc. D'autre part, sur une échelle de vibrations du plus aigu au plus grave, un son occupe une position qui définit sa hauteur. En quatrième lieu, un

son peut être produit avec des degrés variables de force d'émission et de réception, auxquels correspondent, par exemple, les pédales d'étouffement et de renforcement d'un piano. Enfin, un son produit un effet différent selon la source qui le produit, par exemple la voix humaine, ou un tuyau d'orgue, ou la corde frottée d'un alto, ou la peau tendue et percutée d'un tambour.

Il convient d'ajouter que les mots possèdent, en commun avec les sons, une importante dimension acoustique, qui peut expliquer certaines évolutions. Pour prendre un exemple, en norvégien du Sud-Est, la sifflante [s] passe à la chuintante [š] dans Oslo, articulé [ošl⇔o]. Pour comprendre ce phénomène, il faut savoir que dans un premier temps, le [s], qui est sourd, assourdit, par assimilation progressive, le [l] qui lui est contigu, et qui devient donc un [l⇔], c'est-à-dire un *l* sourd, avec frottement de l'air sur les deux tranches de la langue. Mais à une seconde étape, c'est un facteur acoustique qui est en jeu. En effet, [š] et [l⇔], bien que différents en termes articulatoires, ont des spectres très proches, possédant à peu près la même structure de formants dans les hautes et les basses fréquences. Un phénomène comparable de parenté acoustique explique le changement de [al] en [aw] en anglais écossais et de [il] et [al] en [iw] et [aw] en portugais brésilien, où l'on entend, pour les noms des deux pays lusophones, [braziw] et [purtugaw], ainsi qu'en anglais américain du centre et du sud-ouest des États-Unis, où l'on entend [miwk] et [fiwm] pour *milk* et *film*.

C'est cette importante dimension acoustique de certains changements linguistiques qui, en dépit des différences radicales entre constituants des sons et constituants des mots, rapproche la parole et la musique. Un autre point commun entre les deux domaines est à observer en examinant une des deux modalités de l'entendre linguistique, à savoir non pas l'entendre catégoriel, qui articule l'énoncé en unités constitutives, mais l'entendre holistique, qui perçoit comme un tout non analysé un énoncé de langue. L'entendre linguistique holistique est une saisie globale d'un flux sonore comme support d'un contenu sémantique, et dans cette mesure, *l'entendre linguistique holistique est une catégorie cognitive assez proche de l'entendre musical*.

Ayant ainsi défini rapidement les caractéristiques contrastées des sons musicaux et des mots des langues humaines, nous pouvons, à présent, nous interroger sur la façon dont la musique occidentale a essayé d'associer ces deux domaines techniques et cognitifs, dont l'invention caractérise au plus profond les sociétés humaines. La manière la plus instructive de poser cette question et de tenter d'y répondre est de consulter l'histoire des conceptions que l'on s'est faites du genre qui pratique depuis les origines cette association même : l'opéra. C'est Gesualdo qui,

un des premiers, dans la seconde moitié du XVI^e^ siècle, détache une ligne textuelle des contrepoints où elle était absorbée. Le mot va même conquérir un rôle primordial par rapport à l'harmonie, dans la première moitié du XVII^e^ siècle, chez Monteverdi, dont le dernier opéra, *Le Couronnement de Poppée* (1642), est aussi le couronnement de cette tradition.

Jean-Baptiste Lully assurera mieux l'union de la musique et de la langue. Il faut souligner, cela dit, que les livrets sont en vers, ces derniers ne requérant pas la même musicalité que la prose. Mais la langue musicale, loin de gêner la poésie, l'enrichit, au contraire, par les timbres de voix, et plus tard par ceux des instruments. Un genre introduit dès cette époque, le récitatif, constituera un intermédiaire heureux entre la poésie dite et la poésie mélodisée.

Au XVIII^e^ siècle, cependant, la tragédie lulliste, perdant de son éclat, commence d'être supplantée par de nouvelles tendances. Ainsi, en 1735, *Les Indes galantes* de Rameau essaient d'émanciper un peu la musique par rapport au livret. Mais il ne s'agit encore que de tentatives, et le ton est donné quand, en 1767, dans la célèbre préface d'*Alceste*, Gluck, alors au sommet de sa gloire viennoise, va jusqu'à écrire :

> Je cherchai à réduire la musique à sa véritable fonction, celle de seconder la poésie, pour fortifier l'expression des sentiments et l'intérêt des situations, sans interrompre l'action et la refroidir par des ornements superflus.

Sans aller jusqu'à conférer à l'écriture dramatique le poids considérable que lui donnait un poète alors illustre, Métastase, dont le talent régnait sur l'*opera seria*, une telle conception est à l'opposé de celle qu'allait illustrer, par sa collaboration avec Da Ponte pour *Don Giovanni* (1787) et avec Schikaneder pour *La Flûte enchantée* (1791), le génie de Mozart, pour qui la dramaturgie est inscrite dans les formes musicales mêmes qui sont définitoires d'un opéra. Le texte poétique, dès lors, doit servir la musique par cela même que chaque langue possède de musique inhérente. C'est ce qui explique qu'une version française de *Don Giovanni* (1787) ou italienne de *La Flûte enchantée* (1791) n'est guère concevable, tout comme, un siècle plus tard, *Boris Godounov* (1869) sera une recréation, et non une glose, du texte de Pouchkine par un Moussorgski s'appuyant sur tous les effets mélodiques du russe, et de même, plus tard encore, la langue tchèque (dans sa variante morave) prêtera sa mélodie à la *Katia Kabanová* (1921) de Janáček, qui, bien avant Bartók, avait recueilli dans les campagnes les chants de son terroir. Ainsi la musique intrinsèquement propre aux langues entre en symbiose avec les sons musicaux dont on a rappelé plus haut les propriétés.

Après Mozart, la conception mozartienne d'un mariage intime entre langue et musique sera également le credo de Wagner, dont l'ouvrage *Opéra et drame* (1851) insiste sur l'idée que l'opéra est un art total où la poésie, comme l'action et les gestes, doit servir la musique. De là la logique d'un fait bien connu : les partitions lyriques que sont les livrets wagnériens ne peuvent avoir d'autre auteur que Wagner lui-même. Bien que la conception verdienne de l'opéra ne soit pas très éloignée de celle de Wagner, Verdi croit davantage à la fusion entre la musique et la poésie, même si cette dernière n'est que prétexte aux grandes mélodies amoureuses ou héroïques dont foisonne son œuvre.

Tous les librettistes de Verdi, de Piave à Boïto, appliquent, dans leur participation aux grands opéras comme *Rigoletto* (1851), *La Force du destin* (1862), *Don Carlo* (1867), *Otello* (1887), les instructions du musicien, qui subordonnent à la musique la trame du livret, la construction dramatique, l'effet scénique, c'est-à-dire quasiment tout… sauf l'écriture des vers, encore que même sur ce point il intervienne parfois, soutenant que la langue poétique italienne est trop ornée, et trop éloignée de la prose, contrairement aux langues poétiques française et allemande. Quant au livret d'*Aïda* (1871), Verdi l'a lui-même, avec le concours d'un versificateur, tiré d'un texte de l'égyptologue Mariette.

Richard Strauss, assez proche des conceptions wagnérienne et verdienne, demandera au poète Hugo von Hofmannsthal, pour *Salomé* (1905), *Le Chevalier à la rose* (1911) ou *Ariane à Naxos* (1919), « un bon drame, riche d'action et de contrastes, avec peu de scènes de masse »[1].

La même conception se retrouve chez les musiciens de la fin du XIX^e^ siècle qui, regroupés dans la Société nationale de musique, voulurent promouvoir la musique française d'opéra, et montrer que la langue française, tout comme la langue allemande, pouvait chanter la musique la plus pure, comme celle de Saint-Saëns dans *Samson et Dalila* (1877) ou de Massenet dans *Esclarmonde* (1889)[2]. Le français avait déjà été illustré dans *Faust* (1859), par le génie de Gounod, et il le fut encore quand furent entendus en 1875, promis à un fabuleux destin mondial malgré l'inepte aveuglement de la critique d'alors, les grands airs de la non moins géniale *Carmen* de Bizet.

Quelque temps plus tard devaient, cependant, intervenir d'importants changements de la relation entre les architectures musicales et le texte littéraire : en 1902, le *Pelléas et Mélisande* de Debussy dilue le livret

1. Cité par Robert Pitrou, « Hofmannsthal et Strauss », *Revue musicale*, 1930, p. 329.
2. *Cf.* Pierre-Jean Rémy, *Dictionnaire amoureux de l'opéra*, Paris, Plon, 2004, p. 588-591.

du poète symboliste Maeterlinck dans des innovations modales, des enchaînements d'accords, des répartitions de timbres vocaux et orchestraux qui le désarticulent complètement. Dix ans plus tard, le *Pierrot lunaire* de Schönberg (1912) accentue encore la mutation, et le XX^e^ siècle se poursuit sur des ouvrages comme le *Saint François d'Assise* (1984), où Messiaen se libère presque complètement du texte des *Fioretti*. Allant plus loin encore, certains musiciens récuseront la possibilité même d'opéras construits d'après des textes poétiques. Boulez écrit ainsi qu'il ne croit pas que l'écriture poétique de Claudel « puisse se prêter facilement à une augmentation musicale. Elle me semble refuser une telle opération. Il y a chez Claudel une respiration de la phrase qui, en soi, est déjà un phénomène musical. On ne pourrait donc créer que des redondances[3] ».

