



Cortical oscillations as temporal reference frames for perception

Anne Kosem

► **To cite this version:**

Anne Kosem. Cortical oscillations as temporal reference frames for perception. Agricultural sciences. Université Pierre et Marie Curie - Paris VI, 2014. English. <NNT : 2014PA066074>. <tel-01069219>

HAL Id: tel-01069219

<https://tel.archives-ouvertes.fr/tel-01069219>

Submitted on 29 Sep 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THESE DE DOCTORAT DE
L'UNIVERSITE PIERRE ET MARIE CURIE

Spécialité
Neurosciences Cognitives

Présentée par

Anne KÖSEM

**CORTICAL OSCILLATIONS AS TEMPORAL
REFERENCE FRAMES FOR PERCEPTION**

Directeur de recherche: Virginie VAN WASSENHOVE

Thèse soutenue le 27 mars 2014

Devant la commission d'examen formée de :

Dr. Rufin VAN RULLEN
Pr. Ole JENSEN
Pr. David POEPPPEL
Dr. Catherine TALLON-BAUDRY
Dr. Florian WASZAK
Dr. Nathalie GEORGE
Dr. Virginie VAN WASSENHOVE

Rapporteur
Rapporteur
Examineur
Examinatrice
Examineur
Examinatrice
Examinatrice

ABSTRACT

The timing of sensory events is a crucial perceptual feature, which affects both explicit judgments of time (e.g. duration, temporal order) and implicit temporal perception (e.g. movement, speech). Yet, while the relative external timing between events is commonly evaluated with a clock in physics, the brain does not have access to this external reference. In this dissertation, we tested the hypothesis that the brain should recover the temporal information of the environment from its own dynamics. Using magnetoencephalography (MEG) combined with psychophysics, the experimental work suggests the involvement of cortical oscillations in the encoding of timing for perception. In the first part of this dissertation, we established that the phase of low-frequency cortical oscillations could encode the explicit timing of events in the context of entrainment, i.e. if neural activity follows the temporal regularities of the stimulation. The implications of brain oscillations for the encoding of timing in the absence of external temporal regularities were investigated in a second experiment. Results from a third experiment suggest that entrainment does only influence audiovisual temporal processing when bound to low-frequency dynamics in the delta range (1-2 Hz). In the last part of the dissertation, we tested whether oscillations in sensory cortex could also 'tag' the timing of acoustical features for speech perception. Overall, this thesis provides evidence that the brain is able to tune its timing to match the temporal structure of the environment, and that such tuning may be crucial to build up internal temporal reference frames for explicit and implicit timing perception.

RÉSUMÉ

La perception explicite du temps écoulé (la durée, l'ordre temporel...) et les jugements implicites des dynamiques de notre environnement (percevoir le mouvement, la parole...) nécessitent l'extraction des relations temporelles entre événements sensoriels. Alors que le temps physique est communément évalué en rapport à un référentiel externe (celui de l'horloge), le cerveau lui n'a pas accès à ce référentiel. Dans cette thèse, nous émettons l'hypothèse que le cerveau génère son propre référentiel temporel à partir des dynamiques neurales. Combinant la magnétoencéphalographie (MEG) aux données psychophysiques, les présents travaux suggèrent que les oscillations corticales sont impliquées dans l'encodage du temps perçu. Une première étude montre que la phase des oscillations corticales basse-fréquences peut encoder l'ordre temporel perçu entre événements sensoriels s'il y a entrainement neural, i.e. si l'activité cérébrale suit les régularités temporelles de la stimulation. L'implication des oscillations cérébrales en l'absence d'entrainement est testée dans une seconde expérience. Les résultats d'une troisième expérience suggèrent que l'entrainement neural n'a d'influence sur le traitement temporel des informations multisensorielles qu'à basse fréquence (1-2 Hz). Un dernier chapitre aborde le rôle l'entrainement neural dans l'encodage la temporalité des informations acoustiques pour la perception de la parole. En conclusion, cette thèse suggère que le cerveau est capable de suivre la structure temporelle du monde extérieur, et que cet ajustement permet la construction d'un référentiel temporel interne pour la perception explicite et implicite du temps.

ACKNOWLEDGMENTS

First and foremost, I am greatly indebted to my thesis supervisor Virginie van Wassenhove. I thank her for giving me the opportunity to work at Neurospin and for providing me with the most favorable environments to perform research. Virginie generously shared with me a lot of her expertise and research insight, for which I am deeply grateful. I thank her for her trust, enthusiasm, and continuous support during my PhD.

The work described in this thesis was accomplished with the help and support of fellow lab-mates and collaborators. I am very grateful to Alexandre Gramfort who helped me discover the wonders of MNE-python; Christophe Pallier for his help on statistical analysis; Anahita Basirat and Leila Azizi who assisted me in designing and acquiring the data of the so-called “pata-pata” experiment; Laetitia Grabot and Karin Pestke who acquired and analyzed data of the PhaseTime experiment. I would like to thank the MEG team, in particular Marco Buatti and Leila Azizi, for their assistance. I thank the Neurospin medical staff for the recruitment of participants. I also thank the DGA and the CEA for their financial support granted through doctoral fellowship.

I thank all the Neurospin fellows whom I had the opportunity to meet. A special thanks goes to Lucie, for being such a great roommate and a wonderful collaborator, which led to numerous publications in Brain Magazine. Baptiste and Catherine, thank you for the extended coffee-break discussions about cognitive science, gender theory and the quesadillas effect. I am indebted to Nawal, who gave me a ‘yeah-baby’ boost when we worked together at the library. My thanks to Nicolas for the good old karaoke times in the open space. I also thank Anna, Brice, Adeline, Clémence, Diana, Lucille, Thiago, Filipa, Clio, Elodie, Murielle, Marie, Mélanie, Jean-Rémi, Simon, Laurianne, Sebastien, Moti, Ioannis, Toussaint, Aurélien, Antonio, Vanna, Antoinette, and Laurence for all the agreeable moments we spent together at Neurospin and beyond the “plateau de Saclay”.

I thank my peers and friends from the Cogmaster: Jean-Rémy the meditative philosopher, Guillaume the anarchist anthropologist, Camillia the NIRS starlette, Cédric the game

maker, Clio the double agent, Louise the demoiselle inconnue, Eléonore the sailor moon, and DJs Romain, Hadrien, & Victor.

Finally, I would like to thank my family and friends, who have encouraged me and patiently listened to my nerdy reflections through all these years. First, I warmly thank my parents for their constant love, care, and support; and H  l  ne for being the best sister in the world. I also thank my sister of heart, Laure, and Ced, for all the joyful moments and their cheerful encouragements. Thomas, no matter hard I try, you will always be the nerdiest (and simply the best). Thank you for your unyielding support, even being miles away. I thank Becky for her evergreen enthusiasm and her correction of this manuscript. Mathieu and Laure, I do not know if we completed our mission, but life was definitely sweeter by your side. Thank you for these great moments. My big thanks also go to Claire, Karine, Romain, Sophie, Rym, Diane, Etienne and all the people who have accompanied me during this journey.

CONTENTS

Chapter 1: Introduction.....	3
1.1. The inner representation of time	3
1.1.1. What is a time experience?.....	3
1.1.2. How does the time experimenter sense time?.....	5
1.1.3. How is event timing encoded?.....	6
1.2. Encoding event timing through neural latency.....	9
1.2.1. Brain access to external timing.....	9
1.2.2. Temporal binding.....	11
1.2.3. Temporal plasticity.....	17
1.3. Encoding event timing with dedicated neural structures.....	20
1.3.1. Specialized brain areas for the encoding of time	20
1.3.2. Delay-tuned cells.....	21
1.3.3. Brain clocks	22
1.4. Encoding event timing with brain oscillations	25
1.4.1. Temporal binding through neural coherence.....	26
1.4.2. Oscillations affect perceptual temporal sampling.....	29
1.4.3. Neural phase codes of events succession.....	33
1.4.4. Phase tagging of perceived event timing?	36
Chapter 2: Phase encoding of explicit timing.....	39
2.1. Introduction	39
2.1.1. Motivation.....	39
2.1.2. Experiment	40
2.1.3. Summary of the results.....	40
2.2. Article.....	41
Anne Kösem, Alexandre Gramfort, Virginie van Wassenhove (accepted). Encoding of event timing in the phase of neural oscillations. Neuroimage.....	41
2.3. Phase coding without entrainment.....	79
2.3.1. Motivation.....	79
2.3.2. Experiment	79
2.3.3. Results.....	83
2.3.4. Discussion	86
Chapter 3: Low resolution of multisensory temporal binding	89
3.1. Introduction	89
3.1.1. Motivation.....	89
3.1.2. Experiment	90
3.1.3. Summary of the results	90
3.2. Article.....	91

Kösem, A., & van Wassenhove, V. (2012). Temporal Structure in Audiovisual Sensory Selection. PloS one, 7(7), e40936.	91
Chapter 4: Phase encoding of implicit timing	123
4.1. Introduction	123
4.1.1. Motivation.....	123
4.1.2. Experiment	125
4.2. Article.....	127
Oscillatory mechanisms of ambiguous speech parsing (in preparation).....	127
Chapter 5: General discussion	159
5.1. Summary of the findings	159
5.2. Two roles for the phase of slow-oscillations in EEG/MEG recordings?	160
5.2.1. Absolute and relative phase	160
5.2.2. Temporal integration and segregation.....	164
5.3. Multi- and ar- rhythmic brain mechanisms of event timing.....	166
5.3.1. Multiplexing of temporal information with multiple brain oscillators	166
5.3.2. Entrainment vs. arrhythmic stimulation	168
5.4. Conclusion	170
References	171

INTRODUCTION

1.1. THE INNER REPRESENTATION OF TIME

Le temps [...] qui pourra le définir ? Et pourquoi l'entreprendre, puisque tous les hommes conçoivent ce qu'on veut dire en parlant de temps, sans qu'on le désigne davantage.

Blaise Pascal, Pensées

I do not define time, space, place and motion, as being well known to all.

Isaac Newton, the Scholium to the Principia

1.1.1. *What is a time experience?*

Time is a familiar concept to every individual. As pointed out by Saint Augustine, despite our understanding of the concept of time, it is hard to define.

What, then, is time? I know well enough what it is, provided that nobody asks me; but if I am asked what it is and try to explain, I am baffled.

To Saint Augustine time is something we know because we constantly experience its effects on the world: time is present when we work, when we twist and when we shout. It is intuitively obvious why time makes “No man ever steps in the same river twice” and its existence cannot be questioned. But when trying to capture its very essence, time suddenly becomes something “secret, lacking in substance and yet almighty” (Thomas Mann, *The Magic Mountain* 1924).

Blaise Pascal wrote in his *Pensées* that the difficulty to define time may relate to its polysemy. Time can either refer to a continuum where things evolve, a flow or movement of things within this continuum, or a measure of the flow and changes of the world. While the two first definitions refer to time as outside experience, the last definition of time will be of interest for the observer of the passage of time or “time experiencer” (Michon, 1985). Indeed time as a measure is the only time we may have access to, while the time outside experience is inaccessible and may not even exist (Augustine; Pöppel, 1997). But what exactly is the time experiencer measuring?

In all existing definitions, time is tied to the succession of physical events. If time is first defined as “a nonspatial continuum” as in The Merriam-Webster dictionary, it has the property to be “measured in terms of events which succeed one another from past through present to future”. If time is rather defined as flow or as “the indefinite continued progress” in the Oxford dictionary, then this progress will concern the “existence and events in the past, present, and future”. If time is referred to as a quantity (“an amount of time”) it is “reckoned by a conventional standard” and is “measured as seconds, hours, and days”. Here again, the dependency of time to events succession is still apparent, as one second is defined as the succession of 9,192,631,770 cesium atom state changes.

Hence time is only observable with respect to its effects on the outside world, and thus experiencing time is only possible by quantifying the change and succession of external events. It entails that a time experience should not exist in the absence of external temporal information. This description appears at odds with our intuition: if we close our eyes, if we isolate from sensory stimulation, we are still able to feel that time is passing by. But as pointed out by William James (James, 1886):

Our heart-beats, our breathing, the pulses of our attention, fragments of words or sentences that pass through our imagination, are what people this dim habitat. [...] In short, empty our minds as we may, some form of changing process remains for us to feel, and cannot be expelled. And along with the sense of the process and its rhythm goes the sense of the length of time it lasts. Awareness of change is thus the condition on which our perception of time's flow depends; but there exists no reason to suppose that empty time's own changes are sufficient for the awareness of change to be aroused. The change must be of some concrete sort -- an outward or inward sensible series, or a process of attention or volition.

Thus, external temporal information is always available for the time experiencer, either through the dynamics of sensory inputs, body rhythms, or mental states. But what exactly is the content of time experience, and how does the time experiencer extract this information for the establishment of perceptual time?

1.1.2. How does the time experiencer sense time?

For humans and other animals, perceiving the temporal relationship between events is essential for their daily behavior (Pöppel, 1997; van Wassenhove, 2009; Wittmann, 2009). Time perception encompasses a multitude of phenomena that can be categorized into implicit and explicit timing types (Michon, 1990). Explicit timing is defined as the conscious representation of the temporal relations between percepts and/or actions. It includes among others the perception of event synchrony, temporal order, or duration. For instance, it occurs while watching a soccer game and estimating the time left before the end of the match. Conversely, implicit timing refers to the dynamics of perception and behavior that are not consciously conceptualized as temporal properties. It includes motion perception, temporal content segmentation, anticipation, and predictability. It is also present while watching a soccer game as we perceive within the dynamics changes of lights the movements of the soccer team and within the fluctuations of the acoustic signal the exclamation of the sport commentator.

Both explicit and implicit timing percepts seem to rely on the accurate extraction of the external dynamics in the world. Yet, Gibson, in his talk entitled ‘Events are perceivable, but time is not’, suggests to the contrary that the time that we perceive has no physical essence. As a parallel to depth perception that does not rely on the physical information in the external world but on the mental reconstruction of perceived objects surfaces, the representation of time is nothing but a mental reconstruction built on the perceived succession of events. As such he concluded that “there is no such thing as the perception of time, but only the perception of events and locomotion” (Gibson, 1975). Hence time perception should not be confounded with perception of time, as time cannot be perceived *per se*. Time experience should not be considered as a perception of external changes, but rather as a change of the perceived external events. This view of time

perception was supported by philosophers Dennett and Moller and synthesized by Roache (Roache, 1999):

Therefore it seems that, at bottom, our representation of the temporal order of two events is nothing more or less than the temporal order of, and the causal relations between, our experiences of those events.

A bigger claim is actually drawn from the view of time as “the ghost of events in the world”. Like other features of perception, subjective event timing is a construct of the observing brain. But what is highlighted here is that unlike color, pitch or any feature of sensation, time is not a proper “sense” per se. That is, unlike most features, there seems to be no apparent devoted sensor and no dedicated sensory brain area for the encoding of time (Ivry and Schlerf, 2008; van Wassenhove, 2009; Wittmann, 2009, 2013). In other words, the time experiencer cannot directly extract the temporal information in the world. The time experiencer first processes external sensory inputs, from which external event timing is estimated, here defined as a particular time point of occurrence of the event. Hence, the perception of event timing is not based on external stimulation timing, but on the timing of mentally-reconstructed sensory events (Pöppel, 1997; van de Grind, 2002). It entails that time should somehow be extracted through the neural processing of sensory inputs.

1.1.3. How is event timing encoded?

For the time experiencer, the perception of time passing by relies on our capacity to extract the successive changes of the environment. Hence, time perception will obviously rely on the temporal information provided by the external world, but it will also be dependent on our ability to extract the temporal information through the processing of the other features of sensation. Of interest here, the time experiencer is not only observing but also undergoing the passage of time. The functioning of his body, like the external world, follows temporal rules.

From an information-theoretic standpoint, the time experiencer, in order to reconstruct the time passing by, relies on sensors and on information systems that have constraints in

time. At the beginning of information processing, sensors present time constants that constrain the temporal resolution of perceived events at different time scales across sensory modalities. For instance, temporal acuity in human vision is already constrained in time at the very first stage of the light processing in the retina: it takes between 35 ms and 75 ms for the pigments that capture light in the photoreceptors to disintegrate and form again (Tyler, 1985). In contrast within mammalian audition, the transduction of the acoustic signal by the hair cells in the cochlea operates in less than 1 millisecond time scale. As a result, the auditory system can capture much finer variations of its input signal than the visual system (Russell and Sellick, 1978).

The funnel effect in the temporal resolution of sensory processing increases as information travels along the brain. At the neuron level, the integration of presynaptic spikes is constrained with the membrane time constant that is for the principal cells in the waking cerebral cortex of about 10-30 ms (Koch et al., 1996; Destexhe et al., 2003). After spiking, neurons can undergo refractory periods from tens to hundreds of milliseconds (Mickus et al., 1999; Henze and Buzsáki, 2001). Between neurons, synaptic transmission can operate at the speed of tens to hundreds of milliseconds (Monyer et al., 1992), and even at the second timescale (Deisz and Prince, 1989; Nakanishi, 1994). In addition to a decrease in temporal resolution, the presence of synapses and the axonal neural transmission time constants can introduce information transmission delays. A consequence is that sensory information reaches the cortex at different times depending on the sensory modality: auditory information reaches the primary auditory cortex as soon as 10 ms after sound emission (Celesia, 1976; Heil and Irvine, 1997), while it takes around 50-70 ms for visual information to reach primary visual cortex (Schmolesky et al., 1998; Zeki, 2001; Raij et al., 2010).

Therefore, a fundamental question is how human and animals extract the temporal dimension with both external and internal temporal contingencies for perception (Michon, 1985). Events timing perception is dependent on the body sensors contingencies and neural processing delays, and it is likely that the time we have access to, i.e. our experience of time, is an inexact mental construct of reality (Michon, 1985).

Consistently, perceived timing departs from real timing. One possible consequence of sensors and neural integrative mechanisms of signal processing is that subtle signal temporal changes are not perceived. Temporal integration constrains are particularly

apparent when comparing timing perception in vision and audition, as we capture much finer temporal variations in the auditory domain than in the visual domain: acoustic changes of 1-2 ms can be detected (Plomp, 1964; Green, 1971), while visual events need to be separated in time by about 25 ms to be segregated (Hirsh and Sherrick Jr., 1961). Crucially thresholds of temporal segregation increase with stimulus complexity, suggesting that neural processing affects temporal acuity: for speech or musical stimuli temporal segregation occurs for temporal distances of 25 ms to 250 ms (Green, 1971; Poeppel, 2003).

A second consequence is that even if sensory events are perceived as distinct events, their temporal information may be distorted. In particular, human perception is tolerant to asynchrony: two visual elements may be perceived as flashed synchronously while they are actually delayed in time (Fahle and Koch, 1995; Verstraten et al., 2000; Blake and Lee, 2005; Chakravarthi and Cavanagh, 2007). In the same manner delayed auditory and visual information may be perceived synchronous (Fujisaki and Nishida, 2005, 2010; Zampini et al., 2005; van Wassenhove et al., 2007; Benjamins et al., 2008; Vroomen and Keetels, 2010; Keetels and Vroomen, 2012).

Third, if timing perception is indeed constrained by body, sensor, and brain structures, then different bodies should lead to different timing percepts: animals with different body architectures do not perceive time the same way (Healy et al., 2013). In an evolutionary perspective, animals with more demanding temporal contingencies (for instance animals making faster movements) have faster sensors (Laughlin and Weckström, 1993). Conversely, if animals experience small variations in time experience, they can also encounter similar temporal constrains. Differences in biological timing between animals that face similar temporal contingencies should not be too big. Consistently neural dynamics are quite robust and show similar temporal scales among mammals (Buzsáki et al., 2013).

One can wonder to what extent sensors and brain temporal contingencies explain the distortions from physical to mental timing. The reported distortions in perceived timing could just be a consequence of forward neural timing delays that are introduced during visual, auditory and other sensory processing. As such neural latencies have been suggested to be the missing link between external and subjective timing, and constitutes

the simplest coding scheme for event time. Yet, extracting event timing from neural timing must pose quite a few challenges that are described in greater details in the next section.

As an alternative to the neural latency code, dedicated networks could take part in the encoding of time. Indeed, even if time perception does not rely on a specific sensory neural architecture, it is still possible that specialized networks read sensory activity in the sole purpose to extract temporal information. Dedicated models of event timing that are addressed in section 3 of the introduction.

Dedicated models emphasize the role of specific brain regions for the decisional aspect of time perception. However, these models rely on temporal information retrieved from the dynamics of brain activity. As such, the brain capitalizes on dynamics to construct an internal representation of time. Crucially, any modulation of this temporal reference should lead in distortions of event timing encoding. In section 4, the role of brain dynamics _ with a focus on brain oscillations _ on perception in time are reviewed. We will further argue in this thesis that cortical oscillations provide the temporal reference frame on which perceptual timing relies.

1.2. ENCODING EVENT TIMING THROUGH NEURAL LATENCY

1.2.1. Brain access to external timing

In the external world, time is commonly evaluated with a clock. Yet the brain has no access to the external clock reference and in general it has no insight into the timing of the external stimulation. Thus, the first challenge for the brain is to detect new sensory onsets from its internal dynamics (Heil, 2004; VanRullen et al., 2005a; Brasselet et al., 2012; Scharnowski et al., 2013). After the emission of sensory input, a specific combination of neural networks is recruited to process the information and emit spike trains of activity; hence a simple solution to detect the onsets of new events would be to detect when the recruited network starts firing (Heil and Irvine, 1997; VanRullen et al., 2005a).

However, this coding scheme for event timing raises many issues. First, detecting stimulus-related neural activity is not trivial. Not only does the brain lack the knowledge of stimulation onsets, but sensory networks are rarely silent before stimulation emission. To the contrary, network activity in the absence of sensory events is undergoing endogenous fluctuations. For instance at the earliest stages of visual processing, retinal ganglions cells show spontaneous activity, i.e. they fire actions potentials in the total absence of light (Bruce et al., 1996; Rodiek, 1998). This apparent ongoing neural activity is present at all stages of visual and other sensory processing (Gerstein and Kiang, 1960; Laufs et al., 2003; Fox and Raichle, 2007). Therefore, the brain must first dissociate endogenous fluctuations and event-related responses from its overall network activity. This dissociation could be done by contrasting ongoing activity in a network to its “default” or “baseline” state (Kuffler, 1953; Gerstein and Kiang, 1960; Coles and Rugg, 1995). Consistently, neuro-physiological and –imaging studies typically observe changes in neural activity by contrasting activity pre- to post-stimulus. Yet, baseline correction is usually done with the knowledge of stimulus timing. Detecting contrasts of activity must be much harder for the brain, knowing that it does not have access to the stimulus external temporal reference. Furthermore, the dissociation between endogenous and stimulus-driven activity is more problematic when processing successive sensory inputs. In natural settings, sensory events are not often isolated in time, and neuronal networks are possibly already recruited at the emission of new stimulation. An important issue for time perception is to link past events to ongoing events in time. The brain needs to find a way to detect incoming events while processing past stimulation in parallel (VanRullen et al., 2005a) and to place new detected onset in their surrounding temporal context.

Second, neural spiking activity following stimulus onset is quite spread in time (Brasselet et al., 2012) and presents some apparent irregular dynamic properties (Wang, 2010). This phenomenon is also observable with neuroimaging techniques, as electric and magnetic event-related neural activity is known to present many peaks that span from tens to hundreds of milliseconds post-stimulation (Coles and Rugg, 1995). As networks are active for a long time period after the emission of the sensory event, it seems hard to define what temporal information in neural activity could encode the perceived onset. A first possibility is that stimulus onset is tagged by the first spike occurring post-stimulus onset (Heil, 2004; Chase and Young, 2007). However, this sparse coding of event timing seems problematic considering that a first spike would be hard to detect if embedded

between endogenous spikes. More problematically, first-spike latency has been shown to be dependent on the content of the stimulus, and not only on its temporal properties (Heil and Irvine, 1997). The same problem extends when looking at following neural activity: despite its apparent irregularity spike timing is specific to the sensory stimulation (Mainen and Sejnowski, 1995), and may actually provide a temporal code for the representation of different sensory attributes (Bair and Koch, 1996; Victor, 2000; Panzeri et al., 2001). If spike timing serves the encoding of both temporal and non-temporal information, one direct consequence is that, for the same temporal onset, neural latencies may differ depending on stimulus properties. For instance, in LGN a high contrast stimulus is processed faster than a low contrast stimulus (Maunsell et al., 1999). V1 cells respond at different latencies according to stimulus orientation (Gawne et al., 1996), and neurons in MT/V5 respond at different latencies depending on motion direction (Raiguel et al., 1999). Similarly, neuroimaging studies have reported changes in the peak latency of Event Related electric Potentials (ERPs) and Events Related magnetic Fields (ERFs) which reflect non-temporal stimulus features (Tepas and Armington, 1962; Roberts et al., 2000). Based on these findings, it appears that absolute neural timing post-stimulus presentation cannot alone provide a reliable index of sensory event timing (Eagleman, 2010).

Additionally, sensory analysis in the brain gets more and more delayed in time compared to stimulus onset due to cumulative neural transmission delays through the processing pathways. Thus, it remains unclear at which stage of the sensory hierarchy neural timing would be a relevant measure for timing perception. While it is argued that neural latency at sensory cortex level encodes event timing (Brasselet et al., 2012), other evidence suggests the existence neural timing codes at subcortical level (Chase and Young, 2007).

1.2.2. Temporal binding

Determining at which stage of the sensory processing the latency of the neural responses matter for the establishment of conscious timing is particularly crucial when comparing different sensory features in time. Indeed, sensory inputs are processed via distinct brain areas (Zeki, 1978; Livingstone and Hubel, 1988; Felleman and Van Essen, 1991) that

receive stimulus information at different latencies (Schmolesky et al., 1998; Zeki, 2001). As presented in the previous section, visual inputs are first processed in the Lateral Geniculate nucleus (LGN), with a peak latency occurring as soon as 30 ms post stimulus onset, followed 10 to 20 ms later by primary visual cortex (V1). Neural latency increases differently based on how the signal bifurcates in the visual hierarchy, so that specialized areas that code for visual contours, motion or color do not received sensory information at the same time (fig. 1.1). How, then, does the brain correctly bind in time synchronous features if they are processed at different latencies?

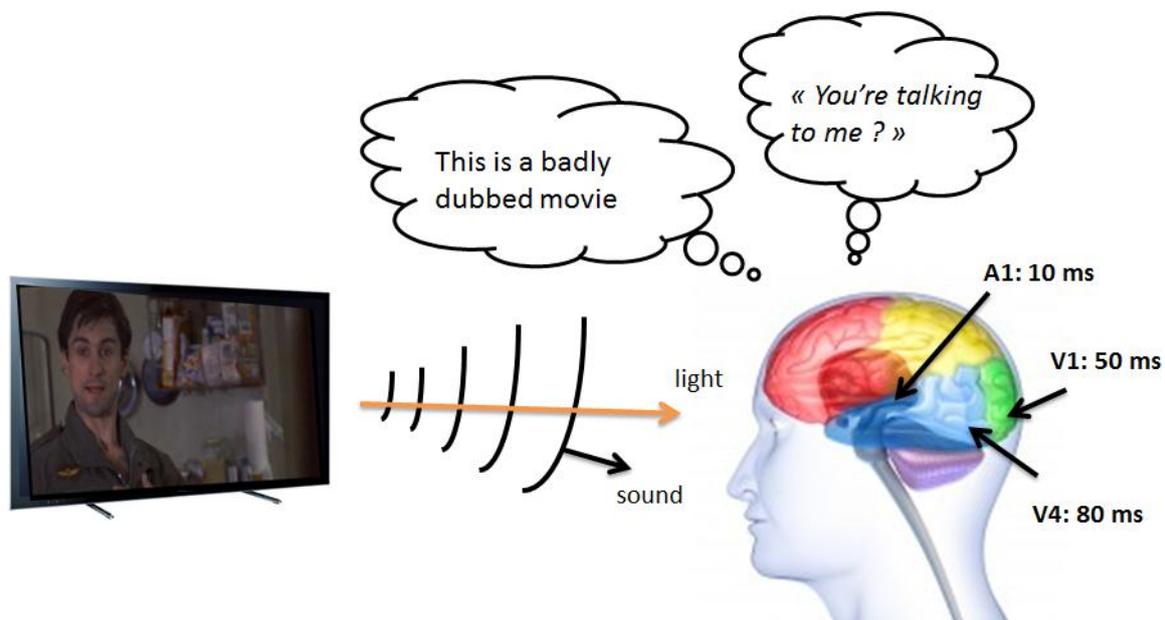


Figure 1.1: Temporal binding problem. How do we perceptually combine in time the different features of an object, if they are processed by distinct brain areas at different instants? And how do neural latencies impact explicit timing perception (e.g. when we perceive asynchrony between lip movements and dubbing when watching a movie) and implicit temporal processing (e.g. when processing speech)?

One view is that awareness of a stimulus feature arises at early stages of sensory processing, such as the awareness of the sensory event, and a fortiori awareness of the sensory event's timing, come as soon as sensory networks that are encoding the feature start firing. In other words, the “differential-latency” hypothesis states that perceived timing between distinct sensory features corresponds to neural timing delays between the two areas encoding each feature. This view suggests that because of neural latencies, the distinct features of a visual object should actually not be perceived as synchronous

(Moutoussis and Zeki, 1997; Purushothaman et al., 1998; Whitney and Murakami, 1998; Zeki and Bartels, 1998; Patel, 2000), and some evidence has been put forward in agreement with this model. For instance, in an illusion first investigated by Moutoussis & Zeki (1997), participants observed a moving square that was changing color and direction (upward or downward) and they were asked to judge the synchrony between changes in motion direction and in color (fig.1.2a).

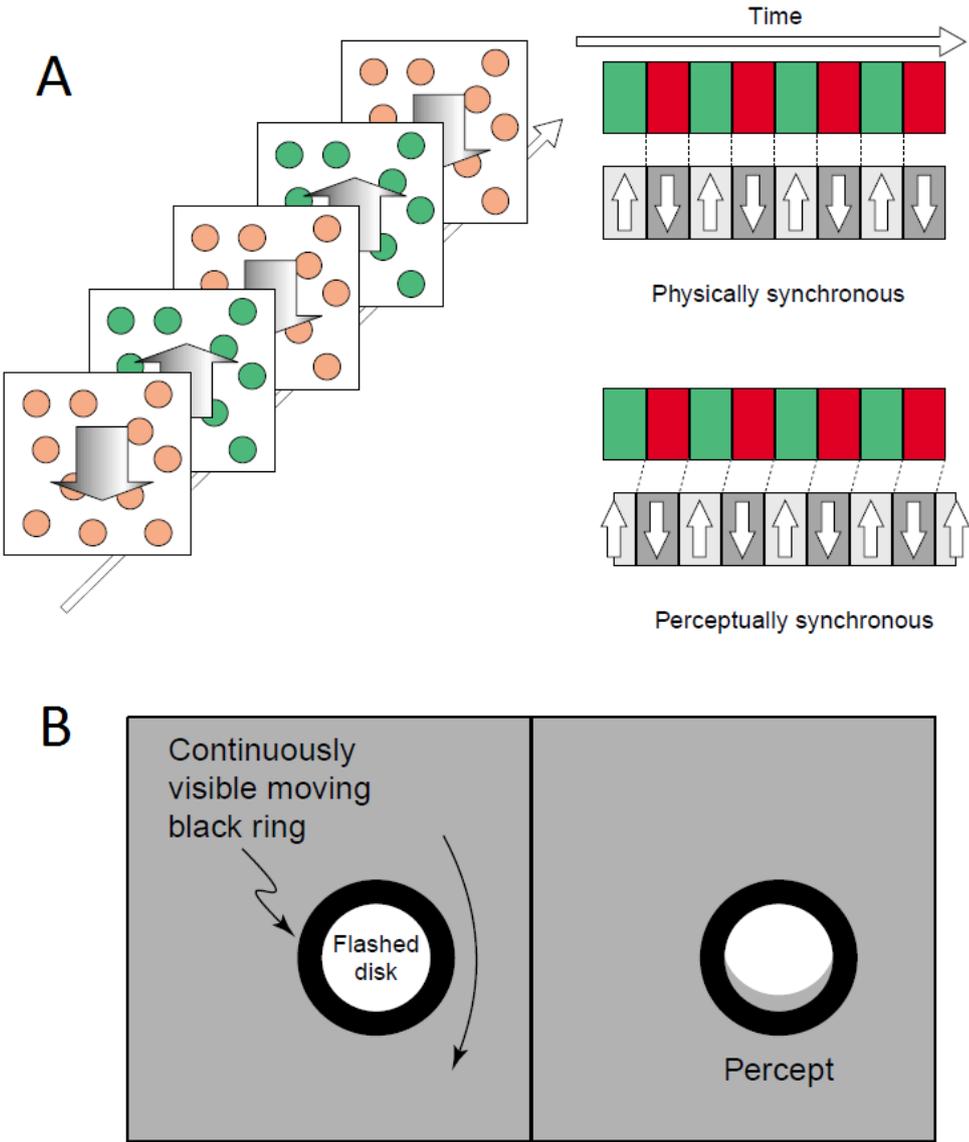


Figure 1.2: Perceptual temporal illusions. (A) Motion color asynchrony. Color dots are displayed on a screen and move upwards or downwards. The perception of synchrony between motion direction changes and color changes do not occur at physical synchrony, but when the change in motion direction occurs approximately 80 ms before the color change. Adapted from (Johnston and Nishida, 2001), (B) Flash-lag effect. A moving visual circle would appear ahead of a static visual disked that is flashed at the same location. Adapted from (Nijhawan, 2002)

Surprisingly perceived synchrony between motion and color features departed from veridical synchrony: color changes had to precede motion direction changes by 70-80 ms to be perceptually matched in time (Moutoussis and Zeki, 1997; Zeki and Bartels, 1998) (fig. 1.2a). The asynchrony of perceptual binding between motion and color remained the same when the timing between the two features was implicitly modulating the perceptual reports (e.g. in a motion aftereffect task) (Arnold et al., 2001). These results thus provided evidence that asynchronous neural processing of color and motion features could have consequences on the perception of event timing. It further suggests that conscious access to feature timing is done at early stages of feature processing (Zeki and Bartels, 1998). Other visual illusions have been interpreted within this differential-latency framework. The flash-lag effect, for example, consists in the perception of asynchrony between a moving and a static visual stimulus even though these stimuli were displayed at the same time (Eagleman and Sejnowski, 2000; Patel, 2000; Nijhawan, 2002; Arnold et al., 2003) (fig. 1.2b), and has also been suggested to rely on differences in neural latencies (Purushothaman et al., 1998; Whitney and Murakami, 1998; Patel, 2000).

Based on these findings, could we conclude that perceptual timing linearly maps onto neural latencies? While the motion-color binding asynchrony suggests independent timing encoding mechanisms for different stimuli features, the reported psychological latencies do not match known neural latencies: the motion area MT/V5 receive inputs around 30 ms after visual onset, while color information is processed 70 ms post stimulus onset in area V4 (Schmolesky et al., 1998; Zeki, 2001). The psychological timing is thus in reverse order compared to neural timing. Moreover, the differential latency hypothesis is counterintuitive and does not match introspection. If different features of an object were perceived at different timings, our perception would be constantly impaired in everyday life. In particular, we would not be able to perceive objects as a whole and would incorrectly interpret the color of a moving car, or the dim and bright surfaces of a same object (Eagleman, 2010). The fact that we are capable of binding correctly the different features of an object in time, and more importantly that temporal congruity is a prerequisite for feature binding (Treisman, 1996; Alais et al., 1998; Blake and Lee, 2005), casts a shadow on the differential latency hypothesis. Clearly a sense of temporal unity is necessary for perceptual binding.

Against the differential latency hypothesis, it was suggested that the brain, to cope with its internal delays, waits for the last processed stimulus to establish event timing (Eagleman, 2010). In other words, if the starting point of feature processing might not explain timing perception maybe the ending point may be more relevant. As evidence for this proposal, most visual temporal illusions present a temporal resolution of 80 ms (Di Lollo, 1980; Eagleman and Sejnowski, 2000; Arnold et al., 2001; Eagleman et al., 2005; Eagleman, 2010), which corresponds to the transmission time of the visual signal from the retina to the slowest visual area V4. This argument has also found support in multisensory temporal binding observations (Kopinska and Harris, 2004). Combining auditory and visual stimuli raises the same problem as combining color and motion features, as auditory and visual areas do not process information at the same latencies: the visual system processes information between 30 ms and 80 ms post stimulus onset (Schmolesky et al., 1998), while A1 receives information only 10 ms post stimulus onset (Celesia, 1976). Like in color-motion feature binding, perceived audiovisual synchrony and temporal order do not occur at pure stimulus synchrony, and the psychophysical delays do not match with neural transmission delays (Slutsky and Recanzone, 2001; Zampini et al., 2005; Van Eijk et al., 2008; Harris et al., 2010; Vroomen and Keetels, 2010). Again, the temporal resolution of audiovisual timing perception is of 80-100 ms, implying that the brain waits for the slow visual processing signal to compute cross-sensory timing.

This proposition does present a few issues. If the brain waits for the last sensory signal to be processed, it presupposes that a specific region in the sensory hierarchy serve as a “temporal buffer”, where perceived synchrony corresponds to neural synchrony. But no clear crossing neural point in visual processing has been found yet (Johnston & Nishida 2001, van de grind 2002). In audiovisual processing, convergent sites in superior colliculus (SC) (Meredith et al., 1987; Stein and Meredith, 1993) and in superior temporal sulcus (STS) (Benevento et al., 1977) may serve the linkage of coincident visual and auditory inputs. However, the coarse temporal resolution of the multisensory neurons in SC and STS does not match with the relative precision of audiovisual timing perception (Stein and Meredith, 1993; Stone et al., 2001; Kopinska and Harris, 2004).

A second and more important flaw that this proposition makes is the assumption that neural latencies recorded by neuroscientists match with the signal latencies at the

neuron's level (Scharnowski et al., 2013). Indeed, these recorded neural latencies are always observed from an external reference frame: they reflect the time it takes to observe event related network activity *compared to stimulus onset timing*. But the neurons do not “know” the timing of external events. They only have access to the signal timing they receive from their connections. Because of the different communication speeds between neurons, it is possible that they perceive the processed information at different latencies than those recorded from the stimulus-centric reference. For instance, a neuron could receive information from higher visual areas more quickly than information from primary visual cortex (fig. 1.3) (Scharnowski et al., 2013). Hence, the reported neural latencies across the brain visual hierarchy may have no validity for the neurons, which might rely on an internal reference of time that is completely distinct (fig. 1.3). As such the perceptual latencies between the different features of perception could indeed reflect the signal processing latencies from the neuron's viewpoint (Scharnowski et al., 2013). This view rehabilitates the differential latency hypothesis, and suggests that neural latencies *from a brain-centric view* could correspond to perceptual latencies. Then, to assess when the neuron starts receiving information of the stimulus input, neuroscientists a priori need the full knowledge of each neuron's connections (Scharnowski et al., 2013), which makes the differential latency hypothesis in its new form hard to test.

