Developmental Science

Developmental Science (2015), pp 1–13 DOI: 10.1111/desc.12325

PAPER

Electrophysiological evidence of phonetic normalization across coarticulation in infants

Karima Mersad[1,2,3] and Ghislaine Dehaene-Lambertz[1,2,3]

1. *INSERM, U992, Cognitive Neuroimaging Unit, Paris, France*
2. *CEA, DSV/I2BM, NeuroSpin Center, Paris, France*
3. *University Paris-Sud, Cognitive Neuroimaging Unit, France*

Abstract

The auditory neural representations of infants can easily be studied with electroencephalography using mismatch experimental designs. We recorded high-density event-related potentials while 3-month-old infants were listening to trials consisting of CV syllables produced with different vowels (/bX/ or /gX/). The consonant remained the same for the first three syllables, followed (or not) by a change in the fourth position. A consonant change evoked a significant difference around the second auditory peak (400–600 ms) relative to control trials. This mismatch response demonstrates that the infants robustly categorized the consonant despite coarticulation that blurs the phonetic cues, and at an age at which they do not produce these consonants themselves. This response was obtained even when infants had no visual articulatory information to help them to track the consonant repetition. In combination with previous studies establishing categorical perception and normalization across speakers, this result demonstrates that preverbal infants already have abstract phonetic representation integrating over acoustical features in the first months of life.

Research highlights

- Preverbal infants can compute automatically consonant representation, independently of the vocalic context.
- A change of phoneme evoked a mismatch response even when the coarticulated vowels were variable and even with no visual articulatory information.
- Infants share with adults a similar neural architecture suitable for computing phonetic representations from the first months of life.

Introduction

The power of language relies on the combinatorial possibilities of its elementary segments, with the phoneme being the smallest combinatorial unit of the linguistic hierarchy. After the first attempts in the 1950s to describe invariant cues allowing a robust identification of phonemes in the speech signal (Liberman, Delattre & Cooper, 1952), researchers realized that phonetic categories are a cerebral construction. Voices, intonations, speech rates and phonetic combinations affect the physical signal. Nonetheless, human adults usually have no difficulty in decoding the words produced by another human speaker, thanks to phonological representations stored in the left posterior temporal and inferior parietal regions (Caplan, Gow & Makris, 1995; Chang, Rieger, Johnson, Berger, Barbaro *et al.*, 2010; Dehaene-Lambertz, Pallier, Serniclaes, Sprenger-Charolles, Jobert *et al.*, 2005; Jacquemot, Pallier, LeBihan, Dehaene & Dupoux, 2003).

In order to learn their native language, infants also need to perceive and manipulate phonemes while disregarding irrelevant acoustic variations. However, the nature of infants' representations of speech is still in question. Do infants and adults share a similar neural architecture, suitable for computing phonetic representations from birth, or do humans develop these specific

Address for correspondence: Karima Mersad, Université de Paris Descartes, 143 Avenue de Versailles, 75016 Paris, France; e-mail: karima.mersad@parisdescartes.fr

© 2015 John Wiley & Sons Ltd

and efficient representations during language acquisition (Dehaene-Lambertz & Gliga, 2004; Kuhl, 2004)? One way to answer this question is to determine the functional and neural properties of a phonetic representation in adults and then to test whether infants compute representations with similar properties.

There is clear evidence that infants are able to characterize speech sounds beyond their acoustic properties. Firstly, they perceive phonemes categorically: at equivalent acoustic distance, they are better able to discriminate phonemes which cross a phonetic boundary than those belonging to the same phonetic category (Eimas, Siqueland, Jusczyk & Vigorito, 1971). Secondly, they have no difficulty recognizing the same phoneme across variations due to voice or intonation (Jusczyk, Pisoni & Mullennix, 1992; Kuhl & Miller, 1982). A third characteristic property of phonetic representation observed in adults is the capacity to recognize a given phoneme independently of its surrounding phonemes. Once again, acoustic variations due, in this case, to coarticulation might prevent the identification of a consonant occurring in different vocalic contexts. For instance, adults identify the same consonant in /di/ and /du/ whereas the movement of the articulators considerably changes the direction of the second formant transition (F2), from rising in /di/ to falling in /du/. Unlike human adults, who have no difficulty generalizing across these vocalic contexts, monkeys, trained to discriminate /d/ and /b/ followed by the vowels /i/ and /e/, cannot generalize this training to the /u/ and /a/ context (Sinnott & Gilmore, 2004), although quails can (Kluender, Diehl & Killeen, 1987), suggesting interesting species differences. Coarticulation might therefore represent an especially challenging case of normalization, notably in the case of stop consonants, given the short duration of the informative cue and the difficulty of finding obvious acoustic correlates of these categories (Liberman, 1996).

To our knowledge, only two studies have examined this question in infants (Bertoncini, Bijeljac-Babic, Jusczyk, Kennedy & Mehler, 1988; Jusczyk & Derrah, 1987). Using a high-amplitude sucking procedure, infants were habituated to a list of CV syllables sharing the same consonant but with different vocalic contexts. Then, during the test period, a new CV token was introduced, either with a new vowel, a new consonant or both a new vowel and a new consonant. Two-month-old infants noticed all changes (Bertoncini *et al*., 1988; Jusczyk & Derrah, 1987), but neonates missed the change of consonants (Bertoncini *et al*., 1988). The authors hypothesized that the consonant segment might be too short to be characterized consistently at birth when many variations are present.

Our goal here was to examine the neural signatures of this behavior using electroencephalography. Oddball designs have been used for several decades to study auditory representations. In such designs, a new sound randomly introduced in a series of repeated sounds evokes an early automatic response, called mismatch negativity (MMN), in adults (Näätänen & Tiitinen, 1998). Depending on which feature of the sound is changed (e.g., a change of frequency, duration, intensity, etc.), the latency and topography of this response on the scalp are slightly, but significantly, different. These topographical differences indicate that close but different networks are involved in the coding of the different features of a sound (Giard, Lavikahen, Reinikainen, Perrin, Bertrand *et al*., 1995). Using sinewave speech, it was even possible to show that perceiving exactly the same stimuli either as CV syllables or as whistles affects both the subjects' overt detection of a change in the presented series and the mismatch response which appears faster and more left-lateralized in speech than in non-speech listening mode (Dehaene-Lambertz *et al*., 2005). Hence, in adults, phonetic representations are computed early on and in parallel with other sound features and can be studied with MMN paradigms.

In infants, a novel sound introduced in a series of repeated sounds also evokes an early automatic mismatch response. Dehaene-Lambertz and Baillet (1998) reported that the infants' mismatch response for syllables varying along the place of articulation dimension has a larger amplitude, notably over the right frontal electrodes, when the change crosses the /ba/–/da/ boundary than when a similar physical change is made within-category. The dipole modeling of the voltage topographies suggested a more posterior and dorsal dipole for the phonetic change than for the acoustic change, congruent with results obtained in adults with fMRI (Celsis, Boulanouar, Doyon, Ranjeva, Berry *et al*., 1999; Dehaene-Lambertz *et al*., 2005; Jacquemot *et al*., 2003) and electrocorticography (Chang *et al*., 2010), as well as with the role of the left posterior temporal regions in phonological processing (Hickok & Poeppel, 2000). In sleeping neonates, the mismatch response was similar whether acoustical variations due to speaker were present or not (Dehaene-Lambertz & Pena, 2001). In these two studies, the recording of a mismatch response sensitive to phonetic properties (categorical perception, voice normalization) beyond acoustical differences suggests that preverbal infants and adults might share a common neural architecture to automatically and rapidly compute phonetic representations from the speech signal.