Ainsi reparaît l'ancienne idée, récurrente depuis le début de cette passionnante histoire des relations entre musique et paroles, que la langue est elle-même porteuse d'une musique qui rend problématique sa mise en musique. Une seule possibilité demeure, au moins pour les musiciens d'aujourd'hui : tenter de connecter, par un strict contrepoint poétique de l'écriture musicale, les syntagmes textuels et les syntagmes musicaux, comme s'y sont efforcés le compositeur Pascal Dusapin et le poète Olivier Cadiot dans *Roméo et Juliette ou la Révolution en chantant* (1989), ou François-Bernard Mâche dans *Kubatum* (1991) sur des chants d'amour en sumérien.

À cette étape de l'histoire du dialogue entre chant musical et texte littéraire, nous sommes loin de l'harmonie lullienne entre le livret et les notes. Les musiciens contemporains vont au-delà même de l'idée mozartienne, wagnérienne, verdienne ou straussienne, du primat de la musique sur la langue. Cela ne signifie pas que l'association des deux soit une permanente et indépassable aporie. Cela signifie seulement qu'elle est l'enjeu d'un défi radical lancé à l'inventivité humaine. Tout ce qui précède essaie de montrer comment divers artistes, à diverses époques, ont tenté de relever ce défi.

3. « Paul Claudel, intolérant et révolté », *in* Alain Galiari, *Six musiciens en quête d'auteur*, Paris, Pro Musica, 1991, p. 13.

Paroles, paroles

par PETER SZENDY

Pour Lia Arrigo

Paroles, paroles.

Des mots, des mots, seulement des mots.

Des mots en l'air – rrrien que des mots, faudrait-il dire en roulant le *r* comme Dalida.

Car, vous l'aurez reconnu, ce titre est celui d'un tube, d'un immense succès qu'elle a chanté avec Alain Delon, qui lui donnait la réplique en parlant, en déclamant.

Mais, avant que Dalida et Delon n'en proposent une version française en 1973, c'est Mina, la grande voix italienne de la variété et du jazz, qui avait, en 1972, chanté *Parole, parole,* accompagnée par l'acteur Alberto Lupo dans le rôle parlé.

Paroles, paroles, paroles.

Des mots, rien que des mots.

Cette petite phrase banale, qui est devenue l'un des refrains les plus chantés au monde ; cette triple répétition triviale du mot « mot », qui donne son titre à un tube que l'on ne prend même plus la peine d'écouter vraiment tant on l'a entendu et réentendu ; cette formule célèbre qui déclenche aussitôt le souvenir de sa mélodie – *Paroles, paroles…* –, cette formule, je voudrais la faire entrer, timidement, là où elle n'a jamais eu droit de cité : dans le temple de la musicologie, là où l'on débat des grands genres et des grands enjeux de la grande musique, la sérieuse, la classique.

Parole, parole, parole : ce refrain, dans la langue de Mina, en italien, résonne et consonne avec une autre formule, qui fut le titre non pas d'une chanson populaire, mais d'un divertissement théâtral : « D'abord la musique, ensuite les mots » (*Prima la musica e poi le parole*), tel était en

effet l'intitulé du petit spectacle que l'empereur Joseph II avait commandé à Salieri, en 1786, pour être joué en même temps que *Der Schauspieldirektor* (« Le directeur de théâtre ») de Mozart. On voyait dans le spectacle de Salieri un musicien (un maître de chapelle, *maestro di cappella)* et un poète *(un poeta)* se disputer au cours des préparatifs pour un opéra.

Je ne résiste pas au plaisir de citer un passage de leur dialogue, au début de la pièce :

> LE POÈTE. – Vous croyez donc que paroles et musique se puissent en quatre jours…
> LE MAÎTRE DE CHAPELLE. – Quant à la musique ne vous en donnez pas la peine, elle est déjà prête ; et vous devez seulement y adapter les paroles.
> LE POÈTE. – Mais c'est la même chose que de faire le vêtement et puis de faire ensuite l'homme auquel il s'adapte.
> LE MAÎTRE DE CHAPELLE. – Vous, messieurs les poètes, vous êtes fous. Mon ami, soyez-en convaincu ; qui donc, croyez-vous, voudra prêter attention à vos paroles ? C'est la musique, aujourd'hui, c'est la musique qu'il nous faut.
> LE POÈTE. – Mais cette musique, il faut pourtant qu'elle exprime le sentiment, bien ou mal.
> LE MAÎTRE DE CHAPELLE. – Ma musique a ceci d'excellent qu'à tout elle peut s'adapter remarquablement[1].

Au fond, entre *Parole, parole* et *Prima la musica e poi le parole,* la distance n'est pas si grande : dans les deux énoncés, il s'agirait de marquer que les paroles n'ont aucune importance, que c'est le chant seul, que c'est la musique seule qui compte. Ou du moins que c'est elle qui prime, qui vient avant (*prima*).

Que ce soit dans les querelles et les considérations esthétiques sur l'opéra ou dans les hypothèses sur l'origine des langues, ces deux formules, si proches par-delà les siècles, pourraient être la matrice de tant de discours, tant de fois repris : d'abord la musique – ou le son, ou le cri – et seulement ensuite les mots.

---

1. « *POETA. - Dunque credete che parole e musica / si possa in quattro dì… MAESTRO. - Circa a la musica / non ve ne date pena, ella è già pronta ; / e voi sol vi dovete / le parole adattar. POETA. - Questo è l'istesso / che far l'abito, e poi / far l'uomo a cui s'adatti. MAESTRO. - Voi, signori poeti, siete matti. / Amico, persuadetevi ; chi mai / credete che dar voglia attenzione / alle vostre parole ? / Musica in oggi, musica ci vuole. POETA. - Ma pure questa musica conviene / ch'esprima il sentimento, o male, o bene. MAESTRO. - La mia musica ha questo d'eccellente, / che può adattarsi a tutto egregiamente.* »

Mais, pour être complète, la grammaire générative des relations possibles entre musique et langage devrait aussi inclure la proposition inverse. Qui, historiquement, peut être mise dans la bouche de Monteverdi, à l'époque de la naissance du genre opératique. Lorsque le chanoine de Bologne, Giovanni Maria Artusi, attaché aux valeurs anciennes de la polyphonie, déclara qu'il n'entendait dans les œuvres vocales de son contemporain qu'« un mélange de voix, une rumeur d'harmonies insupportables aux sens », Monteverdi fit rédiger par son frère une *Déclaration,* imprimée en tête de son *Cinquième Livre* de madrigaux en 1605, dans laquelle il exposait son intention en ces termes : « faire en sorte que le discours soit maître de l'harmonie et non serviteur (*far che l'oratione sia padrona del armonia e non serva*) ».

Nous voici donc face à trois phrases ou quasi-phrases : 1. *Parole, parole* ; 2. *Prima la musica e poi le parole* ; 3. *Che l'oratione sia padrona del armonia e non serva*.

La première dit apparemment la même chose que la deuxième, et toutes deux semblent dire le contraire de la troisième.

Apparemment.

Mais allons y voir de plus près, car il se pourrait bien que la formule de Mina, longtemps après celles de Monteverdi et de Salieri, déplace radicalement les prémisses mêmes de la question telle qu'elle est traditionnellement posée en termes de *primauté* de la musique ou de la langue.

Je dois ici rappeler brièvement l'analyse que j'ai proposée ailleurs[2] de cette merveilleuse chanson à moitié parlée qu'est *Parole, parole.*

Il suffit d'écouter, à la lettre, si j'ose dire. Il suffit d'écouter *au mot* ce qui se dit et ce qui se musique. Car, tandis que Mina *chante* des mots (« des mots, seulement des mots »), l'acteur Alberto Lupo lui donne la réplique en *parlant.* Ils se partagent donc les rôles : certes, elle est, comme dirait Boris Vian, la môme, tandis qu'il est le gars, ces deux éternels protagonistes des chansons d'amour ; mais surtout, elle sera le Chant et lui sera le Mot. Si bien que, à prêter l'oreille à ce qui se joue dans *Parole, parole,* on voit apparaître un petit théâtre allégorique à deux voix, à l'image de celui de Salieri : un divertissement théâtral mettant en scène le dialogue entre le Parlé et le Chanté personnifiés.

Alberto Lupo joue en effet le rôle du Langage. Il interprète le Langage lui-même, il donne voix aux éternelles plaintes, aux sempiternels problèmes du Langage : « *Non vorrei parlare,* dit-il, je ne voudrais pas parler », je voudrais pouvoir arrêter le défilé de mes signifiants, soit dans le

2. Peter Szendy, *Tubes. La philosophie dans le juke-box*, Paris, Éditions de Minuit, 2008.

silence, soit dans le pur lyrisme du chant ou du cri, soit dans la référence réussie d'une signification pleine. C'est ce qui semble d'ailleurs se produire un instant quand, par la voix d'Alberto Lupo, le Langage lui-même, dans l'un de ses tours ou tropes les plus usés, paraît se rassembler dans sa force de dénotation, dans son pouvoir de convoquer les choses qu'il désigne : « *Tu sei come il vento che porta i violini e le rose,* dit le Langage, tu es comme le vent qui apporte les violons et les roses » – et aussitôt on les entend qui surgissent comme par magie, lesdits violons, convoqués dans l'orchestre par la puissance performative d'une parole pleine.