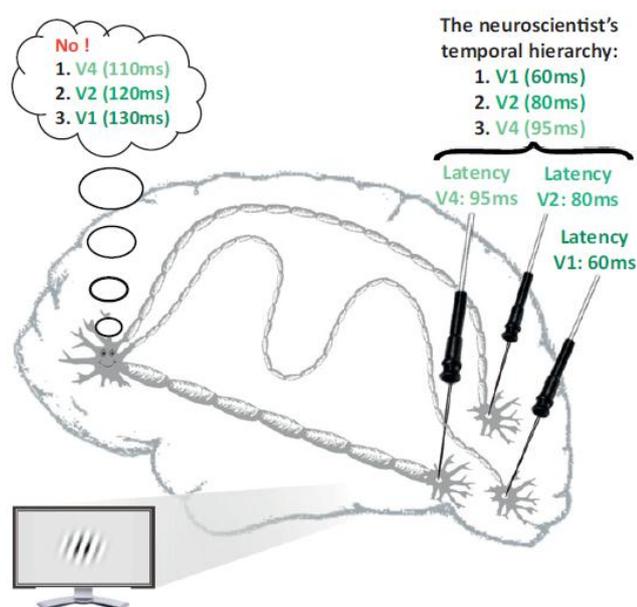


Figure 1.3: Neurorelativity. The known neural latencies at different sensory sites might have no validity at the neuron's level. For instance, it may have a privileged fast access to the apparently late sensory signal coming from V4, while information from V1 and V2 may be relayed via slow and indirect neural routes. Adapted from (Scharnowski et al., 2013)

1.2.3. Temporal plasticity

To circumvent the “neurorelativity” issue (fig. 1.3), a solution can be found in the plasticity of subjective event timing. While a pure internalization of timing could have the advantage of giving a measure of time that is constant and reliable with experience, contrary evidence suggests that the subjective metric of time presents some tuning properties to match external timing information. As defined by Michon (1985), tuning refers to the property of the system to keep track of the correspondence between the timing in the external world and the timing provided with the internal reference, “keeping the two series in synchrony is precisely what tuning is about” (Michon, 1985).

Tuning in that sense is close to temporal expectation, i.e. a mechanism that uses past stimulation to build up predictions on the timing of incoming events. The temporal prediction of incoming sensory stimulation is a well-known phenomenon that modulates our perception of implicit timing, as we are able to predict the position of moving objects (Coull and Nobre, 1998; Nobre, 2001; Coull et al., 2004; Nobre et al., 2007), or correctly detect percepts after a reliable temporal cue (Treisman and Howarth, 1959; Olson and Chun, 2001). Temporal expectations can be built on rhythmic information, and perception is enhanced at specific points in time which are in phase with the external rhythm (Jones, 1976; Jones and Boltz, 1989; Large and Jones, 1999). For example, after rhythmic stimulation a near-threshold visual stimulus is more likely to be detected if it is displayed at multiples of the rhythmic period (Mathewson et al., 2010) and auditory pitch sounds can be better discriminated (Barnes and Jones, 2000).

Temporal predictability also impacts our explicit estimation of events timing (Barnes and Jones, 2000). One classical illustration is the distance effect in audiovisual perception (Sugita and Suzuki, 2003; Fujisaki et al., 2004; Vroomen et al., 2004; Miyazaki et al., 2006; Yamamoto et al., 2012). Because of the difference in light and sound speeds, synchronous audiovisual stimuli arrive asynchronous at the position of the observer. At 15 meter distance, for instance, the sound is already delayed by about 45 ms to the visual information. Yet participants are able to modulate their perceptual judgment of audiovisual synchrony to compensate the external transmission delays (Sugita and Suzuki, 2003). The auditory sound speed may be modulated by both distance as well as transmission medium. The sound speed in air in particular is modulated by temperature,

such as the sound of an object 100 meter away takes around 286 ms to arrive to the ear of the observer at 30°C, while it takes 307 ms to run the distance at -10°C. It has thus been hypothesized that compensation mechanisms take place to overcome the variability in transmission delays. Studies have shown that the repeated presentation of asynchronous stimuli engender the recalibration of audiovisual synchrony (Fujisaki et al., 2004; Vroomen et al., 2004; Di Luca et al., 2009; Heron et al., 2010; Roseboom and Arnold, 2011) and auditory-motor (Praamstra et al., 2003; Cai et al., 2012).

We can therefore conclude that the brain mechanisms for encoding event timing should be flexible and should account for the variability in temporal reports. Hence, if neural latencies are the hallmark of event timing, then any variation in the perceptual timing of a stimulus should correlate with a variation in the stimulus evoked response in the corresponding sensory area. To test this, neuroscientists capitalized on a known perceptual illusion of time called “prior-entry effect”. This effect implies that attended stimuli reach conscious access more quickly than unattended stimuli (Titchener, 1908; Spence and Parise, 2010). Vibell and colleagues (2007) tested whether prior entry was caused by a speeding up of neural latencies with electroencephalography (EEG) in an audio-tactile task. They indeed found temporal facilitation of visual processing when the stimulus was attended (Vibell et al., 2007), yet the changes of neural latencies were quite small (from 3-4 ms for the N1 to 14 ms for the P300) and did not match quantitatively with perceptual reports (perceptual latency changes of 38 ms) (fig. 1.4a). Two years before, an audiovisual prior entry study failed to observe latency shifts in evoked activity, and even reported changes in amplitude of the evoked response (McDonald et al., 2005) (fig 1.4b). This study suggests that timing encoding is not reflected by neural timing, and that the attentional gain of sensory processing can translate into temporal facilitation of sensory processing.

Looking at the timing of EEG activity at seminal peak latencies (N1, P1, P300) follows the idea that the average neural timing in sensory areas accounts for event timing encoding. However, event timing could either be encoded within the timing of a subset of neurons in sensory areas (Chase and Young, 2007; Brasselet et al., 2012) and then could not be observed with Evoke Related Potential (ERP) analysis. Specific networks in sensory networks may have fast latencies following stimulus onset that are “stereotyped”,

i.e. independent of non-temporal stimulus properties (Brasselet et al., 2012), and in this manner may provide a good tag of events' timing. However, two problems arise from this possible code of event timing. First, if the response of these “stereotyped neurons” is static as independent on non-temporal factors, then they cannot explain any plastic change in perceived event timing. Second, stereotyped neurons cannot be used for multi-features temporal integration as the encoding is sensory specific and does not contain information about its relative temporal position according to past and future events.

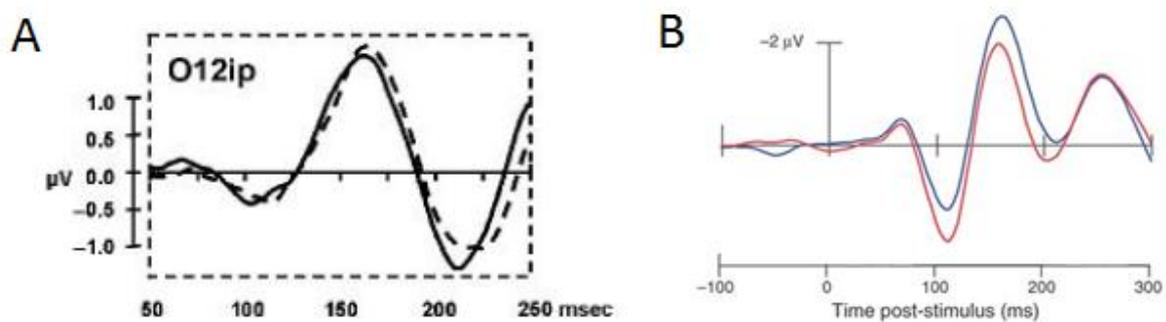


Figure 1.4: Changes in perceived timing are not reflected by changes in neural timing. (A) In a visuo-tactile prior entry task, paying attention to touch (dashed line) compared to vision (continuous) slightly delays the visual evoked response (4 ms for N1, 14 ms for P300), yet this delay is much smaller than perceptual delays (38 ms). (B) In an audiovisual prior-entry task, paying attention to sound (red) compared to vision (blue) reduces the amplitude of the visual evoked response. Adapted from (McDonald et al., 2005; Vibell et al., 2007)

To summarize, it appears that neural latencies _ recorded at the stimulus onset reference frame _ are not good indicators of subjective event timing. Yet, there is no need for a linear relation between neural mechanisms and mental states. As mentioned before, non-temporal properties of sensation could be encoded within the timing of neural activity; conversely, could time be encoded with non-temporal neural features? The next section addresses the possibility of dedicated mechanisms for the encoding of timing.

1.3. ENCODING EVENT TIMING WITH DEDICATED NEURAL STRUCTURES

1.3.1. Specialized brain areas for the encoding of time

If neural time can represent space, motion and other qualities of perception _ and as neural transmission delays do not seem to reflect subjective time _ would it be relevant to consider a non-temporal code for event timing perception?

In particular, encoding time through the activity of a dedicated network would have the ability to easily account for the amodal aspect of time, i.e. that participants are able to estimate the temporal relations between events from distinct sensory modalities. While it is clear that there is no sensory area dedicated to time perception, the existence of brain structures dedicated to the encoding of time is still debated (Treisman et al., 1990; Harrington et al., 1998; Lewis and Miall, 2003, 2006; Coull et al., 2004; Buhusi and Meck, 2005; Ivry and Schlerf, 2008; Wittmann, 2013; Morillon et al., 2009; van Wassenhove, 2009; Wittmann, 2009).

Neuroimaging studies suggest the involvement of several brains structures during time perception: cerebellum has an important role in motor timing, coupled with supplementary motor area (SMA) and motor cortex (Harrington et al., 1998; Ivry and Schlerf, 2008; Schwartz and Kotz, 2013). Basal ganglia and thalamus may provide metrics for absolute timing (Buhusi and Meck, 2005; Schwartz and Kotz, 2013). Parieto-frontal areas are involved in the conscious discrimination of temporal information (Coull et al., 2004; Nobre et al., 2007). Insula may register physiological states of the subject to encode duration (Wittmann and Paulus, 2008; Craig, 2009; Wittmann, 2009, 2013). Finally, early sensory and multisensory areas are also activated when discriminating relative timing or rate of external information (Dhamala et al., 2007; Noesselt et al., 2007; van Wassenhove and Nagarajan, 2007), and in particular the auditory cortex can be recruited even in the absence of sound (Coull et al., 2000). While these structures are specialized in the encoding of a certain aspect of time perception, evidence suggests together they contribute to the final time percept (Lewis and Miall, 2003). It should be noted that many of the reported experiments capture either encoding, emotional, and

decisional aspects of timing reports, and may reflect multiple aspects of temporal processing such as duration, temporal prediction, temporal order, or synchrony perception. It is probable that these different facets of time perception recruit different areas and/or exploit different neural mechanisms. Here we review two dedicated mechanisms that could serve the encoding of relative event timing.

1.3.2. Delay-tuned cells

As explained previously, synchrony perception could be achieved via neural signals converging at specific hubs whose role is to estimate the coincidence of sensory inputs. In multisensory context, SC (Meredith et al., 1987) and dedicated structures in cortex, such as STS or insula (Benevento et al., 1977), have been shown to respond to synchronous multisensory stimuli. Additionally, these areas seem to play a prominent role in temporal audiovisual binding. Of interest, the multisensory neurons in these brain regions have different temporal tuning properties (Meredith et al., 1987). As multisensory neurons are sensitive to different temporal delays, audiovisual timing could be extracted from the whole population code (Roach et al., 2011; Cai et al., 2012; Heron et al., 2012). Computational models further suggest that the adaptation of these “delay-tuned neurons” could underlie reports of changes in subjective timing, such as during temporal recalibration (Roach et al., 2011; Cai et al., 2012) (fig. 1.5).

While “time cells” _ which encode long durations stored in working memory _ have been found in the hippocampus (Pastalkova et al., 2008; MacDonald et al., 2011; Kraus et al., 2013), the existence of similar mechanisms for the encoding of short perceptual timing in sensory and multisensory areas remain to be empirically tested. Other structures, in particular the cerebellum, have been hypothesized to encode time in a “interval-based fashion”, so that each temporal relation between task-dependent stimuli could be stored following a columnar organization (Ivry, 1996). Yet the main argument against the existence of delay-tuned neurons (or interval-based temporal code) is its mechanistic cost: a priori this mechanism needs infinity of cells to compute all possible delays between all possible perceived duets of features; as such it is not a parsimonious computational process (Brasselet et al., 2012).

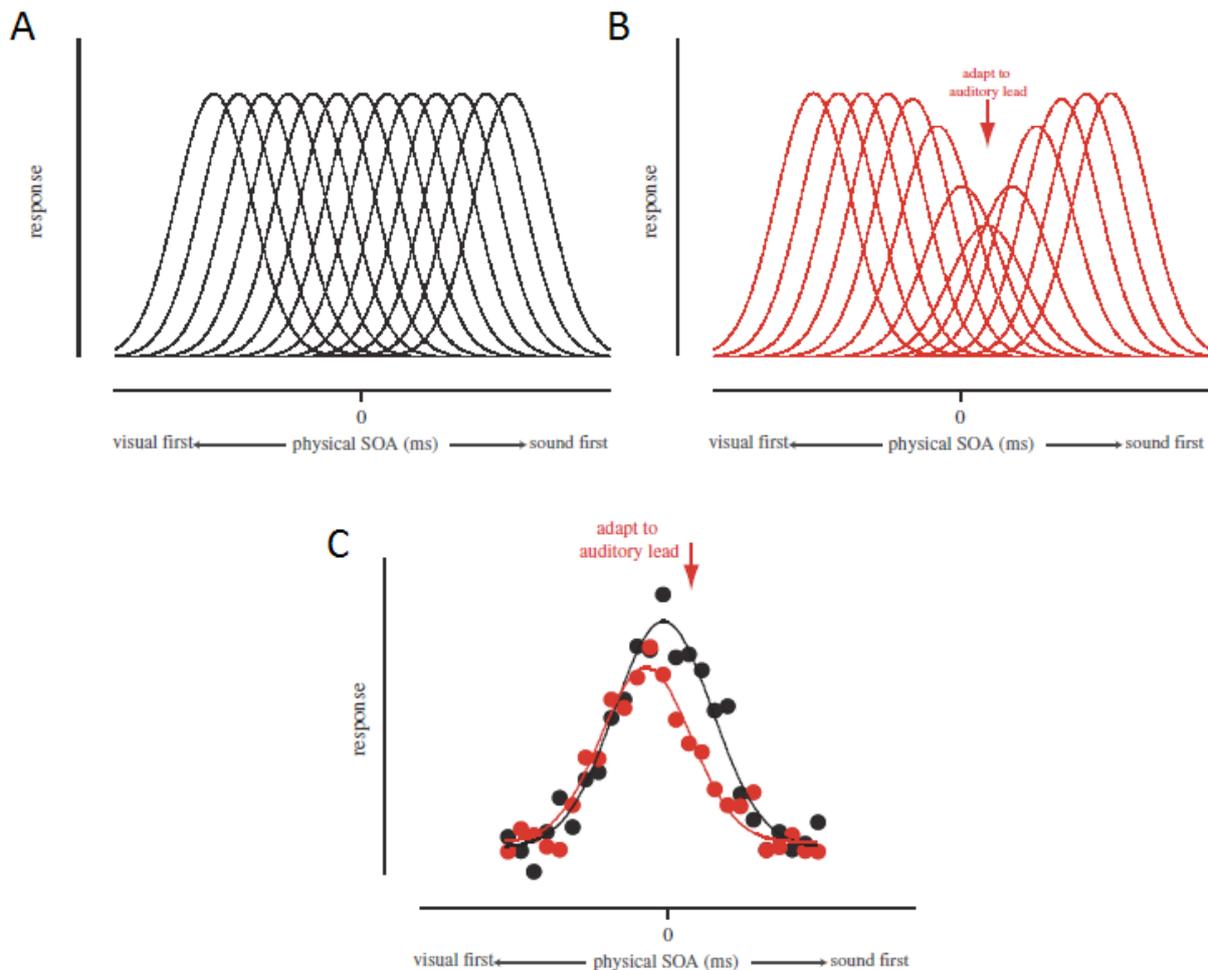


Figure 1.5: Delay-tuned cells. (A) The interval-based model is based on a population code composed of neurons to different audiovisual delays (B) Neurons can adapt to a repeated audiovisual delay and change their tuning curves accordingly (C) After adaptation, the population code of delay-tuned cells can account for temporal recalibration, which consists in the bias in synchrony perception after the repeated exposure to asynchronous audiovisual stimuli. Adapted from (Roach et al., 2011)

1.3.3. Brain clocks

If time is measured in the external world with a clock, why wouldn't the brain build its own clock for timing perception? Following this idea, another dedicated model for time perception suggests that the brain relies on an internal clock for estimating the passage of time (Treisman, 1963; Church, 1984). In this model, the clock is composed of a pacemaker, a counter and a comparator. The pacemaker consists in an oscillator ticking at a default frequency (Treisman et al., 1990, 1992). The counter sums the number of ticks emitted by the pacemaker since stimulus onset, and the comparator receives

information from the counter to estimate the elapsed duration by comparing this representation of time with those stored in working memory (fig. 1.6).

Crucially the pacemaker is the mechanism that serves for the encoding of temporal information, while the counter and the comparator are linked more to the accumulation of evidence and the decision on the perceptual content of time. The pacemaker thus provides the temporal reference on which the brain estimates relative timing between events. Importantly, its default frequency is not fixed; it can be modified by the arousal state of the individual or the temporal statistics of external inputs. As such the internal clock model can account for variability in explicit duration judgments, and can explain some characteristics of implicit temporal prediction as well (Schwartz and Kotz, 2013). Further evidence suggests that the internal clock should not only operate at one specific frequency (Church, 1984; Treisman, 1984; Treisman et al., 1990, 1992; Buhusi and Meck, 2009), and implies the presence of multiple fixed pacemakers for accurate frequency estimations (Buhusi and Meck, 2005, 2009). Additionally, the presence of adaptive pacemakers could account for changes in external timing depending of the properties of sensory stimuli (Treisman, 1984; Treisman et al., 1990).

Pacemakers are instruments that fluctuate in periodic fashion. Neural oscillations have thus been hypothesized to be the best candidates in the role of pacemakers for conscious time estimation (Treisman et al., 1990; Pöppel, 1997; Buhusi and Meck, 2005, 2009). While previous accounts suggested that the Substantia Nigra could play the role of the pacemaker for time perception in the absence of stimulation (Meck, 1996), the Striatal Beat Model (fig. 1.7) suggests that cortical oscillations may also play the role of pacemakers when judging the duration between successive events (Matell and Meck, 2004). During timing estimation, cortical oscillations are read by striatum whose role is to detect stimulus onset and offset according to oscillatory coincidence between the different cortical oscillators. In turn, it sends back information to the cortex to correct the pacemakers if necessary. Hence, it appears that basal ganglia have a dedicated role in time estimation, which is to “read” timing through the behavior of non-dedicated-to-time cortical oscillations and to correct timing in cortex. Furthermore, according to this model, it appears that cortical oscillations may represent the internal temporal reference with which timing judgments are elaborated.

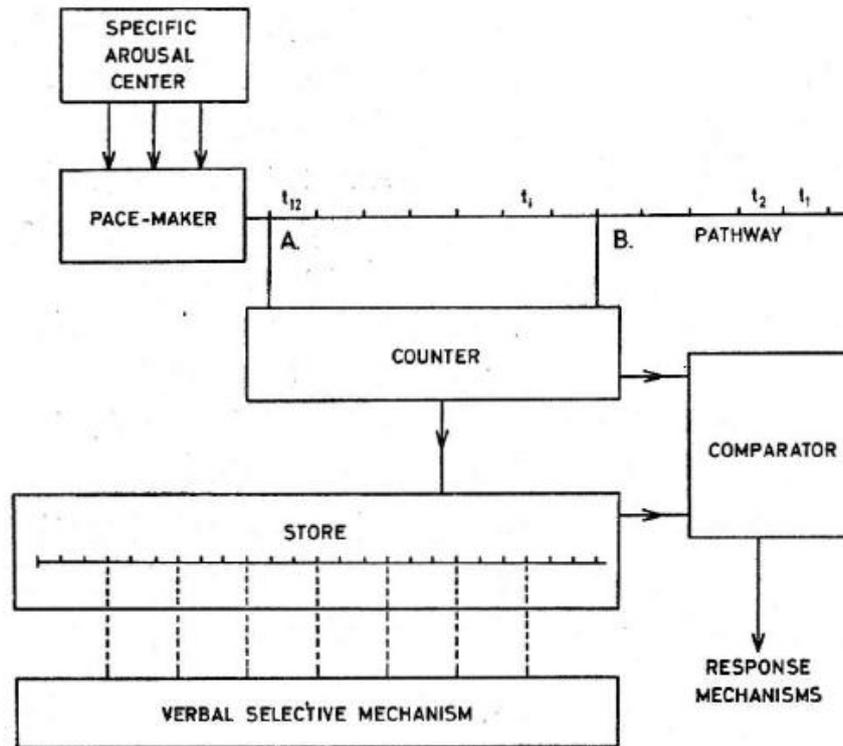


Figure 1.6: The clock model. The clock is composed of a pacemaker that emits regular pulses at a constant interval. The frequency of the pacemaker can be modulated by arousal factors. The counter records the numbers of pulses emitted by the pacemaker between two successive events. The resulting number of pulses is stored in working memory and could be either decoded as a verbal estimate or compared to previous stored items to judge the interval duration. Adapted from (Treisman, 1963)

In summary, two main dedicated mechanisms have been put forward for encoding the relative timing between events. A first model is based on the interval-based code where specific neurons are tuned to specific delays between specific sensory features. This mechanism seems very costly as there must be one cell for each relative duration and for each possible feature of sensation. The second model considers the existence of a brain clock which is composed of a pacemaker (the metric of time), a counter and a comparator. Interestingly, while counter and comparator mechanisms are suggested to rely on dedicated mechanisms; evidence implies that these mechanisms read the dynamics of cortical activity for the establishment of relative time. Particularly, it implies that neural oscillations in networks non-dedicated to the encoding of time _ in particular sensory areas _ could provide the metrics for conscious time. In the next section, we explore the putative role of cortical oscillations in the encoding of relative event timing for perception.

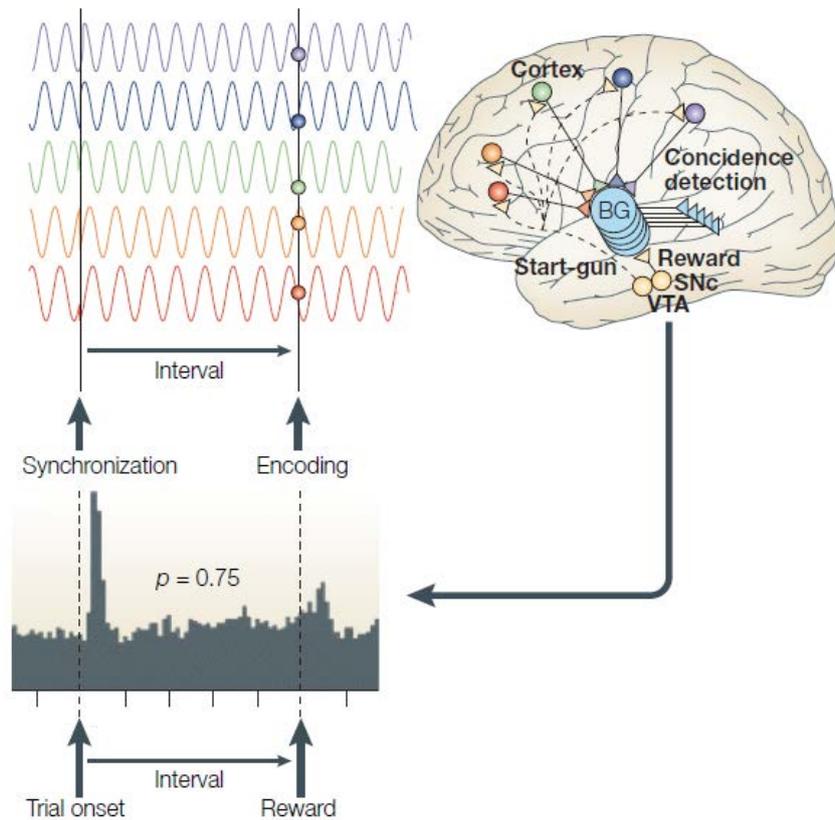


Figure 1.7: The Striatal Beat Model. Cortical oscillations (top panel) project onto basal ganglia (bottom panel), whose role is to detect coincidence in oscillatory activity (e.g. stimulus onset) and to continuously compare the states of cortical oscillations with the expected state detected at the time of the reward. Adapted from (Buhusi and Meck, 2005).

1.4. ENCODING EVENT TIMING WITH BRAIN OSCILLATIONS

As discussed in section 2, one main challenge for the brain is to recover the external timing of the world from its own dynamics. A first hypothesis states that the brain does not compensate for neural transmission delays from sensors to sensory areas, and thus that perceived timing equates neural timing. However, several bodies of work showed that neural transmission delays from stimulus onset to sensory area were bad indicators of perceived timing. Furthermore, many perceptual effects suggest that the time experiencer compensates for its internal neural temporal delays. Therefore, it has been suggested that perceived timing is encoded at a later stage of sensory processing via dedicated structures (Ivry and Schlerf, 2008). In this theoretical framework, timing

compensation could operate through the crossing of sensory informations at specific networks that are specialized for the encoding of timing. Yet it seems that most proposals of dedicated networks for the encoding of time depict networks that act as “readers” of temporal information. Interestingly, this temporal information might be retrieved from the dynamics of non-dedicated brain areas, including sensory networks. This means that the brain feeds on its internal temporal fluctuations to recover time. This view is close to what William James suggested, when stating that time perception could rely on “outward sensible series”, “heart beats” or “breathing”, but also on “the pulses of our attention, fragments of words or sentences that pass through our imagination”. Hence subjective timing could rely on the timing of mental activity. From a neuroscientific point of view, this suggests that ongoing brain dynamics provide the temporal grounds for the inner representation of time. Here, we review evidence that, among indices of brain dynamics, neural oscillatory activity is a relevant candidate to provide the ongoing metrics of time that could be used for sensory processing.

1.4.1. Temporal binding through neural coherence

Neural activity at neuron, network, or area level is characterized by well-described intrinsic periodic fluctuations that span across multiple time scales (Buzsáki and Draguhn, 2004; Roopun et al., 2008; Wang, 2010). These neural oscillations are observed through Local Field Potential (LFP) recordings within the brain, or with electroencephalography (EEG) and magnetoencephalography (MEG) that record electrical and magnetic fields coming out of the scalp of the subject. Although brain rhythms range from 0.02 Hz to 600 Hz (Buzsáki and Draguhn, 2004), they are usually classified within distinct frequency bands: infra-slow oscillations concern rhythms below 1 Hz, delta oscillations corresponds to rhythms between 1-3Hz, theta oscillations span between 3-8 Hz, alpha oscillations between 8-13 Hz, beta oscillations between 15-25 Hz, gamma oscillations between 30 and 120 Hz, and ripples above 150 Hz.

Neural oscillations typically reflect the *synchronous* fluctuating activity of a neural ensemble (Varela and Lachaux, 2001; Buzsáki, 2004, 2010; Lakatos et al., 2005): the presence of oscillations in the signal suggests that the overall activity of the neurons in the networks is grouped at certain periodic time points. Thus neural oscillations constitute a

marker of temporal coherency in local networks. Crucially, following the Hebbian rule “cells that wire together fire together”, it is suggested that the networks that are encoding a common perceptual object – or “cell assembly” (Buzsáki, 2010) – should fire within the same amount of time. If oscillations modulate neural synchrony, then they should provide mechanistic means for neurons that process the same attribute to communicate (Fries, 2005; Sejnowski and Paulsen, 2006; Buzsáki, 2010). The Hebbian rule is not restricted to local processing; it applies also to the coordination of distant brain regions that are encoding the same object. Then, neural oscillations should play a prominent role in the temporal binding of different sensory features (Pöppel et al., 1990; Engel et al., 1991c, 1999; Senkowski et al., 2008; Pöppel, 2009)

Gamma oscillations were first targeted to explain temporal binding. These oscillations are imputed to reveal the coordinated activity within small cell assemblies (Gray et al., 1989; Engel et al., 1991a, 1991b, 1991c; Engel and Singer, 2001b; Fries, 2005, 2009; Sirota et al., 2008), considering that they reflect fluctuations of neural activity of the same time scale than the shortest recorded postsynaptic times constants (in the order of 10 ms–30 ms) (Bragin et al., 1995). Hence, one gamma oscillation cycle should provide the smallest encoding unit in the neural signal, and should reflect the temporal coherence of a fundamental cell assembly (Buzsáki, 2010). The first evidence linking gamma oscillation synchrony to feature binding has been observed in cat visual cortex. Synchronization of neural activity at gamma rate was observed within a visual column in primary visual cortex (Eckhorn et al., 1988), between two visual columns (Gray et al., 1989; Engel et al., 1991b), across visual hemispheres (Engel et al., 1991c), and across distinct visual areas (Engel and Kreiter, 1991). Most importantly, gamma synchronization between two regions of cortex was only observed when they were encoding features from the same object, and not from distinct perceptual objects (Engel et al., 1991a, 1991b) (fig. 1.10). Following studies on primate and human individuals confirmed that gamma synchronization between distinct brain regions is associated with cognitive binding (Tallon-Baudry et al., 1996; Roelfsema et al., 1997; Rodriguez et al., 1999; Tallon-Baudry and Bertrand, 1999; Womelsdorf et al., 2006). Gamma oscillations synchrony between remote brain areas thus reflects information transfer between these regions. As far as sensory cortices are concerned, an increase in oscillation phase coherence can reflect temporal binding between different sensory attributes. Accordingly,

similar gamma activity synchronization has been observed across sensory regions and is related to cross-sensory binding mechanisms (Senkowski et al., 2008).

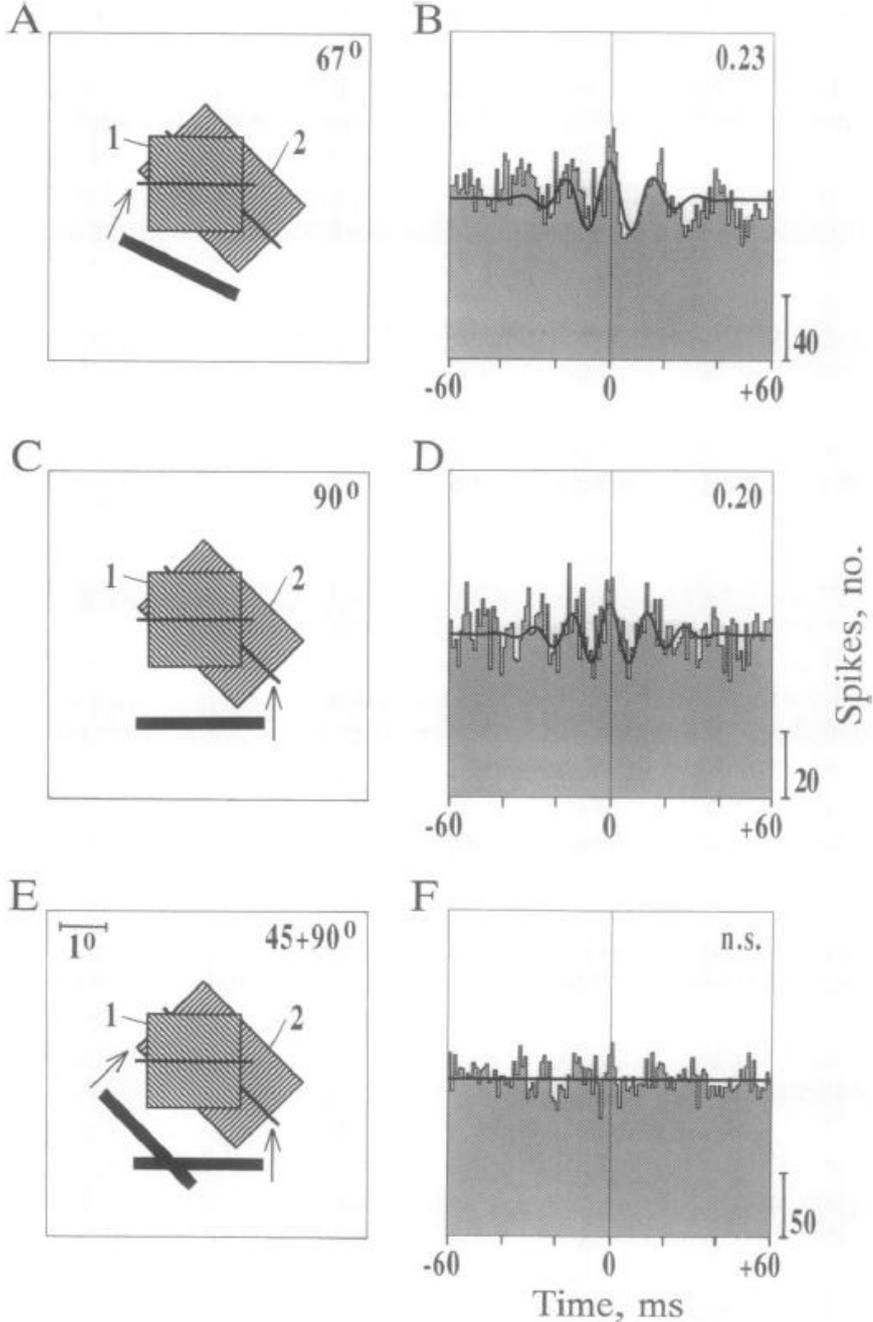


Figure 1.10: Temporal binding through gamma oscillation coherence. (A and C) Stimulating visual columns 1 and 2 with single bars entailed (B and D) the inter-columnar synchronization of neural excitability at gamma frequency. (D) Stimulating the two columns with two distinct bars cause (E) no inter-columnar synchronization. Adapted from (Engel et al., 1991b)

However, temporal binding by gamma synchrony between distant brain areas should be limited by anatomical conduction delays (von Stein and Sarnthein, 2000). It has thus been suggested that slower oscillations take part in communication mechanisms (von Stein and Sarnthein, 2000; Varela and Lachaux, 2001; Siegel et al., 2012), including beta (von Stein and Sarnthein, 2000; Hipp et al., 2011), alpha (Palva and Palva, 2011), theta and even slower oscillations (Schroeder and Lakatos, 2009). As such different oscillators with different time scales take part in temporal binding mechanisms.

1.4.2. Oscillations affect perceptual temporal sampling

Oscillatory activity reflects the temporal coherence of a neural ensemble, and also corresponds to cyclic fluctuations of a network's excitability (Bishop, 1932; Lakatos et al., 2005). The phase of the neural oscillation is correlated with the firing rate of the network, independently of the oscillation frequency (Lakatos et al., 2005) (fig. 1.8). Thus, neural oscillations of different frequencies provide different windows of excitability that can serve to parse the neural signal into relevant encoding units. The networks that are active within one oscillation cycle may encode one perceptual entity _ or “neural word”; and the combination of succeeding neural words can constitute a neural “sentence” (Buzsáki, 2010).

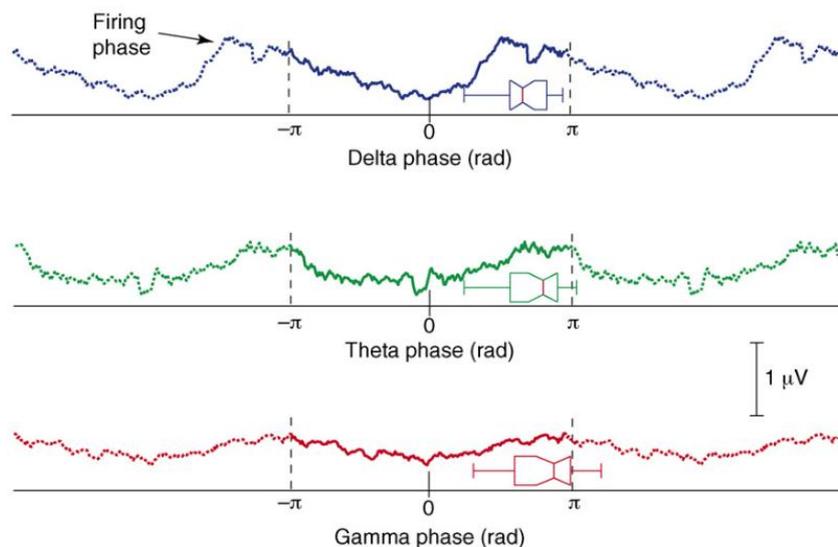


Figure 1.8: Brain oscillations reflect neural excitability. Multiunit activity (MUA) amplitude from primary auditory cortex in awake macaque monkey is modulated by the phase of delta (1–4 Hz, blue), theta (5–7 Hz, green) and gamma (25–50 Hz, red) oscillations. Adapted from (Lakatos et al., 2005; Schroeder et al., 2008)

Coding information through neural oscillations presents many advantages. First, silencing activity periodically is advantageous in terms of resources costs, considering that the generation of action potentials is energetically expensive (Laughlin and Sejnowski, 2003). Second, because it provides syntactic rules to silence neural activity, it constitutes a code of well-separated messages that should be easier to interpret than uninterrupted signals. As such, an oscillation code should transmit information more effectively than continuous spiking activity (Sakurai, 1999; Wickelgren, 1999).

Yet, the chunking of the neural signal into units of information has a crucial consequence for perceptual coding schemes. It suggests that sensory input is not processed continuously, but discretized into quanta of information (Stroud, 1967). Numerous studies confirm that sensory information is sampled during sensory processing (Purves, 1996; VanRullen and Koch, 2003; VanRullen et al., 2006; Fiebelkorn et al., 2011; Giraud and Poeppel, 2012; VanRullen and Macdonald, 2012; Sokoliuk and VanRullen, 2013). The most convincing illustration of the sampling of sensory information comes from the invention of the cinema. Perceiving continuity when watching a movie necessitates the humans' inability to perceive the discrete succession of images. This suggests that the human visual system does not process stimulus dynamics continuously, but integrates the signal in bits of information. The fact we perceive continuity in the visual world constitutes another introspective evidence of the existence of discrete sensory sampling processes. If we were able to perceive the continuous flow of visual information, we would be aware of the abrupt changes in visual scenes due to saccadic eye movements (Andrews and Coppola, 1999).