Here, we wanted to further explore the sensitivity of the mismatch response to phonetic properties and

© 2015 John Wiley & Sons Ltd

confirm whether infants were able to compute phonetic representation independently of the coarticulation context (Bertoncini *et al*., 1988; Jusczyk & Derrah, 1987). Should this prove to be the case, we should record a mismatch response when a change of consonant occurs after several repetitions, even if the consonant is systematically associated with a different vowel. We thus presented 3-month-old infants with trials comprising four successive CV syllables, each syllable crucially having a different vowel. The first three syllables (called context syllables) shared the same consonant /b/ (or /g/) and were followed by a test syllable which either shared the same consonant (congruent trials) or not (incongruent trials). If infants of this age were only able to represent the syllable as a whole, they would detect no repetition and then would not perceive a greater change in incongruent trials relative to congruent trials. In that case, we should record no mismatch response. Conversely, if infants were able to compute a phonetic representation, the repeated presentation of a consonant in a sequence of syllables, followed by a change of consonant, should elicit a mismatch response, visible when congruent and incongruent trials are compared.

Our second goal was to investigate whether visual articulatory information facilitates phonetic encoding. To account for human perceptual constancy in spite of the variability of speech acoustic patterns, Liberman (1996) proposed a theory in which the speech coding unit is based on motor/articulatory schemes rather than on auditory patterns. This hypothesis, which is still hotly debated, was given a physiological basis with the discovery of mirror neurons in the inferior frontal regions of the macaque (Rizzolatti & Craighero, 2004). These multi-modal neurons not only fire when an action is done by the monkey, but also when the monkey sees the action or hears the sound of the action (Kohler, Keysers, Umilta, Fogassi, Gallese *et al*., 2002; see also Romanski & Goldman-Rakic, 2002). Demonstrating the existence of these mirror neurons in humans can only rely on indirect evidence, and their role in speech perception is disputed (D'Ausilio, Pulvermuller, Salmas, Bufalari, Begliomini *et al*., 2009, vs. Lotto, Hickok & Holt, 2009).

The immaturity of the motor system during the first months of life, manifested in infants' poor vocal productions, seems to be a definitive counter-argument to explaining speech perception capacities by sophisticated motor representations, unless it is assumed that infants possess an innate representation of the gestures allowed by human physical articulators to produce speech. However, one scenario, more in line with the physiological properties of mirror neurons, would be that infants rapidly set up speech motor units by combining visual perception of their caregivers' articulatory movements with their own elementary productions. Infants are indeed able, from birth onwards, to associate the movements they see with their own movements (Chen, Striano & Rakoczy, 2004; Meltzoff & Moore, 1977, 1989). Importantly, they are also able to link the articulatory gesture they see with the appropriate sound (Bristow, Dehaene-Lambertz, Mattout, Soares, Gliga *et al*., 2009; Kuhl & Meltzoff, 1982; Patterson & Werker, 2003). The left inferior frontal region, where mirror neurons have been postulated (Kohler *et al*., 2002), has been seen activated in infants during speech perception in several brain imaging studies (Bristow *et al*., 2009; Dehaene-Lambertz, Hertz-Pannier, Dubois, Meriaux, Roche *et al*., 2006; Mahmoudzadeh, Dehaene-Lambertz, Fournier, Kongolo, Goudjil *et al*., 2013) and might be the crucial converging hub between the different representations of speech, thanks to its connections with the appropriate brain regions (i.e. the superior temporal and inferior parietal regions and the ventral regions of the temporal lobe), notably through the arcuate and the fronto-occipital fasciculi (Turken & Dronkers, 2011).

Therefore, to examine the role of visual articulatory gesture in syllable perception, we presented trials in which an articulating face was associated with the syllables during the context phase (mouth movement trials – hereafter M_mov). We contrasted this condition with trials in which the mouth was hidden by a surgical mask but the eyes were blinking at the onset of the auditory syllable (blinking eyes trials – hereafter E_blink) during the context phase. If the auditory information is by itself sufficient to detect the repetition and change of the consonant, a mismatch response should be recorded in the E_blink condition. However, if infants need to recover the articulatory pattern to succeed, there are two alternatives: (1) the mismatch response might be present only in the M_mov condition, in which the articulating face can help them to recover the underlying gesture; (2) we may observe no mismatch response in any modality, in line with the hypothesis that the infants, who are not able to produce the syllables themselves at the age at which we tested them, have no motor representation.

Materials and Method

Participants

Twenty-five full-term French infants (12 girls and 13 boys), living in a French- speaking environment, were tested at a mean age of 12 weeks 2 days (10 w. 3 d. to 13

© 2015 John Wiley & Sons Ltd

w. 1 d.). An additional 13 infants did not provide exploitable data: 10 were too agitated during the test to obtain clean EEG data, two refused to wear the EEG sensor net and one infant fell asleep before the test. The study was approved by the regional ethical committee for biomedical research and parents gave their written consent for the protocol.

Stimuli

To test whether facial movements help infants during the context phase (the first three syllables of each trial), we chose vowels that make the clearest visual distinction between /b/ and /g/. As /a/ is the best vowel for visual identification of the initial consonant (Massaro, Cohen & Gesi, 1993), we chose the vowels (/ɛ/, /ã/, /ɛ̃/) belonging to the same viseme category to constitute our set of context vowels (Montgomery & Jackson, 1983). As the test vowel, we choose /i/, a vowel located in a different corner of the vowel triangle, in order to maximize acoustic differences between the test and context vowels. As shown in Figure 1, F2 transition varied significantly in direction and in duration within the context syllables, and consequently was not a reliable cue to recover an invariant acoustic property of the consonant.

Auditory stimuli

Eight syllables were produced in a natural manner by a French female speaker (/ga/: 355 ms, /gɛ/: 396 ms /, /gã/: 374 ms, /gɛ˜ /: 384 ms, /ba/: 342 ms, /bɛ/: 356 ms, /bã/: 333 ms, and /bɛ˜/: 348 ms) to be used as context stimuli. As variability was welcome, no effort was made to control the duration of these syllables. By contrast, the two syllables (/bi/: 354 ms, /gi/: 357 ms) used as test stimuli, produced by the same speaker, were equalized in duration (356 ms) and intonation. All syllables were matched for subjective intensity.

All stimuli were recorded on the left channel and a click was positioned on the right channel at the exact time-point of the syllable onset. The left channel was connected to the audio amplifier to present the sound in mono mode to the subject while the right channel was connected to the EEG amplifiers through the DIN port as a TTL signal. This method provided a precise relation between EEG recordings and the sound as the brain voltage and the trigger signal were recorded simultaneously with the same temporal resolution when the file was played by the PC soundcard.

Visual stimuli

The same woman was filmed articulating /ba/ and /ga/ against a light blue background. Five frames were extracted from each clip: (i) mouth closed, (ii) lips lightly joined, (iii) lips tightly joined, (iv) mouth semi-extended, (v) mouth fully extended (Figure 2). These images were used to create natural articulatory movements with precise onsets during the context part of the M_mov condition. Two still images of the same woman with her eyes open and closed were extracted from the movie. For these two images, her mouth was hidden by a surgical mask. These frames were used during the context part of the E_blink auditory condition. Finally, a colored 'bullseye' was the visual stimulus for the test period in all conditions.