Mais c'est là précisément ce que le Chant, incarné par la voix de Mina, refuse aussitôt. Car Mina, qui joue le rôle de la Mélodie, voire de la Musique, cherche à se défaire des mots pour s'élancer dans le refrain, dans le fredon absolu. « *Caramelle, non ne voglio più* », chante le Chant ; c'est-à-dire, dans la version française de Dalida : « Moi, les mots tendres enrobés de douceurs se posent sur ma bouche, mais jamais sur mon cœur. »

La scène – cette scène d'amour ou de ménage entre le Langage et le Chant –, la scène s'accélère maintenant, elle prend un tour plus dramatique avec la ritournelle du refrain : « *Una parola ancora,* un mot encore », un dernier mot, supplie le Langage (« une parole encore », énonce mot à mot Alain Delon), tandis que la Mélodie lui répond, comme un écho multiplié qui vide les mots de leur pouvoir : « *Parole, parole, parole,* des mots, seulement des mots, rien que des mots… »

Les mots ne valent rien, voudrait chanter le Chant. Ou plutôt : ils *devraient* ne rien valoir *pour que le chant puisse s'élever,* s'envoler comme chant. « Paroles, paroles, paroles…, encore des paroles que tu sèmes au vent », chante ainsi Dalida dans la grande envolée lyrique qui clôt le refrain. Et Alain Delon de lui répondre, en relançant le couplet, en revenant au point de départ : « Voilà mon destin, te parler, te parler comme la première fois. »

Mais quelque chose, ici, s'est perdu, dans la version française qui, un an après l'originale, transportait notre scène dans cette langue dont Rousseau déclarait sans ambages : « …il n'y a ni mesure ni mélodie dans la Musique française, parce que la langue n'en est pas susceptible[3]. »

Si l'on prête en revanche l'oreille à la version de Mina, ce qu'on entend, c'est tout autre chose : c'est, ni plus ni moins, une façon inouïe de redonner à penser la question classique et apparemment usée des rapports entre musique et langage. Une manière inédite de la relancer, de la

3. Jean-Jacques Rousseau, *Lettre sur la musique française,* dans *Œuvres,* t. V, Paris, Gallimard, « Bibliothèque de la Pléiade », 1995, p. 328.

réinventer. Car, en ce même point où le refrain s'achève pour que tout recommence, Mina n'évoque pas des paroles, des mots simplement semés au vent, dépensés pour rien, inutiles, comme si l'enjeu du chant était banalement de faire taire le langage en le renvoyant à l'ineffable. Non, ce qu'elle énonce, c'est littéralement ceci : « *parole, parole, soltanto parole, parole tra noi.* » C'est-à-dire : « des mots, seulement des mots, des mots entre nous ».

Entendons-la bien.

Entendons *mot à mot* ce que donne ici à penser le Chant en personne : *entre nous,* affirme-t-il, donc entre le Chant et le Parlé, entre la Musique et le Langage, il n'y a que ça : des mots, des mots. L'entre, la différence ou l'écart entre Musique et Langage, c'est encore le langage. Le langage comme renvoi, comme ce qui diffère sa propre plénitude, son silence ou son devenir-chant.

À rebours du cliché rousseauiste qui voudrait que la langue italienne soit plus simplement et pleinement « musicale » que la française, ce que dit ou chante Mina, en italien, c'est que la différence sur fond de laquelle s'enlèvent ou s'emportent Musique et Langage, c'est la langue.

Avec Mina, il ne s'agit donc plus tant de savoir ce qui viendrait avant, entre paroles et musique, entre musique et langage. La question n'est plus celle de la primauté de l'un ou de l'autre – et c'est toute la position ou la posture musicologique classique qui s'en trouve bouleversée.

On pourrait en effet le montrer aisément : la musicologie, qu'elle le veuille ou non, reste, comme telle, prise dans le problème de la précédence, entre musique et langage. Son nom même en témoigne : la *musicologie* est cette discipline qui devrait être à même de traduire le fait musical dans la langue, dans le *logos* d'un discours qui puisse en rendre raison ; et qui d'ailleurs, pour ce faire, s'est toujours largement inspiré de la linguistique ou de la sémiotique, c'est-à-dire de modèles langagiers *à partir desquels* la musicalité de la musique pourrait faire sens.

Or, ce que Mina propose, c'est d'entendre autre chose dans ce vieux nom : la *musicologie,* chante-t-elle en quelque sorte, n'est pas, ne peut pas être ce qui, après coup, ordonne la musique selon l'ordre de la langue ; non, c'est plutôt que, en se produisant, dans le mouvement même de la différence qui la produit comme telle, *la musique s'écarte de la langue dans la langue.* Ou encore, ce qui revient au même : c'est dans l'espacement de la langue, en tant qu'elle ne coïncide jamais avec elle-même, que s'ouvre et s'emporte à la fois l'origine sans origine de la musique.

Singulière *mélologie* que celle de Mina analysant son propre tube, se faisant la musicologue d'elle-même et de son chant. Singulier *mélologue,*

pourrait-on dire en reprenant ce mot de Berlioz[4], singulier mélologue que *Parole, parole,* cette chanson parlée-chantée, que l'on aura peut-être entendue vraiment pour la première fois, même si on l'avait écoutée tant de fois.

Après avoir laissé entrer Mina dans le temple de la musicologie, où elle n'aura pas manqué de faire trembler quelques assises séculaires ; après l'avoir écoutée renvoyer dos à dos les tenants de Salieri ou de Monteverdi ou de Rousseau, je voudrais pour finir, car ce n'est que justice, redonner la parole à ce dernier. Comme si Rousseau, oui, Rousseau lui-même, présentait Mina et Alberto Lupo. Comme s'il les introduisait sur un plateau télévisé – car c'est là, en tant que générique de fin pour l'émission *Teatro 10,* que tout aura commencé.

Comme si Rousseau, donc, dans le rôle de Monsieur Loyal, désignait ces deux-là – le Chant et le Langage –, comme s'il les montrait au public en disant, en guise de prélude à la scène de ménage qui va suivre (je le cite en scandant lentement ses mots) :

« Voilà comment le chant devint – par degrés – un art entièrement séparé de la parole – dont il tire son origine[5]… »

Et encore une fois, à se répéter dans un montage en boucle, comme un tube, aussi souvent et aussi lentement qu'on le devra :

« … de la parole – dont il tire son origine – voilà comment le chant devint – par degrés – un art entièrement séparé de la parole – dont il tire son origine… »

4. . Qui qualifiait ainsi son œuvre intitulée *Lélio,* à la fois chantée et parlée, entre mélodrame et cantate.
5. Jean-Jacques Rousseau, *Essai sur l'origine des langues,* dans *Œuvres,* t. V, *op. cit.,* p. 427.

L'émotion dans le langage musical

par Emmanuel Bigand

Pour les sciences cognitives, le pouvoir expressif de la musique relève d'un paradoxe. Les stimuli musicaux ne réfèrent à aucune réalité du monde extérieur et n'ont aucune implication biologique immédiate. Ils n'en demeurent pas moins expressifs pour l'auditeur. Sur quels processus psychologiques reposent les sentiments et les émotions qu'ils nous procurent ? Dans ce chapitre, nous considérerons quelques conditions cognitives qui contribuent à l'émergence de cette expression musicale, et nous préciserons en quoi ces conditions pourraient être communes à la musique et au langage. Dans chacune de ces formes de communication, le sens repose sur des organisations syntaxiques qui inscrivent l'auditeur dans une attitude active visant à intégrer les informations entendues pour mieux comprendre et anticiper les informations à venir. En musique, l'émotion repose sur le jeu qu'introduit le compositeur avec l'auditeur par l'intermédiaire de ces attentes.

L'émotion musicale est une notion paradoxale à plus d'un titre. Le concept même est l'objet d'affirmations contradictoires parmi les compositeurs. Certaines soulignent que l'expression et l'émotion constituent les objectifs ultimes visés par la musique, alors que d'autres considèrent comme Stravinsky[1] que « la musique dans son essence (est) impuissante à exprimer quoi que ce soit : un sentiment, une attitude, un état psychologique, un phénomène de la nature, etc. L'expression n'a jamais été la propriété immanente de la musique ». Stravinsky ne nie pas le pouvoir évocateur de la musique, mais il considère qu'il

1. Stravinsky (1930), p. 116.

résulte principalement de projections de l'esprit. Si maintenant on accepte l'idée que la musique réfère (ne serait que par projection) à autre chose que l'univers des sons qui la constituent, la question de sa signification révèle un nouveau paradoxe. La musique est en effet le seul langage qui est à la fois intelligible et intraduisible dans une autre langue.

Pour la psychologie, le point le plus surprenant réside dans le pouvoir évocateur de la musique. Comment un stimulus peut-il avoir un impact émotionnel aussi important, alors qu'il ne renvoie à aucune réalité du monde extérieur et n'a aucune implication adaptative immédiate ? Ce pouvoir s'observe sous de multiples formes, dès le plus jeune âge jusqu'à la fin de la vie. La musique régule les émotions du bébé humain et elle stabilise les humeurs des plus anciens. On l'emploie d'ailleurs à cette fin pour réduire les traitements chimiothérapeutiques des patients Alzheimer. Chez les adultes, elle peut avoir un effet dynamisant, très utile pour la réalisation d'exercices sportifs intenses et elle contribue à lutter contre la douleur, le stress et l'angoisse. De ce fait, la musique est (ou a été) utilisée pour stimuler l'organisme dans des contextes aussi différents que les salles d'aérobic ou les champs de bataille.