If the temporal resolution of perception relies on an endogenous neural oscillation that acts as a "sampler", it implies that sensory information could only correctly be perceived if it fluctuates at a relatively slow rate. If stimulation dynamics are as fast as half the frequency of the neural sampler, temporal aliasing should be observed. Aliasing occurs when a signal is discretely sampled at a rate that is insufficient for capturing all the variations in the signal. A typical example of this phenomenon often observed in films is the wagon-wheel illusion. When the rotation speed of a filmed wheel is slower than twice the movie frame rate, the wheel is perceived rotating backward while the vehicle is clearly moving forward (Purves, 1996). Consistent with a neural sampling mechanism of visual stimulation, this illusion can be observed in reality (Purves, 1996; VanRullen et al.,

2005b). However, the percept of the illusory backward rotation is however not stable but alternates with the true forward rotation over time (Purves, 1996; VanRullen et al., 2005b). This effect has been quantified in studies showing that a 10 Hz rotation maximized the effect (VanRullen et al., 2005b). Additionally, EEG data confirmed the implication of neural oscillations, showing that alpha power (13 Hz) predicted reversals of perceived rotation (VanRullen et al., 2006). Interestingly, oscillations in the alpha band (8-13 Hz) are dominant in occipital and parietal regions. The brain might thus capitalize on this prominent oscillation to sample sensory signal into perceptual coherent units. Consistently, several studies showed the association between alpha oscillations activity and visual temporal sampling. In particular synchrony judgments between two static flashes (Varela et al., 1981; Gho and Varela, 1988), and between a moving and a static stimulus (Chakravarthi and Vanrullen, 2012) in the flash-lag effect are predicted by pre-stimulus alpha phase.

The present results suggest that one cycle of a neural oscillation encodes for one temporal sample of sensory information. But it has also been suggested that oscillation phases provide deeper insight into the fine-grained coding of sensory information within one duty cycle. As neural oscillations shape neural excitability in time (Lakatos et al., 2005), it entails that perceptual excitability should be maximal at some particular phase of the oscillation cycle, and minimal at the opposite phase. Neuroimaging results validate this prediction, by showing that perceptual detection and discrimination is determined by pre-stimulus oscillation phase (Monto et al., 2008; Busch et al., 2009; Romei et al., 2009, 2012; Mathewson et al., 2010; Chakravarthi and Vanrullen, 2012; Henry and Obleser, 2012; Neuling et al., 2012a). For instance, the detection of a near threshold visual stimulus could be predicted by pre-stimulus alpha phase in visual and frontal cortex (Busch et al., 2009; Mathewson et al., 2010) (fig. 1.9). Similarly, applying a rhythmic external current via oscillating transcranial direct current stimulation at 10 Hz can generate oscillatory activity in the auditory cortex; and the phase at stimulus onset of the engendered oscillation predicts auditory detection (Neuling et al., 2012a).

The shaping of perceptual excitability by slow oscillations suggests that periods of excitability are followed by periods of inhibition. Interestingly, the peaks of slow oscillations are usually associated with periods of low neural excitability (Lakatos et al., 2005; Haegens et al., 2011), suggesting that slow oscillations may actually reflect the

silencing of neural excitability. In line with this idea, strong slow oscillating components in the signal have been associated with sensorimotor inhibition (Linkenkaer-Hansen, 2004; van Dijk et al., 2008; Jensen and Mazaheri, 2010). Moreover, studies show evidence that stronger slow oscillation power entails more pronounced cyclic stimulus inhibition (Jensen and Mazaheri, 2010; Haegens et al., 2011; Mathewson et al., 2011, 2012; Joundi et al., 2012).

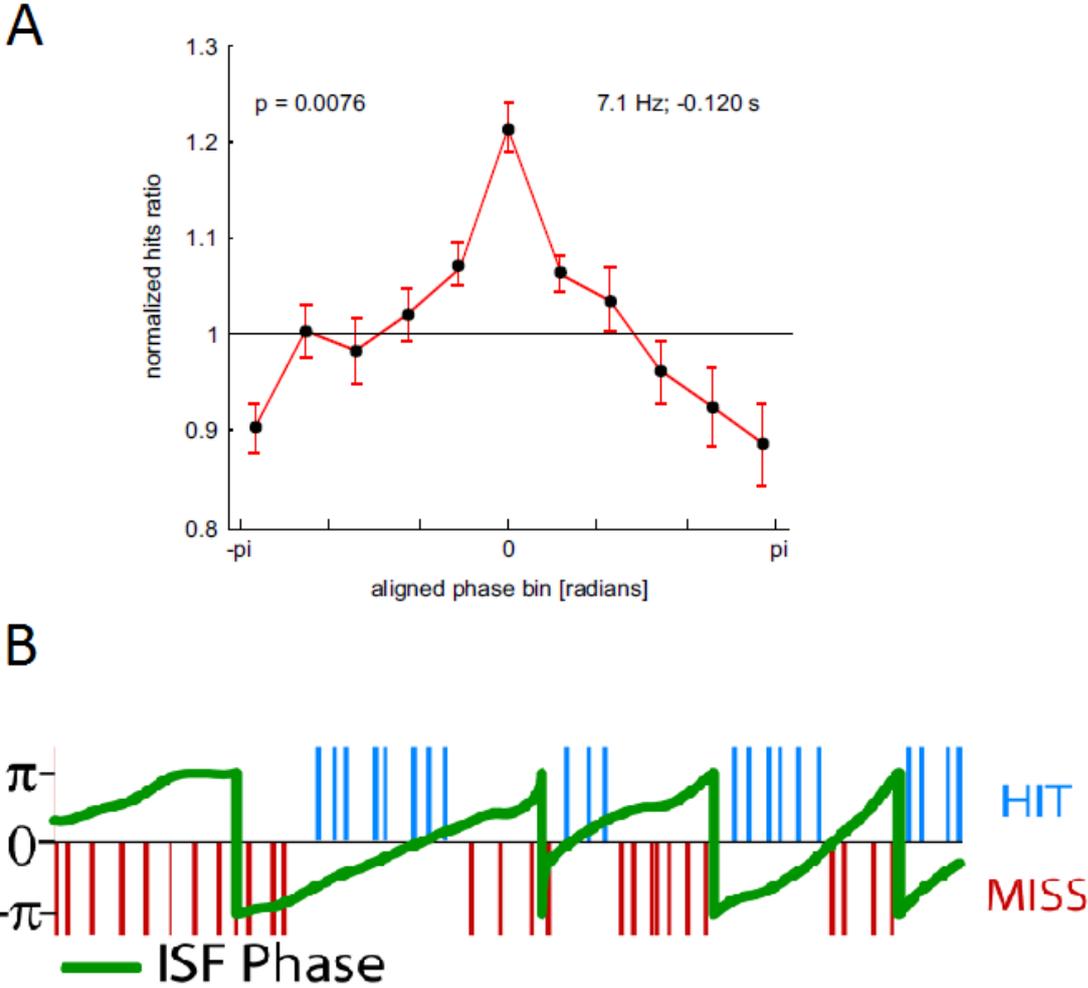


Figure 1.9: Cortical oscillations shape perception in time. (A) Endogenous pre-stimulus alpha phase in fronto-central areas predict whether a visual stimulus will be detected or missed. Adapted from (Busch et al., 2009). (B) Infra-slow frequency (ISF, <1 Hz) oscillation phase has also an impact on visual detection. Adapted from (Monto et al., 2008).

Brain oscillations are ubiquitous to neural activity, and are present in the total absence of external stimulation (Buzsáki and Draguhn, 2004). However, there is strong evidence that brain oscillations are modulated in the presence of sensory inputs, first by phase reset (Fiebelkorn et al., 2011; Thorne et al., 2011; Romei et al., 2012; Thorne and Debener, 2014), and second via entrainment (Schroeder and Lakatos, 2009). Neural entrainment reflects the propensity of sensory cortices to tune to the dynamics of the world and to follow environmental rhythms (Rees et al., 1986; Hari, 1989). Knowing that oscillations shape perception in time, neural entrainment thus provide a plausible mechanism for improving perceptual performances at expected temporal points (Lakatos et al., 2008; Stefanics et al., 2010b; Henry and Obleser, 2012; Mathewson et al., 2012; Cravo et al., 2013; Graaf et al., 2013), and could present a valuable explanation for known psychophysical results on rhythmic expectation (Jones and Boltz, 1989; Barnes and Jones, 2000; Mathewson et al., 2010).

Overall, the present findings point to a clear role of cortical oscillations in the temporal sampling of perception. Yet, reducing oscillations to temporal “samplers” could have detrimental consequences on perception. Sampling implies that sensory information which occurs within the sample period is reduced to one unit of neural information. Hence, if a cognitive function is based on the recruitment of a neural oscillation, it should entail that the sensory dynamics that are faster than the oscillation frequency could not be retrieved anymore. Yet, to the contrary, growing evidence suggests that fine-grained temporal information across different time scales could still be processed owing to the parallelization of multiple neural oscillators.

1.4.3. Neural phase codes of events succession

A fundamental property of neural syntax code proposed by Buzsáki (2010) is that the smallest units of the neural codes, e.g. the gamma oscillations bursts (Gray et al., 1989; Engel et al., 1991a, 1991b, 1991c; Engel and Singer, 2001b; Fries, 2005, 2009; Sirota et al., 2008), are embedded together to form more complex units of information. Thus, the temporal sequencing of distinct neural assemblies is at the core of the multiplexing of information in neuroscience. To do so, gamma oscillations bursts are grouped by slower

oscillations (Bragin et al., 1995; Lakatos et al., 2005; Canolty et al., 2006; Sirota et al., 2008; Buzsáki, 2010; Canolty and Knight, 2010; Siegel et al., 2012).

As reviewed previously, the nesting of fast oscillations within slower oscillations could reflect the temporal integration (or sampling) of smaller units of information into a more complex unit. Integration means that all information contained in the smaller units is aggregated, and as such that they cannot be retrieved separately anymore. It predicts in particular that the temporal information of each smaller unit of information should be lost. Some reports support this view by showing that delayed stimuli were perceived as synchronous if they arrived within the same cycle of a slow oscillation (Varela et al., 1981; Gho and Varela, 1988). Yet, several studies suggest that information relayed by fast oscillations is not “blurred” if coupled to slow oscillations. To the contrary slow and fast oscillations may work in parallel, and encode distinct aspects of perception in a hierarchical scheme (Lakatos et al., 2005; Schroeder et al., 2008; Ghitza, 2011; Giraud and Poeppel, 2012).

Here, we distinguish chunking (or parsing) from integration (or sampling). In opposition to integrative mechanisms, chunking mechanisms “only” serve the cutting of the signal into relevant temporal units. Chunking amounts to delineating a temporal window of processing, without merging the information that is present within this time window. As such, chunking provides relevant temporal markers for one cognitive processing, without deteriorating finer temporal information. A particularly relevant case where oscillatory chunking may operate is during speech processing. It has been hypothesized that the comprehension of spoken sentences relies on banks of neural oscillators with different frequencies (delta, theta, and gamma oscillations) that chunk in parallel the different units of language (phonemes, syllables and words) in the acoustic signal (Ghitza, 2011; Giraud and Poeppel, 2012).

If slow oscillations serve as a means for temporal binding, some information at higher time scales could still be preserved for correct sensory processing. High-frequency oscillations coupled to low-frequency oscillations should provide additional information on stimulus properties. The position of fast temporal dynamics within the slow oscillation cycle is not insignificant. It can constitute a powerful temporal code for perception (Montemurro et al., 2008; Kayser et al., 2009, 2012; Panzeri et al., 2010; Ng et al., 2013): the timing of spiking activity according to the phase of slow ongoing

oscillations is shown to be a reliable decoder of sensory content (Montemurro et al., 2008; Kayser et al., 2009; Panzeri et al., 2010; Ng et al., 2013). More precisely, different phases of slow oscillations provide temporal tags to partition neural spiking activity into meaningful bits of information. This encoding scheme is almost as efficient as using temporal windows that are locked to stimulus onset (Kayser et al., 2009, 2012; Ng et al., 2013). Therefore, owing to neural oscillations, the brain is able to extract stimulus information in time without the *a priori* knowledge of external timing.

Crucially for temporal perception, the ordering of gamma bursts within the slower oscillation cycle may have the general role of representing sensory event succession (Lisman, 2005; Lisman and Jensen, 2013). This hypothesis is supported by numerous findings of phase precession mechanisms in rat hippocampus during navigation (fig. 1.11). When the rat moves in a maze, it activates neurons in the hippocampus that encodes for a particular position in space. Phase precession refers to the fact that each of these “place cells” fire within a specific gamma burst that is coupled to a slow endogenous theta rhythm (O’Keefe and Recce, 1993; Bragin et al., 1995; Skaggs et al., 1996). Remarkably, place cells that fire in successive gamma cycles encode neighboring locations and the ordering of place cells activation within one theta cycle reflects the sequential order of locations explored by the rat in the maze (Jensen and Lisman, 1996, 2000; Skaggs et al., 1996; Pastalkova et al., 2008).

Phase precession mechanisms have also been reported between hippocampal theta and prefrontal areas (Jones and Wilson, 2005), and within cortical areas such as entorhinal cortex (Hafting et al., 2008). It has been hypothesized that phase precession provides a general mechanism for computing the temporal order of items (Lisman, 2005; Nadasdy, 2010; Lisman and Jensen, 2013). Computational models even suggest that interactions between sensory cortices and the hippocampus might be at the origins of known phase precession mechanisms (Nadasdy, 2010). Recent results further imply that phase precession is used in visual cortex as a mean to order visual elements by saliency (Bonfond and Jensen, 2012). Even if the represented information is not temporal per se, an ordered saliency map in alpha cycle could serve the prioritization of conscious access to the more salient stimuli (Jensen et al., 2012), and as such could have an influence in its perceived relative temporal order.

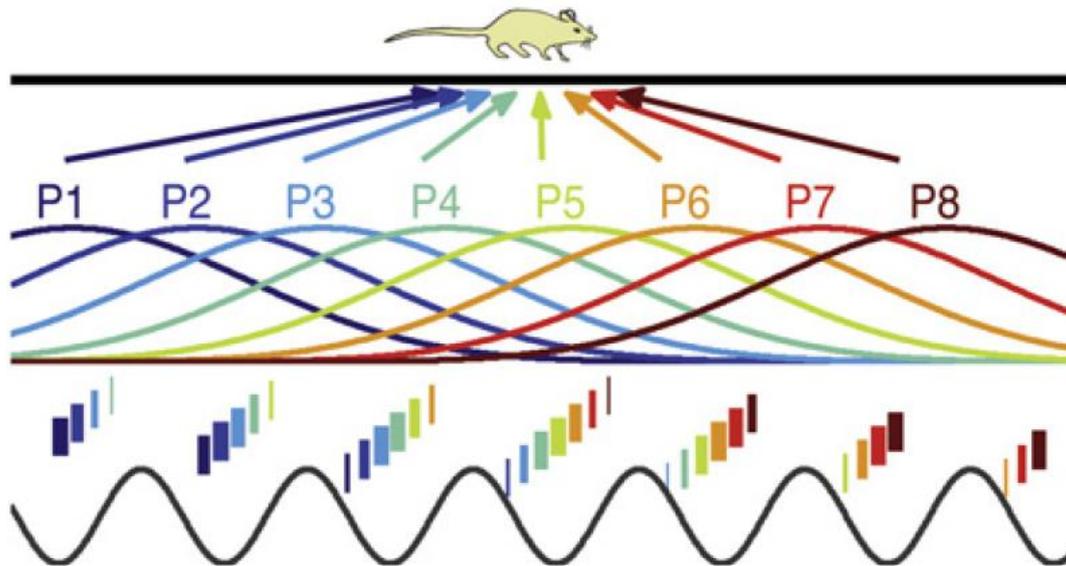


Figure 1.11: Phase precession mechanisms permit the temporal ordering of memorized items. When the rat traverses the maze, it activates different places cells with encodes neighboring locations (P1 to P8). Place cells fire (colored bars) at different phases of the hippocampal theta. The order reflects the path taken by the animal. Adapted from (Buzsáki, 2010)

The present findings suggest that the brain could use neural oscillations as an internal reference of time. Specifically, the phase of firing of neural networks according to the slow oscillation internal inference participates in the representation of sensory content (Panzeri et al., 2010; Kayser et al., 2012). In particular, the phase of cortical oscillations could represent the perceived order of successive events (Lisman, 2005; Lisman and Jensen, 2013). Knowing that oscillations may also constitute time metrics for perception (Matell and Meck, 2004; Buhusi and Meck, 2005), could the relative phase between successive events also encode relative timing?

1.4.4. Phase tagging of perceived event timing?

It appears that the brain uses oscillations in cortex as internal reference frames for time perception. First, cortical oscillations may provide internal metrics of time for duration perception. Second, the time course of stimulus-related neural activity in relation to the phase of slower oscillations reflects the order of presentation of sensory events. On one

hand brain oscillations are metronomes, and on the other hand oscillations line up percepts in temporal order. Hence, a simple way for the brain to compute relative timing between events would be to measure the *phase distance* at which fire the cell assemblies that encode each stimulus. Therefore, the relative phase should not only reflect temporal order but also the *temporal distance* between sensory events. The resulting measure of relative event timing would certainly depart from the veridical timing between sensory events. In this thesis, we tested the hypothesis that phase distance between sensory events processing reflects the *perceived timing* between these events. In other words, subjective event timing should be given by the phase position of the event-related neural response within the internal clock reference.

It is well-known that temporal perception is very variable within and across individuals. This variability both concerns explicit and implicit temporal judgments. With three experiments, we specifically questioned whether variability in explicit and implicit timing perception could be explained by phase shifts of stimulus responses within the “pacemaker” oscillation.

If these mechanisms do exist, then which oscillation serves as a temporal reference? One possibility is that the brain uses prevalent endogenous rhythms such as hippocampal theta or parieto-visual alpha. Here, following the idea that the brain tunes its reference of time to external temporal information, we tested the hypothesis that the brain uses entrainment as a mean to infer time perception in a plastic fashion. The phenomenon of entrainment is particularly interesting because (1) it is present in many ecological situations (Schroeder and Lakatos, 2009), (2) it is a situation where oscillatory activity is prominent in sensory cortices and then reliable (Regan, 1966; Rees et al., 1986; Hari, 1989; Capilla et al., 2011), (3) it is not a passive response to stimulation, as it is modulated by cognitive factors such as attention (Lakatos et al., 2008; Besle et al., 2011; Nozaradan et al., 2011; Zion Golumbic et al., 2013), or temporal expectation (Stefanics et al., 2010b; Hsu et al., 2013). Hence variability in entrainment could be observed within participants and between participants (Besle et al., 2011). As such we tested whether this variability could explain variability in timing perception.

The next four chapters will cover the following aspects:

Chapter 2 will concern the link between oscillations and explicit timing. The results confirmed that variations in the entrained oscillation in sensory cortices predict how individuals perceive audiovisual timing. In a pilot experiment we also questioned the existence of endogenous brain clocks in the absence of stimulation.

While chapter 2 focuses on the effect of slow oscillatory entrainment in the emergence of time, in Chapter 3 we questioned whether entrainment to fast rhythms could also provide a temporal code for audiovisual perception. The reported findings suggest that rhythmic audiovisual binding mechanisms are limited to low-frequency oscillations in the delta range (1-2 Hz).

In chapter 4, we investigated the role of oscillatory entrainment in the tagging of event timing in an implicit temporal task, e.g. in a speech chunking task.

Finally, Chapter 5 provides a general commentary on the role of oscillatory phase in the temporal tagging of events.

PHASE ENCODING OF EXPLICIT TIMING

2.1. INTRODUCTION

2.1.1. Motivation

In this experiment, we tested the hypothesis that the brain capitalizes on external temporal regularities and uses entrainment to build up a temporal reference frame for explicit timing. We focused in particular on audiovisual timing for the following reasons:

- Audiovisual perception is a classic situation in which the temporal binding problem is at stake. As explained in the introduction, auditory and visual inputs do not reach primary sensory areas at the same time (Heil and Irvine, 1997; Schmolesky et al., 1998). Audiovisual input timing is also highly modulated by external factors due to the different speeds between light and sound. As such, the brain should use strategies to flexibly compensate for these variable external delays. Accordingly, perceived audiovisual timing has been shown to be influenced by context (Sugita and Suzuki, 2003), experience (Fujisaki et al., 2004; Vroomen et al., 2004; Yamamoto et al., 2012), and importantly, it is strongly variable between individuals (Stone et al., 2001; Love et al., 2013).
- Neural entrainment is cross-modal. The presentation of slow auditory rhythms will entrain delta (1-3Hz) oscillations in both auditory and visual areas (Lakatos et al., 2008; Besle et al., 2011). Similarly, auditory cortex can be entrained to visual rhythms (Besle et al., 2011; Gomez-Ramirez et al., 2011). As such, low-frequency oscillatory entrainment has been proposed to support audiovisual temporal binding by driving cross-modal attentional selection (Schroeder and Lakatos, 2009).

- Auditory and visual regions are well separated in cortex, allowing the segregation of their neural signatures more easily with MEG techniques.

2.1.2. Experiment

In this MEG study, participants underwent a series of adaptations to asynchronous audiovisual events. Presenting repeatedly delayed audiovisual information is known to induce perceptual changes in audiovisual timing (Fujisaki et al., 2004; Vroomen et al., 2004; Yamamoto et al., 2012). It indicates that long rhythmic exposure to delayed stimuli modifies the brain's internal reference frame of time. Of interest, the rhythmic presentation of audiovisual stimuli at a 1 Hz frequency will likely entrain auditory and visual cortices. Therefore, we investigated whether entrainment would be a passive response to sensory stimulus, or whether it could be modulated through adaptation. Furthermore, if the entrained oscillation serves as a reference frame for audiovisual timing, it should entail that any change in the entrainment response should be commensurate with a change in audiovisual timing.

2.1.3. Summary of the results

The repeated presentation of asynchronous audiovisual stimuli led to non-stationary 1Hz entrainment in sensory cortices: the preferential phase at this frequency was significantly shifted between the beginning and the end of the adaptation period. Individuals' reports of subjective simultaneity linearly mapped onto the phase shifts of auditory but not visual neural responses (fig 2.1). Overall, the present results provide evidence that entrained oscillations play a role in the subjective estimation of audiovisual timing. This further suggests that auditory cortex recalibrates its timing to the visual response that serves as a temporal anchor.



Figure 2.1: Summary of the results. Phase shifts were observed during the beginning and the end of the adaptation period, while no difference was observed in the sensory evoked responses. These results are seen through the external reference frame centered on event onset (left panel). If we now place ourselves on the entrained oscillation reference frame (right panel), we see that evoked responses are clearly shifted in time according to this reference. The results of this experiment suggest that this shift correspond quantitatively to the perceived shift in timing between auditory and visual information.

2.2. ARTICLE

Anne Kösem, Alexandre Gramfort, Virginie van Wassenhove (accepted). Encoding of event timing in the phase of neural oscillations. Neuroimage

ENCODING OF EVENT TIMING IN THE PHASE OF NEURAL OSCILLATIONS

Anne Kösem¹⁻³, Alexandre Gramfort^{2, 4, 5}, Virginie van Wassenhove^{1-3*}

¹INSERM, U992, Cognitive Neuroimaging Unit, F-91191 Gif/Yvette, France.

²CEA, DSV/I2BM, NeuroSpin Center, F-91191 Gif/Yvette, France.

³Univ Paris-Sud, Cognitive Neuroimaging Unit, F-91191 Gif/Yvette, France.

⁴INRIA, Parietal team, Saclay, F-91191 Gif-sur-Yvette, France.

⁵LNAO, NeuroSpin, CEA Saclay, F-91191 Gif-sur-Yvette, France.

ABSTRACT

Time perception is a critical component of conscious experience. To be in synchrony with the environment, the brain must not only deal with differences in the speed of light and sound but also with its computational and neural transmission delays. Here, we asked whether the brain could actively compensate for temporal delays by changing its processing time. Specifically, can changes in neural timing or in the phase of neural oscillations index perceived timing? For this, a lag-adaptation paradigm was used to manipulate participants' perceived audiovisual (AV) simultaneity of events while they were recorded with magnetoencephalography (MEG). Desynchronized AV stimuli were presented rhythmically to elicit a robust 1 Hz frequency-tagging of auditory and visual cortical responses. As participants' perception of AV simultaneity shifted, systematic changes in the phase of entrained neural oscillations were observed. This suggests that neural entrainment is not a passive response and that the entrained neural oscillation shifts in time. Crucially, our results indicate that shifts in neural timing in auditory cortices linearly map participants' perceived AV simultaneity. To our knowledge, these results provide the first mechanistic evidence for active neural compensation in the encoding of sensory event timing in support of the emergence of time awareness.

KEYWORDS: MEG, oscillatory entrainment, temporal order, simultaneity, internal clock.

1. INTRODUCTION

While dedicated neural structures for time perception have been described (Buhusi and Meck, 2005; Coull et al., 2004; Harrington et al., 1998; Ivry and Schlerf, 2008; Morillon et al., 2009; Treisman et al., 1990; van Wassenhove, 2009; Wittmann, 2013, 2009), the encoding of sensory event timing has been proposed to result from the intrinsic dynamics of neural populations not necessarily dedicated to temporal processing (Johnston and Nishida, 2001; Karmarkar and Buonomano, 2007; van Wassenhove, 2009). For instance, the timing of a colored visual patch could be encoded in the dynamics of the neural population dedicated to the analysis of color (Karmarkar and Buonomano, 2007; Moutoussis and Zeki, 1997). In this non-dedicated view, the latency of neural responses could provide an index for event timing (Johnston and Nishida, 2001; Zeki and Bartels, 1998). Under this latency code hypothesis, timing mechanisms are based on the changes of neural routing delays in sensory areas coding for a specific sensory attribute (Moutoussis and Zeki, 1997; Zeki and Bartels, 1998). To date however, electroencephalographic (EEG) studies have reported little-to-no correspondence between neural latencies and participants' perceived timing (McDonald et al., 2005; Vibell et al., 2007), and rather suggest that it is the phase of neural oscillations that plays a crucial role in the encoding of visual event timing (Chakravarthi and Vanrullen, 2012; Gho and Varela, 1988).

We here provide further evidence that the encoding of event timing is realized in the phase of neural oscillations (in auditory cortex). It is well known that distinct phases of low-frequency neural oscillations are associated with periods of high and low neural excitability (Buzsáki, 2010; Lakatos et al., 2008). These fluctuations have been shown to impose temporal constraints on the “what” of perception by modulating the perceptual detection threshold of various stimuli (Busch et al., 2009; Fiebelkorn et al., 2013; Henry and Obleser, 2012; Monto et al., 2008; Neuling et al., 2012). They have also been proposed to serve parsing and informational chunking of sensory information over time (VanRullen and Koch, 2003) notably for complex temporal structures such as speech (Giraud and Poeppel, 2012). Indeed, neural oscillations are known to be entrained to external rhythms (Rees et al., 1986; Regan, 1966) and this entrainment may allow the alignment of cortical processing to the timing of sensory events (Giraud and Poeppel, 2012; Schroeder and Lakatos, 2009). As such, this mechanism naturally provides a means for the brain to internalize external temporal regularities (Schroeder and Lakatos, 2009). In line with this proposal, the phase of low-frequency neural oscillations has been shown to reflect temporal expectancy or predictability of event timing (Stefanics et al., 2010). Here, we hypothesize that the brain could use

oscillatory entrainment to establish a temporal reference frame and we thus ask whether the phase of entrained neural oscillations actually encodes the “when” of perception. Specifically, the preferred phase of oscillatory entrainment is known to be context-dependent (Besle et al., 2011; Gomez-Ramirez et al., 2011; Lakatos et al., 2008; Rees et al., 1986), suggesting that neural entrainment may not be a passive neural response. Additionally, preferential phases of entrained neural oscillations are subject-specific (Besle et al., 2011), making this neural index particularly well-suited for investigating the highly subjective and variable nature of time perception.

To test the specific hypothesis that the phase of an entrained neural oscillation directly informs on the variability of conscious timing, we transiently shifted participants’ perceived timing using a lag-adaptation paradigm (Fujisaki et al., 2004; Miyazaki et al., 2006; Vroomen et al., 2004). Figure 1 provides an overview of the experimental paradigm. During the induced changes of perceived timing, participants’ brain activity was recorded with magnetoencephalography (MEG). During a given lag-adaptation block, audiovisual stimuli were presented rhythmically to induce an entrainment of oscillatory activity in sensory cortices. Analysis of MEG data showed that the preferential phase of entrained neural oscillations shifted during adaptation. Crucially, phase shifts of neural oscillatory entrainment in auditory cortex mirrored individuals’ perceived simultaneity.

2. MATERIALS AND METHODS

2.1. Participants

Nineteen participants (7 females, mean age: 24 years old) took part in the study. All had normal or corrected-to-normal vision, normal color vision and normal hearing, and were naive as to the purpose of the study. Each participant provided a written informed consent in accordance with the Declaration of Helsinki (2008) and the Ethics Committee on Human Research at NeuroSpin (Gif-sur-Yvette, France). Three subjects were excluded from the study: one subject did not finish the experiment, and two were unable to perform the temporal order judgment task properly. A total of sixteen participants were thus considered for MEG analyses.

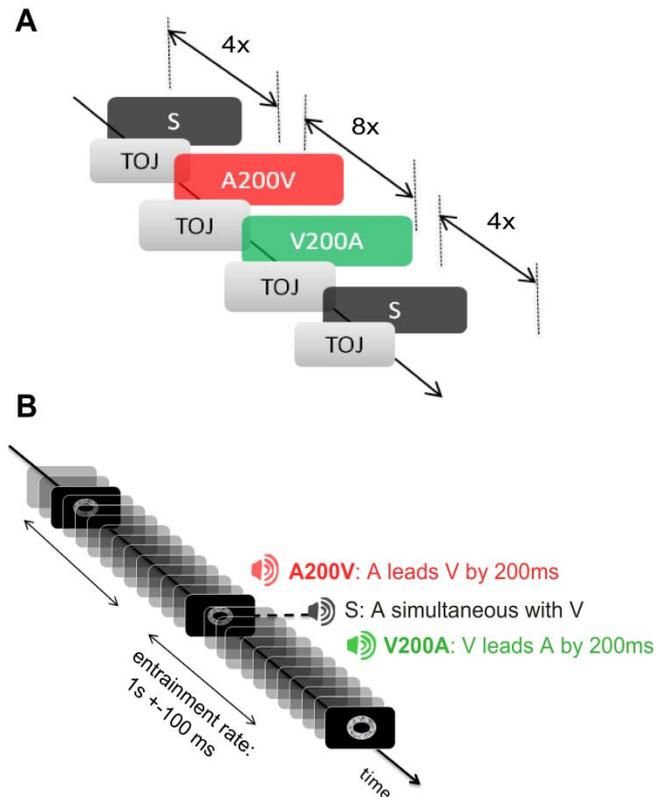


Fig. 1: Experimental design. (A) Three different audiovisual (AV) lag-adaptations were tested: simultaneous AV presentation (S, black), sound leading visual by 200 ms (A200V, red) and visual leading sound by 200 ms (V200A, green). Each lag-adaptation block was followed by a temporal order judgment (TOJ) block during which participants reported which of the auditory or visual event occurred first. One MEG session comprised eight blocks of each lag-adaptation (S, V200A and A200V). S blocks were run at the beginning and at the end of the MEG session; A200V and V200A alternated within the session. (B) In all lag-adaptation blocks, 65 AV stimuli were presented at an average rate of 1 Hz with a random jitter of ± 100 ms. This experimental manipulation was designed to elicit neural entrainment at 1 Hz in sensory cortices.

2.2. Stimuli

The experiment was written in Matlab using the Psychophysics toolbox (Brainard, 1997). Visual stimuli consisted of disks lasting 16.7 ms (1 frame). A visual annulus (9.5° of visual angle) consisted in the superposition of circles with different shades of gray. Visual stimuli were projected at a 60Hz refresh rate onto a screen placed 90 cm away from participants seated under the MEG dewar. Auditory stimuli consisted of 16 ms duration white noise (incl. 5 ms fade-in and fade-out). Auditory stimuli were presented via Etymotic earphones (Etymotic Research Inc., USA).

2.3. Procedure

Two types of blocks were used in this experiment namely, lag-adaptation (3 conditions: S, A200V or V200A) and temporal order judgment (TOJ) blocks. In the lag-adaptation blocks (Fig. 1b), a series of simultaneous (S) or desynchronized audiovisual events were displayed (A200V: audition leading vision by 200 ms or V200A: vision leading audition by 200 ms). During the lag-adaptation block, a stream of 65 AV stimuli was presented. The stream of AV events was displayed at an average rate of 1 Hz; the stimulus onset asynchrony (SOA) between two successive auditory or visual stimuli was randomly chosen from a normal distribution with a mean of 1s and a standard deviation of 100 ms: thus, each SOA has 95% probability to fall between 804 ms and 1196 ms. The first 20 AV events and the last 15 AV events in the stream were made up of stimuli with a constant temporal lag. Three lags were tested: in the S condition, AV stimuli were synchronously displayed (lag = 0 ms); in the A200V condition, the sound preceded the visual stimulus by 200 ms and in the V200A condition, the visual stimulus preceded the sound by 200 ms. During the lag-adaptation block, participants were asked to count the number of temporal deviants that were introduced in the middle part of the lag-adaptation block. Temporal AV deviants consisted of desynchronized AV stimuli that deviated from the constant lag introduced at the beginning of the block. This task was introduced to insure participants attended the temporal dimension of the AV stream which was reported to enhance temporal recalibration effects (Heron et al., 2010). Crucially however, only the first 20 and last 15 AV stimuli are reported here namely the periods during which no temporal deviants were introduced. Each lag-adaptation block was systematically followed by a TOJ in which participants' subjective simultaneity of AV events was evaluated. In the TOJ blocks, AV stimuli were displayed with delays ranging from +/-317, +/-217, +/-133, +/-67, and 0 (a negative delay corresponds to the auditory leads and a positive delay corresponds to visual leads). After each presentation of an AV pair, participants had to judge which of the sound or the visual event appeared first in a two alternative forced choice (2-AFC). Each condition was tested four times leading to a total of 36 trials in the TOJ blocks. The experiment started and ended with 4 S blocks (i.e. 4 times S + TOJ). Other blocks were alternated between A200V and V200A condition. In total, each condition was run in 8 lag-adaptation + TOJ blocks leading to a total of 24 blocks (Fig. 1a).

2.4. Data acquisition and preprocessing

2.4.1. MEG data acquisition

Brain magnetic fields were collected in a magnetically shielded room using the whole-head Elekta Neuromag Vector View 306 MEG system (Neuromag Elekta LTD, Helsinki) equipped with 102 triple-sensor elements (two orthogonal planar gradiometers and one magnetometer per location). Participants were seated in upright position. Participants' head position was measured before each block with four head position coils (HPI) placed over frontal and mastoid areas. Three fiducial points (nasion, left and right pre-auricular areas) and were used during the digitization procedure to help coregistration with anatomical MRI. MEG recordings were sampled at 1 kHz; band-pass filtered between 0.03 Hz and 330 Hz and used Maxshield. The electro-oculograms (EOG, horizontal and vertical eye movements) and electrocardiogram (ECG) were simultaneously recorded with MEG. Before each experiment, a so-called empty room recording of about 1 minute with no subject sitting under the dewar was acquired for the computation of the noise covariance matrix.

2.4.2 MEG data preprocessing

Signal Space Separation (SSS) method was applied to decrease the impact of external noise (Taulu et al., 2003). SSS correction, head movement compensation, and bad channel rejection was done using MaxFilter Software (Elekta Neuromag). Signal-space projection (SSP) were computed by principal component analysis (PCA) using Graph software (Elekta Neuromag) to correct for eye-blinks and cardiac artifacts (Uusitalo and Ilmoniemi, 1997). A rejection criterion for epochs was applied for gradiometers with amplitude exceeding $4000 \text{ e}^{-13}\text{T/m}$.

2.4.3. Structural MRI acquisition

Magnetic Resonance Imaging (MRI) was used to provide high-resolution structural image of each individual's brain. The anatomical MRI was recorded using a 3-T Siemens Trio MRI scanner. Parameters of the sequence were: voxel size: 1.0 x 1.0 x 1.1 mm; acquisition time: 466 s; repetition time TR = 2300 ms; and echo time TE= 2.98 ms.

2.4.5 Anatomical MRI segmentation

Volumetric segmentation of participants' anatomical MRI and cortical surface reconstruction

was performed with the FreeSurfer software (<http://surfer.nmr.mgh.harvard.edu/>) (Dale et al., 1999; Fischl and Dale, 2000). These procedures were used for group analysis with the MNE suite software (<http://www.martinos.org/mne/>). Individuals' current estimates were registered onto the FreeSurfer average brain for surface based analysis and visualization.

2.4.6. Co-registration procedure (MEG-aMRI)

The co-registration of MEG data with the individual's structural MRI was carried out by realigning the digitized fiducial points with MRI slices. Using `mne_analyze` within the MNE suite, digitized fiducial points were aligned manually with the multimodal markers on the automatically extracted scalp of the participant. To insure reliable co-registration, an iterative refinement procedure was then used to realign all digitized points (about 30 more supplementary points distributed on the scalp of the subject) with the individual's scalp.

2.5. Data analysis

2.5.1. MEG source reconstruction

Individual forward solutions for all source locations located on the cortical sheet were computed using a 3-layers boundary element model (BEM) (Hämäläinen and Sarvas, 1989) constrained by the individual's anatomical MRI. Cortical surfaces extracted with FreeSurfer were sub-sampled to about 5,120 equally spaced vertices on each hemisphere. The noise covariance matrix for each individual was estimated from the raw empty room MEG recordings preceding the individual's MEG acquisition. The forward solution, noise covariance and source covariance matrices were used to calculate the dSPM estimates (Dale et al., 1999). The inverse computation was done using a loose orientation constraint (loose = 0.2, depth = 0.8) (Lin et al., 2006). The cortically constrained reconstructed sources were then registered, morphed, onto the FreeSurfer average brain for group-level statistical analysis that was performed with MNE-python (Gramfort et al., 2013a, 2013b).

2.5.2. Labels of interest

We restricted the analysis to labels of interest in auditory and visual sensory cortices in the right hemisphere on the average Freesurfer brain after morphing. Known hemispheric asymmetries in auditory cortex have consequences on the signal-to-noise ratio of MEG recordings across hemispheres (Shaw et al., 2013). Consistent with this, a great majority of our participants showed a higher and more reliable SNR in the right hemisphere. Labels were drawn individually based on the following two criteria: (i) maximal amplitude of the M100 response to auditory (resp. visual) stimulus for the auditory (resp. visual) label; (ii) consistency with functional anatomy. Individuals' labels are presented in figure S1 superimposed on the Freesurfer average brain.

2.5.3. Event-related fields and source reconstruction

Event-related fields (ERF) were computed by averaging 15 trials at the beginning and at the end of a lag-adaptation block. Data were gathered across the 8 lag-adaptation blocks for each asynchrony condition (S, A200V, V200A). For auditory ERF, the stimulus onset was locked to the sound onset; for visual ERF, the stimulus onset was locked to the visual stimulus. Data were segmented in epochs of 1s (400 ms pre- and 600 ms post-stimulus onset). Baseline correction was applied using the first 200 ms of the epoch (-400 to -200 ms pre-stimulus onset). The inverse solver used to localize the sources was then applied on the averaged normed evoked data. The normalization procedure was done to alleviate source cancellation when averaging sources within a label of interest, and across subjects (Gross et al., 2013). The comparisons of evoked responses between conditions were computed using a non-parametric permutation test. Correction for multiple comparisons was performed with cluster level statistics using as base statistic Student t-test computed at each time sample (Maris and Oostenveld, 2007). Only temporal clusters with corrected p-value ≤ 0.05 are reported.