Procedure

Infants wearing the EEG net were seated on a parent's lap facing a projector screen with a loudspeaker positioned behind the screen (mono presentation). The screen was located approximately 80 cm away from the infant's face (visual angle = 0.184 radian). Auditory stimuli were presented at 68 decibels. An experimenter (KM) monitored an online video of the infant looking at

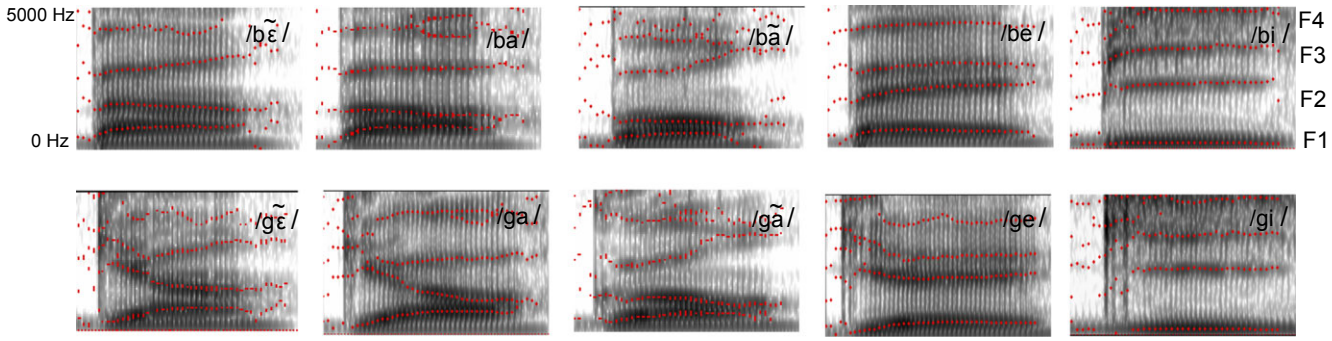

Figure 1 *Spectrograms and formant transitions of the syllables used as context and test auditory stimuli. In these naturally produced syllables, F2 transition is variable in direction and duration.*

© 2015 John Wiley & Sons Ltd

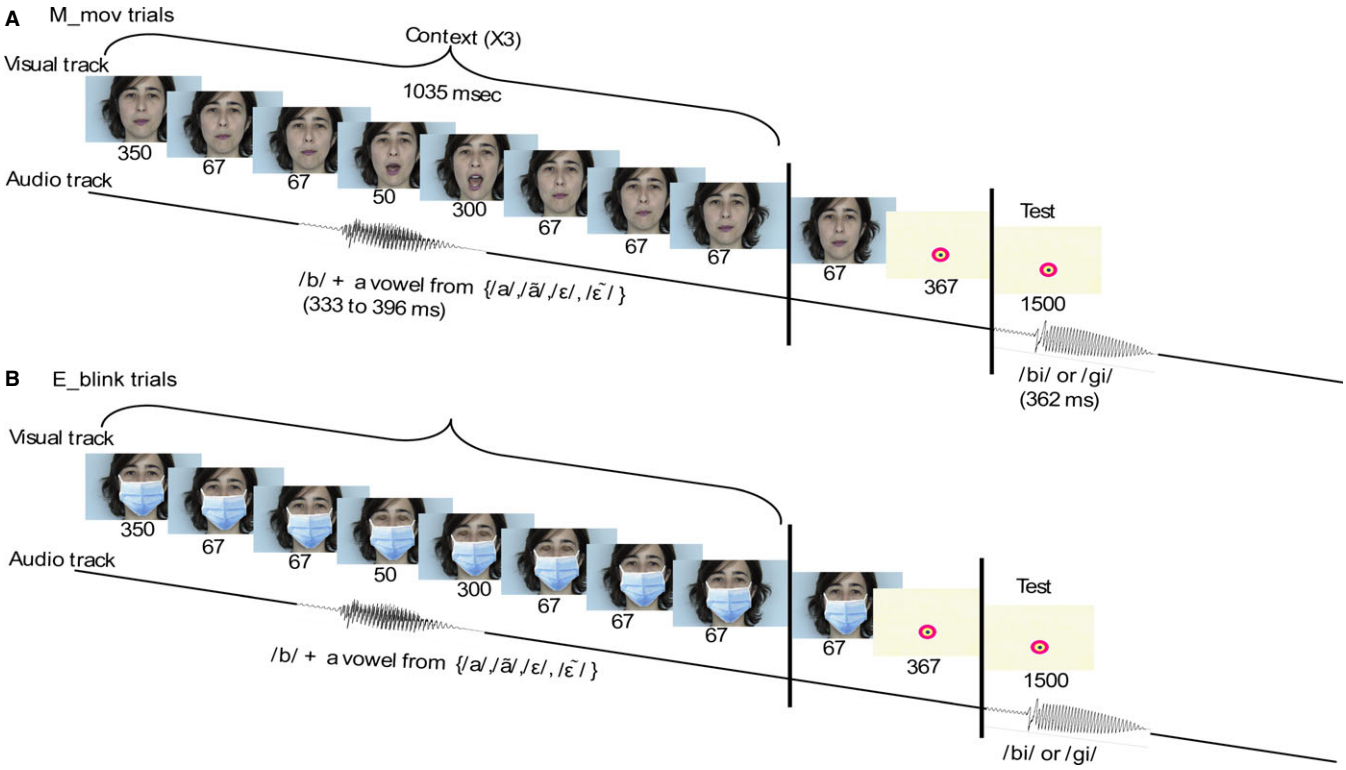

Figure 2 *Trial structure. (a) Mouth movement (M_mov) trials: The trial began with the presentation of a face articulating three times in succession /b/ or /g/ + a vowel randomly chosen from {/a/, /ã/, /ɛ/ or /ɛ̃/}, for example /ba/, /bɛ/, /bɛ̃/ (context stimuli sharing the same viseme). Then, the face was replaced by a bullseye followed by the auditory test syllable (/bi/ or /gi/). (b) In blinking eyes (E_blink) trials, the auditory track was similar to the M_mov trials but the face was presented with the mouth hidden with a surgical mask and the eyes blinked at the onset of the syllable. Relative to context syllables, the consonant changed in the test part of incongruent trials but remained the same for congruent trials.*

the screen. If the infant looked away from the screen, the experiment was paused and the infant's gaze attracted back to the screen by pointing or tapping on the screen before the experiment resumed. If this was not possible, the experiment was terminated.

Each trial began with a still image of a woman with her mouth fully closed, presented for 533 ms, then four auditory syllables were successively presented (SOA = 1035 ms). The first three syllables constituted the phonetic *context* of the trial. They shared the same first consonant (/b/ or /g/) but were coarticulated with three different vowels, each of them being randomly chosen at each trial from the set {/a/, /ɛ/, /ɛ̃/, /ã/}. The last syllable of the trial was the *test* syllable (/bi/ or /gi/) whose consonant was either congruent or not congruent with the phonetic context. Thus each trial comprised a context and a test part.

Depending on the visual information given during the *context part* (the first three syllables of the trial), two types of trials were contrasted. In M_mov trials, the mouth began to open at the onset of the auditory syllable, with the mouth fully open 117 ms after the auditory syllable onset. Note that we used a single articulatory movement /ba/ (or /ga/) for all the syllables presented in the context phase, as /ba/ and /ga/ have the clearer visemes for /b/ and /g/ (Massaro *et al.*, 1993) and the other vowels used in the context have visemes close to /a/. Our goal was to give unambiguous visual information which showed articulation repetition and clearly differentiated the two consonants. In E_blink trials, there was no articulatory movement and the mouth was hidden by a surgical mask, but the eyes blinked at the onset of the auditory syllable and remained closed for the entire duration of the auditory syllable.

In each trial, the context part was followed by a test part where the face was replaced by a brightly colored 'bullseye' in all conditions, to present the test syllable in a visually neutral context. The test syllable occurred 367 ms after the bullseye (for a precise time-course of the trials, see Figure 2). The bullseye remained until the beginning of the next trial (1500 ms).