L'étude scientifique des émotions musicales est relativement récente, et l'on commence seulement à entrevoir l'impact profond des stimuli musicaux sur le cerveau humain. Sur quels processus reposent ces émotions ? Dans quelle mesure peuvent-ils avoir un lien, même diffus, avec ceux intervenant dans le langage ?

Quelques remarques préliminaires sont nécessaires avant d'aborder ces questions. La notion même d'émotion recouvre des significations diverses. Dans le présent texte, nous la considérerons comme synonyme de sentiment, d'affect ou encore d'expression, malgré les différences conceptuelles qui peuvent être faites entre ces termes. Il convient également de séparer deux grandes sources d'émotions. Certaines sont extramusicales. Elles proviennent de situations qui ont été associées à l'œuvre par le hasard des circonstances. L'émotion évoquée résulte alors d'une association conditionnée par l'expérience, qui aurait pu, logiquement, se produire avec n'importe quelle pièce de musique. Cette source d'émotion, qui n'est pas liée au langage propre de l'œuvre, est non négligeable. La musique accompagne naturellement les différentes circonstances de la vie. Elle fonctionne cognitivement comme un bloc-notes qui fixe en mémoire les expériences émotionnelles multiples qui jalonnent ces circonstances. En situation thérapeutique, elle contribue à réveiller la mémoire autobiographique des patients atteints de troubles mnésiques. L'écoute d'un morceau de musique ne produit donc pas nécessairement

une émotion musicale. Elle peut simplement réveiller les émotions liées à ces expériences vécues. Ceci rend délicate l'interprétation des travaux conduits avec des morceaux que les sujets connaissent très bien. Par exemple, lorsqu'on demande à des auditeurs de venir écouter dans un scanner les musiques qu'ils affectionnent le plus, on constate que ces morceaux ont un impact profond sur le cerveau émotionnel[2]. Cet effet n'est cependant pas nécessairement d'origine musicale. Il pourrait tout aussi bien provenir d'émotions produites par des circonstances extérieures qui ont été fortuitement associées à l'œuvre.

L'importance des émotions extramusicales ne doit pas masquer l'existence d'autres sources d'émotions. La musique a la capacité d'imposer des états psychologiques variés à un large groupe d'auditeurs. Le pouvoir de cohésion sociale qui est souvent attribué à la musique, et qui pourrait constituer l'une de ses principales fonctions adaptatives, réside entièrement sur la capacité des stimuli musicaux à mettre à l'unisson émotionnel une foule entière. Dans les cas extrêmes, la musique parvient à déclencher des scènes pouvant aller jusqu'à l'hystérie collective, qu'il s'agisse des scènes de transe ou des concerts des Beatles. Il est d'ailleurs instructif d'analyser combien les pouvoirs politiques et religieux ont cherché, tout au long de l'histoire et jusqu'à aujourd'hui, à contrôler les productions musicales afin de canaliser les émotions qu'elles pouvaient engendrer. Quelles caractéristiques sonores parviennent à déclencher des émotions aussi fortes ?

Deux possibilités sont généralement avancées. Le pouvoir expressif de la musique pourrait reposer sur des propriétés immanentes des sons musicaux. Certaines dynamiques sonores, portant tout à la fois sur des variations de contour de hauteur, d'intensité ou de la qualité spectrale, sembleraient être universellement associées à des émotions spécifiques. Ces liens entre émotion et son pourraient trouver leur origine dans les modifications acoustiques de la voix qui font suite aux changements d'état émotionnel du sujet. C'est en constatant que les qualités sonores de la voix changent en fonction de ces états que des liens profonds entre émotion et son se seraient forgés dans nos cerveaux[3]. Les similitudes entre les traits expressifs de la voix et ceux de la musique ont été amplement soulignées. Très récemment, l'observation de fortes correspondances entre les intervalles de hauteurs séparant les formants de la parole et les intervalles musicaux[4] a renforcé l'idée que la musique puiserait certains de ses

2. Blood & Zatorre (2001).
3. Darwin (1898).
4. Ross *et al.* (2007).

traits expressifs dans la voix, qu'il s'agisse d'effets liés aux contours intonatifs ou aux intervalles.

La contribution des caractéristiques acoustiques des sons sur l'émotion musicale reste mal comprise, mais il est indéniable que ces caractéristiques ont un pouvoir immédiat sur l'auditeur. Dans des études récentes, nous avons présenté des extraits musicaux dont la durée était progressivement réduite pour atteindre des valeurs aussi brèves que 50 et 25 millisecondes. On sait que l'auditeur peut encore reconnaître la voix humaine à des durées aussi courtes. Peut-il également identifier l'émotion induite par la musique ? Pour répondre à cette question, nous demandions aux auditeurs de grouper les extraits musicaux qui induisaient des émotions similaires[5]. Les groupements effectués étaient ensuite convertis en une matrice de cooccurrence qui pouvait être interprétée comme une matrice de proximité émotionnelle : les extraits le plus souvent groupés correspondent à des extraits qui induisent des émotions similaires. L'analyse multidimensionnelle des valeurs de cette matrice permet de visualiser dans un espace bi- ou tridimensionnel les dimensions psychologiques qui sous-tendent l'espace des émotions ainsi que la place spécifique des pièces dans cet espace. L'un des objectifs des études les plus récentes était d'analyser combien la structure de cet espace change lorsque la durée des extraits augmente de 25 millisecondes à 20 (ou plus) secondes. Il va de soi que la nature des émotions évolue avec cette durée. Le plus surprenant est toutefois de constater que certaines structures de cet espace sont présentes dès les premières 25 millisecondes, et que l'ensemble se stabilise dès la première seconde de musique. Ce résultat suggère que les propriétés psychoacoustiques des sons induisent instantanément des expressions consistantes pour les auditeurs. Ces propriétés activeraient une voie émotionnelle dont la rapidité pourrait être comparable à celle des émotions engendrées par des stimuli biologiquement pertinents.

La musique ne se réduit cependant pas à une suite de sons plus ou moins chatoyants qui activeraient, moment par moments, des voies émotionnelles rapides. Comme le note le musicologue Leonard Meyer[6], « une symphonie de Beethoven n'est pas une sorte de banana split musical ». La musique est une structure sonore composée qui met en œuvre des systèmes complexes de relations entre les signaux acoustiques. Pour de nombreux auteurs, ces systèmes de relations, plus encore que les propriétés immanentes des sons, sont porteurs d'expression et d'émotion. Il est

5. Bigand *et al.* (2005).
6. Meyer (1956), p. 6.

d'ailleurs possible de faire de la musique avec n'importe quel type de sources sonores, comme le démontrent les musiques dites traditionnelles, voire même avec les objets de l'environnement extérieur, comme l'illustre le cas de la musique concrète. L'important réside dans l'organisation temporelle des contrastes et des similitudes qui existent entre ces objets sonores. Une aléatorisation complète de ces sons anéantirait l'expérience émotionnelle, ou du moins la réduirait fortement. Quel pourrait être le pouvoir émotionnel d'une musique qui ne serait pas composée en fonction d'un ensemble de règles au sein d'un langage spécifique ? En l'absence de toute technique de composition, les œuvres seraient courtes, et les émotions induites très anecdotiques. De façon similaire, que resterait-il de notre langue maternelle si nous lui ôtions son organisation syntaxique ? Les émotions musicales ont la richesse que nous leur connaissons parce qu'elles reposent sur des langages qui ont été conçus et mûris au fil des siècles par des générations de compositeurs, de musicologues et de scientifiques, dans le but de rendre les œuvres aussi expressives que possibles. Comprendre les liens étroits qui unissent émotion et syntaxe en musique est un enjeu essentiel pour comparer la musique au langage. Le fait que la présence d'une syntaxe puisse être une condition nécessaire à la production de significations et d'émotions riches et variées a des implications théoriques importantes. La syntaxe serait-elle une condition *sine qua non*, à l'origine de tout dialogue humain ? Si oui, quelle fonction sociocognitive permet-elle d'assurer ? Avant de revenir sur ces questions, il importe d'analyser plus en détail comment l'émotion prend sa source dans le langage musical. Nous considérons pour ce faire le cas de la musique occidentale tonale, puis, plus brièvement, celui de la musique contemporaine.