2.5.4. Power spectrum analysis

Low-frequency components in the frequency spectra could either originate from neural entrainment to the 1Hz stimulation or from noise having a power spectrum density with 1/f distribution. To substantiate a peak neural entrainment at 1Hz, the 1/f component was removed by subtracting at each frequency bin the mean power of the neighboring frequency values (4 frequency values were: $[f_0 - 0.14\text{Hz} ; f_0 - 0.07 \text{ Hz}; f_0 + 0.07 \text{ Hz}; f_0 + 0.14 \text{ Hz}]$) (Nozaradan et al., 2011).

2.5.5. Phase analyses

The first 5 trials of each lag-adaptation block were discarded as they established transient episodes before establishment of the steady-state regime (Capilla et al., 2011). Hence, an equal number of 15 trials (or stimuli) at the beginning and at the end of an lag-adaptation period were considered for analysis. Single trial data were convolved with a 3-cycle Morlet wavelet centered at 1 Hz with a full width at half maximum of the power in the frequency domain of [0.7 Hz, 1.3 Hz] (Keil et al., 2013) (Fig. S1). Epoch lengths were 4s and centered on the visual or on the auditory stimulus onset. From the coefficients obtained with wavelet convolution the instantaneous phase at visual or auditory onset was extracted (as indicated in text where relevant). Subsequent analyses were done on the distribution of phase values across trials gathered across specific condition, namely: beginning or end of a lag-adaptation period, conditions S, A200V or V200A and for each participant.

2.5.6. Phase uniformity test

Phase distributions were submitted to Rayleigh's test for uniformity of phase data (Fisher, 1995). A significant Rayleigh test ($p \leq 0.05$) indicates that the distribution of phases show a phase preference. If significance was reached, the circular mean of each distribution was computed and used for phase/ERFs and phase/behavior correlation analyses.

For each participant and each condition, a Rayleigh test was performed. If all three conditions passed the test, entrainment was considered to be true for that participant. When separating trials between the beginning and the end of a lag-adaptation period, a criterion of two out of three conditions passing the Rayleigh test was considered evidence for entrainment.

The Phase-Locking Value (PLV) (Lachaux et al., 1999) is defined as:

$$PLV(t) = \frac{1}{K} \left| \sum_{k=1}^K e^{j\theta(t,k)} \right|$$

where K is the number of trials, and $\theta(t,k)$ is the instantaneous phase at time t and trial k . PLVs were computed to assess intra-subject variability in the preferential phase.

2.5.7. Statistical comparison of phase distributions

To assess statistical significance of phase shifts between 2 conditions (e.g. A and B), bootstrap measure of 95% confidence interval was used on the phase distribution of the paired differences A-B (Fisher, 1995). Phase distributions A and B were statistically different if the mean of the difference was statistically different from zero, i.e. if zero lies outside the measured confidence interval ($p \leq 0.05$).

2.5.8. Psychophysics - Point of Subjective Simultaneity

The percentage of “visual first” responses during the TOJ task following the lag-adaptation period (S, V200A, and A200V) were plotted as a function of AV asynchrony and fitted with a logistic regression to a sigmoid function of the form:

$$\text{fit} = \frac{1}{1 + \exp\left(\frac{\text{data} - \text{PSS}}{\text{JND}}\right)}$$

From each individual fit, the Point of Subjective Simultaneity (PSS) value and the Just Noticeable Difference (JND) were estimated. The PSS corresponds to the AV asynchrony at which an individual responds at chance level (50%) in a TOJ task and thus taken as a true subjective simultaneity estimate (Vroomen and Keetels, 2010).

3. RESULTS

Participants underwent a series of alternating lag-adaptation and TOJ blocks while being recorded with MEG. Three AV delays were tested (Fig. 1a-b): simultaneous AV presentations (S, control condition), a sound leading a visual event by 200 ms (A200V) and a visual event leading a sound by 200 ms (V200A). Each TOJ block allowed establishing an individual’s psychometric curve following each lag-adaptation block as well as deriving the progression of the individual’s Point of Subjective Simultaneity (PSS). Our hypothesis was that changes in neural activity during lag-adaptation would predict changes in subjects’ perceived simultaneity. First, we tested the latency code hypothesis by comparing the event-related responses at the beginning and at the end of the lag-adaptation. We then tested the phase code hypothesis by comparing the phase of the entrained neural oscillation at the beginning and the

end of the adaptation (Supplementary Fig. S1).

3.1. Stable evoked activity in sensory cortices during adaptation

Auditory and visual Event-Related Fields (ERFs) were source-reconstructed. The resulting time-source series were separately averaged in the auditory and visual labels. In all three lag-adaptation conditions, visual evoked responses were comparable with no significant modulations of the visual evoked responses by auditory stimuli (Fig. 2a-c). To the contrary, clear modulations of the auditory evoked responses by the presence of visual stimuli were observed (Fig. 2d-f) in both A200V and V200A. The auditory evoked response profiles in A200V and V200A

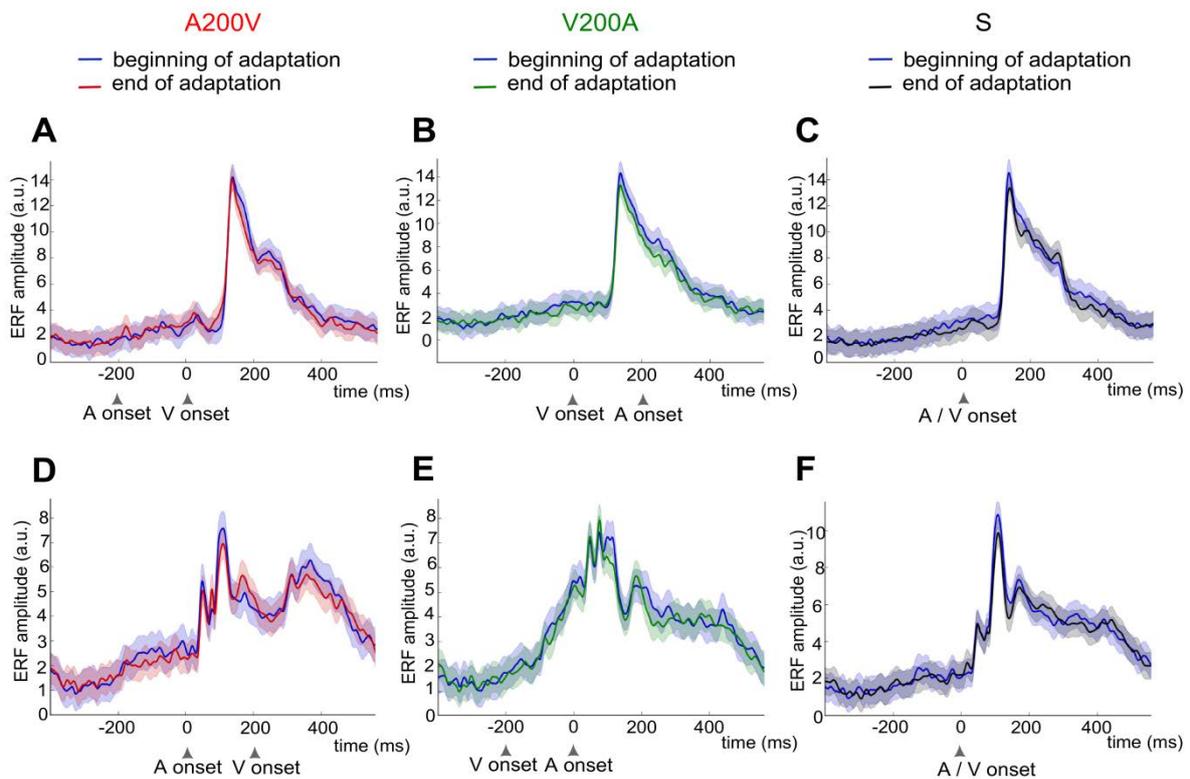


Fig. 2: Visual and auditory evoked responses before and after lag-adaptation. Auditory and visual evoked responses were obtained by separately time-averaging the first 15 trials at the beginning (blue) and at the end (A200V: red; V200A: green; S: black) of a given lag-adaptation block. Evoked responses were time-locked to the visual and auditory onsets in visual (A-C) and auditory cortices (D-F), respectively. Overall, no significant differences in the auditory or visual Evoked responses profiles were observed in the course of the lag-adaptation blocks, irrespective of the condition. Additionally, no significant influence of auditory stimuli was observed in the visual ERFs (A-C). In contrast, auditory Evoked responses in A200V and V200A significantly differed ($p \leq 0.05$): in V200A, a ramping modulation of the auditory Evoked responses was observed ~ 70 ms prior to sound onset whereas in A200V, a significant modulation of the auditory Evoked responses at ~ 300 ms post- sound onset was observed.

significantly differed ($p \leq 0.05$): in V200A, a ramping modulation of the auditory evoked response was observed ~ 70 ms prior to sound onset whereas in A200V, a significant modulation of the auditory evoked response at ~ 300 ms post- sound onset was observed. The modulations of the auditory evoked responses remained steady throughout the lag-adaptation block and did not significantly differ between the beginning and the end of a given lag-adaptation block. If, as hypothesized, changes in perceived timing were caused by changes in the neural timing of auditory and visual cortices during lag-adaptation, the latency of evoked activity did not appear to be a good candidate to capture this change. These observations are in agreement with previous findings on evoked-related-potential literature (McDonald et al., 2005).

3.2. Non-stationarity of the entrained neural oscillations during lag-adaptation

As predicted by the rate of AV stimulation during lag-adaptation, neural activity over long time scales displayed periodic fluctuations at 1Hz i.e. oscillatory entrainment or frequency-tagging: a characteristic frequency peak at 1Hz was clearly observable in auditory and visual power spectra (Fig. 3b, d) and in single-trials data (Fig. 3a, c). No significant changes in 1 Hz power were found between the beginning and the end of a given lag-adaptation period irrespective of the experimental condition.

Additionally, 1Hz oscillatory activity showed a significant phase-locking in both sensory cortices. Phase preferences of the 1 Hz oscillation were tested using a Rayleigh test against uniformity ($p < 0.05$). At the beginning of all lag-adaptation periods (S, A200V, and V200A), phase preferences were found to be significant in both sensory cortices for 15 participants; at the end of all lag-adaptation periods (S, A200V, and V200A), significant phase preferences were found for all participants. For all conditions and in both sensory cortices, the observed Phase Locking Values (PLV, index the variance of phase distributions) did not significantly differ between the beginning and the end of each lag-adaptation period (Supplementary Table S1). Altogether, these results show the existence of robust phase preferences in all conditions throughout the course of the experiment.

However, and unlike the power of the entrained 1Hz oscillation, the phase of the 1Hz oscillatory component did not appear to be stationary over the course of the lag-adaptation

period. Specifically, the neural responses evoked by the stimulus presentation arrived at different phases of the 1Hz oscillation (Fig. 4a, b). In the non-zero lag adaptations (A200V and V200A), phases of the 1 Hz oscillation shifted in opposite directions in visual and auditory cortices whereas in the control condition (S), no phase shifts were observed (Fig. 4c). Within a given block, stimuli presented at the beginning of the adaptation were identical to those presented at the end: as

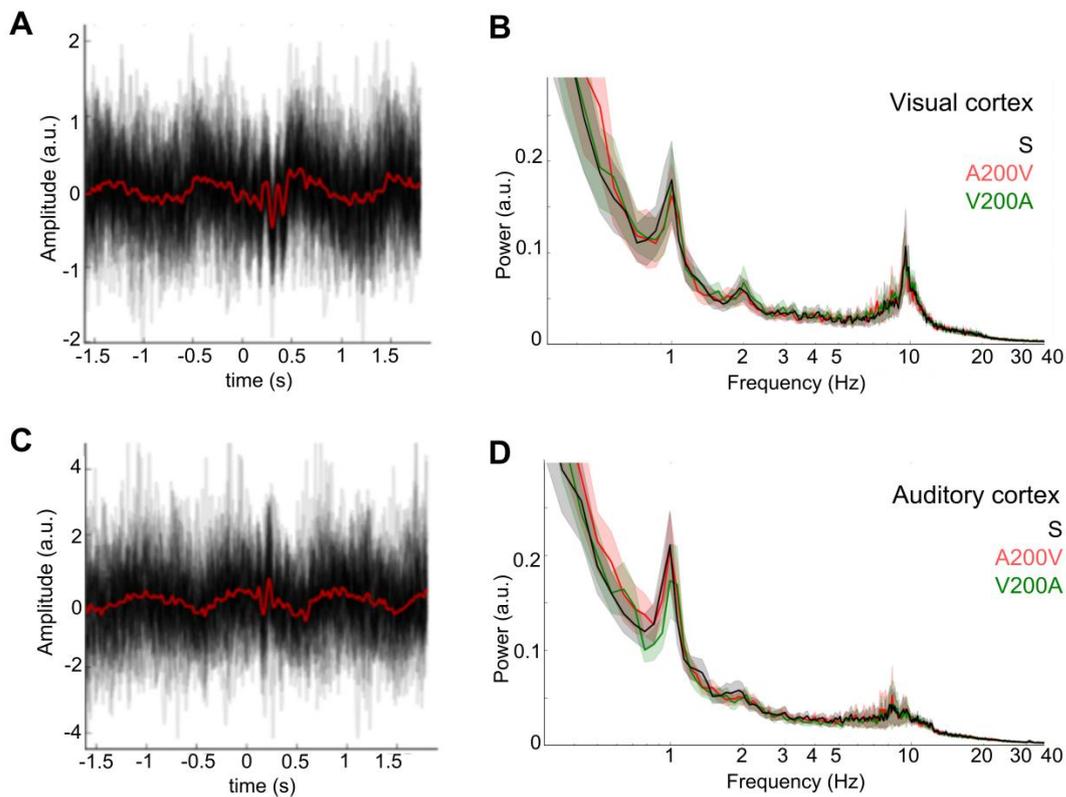


Fig. 3: Oscillatory entrainment and jitter procedure. In all lag-adaptation blocks, AV stimuli were presented at a rate of 1 Hz +/- 100 ms. The temporal jitter was introduced to prevent full neural response time-locking and enable the dissociation of the oscillatory component from the evoked response (Lakatos et al., 2008). In both (A) and (C): one participant's superimposed single trial data (black) with the visual evoked response (red) at the end of A200V in visual cortex (A) and in auditory cortex (C). As can readily be seen in these two examples, the time-locked averaging of one stimulus (at zero) prevented to see the evoked responses of the preceding and following stimuli (which would be expected at about -1 and +1 sec, respectively). As predicted, the temporal jittering procedure massively reduced the temporally adjacent evoked responses. Despite the absence of evoked response at -1 and +1 s, a clear single-trial oscillatory component at 1 Hz could be seen in both sensory cortices. **In both (B) and (D):** Frequency power spectra of neural responses in visual (B) and auditory (D) cortices for all lag-adaptation blocks. After 1/f correction (see section 2.5.4.), a significant 1 Hz peak was readily observable (on sample t-test against $H_0 = \text{zero power}$, $p < 0.01$) in both sensory cortices. No significant differences in 1 Hz power were observed between the beginning and the end of the lag-adaptation periods irrespective of the experimental conditions. In visual cortices (panel B): S: $t(15) = -0.9$, n.s.; A200V: $t(15) = 0.6$, n.s.; and V200A: $t(15) = -0.7$, n.s.. In auditory cortices (panel D): S: $t(15) = -1.8$, n.s.; A200V: $t(15) = 0.05$, n.s.; V200A: $t(15) = 1.7$, n.s..

previously reported, and consistent with the steady stimulation, no significant differences were observed when contrasting the evoked responses at the beginning and at the end of a given lag-adaptation period (Fig. 2). If neural entrainment were a passive neural response, no changes in the preferential phase would be predicted. Hence, the observed phase shifts in the entrained oscillatory response suggest an active modulation of the entrained 1 Hz oscillation not easily accounted for by the unchanged event-related responses.

Additional analyses were performed supporting the independence of evoked activity and neural oscillatory phase shifts. If the evoked responses impacted the phase of neural oscillations at 1Hz, a similar pattern of phase shifts in neighboring frequency regions should be found by virtue of evoked response being fixed-latencies and strongly phase-locked signals. Weak-to-no phase locking and no significant phase shifts were observed for 2 or 3 Hz neural oscillations (Supplementary Fig. S2).

3.3. Encoding of subjective timing in the phase of neural oscillations

Our main hypothesis states that shifts in the phase of neural oscillations during adaptation may reflect active changes in subjective timing. To test this, we compared the shifts in the phase of neural oscillations with perceptual reports (Fig. 5a, b). Consistent with previous reports on TOJ paradigms (Love et al., 2013; Van Eijk et al., 2008), the average PSS value in the zero-lag adaptation condition (S) was biased towards sound-leading asynchronies: on average, participants required the auditory event to lead the visual event by 38 ms to consider them as simultaneous (Fig. 5b). Following lag-adaptation to A200V and V200A, participants required the sound to lead the visual event even more with PSS values of -87 ms (significant main effect of lag-adaptation: $F(2,30) = 10.1, p < 0.001$, significant contrast $PSS_{A200V} - PSS_S: p = 0.002$) and -69 ms ($PSS_{V200A} - PSS_S: p = 0.03$), respectively (Fig. 5b). Thus, in both A200V and V200A, the sound needed to be heard before the visual event to be perceived as simultaneous, but shifts in perception were more pronounced in A200V than in V200A ($PSS_{A200V} - PSS_S, p = 0.01$). From a neural processing point of view, these results suggest that auditory analysis may be delayed during lag-adaptation and/or visual analysis advanced in time. As seen in Fig. 4c, the average phase shifts during A200V and V200A lag-adaptations were consistent with these predictions: lag-adaptation lead to a negative shift of the phase of auditory entrained oscillation i.e. the 1 Hz response in auditory cortex shifted forward in time; conversely, the phase of the visual entrained oscillation positively shifted suggesting that the

1Hz entrained response shifted backward in time.

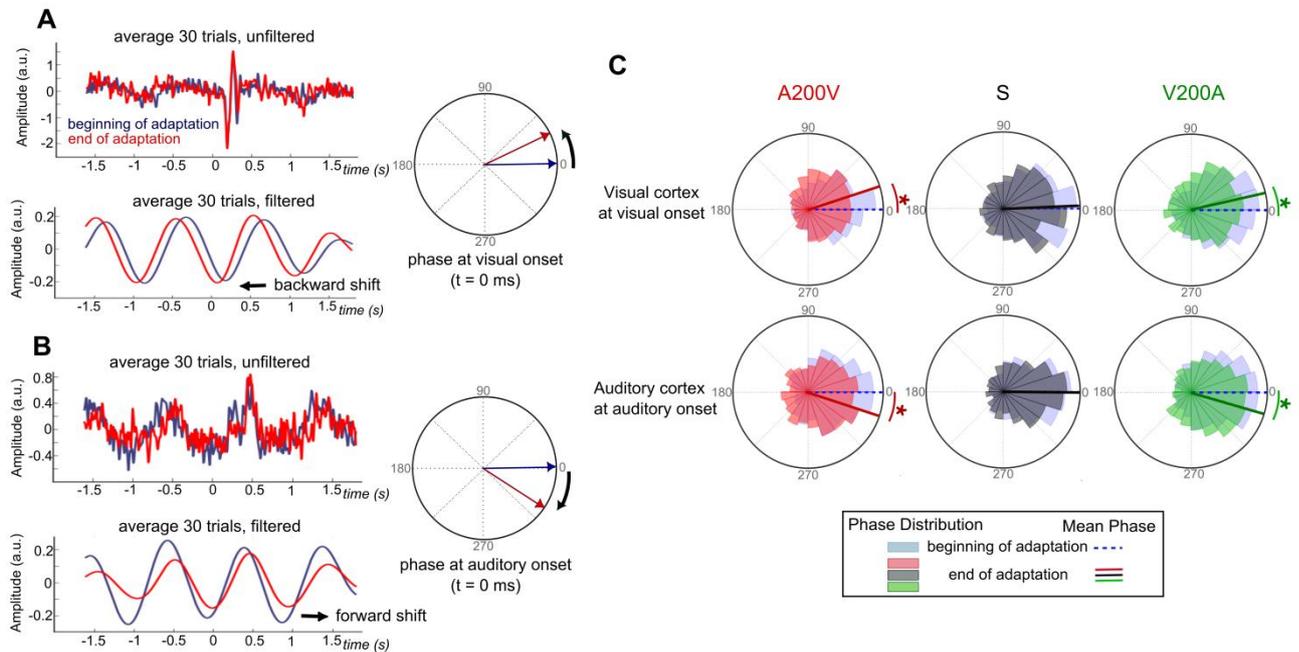


Fig. 4: 1 Hz neural response: oscillatory phase differences during lag-adaptation. (A) One participant's visual evoked response obtained at the beginning (blue) and at the end (red) of an A200V lag-adaptation block. The top graph shows the unfiltered visual evoked response; the bottom graph shows the same visual evoked response band-pass filtered from 0.5 to 1.5 Hz. The 1 Hz oscillatory component at the end of the lag-adaptation block (red) shows a backward shift in time with respect to the same oscillatory component at the beginning of the lag-adaptation (blue): this backward temporal shift is quantified as an increase in the mean instantaneous phase value across the 30 single trials used to compute the evoked response (right panel). (B) One participant's auditory evoked response at the beginning (blue) and at the end (red) of an A200V lag-adaptation block. The unfiltered and 0.5-1.5 Hz band-pass filtered auditory evoked responses are depicted in the top and bottom graph, respectively. The 1 Hz oscillatory component at the end of the lag-adaptation block (red) shows a forward shift in time with respect to the same oscillatory component at the beginning of the lag-adaptation (blue). The mean instantaneous phase across the 30 trials used to build the auditory evoked response show a decrease between the beginning and the end of the block (right panel). (C) Instantaneous phase distribution and preferential instantaneous phase of the entrained 1 Hz neural oscillatory response at the beginning (light gray) and at the end (colored) of a given lag-adaptation block (S: black; A200V: red; V200A: green). Phase distributions were computed at visual onset in visual cortices (top) and at sound onset in auditory cortices (bottom). Phase distributions were individually normalized to the preferred instantaneous phase observed at the beginning of a given lag-adaptation block. Hence, all phase distributions at the beginning of a given block are centered on zero. In S, the phase distributions remained stable over time in both auditory (+1°, 95% confidence interval (CI) = [-6°, +5°]) and visual (-1°, CI = [-6°, +8°]) cortices. In the desynchronized blocks, the mean instantaneous phase shifted in opposite directions in the auditory and visual cortices during lag-adaptation: specifically, in A200V the preferential phase in visual cortices increased (+19° or -53 ms, CI = [12°, 26°]) suggesting a backward shift in time of the entrained 1 Hz oscillatory response, whereas a forward shift in time was observed in auditory cortices (-19° or +53 ms, CI = [-27°, -9°]). Conversely, in V200A the mean instantaneous phase in visual cortices increased (+16° or -44 ms, CI = [7°, 25°]) but decreased in auditory cortices (-19° or +53 ms, CI = [-26°, -13°]). Hence, and as predicted, lag-adaptation to simultaneous AV stimuli (S) did not affect the phase of the entrained 1 Hz neural oscillation in sensory cortices whereas desynchronized AV stimuli (A200V and V200A) shifted the preferential phase distribution in opposite direction in the auditory and visual cortices.

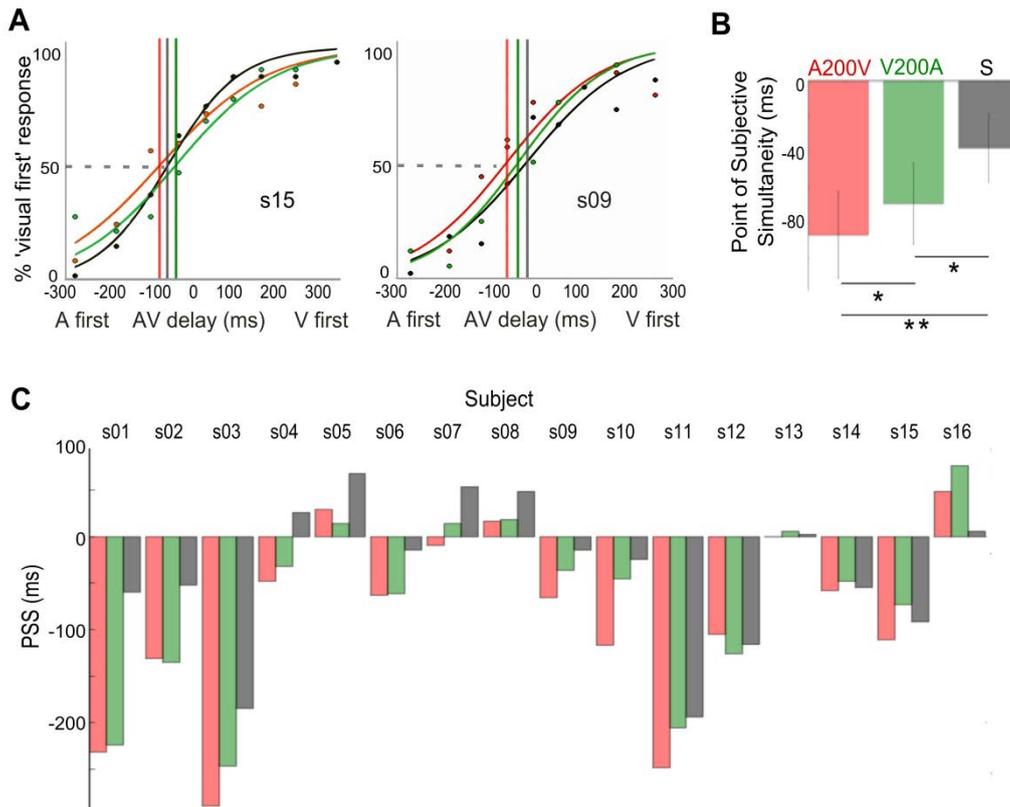


Fig. 5: Psychophysical results. (A) Two individuals' psychometric curves fitted to the percentage of "visual first" responses as a function of AV delays in the TOJ blocks. The negative and positive AV delays are audio leads (visual lags) and audio lags (visual leads), respectively. The individual's Points of Subjective Simultaneity (PSS) observed after each lag-adaptation block are indicated by a colored line: TOJ obtained in S, A200V and V200A are in black, red and green, respectively. (B) Grand average PSS ($n=16$). The control condition S showed a mean TOJ value of -38 ± 19 ms, suggesting that on average a sound had to be presented about 38 ms before a visual event to be perceived as synchronous. Following lag-adaptations to A200V and V200A, significant shifts of PSS towards audio leads were observed compared to the control condition (A200V: -87 ± 24 ms; V200A: -69 ± 23 ms). Errors bars reflect s.e.m. (C) Summary of all individuals' PSS. As can readily be observed, a large inter-individual variability was obtained in the individuals' PSS values (control blocks (S)) and in the propensity of a given individual to temporally shift his or her natural PSS.

It is noteworthy that while the direction of the PSS shifts observed after lag-adaptation to A200V was consistent with seminal reports on temporal recalibration (Fujisaki et al., 2004; Vroomen et al., 2004), the PSS following V200A lag-adaptation did not shift in the direction predicted by seminal temporal recalibration effects. Nevertheless, the latter finding remains consistent with other lag-adaptation reports (Miyazaki et al., 2006; Yamamoto et al., 2012) and major differences in experimental design including the choice of task and the absence of re-adapting trials in the TOJ assessment task could account for these differences (Cai et al., 2012; Yamamoto et al., 2012). Specifically, remote and recent stimulation histories are known to bias in opposite ways the perception of incoming stimuli (Chopin and Mamassian, 2012)

and consistent with this, PSS shifts are distinctly influenced by the presence or the absence of re-adapting trials (Cai et al., 2012). Here, we did not include re-adapting trials during the TOJ blocks and stimuli were randomly chosen between -316 ms and + 316 ms, hence the primary influence on lag adaptation was the remote stimulation history i.e. the lag-adaptation trials.

Here, we used lag-adaptation to generate shifts in an individual's subjective timing. As expected, a large inter-individual variability was observed in the individual's default PSS (control S, zero-lag adaptation) and in the propensity of an individual to temporally adapt to desynchronized AV stimuli (Fig. 5a, c). In order to test whether the shifts in the phase of the 1 Hz neural oscillation were commensurate with subjective simultaneity, we capitalized on this inter-individual variability and compared individuals' PSS with the shifts in the entrained neural oscillation in each sensory cortex. Strikingly, the phase shifts of the 1Hz auditory oscillatory neural response significantly correlated with participants' subjective simultaneity whereas no such correlations were observed in visual cortices (Fig. 6). Specifically in auditory cortex, the more negative the phase of 1 Hz neural oscillation, the more the sounds needed to lead visual events to be perceived as simultaneous. Thus a negative shift in phase, which corresponds to a forward neural timing i.e. a delay of 1Hz auditory activity in time, is associated with a shift in perceived simultaneity towards auditory-leads asynchronies. Note that shifts in perception and shifts in neural timing change together in coherent directions: a delay in auditory processing as measured by a negative phase shifts should correspond to a perceived delay in auditory event timing; thus to perceive simultaneity the sound needs to be advanced in time i.e. the PSS is shifted towards auditory-leads asynchronies. Conversely, a positive phase shift moves the processing of auditory events backward in time thereby sounds have to lag visual events to be perceived as simultaneous. As reported in Figure 6, the slope of the regression between perceived timing shift and neural timing shift was 1.2, suggesting that neural timing and perceived timing shifts are quantitatively similar. Additionally, the obtained regression predicts that a zero-phase shift in the 1 Hz neural oscillation observed in auditory cortex (i.e. stationarity or stable phase preference through time) should map onto a PSS of -45 ms. This value was very close to the mean PSS obtained experimentally in the control condition S (namely, -38 ms).

Overall these results suggest that AV simultaneity relies on asymmetrical cross-talks between auditory and visual sensory cortices, namely: auditory cortices actively adjust the timing of auditory events to match that of visual inputs.

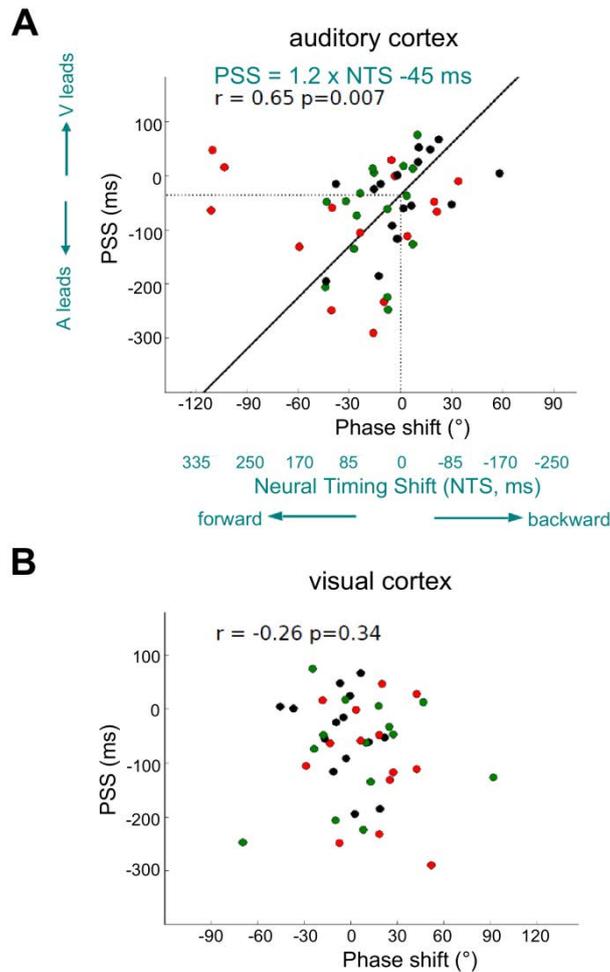


Fig. 6: Phase shifts reflect subjective timing. Individuals' PSS were plotted as a function of the difference in mean instantaneous phase (end minus beginning in a given lag-adaptation period) in auditory (A) and visual cortices (B). Each data point corresponds to an individual duplet, namely: the mean phase difference obtained in a given lag-adaptation period (S: black; A200V: red; V200A: green) and the associated individual's PSS measured during the following TOJ block. A linear regression was computed on a per individual basis between the mean PSS and the circular mean of the instantaneous phases of the 1 Hz neural oscillation obtained across lag-adaptation blocks. A significant correlation was found between the phase shifts of the entrained 1 Hz neural oscillation in auditory cortices and individuals' PSS ($r = 0.65$, $p < 0.01$) whereas no such correlation was found in visual cortices ($r = -0.26$, n.s.).

4. DISCUSSION

Shifts in perceived AV simultaneity following lag-adaptation (Fujisaki et al., 2004; Miyazaki et al., 2006; Vroomen et al., 2004; Yamamoto et al., 2012) have been hypothesized to originate from mechanisms capable of adjusting the neural processing time across sensory modalities (Fujisaki et al., 2004; Moutoussis and Zeki, 1997; Stone et al., 2001; Sugita and Suzuki, 2003; Zeki and Bartels, 1998). In support of this hypothesis, our study reveals that such mechanisms may be implemented as phase shifts of neural oscillations: contrasting the

sensory responses before and after AV lag-adaptation provided no evidence for a latency code hypothesis and instead revealed significant phase shifts of the entrained 1 Hz neural oscillations. Crucially, it is the phase shifts of the auditory response that linearly predicted participants' shifts of subjective AV simultaneity and no systematic mapping was found between visual responses and subjective AV timing. The present findings thus suggest that auditory cortex temporally calibrates its window of analysis with respect to vision and that event timing linearly maps onto the phase of entrained neural oscillations.

4.1. Neural oscillations as pacemakers for the encoding of time

The “internal clock” is a prominent model of time perception which is classically composed of a pacemaker (ticking mechanism), an accumulator (of ticks) and a counter (Church, 1984; Treisman, 1963). Of particular interest here, the pacemaker consists of an oscillator ticking at a frequency that can be modulated depending on the temporal properties of sensory stimuli (Buhusi and Meck, 2009; Treisman, 1984; Treisman et al., 1992, 1990): specifically, external temporal regularities can impose modulations of the pacemaker frequency so as to entrain the internal clock (Treisman et al., 1992). Similarly, intrinsic neural oscillations match the temporal scales of perceptual phenomena (Buzsáki and Draguhn, 2004; Roopun et al., 2008; van Wassenhove, 2009; Wang, 2010) and can be entrained to external rhythms (Rees et al., 1986; Regan, 1966). As such, neural oscillations have been hypothesized as natural pacemakers for conscious time estimation (Buhusi and Meck, 2005; Pöppel, 1997; Treisman et al., 1990; Varela et al., 1981). However, within this framework, a major problem for the brain is to determine *when* events occur with respect to its internal frame of reference. Our results suggest that the timing of events could automatically be encoded in the phase of a recruited pacemaker or entrained oscillation (thereby acting as a temporal frame of reference for cortex) and that the variation of the pacemaker's phase over time results in variation of perceived timing.

4.2. A canonical role for the phase of intrinsic and entrained neural oscillations?

It is noteworthy that we specifically targeted the delta range using neural entrainment or frequency-tagging. EEG studies have previously tested the idea that the order of visual events were coded in the phase of the alpha oscillatory component (Gho and Varela, 1988) and recent studies have pointed out to the role of theta/alpha in temporal visual illusions

(Chakravarthi and Vanrullen, 2012; VanRullen et al., 2006). These studies suggest that intrinsic oscillations are recruited for the encoding of events in the absence of external temporal regularities. Recent hypotheses further extend the notion that the phase of low-frequency neural oscillations is crucial for the encoding of order - for instance with the implication of the theta band in working memory (Lisman & Jensen, 2013) - or even for temporal parsing - for instance, in speech (Giraud & Poeppel, 2012). The temporal encoding mechanisms described in our experiment are *de facto* constrained by the rhythmicity of the external inputs; in turn, however, the encoding of event timing may capitalize on the temporal features provided by external stimulation to build a temporal reference frame or pacemaker consistent with the rhythms provided by the external sensory world. While delta oscillations have been previously linked to temporal predictability (Stefanics et al., 2010), further investigations need to be done to test their implication in the encoding of event timing when no rhythmic stimulation or external temporal reference frame is provided.

4.3. Evoked activity and attention to time

Latency-based descriptions of cognitive functions classically use event-related potentials / fields (ERP and ERF, respectively) to describe the timing at which mental operations take place in cortex (Coles and Rugg, 1995; Madl et al., 2011). The auditory and visual evoked responses were thus expected to partly reflect participants' perceptual shifts in AV simultaneity taking place during lag-adaptation. However, and surprisingly, no significant changes in the amplitude or in the latency of the evoked responses were observed in the course of lag-adaptation, albeit clear visual modulations of the auditory responses were seen. Previous EEG studies using auditory and tactile stimuli during a TOJ task reported amplitude modulation (McDonald et al., 2005) or latency shifts (Vibell et al., 2007) of the evoked sensory responses as a function of which sensory modality was attended. It was notably reported that attention could speed up the processing of the attended sensory modality. Here, no systematic changes in the evoked profiles were observed suggesting that during lag-adaptation participants equally paid attention to the auditory and visual events as per task requirements (cf. Methods).

Although the lack of significant lag-adaptation suppression (Grill-Spector et al., 2006) of the evoked responses was surprising, it is classically known that attention can attenuate the effect of neural suppression (Gazzaley et al., 2005). Here, participants were asked to pay attention to any deviants presented in the auditory, visual or audiovisual modalities during lag-

adaptation. The lack of repetition suppression in both sensory cortices may thus be an index of successful attentional orienting. Additionally, recent findings have shown that the more temporally predictable, the higher the repetition suppression effects notably in the auditory responses (Costa-Faidella et al., 2011; Summerfield et al., 2011, 2008). In the context of predictive coding models, it has also been suggested that repetition and expectation were dissociable (Todorovic and de Lange, 2012). The current experimental design did not allow us to dissociate the factor of predictability and expectation but these observations provide an alternative speculation, namely that paying attention to time may alleviate neural lag-adaptation.

One relevant point here is the integrative vs. segregative nature of the task with regards to multisensory processing: experimental paradigms using multisensory integration have classically reported an increase of sensory evoked responses, for instance when using the sound-induced flash illusion (Mishra et al., 2007; Watkins et al., 2006). In a TOJ task however, the segregation of auditory and visual information is a pre-requisite for successful ordering of auditory and visual events in time. Participants were repeatedly presented with AV lags at an entrainment rate consistent with automatic multisensory integration (Kösem and van Wassenhove, 2012): as such, audiovisual binding was reinforced in this task and an increased evoked response would have been expected. However, and at the same time, a decrease of the evoked responses were expected by virtue of neural suppression (Grill-Spector et al., 2006). Hence, one possible explanation for the absence of significant modulations of the sensory evoked responses in the course of adaptation is the competition between the integrative and segregative processes in this particular experimental design. Additional work will be required to further address this working hypothesis.