Trials were presented in alternating blocks of 30 trials sharing the same phonetic context (/b/ or /g/) and the same modality (E_blink or M_mov). Each block

© 2015 John Wiley & Sons Ltd

comprised half congruent and half incongruent trials presented randomly. The order of blocks was counterbalanced across subjects (2 modalities × 2 phonetic contexts = 120 trials). If the infant was still interested, the same order of blocks was presented a second time. In order to be included in the analyses, infants had to undergo at least two blocks (one in each modality) and, on average, the included infants completed 108 trials.

ERP recordings

Scalp voltages were recorded from a Geodesic sensor net (EGI, 129 channels, amplifiers N200) referenced to the vertex. They were amplified, digitized at 250 Hz, and filtered between 0.5 and 20 Hz. The EEG was then segmented into epochs starting 3500 ms before (to include all context syllables – S1, S2, S3) and ending 1200 ms after the test syllable (S4). As onset, we used the click signal of the auditory files, sent to the EEG recording device when the audio file was played.

For each epoch, channels contaminated by eye or motion artifact were rejected if one of the following two criteria was met: (1) local deviation larger than 150 microvolts, or (2) mean of the voltage higher/lower than 3 standard deviations from the mean computed over all epochs. Trials in which more than 40% of the channels were contaminated were excluded. Channels with fewer than 15 trials in one condition were rejected for the entire recording and removed from further analysis. On average, 2.6 bad channels/infant were excluded. Furthermore, three channels of the net (the nasion and the two infra-occular electrodes) were systematically bad, or not used, and thus rejected in all subjects. However, given that the net had 128 electrodes, the head coverage remained acceptable in each infant (around 122 channels in each infant).

After artifact rejection, 84 trials on average were retained per infant. As we were not expecting differences relative to the precise phonetic contrast, we collapsed the /b/ and /g/ context and the artifact-free trials were averaged in the four conditions (Modalities (2 levels) × Congruency (2 levels)) with a mean number of trials per condition: 21.2, 21.9, 20.5, 20.7 for the E_blink congruent, E_blink incongruent, M_mov congruent and M_mov incongruent conditions, respectively. To obtain reference-free data, the averages were re-referenced to the mean voltage at each data-point (average-reference (Dien, 1998)) and finally baseline-corrected (−3500 ms to 0 ms) in each infant.

Statistical analyses

Statistical analyses are an issue for high-density recordings because of the large number of possible comparisons which risks family-wise errors. We used two approaches to circumvent this problem. Firstly, we used classical ERP analyses driven by a prioris based on previous studies of auditory discrimination in infants. When infants perceive an auditory change, an early mismatch response (MMR) is recorded, which may be followed by a late slow wave (LSW) (Dehaene-Lambertz & Dehaene, 1994; Friederici, Friedrich & Weber, 2002; Kushnerenko, Ceponiene, Balan, Fellman, Huotilaine *et al.*, 2002). We used the known characteristics of these responses to define the studied time-windows and clusters of electrodes.

Although sensitive, these analyses driven by the literature are limited. They might ignore new and unexpected effects. This is why we completed these analyses with a data-driven approach in which we reduced the number of possible comparisons by summarizing the voltage of the difference between two conditions considered at each time-sample as the mean across all electrodes, of the absolute value of the difference between the two factors studied. We ensured statistical significance by evaluating the null hypothesis distribution in our population through iterative permutations of the condition labels. This method is robust and helpful to determine the precise time-window of an effect for a specific experiment, but it lacks sensitivity because, if the difference between conditions only concerns a few electrodes, it might be diluted by the process of averaging across all electrodes. These two methods are thus complementary.

Literature-driven analyses (MMR)

In auditory oddball studies, the change of stimulus in series of repeated stimuli evoked a first automatic mismatch response (MMR), recorded even when infants were asleep (Dehaene-Lambertz & Pena, 2001; Friederici *et al.*, 2002; Kushnerenko, Winkler, Horvath, Naatanen, Pavlov *et al.*, 2007). In previous studies using the same AAAX paradigm at the same age as here, we recorded a mismatch response consisting of a predominantly right frontal positivity synchronous with a posterior negativity around the latency of the second auditory peak (Dehaene-Lambertz & Gliga, 2004). Some studies using more classical oddball paradigms (i.e. random presentation of a deviant sound among repeated sounds) may report different polarities: anterior negativity and posterior positivity (e.g. Kushnerenko, Cheour, Ceponiene, Fellman, Renlund *et al.*, 2001). The mechanisms explaining these changes of polarity between studies are still not understood. In any case even when polarities are reversed, source modeling located the main sources in the superior temporal regions (Dehaene-Lambertz &

© 2015 John Wiley & Sons Ltd

Baillet, 1998; Dehaene-Lambertz & Dehaene, 1994) as is the case in adults. Source modeling of EEG voltage is confirmed by more spatially accurate imaging techniques: MEG recording in newborns (Huotilainen, Kujala, Hotakainen, Shestakova, Kushnerenko *et al.*, 2003) and a recent NIRS study in preterm infants (Mahmoudzadeh *et al.*, 2013).

As can be seen in Figure 3, the stimuli elicited the classic auditory response at this age, which consists of two positive peaks over the lateral frontal regions with a polarity reversal on the occipito-parietal regions. Their latency measured on the test syllable across all conditions was 352 and 488 ms. We examined the 2D maps of the difference between all standard and deviant trials around the second peak latency. A classical topography of mismatch response was observed with maxima over the right frontal and left parietal regions. We selected two clusters on these maxima: 17 sensors over the right frontal area (comprising FZ, F4 and C4) and 12 sensors over the left occipito-temporal area (between and above O1–T5), in agreement with published studies. We averaged the voltage on these clusters for 200 ms after each syllable comprising the peak and descending slope of the second auditory peak (400–600 ms post-syllable onset) in each infant, and entered these values in an ANOVA with Cluster (positive and negative cluster), Congruency (congruent and incongruent), Syllable position (mean of (S1, S2, S3) and S4) and Modality (E_blink and M_mov) as within-subject factors. Note that because the two clusters of electrodes represent the two poles of the same effect, the relevant comparison to identify a condition effect is the interaction Cluster × Congruency.

To be sure that the putative difference between standard and deviant trials occurring in response to S4 was significantly more important than any spurious noise remaining in the individual averages, we introduced a syllable position