Le musicologue Leonard Meyer[7] a émis plusieurs hypothèses sur les liens entre syntaxe et émotions musicales, hypothèses auxquelles les recherches récentes redonnent une nouvelle actualité. Selon l'auteur, l'émotion se produit lorsqu'une tendance à répondre est provisoirement arrêtée, bloquée ou détournée. Un stimulus qui engendre, à l'intérieur de l'organisme, une tendance à agir ou à penser est donc potentiellement porteur d'émotion. Selon Meyer, la syntaxe musicale confère à la musique les caractéristiques requises pour induire de telles tendances. Lorsque l'on est familiarisé avec un langage musical, tout contexte musical crée des attentes perceptives qui portent sur des événements sonores à venir, ou sur des structures plus abstraites de relations entre les sons[8]. Ces attentes

7. Meyer (1956).
8. Jackendoff (2000).

ne sont pas conscientes, sauf peut-être pour un auditeur expert analysant auditivement une œuvre ; mais on peut facilement les mettre en évidence expérimentalement[9]. L'auditeur est automatiquement projeté en avant dans un processus actif d'anticipation des événements. Des anticipations similaires s'observent avec le langage. Une part importante du dialogue humain repose sur cette faculté d'anticiper la suite du discours de l'autre, ce qui conduit même dans certains cas, à finir ses phrases. La communication suppose une correspondance minimale entre les attentes du locuteur et celles de celui qui l'écoute (ou le lit). Selon Meyer, ce processus actif est à la base de nos émotions musicales. Celles-ci prennent forme dans le jeu que le compositeur, l'improvisateur ou, dans une mesure moindre, l'interprète entretiennent avec ces attentes. Celles-ci sont généralement résolues dans le déroulement de l'œuvre à la suite de péripéties musicales multiples, chacune donnant naissance à d'autres attentes plus locales selon des procédés d'une variété infinie. Selon Meyer, un contexte musical qui n'engendrerait aucune attente perceptive ne pourrait pas susciter d'états émotionnels.

Les œuvres regorgent d'exemples illustrant cette notion. La figure 1 (*haut*) représente un passage du dernier mouvement de la quatrième symphonie de Tchaïkovski. De nombreux éléments musicaux sont présents au début de l'exemple (dont notamment une longue pédale sur l'accord de septième de dominante), qui vont transformer l'intuition naissante de l'imminence d'un événement en la certitude d'une explosion orchestrale. Au milieu de l'extrait, des modifications rythmiques et intensives indiquent sans équivoque que cette entrée fracassante de l'orchestre va se produire à un instant « t » bien précis, matérialisé sur la figure par la flèche. Nul doute que si vous écoutiez l'extrait schématiquement représenté par la forme d'onde de la figure 1 (*haut*), cette attente vous paraîtrait parfaitement résolue. Le fait est que cette représentation ne correspond pas à la pièce réelle. Celle-ci est représentée dans la figure 1 (*bas*). Le décalage de la flèche représente comment le compositeur joue avec l'attente qu'il a suscitée dans le début de l'extrait. L'explosion orchestrale se produit bien, mais avec un léger retard par rapport à ce qui aura été anticipé.

9. Pour des revues, voir Bigand (2005) et Bigand & Poulin-Charonnat (2006).

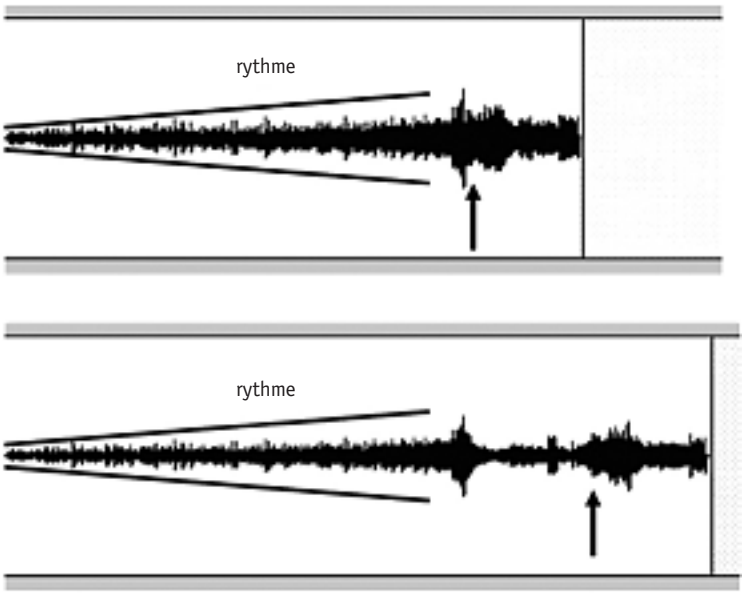


Figure 1 : *Illustration de la notion d'attente perceptive. Dans cet extrait du mouvement final de la quatrième symphonie de Tchaïkovski, l'entrée orchestrale anticipée pour un instant précis (en haut) est, en réalité, retardée par le compositeur (en bas)*[10].

Cet exemple, volontairement très simple, pourrait être répliqué avec un très grand nombre d'œuvres, qu'elles soient de style savant ou populaire. Il illustre combien cette notion d'attente est au cœur des intuitions musicales de l'auditeur. Du point de vue musicologique, cette notion correspond au concept de *directionalité*[11]. Sans directionalité, la musique n'instaurerait aucune tension ni détente, et ne susciterait aucune attente. Elle ne posséderait pas cette qualité dynamique qui donne l'impression que les sons progressent d'un point vers un autre dans le temps. De nombreux paramètres sonores contribuent à engendrer ces attentes. Dans l'exemple ci-dessus, le rythme, l'intensité, l'harmonie et la répétition des éléments thématiques sont des éléments cruciaux. Dans le cas de la musique occidentale tonale, l'harmonie joue un rôle primordial. Les règles de l'harmonie établissent un lien très fort entre syntaxe, attente perceptive et émotion.

Le système musical occidental tonal repose en effet sur une organisation très stricte des hauteurs. Cette organisation porte sur trois catégo-

10. Cet exemple est consultable sur le site web http://leadserv.u-bourgogne.fr/rubrique83.html/.
11. Kramer (1988) ; Lalitte (2004).

ries d'objets sonores : les notes, les accords, et les tonalités[12]. Les notes sont organisées en ensembles de sept sons (les tonalités), à l'intérieur desquels opèrent des hiérarchies de hauteurs. Certaines notes de la tonalité jouent le rôle de point d'ancrage cognitif. Ces notes attirent les autres, qui semblent graviter autour d'elles. Dans la tonalité de *do* majeur, les notes *do* (la tonique) puis *sol* et *mi* sont des notes hiérarchiquement importantes. L'occurrence d'autres notes instaure des tensions. Un auditeur familier avec ce langage attendra la résolution de ces tensions. Les hiérarchies changent d'une tonalité à l'autre. Ainsi, dans la tonalité de *la* majeur, la note *mi* sera encore hiérarchiquement importante mais la note *do* sera une note ornementale créant des tensions qui devront être résolues. La fonction syntaxique d'une note dépend donc de la tonalité dans laquelle elle apparaît. De cette façon, une mélodie donnée peut être entendue de façon très différente selon qu'elle est comprise dans telle tonalité ou dans telle autre[13]. Dans ce type de cas, nous pouvons montrer que les tensions et les détentes perçues par les auditeurs sont très différentes, même si la mélodie est composée des mêmes rythmes et des mêmes sons.

Les notes s'organisent en accords qui entretiennent à leur tour des relations hiérarchiques fortes. Les accords constitués sur les premier, quatrième et cinquième degrés de la gamme (respectivement appelés, accords de tonique, I, de sous-dominante, IV et de dominante, V) jouent dans le domaine harmonique le rôle de points d'ancrage pour la perception. Les accords construits sur les autres degrés doivent se résoudre sur ces degrés dits « forts ». Ne pas respecter ces hiérarchies crée des tensions qui sont très expressives. Tel est le cas, sans doute très simple, de la cadence rompue qui, au lieu d'aller vers l'accord de tonique (I), se termine sur un accord hiérarchiquement moins important (à savoir, l'accord construit sur le sixième degré de la gamme). Pour un morceau donné, il est possible de composer un nombre quasi infini de suites, chacune surprenant l'auditeur de façon expressive et spécifique. Dans la musique, tout comme dans le langage, il est en fait impossible d'entendre le début d'une pièce musicale sans projeter, ne serait-ce qu'implicitement, son attention vers les événements qui viendront résoudre les tensions créées.

Un troisième niveau d'attente est engendré par les relations qu'entretiennent les tonalités. Dans la musique occidentale tonale, il existe vingt-quatre tonalités majeures et mineures. La proximité psychologique de deux tonalités dépend principalement du nombre de change-

12. Krumhansl (1990) ; Lerdahl (2001).
13. Bigand (1993).

ments hiérarchiques qui se produisent lorsque l'on passe d'une tonalité à l'autre[14]. Les tonalités proches possèdent des notes et des accords communs et leurs points d'ancrage cognitifs sont similaires. Ainsi, les tonalités de *do* et de *sol* majeurs seront proches parce qu'elles comportent six notes identiques et que les notes *do* et *sol* sont hiérarchiquement importantes dans les deux cas. Les tonalités de *do* majeur et de *la* mineur seront proches pour les mêmes raisons Dans ces deux tonalités, les notes *do* et *mi* fonctionnent comme des points de référence pour la perception. Les tonalités de do majeur et de do mineur seront également proches bien qu'elles partagent moins de notes communes. L'important dans ce cas, est que les notes *do* et *sol* conservent les mêmes fonctions syntaxiques de tonique et de dominante dans ces tonalités. Ces deux points de référence ne bougeant pas, le passage d'une tonalité à l'autre n'implique pas de grandes modifications cognitives, bien que ces tonalités partagent peu de notes. Le seul point de référence qui change fortement porte sur le troisième degré de la gamme. La note *mi* est hiérarchiquement importante dans le premier cas, et elle devient ornementale dans le second. Une évolution inverse se produit pour la note *mi* bémol. Ces changements, qui s'ajoutent à quelques autres plus fins, rendent compte des effets expressifs des modes majeur et mineur, bien établis dans la littérature[15].