4.4. Phase of neural oscillations: encoding time (or space?)

Multisensory integration is known to capitalize on the spatiotemporal coincidence of sensory events (Colonius and Diederich, 2010; Meredith et al., 1987) and visual capture of auditory spatial representation is a classic phenomenon (i.e. ventriloquism, (Alais and Burr, 2004; Lewald and Guski, 2003; Slutsky and Recanzone, 2001)) well accounted for by Bayesian models of multisensory integration (Alais and Burr, 2004; Burr and Alais, 2006; Ernst and Bühlhoff, 2004; Witten and Knudsen, 2005). More generally, vision tends to be most reliable in encoding spatial cues whereas audition provides the most reliable temporal cues. In the experimental design used here, visual events were displayed at a constant distance on the

monitor screen whereas sounds were presented via headphones. By virtue of spatiotemporal coincidence and given the consistent AV timing over lag-adaptation, the auditory distance would have to be adjusted to visual information. In a scheme analogous to the calibration of auditory spatial representation by vision in the barn owl (Knudsen and Brainard, 1991), the observed non-stationary phase shifts in auditory cortices could thus reflect an automatic means to fine tune spatiotemporal coincidence across sensory modalities. Specifically, auditory spatial uncertainty could be compensated for by the stable spatiotemporal reference frame established in vision. From this viewpoint, the auditory system would not act as a timer per se; rather, the distance of auditory events would be actively made compatible with visual inputs to form an integrated AV percept. Hence, shifts in AV simultaneity may reflect the compensation of temporal delays in audition. Such mechanism would predict what we observed, namely that the entrainment of auditory and visual cortices is asymmetrical when attention is directed to the timing of events (i.e. to the dominant sensory modality for timing, namely audition); it also predicts that response times for audition (but not for vision) vary during AV delay exposure (Navarra et al., 2009) as well as the correlation between reaction times and visual to auditory phase-reset previously described (Thorne et al., 2011). Our results further support a recent discussion on the functional asymmetry between the sampling of acoustic and visual information over time (Thorne et al., 2011), namely: while the visual system may naturally rely on endogenous rhythms (e.g. alpha oscillations, Gho and Varela, 1988; Jensen et al., 2012; VanRullen and Koch, 2003; Varela et al., 1981) and overt sampling (e.g. (micro)saccades, Schroeder and Lakatos, 2009), the auditory system may necessitate temporal-locking to incoming acoustic inputs to accurately represent information over time (Giraud and Poeppel, 2012; Henry and Obleser, 2012; Stefanics et al., 2010; Thorne et al., 2011). Hence, while visual timing may rely on an internally generated temporal reference frame, audition may require the establishment of a temporal reference frame on the go and locked to the temporal statistics of the auditory environment.

4.5. Neural oscillations: multiplex encoding of information

Our results suggest that, in cortex, the phase of neural oscillations may provide an automatic means to flag events in a brain's referential time – i.e. provide the needed brain-centric view of time (Scharnowski et al., 2013). The encoding of spatiotemporal information in the phase of neural oscillations has been described in the hippocampus in which mechanisms of phase precession encode spatial locations as the animal navigates in a maze (Buzsáki, 2002; Lisman,

2005; Skaggs et al., 1995). Phase precession mechanisms may not be exclusive to hippocampal networks nor to spatial processing and may serve a more general purpose such as encoding events for working memory while preserving temporal order (Lisman, 2005; Lisman and Jensen, 2013). In particular, the content of a sensory event is encoded by the neural assembly firing within a certain gamma cycle (Lisman, 2005; Lisman and Jensen, 2013) while the relative timing of the event is encoded in the phase of the theta oscillation (Lisman, 2005; Lisman and Jensen, 2013).

In cortex, low-frequency neural oscillations are known to regulate the excitability of neural ensembles such that specific phases of low-frequency neural oscillations are associated with periods of high and low neuronal excitability (Buzsáki, 2010; Lakatos et al., 2008; Schroeder and Lakatos, 2009): the phase of low-frequency neural oscillations modulates the power of high-frequency neural oscillatory responses, a mechanism known as phase-power or cross-frequency coupling (Canolty et al., 2006). Neural synchronizations in higher frequency ranges (e.g. gamma range, >40 Hz) provide a reliable index of feature binding within and across sensory modalities (Arnal et al., 2011; Engel et al., 1991; Roelfsema et al., 1997; Senkowski et al., 2008; Tallon-Baudry and Bertrand, 1999). In multisensory integration, low-frequency neural oscillations (delta, 1-2Hz) play a crucial role in the temporal selection (Besle et al., 2011; Fiebelkorn et al., 2013; Gomez-Ramirez et al., 2011; Lakatos et al., 2008; Schroeder and Lakatos, 2009) and in the integration of AV information (Fiebelkorn et al., 2011; Kösem and van Wassenhove, 2012; Luo et al., 2010).

Hence, the phase of low-frequency oscillations may provide the fine-grained temporal resolution needed for the segregation of AV event timing and conscious timing while preserving integration processes through neural synchronization (necessary in building the mental representation of a multisensory AV object).

In such scheme, the informational chunking operates over an oscillatory cycle by eliciting temporal windows of high neural excitability for integration (Panzeri et al., 2010; van Wassenhove, 2009), while the phase of neural oscillations provides the temporal stamping operation needed to preserve the timing of operations in parallel systems. As such, the same informational content can be encoded in a multiplexed manner with (i) integration operating on those sensory attributes used in the building of an internal object (Engel and Singer, 2001; Treisman, 1996) while (ii) segregation - or temporal stamping – provides the automatic encoding of event timing. Such temporal encoding framework comes in support of intrinsic and non-dedicated models of time perception over small time scales (Karmarkar and Buonomano, 2007). Our data further converge with recent findings showing that accurate

phase encoding of the temporal structure of sensory events affords predictability (Schroeder and Lakatos, 2009; Stefanics et al., 2010) and support recent phase-coding approaches in computational neurosciences (Nadasdy, 2010).

4.6. Conclusion

We showed that perceived simultaneity linearly maps onto the phase of neural oscillations in auditory cortex. Our findings complement recent findings showing that accurate phase encoding of temporal event structure affords predictability (Schroeder and Lakatos, 2009; Stefanics et al., 2010) and enhances task performance (Busch et al., 2009; Monto et al., 2008; Neuling et al., 2012; Romei et al., 2012; Varela et al., 1981). Our results further suggest that mechanisms analogous to phase precession in hippocampus may be used in cortex for the encoding of event timing. Specifically, the phase of slow oscillatory activity in sensory areas may provide a canonical means to organize sensory inputs in time. Future work will address the possibility that a canonical function of neural oscillations is the encoding of event timing serving the emergence of psychological time.

ACKNOWLEDGMENTS

This work was supported by an ERC-YStG-263584, an ANR10JCJC-1904 and an IRG-249222 to V.vW. The authors report no conflict of interest. We are grateful to the NeuroSpin infrastructure groups, for their support in participant recruitment; and to anonymous reviewers for their comments on prior version of the manuscript.

REFERENCES

- Alais, D., Burr, D., 2004. The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol* 14, 257–62.
- Arnal, L.H., Wyart, V., Giraud, A.-L., 2011. Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. *Nat Neurosci* 14, 797–801.
- Besle, J., Schevon, C.A., Mehta, A.D., Lakatos, P., Goodman, R.R., McKhann, G.M., Emerson, R.G., Schroeder, C.E., 2011. Tuning of the human neocortex to the temporal dynamics of attended events. *J Neurosci* 31, 3176–3185.

- Brainard, D.H., 1997. The Psychophysics Toolbox. *Spat Vis* 10, 433–436.
- Buhusi, C. V, Meck, W.H., 2005. What makes us tick? Functional and neural mechanisms of interval timing. *Nat Rev Neurosci* 6, 755–65.
- Buhusi, C. V, Meck, W.H., 2009. Relativity theory and time perception: single or multiple clocks? *PLoS One* 4, e6268.
- Burr, D., Alais, D., 2006. Combining visual and auditory information. *Prog Brain Res* 155, 243–58.
- Busch, N. a, Dubois, J., VanRullen, R., 2009. The phase of ongoing EEG oscillations predicts visual perception. *J Neurosci* 29, 7869–7876.
- Buzsáki, G., 2002. Theta oscillations in the hippocampus. *Neuron* 33, 325–40.
- Buzsáki, G., 2010. Neural syntax: cell assemblies, synapsembles, and readers. *Neuron* 68, 362–385.
- Buzsáki, G., Draguhn, A., 2004. Neuronal oscillations in cortical networks. *Science* 304, 1926–1929.
- Cai, M., Stetson, C., Eagleman, D.M., 2012. A neural model for temporal order judgments and their active recalibration: a common mechanism for space and time? *Frontiers in psychology* 3, 470.
- Canolty, R.T., Edwards, E., Dalal, S.S., Soltani, M., Nagarajan, S.S., Berger, M.S., Barbaro, N.M., Knight, R.T., 2006. High gamma power is phase-locked to theta oscillations in human neocortex. *Science* 313, 1626–1628.
- Capilla, A., Pazo-Alvarez, P., Darriba, A., Campo, P., Gross, J., 2011. Steady-state visual evoked potentials can be explained by temporal superposition of transient event-related responses. *PLoS One* 6.
- Chakravarthi, R., Vanrullen, R., 2012. Conscious updating is a rhythmic process. *Proc Natl Acad Sci U S A* 109, 10599–604.
- Chopin, A., Mamassian, P., 2012. Predictive properties of visual adaptation. *Curr Biol* 22, 622–6.
- Church, R.M., 1984. Properties of the Internal Clock. *Ann N Y Acad Sci* 423, 566–582.
- Coles, M., Rugg, M., 1995. Event-related brain potentials: An introduction, in: Coles, M., Rugg, M. (Eds.), *Electrophysiology of Mind: Event-Related Brain Potentials and Cognition*. Oxford University Press, London, pp. 1–26.
- Colonus, H., Diederich, A., 2010. The optimal time window of visual-auditory integration: a reaction time analysis. *Front Integr Neurosci* 4, 11.

- Costa-Faidella, J., Baldeweg, T., Grimm, S., Escera, C., 2011. Interactions between “what” and “when” in the auditory system: temporal predictability enhances repetition suppression. *J Neurosci* 31, 18590–7.
- Coull, J.T., Vidal, F., Nazarian, B., Macar, F., 2004. Functional anatomy of the attentional modulation of time estimation. *Science* 303, 1506–8.
- Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* 9, 179–194.
- Engel, a K., König, P., Singer, W., 1991. Direct physiological evidence for scene segmentation by temporal coding. *Proc Natl Acad Sci U S A* 88, 9136–9140.
- Engel, A.K., Singer, W., 2001. Temporal binding and the neural correlates of sensory awareness. *Trends Cogn Sci* 5, 16–25.
- Ernst, M.O., Bühlhoff, H.H., 2004. Merging the senses into a robust percept. *Trends Cogn Sci* 8, 162–9.
- Fiebelkorn, I.C., Foxe, J.J., Butler, J.S., Mercier, M.R., Snyder, A.C., Molholm, S., 2011. Ready, set, reset: stimulus-locked periodicity in behavioral performance demonstrates the consequences of cross-sensory phase reset. *J Neurosci* 31, 9971–81.
- Fiebelkorn, I.C., Snyder, A.C., Mercier, M.R., Butler, J.S., Molholm, S., Foxe, J.J., 2013. Cortical cross-frequency coupling predicts perceptual outcomes. *Neuroimage* 69, 126–37.
- Fischl, B., Dale, A.M., 2000. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc Natl Acad Sci U S A* 97, 11050–11055.
- Fisher, N.I., 1995. *Statistical Analysis of Circular Data*. Cambridge University Press, Cambridge.
- Fujisaki, W., Shimojo, S., Kashino, M., Nishida, S., 2004. Recalibration of audiovisual simultaneity. *Nat Neurosci* 7, 773–778.
- Gazzaley, A., Cooney, J.W., Rissman, J., D’Esposito, M., 2005. Top-down suppression deficit underlies working memory impairment in normal aging. *Nat Neurosci* 8, 1298–300.
- Gho, M., Varela, F.J., 1988. A quantitative assessment of the dependency of the visual temporal frame upon the cortical rhythm. *J Physiol (Paris)* 83, 95–101.
- Giraud, A.-L., Poeppel, D., 2012. Cortical oscillations and speech processing: emerging computational principles and operations. *Nat Neurosci* 15, 511–7.
- Gomez-Ramirez, M., Kelly, S.P., Molholm, S., Sehatpour, P., Schwartz, T.H., Foxe, J.J., 2011. Oscillatory Sensory Selection Mechanisms during Intersensory Attention to Rhythmic Auditory and Visual Inputs: A Human Electrographic Investigation. *J Neurosci* 31, 18556–18567.

- Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., Hämäläinen, M., 2013a. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroinformatics* 7.
- Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Parkkonen, L., Hämäläinen, M.S., 2013b. MNE software for processing MEG and EEG data. *Neuroimage*.
- Grill-Spector, K., Henson, R., Martin, A., 2006. Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn Sci* 10, 14–23.
- Gross, J., Baillet, S., Barnes, G.R., Henson, R.N., Hillebrand, A., Jensen, O., Jerbi, K., Litvak, V., Maess, B., Oostenveld, R., Parkkonen, L., Taylor, J.R., van Wassenhove, V., Wibral, M., Schoffelen, J.-M., 2013. Good practice for conducting and reporting MEG research. *Neuroimage*.
- Hämäläinen, M.S., Sarvas, J., 1989. Realistic conductivity geometry model of the human head for interpretation of neuromagnetic data. *IEEE Trans Biomed Eng* 36, 165–171.
- Harrington, D.L., Haaland, K.Y., Knight, R.T., 1998. Cortical Networks Underlying Mechanisms of Time Perception. *J Neurosci* 18, 1085–1095.
- Henry, M.J., Obleser, J., 2012. Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proc Natl Acad Sci U S A* 109, 20095–100.
- Heron, J., Roach, N.W., Whitaker, D., Hanson, J.V.M., 2010. Attention regulates the plasticity of multisensory timing. *Eur J Neurosci* 31, 1755–1762.
- Ivry, R.B., Schlerf, J.E., 2008. Dedicated and intrinsic models of time perception. *Trends Cogn Sci* 12, 273–80.
- Jensen, O., Bonnefond, M., VanRullen, R., 2012. An oscillatory mechanism for prioritizing salient unattended stimuli. *Trends in Cognitive Sciences* 16, 200–206.
- Johnston, A., Nishida, S., 2001. Time perception: brain time or event time? *Curr Biol* 11, R427–30.
- Karmarkar, U.R., Buonomano, D. V., 2007. Timing in the absence of clocks: encoding time in neural network states. *Neuron* 53, 427–38.
- Keil, A., Debener, S., Gratton, G., Junghöfer, M., Kappenman, E.S., Luck, S.J., Luu, P., Miller, G.A., Yee, C.M., 2013. Committee report: Publication guidelines and recommendations for studies using electroencephalography and magnetoencephalography. *Psychophysiology* 51, 1–21.
- Knudsen, E., Brainard, M., 1991. Visual instruction of the neural map of auditory space in the developing optic tectum. *Science* 253, 85–87.
- Kösem, A., van Wassenhove, V., 2012. Temporal structure in audiovisual sensory selection. *PLoS One* 7, e40936.

- Lachaux, J.P., Rodriguez, E., Martinerie, J., Varela, F.J., 1999. Measuring phase synchrony in brain signals. *Hum Brain Mapp* 8, 194–208.
- Lakatos, P., Karmos, G., Mehta, A.D., Ulbert, I., Schroeder, C.E., 2008. Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science* 320, 110–3.
- Lewald, J., Guski, R., 2003. Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Cognitive brain research* 16, 468–478.
- Lin, F.-H., Belliveau, J.W., Dale, A.M., Hämäläinen, M.S., 2006. Distributed current estimates using cortical orientation constraints. *Hum Brain Mapp* 27, 1–13.
- Lisman, J., 2005. The theta/gamma discrete phase code occurring during the hippocampal phase precession may be a more general brain coding scheme. *Hippocampus* 15, 913–22.
- Lisman, J.E., Jensen, O., 2013. The Theta-Gamma Neural Code. *Neuron* 77, 1002–1016.
- Love, S.A., Petrini, K., Cheng, A., Pollick, F.E., 2013. A psychophysical investigation of differences between synchrony and temporal order judgments. *PLoS One* 8, e54798.
- Luo, H., Liu, Z., Poeppel, D., 2010. Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS Biol* 8, e1000445.
- Madl, T., Baars, B.J., Franklin, S., 2011. The timing of the cognitive cycle. *PLoS One* 6, e14803.
- Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods* 164, 177–190.
- McDonald, J.J., Teder-Sälejärvi, W.A., Di Russo, F., Hillyard, S.A., 2005. Neural basis of auditory-induced shifts in visual time-order perception. *Nat Neurosci* 8, 1197–202.
- Meredith, M., Nemitz, J., Stein, B., 1987. Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *J Neurosci* 7, 3215–3229.
- Mishra, J., Martinez, A., Sejnowski, T.J., Hillyard, S.A., 2007. Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. *J Neurosci* 27, 4120–31.
- Miyazaki, M., Yamamoto, S., Uchida, S., Kitazawa, S., 2006. Bayesian calibration of simultaneity in tactile temporal order judgment. *Nat Neurosci* 9, 875–7.
- Monto, S., Palva, S., Voipio, J., Palva, J.M., 2008. Very slow EEG fluctuations predict the dynamics of stimulus detection and oscillation amplitudes in humans. *J Neurosci* 28, 8268–8272.
- Morillon, B., Kell, C.A., Giraud, A.-L., 2009. Three stages and four neural systems in time estimation. *J Neurosci* 29, 14803–11.

- Moutoussis, K., Zeki, S., 1997. A direct demonstration of perceptual asynchrony in vision. *Proc Biol Sci* 264, 393–9.
- Nadasdy, Z., 2010. Binding by asynchrony: the neuronal phase code. *Front Neurosci* 4.
- Navarra, J., Hartcher-O'Brien, J., Piazza, E., Spence, C., 2009. Adaptation to audiovisual asynchrony modulates the speeded detection of sound. *Proc Natl Acad Sci U S A* 106, 9169–9173.
- Neuling, T., Rach, S., Wagner, S., Wolters, C.H., Herrmann, C.S., 2012. Good vibrations: Oscillatory phase shapes perception. *Neuroimage* 63, 771–778.
- Nozaradan, S., Peretz, I., Missal, M., Mouraux, A., 2011. Tagging the neuronal entrainment to beat and meter. *J Neurosci* 31, 10234–10240.
- Panzeri, S., Brunel, N., Logothetis, N.K., Kayser, C., 2010. Sensory neural codes using multiplexed temporal scales. *Trends Neurosci* 33, 111–20.
- Pöppel, E., 1997. A hierarchical model of temporal perception. *Trends Cogn Sci* 1, 56–61.
- Rees, A., Green, G.G.R., Kay, R.H., 1986. Steady-state evoked responses to sinusoidally amplitude-modulated sounds recorded in man. *Hear Res* 23, 123–133.
- Regan, D., 1966. Some characteristics of average steady-state and transient responses evoked by modulated light. *Electroencephalogr Clin Neurophysiol* 20, 238–248.
- Roelfsema, P., Engel, A., Konig, P., Singer, W., 1997. Visuomotor integration is associated with zero time-lag synchronization among cortical areas. *Nature* 385, 157–161.
- Romei, V., Gross, J., Thut, G., 2012. Sounds reset rhythms of visual cortex and corresponding human visual perception. *Curr Biol* 22, 807–813.
- Roopun, A.K., Kramer, M.A., Carracedo, L.M., Kaiser, M., Davies, C.H., Traub, R.D., Kopell, N.J., Whittington, M.A., 2008. Temporal Interactions between Cortical Rhythms. *Front Neurosci* 2, 145–54.
- Scharnowski, F., Rees, G., Walsh, V., 2013. Time and the brain: neurorelativity: The chronoarchitecture of the brain from the neuronal rather than the observer's perspective. *Trends Cogn Sci* 17, 51–2.
- Schroeder, C.E., Lakatos, P., 2009. Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci* 32, 9–18.
- Senkowski, D., Schneider, T.R., Foxe, J.J., Engel, A.K., 2008. Crossmodal binding through neural coherence: implications for multisensory processing. *Trends Neurosci* 31, 401–409.
- Shaw, M.E., Hämäläinen, M.S., Gutschalk, A., 2013. How anatomical asymmetry of human auditory cortex can lead to a rightward bias in auditory evoked fields. *Neuroimage* 74, 22–29.

- Skaggs, W.E., Knierim, J.J., Kudrimoti, H.S., McNaughton, B.L., 1995. A model of the neural basis of the rat's sense of direction. *Adv Neural Inf Process Syst* 7, 173–80.
- Slutsky, D.A., Recanzone, G.H., 2001. Temporal and spatial dependency of the ventriloquism effect. *Neuroreport* 12, 7–10.
- Stefanics, G., Hangya, B., Hernádi, I., Winkler, I., Lakatos, P., Ulbert, I., 2010. Phase entrainment of human delta oscillations can mediate the effects of expectation on reaction speed. *J Neurosci* 30, 13578–13585.
- Stone, J. V, Hunkin, N.M., Porrill, J., Wood, R., Keeler, V., Beanland, M., Port, M., Porter, N.R., 2001. When is now? Perception of simultaneity. *Proc Biol Sci* 268, 31–38.
- Sugita, Y., Suzuki, Y., 2003. Implicit estimation of sound-arrival time. *Nature* 421.
- Summerfield, C., Trittschuh, E.H., Monti, J.M., Mesulam, M.M., Egner, T., 2008. Neural repetition suppression reflects fulfilled perceptual expectations. *Nat Neurosci* 11, 1004–6.
- Summerfield, C., Wyart, V., Johnen, V.M., de Gardelle, V., 2011. Human Scalp Electroencephalography Reveals that Repetition Suppression Varies with Expectation. *Front Hum Neurosci* 5, 67.
- Tallon-Baudry, C., Bertrand, O., 1999. Oscillatory gamma activity in humans and its role in object representation. *Trends Cogn Sci* 3, 151–162.
- Taulu, S., Kajola, M., Simola, J., 2003. Suppression of Interference and Artifacts by the Signal Space Separation Method. *Brain Topogr* 16, 269–275.
- Thorne, J.D., De Vos, M., Viola, F.C., Debener, S., 2011. Cross-modal phase reset predicts auditory task performance in humans. *J Neurosci* 31, 3853–61.
- Todorovic, A., de Lange, F.P., 2012. Repetition suppression and expectation suppression are dissociable in time in early auditory evoked fields. *J Neurosci* 32, 13389–95.
- Treisman, A., 1996. The binding problem. *Curr Opin Neurobiol* 6, 171–8.
- Treisman, M., 1963. Temporal discrimination and the indifference interval: Implications for a model of the “internal clock”. *Psychological Monographs: General and Applied* 77, 1.
- Treisman, M., 1984. Temporal Rhythms and Cerebral Rhythms. *Ann N Y Acad Sci* 423, 542–565.
- Treisman, M., Faulkner, A., Naish, P.L., Brogan, D., 1990. The internal clock: Evidence for a temporal oscillator underlying time perception with some estimates of its characteristic frequency. *Perception* 19, 705–743.
- Treisman, M., Faulkner, A., Naish, P.L.N., 1992. On the Relation Between Time Perception and the Timing of Motor Action: Evidence for a Temporal Oscillator Controlling the

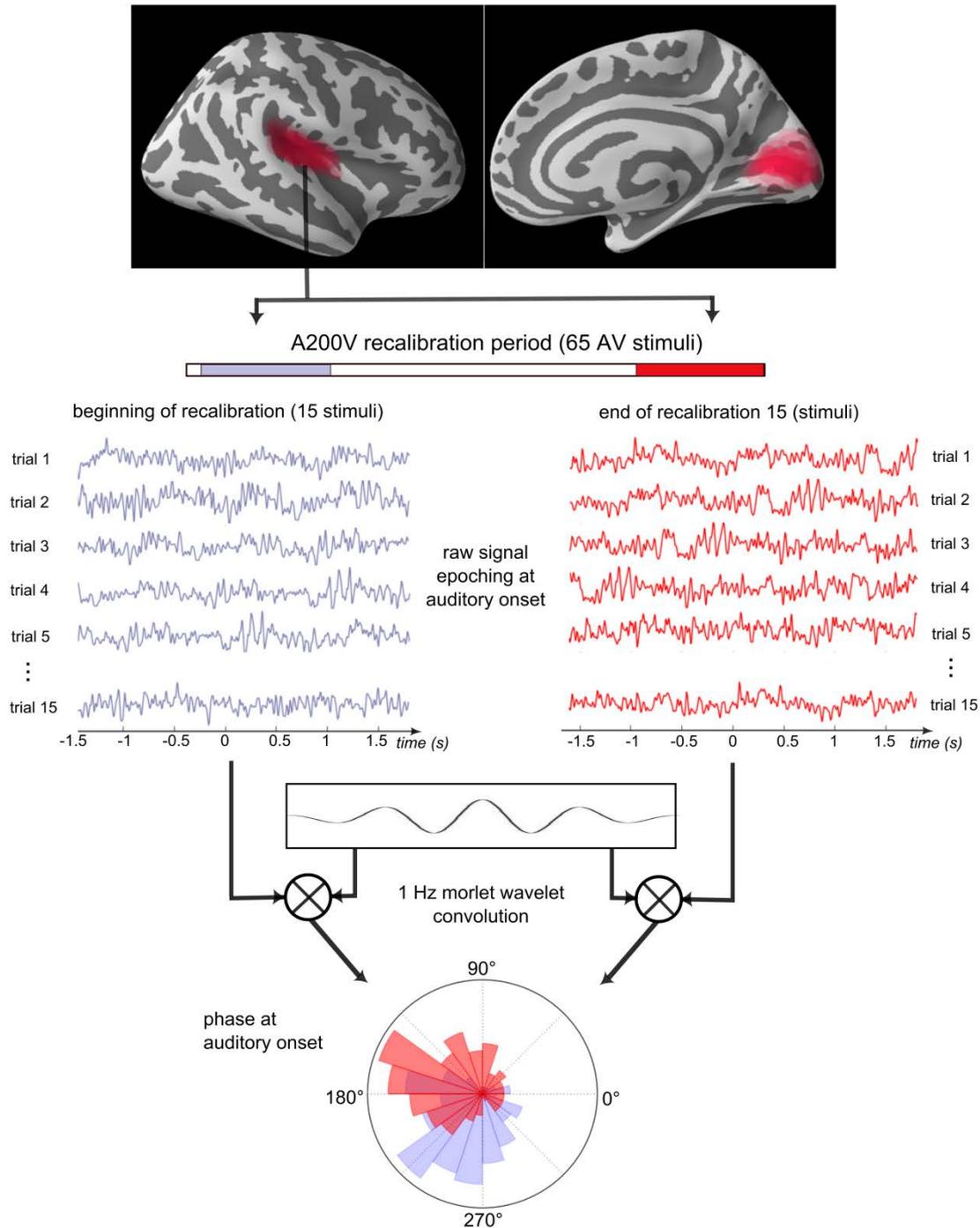
- Timing of Movement. *The Quarterly Journal of Experimental Psychology Section A* 45, 235–263.
- Uusitalo, M.A., Ilmoniemi, R.J., 1997. Signal-space projection method for separating MEG or EEG into components. *Med Biol Eng Comput* 35, 135–140.
- Van Eijk, R.L.J., Kolhtrausch, A., Juola, J.F., Van de Par, S., 2008. Audiovisual synchrony and temporal order judgments: Effects of experimental method and stimulus type. *Perception & Psychophysics* 70, 955–968.
- Van Wassenhove, V., 2009. Minding time in an amodal representational space. *Philos Trans R Soc Lond B Biol Sci* 364, 1815–1830.
- VanRullen, R., Koch, C., 2003. Is perception discrete or continuous? *Trends Cogn Sci* 7, 207–213.
- VanRullen, R., Reddy, L., Koch, C., 2006. The continuous wagon wheel illusion is associated with changes in electroencephalogram power at approximately 13 Hz. *J Neurosci* 26, 502–7.
- Varela, F., Toro, A., John, E.R., Schwartz, E., 1981. Perceptual framing and cortical alpha rhythm. *Neuropsychologia* 19, 675–686.
- Vibell, J., Klinge, C., Zampini, M., Spence, C., Nobre, A.C., 2007. Temporal order is coded temporally in the brain: early event-related potential latency shifts underlying prior entry in a cross-modal temporal order judgment task. *J Cogn Neurosci* 19, 109–20.
- Vroomen, J., Keetels, M., 2010. Perception of intersensory synchrony: a tutorial review. *Atten Percept Psychophys* 72, 871–884.
- Vroomen, J., Keetels, M., de Gelder, B., Bertelson, P., 2004. Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Brain Res Cogn Brain Res* 22, 32–5.
- Wang, X., 2010. Neurophysiological and computational principles of cortical rhythms in cognition. *Phys Review* 90, 1195–1268.
- Watkins, S., Shams, L., Tanaka, S., Haynes, J.-D., Rees, G., 2006. Sound alters activity in human V1 in association with illusory visual perception. *Neuroimage* 31, 1247–1256.
- Witten, I.B., Knudsen, E.I., 2005. Why seeing is believing: merging auditory and visual worlds. *Neuron* 48, 489–96.
- Wittmann, M., 2009. The inner experience of time. *Philos Trans R Soc Lond B Biol Sci* 364, 1955–67.
- Wittmann, M., 2013. The inner sense of time: how the brain creates a representation of duration. *Nat Rev Neurosci* 14, 217–23.
- Yamamoto, S., Miyazaki, M., Iwano, T., Kitazawa, S., 2012. Bayesian calibration of simultaneity in audiovisual temporal order judgments. *PLoS One* 7, e40379.

Zeki, S., Bartels, A., 1998. The asynchrony of consciousness. *Proc Biol Sci* 265, 1583–5.

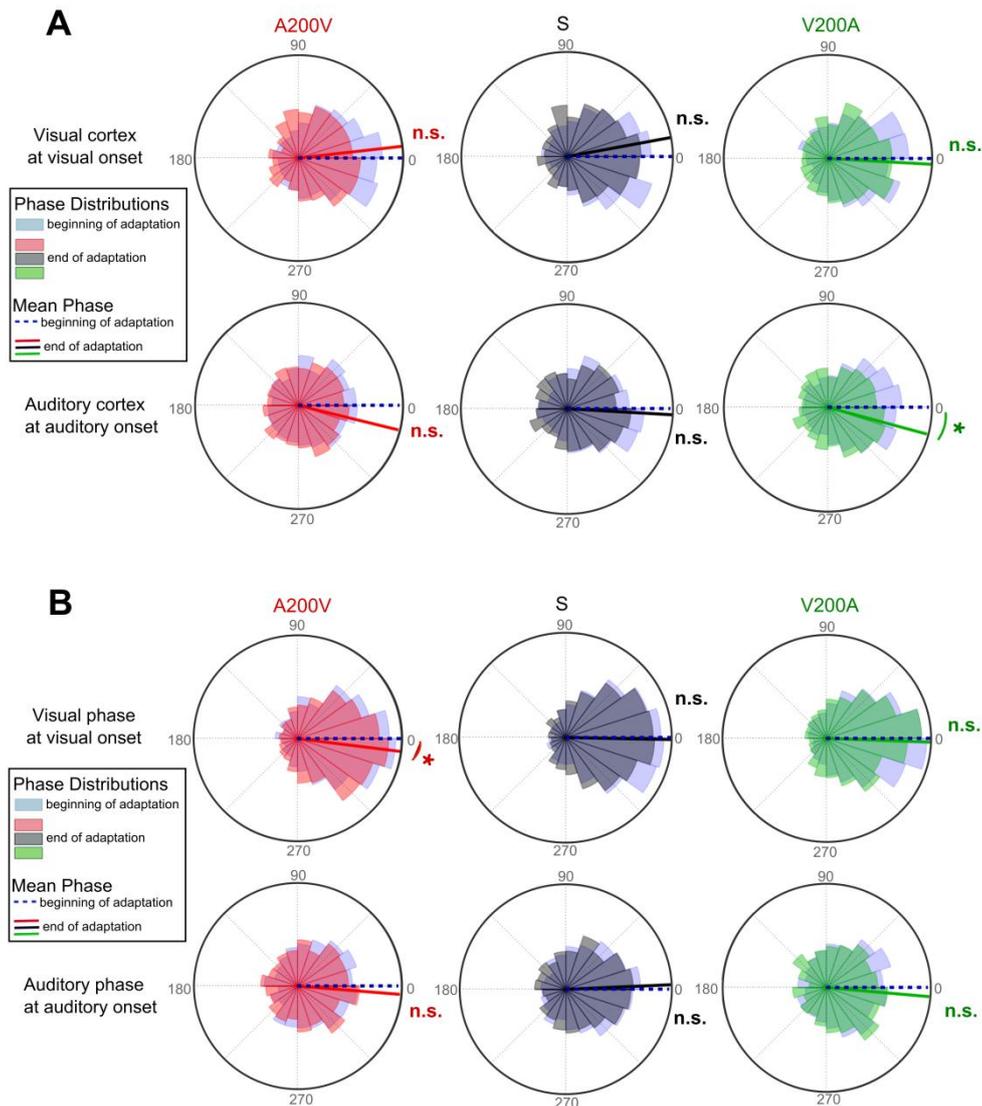
SUPPLEMENTARY MATERIAL

		S	A200V	V200A
Auditory cortices	beginning	0.39 $t(15) = -1.5$	0.35 $t(15) = -0.1$	0.38 $t(15) = 1.6$
	end	0.42 ns	0.35 ns	0.34 ns
Visual cortices	beginning	0.35 $t(15) = 0.4$	0.35 $t(15) = 0.1$	0.36 $t(15) = 0.7$
	end	0.34 ns	0.35 ns	0.34 ns

Supplementary Table S1: Grand average Phase-locking Values (PLVs) of 1Hz neural oscillations in auditory and visual cortices. PLVs were measured at auditory and visual onsets in auditory and visual cortices, respectively. No significant decreases in PLVs were observed between the beginning and the end of the lag-adaptation. As PLVs reflect consistency in the phase-locking of 1Hz oscillations across trials, these results show that the robustness of the 1 Hz entrainment was preserved throughout the experimental trials.



Supplementary Figure S1: Phase analysis. The red patches indicate the overlapping of auditory (left) and visual (right) labels obtained for each individual and morphed onto a common Freesurfer averaged brain (see section 2.5.2.). 15 trials at the beginning (blue) and at the end (red) of each lag-adaptation block were compared in visual and auditory cortices. Single-trial source time series were convolved with a 1 Hz Morlet wavelet. The instantaneous phase was extracted from the resulting complex time series at the center of each epoch (for illustration, here at the auditory onset). For graphical reports, phase distributions were segmented into 20 degrees bins.



Supplementary Figure S2: Phase-shifts are specific to the frequency of the entrained neural oscillation. In order to test whether the reported phase shifts were specific to the entrained 1 Hz neural oscillation, identical analyses were carried out on the same neural responses filtered at 2 Hz (A) and 3 Hz (B). In particular, if the observed phase shifts at 1 Hz were confounded by the evoked responses, identical PLVs and instantaneous phase shifts should be observed at higher frequencies by virtue of the wide spectral impact of evoked responses. Phase distributions and preferential instantaneous phase at 2 Hz and 3 Hz were computed at the beginning (light gray) and at the end (colored) of a given lag-adaptation block (S: black; A200V: red; V200A: green). Phase distributions were computed at visual onset in visual cortices and at sound onset in auditory cortices. Phase distributions were individually normalized to the preferred instantaneous phase observed at the beginning of a given lag-adaptation block. Hence, all Phase distributions at the beginning of a given block are centered on zero. No significant changes in 2Hz oscillation phase were seen between the beginning and the end of the lag-adaptation for A200V or for S in visual and auditory cortices. Only one significant difference in the mean instantaneous phase distribution was observed for the V200A condition in auditory cortices. Note however that 2 Hz is a harmonic of 1 Hz and may actually be a relevant spectral region to consider (albeit outside the scope of this report). No significant changes in 3Hz oscillatory behavior were seen except in visual cortex in condition A200V.

2.3. PHASE CODING WITHOUT ENTRAINMENT

2.3.1. *Motivation*

Entrainment is an ecological phenomenon that occurs when we listen to speech, to music, when we walk, and so on. Yet it is not a prevailing situation in the external world. In many cases, sensory stimulation does not contain any significant temporal information to build up temporal predictions; and by far time experiencers are still able to apprehend the temporality of their environment (Michon, 1985; Stone et al., 2001; Kopinska and Harris, 2004).

In this pilot experiment, we tested whether endogenous fluctuations in auditory and visual cortices could provide a temporal reference frame for timing perception in the absence of explicit entrainment. In the previous experiment delta oscillations (1 Hz) were targeted. This was justified owing to previous reports, indicating a prominent role of these oscillations in audiovisual temporal binding (Lakatos et al., 2008; Schroeder and Lakatos, 2009; Besle et al., 2011; Gomez-Ramirez et al., 2011). Interestingly, these oscillations were found to impact audiovisual perception of temporally isolated audiovisual stimuli (Fiebelkorn et al., 2011, 2013). It thus suggests that endogenous ~1 Hz oscillations have a specific role in audiovisual temporal perception, and might _ with or without entrainment _ provide a reference frame for subjective audiovisual timing. It is also possible that visual alpha oscillations take part in the encoding of event timing (Varela et al., 1981; Gho and Varela, 1988; Chakravarthi and Vanrullen, 2012). Finally, endogenous physiological, subcortical, or non-sensory cortical oscillations may take over temporal processing in the absence of external temporal regularities (James, 1886; Meck, 1996; Matell and Meck, 2004; Wittmann and Paulus, 2008).

2.3.2. *Experiment*

Subjects

10 participants (4 female, mean age: 25 years old) took part in this study. All had normal or corrected-to-normal vision, normal color vision and normal hearing, and were naive as to the purpose of the study. Each participant provided a written informed consent in accordance with the Declaration of Helsinki (2008) and the Ethics Committee on Human Research at NeuroSpin (Gif-sur-Yvette, France).

Experimental Paradigm

The audiovisual stimuli were here composed of a sound via Etymotic earphones (Etymotic Research Inc., USA), and of a visual stimulus flashed with a LED. Sound and flash both had a duration of 30 ms.

After 5-min familiarization to the task, subjects had to estimate the temporal order of audiovisual stimuli with various delays, ranging from +/-300, +/-240, +/-160, +/-120, +/-60, and 0 (a negative delay corresponds to auditory leads asynchronies). Each audiovisual delay was presented 10 times in random order. Each audiovisual stimulus was presented between 0.8 s and 1.2 s after button press. This block was performed to derive the psychometric curve of the subject. From this curve, the Point of Subjective Simultaneity (PSS) and the delays corresponding to 25 % and 75% of “Visual-First” responses (Just-Noticeable Differences or JND, Stekelenburg & Vroomen) were extracted.