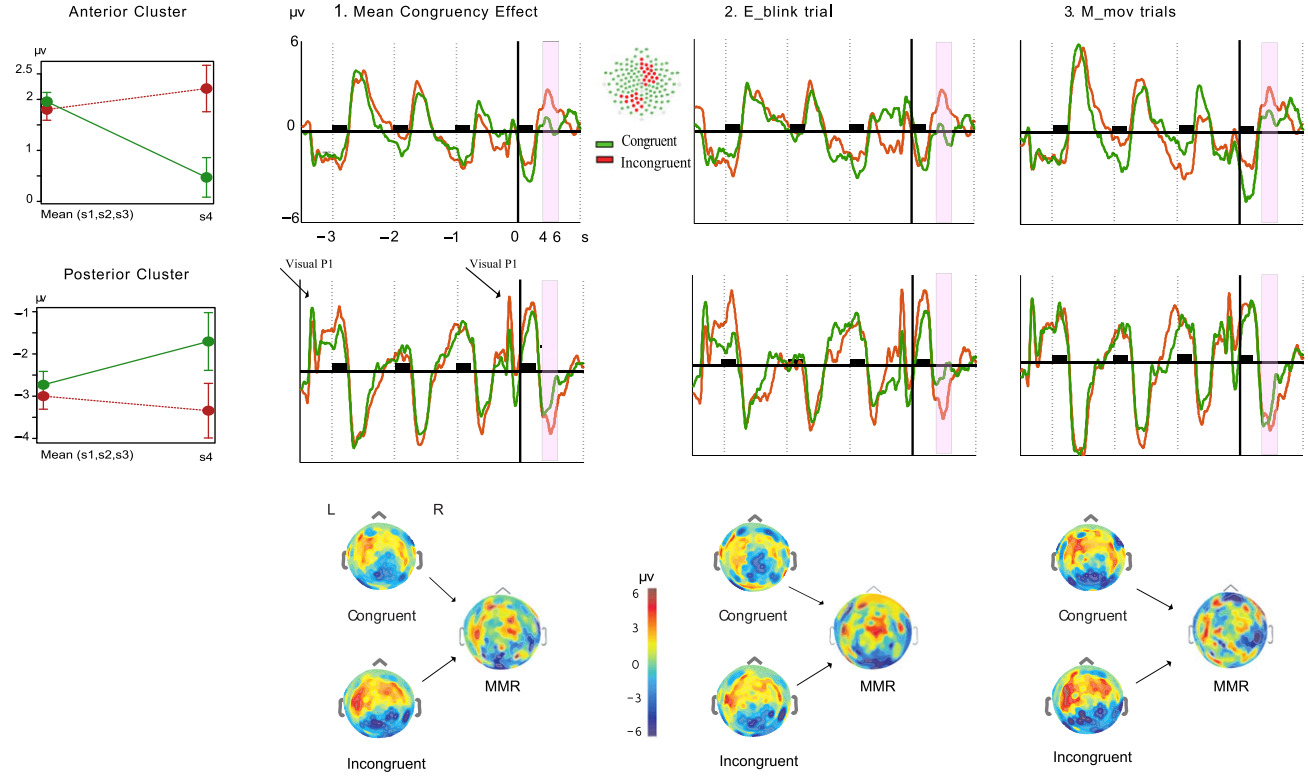

Figure 3 *Mismatch response in E_blink and M-mov trials. Panel 1: Main effect (E_blink and M-mov conditions are merged), Panel 2: E_blink trials, Panel 3: M_mov trials. First two rows: grand average of congruent and incongruent trials recorded from right anterior (first row) and left posterior (second row) clusters of electrodes. The selected electrodes are indicated as red dots on the electrodes net topography between Panels 1 and 2. Red rectangles on the waveform represent the statistical time-window and the black horizontal rectangles represent the auditory stimuli. On the posterior cluster B, the transition between the face and the bullseye is clearly visible as a visual N1 P1 N2 complex (arrow). The last rows present the topographies of the evoked potentials in response to the auditory test stimulus in congruent and incongruent conditions and the difference between the two conditions (MMR) averaged on the statistical time-window. A significant MMR is clearly recorded in the blinking eyes condition but is weaker in the mouth movement condition. Left panel. Mean and standard error of the voltage of Congruent versus Incongruent trials for context and test syllables computed across the time-windows and clusters of electrodes selected for the MMR.*

© 2015 John Wiley & Sons Ltd

factor. As the context syllables were the same in deviant and standard trials, variability in the responses to S1-S2-S3 between congruent and incongruent trials reflects unwanted noise. A significant MMR should thus be revealed by a significant interaction Cluster × Congruency observed in response to S4, but not to S1-S2-S3, that is, by a triple interaction Syllable × Cluster × Congruency. Similarly, to assess an effect of modality, the context syllables are used to control for any difference of general attention across the blocks, which should similarly affect the context and test syllables.

As an exploratory and follow-up analysis, we also examined whether a significant interaction Modality × Congruency over the same time-window might be observed on other channels than the classical channels recording the mismatch response. As can be seen in Figure 3, the MMR response appeared to be more intense on the medial electrodes in the E_blink modality. We thus selected 11 electrodes in front of CZ (positive cluster) and 10 electrodes around OZ (negative cluster) and performed the same analysis as above during the same time-windows.

Literature-driven analyses (LSW)

When infants are awake, a second response, a negative wave, or late slow wave (LSW), has sometimes been reported around 900 ms over frontal areas (Basirat, Dehaene & Dehaene-Lambertz, 2014; Dehaene-Lambertz & Dehaene, 1994; Friederici *et al.*, 2002). We indeed observed a weak negative response developed over the right anterior frontal region around 900 ms. We thus selected 10 sensors over this region (comprising FZ and F4) and performed an ANOVA similar to the above-described ANOVA on the voltage averaged across these sensors and for 100 ms after each syllable (820–920 ms) in each infant, with the same factors (except cluster).

Data-driven permutation analyses

We reduced the spatial dimension of the data by computing at each time-point the mean across all electrodes of the absolute value of the voltage difference between conditions (congruent and incongruent) in each infant. We then compared the grand average obtained across infants to surrogate data obtained by permuting the labels of the conditions within subject, followed by the same processing steps as for the real data. In each infant, 1000 permutations were carried out, contributing to 1000 grand-averages across infants. The *p*-values were determined at each time-point as the number of surrogate grand averages above the real grand-average divided by the number of permutations. This method provided accurate assessment of the null hypothesis in our data set. This second analysis should identify without priors the significant time-points when the two conditions differed. We carried out this analysis on the difference between congruent and incongruent trials independently of the modality, then in each modality, and finally on the difference between the differences obtained in each modality.

Results

MMR time-window: 400–600 ms

If infants were able to identify the same phoneme in spite of the variable vocalic context, we expected the change of phoneme to induce, around the second peak latency, an MMR (Dehaene-Lambertz & Dehaene, 1994), attested here by a significant Congruency × Cluster × Syllable position interaction. This was confirmed by the ANOVA result ($F(1, 24) = 6.92$, $p = .01$, $\eta^2_{partial} = .22$; see Figure 3 for effect sizes). Post-hoc analyses restricted to each syllable position confirmed that a significant effect of congruency was indeed observed for the test syllable S4 (Cluster × Congruency: $F(1, 24) = 10.86$, $p = .003$, $\eta^2_{partial} = .31$), and not for the context syllables (Cluster × Congruency: $F(1, 24) < 1$). Note, however, that a baseline taken during the whole context part might affect this comparison. We thus confirmed that there was no spurious effect of congruency during the context when taking a 100 ms baseline before S1: Cluster × Congruency: $F(1, 24) < 1$.

Post-hoc analyses on each cluster showed a significant interaction of Syllable position × Congruency at the anterior cluster ($F(1, 24) = 11.94$, $p = .002$) but not at the posterior cluster ($F < 1$). Analyses restricted to S4 showed a significant effect of congruency at the anterior cluster ($F(1, 24) = 13.10$, $= .001$, $\eta^2_{partial} = .35$, incongruent/congruent trials: $M = 2.19$ μv, ($SE = .45$) vs. .44 μv ($SE = .38$)), whereas it was only marginal at the posterior cluster ($F(1, 24) = 3.55$, $p = .07$, $\eta^2_{partial} = .013$, incongruent/congruent trials: $M = -3.34$ μv ($SE = .64$) vs. -1.70 μv ($SE = .68$)). None of these effects had a significant interaction with Modality (E_blink vs. M_Mov).