Par-delà l'effet expressif des modes, ces changements de tonalités ont des implications fortes sur la formation des attentes perceptives. Moduler d'une tonalité à une autre ne constitue pas un geste neutre. La modulation implique certaines continuations plutôt que d'autres. L'auditeur qui perçoit une modulation ne peut s'empêcher d'anticiper les tonalités à venir. La musique invite à un voyage dans l'espace formé par ces vingt-quatre tonalités et, à chaque étape, l'auditeur anticipe les suites possibles de ces déplacements[16]. Là encore les possibilités de jouer avec les attentes de l'auditeur sont considérables. On en trouvera un exemple simple dans le passage musical suivant (Figure 2), issu d'un poème symphonique de Liszt (*Tasso, lamento e trionfo*). Une première déclamation orchestrale s'achève sur une harmonie de septième diminuée, qui est longuement entretenue par l'orchestre. Cet accord est syntaxiquement très particulier puisqu'il peut conduire vers quatre tonalités majeures ou mineures. L'incertitude syntaxique est donc totale lorsque la musique est en suspension sur cet accord. Dans cet exemple, le compositeur utilise magnifiquement cet accord pour entretenir un suspense intense, qui sera

14. Lerdahl (2001).
15. Della Bella *et al.* (2001).
16. Lerdahl (2001).

renforcé par la répétition de la déclamation. Cette seconde déclamation, bien qu'un peu différente musicalement, aboutit encore à cette harmonie suspensive. L'auditeur comprend très bien que quelque chose va se produire, et, étant donné les couleurs orchestrales de l'extrait, il est facile d'anticiper que cet événement sera dramatique. L'attente est renforcée par les faux appels à la résolution qui interviennent après la seconde déclamation, mais qui conduisent tous vers des harmonies instables. Un geste déterminé émerge enfin des violoncelles et contrebasses, qui emporte l'ensemble de l'orchestre dans une envolée colérique fulgurante. L'expression de cette intervention orchestrale est assez forte pour être perceptible hors contexte, mais il va de soi que replacée dans cette mise en scène sonore, elle présente un caractère stupéfiant qui, à n'en point douter, provoquerait les fameux frissons dans le dos, si caractéristiques des émotions musicales.

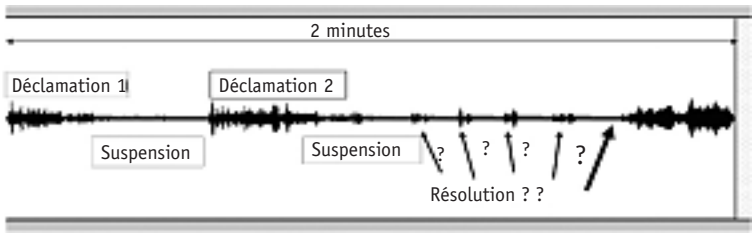


Figure 2 : *Dans ce poème symphonique de Liszt, la syntaxe musicale entretient pendant presque deux minutes un climat d'attente qui se résout, après plusieurs hésitations, par une entrée fulgurante de l'orchestre*[17].

Le système musical tonal définit donc des règles sur la dimension des hauteurs, règles que l'on peut qualifier de syntaxiques dans la mesure où elles conditionnent l'organisation temporelle des événements. De ce fait, le début d'un processus musical implique certains types de continuations plutôt que d'autres, exactement au même titre qu'une phrase. Un auditeur familiarisé avec un langage musical anticipe ces continuations. Dans le cas de la musique tonale, ces anticipations sont d'autant plus fortes que les hiérarchies de hauteurs se combinent avec les hiérarchies métriques définies sur la dimension du temps. Lerdahl et Jackendoff[18] ont formalisé comment la combinaison de ces deux hiérarchies contribuait à définir des relations de tensions et détentes musicales à plusieurs échelles de temps. Ces hiérarchies de tensions et détentes donnent lieu à des

17. Cet exemple est consultable sur le site web http://leadserv.u-bourgogne.fr/rubrique83.html/.
18. Lerdahl et Jackendoff (1983).

enchâssements d'attentes perceptives qui structurent le temps musical. L'auditeur anticipe donc, à plusieurs échelles de temps, le contenu des événements à venir et le moment le plus propice de leur apparition. Il est donc très facile pour un compositeur de jouer avec ces attentes perceptives et de créer une infinité d'effets de surprise ayant chacun des qualités expressives spécifiques.

Les études de neurosciences cognitives ont récemment mis en évidence l'existence de ces attentes, et ont démontré leurs liens avec les réponses émotionnelles. On peut sonder la nature des attentes de l'auditeur dans des études où l'on présente un contexte musical suivi d'un accord cible qui peut être plus ou moins plausible syntaxiquement. Par exemple, un accord de tonique est hautement plausible à la fin d'une séquence musicale, bien plus qu'un accord de sous-dominante, qu'un accord de sixième degré. L'auditeur n'a généralement pas conscience de ces attentes, mais elles peuvent être mises en évidence par des mesures très simples. Ainsi, dans une étude, nous présentions des séquences d'accords chantées avec des phonèmes sans signification. La fin de la séquence finissait toujours sur le phonème « di » ou le phonème « dou ». Les auditeurs devaient appuyer le plus vite possible sur une touche de l'ordinateur identifiant l'un ou l'autre de ces phonèmes. Il apparaît que cette discrimination phonétique s'effectue plus rapidement, et avec moins d'erreurs, lorsque le dernier phonème est chanté sur une harmonie (un accord) qui est syntaxiquement très reliée au contexte[19]. La discrimination phonétique étant objectivement la même, seuls des effets d'attente perceptive peuvent expliquer cette différence. Des résultats comparables s'observent en psycholinguistique. Par exemple, lorsque l'on présente des phrases qui finissent sur des mots ou des non-mots, il est plus facile de discriminer les mots des non-mots lorsque le mot cible est relié au contexte de la phrase. Ainsi, reconnaître que le mot « lapin » est un mot et le mot « laprin », un non-mot, sera plus facile et plus rapide dans la phrase « le chasseur tira sur le lapin (laprin) » que dans la phrase « le cycliste roula sur le lapin (laprin) ». Plusieurs études ont montré que des phénomènes comparables s'observent en musique. Nous avons ainsi observé que, dans le chant, le traitement d'un mot cible dépend de sa relation sémantique avec le début de la phrase (le mot est plus facilement et plus rapidement traité lorsqu'il est relié), mais que ces effets d'amorçage sont modulés par la syntaxe musicale. Les effets d'amorçage linguistiques sont d'autant plus forts que l'accord sur lequel le mot cible est chanté est attendu du point de vue de la syntaxe musicale[20].

19. Bigand *et al.* (2001).
20. Poulin-Charonnat *et al.* (2005).

Le cerveau répond à ces violations d'attentes syntaxiques de façon très semblable dans la musique et le langage. Ainsi, Patel *et al.*[21] ont montré que lorsqu'on augmente l'importance des fautes de syntaxe dans une phrase, on observe une composante positive dans les potentiels électriques évoqués, 600 millisecondes après l'occurrence du mot erroné. L'intensité de cette composante augmente avec l'intensité de la faute[22]. Une réponse électrophysiologique semblable est obtenue lorsque l'on joue un accord qui enfreint de plus en plus fortement la syntaxe musicale. De nombreuses études réalisées par Koelsch et ses collaborateurs[23] ont également mis en évidence une onde négative précoce dans les potentiels évoqués (ERAN pour *early right anterior negativity*), qui apparaît 300 millisecondes après une violation syntaxique en musique, et qui est très similaire à l'onde ELAN (*early left anterior negativity*) associée à l'occurrence d'un mot erroné dans une phrase. Les études d'imagerie cérébrale ont également permis de cartographier les zones cérébrales actives lorsque l'on présente aux auditeurs des événements musicaux qui respectent les règles de la syntaxe tonale et celles qui sont actives lorsque ces événements ne les respectent pas. Le recouvrement de ces zones avec celles impliquées dans le traitement du langage a été rapporté par plusieurs auteurs[24].

Quelles relations ces attentes liées à la structure du langage peuvent-elles entretenir avec les émotions musicales ? Dans une étude récente d'imagerie cérébrale, Tillmann et ses collaborateurs[25] ont présenté des pièces musicales qui pouvaient ou non contenir des fautes fines de syntaxe musicale. Il est apparu que ces stimuli activent un réseau assez vaste, qui comprend certaines zones du langage, mais aussi des zones telles que le cortex orbitofrontal, que l'on sait être activées par les musiques intensément plaisantes[26]. Ce résultat suggère un lien possible entre les violations d'attentes syntaxiques et les réponses émotionnelles. Ce lien a été approfondi par plusieurs études dans lesquels des erreurs d'harmonie étaient également introduites dans les stimuli musicaux. Les réponses émotionnelles de nature comportementale (par des jugements subjectifs) et physiologique étaient enregistrées. Il en allait de même pour les potentiels évoqués associés à ces erreurs de syntaxe. Koelsch et ses collaborateurs[27] ont ainsi montré que l'occurrence d'une cadence rompue ou

21. Patel *et al.* (1998).
22. *Cf.* aussi Regnault *et al.* (2001).
23. Koelsch *et al.* (2000) ; Poulin-Charronnat *et al.* (2006).
24. Koelsch *et al.* (2000) ; Mass *et al.* (2001) ; Koelsch *et al.* (2006).
25. Tillmann *et al.,* (2006).
26. Blood & Zatorre (2001).
27. Koelsch *et al.* (2008).

l'apparition d'une sixte napolitaine non résolue entraînaient une modification sensible des jugements émotionnels, un changement graduel de la conductance de la peau, et une augmentation graduelle de l'ERAN. Ces résultats ont été répliqués avec un matériel musical issu du répertoire, et non construit pour les besoins de l'expérience[28]. En réanalysant des données d'imagerie cérébrale, Koelsch et ses collaborateurs[29] ont observé que l'occurrence d'un accord syntaxiquement peu attendu (une sixte napolitaine non résolue) provoque une activité accrue de l'amygdale. L'ensemble de ces résultats suggère que les violations d'attente musicale sont bien associées à des réponses émotionnelles pour l'auditeur.