These 3 delays were used in the main experiment. As such the delays presented in this session were subject-specific. This choice was motivated by the psychophysical results in the first experiment. Knowing the large inter-subject variability in temporal order judgments, using fixed audiovisual delays (for instance 0 ms and +/-100ms) might have led to unbalanced temporal order responses. For instance, a subject with a PSS of -150 ms (which occurred in our first experiment) would have mainly responded “Visual-first” responses with this setting. To avoid this issue, and to maximize our chances to get 50 % of each response, we calibrated the audiovisual delays according to each subject’s psychometric curve.

At this stage of the experiment the task remained identical with each audiovisual stimulus presented only 3s to 5s after button press. Participants were asked to remain still

during the absence of stimulation and to keep their gaze fixated on the LED. This long period of silence was introduced to measure the phase of low-frequency oscillations before stimulus onset (fig. 2.2a). 120 trials were presented for PSS, 25% JND, and 75% JND conditions, for a total of 360 trials. This session lasted 30 min and was divided into 3 blocks of 8-9 min.

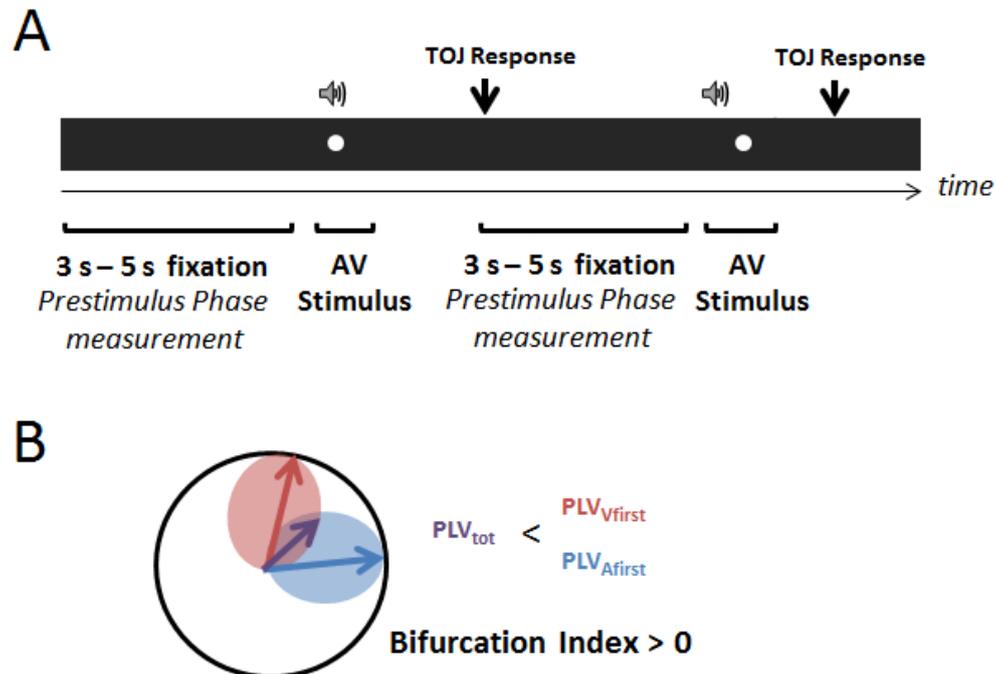


Figure 2.2: Paradigm of control experiment (A) and Hypothesis (B). (A) Participants had to judge the temporal order of temporally isolated audiovisual stimuli. The next stimulus appeared 3 to 5 seconds after button press. This was done to measure the pre-stimulus phase of slow oscillations. (B) If events are mapped in time according to an endogenous “pacemaker” oscillation, then the phase of this oscillation should predict perceived temporal order response. By separating data according to subject’s response, two distinct phase distributions with strong phase locking and distinct preferential phases should be observed. This situation corresponds to a positive Bifurcation Index (BI) value.

MEG analysis

The method used for MEG data recording and preprocessing was identical to the method used in the previous experiment in accordance with typical MEG processing guidelines (Gross et al., 2013). Source space data analysis was performed with MNE-python

software (Gramfort et al., 2013). Data analysis was restricted to the transverse temporal label and t primary visual cortices label taken from the “aparc” parcellation in the right hemisphere (fig. 2.3.).

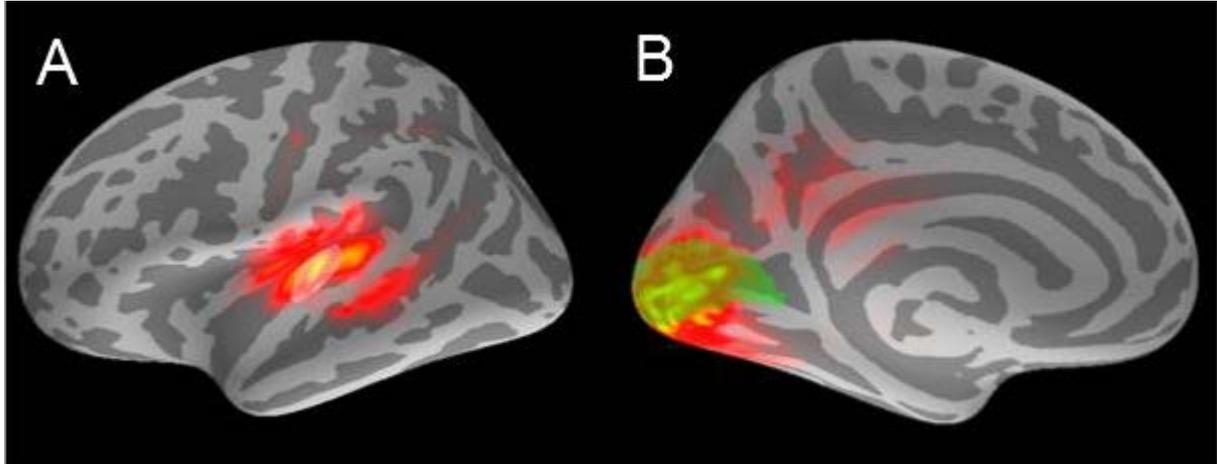


Figure 2.3: Source reconstruction of the (A) auditory and (B) visual evoked response. The white and green labels correspond to the region of interest for the computation of the Bifurcation Index.

The Bifurcation Index (BI) (Busch et al., 2009) was used to test across different frequency bands whether the phase of prestimulus oscillatory activity could predict subject’s response of audiovisual temporal order. For each frequency f and time point t , the BI was computed as follows:

$$BI(f,t) = (PLV_{A_{first}}(f,t) - PLV_{tot}(f,t)) * (PLV_{V_{first}}(f,t) - PLV_{tot}(f,t))$$

Where the PLV is the Phase Locking Value as defined by (Lachaux et al., 1999). $PLV_{A_{first}}$ and $PLV_{V_{first}}$ correspond to the PLVs of all the trials for which the participant responded “auditory stimulus-first” and “visual stimulus-first” respectively. PLV_{tot} is the PLV calculated across all trials indistinctly of participant’s responses.

A null BI either reflects similar phase distributions (same PLV and same preferential phase) between the two conditions of interest or phase locking in none of the conditions.

A negative BI corresponds to a difference in PLV between conditions, and a positive BI reflects that the preferential phase is different between the two conditions. Hence, a positive BI before stimulus presentation would indicate that the preferential phase of an endogenous oscillation predicts perceived timing (fig. 2.2b).

The BI was computed for PSS conditions, e.g. for audiovisual delays at which subjects discriminated temporal order at change level during phase 2. It was computed for each time point in the interval [-2s, 0s] prior to the first stimulus onset (that differs between subjects depending on their PSS). The frequencies of interest were chosen from a log linear scale between 1 Hz and 50 Hz, e.g. 1, 1.4, 1.8, 2.5, 3.3, 4.5, 6.1, 8.2, 11, 15, 20, 27, 37, and 50 Hz. We used a log scale to privilege the investigation of low frequency oscillations (<15 Hz).

The significance of the BI index was assessed for each subject using a bootstrap procedure. Surrogate BI were computed from two sets of trials that were randomly chosen from the whole data set (irrespective of the response “A-first” or “V-first”). This procedure was repeated 100 times to produce a distribution of surrogate BI data. The statistical significance threshold was established for each subject by looking at the 95th percentile of the surrogate BI distribution at each time-frequency point.

2.3.3. Results

Each subject presented various profiles in the BI prior to stimulus onset in both auditory (fig. 2.4) and visual (fig. 2.5) cortices. While various positive peaks in BI were observed, these peaks did not reach statistical significance (uncorrected p_value for multiple comparison > 0.05).

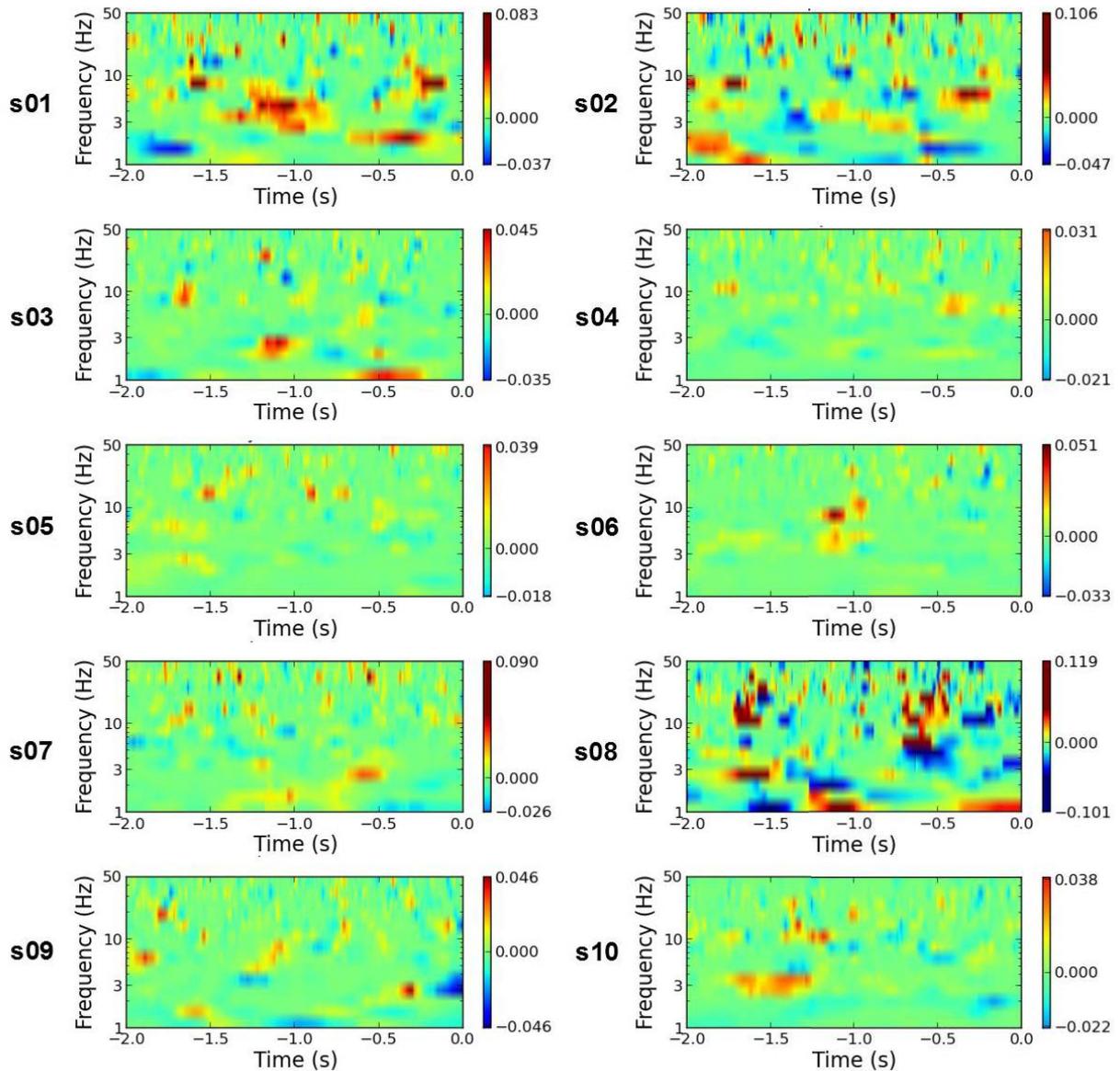


Figure 2.4: Bifurcation Index in auditory cortex for each subject. Each panel corresponds to the BI calculated for each subject before the presentation of audiovisual stimuli in the PSS condition. For subjects s01 to s08 who had a negative PSS, the sound appeared before the flash in the PSS condition. The BI was thus calculated prior to sound onset. Subjects s09 and s10 had a positive PSS, accordingly their BI at calculated prior to flash onset. Color maps are restricted to $[-0.05, +0.05]$ BI values. In addition, color bars indicate the maximal and minimal values for each BI plot. We observed a large variability in the maximal BI values across subjects. We think this variability is due to the difference in the number of trials that were selected to compute the BI for each subject. For example, s01, s02 and s08 had few trials (~ 20 trials per condition) and had large BI values. We suspect these large values are caused by the inaccurate estimation of the PLV in each condition. Overall, no significant positive BI peaks were observed at the subject level (uncorrected p -value for multiple comparison > 0.05). And no clear BI profile appears consistent across subjects.

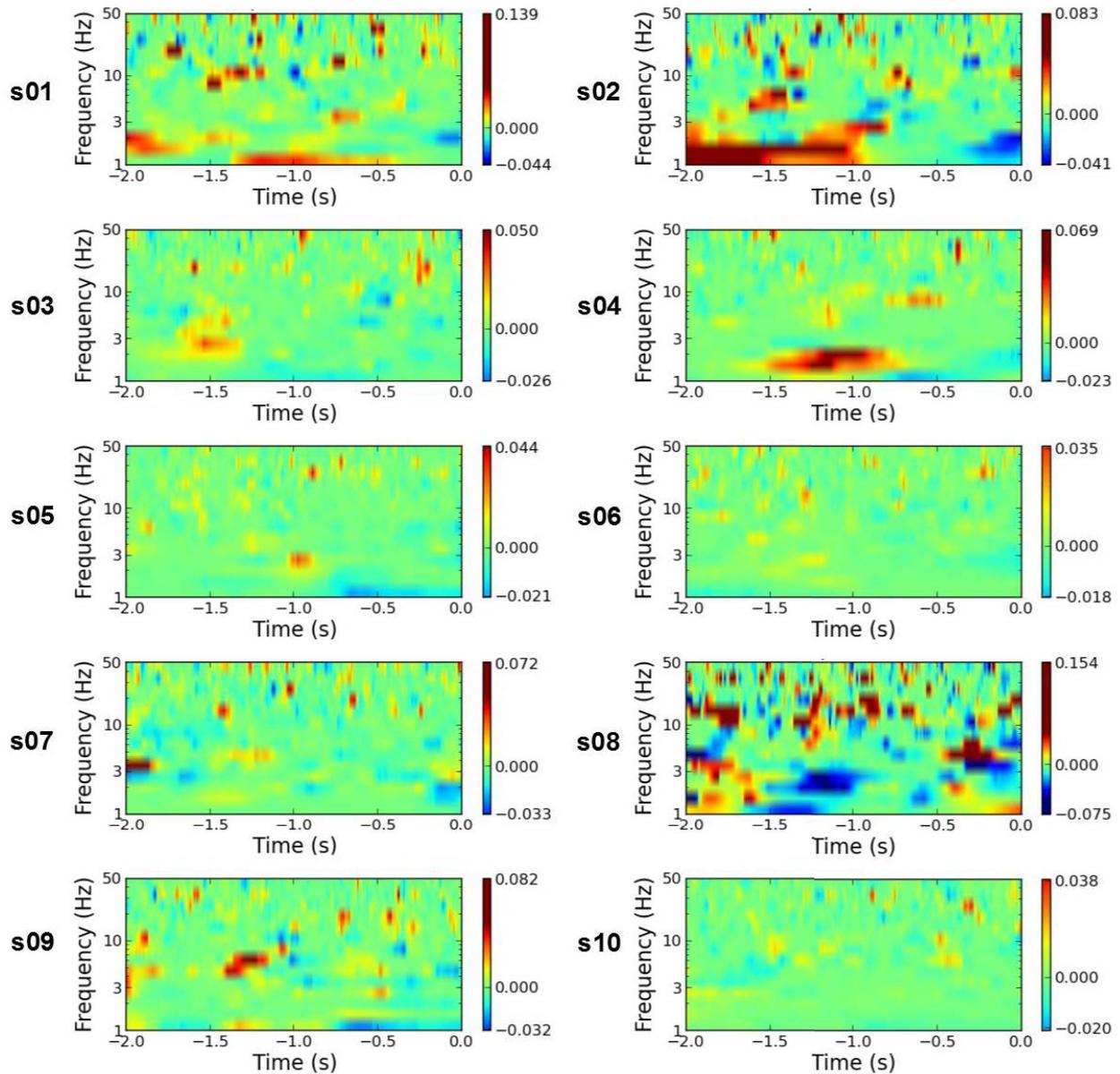


Figure 2.5: Bifurcation Index in visual cortex for each subject. Each panel corresponds to the BI calculated for each subject before the presentation of audiovisual stimuli in the PSS condition. For subjects s01 to s08 who had a negative PSS, the sound appeared before the flash in the PSS condition. The BI was thus calculated prior to sound onset. Subjects s09 and s10 had a positive PSS, accordingly their BI at calculated prior to flash onset. Color maps are restricted to $[-0.05, +0.05]$ BI values. In addition, color bars indicate the maximal and minimal values for each BI plot. Again, we observed a large variability in the maximal BI values across subjects, and in particular s01, s02 and s08 had large BI values. Overall, no significant positive BI peaks were observed at the subject level (uncorrected p -value for multiple comparison > 0.05). And no clear BI profile appears consistent across subjects.

2.3.4. Discussion

The preliminary results of this experiment remain inconclusive regarding to the existence of endogenous “pacemaker” oscillations in visual nor in auditory cortices.

Please note that the number of trials in our experiment are probably a limiting factor for the accurate computation of the BI. In our experiment, 120 trials were presented with the PSS delay, leading to ideally 60 trials for which the sound was perceived first, and 60 trials for which the flash was perceived first. This relatively few amount of trials was chosen as a compromise between amount of collected trials the and the arduousness of the task. In addition to that, some participants (such as s01, s02 and s08) preserved some bias in temporal order perception, so that the proportion of perceived sound-first and perceived flash-first trials was imbalanced at the PSS condition. To cope with this imbalanced data issue, we matched the number of trials in the two data sets by rejecting randomly chosen trials of the largest dataset. This procedure came at the cost of BI precision cost. Hence for these subjects in particular, the reported BI may not be trustable.

For the other subjects, even if not significant, the peaks in th BI profiles corresponded to non uniform phase distributions according to perceived temporal order. For instance, for subject s04 the preferential phase in visual cortices at 2 Hz frequency 1 s before stimulus presentation seemed to predict subsequent temporal order judgment. Current analysis is performed to test the influence of neural oscillations at frequencies that are subject-specific. This analysis is justified by the hypothesis that timing perception, in the absence of entrainment, could recruit endogenous oscillations that are specific to subject’s past experience.

Nevertheless, the fact that we did not observed clear BI profiles in the other subjects could also question whether sensory timing could be retrieved through the dynamics of sensory areas in the *absence of external temporal regularities*. It may further suggest that the reported phase temporal tagging effects reported in the previous experiment may be restricted to neural entrainment situations. This interpretation would be in agreement with the idea that oscillatory timers could be implemented outside sensory areas. Temporal pacemakers could originate from subcortical rhythms in thalamus (Bushara et

al., 2001; Schwartz and Kotz, 2013), hippocampus (Sirota et al., 2008) or substantia nigra (Meck, 1996). In addition, the Striatal Beat Model suggests that striatal neurons might read temporal information within the activity of a large range of areas (Matell and Meck, 2004; Buhusi and Meck, 2005). In particular, frontal neurons that inherently fire rhythmically could serve as “absolute” markers of time perception in the absence of clear external temporal input (Matell and Meck, 2004; Buhusi and Meck, 2005). We also test this alternative hypothesis: we currently perform source space analysis to investigate the potential existence of temporal sensory pacemakers outside primary areas.

LOW RESOLUTION OF MULTISENSORY TEMPORAL BINDING

3.1. INTRODUCTION

3.1.1. Motivation

Chapter 2 demonstrates a prominent role of low-frequency oscillations (1 Hz) entrainment in audiovisual timing perception.

The choice of 1 Hz as the frequency of entrainment was motivated by recent work which emphasized the implication of delta oscillations in the temporal expectancy of audiovisual features. Oscillations in the delta band in auditory cortex can phase-lock to the time of arrival of target sound (Luo et al., 2010; Stefanics et al., 2010b; Besle et al., 2011; Cravo et al., 2013; Hsu et al., 2013) and the likelihood of phase-locking steadily increases with the increased probability of the appearance of the target (Stefanics et al., 2010b). As such, low-frequency oscillations are assumed to take part in the construction of sensory temporal predictions. In multisensory context, visual (auditory) inputs can cause significant delta phase reset in primary auditory (visual) cortex (Lakatos et al., 2008; Besle et al., 2011; Gomez-Ramirez et al., 2011) but phase-resetting is only seen if the visual (auditory) stimuli are attended. If they are ignored, these stimuli have no detectable impact on cortical oscillations (Lakatos et al., 2008; Besle et al., 2011). These results highlight the role of low-frequency oscillations under the process of attentive selection in time (Lakatos et al., 2008; Besle et al., 2011), by applying goal-directed (top-down) attentional selection to either auditory or visual inputs.

Yet, audiovisual attentional selection can also be tuned to the presence of salient unexpected stimuli (Van der Burg et al., 2008, 2011; Talsma et al., 2010). In particular, the search of a dynamic visual target is facilitated in the presence of salient sound beeps that are synchronous to target's changes of appearance (Van der Burg et al., 2008). The modulation of visual search by sound was characterized by early modulations in parietal and primary sensory areas (Van der Burg et al., 2011). Hence, in line with previous early multisensory effects in stimulus evoked responses (Giard and Peronnet, 1999; van Wassenhove et al., 2005), stimulus-driven attention engages fast multisensory binding processes (50-100 ms post stimulus onset). If multisensory integration relies on fast communicating mechanisms, it is then possible that selective attention in time could also capitalize on the fine-grained dynamics of the world, and not only on its very slow fluctuations as previously suggested. Therefore, entrainment to fast temporal regularities could also facilitate audiovisual temporal binding. Here we thus asked the range for which temporal regularities facilitate audiovisual integration.

3.1.2. Experiment

In this experiment we modified the visual search task paradigm introduced by van der Burg and colleagues (2008). Specifically, we used a dynamic visual conjunction search task, with auditory cues that are synchronized with the color change of the target (horizontal or vertical bar). Here, sounds and visual elements were displayed at a pseudo-regular rate which varied between 0.6 Hz and 10 Hz. If the temporal synchrony between the auditory and visual streams of stimuli is sufficient to induce binding, then visual search should be facilitated for all temporal rates. However, if audiovisual temporal binding is limited by the temporal constraints of the recruited mechanisms, then visual search should also be bound to rather slow stimuli dynamics.

3.1.3. Summary of the results

Visual search was improved when the sound was synchronized to the visual target's dynamic changes. In line with previous reports, these results support that rhythmic information could afford predictability of sensory information, and also indicate that

stimulus-driven temporal attentional selection can be crossmodal. However, the benefits from auditory cues could only be observed for temporal rates below 1.4 Hz. These findings further support the hypothesis that audiovisual binding mechanisms are restricted to slow-dynamics in the delta range.

3.2. ARTICLE

Kösem, A., & van Wassenhove, V. (2012). Temporal Structure in Audiovisual Sensory Selection. PloS one, 7(7), e40936.

TEMPORAL STRUCTURE IN AUDIOVISUAL SENSORY SELECTION

Anne Kösem¹⁻³, Virginie van Wassenhove¹⁻³

¹ INSERM, U992, Cognitive Neuroimaging Unit, F-91191 Gif/Yvette, France

² CEA, DSV/I²BM, NeuroSpin Center, F-91191 Gif/Yvette, France

³ Univ Paris-Sud, Cognitive Neuroimaging Unit, F-91191 Gif/Yvette, France

ABSTRACT

In natural environments, sensory information is embedded in temporally contiguous streams of events. This is typically the case when seeing and listening to a speaker or when engaged in scene analysis. In such contexts, two mechanisms are needed to single out and build a reliable representation of an event (or object): the temporal parsing of information and the selection of relevant information in the stream. It has previously been shown that rhythmic events naturally build temporal expectations that improve sensory processing at predictable points in time. Here, we asked to which extent temporal regularities can improve the detection and identification of events across sensory modalities. To do so, we used a dynamic visual conjunction search task accompanied by auditory cues synchronized or not with the color change of the target (horizontal or vertical bar). Sounds synchronized with the visual target improved search efficiency for temporal rates below 1.4 Hz but did not affect efficiency above that stimulation rate. Desynchronized auditory cues consistently impaired visual search below 3.3 Hz. Our results are interpreted in the context of the Dynamic Attending Theory: specifically, we suggest that a cognitive operation structures events in time irrespective of the sensory modality of input. Our results further support and specify recent neurophysiological findings by showing strong temporal selectivity for audiovisual integration in the auditory-driven improvement of visual search efficiency.

KEYWORDS: multisensory perception; attentional selection; binding problem; neural oscillation; perceptual grouping; dynamic attending theory.

1. INTRODUCTION

Many ecologically relevant events (such as speech, auditory and visual scenes, music...) present natural periodicities or statistical temporal regularities [1-4]. These temporal regularities provide useful cues to help parse and structure events out of complex sensory streams notably by building strong temporal expectations on the upcoming sensory inputs. For instance, it has previously been shown that a steady rhythmic presentation improves the detection of an event in a stream [5-7]: the detection of an auditory (visual) event is improved when it appears one period after the last auditory (visual) event, but impaired when it is presented earlier or later than at the instant predicted on the basis of the previous stimulation rate [5-7]. Such results have been interpreted in the context of the Dynamic Attending Theory (DAT) [8] (Figure S1).

The DAT provides a mechanism for selective attention in time i.e. for the parsing of objects based on their inherent temporal structure or based on the temporal structure of an internal oscillator. One strong assumption of the DAT is that the brain can not only keep track of temporal regularities (or environmental rhythms) but also predict, on this basis, the arrival time of a transient event that fluctuates at the same rate. As such, the ‘temporal context’, defined as the relative timing between past AV events, becomes an important factor for attentional selection in time. One implementation of attentional selection in time heavily relies on oscillatory mechanisms that lock to the temporal structure of sensory events [8,9]. This process can be compared to an expectancy profile that naturally allows attention to be engaged at the very point in time at which a stimulus is anticipated to appear or change [5-7]. This temporally precise allocation of attention could bear functional relevance for the early encoding and selection of features across sensory modalities. Thus, DAT sketches an attentional-tracking mechanism over time that is understudied yet offers interesting complementary views to more traditional space-, feature- or object-tracking approaches in the study of attention [10,11]. Here, we asked whether the DAT could be extended across sensory modalities and whether temporal regularities can be shown to operate automatically in the selection of appropriate audiovisual (AV) events.

A first motivation for this experimental work is that the temporal structure of events is a well-known constraint for multisensory integration [12-19]; yet, previous studies have provided contradictory results regarding the automaticity of attentional selection for synchronous

streams of AV stimuli. Using visual search paradigms with dynamic stimuli, the presence of rhythmic auditory stimuli synchronized with visual targets can either improve [20, 21] or have no effect [22] on visual search efficiency. One major difference in these studies was the rate at which AV events were displayed: no AV search efficiency was observed for 10Hz [22] but improvements were reported for 1.1Hz [20]. Second, recent neurophysiological findings have suggested that tracking the temporal structure of AV events likely operates in particular temporal regimes [23-25]. Neural oscillations are classically known to entrain to rhythmic stimuli [26, 27], thereby providing a direct mechanistic implementation for the DAT (Figure S1): neural entrainment modulates the excitability of tuned neural population *through time*. As such, the processing of events in phase with the entrained oscillation is facilitated due to higher neural excitability [28]. In AV context, it has been shown that rhythmic sounds lead to neural entrainment not only in auditory cortex but also in visual cortices [24]; this suggests that rhythmic auditory stimulation can modulate visual processing at relevant points in time and could presumably affect visual detection rates. Third, the AV modulation of neural excitability across sensory cortices has been demonstrated for neural oscillations of 1-2 Hz or “delta band” [23-25]. The specific involvement of slow frequency oscillations (1-2 Hz range) may optimize early AV integration for stimuli in that dynamic range.

The temporal rates between studies [20] and [22], the known importance of AV transience for multisensory integration [21, 29-31], and the temporal structure of AV events [12, 17, 18] can all be limiting factors for automaticity in AV integration. Hence, we were interested in the effect of AV temporal rate and the temporal context it confers to visual search efficiency. Specifically, on neurophysiological grounds, the existence of a temporal threshold (1-2 Hz) on the automaticity of AV integration is here predicted.

To test this hypothesis, we build on the visual conjunction search paradigm developed by van der Burg and colleagues [20]: a horizontal or vertical bar (visual target) surrounded by distracters of various orientations changed colors at particular temporal rates (Figure 1, Video S1). We used seven temporal rates and three set sizes to test whether the rate at which the visual target changed color alone (V), with a synchronized sound (AVc) or with a sound synchronized with a distracter (AVi) was a determining factor for search efficiency.

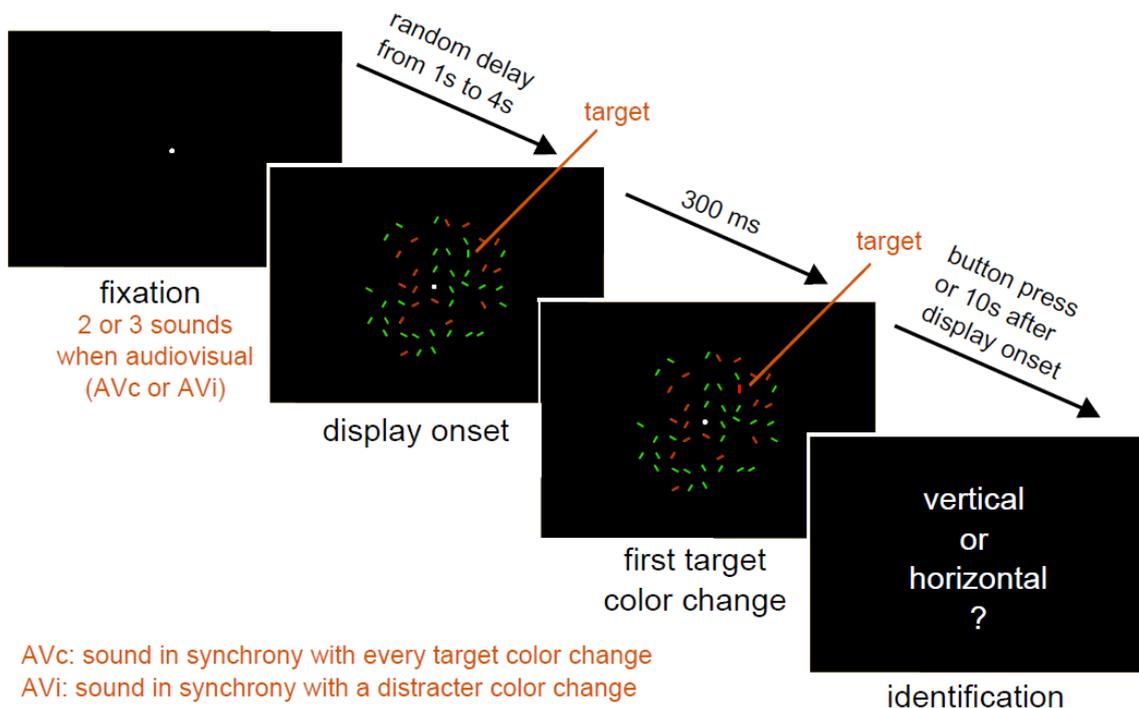


Figure 1: Experimental Paradigm. Each trial started with a fixation point lasting between 1 and 4 seconds (randomized across trials). In the audiovisual conditions (AVc, AVi) two or three sounds appeared before the visual display in order to avoid a surprise effect at the onset of the first sound. This was followed by the visual display with or without a sound (AVc and AVi or V, respectively). Participants were asked to find a horizontal or a vertical bar in the visual display while maintaining their gaze on the fixation point at all times. They were asked to answer as fast and as accurately as possible by pressing the space bar on the keyboard. One trial lasted a maximum of 10 seconds during which the participant was expected to have detected the target. After detection, participants were asked to identify the orientation of the detected target (vertical or horizontal). If the participant had not detected the target, he was nevertheless asked to make a guess. Therefore, this design allowed quantifying two dependent variables: reaction times (RTs – with a 10 sec imparted limit for the participant’s detection) and identification rate. In subsequent analysis, trials in which the target was not detected within 10 s were discarded for RTs. The experiment was run in 3 pseudo-randomized blocks corresponding to the display condition (V, AVc and AVi).

2. RESULTS

Visual search efficiency was quantified in terms of RTs and identification rate: for each trial, participants were asked to press a button as fast as possible when they saw the target; after detection, they reported its orientation in a 2-Alternative-Forced-Choice (2-AFC, “vertical” or “horizontal”) allowing the assessment of the correct identification rate. Statistical analysis was performed using a linear mixed effects model for RTs and a logistic regression model for identification rate [32, 33]. The fixed factors were display condition (3: AVc, AVi, and V), set

size (continuous factor) and temporal rate (7 discrete levels). Subjects (n=24) were a random effect. Significant fixed factors were assessed by means of a regression model simplification using the Akaike Information Criterion (AIC). Each model's goodness of fit was evaluated against the inclusion of each variable and interactions across variables of interests. Table 1 summarizes the comparison of the obtained models. The 'set size', 'temporal rate' and their interaction with the predictor 'display condition' accounted for a significant amount of variance on RTs and identification rate (Table 1: model 5 and 4, respectively). The interpretation of all models preceding model 5 and 4 are provided in text. The additional statistical analyses were conducted with the regression models containing all significant predictors and interactions for each dependent variable: namely, model 5 for RTs and model 4 identification rates. Specifically, all regression coefficients used to assess statistical significance (t-tests, Wald tests (yielding Z)) were directly drawn from these two models.

2.1. Transient sounds affect visual search efficiency irrespective of temporal stimulation rate

Participants were faster and more accurate at detecting the target in AVc than in V (significance of contrast coefficient AVc vs. V for RTs: $t = 3.2$, $p < 0.001$; for identification: $Z = -5.8$, $p < 0.001$) but slower and less accurate in detecting the target in AVi than in V (significance of contrast coefficient AVi vs. V for RTs: $t = -5.9$, $p < 0.001$; for identification: $Z = 8.5$, $p < 0.001$). An AV congruency effect (contrast AVc vs. AVi) was observed in both RTs and identification rates (RTs: $t = 9.7$, $p < 0.001$; identification: $Z = -13.3$, $p < 0.001$). These results suggest that a transient sound facilitates the detection of a synchronized visual target but also impairs target detection when it is synchronized with a distracter, in line with prior reports [20].

A main effect of set size was found (Figure 2; Table 1, "model 2: model 1+ set size"): across all display conditions, search efficiency decreases when the number of distracters increases (RT slope = 7 ± 2 ms / item, $t = 4.3$, $p < 0.001$). As shown in Figure 2, the set size impaired RTs more in AVi than in AVc (AVi slope value: 11 ± 3 ms / item; AVc slope value: 4 ± 3 ms / item; $t = 2.7$, $p < 0.01$). When gathering data across all temporal rates, no significant effect was found between the AVc and V slopes (V slope value 11 ± 2 ms / item; $t = 1.8$, $p = 0.058$)

or between the AVi and V slopes ($t = -0.8, p = 0.4$). Thus, the number of distracters influences the visual search less in AVc than in AVi or in V.

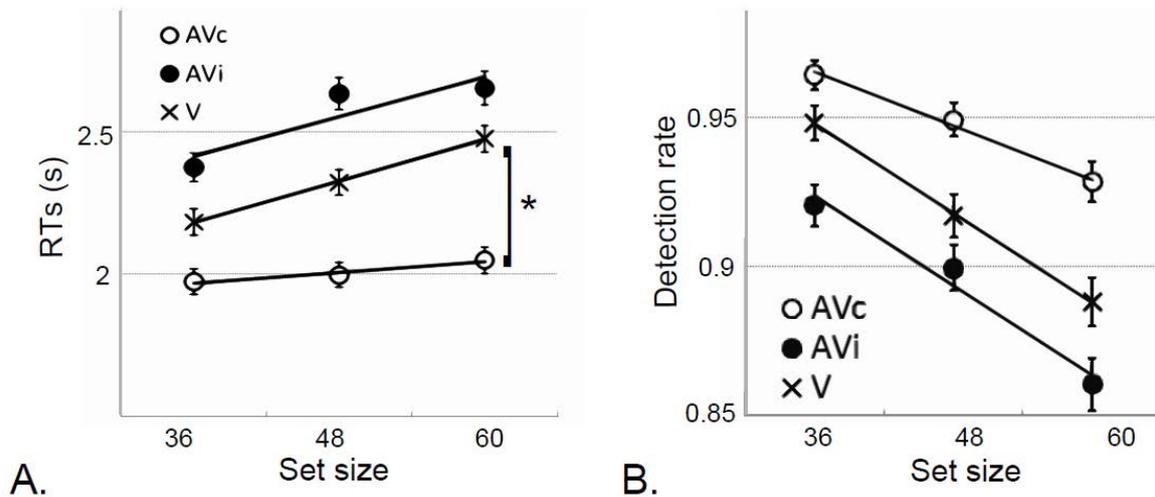


Figure 2: Effect of number of distracters on RT and identification rate collapsed over all temporal rates.

Mean response times (A) and detection rates (B) per condition and per subject as a function of set size. Bars denote two SEM. A significant interaction was found between display condition and set size for RTs. The slope of the curve $RTs = f(\text{set size})$ was significantly lower in condition AVc than in condition AVi. The number of distracters affected the visual search less when a sound was synchronized with a target color change than in the absence of sound (V) or in desynchronized condition (AVi).

However, RTs cannot be taken as definite evidence for improvements in perceptual processing [34, 35]. No significant interaction between display condition and set size was observed for identification (Table 1) and the slopes for correct identification did not significantly differ across modalities when temporal rates were taken out of the model (Figure 2). Identification rates decreased with increasing number of distracters whereas the search remained most efficient in AVc and least efficient in AVi. According to our hypothesis, efficient AV search may not occur across all temporal rates (cf. main effects of temporal rates in RT and identification rate in Table 1, model 3) and we thus turn to the specific effects of temporal rate on visual search efficiency.