As an exploratory and follow-up analysis, we examined whether a significant interaction of Modality × Congruency over the same time-window might be observed on other channels. As the MMR in the E_blink modality appeared to extend more on medial electrodes, two new clusters of electrodes were selected. On these clusters, a significant Syllable position × Cluster × Congruency × Modality interaction was observed ($F(1, 24) = 26.14$, $p < .001$) due to a significant interaction of

© 2015 John Wiley & Sons Ltd

Cluster × Congruency × Modality in response to S4 (F(1, 24) = 17.44, p < .001). Post-hoc analyses confirmed a significant MMR (i.e. Cluster × Congruency interaction) in the E_blink modality (F(1, 24) = 13.89, p = .001) but not in the M_Mov modality (F(1, 24) < 1). For each cluster, the interaction Syllable position × Congruency × Modality was significant (central cluster: F(1, 24) = 13.87, p = .001; occipital cluster: F(1, 24) = 10.29, p = .003) due to a larger mismatch response for the E_Blink modality (Congruency × Modality interaction in response to S4 on the central cluster: F(1, 24) = 11.22, p = .003; occipital cluster: F(1, 24) = 8.68, p = .007).

Late frontal negativity: 820–920 ms

The early response was followed by a late frontal negativity (Figure 4) similar in timing and topography to what was observed in other studies with attentive infants (Basirat *et al.*, 2014; Dehaene-Lambertz & Dehaene, 1994; Friederici *et al.*, 2002). This late response was weak. Although the effect of congruency was significant only for the test syllable (S4: F(1, 24) = 4.32, p = .04; context phase: F(1, 24) < 1), the interaction Congruency × Syllable position did not reach significance (F(1, 24) = 2. 88, p = .10). There was no effect of Modality (F(1, 24) < 1).

Permutation analyses

This analysis identified only one significant time-period (p < .05 from 560 to 630 ms) for the main effect of congruency (Figure 5). This time-period is shorter but around the same time-period as the one selected above guided by the literature. The significant time-periods were (610 to 630 ms) and (600 to 620 ms) for the E_blink and M_mov conditions, respectively. There was no significant point at a later time-window, which confirms that the late frontal negativity was a weak response. The same analysis carried out on the differences between modalities did not find significant time-windows.

Discussion

As is classically described when a change of sound is detected after a series of repeated sounds, we recorded in our 3-month-old infants an auditory MMR beginning around the latency of the second peak of the infants' auditory response (see Figure 3) but becoming significant only around 560 ms in the permutation analysis. This significant difference between congruent and incongruent trials demonstrates that infants detected the repetition and change of the initial consonant in the CV syllables presented despite the variation of vowels.

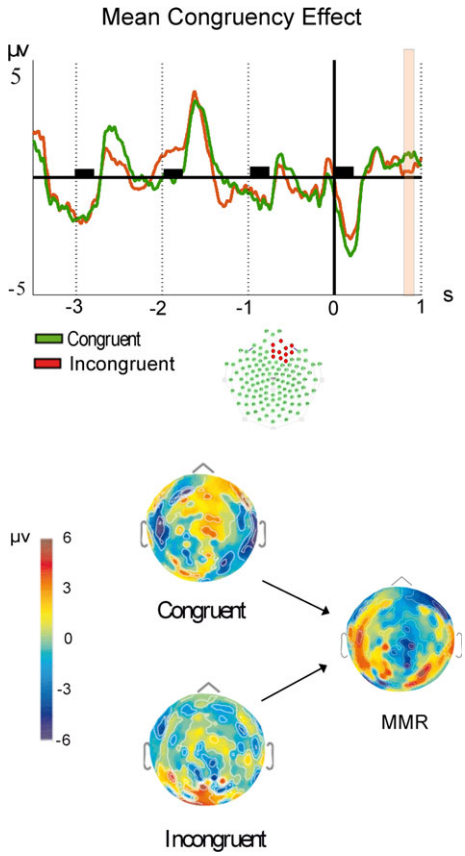

Figure 4 *Late frontal negative response. Upper row: grand average of congruent and incongruent trials recorded from right frontal cluster of electrodes for both modalities merged. The topographies represent the evoked potentials in response to the auditory test stimulus in congruent and incongruent conditions and the difference between the two conditions averaged on the statistical time-window.*

Some experiments have reported a late slow wave following the MMR in infants who were awake (Friederici *et al.*, 2002). This response, related to an orientation of attention to a novel event (Csibra, Kushnerenko & Grossmann, 2008; Dehaene-Lambertz & Dehaene, 1994), was not observed in the present data, or if present, this response was not robust. Although a clear mismatch response was observed, it was not sufficient to clearly elicit an attentional orientation response in such a constantly varying vowels context.

A phonetic representation already present in the infant brain

Auditory mismatch responses have been described in adults as an automatic detection of an auditory change based on statistics automatically performed on the previous stimuli. We demonstrate here that these statis-

© 2015 John Wiley & Sons Ltd

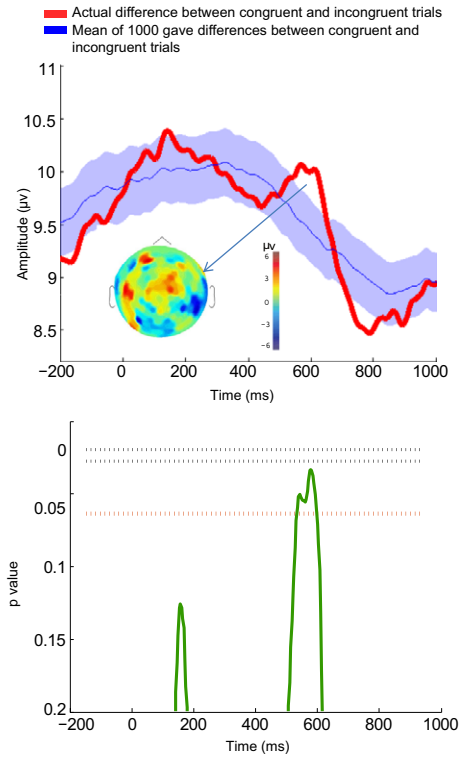

Figure 5 *Permutation analyses. Upper panel: the red line represents the mean voltage difference between congruent and incongruent conditions computed across all channels and across all infants, and the blue line the mean of the surrogate data with their distribution in shaded blue. The presented topography, averaged on the time-window captured by the permutation analysis, nicely fits with the one selected from the literature. Lower panel: the green line represents the p-value (< .05 between 560 and 630 ms).*

tics can be performed by infants as young as 3 months within CV syllables on an acoustic-invariant representation of the first phoneme.

A phonetic representation includes two fundamental properties: categorical perception and normalization across non-relevant acoustic variations, such as those produced by speakers or by the coarticulation context. Previous electrophysiological studies in preverbal infants have reported categorical perception (Dehaene-Lambertz & Baillet, 1998) and voice normalization (Dehaene-Lambertz & Pena, 2001). Here we add evidence that infants can also compute automatically consonant representation, independently of the vocalic context, even in the difficult context of place of articulation. This is not a trivial performance as the information on the place of articulation is a very short cue, based on the variations of F2 transition, which is easily lost in noisy conditions and poorly perceived by pathological subjects, such as adults after a left stroke (Caplan *et al.*, 1995) or language impaired/dyslexic children (Kraus, McGee, Carrel, Zecker, Nicol *et al.*, 1996).

Furthermore, the significant mismatch response in the blinking eyes condition, where no articulatory movement was presented, demonstrates that auditory information was by itself sufficient for infants to recover the consonant identity, as shown with behavioral methods in 2-month-olds (Bertoncini *et al.*, 1988; Jusczyk & Derrah, 1987). At the age they were tested, infants' productions are still poor, consisting mainly of vocalizations, and they are still not able to produce the two phonemes, /b/ and /g/, that they discriminate here. Sinnott and Gilmore (2004) showed that monkeys have difficulty generalizing a /b/–/d/ discrimination learned in an /a/–/u/ context to an /i/–/e/ context, and hypothesized that monkeys focus on the second formant transition which largely separates /ba/ and /gu/ but is far less informative for /bi/ and /ge/. Assuming that human adults' performances were reached through speech production practice, the authors suggested that monkeys might model a preverbal human infant. Our results do not support this hypothesis. The ability to produce these consonants is not a prerequisite for humans to correctly perceive the place of articulation. This suggests that, at least for the place of articulation continuum, categorization does not rely on the same perceptive cues among primates.