Le système musical occidental présente des caractéristiques structurelles avantageuses pour engendrer, par l'intermédiaire de ses hiérarchies de hauteurs et de métrique, des attentes perceptives. La syntaxe musicale tonale démultiplie les possibilités expressives des discours musicaux en facilitant le jeu avec ces attentes, ce que confirment *in situ* les études neuroscientifiques récentes. Il serait cependant erroné de penser que seul notre langage musical tonal offre une telle possibilité. Les hiérarchies de hauteurs constituent un atout pour définir des syntaxes musicales expressives, mais ce n'est certainement pas la seule possibilité. Le musicologue P. Lalitte a réalisé une étude étonnante de ce point de vue[30]. Il a présenté à des auditeurs, musiciens et non musiciens, deux sonates pour piano de Beethoven et une version atonale de ces sonates. Ces versions atonales représentent approximativement ce qu'une rencontre entre les compositeurs Ligeti et Beethoven aurait pu produire musicalement. Les versions atonales détruisaient toutes les hiérarchies de hauteurs liées à la tonalité, mais elles préservaient l'ensemble des autres paramètres sonores (rythme, variations intensives, etc). Il va de soi que le caractère musical de ces œuvres change considérablement, mais il est surprenant de constater que la disparition des hiérarchies de hauteur, loin de remodeler entièrement l'expérience émotionnelle des deux groupes d'auditeurs, n'entraîne pas de modification forte dans la dynamique temporelle des émotions ressenties. Ce résultat souligne que des liens forts entre syntaxe et émotion persistent dans les musiques qui ne recourent pas aux hiérarchies de hauteurs, telles que les musiques contemporaines. La syntaxe est définie dans ce cas par d'autres paramètres sonores, ce qui ne l'empêche pas d'induire des attentes perceptives[31] et des émotions, ainsi que le montre cette étude.

Les recherches à venir contribueront certainement à mieux comprendre les liens entre la syntaxe et les émotions dans les musiques occi-

28. *Ibid.*
29. *Ibid*.
30. Lalitte *et al.* (2009).
31. Lerdahl (2001).

dentales tonale et contemporaine, ainsi que pour d'autres musiques traditionnelles. S'il était confirmé que la syntaxe joue un rôle fondamental pour rendre les signaux acoustiques expressifs dans la musique et le langage, nous pourrions nous demander si ce type d'organisation ne répond pas à une contrainte sociocognitive fondamentale, à l'origine de tout dialogue humain. Communiquer avec autrui suppose en effet un recouvrement minimum entre les attentes des interlocuteurs. Il ne peut pas y avoir de communication en l'absence d'un système qui ajuste et régule ces attentes. Une syntaxe partagée faciliterait cette régulation. Cette fonction est essentielle pour la musique comme pour le langage, et il n'est de ce fait pas illégitime de penser que des ressources neuronales partiellement communes aient pu être investies pour la mettre en œuvre.

RÉFÉRENCES BIBLIOGRAPHIQUES

– Bigand E. (2005), « L'oreille musicale peut-elle se développer par l'écoute passive de la musique ? », *Revue de neuropsychologie*, 14, p. 191-221.

– Bigand E. (1997), « Perceiving musical stability : The effect of tonal structure, rhythm and musical expertise », *Journal of Experimental Psychology, Human Perception and Performance*, 23, p. 808-812.

– Bigand E., Poulin-Charronnat B. (2006), « Are we all "experienced listeners" ? », *Cognition*, 100, p. 100-130.

– Bigand E., Vieillard, S., Madurell F. et Marozeau J. (2005), « Multidimensional scaling of emotional responses to music : The effect of musical expertise and excerpts' duration », *Cognition & Emotion*, 8, p. 1113-1139.

– Bigand. E, Tillmann B., Poulin B., D'Adamo D. et Madurell F. (2001), « The effect of harmonic context on phoneme monitoring in vocal music », *Cognition*, p. 11-20.

– Blood A. J., Zatorre R. J. (2001), « Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward & emotion », *PNAS*, 98, p. 11818-11823.

– Dalla Bella S., Peretz I., Rousseau L. et Gosselin N. (2001), « A developmental study of the affective value of tempo and mode in music », *Cognition*, 80, B1-B10.

– Darwin C. (1898), *The Expression of the Emotions in Mans and Animals*, Appleton and Company.

– Jackendoff R. (1991), « Musical parsing and musical affect », *Music Perception*, 9, p. 199-230.

– Koelsch S., Gunter T., Friederici A. et Schröger E. (2000), « Brain indices of music processing : "Non-musicians are musical" », *Journal of Cognitive Neuroscience,* 12 (3), p. 520-541.

– Koelsch S., Gunter T., Cramon D. Y. von, Zysset S., Lohmann G. et Friederici A. (2000), « Bach speaks : A cortical language network serves the processing of music », *NeuroImage*, 17, p. 956-966.

– Koelsch S., Fritz T., Cramon D. Y. von, Müller K., Friederici A. D., (2006), « Investigating emotion with music : an fMRI study », *Human Brain Mapping*, 27 (3), p. 239-250.
– Koelsch S., Fritz T., Schlaug G., (2008), « Amygdala activity can be modulated by unexpected chord functions during music listening », *Neuroreport*.
– Koelsch S., Kildnes S., Steinbeis N., Schelinski S. (2008), « Effects of unexpected chords and of performer's expression on brain responses and electrodermal activity », *Plusone*, 7, p. 2631.
– Kramer Jonathan D. (1988), *The Time of Music*, New York-Londres, Schirmer Books.
– Krumhanls C. (1990), *The Cognitive Fundation of Musical Pitch*, New York, Oxford University Press.
– Lalitte P., « Le carré sémiotique des formes musicales : un modèle d'analyse des formes musicales de la deuxième moitié du XX^e siècle », *Analyse et contextualisation*, actes de la rencontre du 24 mai 2003 réunis et présentés par M. Battier et D. Pistone, série « Conférences et séminaires », n° 16, 2004, p. 51-62.
– Lalitte P., Bigand E., Kantor J. (2009), « On listening to atonal variants of two piano sonatas by Beethoven », *Music Perception.*
– Lerdahl F. (2001), *Tonal Pitch Space*, New York, Oxford University Press.
– Maess B., Koelsch S., Gunter T., Friederici A. (2001), « Musical syntax is processed in Broca's area », *Nature Neuroscience,* 4 (5), p. 541-545.
– Meyer L. (1958), *Emotion and Meaning in Music*, Chicago, The University of Chicago Press.
– Patel A. D., Gibson E., Ratner J., Besson M. et Holcomb P. J. (1998), « Processing syntactic relations in language and music : An event-related potential », *Journal of Cognitive Neurosciences*, 10, p. 717-733.
– Poulin-Charronnat B., Bigand E., Madurell F., Peereman R. (2005), « Musical structure modulates semantic priming in vocal music », *Cognition*, 94 (3), B67-B78.
– Poulin-Charronnat B., Bigand E. et Koelsch S. (2006), « Processing of musical syntax tonic versus subdominant : An event-related potential study », *Journal of Cognitive Neuroscience*, 18 (9), p. 1545-1554.
– Regnault P., Bigand E. et Besson M. (2001), « Different brain mechanisms mediate sensitivity to sensory consonance and harmonic context : Evidence from auditory event related brain potentials », *Journal of Cognitive Neurosciences*, 13, p. 1-17.
– Ross D., Choi J., Purves D. (2007), « Music interval in speech », *PNAS*, 23, p. 9852-9857.
– Sammler D., Grigutsch M., Fritz T. et Koelsch S. (2007), « Music and emotion : Electrophysiological correlates of the processing of pleasant and unpleasant music », *Psychophysiology*, 44, p. 293-304.
– Stravinsky I. (1930), *Chroniques de ma vie*, Paris, Denoël.
– Tillmann B., Koelsch S., Escoffier N., Bigand E., Lalitte P., Friederici A. D. et Cramon D. Y. von (2006), « Cognitive priming in sung and instrumental music : Activation of inferior frontal cortex », *NeuroImage*, 31, p. 1771-1792.

Présentation des auteurs

Simha AROM est ethnomusicologue et directeur de recherche émérite au CNRS. Ses travaux portent sur la systématique musicale des polyphonies d'Afrique centrale, ainsi que sur l'organisation temporelle, la modélisation et les aspects cognitifs des musiques de l'oralité. Nombre de compositeurs – parmi lesquels Luciano Berio, György Ligeti et Steve Reich – ont exploité dans leurs œuvres des procédés musicaux qu'il a mis au jour. Il est notamment l'auteur de *La Fanfare de Bangui. Itinéraire enchanté d'un ethnomusicologue* (2009).