RTs	Df model	AIC	ChisqChi	Df	Pr(>Chisq)
Regression models					
model 1: display condition + (1 subject)	5	55263			
model 2: 1+ set size	6	55243	22.92	1	1.69e-16 ***
model 3: 2+ temporal rate	12	55118	136.85	6	< 2.2e-16 ***
model 4: 3+ display condition * temporal rate	24	55090	51.43	12	7.84e-7 ***
model 5: 4+ display condition * set size	26	55086	8.46	2	0.0015 *
model 6: 5+ set size * temporal rate	32	55091	7.03	6	0.32
IDENTIFICATION	Df Model	AIC	ChisqChi	Df	Pr(>Chisq)
Regression models					
model 1:display condition + (1 subject)	4	8055.3			
model 2: 1+ set size	5	7965.9	91.3227	1	< 2.2e-16 ***
model 3: 2+ temporal rate	11	7923.1	54.8172	6	5.047e-10 ***
model 4:3+ display condition * temporal rate	23	7875.9	71.2673	12	1.854e-10 ***
model 5: 4+ display condition * set size					
model 6:5+ set size * temporal rate	25	7877.8	2.0273	2	0.362
	31	7885.4	4.3897	6	0.624

Table 1: Summary of linear mixed regression analyses. Regression model minimization used the Akaike Information Criterion (AIC) and likelihood ratio. Three factors were analyzed: display condition (3 levels), set size (continuous factor) and temporal rate (7 discrete levels) plus one random effect (24: participants). Six models were tested to explain the data with increasing order of complexity, namely: model (1): the effect of display condition; model (2): model 1 + set size; model (3): model 2 + temporal rate; model (4): model 3 + display condition x set size; model (5): model 4 + display condition x temporal rate; model (6): model 5 + set size x temporal rate. Bold models designate those variables significantly contributing to model estimate.

2.2. Temporal rates and attentional selection

Temporal rates accounted for a significant amount of RT and identification rate variance (Table 1, model 3): significant effects were observed for both RT and identification rates between the different temporal rates (Tables S1 and S2, respectively). Overall, participants were faster and more accurate at temporal rates below 1.4 Hz compared to rates above 3.3 Hz irrespective of modality and set size (Figure 3, Table S1 and S2). Abrupt visual onsets are known to capture exogenous attention [36]. In this paradigm, as the temporal rate increases so does the number of color changes: this could lead to a larger temporal crowding effect in which individuating the visual target in time may become particularly challenging [37]. Interestingly, temporal rates significantly affected both RTs and identification rates in AVc and in V but not in AVi (Tables S3 and S4, respectively). The temporal margin introduced in our paradigm could have diminished the temporal crowding effect and benefited target identification in AVc but impaired it in AVi. However, the lack of temporal rate effect on the identification rate in AVi suggests that, consistent with the observed slower RTs, auditory information may either be disregarded or compete with desynchronized visual information (in this case, the target).

2.2.1. Search efficiency and temporal stimulation rate

RTs were significantly faster in AVc than in V for almost all temporal rates (significance of contrast coefficients AVc vs. V at each temporal rate: $t_{0.6\text{ Hz}} = 3.5$, $p < 0.001$; $t_{0.8\text{ Hz}} = 2.4$, $p < 0.05$; $t_{1.1\text{ Hz}} = 1.95$, $p = 0.05$; $t_{1.4\text{ Hz}} = 2.7$, $p < 0.01$; $t_{2\text{ Hz}} = 3.2$, $p < 0.01$. $t_{3.3\text{ Hz}} = 2.3$, $p < 0.01$. $t_{10\text{ Hz}} = 1.95$, $p = 0.05$) but slower in AVi compared to V only for temporal rates below 3.3 Hz (significance of contrast coefficients AVi vs. V at each temporal rate: $t_{0.6\text{ Hz}} = -5.3$, $p < 0.001$; $t_{0.8\text{ Hz}} = -5.8$, $p < 0.001$; $t_{1.1\text{ Hz}} = -5.8$, $p < 0.001$; $t_{1.4\text{ Hz}} = -4.2$, $p < 0.001$; $t_{2\text{ Hz}} = 3.2$, $p < 0.01$. $t_{3.3\text{ Hz}} = -1.3$, ns; $t_{10\text{ Hz}} = -1.1$, ns). Note however that the limit of 3.3 Hz in AVi is not due to the interaction of RTs with temporal rate but rather to the slowing down of RTs in AVc and V with increasing temporal rates (cf. Fig. 3a). The constant RT difference (about 320 ms) across temporal rates between AVc and V support an effect of overall alertness affecting the central decision stage [38].

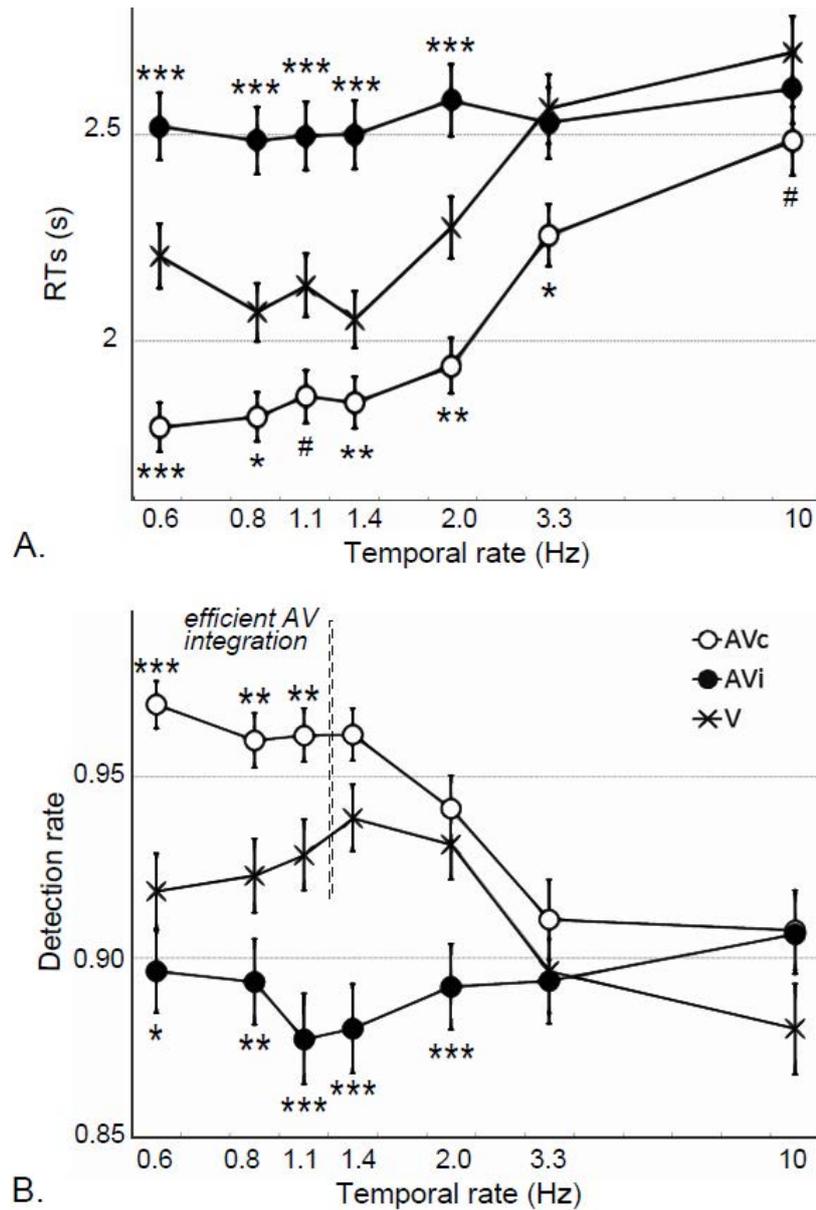


Figure 3: Main effect of temporal rates collapsed across all set sizes on reaction times and identification rate. Mean response times (A) and detection rates (B) per condition (V: crosses, AVc: filled circles and AVi: open circles). Bars denote are two SEM. A sound synchronized with the visual target color change fastens RTs for all temporal rates (Fig. 3a) and improves target detection only below 1.4 Hz (Fig. 3b). The level of significance between AV (AVc and AVi) and V conditions are reported as follows: *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$, #: $p < 0.055$.

Identification was significantly worse in AVi compared to V, below 3.3 Hz (AVi vs. V: $Z_{0.6\text{Hz}} = 2.5$, $p < 0.05$; $Z_{0.8\text{Hz}} = 2.9$, $p < 0.01$; $Z_{1.1\text{Hz}} = 3.8$, $p < 0.001$; $Z_{1.4\text{Hz}} = 5.2$, $p < 0.001$; $Z_{2\text{Hz}} = 4.3$, $p < 0.001$; $Z_{3.3\text{Hz}} = 1.4$, $p = 0.15$; $Z_{10\text{Hz}} < 1$); for temporal rates below 1.4 Hz, identification in AVc was significantly better than in V (AVc vs V: $Z_{0.6\text{Hz}} = -3.5$, $p < 0.001$;

$Z_{0.8\text{ Hz}} = -3.3, p < 0.01$; $Z_{1.1\text{ Hz}} = -2.7, p < 0.01$; $Z_{1.4\text{ Hz}} = -1.8, p = 0.07$; $Z_{2\text{ Hz}} < 1$; $Z_{3.3\text{ Hz}} = -1.9, p = 0.055$; $Z_{10\text{ Hz}} = -1.4, p = 0.2$).

These results suggest that true AV benefits in search efficiency (considering both RT and identification rate) are constrained to temporal rates below 1.4 Hz: informational gain could be uniquely obtained in the range of temporal rates in which auditory information affects visual analysis and/or the robustness of the target representation and within which the temporal structure of events can be tracked.

2.2.2. Two search regimes based on temporal rate

To establish whether AV search operates in two modes based on the temporal structure of events (namely, one of automatic AV integration at low temporal rates and one of AV competition at higher temporal rates), data were divided into two groups (temporal rates below and above 1.4 Hz; preliminary statistical analysis was conducted to determine this grouping albeit details are not reported here for sake of clarity).

A main effect of display condition on RTs was found for the below 1.4 Hz group (AVc vs. AVi: $t = 14.77, p < 0.001$; AVc vs. V: $t = 4.58, p < 0.001$; AVi vs. V: $t = -9.99, p < 0.001$) and the above 1.4 Hz group (AVc vs. AVi: $t = 11.54, p < 0.001$; AVc vs. V: $t = 5.09, p < 0.001$; AVi vs. V: $t = -6.27, p < 0.001$). Similarly, a main effect of display condition was found for identification rates in the below 1.4 Hz group (AVc vs. AVi: $Z = -10.14, p < 0.001$; AVc vs. V: $Z = -5.41, p < 0.001$; AVi vs. V: $Z = 5.61, p < 0.001$) and in the above 1.4 Hz group (AVc vs. AVi: $Z = -7.71, p < 0.001$; AVc vs. V: $Z = -2.06, p < 0.05$; AVi vs. V: $Z = 5.62, p < 0.001$). However, differences in identification rate between AVc vs. V and AVi vs. V conditions obtained in the below 1.4 Hz group were twice as large as those obtained in the above 1.4 Hz group (Figure 4).

More importantly, the effect size was only found to be significant for the below 1.4 Hz group between the conditions of interests (Figure 4), namely RTs and identification rates were less affected by the number of distracters in AVc than in V in this group (RT: AVc slope value: 1 ± 3 ms / item, V slope value: 11 ± 3 ms / item; $t = -2.1, p < 0.05$; identification AVc slope value: -0.1 ± 0.02 % /item, V slope value: -0.2 ± 0.07 % /item: $z = 2.1, p < 0.05$). The slope of RTs was also found to be much steeper in AVi than in AVc (AVi slope value: 13 ± 4 ms/item; $t = -2.0, p < 0.05$). Additionally, no difference in slopes was found between AVi and

V (RTs: $t = 0.3$, ns, identification: $z = 0.2$, ns). The major decrease in AVc slope compared to other conditions suggests that the target is more immune to the presence of visual distracters in this display condition; 1.4 Hz appears to be a temporal rate below which a synchronous sound automatically improves visual search efficiency by helping the individualization or segregation of visual targets presented in a dynamic stream of events.

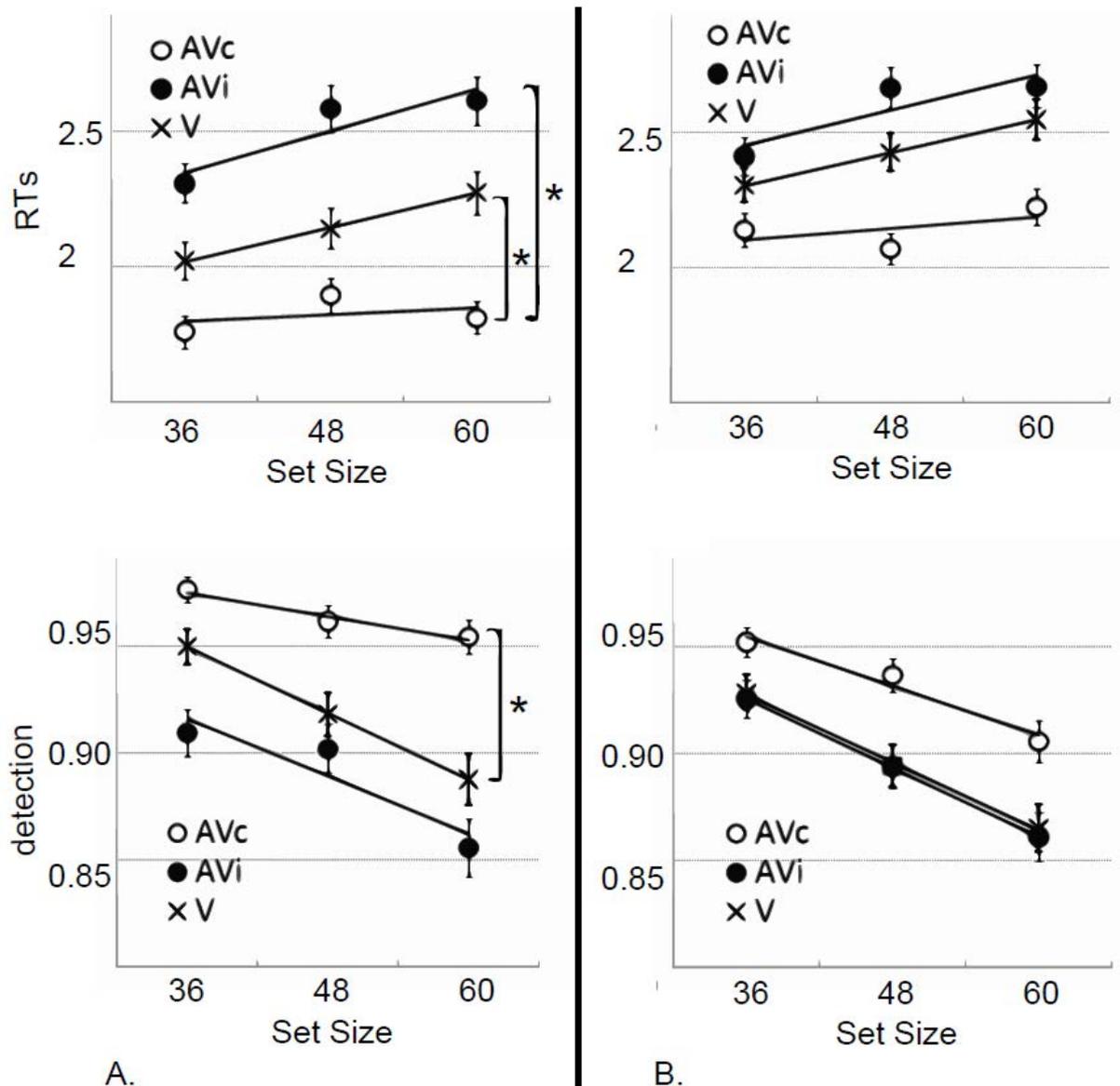


Figure 4: Grouped RTs and identification rates as a function of set size. Grouped RTs (upper panels) and identification rates (lower panels) as a function of set size for temporal rates under (A) and above (B) 1.4Hz. Bars are two SEM. In the below 1.4Hz group, slopes in AVc are shallower than in AVi and V conditions; this suggests that visual search is less impaired by distracters in AVc. In the above 1.4Hz group, no significant differences in slopes were found.

3. DISCUSSION

In this study, we first replicated prior findings showing that RT and identification of a visual target in a dynamic conjunction search task is more efficient when a transient sound is synchronized with the visual target color change (AVc) at 1.1 Hz [20]. Additionally, we showed that visual search efficiency is impaired in AVi - sound synchronized with a visual distracter color change - compared to V or AVc. Crucially, we showed the existence of two temporal regimes: one in which AVc search reveals an automaticity profile for temporal rates below 1.4 Hz (no effect of number of distracters on RT or identification rate), and the other in which a competition profile is seen above 1.4 Hz.

The 1.4 Hz temporal limit found here may provide an insight on how to disambiguate conflicting results in the literature: whereas some studies have supported the existence of automaticity in AV integration [20], others have postulated that AV integration was post-attentional [22]. Our first working hypothesis was that both studies used a very different temporal rate (1.1 Hz in the former, ~10 Hz in the latter). We now show that these temporal rates are indeed in and out of the efficiency search range observed here, respectively. Thus, our study suggests that the presence of transient events may be a necessary [21] but insufficient condition for automatic AV integration: specifically, we suggest that *shared temporal structure* between auditory and visual events matters.

3.1. AV attentional selection depends on temporal structure

In this paradigm, two properties were shared between audition and vision: first, the transience of a sound aligned in time with the abrupt visual color change of the target or distracter and second, the temporal context, namely the relative temporal history between past AV events. One question is thus whether AV synchrony is a determining factor in sensory selection [20,21] or whether the temporal structure of sensory events also matters for the observed improvements in AV search efficiency.

First, it has previously been argued that AV synchrony and the transience of events were critical factors for improved visual search efficiency [21]. If improved search efficiency solely

relied on instantaneous AV integration (i.e. independently of the temporal context), the temporal stimulation rate effects could be argued to be a consequence of visual temporal crowding effects. Specifically, as the number of visual event changes increase (with faster temporal rate), visual search efficiency should decrease (slower RT, poorer identification rate). Such temporal crowding effect in vision should equally affect search efficiency irrespective of the modality of presentation (V, AVc, AVi). Our results clearly suggest that this is not the case: the profiles observed as a function of distracter number under different temporal rates indicate specificities both in the RT and in the identification rates with distinct patterns under and above 1.4 Hz. It could be argued that at higher temporal rates, the search becomes inefficient due to the inability to extract temporal anchors from the acoustic stream. Recent evidence [39] suggests that multisensory integration is pre-attentive when the spatial location of a visual stimulus is already resolved. The discrepancy between Alsius *et al.* [39] and van der Burg *et al.* [20] were deemed to rely on the nature of audiovisual stimuli: AV speech in the former case, transients AV events in the latter. If transient sounds can provide strong temporal anchors for the parsing of visual events thereby enhancing visual spatial search, temporal information extracted from auditory speech stream may be more subdued and less informative for visual segmentation [39]. In our experiment, transient tones were used (similar to [20]). It could be argued that temporal cues provided by the sound become weaker as temporal rate increases, thereby alleviating any benefit for visual search. However, the fastest temporal rate in our experiment was 10 Hz (one 15 ms tone every 100 ms on average). Even at this high display rate, sounds are perceptually discrete and preserve the ability to affect in a non-random fashion the discrimination of temporal visual structure [12]. One interesting question is thus whether systematic manipulation of the salience in the natural temporal modulation of AV speech could affect the results reported in Alsius *et al.* [39].

Second, the temporal distance between visual events (target and distracters alike) was carefully controlled so that when a sound occurred, only one auditory and one visual event could be integrated at a time (see Methods). For most temporal rates tested here (except arguably 10 Hz, see Methods) the temporal distance was large enough for the visual target and the sound cue to integrate. If improvement in visual search efficiency solely relied on the integration of a single AV occurrence, no temporal rate effect should be observed. Hence, the significant temporal rate effects suggest that AV synchrony is not the sole factor in the sensory selection process: namely, the *temporal context* plays a critical role (in vision see also[40-42]).

The role of temporal context can further provide an account of the empirical discrepancies regarding the effect of a temporally incongruent sound in visual search. In our study, the sound-distracter pairing was temporally uncorrelated with the target's color change, i.e. no temporal structure was shared between distracters and targets. In van der Burg *et al.* [20], the authors reported that a sound could improve visual search even when synchronized with a visual distracter: when a sound-distracter's color change shortly preceded the target's color change, efficient search was observed. However, in this control condition, the sound preserved its temporal cueing property with respect to the target: the sound-distracter pair and the target was systematically and on average separated by 200 ms. It is thus not entirely surprising that the sound-distracter pair kept on improving the target detection in [20] since temporal correlations between sound-distracter and target color changes were maintained.

For these reasons, we suggest that the improvement (impairment) observed in AVc (AVi) search efficiency below 1.4 Hz originates from AV perceptual grouping in time. One important issue in multisensory integration is whether the identity of AV information matters in the integration process. This has led to two hypotheses: “multisensory enhancement hypothesis” or “perceptual grouping” hypothesis [29, 30]. Although our paradigm essentially addressed the former issue, our results are consistent with the ‘common-fate’ Gestalt principle, namely, features that have the same dynamics are more likely to be perceived as belonging to the same perceptual object. This has previously been demonstrated in visual [40-42] and in auditory grouping [43-44]. Our data add to this literature and suggest that a similar principle may be driving AV integration for certain temporal regimes. AV synchrony [20] but also, and crucially, AV temporal structure are fundamental to perceptual grouping in AV integration.

3.2. AV temporal prediction benefits visual encoding

In line with - and as a multisensory extension of – the original DAT proposal [8] (Figure S1 panel A), selective attention can fluctuate in time and predict the arrival of future events based on the rate of presentation of the preceding stimuli. This, we suggest, may occur irrespective of the sensory modality of input. Additionally, the rhythmic occurrence of AV events may

enable the elicitation of a (AV) temporal expectancy profile. It is noteworthy that two to three auditory events occurred before the visual display was shown thereby enabling a temporal expectancy profile to emerge even before the visual target first changed color. This may partly account for the overall faster RT observed in our study as compared to original findings [20].

Importantly, at rates below 1.4 Hz, both RTs and identification rates improved suggesting specificity in the AVc integration process. As observed within modalities [5-7], AV stimulus repetition may improve the precision of visual encoding. For instance, in agreement with the repetition-expectation effect [45], the extraction of visual information may be enhanced by the temporal predictability of the visual target based on the AVc temporal rate. Said differently, the shared temporal structure of AV events enables precise temporal prediction of the timing of the visual target change (Figure S1, panel B) . In AVi condition, AV integration is prevented as sounds cue for the color change of a distracter (Figure S1, panel C). Additional experiments are needed to explore to which extent auditory and visual streams may enter in competition for attentional selection at these rates.

Like the DAT, alternative approaches based on the temporal statistics of events [46, 47] predict the establishment of an expectancy profile after the presentation of rhythmic stimuli. Based on the interval-based mechanism of perceptual timing [46], the extraction of temporal properties of a stimulus relies on the memory of interval durations between previous stimuli, not on the synchrony of events entrained by an internal oscillator. In our study, the arrival time of the next AV event would thus be computed based on the distribution of previous temporal delays between AV events. The central tendency and the dispersion of the distribution could encode the nature and the strength of the temporal expectancy. However, our main finding cannot be accounted for by these models: specifically, in case of jittered rhythmic stimulation, such models predict an increase in strength and accuracy of the temporal prediction. In this paradigm, this would translate into a more efficient search for high temporal rates – considering that more events are displayed per second. The opposite effect was found here.

3. 3. The attentional selection threshold is consistent with neurophysiological findings

Recent neurophysiological findings have suggested that attentional selection across sensory modalities may be implemented as entrainment of neural oscillations whether stimuli are rhythmic or present inherently a complex temporal structure [23-25,28, 48-50]. Using simple rhythmic AV stimuli, the neural entrainment of auditory and visual cortices has been demonstrated in the 1-2 Hz range (“delta band”) [23- 25].

One mechanistic view of brain function is that cortical oscillations naturally impose their temporal granularity on the parsing of sensory information. This has been shown in speech [51], in vision [48, 52] and extended to AV parsing [3]. If AV attentional selection operates in the 1-2 Hz range as suggested by monkey neurophysiology work [23- 25], this mechanism should bear functional relevance to the central question of automaticity in AV integration [53]. These findings constitute a major prediction for the existence of a temporal boundary for AV attentional selection.

The 1.4 Hz limit in AV search efficiency is thus in line with neurophysiological predictions: specifically, neural entrainment above that temporal modulation would lead to a processing bottleneck of event tracking in time (Figure S1, panel B, cases illustrating the > 1.4 Hz). Neural entrainment is characterized by an increased neural excitability at a particular phase of the entrained oscillation: if (i) neural entrainment is conceived as the mechanistic implementation of the expectancy profile hypothesized in the DAT [7] and (ii) auditory stimuli can entrain oscillations in visual cortices [24], then our results suggest that the encoding of visual events co-occurring with the sound will be more efficient at the time predicted by the auditory stimuli. In this context, the encoding of a visual event is as efficient for a target as for a distracter, as long as it shares its temporal structure with the auditory stream. Hence, when the visual event is a target, RTs and identification benefit from this automatic attention selection mechanism; when the visual event is a distracter, this mechanism impairs efficient detection of the target. In AVi, the automaticity of temporal parsing induced by the auditory rhythm hinders, and perhaps competes with, the detection of the visual stream that does not share the same temporal structure. Indeed, in AVi, the visual target stream cannot be tracked automatically and requires additional attentional resources as can readily be seen with the RTs increase and the lower identification rate irrespective of temporal rates (Figure 3). This attentional selection mechanism provides specific and testable neurophysiological predictions of increased (AVc) and decreased (AVi) search efficiency - or decreased and increased AV competition, respectively (Figure S1, panel B and C).

One limitation remains in both our study and early neurophysiological findings [23- 25], namely: the temporal structure of events is imposed by the stimulation rate and not necessarily imposed by ongoing neural oscillations. Automaticity is demonstrated in the context of salient entrainment of AV stimuli but it is unclear whether a similar limit on the automaticity of attentional selection would be observed for AV stimuli with more complex dynamics (e.g. as in AV speech [39]) or using a very different paradigmatic approach (e.g. as in [29-30]). Nevertheless, a recent behavioral study [54] using non-rhythmic stimulation further suggests that similar selection attentional mechanisms can affect perception: in a visual detection task, hit rates were shown to change periodically through time and maximum hit rates were phase-locked to the sound onset. The authors reported that the hit rates periodicity approximated 1 Hz although different temporal rates were not explicitly tested.

3.5.Conclusion

In a visual conjunction search paradigm, sounds can improve and impair search efficiency when synchronized or desynchronized with a visual target, respectively. Major improvements in search efficiency are limited to temporal stimulation rate slower than 1.4 Hz whereas impairments are consistent across temporal rates. Our results are interpreted in the context of the DAT [8] in the temporal frequency range predicted by monkey and human neurophysiology [23, 24, 25]: specifically, brain rhythms in the 1-2 Hz range naturally impose a limit on the attentional selection of events in time irrespective of sensory inputs. This can be considered a temporal Gestalt that operates at a slow rate across sensory modalities and enables automatic audiovisual integration.

4. MATERIAL AND METHODS

4.1.Subjects

Twenty-four volunteers (13 females, mean age: 22.5 years old) participated in the study. All had normal, corrected-to-normal vision, normal color vision and normal hearing, and were

naive as to the purpose of the study. Each participant provided an informed consent in accordance with the Declaration of Helsinki (2008) and the Ethics Committee on Human Research at NeuroSpin (Gif-sur-Yvette, France).

4.2. Stimuli

Experiments were run in a darkened soundproof cabin. Participants were positioned on a headrest apparatus 70 cm away from a Viewsonic CRT monitor (19", 60 Hz). Auditory stimuli were presented via two speakers located on each side of the monitor. Visual stimuli consisted of an array of colored bars displayed on a black background (Figure 1). All bars were the same size (length: 0.57° ; width: 0.19°) and randomly placed on a circular display with maximal eccentricity at 30° . All bars had random orientations except for the target which was vertical or horizontal. In each trial, the set size was 36, 48 or 60. A target could never appear within a radius of 3° around the white fixation point. In the initial frame, a color (red and green) was randomly assigned to each bar. All bars changed color through time. The timing of color changes was manipulated so that they always occurred at a given average temporal rate within one trial (but differed across trials). The temporal rates (F) tested were 0.56, 0.77, 1.1, 1.4, 2, 3.3 and 10 Hz. For a given trial presented at F, the delay between two color changes of a given bar was randomly chosen following a normal distribution with a mean of $1/F$ and a standard deviation of $1/4F$. Three modalities of presentation were examined. In V, visual stimuli were displayed without any sound. In AV, a 15 ms (incl. 5 ms fade-in and -out) 2 kHz tone (44.1 kHz sample rate, 16 bit, mono) was synchronized with the color changes of a given bar in the display. In AVc, the sound was synchronized with the color change of the target; in AVi, the sound was synchronized with a randomly chosen distracter (the same one within a trial). Importantly, a sound had to be synchronized with only one bar at a time: to minimize the perceived synchrony between the color changes of the distracters and the sound, a temporal margin surrounding the sound/target onset was introduced during which no bars could change color. This temporal margin was scaled on the tested F: ± 16.7 ms for 10 Hz, ± 50.1 ms for 3.3 Hz, ± 83.5 ms for 2.0 Hz, and ± 117 ms for the remaining rates. In V and AVc, the temporal protection margin was applied to the target; in AVi, it was applied to the distracter.

4.3.Procedure

Participants were asked to find as fast and as accurately as possible the target while maintaining their gaze on a central fixation point. Each trial started with the presentation of the fixation point for a random duration (1-4 seconds) followed by the visual display. In AVc and AVi, two or three sounds were played before the visual onset to avoid surprise effects at the onset of the first sound. The presence of a sound synchronized with the target was expected to improve the speed of target detection [20]. In our paradigm, the sound onset was directly tied to F, namely, the higher the F, the earlier the auditory onset. To avoid a confounded faster RT, the first color change of the target occurred systematically at 300 ms after the display onset in all conditions. After detection or after 10 s has elapsed, participants reported the orientation of the target in a 2-Alternative-Forced-Choice (vertical or horizontal). The efficiency in visual search was quantified in terms of reaction times (RTs) and correct detection rate. We excluded the RTs from trials in which the target was not detected within 10 s (14% of the trials) from the data analysis. Each condition was repeated 15 times. The experiment was run in 3 blocks corresponding to the modality of presentation. Participants were told to ignore sounds as they were irrelevant to the task. The order of block presentation was counterbalanced across participants. The first block was used as a training session for all participants. The analysis focused on the last two blocks, when participants had reached asymptote on the task.

4.4 Statistical analysis

Statistical analysis was performed using Linear Mixed Effects models [32] with R [55] (R Foundation for statistical computing). Linear mixed models can be thought of as a generalization of linear regression models: in mixed regression models, data are not aggregated, and statistics are made on all observations. Specifically, participants were considered as a random effect and separate regression models were fitted to the entire data set (i.e. one for each participant). This approach increases statistical power without over-fitting the data. On the contrary, classical regression models and repeated measures ANOVAs are

based on the comparisons of measured means according to variables of interest (or fixed factors). Hence, unlike repeated measures ANOVAs in which comparisons are made between averaged data (information carried out by each observation is lost), in the mixed models used here, each observation is taken into account while considering the variability between subjects as a random effect. Additionally, the analysis of a categorical dependent variable (e.g. identification rate) is possible using a logistic mixed regression whereas ANOVAs may bring spurious results [33].

Thus, we selected this method as it is best suited for this study: fixed factors were display condition (3: AVc, AVi, and V), set size (continuous factor) and temporal rate (7 discrete levels). Subjects (n=24) were a random effect. We considered ‘set size’ as a continuous factor because the RTs and identification rates as a function of set size fit well with the assumptions of a linear regression. We considered ‘temporal rate’ as a discrete factor because the dependency of RTs and identification rates as a function of temporal rate is not linear. Significant fixed factors can be assessed in two ways: (i) a regression model simplification using the Akaike Information Criterion (AIC) or (ii) the likelihood ratio using Chi square. The AIC is a measure that optimizes model fit by taking into account the amount of explained variance as well as the degrees of freedom. This procedure ensures that the obtained model achieves the best fit to the data with the minimum number of predictor variables. When two models are compared, the AIC provides information about whether the predictors added in the second model account for a significant amount of variance in the dependent variable. The best model corresponds to the minimal AIC. For instance, in the reported tables (e.g. Table 1), the list of models is provided along with their respective AIC index. The model that best fit the data is the one with the minimal AIC, here model 5 (for the RTs) and model 4 (for the identification rates). Consistent with this, the best models can also be found using Chi square.

The best model using the likelihood measure is defined by a significant Chi square test ($P < \chi^2$) comparing one model in the list to the next (e.g.: model 1 vs. 2, then model 2 vs. 3 and so on). The last comparison providing a significant effect points to the best model: namely, in our example, model 5 (RTs) and 4 (identification rates). The “ χ^2 ” value corresponds to twice the difference of the log likelihood of the two models. Both AIC and χ^2 values are reported in Table 1.

Simpler models (for instance, let’s consider models 1 to 4 for RTs in Table 1) do provide crucial information. Low AIC or significant Chi square tests for these models are interpreted as follows: the factor of interest (e.g. model 2, factor of set size) significantly impacts the

model fit irrespective of all other factors – and hence, has a significant effect in our paradigm. This is analogous to stating a “main effect” for the more classic ANOVA approach. Here, the procedure is iterative such that adding another factor may enable better model fit (e.g. model 3 and so on) leading to the preferred model that explains most of the data (model 5 in our example). Hence, all factors up to model 5 (including their interactions) showed significant RTs effect.

The lme4 package [56] was used to obtain parameter estimates and the language package [32] was used to obtain the reported p-values. The ‘lmer’ function yielded regression coefficients and related t statistics (exclude degrees of freedom), and p-values were derived from a Markov Chain Monte Carlo (MCMC) [32]. Statistical tests were carried out on the contrast coefficients resulting from the selected linear mixed effect model. For instance, contrasting two levels for the display condition (AVc vs. V) yields a contrast coefficient submitted to a t-test for RTs and Wald test for identification rates. As a rule of thumb, statistical tests reported here vary according to the dependent variables, namely Student t-tests for the RTs and Wald tests for the identification rates [33].

ACKNOWLEDGMENTS

This work was supported by a Marie Curie IRG-249222 to VvW and by Fondation EDF and CEA/DGA fellowships to AK. We thank two anonymous reviewers for their insightful remarks on previous versions of the manuscript. We would like to thank Christophe Pallier for his precious help on statistical analysis, Jean Vroomen and Salvador Soto-Faraco for insightful discussions on preliminary findings of this research. This work was first made public at the 11th International Conference on Cognitive Neurosciences conference in Mallorca, Spain (2011).

REFERENCES

1. Lewicki MS (2002) Efficient coding of natural sounds. *Nature neuroscience* 5: 356-363.
2. Simoncelli EP, Olshausen BA (2001) Natural image statistics and neural representation. *Annual review of neuroscience* 24: 1193-1216.

3. Schroeder CE, Lakatos P, Kajikawa Y, Partan S, Puce A (2008) Neuronal oscillations and visual amplification of speech. *Trends in Cognitive Sciences* 12: 106-113.
4. Winfree AT (2000) *The geometry of biological time*. New York: Springer. 777 p.
5. Jones MR, Johnston HM, Puente J (2006) Effects of auditory pattern structure on anticipatory and reactive attending. *Cognitive psychology* 53: 59-96.
6. Mathewson KE, Fabiani M, Gratton G, Beck DM, Lleras A (2010) Rescuing stimuli from invisibility: Inducing a momentary release from visual masking with pre-target entrainment. *Cognition* 115: 186-191.
7. Large EW, Jones MR (1999) The dynamics of attending: How people track time-varying events. *Psychological review* 106: 119-159.
8. Jones MR (1976) Time, our lost dimension: toward a new theory of perception, attention, and memory. *Psychological review* 83: 323-355.
9. Barnes R, Jones MR (2000) Expectancy, attention, and time. *Cognitive psychology* 41: 254-311.
10. Treisman AM, Gelade G (1980) A feature-integration theory of attention. *Cognitive psychology* 12: 97-136.
11. Yantis S (1992) Multielement visual tracking: Attention and perceptual organization. *Cognitive psychology* 24: 295-340.
12. Guttman SE, Gilroy LA, Blake R (2005) Hearing what the eyes see: auditory encoding of visual temporal sequences. *Psychological science* 16: 228-235.
13. Fujisaki W, Nishida S (2005) Temporal frequency characteristics of synchrony-asynchrony discrimination of audio-visual signals. *Experimental brain research Experimentelle Hirnforschung Experimentation cerebrale* 166: 455-464.
14. Fujisaki W, Nishida S (2007) Feature-based processing of audio-visual synchrony perception revealed by random pulse trains. *Vision research* 47: 1075-1093.
15. Fujisaki W, Nishida S (2010) A common perceptual temporal limit of binding synchronous inputs across different sensory attributes and modalities. *Proceedings Biological sciences / The Royal Society* 277: 2281-2290.