The visual, motor and auditory components of speech

Speech has three components: auditory, visual and proprioceptive/motor. These three modalities have a very different developmental calendar. Auditory perception is functional during gestation, visual information becomes available at birth, and it takes several months to achieve efficient control of complex articulatory movements. How are these three components integrated in a common phonetic representation?

In infants, neonatal imitation capacities may be an argument supporting innate motor representations of speech. Indeed, without any training, human neonates imitate facial movements (Meltzoff & Borton, 1979) and try to produce sounds congruent with the auditory-vocal models they are exposed to (Chen *et al.*, 2004; Kuhl & Meltzoff, 1982). In this context, we would have expected visual articulatory cues to significantly improve infants' perception relative to eye-blink. This was not the case and, if anything, we obtained the reverse effect, that is, a clear MMR in the Eye-Blink blocks and a weak effect in the blocks with visible mouth movements. Although an innate representation of the gesture pattern of all human sounds directly linked to auditory perception, with no

© 2015 John Wiley & Sons Ltd

need for visual information, could still be proposed, the fact that quails can learn to correctly categorize /b/, /d/ and /g/ followed by different vowels (Kluender *et al*., 1987) confirms that the auditory signal is by itself sufficient to categorize these consonants.

We should be cautious in interpreting the significant difference between the MMRs in the two modalities (E_blink and M_mov), as we had no priors for this comparison and as permutation analyses revealed no effect for this comparison. Thus, further testing is needed for a definitive conclusion, notably about a stronger response in the pure auditory modality. At least, the visual articulatory information seems not to be helpful to infants at this age. Although there is evidence that audio-visual mapping occurs rapidly during the first months of life, this was shown for simple vowel gestures (Chen *et al*., 2004; Kuhl & Meltzoff, 1982), or for temporal synchrony between speech and lip movements (Dodd, 1979). Brief phonemes with subtle facial differences might need more time to be mapped, as suggested by the strengthening of the McGurk effect after the age of 4 months (Kushnerenko, Teinonen, Volein & Csibra, 2008; Patterson & Werker, 2003; Rosenblum, Schmuckler & Johnson, 1997). Here the varied vowel context in the auditory sounds, given the poor temporal audio-visual integration capacities at this age (Lewkowicz, 2003), might have complicated the infants' task when visual information was presented and weakened the phonetic representation formed during the context part of the M_mov trials.

Through trial and error, infants might progressively and actively try to match their own productions with the stored auditory-visual template, and learn the motor gesture for a correct match (Kuhl & Meltzoff, 1996). This learning might be favored by the rapid maturation of the dorsal linguistic pathway which catches up with the ventral pathway at around 4 months (Leroy, Glasel, Dubois, Hertz-Pannier, Thirion *et al*., 2011). Yeung and Werker (2013) showed that chewing and sucking interact with audio-visual perception of vowels in 4.5-month-old infants, suggesting a common sensory-motor representation for the vowels /a/ and /i/. Inferior frontal activation – observed, for example, with EEG when 2-month-old infants matched a previously seen mouth movement with an auditory vowel (Bristow *et al*., 2009), and reported with MEG when 7- and 12-month-old infants listened to difficult speech contrasts (Kuhl, Ramírez, Bosseler, Lin & Imada, 2014) – might represent the signature of the progressive motor involvement in speech perception. Further studies are needed to specify whether the frontal region is automatically activated by phoneme perception or only recruited when listening conditions are difficult, or when imitation is triggered.

Conclusion

Infants are able to learn their native language relatively fast, probably due to a favorable organization of the human brain. Our goal is thus to describe the neural machinery facilitating language acquisition. A phonetic module allowing rapid access to the elementary building block of speech, the phoneme, despite acoustical variations, might be one of the important features of this cerebral apparatus.

Acknowledgements

This study was supported by the Ecole des Neurosciences Paris Ile-de-France (DIM Cerveau & Pensée), the Fondation de France and the McDonnell foundation.

References

Basirat, A., Dehaene, S., & Dehaene-Lambertz, G. (2014). A hierarchy of cortical responses to sequence violations in three-month-old infants. *Cognition*, **132** (2), 137–150.

Bertoncini, J., Bijeljac-Babic, R., Jusczyk, P.W., Kennedy, L., & Mehler, J. (1988). An investigation of young infants' perceptual representations of speech sounds. *Journal of Experimental Psychology: General*, **117**, 21–33.

Bristow, D., Dehaene-Lambertz, G., Mattout, J., Soares, C., Gliga, T. *et al.* (2009). Hearing faces: how the infant brain matches the face it sees with the speech it hears. *Journal of Cognitive Neuroscience*, **21** (5), 905–921.

Caplan, D., Gow, D., & Makris, N. (1995). Analysis of lesions by MRI in stroke patients with acoustic-phonetic processing deficits. *Neurology*, **45** (2), 293–298.

Celsis, P., Boulanouar, K., Doyon, B., Ranjeva, J.P., Berry, I. *et al.* (1999). Differential fMRI responses in the left posterior superior temporal gyrus and left supramargynal gyrus to habituation and change detection in syllables and tones. *NeuroImage*, **9**, 135–144.

Chang, E.F., Rieger, J.W., Johnson, K., Berger, M.S., Barbaro, N.M. *et al.* (2010). Categorical speech representation in human superior temporal gyrus. *Nature Neuroscience*, **13** (11), 1428–1432.

Chen, X., Striano, T., & Rakoczy, H. (2004). Auditory-oral matching behavior in newborns. *Developmental Science*, **7** (1), 42–47.

Csibra, G., Kushnerenko, E., & Grossmann, T. (2008). Electrophysiological methods in studying infant cognitive development. In C.A. Nelson & M. Luciana (Eds.), *Handbook of*

© 2015 John Wiley & Sons Ltd

developmental cognitive neuroscience (2nd edn., pp. 247–262). Cambridge, MA: MIT Press.

D'Ausilio, A., Pulvermuller, F., Salmas, P., Bufalari, I., Begliomini, C. *et al.* (2009). The motor somatotopy of speech perception. *Current Biology*, **19** (5), 381–385.

Dehaene-Lambertz, G., & Baillet, S. (1998). A phonological representation in the infant brain. *NeuroReport*, **9** (8), 1885–1888.

Dehaene-Lambertz, G., & Dehaene, S. (1994). Speed and cerebral correlates of syllable discrimination in infants. *Nature*, **370** (6487), 292–295.

Dehaene-Lambertz, G., & Gliga, T. (2004). Common neural basis for phoneme processing in infants and adults. *Journal of Cognitive Neuroscience*, **16** (8), 1375–1387.

Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., Meriaux, S., Roche, A. *et al.* (2006). Functional organization of perisylvian activation during presentation of sentences in preverbal infants. *Proceedings of the National Academy of Sciences, USA*, **103** (38), 14240–14245.

Dehaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A. *et al.* (2005). Neural correlates of switching from auditory to speech perception. *NeuroImage*, **24** (1), 21–33.

Dehaene-Lambertz, G., & Pena, M. (2001). Electrophysiological evidence for automatic phonetic processing in neonates. *NeuroReport*, **12** (14), 3155–3158.

Dien, J. (1998). Issues in the application of the average reference: review, critiques, and recommendations. *Behavior Research Methods, Instruments, & Computers*, **30** (1), 34–43.