Anne BARGIACCHI est interne en psychiatrie, actuellement en thèse de sciences à l'école doctorale de l'université Paris-VI « Cerveau, cognition et comportement ». Elle participe aux recherches en imagerie cérébrale dans les troubles du développement de l'enfant, en particulier l'autisme, à l'U797 Inserm-CEA à Orsay.

Emmanuel BIGAND est professeur de psychologie cognitive et directeur du Laboratoire d'étude de l'apprentissage et du développement (LEAD) à l'université de Bourgogne. Sa double formation, en musique/musicologie et psychologie cognitive, l'a conduit à développer de nombreux travaux sur les fondements cognitifs des aptitudes musicales. Il est notamment l'auteur de *L'Organisation perceptive d'œuvres musicales tonales* (1995) et coauteur de *Penser les sons. Psychologie cognitive de l'audition* (1994).

Jacques BOUVERESSE est professeur au Collège de France (Philosophie du langage et de la connaissance). Il a publié récemment : *Langage, perception et réalité. I. La perception et le jugement. II. Physique, phénomé-*

nologie et grammaire (1994-2004) ; *Peut-on ne pas croire ? Sur la vérité, la croyance et la foi* (2007) ; *La connaissance de l'écrivain. Sur la littérature, la vérité et la vie* (2008).

Roger CHARTIER est professeur au Collège de France (Écrit et cultures dans l'Europe moderne) et professeur associé à l'Université de Pennsylvanie. Ses recherches portent sur l'histoire du livre, de la lecture et, plus particulièrement, sur les relations ente culture écrite et littérature. Ses deniers ouvrages publiés sont *Inscrire et effacer. Culture écrite et littérature : XI^e^-XVIII^e^ siècle* (2005) et *Écouter les morts avec les yeux* (2008) et *Au bord de la falaise. L'histoire entre certitudes et inquiétude.* (2009)

Stanislas DEHAENE est professeur au Collège de France (Psychologie cognitive expérimentale). Ses recherches visent à élucider les bases cérébrales des opérations les plus fondamentales du cerveau humain : lecture, calcul, raisonnement, prise de conscience. Il est notamment l'auteur de *La Bosse des maths* (2003), *Vers une science de la vie mentale* (2006), et *Le Neurones de la lecture* (2007).

Ghislaine DEHAENE-LAMBERTZ est pédiatre, directrice de recherche au CNRS, et responsable de l'équipe de neuro-imagerie et développement dans l'unité Inserm U562. Elle étudie les bases cérébrales des fonctions cognitives de l'enfant, et notamment les particularités de l'organisation cérébrale du nourrisson qui favorisent l'acquisition du langage dans notre espèce.

Michael EDWARDS est professeur honoraire au Collège de France (Étude de la création littéraire en langue anglaise), ainsi que poète en français et en anglais. Il est l'auteur de nombreux ouvrages sur la philosophie de la création littéraire (d'Homère à nos jours) et artistique (peinture, sculpture, musique). Il a récemment publié *Shakespeare et l'œuvre de la tragédie* (2005), *Le Génie de la poésie anglaise* (2006) et *De l'émerveillement* (2008).

Dan GNANSIA est chercheur au Laboratoire de psychologie de la perception (CNRS, Paris), et auteur d'une thèse intitulée *Intelligibilité dans le bruit et démasquage de la parole chez les sujets entendants, mal entendants et implantés cochléaires* (2009).

Claude HAGÈGE est professeur honoraire au Collège de France (Théorie linguistique). Il s'est attaché à mettre en évidence les propriétés communes des langues, à lier les traits généraux et la recherche typolo-

gique. Dans ses travaux récents, il s'est efforcé de construire un modèle théorique rendant compte de la relation entre l'homme et le langage. Il a notamment publié *Halte à la mort des langues* (2001), *Combat pour le français : au nom de la diversité des langues et des cultures* (2006), *Dictionnaire amoureux des langues* (2009)

Martine HAUSBERGER est directrice du Laboratoire d'éthologie animale et humaine (Ethos) de l'université de Rennes-I. Ses travaux, qui portent notamment sur le chant des étourneaux, combinent éthologie, psychologie cognitive et neurosciences. Elle est notamment coauteur de *Social influences on vocal development* (1997).

Régine KOLINSKY est maître de recherche du Fonds de la recherche scientifique de Belgique (FNRS) et directrice de l'unité de recherche en neurosciences cognitives de l'Université libre de Bruxelles. Ses travaux portent notamment sur la perception de la parole, sur les conséquences cognitives et cérébrales de l'apprentissage de la lecture, et sur les traitements comparés de la musique et du langage. Elle a notamment codirigé *La Reconnaissance des mots dans les différentes modalités sensorielles* (1991).

Christian LORENZI est professeur en psychologie expérimentale (université Paris-Descartes), directeur adjoint du Département d'études cognitives de l'École normale supérieure de Paris, et le Groupe de recherche en audiologie expérimentale et clinique (GRAEC). Il mène des recherches en psychoacoustique, neurophysiologie, audiologie expérimentale, neuropsychologie et modélisation informatique Ses travaux portent notamment sur la perception auditive de la structure temporelle des sons.

Helen J. NEVILLE est professeur de psychologie et de neurosciences à l'Université d'Eugene (Oregon, États-Unis), où elle dirige le Laboratoire sur le développement du cerveau et le Centre des neurosciences cognitives. Ses travaux portent sur les contraintes biologiques et le rôle de l'expérience dans le développement neurosensoriel et neurocognitif des humains.

Pierre-Yves OUDEYER est chercheur à l'Inria Bordeaux-Sud-Ouest, où il est responsable de l'équipe Flowers en robotique développementale et sociale. Ses travaux portent sur les mécanismes qui permettent aux humains et aux machines de développer des capacités perceptuelles, motivationnelles, comportementales et sociales afin de pouvoir partager des représentations culturelles et interagir naturellement dans le monde réel.

Il est notamment l'auteur de *Self-Organization in the Evolution of Speech* (2006).

Isabelle PERETZ est professeure au département de psychologie de l'Université de Montréal, et codirectrice du laboratoire international de recherche sur cerveau, musique et son (BRAMS) à Montréal. Ses recherches portent sur le potentiel musical de la population générale, en passant par l'étude de son organisation cérébrale, sa transmission génétique et sa spécificité à l'égard du langage. Elle est notamment coauteure de *The Cognitive Neuroscience of Music* (2002).

Christine PETIT est professeur au Collège de France (Génétique et physiologie cellulaire) et dirige le Laboratoire de génétique et physiologie de l'audition à l'Institut Pasteur. Ses recherches, qui ont porté notamment sur le développement et le fonctionnement de la cochlée, ont ouvert le champ des surdités héréditaires à l'analyse génétique et elle a identifié avec son équipe un grand nombre de gènes impliqués. Elle travaille actuellement à élucider la pathogénie moléculaire de plusieurs formes génétiques de surdité.

Jean-Claude RISSET a mené parallèlement une carrière de chercheur et de compositeur. Pionnier de la synthèse des sons aux Bell Laboratories dans les années 1960, il a effectué des recherches sur le son musical et sa perception pour exploiter musicalement ses ressources nouvelles : synthèses imitatives, composition du son, musiques mixtes, paradoxes et illusions acoustiques, duo pour un pianiste.

Luigi RIZZI est professeur de linguistique à l'Université de Sienne, où il dirige le Centre interdépartemental d'études cognitives sur le langage (CISCL). Ses travaux portent sur la théorie de la syntaxe et la syntaxe comparative des langues romanes et germaniques ; il a contribué, en particulier, au développement de l'approche paramétrique de la syntaxe comparative, à la théorie de la localité et à l'étude des représentations syntaxiques. Il s'intéresse aussi à l'acquisition du langage, tout particulièrement au développement de la morphosyntaxe chez l'enfant.

Xavier RODET est responsable de l'équipe Analyse et synthèse à l'Institut de recherche et coordination acoustique/musique (Ircam). Ses travaux portent notamment sur le traitement numérique du signal de parole, le traitement automatique et la reconnaissance de la parole, l'analyse et la synthèse de la voix chantée, et l'informatique musicale. Il a tra-

vaillé également sur les modèles physiques des instruments musicaux. Il a développé de nouvelles méthodes pour l'analyse et la synthèse musicales.

Peter SZENDY enseigne l'esthétique et la philosophie à l'université de Paris-X (Nanterre) ; il est également conseiller pour les programmes de la Cité de la musique à Paris. Ses travaux portent sur la lecture, sur l'écoute, et sur l'histoire des techniques et des corps. Il a récemment publié : *Écoute, une histoire de nos oreilles* (2001) ; *Membres fantômes. Des corps musiciens* (2002) ; *Les prophéties du texte-Léviathan. Lire selon Melville* (2004) *; Sur écoute. Esthétique de l'espionnage* (2007) ; *Tubes. La philosophie dans le juke-box* (2008).

Monica ZILBOVICIUS est psychiatre, directrice de recherche à l'Inserm, et responsable de la recherche en imagerie cérébrale dans les troubles du développement de l'enfant à l'U797 Inserm-CEA à Orsay et à l'hôpital Necker-Enfants malades. Ses travaux portent notamment sur l'autisme.

Table

Cet ouvrage a été transcodé et mis en pages
chez NORD COMPO (Villeneuve-d'Ascq)
N° d'impression :
N° d'édition : 7381-2348-X
Dépôt légal : octobre 2009

Imprimé en France