Dodd, B. (1979). Lip reading in infants: attention to speech presented in- and out-of-synchrony. *Cognitive Psychology*, **11** (4), 478–484.

Eimas, P.D., Siqueland, E.R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, **171** (3968), 303–306.

Friederici, A.D., Friedrich, M., & Weber, C. (2002). Neural manifestation of cognitive and precognitive mismatch detection in early infancy. *NeuroReport*, **13** (10), 1251–1254.

Giard, M.H., Lavikahen, J., Reinikainen, K., Perrin, F., Bertrand, O. *et al.* (1995). Separate representation of stimulus frequency, intensity, and duration in auditory sensory memory: an event-related potential and dipole-model analysis. *Journal of Cognitive Neuroscience*, **7** (2), 133–143.

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, **4** (4), 131–138.

Huotilainen, M., Kujala, A., Hotakainen, M., Shestakova, A., Kushnerenko, E. *et al.* (2003). Auditory magnetic responses of healthy newborns. *NeuroReport*, **14** (14), 1871–1875.

Jacquemot, C., Pallier, C., LeBihan, D., Dehaene, S., & Dupoux, E. (2003). Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience*, **23** (29), 9541–9546.

Jusczyk, P.W., & Derrah, C. (1987). Representation of speech sounds by young infants. *Developmental Psychology*, **23**, 648–654.

Jusczyk, P.W., Pisoni, D.B., & Mullennix, J. (1992). Some consequences of stimulus variability on speech processing by 2-month-old infants. *Cognition*, **43** (3), 253–291.

Kluender, K.R., Diehl, R.L., & Killeen, P.R. (1987). Japanese quail can learn phonetic categories. *Science*, **237** (4819), 1195–1197.

Kohler, E., Keysers, C., Umilta, M.A., Fogassi, L., Gallese, V. *et al.* (2002). Hearing sounds, understanding actions: action representation in mirror neurons. *Science*, **297** (5582), 846–848.

Kraus, N., McGee, T.J., Carrel, T.D., Zecker, S.G., Nicol, T.G. *et al.* (1996). Auditory neurophysiologic responses and discrimination deficits in children with learning problems. *Science*, **273**, 971–973.

Kuhl, P.K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience*, **5** (11), 831–843.

Kuhl, P.K., & Meltzoff, A.N. (1982). The bimodal perception of speech in infancy. *Science*, **218** (4577), 1138–1141.

Kuhl, P.K., & Meltzoff, A.N. (1996). Infant vocalizations in response to speech: vocal imitation and developmental change. *Journal of the Acoustical Society of America*, **100** (4 Pt 1), 2425–2438.

Kuhl, P.K., & Miller, J.D. (1982). Discrimination of auditory target dimensions in the presence or absence of variation in a second dimension by infants. *Perception & Psychophysics*, **31** (3), 279–292.

Kuhl, P.K., Ramírez, R.R., Bosseler, A., Lin, J.F., & Imada, T. (2014). Infants' brain responses to speech suggest analysis by synthesis. *Proceedings of the National Academy of Sciences, USA*, **111** (31), 11238–11245.

Kushnerenko, E., Ceponiene, R., Balan, P., Fellman, V., Huotilaine, M. *et al.* (2002). Maturation of the auditory event-related potentials during the first year of life. *NeuroReport*, **13** (1), 47–51.

Kushnerenko, E., Cheour, M., Ceponiene, R., Fellman, V., Renlund, M. *et al.* (2001). Central auditory processing of durational changes in complex speech patterns by newborns: an event-related brain potential study. *Developmental Neuropsychology*, **19** (1), 83–97.

Kushnerenko, E., Teinonen, T., Volein, A., & Csibra, G. (2008). Electrophysiological evidence of illusory audiovisual speech percept in human infants. *Proceedings of the National Academy of Sciences, USA*, **105** (32), 11442–11445.

Kushnerenko, E., Winkler, I., Horvath, J., Naatanen, R., Pavlov, I. *et al.* (2007). Processing acoustic change and novelty in newborn infants. *European Journal of Neuroscience*, **26** (1), 265–274.

Leroy, F., Glasel, H., Dubois, J., Hertz-Pannier, L., Thirion, B. *et al.* (2011). Early maturation of the linguistic dorsal pathway in human infants. *Journal of Neuroscience*, **31** (4), 1500–1506.

Lewkowicz, D.J. (2003). Learning and discrimination of audiovisual events in human infants: the hierarchical relation between intersensory temporal synchrony and rhythmic pattern cues. *Developmental Psychology*, **39** (5), 795–804.

Liberman, A.M. (1996). *Speech: A special code*. Cambridge, MA: Bradford Books/MIT Press.

© 2015 John Wiley & Sons Ltd

Liberman, A.M., Delattre, P., & Cooper, F.S. (1952). The role of selected stimulus-variables in the perception of the unvoiced stop consonants. *American Journal of Psychology*, **65** (4), 497–516.

Lotto, A.J., Hickok, G., & Holt, L.L. (2009). Reflections on mirror neurons and speech perception. *Trends in Cognitive Sciences*, **13** (13), 110–114.

Mahmoudzadeh, M., Dehaene-Lambertz, G., Fournier, M., Kongolo, G., Goudjil, S. *et al.* (2013). Syllabic discrimination in premature human infants prior to complete formation of cortical layers. *Proceedings of the National Academy of Sciences, USA*, **110** (12), 4846–4851.

Massaro, D.W., Cohen, M.M., & Gesi, A.T. (1993). Long-term training, transfer, and retention in learning to lipread. *Perception & Psychophysics*, **53** (5), 549–562.

Meltzoff, A.N., & Borton, R.W. (1979). Intermodal matching by human neonates. *Nature*, **282** (5737), 403–404.

Meltzoff, A.N., & Moore, M.K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, **198** (4312), 74–78.

Meltzoff, A.N., & Moore, M.K. (1989). Imitation in newborn infants: exploring the range of gestures imitated and the underlying mechanisms. *Developmental Psychology*, **25**, 954–962.

Montgomery, A.A., & Jackson, P.L. (1983). Physical characteristics of the lips underlying vowel lipreading performance. *Journal of the Acoustical Society of America*, **73** (6), 2134–2144.

Näätänen, R., & Tiitinen, H. (1998). Auditory information processing as indexed by the mismatch negativity. In M. Sabourin, F. Craik & M. Robert (Eds.), *Advances in psychological science* (Vol. 2, Biological and cognitive aspects) (pp. 145–170). New York: Psychology Press.

Patterson, M.L., & Werker, J.F. (2003). Infants match phonetic information in lips and voice. *Developmental Science*, **6**, 191–196.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, **27**, 169–192.

Romanski, L.M., & Goldman-Rakic, P.S. (2002). An auditory domain in primate prefrontal cortex. *Nature Neuroscience*, **5** (1), 15–16.

Rosenblum, L.D., Schmuckler, M.A., & Johnson, J.A. (1997). The McGurk effect in infants. *Perception & Psychophysics*, **59** (3), 347–357.

Sinnott, J.M., & Gilmore, C.S. (2004). Perception of place-of-articulation information in natural speech by monkeys versus humans. *Perception & Psychophysics*, **66** (8), 1341–1350.

Turken, A.U., & Dronkers, N.F. (2011). The neural architecture of the language comprehension network: converging evidence from lesion and connectivity analyses. *Frontiers in Systems Neuroscience*, **5**, 1.

Yeung, H., & Werker, J.F. (2013). Lip movements affect infants' audiovisual speech perception. *Psychological Science*, **24** (5), 603–612.

Received: 1 December 2013
Accepted: 2 April 2015

© 2015 John Wiley & Sons Ltd