16. Benjamins JS, van der Smagt MJ, Verstraten FA (2008) Matching auditory and visual signals: is sensory modality just another feature? *Perception* 37: 848-858.
17. Shipley T (1964) Auditory Flutter-Driving of Visual Flicker. *Science* 145: 1328-1330.
18. Welch RB, DuttonHurt LD, Warren DH (1986) Contributions of audition and vision to temporal rate perception. *Perception & psychophysics* 39: 294-300.
19. Gebhard JW, Mowbray GH (1959) On discriminating the rate of visual flicker and auditory flutter. *The American journal of psychology* 72: 521-529.
20. van der Burg E, Olivers CNL, Bronkhorst AW, Theeuwes J (2008) Pip and pop: Nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology-Human Perception and Performance* 34: 1053-1065.
21. van der Burg E, Cass J, Olivers CN, Theeuwes J, Alais D (2010) Efficient visual search from synchronized auditory signals requires transient audiovisual events. *PloS one* 5: e10664.
22. Fujisaki W, Koene A, Arnold D, Johnston A, Nishida S (2006) Visual search for a target changing in synchrony with an auditory signal. *Proceedings Biological sciences / The Royal Society* 273: 865-874.
23. Gomez-Ramirez M, Kelly SP, Molholm S, Sehatpour P, Schwartz TH, et al. (2011) Oscillatory Sensory Selection Mechanisms during Intersensory Attention to Rhythmic Auditory and Visual Inputs: A Human Electrographic Investigation. *Journal of neuroscience* 31: 18556-18567.
24. Lakatos P, Karmos G, Mehta AD, Ulbert I, Schroeder CE (2008) Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science* 320: 110-113.
25. Besle J, Schevon CA, Mehta AD, Lakatos P, Goodman RR, et al. (2011) Tuning of the human neocortex to the temporal dynamics of attended events. *Journal of neuroscience* 31: 3176-3185.
26. Galambos R, Makeig S, Talmachoff PJ (1981) A 40-Hz potential recorded from the human scalp. *Proceedings of the National Academy of Sciences of the United States of America*. 78: 2343-2347

27. Regan D (1966) Some characteristics of average steady-state and transient responses evoked by modulated light. *20*: 238-248
28. Schroeder CE, Lakatos P (2009) Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences* 32: 9-18.
29. Chen YC, Yeh SL (2009) Catch the moment: multisensory enhancement of rapid visual events by sound. *Experimental Brain Research*. 198: 209-219
30. Chen YC, Yeh SL (2008) Visual events modulated by sound in repetition blindness. *Psychonomic bulletin & review*. 15: 404-408
31. Shams L, Kamitani Y, Shimojo S (2000) What you see is what you hear. *Nature*. 408: 788
32. Baayen RH, Davidson DJ, Bates DM (2008) Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59: 390-412.
33. Jaeger TF (2008) Categorical Data Analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language* 59: 434-446.
34. Santee JL, Egeth HE (1982) Do reaction time and accuracy measure the same aspects of letter recognition? *Journal of experimental psychology Human perception and performance* 8: 489-501.
35. McDonald JJ, Teder-Salejarvi WA, Hillyard SA (2000) Involuntary orienting to sound improves visual perception. *Nature* 407: 906-908.
36. Yantis S, Jonides J (1984) Abrupt visual onsets and selective attention: evidence from visual search. *Journal of experimental psychology Human perception and performance* 10: 601-621.
37. Verstraten FA, Cavanagh P, Labianca AT (2000) Limits of attentive tracking reveal temporal properties of attention. *Vision research* 40: 3651-3664
38. Posner MI, Boies SJ (1971) Components of Attention. *Psychological review* 78: 391-&.
39. Alsius A, Soto-Faraco S (2011) Searching for audiovisual correspondance in multiple speaker scenarios. *Experimental Brain Research* 213: 175-183
40. Blake R, Lee SH (2005) The role of temporal structure in human vision. *Behav Cogn Neurosci Rev* 4: 21-42.

41. Lee SH, Blake R (1999) Visual form created solely from temporal structure. *Science* 284: 1165-1168.
42. Kandil FI, Fahle M (2001) Purely temporal figure-ground segregation. *The European journal of neuroscience* 13: 2004-2008.
43. French-St George M, Bregman AS (1989) Role of predictability of sequence in auditory stream segregation. *Perception & psychophysics* 46: 384-386.
44. Turgeon M, Bregman AS (2001) Ambiguous musical figures. Sequential grouping by common pitch and sound-source location versus simultaneous grouping by temporal synchrony. *Annals of the New York Academy of Sciences* 930: 375-381.
45. Huang L, Pashler H (2005) Expectation and repetition effects in searching for featural singletons in very brief displays. *Perception & psychophysics* 67: 150-157.
46. Keele SW, Nicoletti R, Ivry RI, Porkony RA (1989) Mechanisms of perceptual timing: beat-based or interval-based judgements? *Psychological Research*. 50: 251-256
47. Drake C, Botte MC (1993) Tempo sensitivity in auditory sequences: evidence for a multiple look model. *Attention, Perception & Psychophysics*. 54: 277-286
48. Busch NA, Dubois J, VanRullen R (2009) The phase of ongoing EEG oscillations predicts visual perception. *Journal of neuroscience* 29: 7869-7876.
49. Luo H, Liu Z, Poeppel D (2010) Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS biology* 8: e1000445.
50. Stefanics G, Hangya B, Hernadi I, Winkler I, Lakatos P, et al. (2010) Phase entrainment of human delta oscillations can mediate the effects of expectation on reaction speed. *Journal of neuroscience* 30: 13578-13585.
51. Luo H, Poeppel D (2007) Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54: 1001-1010.
52. VanRullen R, Reddy L, Koch C (2006) The continuous wagon wheel illusion is associated with changes in electroencephalogram power at approximately 13 Hz. *Journal of neuroscience*. 26: 502-507.

53. Talsma D, Senkowski D, Soto-Faraco S, Woldorff MG (2010) The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Science*. 14: 400-410
54. Fiebelkorn IC, Foxe JJ, Butler JS, Mercier MR, Snyder AC, et al. (2011) Ready, set, reset: stimulus-locked periodicity in behavioral performance demonstrates the consequences of cross-sensory phase reset. *Journal of neuroscience*. 31: 9971-9981.
55. R development core team (2007). R: a language and environment for statistical computing. Vienna: R Foundation for statistical computing. <http://www.R-project.com>
56. Bates DM, Sarkar D (2007) lme4: Linear mixed-effects models using S4 classes. R package version 0.99875-6

SUPPLEMENTARY MATERIAL

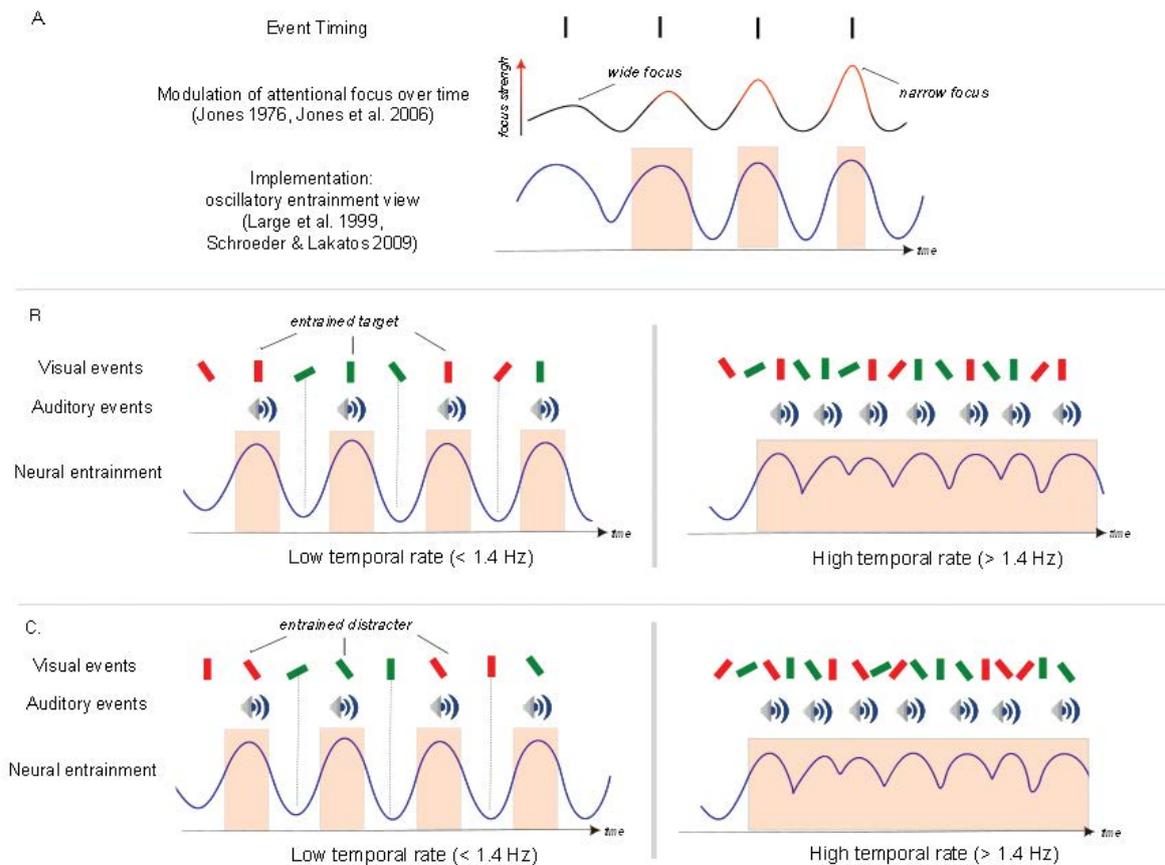


Figure S1: Dynamic Attending Theory (DAT), neural implementation as oscillatory entrainment and relevance for findings on AV selective attention. (A) The DAT [8] postulates that attention is a dynamical process which oscillates in time and entrains to the temporal structure of events. Event Timing: dynamics of stimuli in a scene. Stimuli need not be isochronous – for illustrative purposes, events are represented with a particular rhythm. Events can be auditory or visual. Modulation of attentional focus over time: a temporal expectation profile builds up over time (i.e. after several occurrence of a same event) leading to a narrowing of attentional focus (from “wide” to “narrow”, [5]). The “narrow foci” are also times of high expectation (temporal prediction). Thus, the attentional profile oscillates between periods of high and low temporal expectation. Implementation: one suggested implementation of the DAT [5,7] is via an oscillatory mechanism represented here as a simple waveform entrained to the rhythm of events. Recent neurophysiological evidence has suggested a similar neural implementation for attentional selection across auditory and visual sensory modalities, specifically with neural oscillations in the 1–2 Hz range [23–25]. In neural terms, high temporal expectations (or narrow attentional foci) are periods of high neural excitability. The encoding of events at the entrained rhythm is more efficient during period of increased neural excitability. For synchronized AV events, the auditory entrainment of oscillations in visual cortices leads to high expectation/excitability periods synchronized to the sound [23–25]. We now illustrate the implications for the AVc and AVi conditions tested in this study. (B) In AVc, the high expectation/excitability period is aligned to the target enabling faster RTs and improved identification rate. (C) In AVi, these periods are aligned with a distracter, leading to slower RTs and poorer identification rate. Temporal rate effects: modulation of visual search efficiency by the temporal rate of AV displays. The working hypothesis was that visual search efficiency would be frequency-specific: search would be efficient in the range of oscillatory entrainment but not above. Data revealed a 1.4 Hz boundary. Neural predictions: oscillations in visual cortex are entrained to the sound. Neural entrainment (alternation of high and low excitability phases) yields an expectancy profile favoring the encoding of visual events synchronized with the sounds (either the target (B), or one distracter (C)). When the temporal rate is above the oscillatory mechanism (B or C, right panels), the sound phase resets the entrained oscillation before it reaches a low excitability state. As the system is continuously solicited, no expectancy profile can be built and visual targets cannot benefit from the sound despite sharing the same temporal structure.

RTs	0.8 Hz	1.1 Hz	1.4 Hz	2.0 Hz	3.3 Hz	10 Hz
0.6 Hz						
<i>contrast coefficient</i>	<i>-0.06</i>	<i>0.01</i>	<i>-0.04</i>	<i>0.1</i>	<i>0.31</i>	<i>0.47</i>
t value, significance	-1 ns	0.2 ns	-0.7 ns	1.7 ns	5.7 ***	8.0 ***
0.8 Hz						
		0.07	0.02	0.16	0.39	0.54
		1.2 ns	0.3 ns	2.7 *	6.6 ***	8.9 ***
1.1 Hz						
			-0.05	0.09	0.33	0.47
			-0.8 ns	1.5 ns	5.4 ***	7.7 ***
1.4 Hz						
				0.14	0.38	0.52
				2.3 *	6.2 ***	8.5 ***
2.0 Hz						
					0.24	0.38
					3.8 ***	6.1 ***
3.3 Hz						
						0.14
						2.2 *

Table S1: Effect of temporal rate on RTs irrespective of display condition (V, AVc and AVi combined). Table shows contrast coefficients (italics, regression coefficients referring to contrast between two levels of one factor) between the different temporal rates and their related t values. Statistics were computed using mixed regression analysis with model 4 (cf. Table 1). Corrected p values were estimated using a Monte Carlo procedure. The reported significance values are as follows: *p,0.05; **p,0.01; ***p,0.001

Identification	0.8 Hz	1.1 Hz	1.4 Hz	2.0 Hz	3.3 Hz	10 Hz
0.6 Hz						
<i>contrast coefficient</i>	<i>-0.1</i>	<i>-0.13</i>	<i>-0.12</i>	<i>-0.16</i>	<i>-0.54</i>	<i>-0.57</i>
Z value, significance	-0.4 ns	-0.9 ns	-1.0 ns	-1.4 ns	-4.8 ***	-5.2 ***
0.8 Hz						
		-0.05	-0.05	-0.08	-0.47	-0.45
		-0.5 ns	-0.5 ns	-0.8 ns	-4.2 ***	-4.5 ***
1.1 Hz						
			0.05	-0.23	-0.42	-0.44
			-0.5 ns	-1.9 ns	-3.8 ***	-4.1 ***
1.4 Hz						
				-0.04	-0.42	-0.45
				-0.4 ns	-4.0 ***	-4.1 ***
2.0 Hz						
					-0.39	-0.42
					-3.5* **	-3.8 ***
3.3 Hz						
						0.03
						0.7 ns

Table S2: Effect of temporal rate on identification rate irrespective of display condition (V, AVc and AVi combined). Table shows contrast coefficients (italics, regression coefficients referring to contrast between two levels of one factor) between the different temporal rates and their related Z values (Wald tests). Statistics were computed using mixed regression analysis with model 4 (cf. Table 1). Corrected p values were estimated using a Monte Carlo procedure. The reported significance values are as follows: *p,0.05; **p,0.01; ***p,0.001

AVc	0.8 Hz	1.1 Hz	1.4 Hz	2.0 Hz	3.3 Hz	10 Hz
0.6 Hz	-0.02 -0.3 ns	0.06 0.7 ns	0.01 0.1 ns	0.13 1.4 ns	0.54 5.7 ***	0.77 8.2 ***
0.8 Hz		0.09 1.0 ns	0.04 0.4 ns	0.15 1.6 ns	0.56 6.0 ***	0.79 8.4 ***
1.1 Hz			-0.05 -0.5 ns	0.07 0.7 ns	0.48 4.9 ***	0.71 7.3 ***
1.4 Hz				0.12 1.2 ns	0.53 5.4 ***	0.76 7.7 ***
2.0 Hz					0.41 4.1 ***	0.64 6.4 ***
3.3 Hz						0.23 2.3 *
AVi	0.8 Hz	1.1 Hz	1.4 Hz	2.0 Hz	3.3 Hz	10 Hz
0.6 Hz	-0.07 -0.6 ns	-0.02 -0.1 ns	-0.004 -0.1 ns	0.06 0.5 ns	0.01 0.1 ns	0.08 0.7 ns
0.8 Hz		0.06 0.1 ns	0.06 0.6 ns	0.13 1.1 ns	0.08 0.7 ns	0.15 1.4 ns
1.1 Hz			0.01 0.1 ns	0.08 0.6 ns	0.03 0.2 ns	0.10 0.8 ns
1.4 Hz				0.06 0.5 ns	0.02 0.2 ns	0.08 0.8 ns
2.0 Hz					-0.05 -0.4 ns	0.02 0.2 ns
3.3 Hz						0.07 0.6 ns
V	0.8 Hz	1.1 Hz	1.4 Hz	2.0 Hz	3.3 Hz	10 Hz
0.6 Hz	-0.08 -0.7 ns	0.01 -0.1 ns	-0.11 -1.1 ns	0.11 1.2 ns	0.43 4.4 ***	0.56 5.6 ***
0.8 Hz		0.07 0.8 ns	-0.03 -0.3 ns	0.20 2.0 *	0.51 5.2 ***	0.64 6.3 ***
1.1 Hz			-0.11 -1.1 ns	0.12 1.2 ns	0.44 4.4 ***	0.56 5.5 ***
1.4 Hz				0.23 2.3 *	0.55 5.5 ***	0.67 6.6 ***
2.0 Hz					0.33 3.1 **	0.45 4.2 ***
3.3 Hz						0.12 1.1 ns

Table S3: Effect of temporal rate on RTs per display condition. Table shows contrast coefficients between temporal rates for each display condition level and their related t values. Statistics were computed using mixed regression analysis with model 4 (cf. Table 1). Corrected p values were estimated using a Monte Carlo procedure. The reported significance values are as follows: *p,0.05; **p,0.01; ***p,0.001.

AVc	0.8 Hz	1.1 Hz	1.4 Hz	2.0 Hz	3.3 Hz	10 Hz
0.6 Hz	-0.13 -0.84 ns	-0.24 -0.97 ns	-0.38 -1.71 ns	-0.72 -3.11 **	-1.28 -5.25 ***	-1.40 -5.82 ***
0.8 Hz		-0.11 -0.13 ns	-0.25 0.89 ns	-0.59 -2.34 *	-1.15 -4.61 ***	-1.27 -5.23 ***
1.1 Hz			-0.19 -0.76 ns	-0.54 -2.23 *	-1.03 -4.52 ***	-1.16 -5.14 ***
1.4 Hz				-0.34 -1.49 ns	-0.84 -3.89 ***	-0.97 -4.52 ***
2.0 Hz					-0.50 -2.46 *	-0.62 -3.15 **
3.3 Hz						-0.13 -0.71 ns
AVi	0.8 Hz	1.1 Hz	1.4 Hz	2.0 Hz	3.3 Hz	10 Hz
0.6 Hz	-0.03 -0.18 ns	-0.20 -1.12 ns	-0.17 -0.98 ns	-0.05 -0.26 ns	-0.03 -0.16 ns	0.14 0.73 ns
0.8 Hz		-0.16 -0.94 ns	-0.14 -0.80 ns	-0.01 -0.08 ns	0 0.01 ns	0.17 0.91 ns
1.1 Hz			0.02 0.14 ns	0.15 0.86 ns	0.17 0.95 ns	0.33 1.90 ns
1.4 Hz				0.13 0.72 ns	0.14 0.81 ns	0.31 1.70 ns
2.0 Hz					0.02 0.09 ns	0.18 0.99 ns
3.3 Hz						0.17 0.89 ns
V	0.8 Hz	1.1 Hz	1.4 Hz	2.0 Hz	3.3 Hz	10 Hz
0.6 Hz	-0.06 -0.28ns	0.02 0.16 ns	0.13 0.64 ns	0.11 0.55 ns	-0.63 -3.36 ***	-0.73 -3.94 ***
0.8 Hz		0.08 0.40 ns	0.19 0.93 ns	0.17 0.83 ns	-0.58 -3.13 **	-0.68 -3.72 ***
1.1 Hz			0.11 0.53 ns	0.09 0.43 ns	-0.67 -3.52 ***	-0.77 -4.11 ***
1.4 Hz				-0.02 -0.09 ns	-0.78 -4.03 ***	-0.89 -4.61 ***
2.0 Hz					-0.76 -3.93 ***	-0.86 -4.51 ***
3.3 Hz						-0.10 -0.63ns

Table S4: Effect of temporal rate on identification rate per display condition. Table shows contrast coefficients between the temporal rates for each display condition and their related Z-values. Statistics were computed using mixed regression analysis with model 4 (cf. Table 1). Corrected p values were estimated using a Monte Carlo procedure. The reported significance values are as follows: *p,0.05; **p,0.01; ***p,0.001.

PHASE ENCODING OF IMPLICIT TIMING

4.1. INTRODUCTION

4.1.1. Motivation

The data from the first study (Chapter 2) suggest that entrainment in auditory cortex inform on the subjective report of explicit audiovisual timing. In addition, these results show that the encoding of timing is already modulated at early stages of sensory processing, i.e. in primary sensory areas. This is consistent with previous reports, which proposed that the modulation of perceived event's timing with rhythmic adaptation is modifying sensory encoding and not decisional biases (Fujisaki et al., 2004; Yamamoto et al., 2012). Importantly, if entrainment affects sensory event encoding, it may not be restricted to explicit temporal tasks. In this third chapter, we tested whether similar time encoding mechanisms could be found in cognitive tasks for which timing is not explicitly computed.

For this, we focused on speech parsing mechanisms. To understand the meaning of a spoken sentence, it is necessary to know when to parse the acoustic signal into representational chunks relevant for speech and language processing, such as phonemes, syllables, or words. This chunking mechanism has been hypothesized to arise through the oscillatory mechanisms in auditory cortex (fig. 4.1) (Poeppel, 2003; Schroeder et al., 2008; Ghitza, 2011; Giraud and Poeppel, 2012). First, the envelope of the speech signal presents some strong rhythmicity at syllabic rate (3-8 Hz). Listening to speech is a classic case of ecological neural entrainment. Accordingly, slow oscillations in the theta range (3-8 Hz) are known to be entrained to the envelope of the stimulus (Ahissar et al., 2001; Howard and Poeppel, 2010; Zion Golumbic et al., 2013). The strong entrainment of slow

oscillations will generate pulsed inhibition of auditory networks excitability, manifested through cross-frequency coupling and specifically through theta phase – gamma amplitude nesting mechanisms. Owing to theta-gamma coupling, neural excitability will align to the relevant features of auditory speech. Altogether, this mechanism affords periods of strong excitability during the encoding of a syllable, and periods of network silence in between syllables. As such, it has been put forward as an efficient code for speech encoding, maximizing information processing at relevant time points (Giraud and Poeppel, 2012).

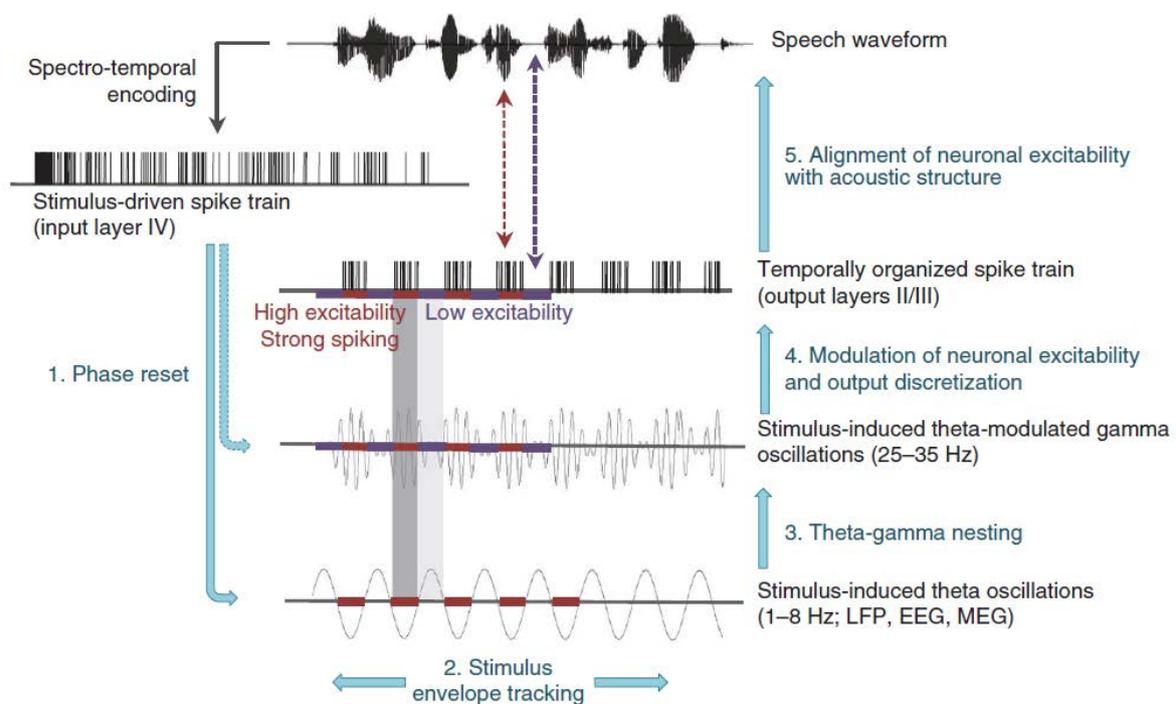


Figure 4.1: Brain oscillations model of speech parsing mechanisms in five steps. 1. Stimulus onset causes phase reset of theta oscillations. 2. Theta oscillations track the speech envelope. 3. Theta modulates gamma activity in time through phase-amplitude coupling. 4. Gamma activity reflects neural excitability. 5. Neural excitability through this process aligns to the fluctuations of the acoustic signal. Adapted from (Giraud and Poeppel, 2012)

However, it also suggests that the coupling between theta and gamma oscillations in auditory cortex should provide information about *when a syllable starts and when it ends*. As such, the minima of the theta modulated gamma power should correspond to the junctions between perceived syllables. This hypothesis predicts that any change of temporal position of the gamma power minima should entail a change in perceived

syllabic junctions. The acoustic signal should be chunked differently, leading to a change in the perceptual content of speech.

4.1.2. Experiment

To test the role of oscillations in the parsing of speech signals, participants listened to ambiguous bistable auditory speech while being recorded with MEG. Specifically, participants listened to a series of words that were repeated at a 3Hz syllabic rate (fig. 4.2.). The repetition of words in sequences is known to induce changes in speech perception called “verbal transformations” (Sato et al., 2006). Of particular interest, these transformations are akin to modulations in the timing of syllable chunking: for instance, the repetition of the word “life” leads to bistable perception with the main concurrent percept being the word “fly”. Similarly in French, the repetition of the word “plan” (map) leads to the alternate percept “lampe” (lamp). Hence the same speech sequence can be perceptually chunked differently, and we asked whether neural oscillations in auditory cortex can account for bistability in auditory speech perception.

If theta-gamma coupling mechanisms are implicated in the chunking of speech into syllables, they should indicate *when* the speech signal is perceptually chunked into percept 1 or into percept 2. Specifically, evidence suggests that there is a fixed relation between slow oscillation phase and network excitability (Lakatos et al., 2005; Haegens et al., 2011). Precisely, maximal network excitability _ associated with gamma power _ should occur at the trough of the slow oscillation. The only possibility for the encoding window to move is to change the phase locking between the slow oscillation and the acoustic signal (fig. 4.2.). Interestingly, the phase of entrainment of slow oscillations is modulated by attention and predictability (Lakatos et al., 2008; Stefanics et al., 2010b). The attentional modulation of low-frequency oscillations has been shown to be crucial when listening to speech in a crowded environment (Zion Golumbic et al., 2013). In addition, we have shown in chapter 2 that entrainment is not steady and could be modulated in time. It is probable that, when listening to an ambiguous speech sequence, the phase of oscillatory entrainment slowly shifts in time. If speech parsing mechanisms uses the entrained neural oscillation to cut the acoustic signal into syllables, then a shift in the phase of entrainment should cause a shift in the syllabic parsing window (fig. 4.2.). In

this experiment, we tested whether bistable chunking of the acoustic signal was caused by a shift in the oscillation that is entrained at the syllabic rate.

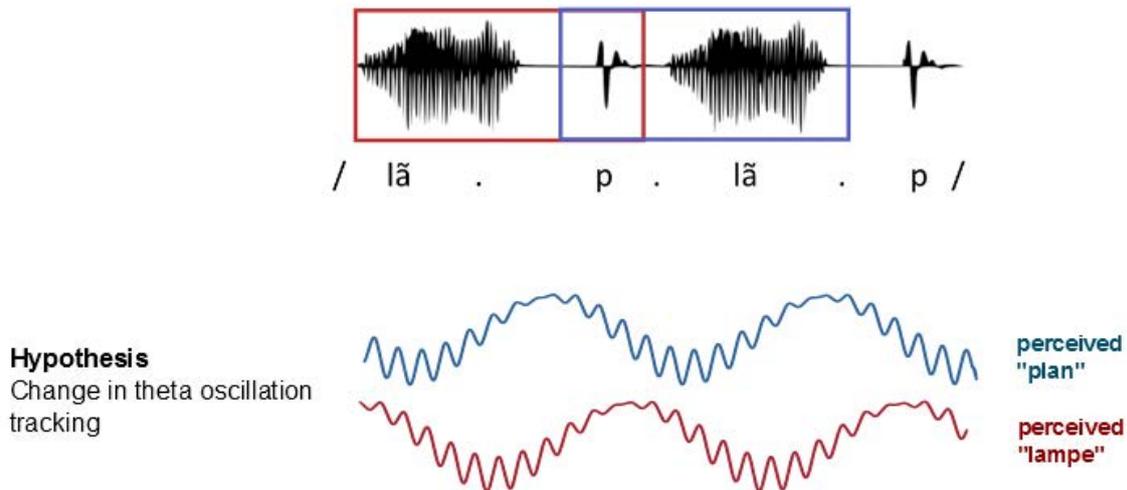


Figure 4.2: Working hypothesis. Verbal transformations can be caused by phase shifts in the theta oscillation entrainment to syllabic rhythm.

4.1.3. Summary of the results

As expected, listening to the repeated syllables caused entrainment to syllabic rhythm in auditory cortices. Changes in words percepts were associated with small phase shifts of the entrained oscillations. These results confirm that slow oscillation entrainment is not only reflecting pure stimulus-driven activity. However, the observed phase shifts corresponded to small shifts in the timing of the entrained oscillation (~8-10 ms). Therefore the results cannot completely validate our first working hypothesis: a temporal shift of 80-100 ms, corresponding to the distance between the 2 consonants' onsets would have been necessary to account for the perceptual shift.

An alternative hypothesis is that the theta phase –gamma amplitude coupling is not fixed. In this case, theta activity would still provide a temporal reference for sensory encoding, and the position of maximal gamma power according to the theta reference could reflect the perceived word (fig. 3.3). This alternative hypothesis is currently under investigation.

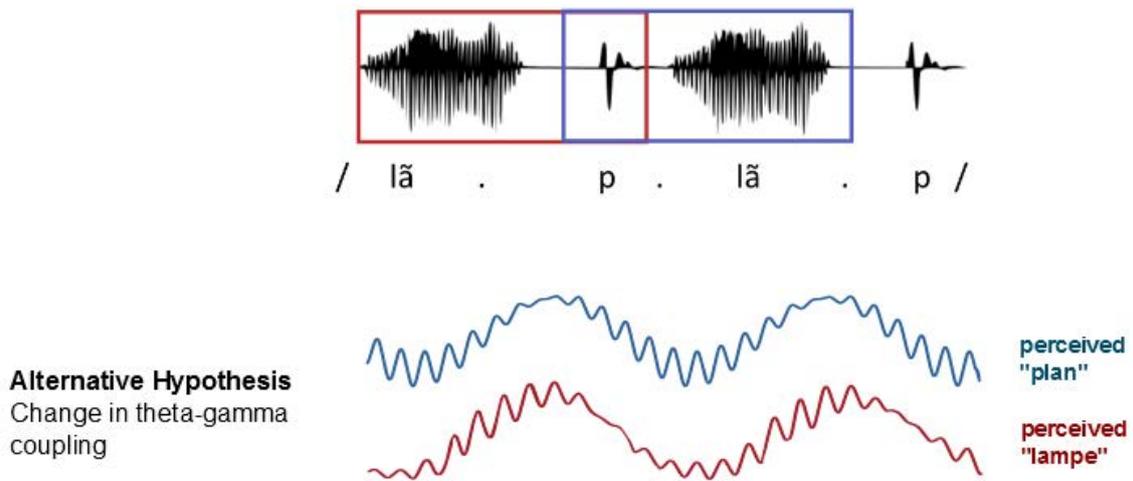


Figure 4.3: Alternative hypothesis. Changes in word perception could be explained from the shift of the peak in gamma power according to the theta temporal reference.

4.2. ARTICLE

Oscillatory mechanisms of ambiguous speech parsing (in preparation)

OSCILLATORY MECHANISMS OF AMBIGUOUS SPEECH PARSING

ABSTRACT

In this MEG study, we investigated the role of cortical oscillations in parsing linguistic units during speech listening. To do so, we used sequences of multistable acoustic signals, which could perceptually be parsed into two distinct speech or non-speech percepts. The changes in perceptual parsing were associated with changes in the preferential phase of neural oscillations in auditory cortex. Crucially, the peak frequency of the observed neural oscillations was indicative of the size of the speech unit: specifically, changes in monosyllabic words perception recruited oscillations at syllabic rate (here, 3Hz) whereas the sub-harmonics were implicated in bi-syllabic word parsing (here, 1.5 Hz). In line with previous reports, these results thus suggest that cortical oscillations reflect at once the tracking of acoustic features and the parsing of distinctive features necessary for speech processing. Surprisingly however, phase shifts of neural oscillations were too small to account for a true shift of the temporal parsing window. As such, the present findings question the role of auditory cortex oscillations as “parsers” of the speech signal.

KEYWORDS

Speech parsing, distinctive features, entrainment, theta, delta, verbal transformations, auditory, multistable perception, phase of neural oscillations, MEG

1. INTRODUCTION

The parsing of the continuous acoustic signal is required in speech to extract the discrete perceptual units or distinctive features used in the elaboration of more complex linguistic units (Poeppel et al., 2008). Speech signals are constituted of a hierarchy of linguistic features that are conveyed at different time scales, namely: words and sentences, evolve at a rate 0.5-3Hz and are composed of syllables produced at rate of 3-8Hz and syllables themselves comprised of phonemes occurring at a rate of 20-50 Hz (Poeppel, 2003). The temporal structure in speech inputs suggests the existence of a hierarchical organization of speech parsing mechanisms, with multiple processors sampling the signal at different time scales in parallel (Poeppel et al., 2008; Ghitza, 2011; Giraud and Poeppel, 2012). Consistent with this proposal, neural oscillations present temporal regularities that match the time scales of speech features (Poeppel, 2003; Ghitza, 2011; Giraud and Poeppel, 2012; Luo and Poeppel, 2012) and neural oscillations at different times scales are not strictly independent of each other: high frequency oscillations can be nested to the envelope of lower frequencies dynamics, suggesting a hierarchy in neural encoding mechanisms (Lakatos et al., 2005; Canolty et al., 2006; Canolty and Knight, 2010; Akam and Kullmann, 2014).

A recent body of work pointing to the role of neural oscillations in speech parsing (Schroeder et al., 2008; Ghitza, 2011; Giraud and Poeppel, 2012) focused on the coupling of neural oscillations in the theta (3-8 Hz) and gamma (30-50 Hz) range. In a recent model proposed by Giraud and Poeppel (2012), theta oscillations are shown to track the temporal regularities of speech in the syllabic range, while gamma oscillations encode phonemic information. Importantly for our study, gamma oscillations have been proposed to be modulated in power by the phase of theta oscillations, meaning that phonemic encoding is modulated in time at the syllabic rate. Hence, owing to the theta Phase-gamma Amplitude Coupling (PAC), the encoding capacity in auditory cortex is aligned to the relevant acoustic features of speech across different time scales (Giraud and Poeppel, 2012).

In this model, the theta tracking at the syllabic rate appears to be the first crucial step to parse speech into meaningful segments. In line with this proposal, theta entrainment to the temporal speech envelope (Poeppel, 2003; Luo and Poeppel, 2007; Howard and Poeppel, 2010) has been closely linked to speech intelligibility (Ahissar et al., 2001; Gross et al., 2013; Hertrich et al., 2013; Peelle et al., 2013; Doelling et al., 2014). Yet, the model makes the finer

prediction that theta imposes periods of shunting of the phonemic processing (as reflected by gamma oscillations), and as such defines syllabic parsing windows. Interestingly, there is evidence that PAC mechanisms have a fixed relationship in which a maximal high-frequency power occurs at the trough of the low-frequency oscillation (Lakatos et al., 2005). Hence, if the theta modulated gamma power defines temporal windows of syllabic encoding, then the minima of gamma power _ or *the peaks of theta oscillations* _ should correspond to perceptual syllabic junctions. Hence, the phase of theta oscillations should be an indicator of *when* the acoustic signal is parsed into syllables.

From the original model (Poeppel, 2003; Ghitza, 2011; Giraud and Poeppel, 2012), any change in the parsing of the same acoustic signal should lead to a different speech percept and hence correspond to a different phase in the theta band. To test this hypothesis, we departed from a perceptual phenomenon called verbal transformation. Verbal transformation consists in the experience a bistable change of percept when listening to a repetitive word stream without pausing (Warren, 1968; Pitt and Shoaf, 2002; Sato et al., 2006; Basirat et al., 2012). Verbal transformation is produced by a change in the timing of syllable segmentation and for instance, the repetition of the word “fly” will alternate with the perception of the word “life”. Hence, and crucially, verbal transformations show that for a given acoustic utterance, the speech signal could be chunked into different units, suggesting that parsing mechanisms may be endogenous and not purely acoustic-based. This perceptual phenomenon is further in line with the prediction that low-frequency oscillations are instrumental for sensory selection (Schroeder and Lakatos, 2009), and can be modulated by the subject’s attention and stimulus expectation (Lakatos et al., 2008; Schroeder and Lakatos, 2009; Stefanics et al., 2010; Besle et al., 2011; Gomez-Ramirez et al., 2011), in particular during speech listening (Zion Golumbic et al., 2013).

We thus tested whether any change in the perceived word during verbal transformations could be predicted by the phase of the entrained low-frequency oscillation in auditory cortex. Participants performed a verbal transformation task in which syllables were repeated at 3Hz in sequences of approximately one minute. In a first experimental block, we asked participants to spontaneously report the words they perceived during each sequence; in a second block (“volitional”), we asked participants to maintain the perception of one of the two possible percepts. Our preliminary results show small modulations in the theta phase with the perceived chunking of the signal, consistent with the hypothesis that theta entrainment reflects both the processing of low-level acoustical features and endogenous parsing. In addition, the

power of the first sub-harmonic in the delta range was noticeable in the perception of bi-syllabic words. Altogether, our preliminary results suggest that delta and theta oscillations characterize both bottom-up acoustic tracking and top-down speech parsing. By far however, the effects we observe are small and question the role of the slow oscillations as *parsers* per se; instead, we emit a complementary hypothesis in which slow oscillations have a role of temporal *taggers* of sensory information.

2. MATERIALS AND METHODS

2.1. Participants

15 participants (5 females, mean age: 23 years old) took part in the study. All were native French speakers who had normal hearing and were naive as to the purpose of the study. Prior to taking part in the study, each participant provided a written informed consent in accordance with the Declaration of Helsinki (2008) and the Ethics Committee on Human Research at NeuroSpin (Gif-sur-Yvette, France).

2.2. Experimental paradigm

2.2.1. Stimuli

Eight auditory sequences (adapted from the work of (Sato et al., 2007; Basirat et al., 2012)) were presented to participants via Etymotic Earphones (Etymotic Research Inc., USA). Four sequences consisted of the repetition of non-words that were either monosyllabic “pse” (/psə/) and “sep” (/səp/) (in one sequence /psə/ was repeated, in the other /səp/ was repeated), or bisyllabic “pata” (/pata/) and “tapa” (/tapa/) (one sequence repeating the pseudoword /pata/, the other /tapa/). Four additional sequences consisted in the repetition of French words that were either monosyllabic “plan” (/plā/, meaning “map”) and “lampe” (/lāp/, meaning “lampe”) or bi-syllabic “képi” (/kepi/) and “piqueur” (/pike/, meaning “to sting”) (two sequences of each defined by the repeated word). All syllables composing the auditory sequences were recorded

(16 bit resolution, 22.05-kHz sampling rate) in a soundproof room by a native French speaker (A.K.). The speaker pronounced each syllable naturally, and maintained an even intonation and vocal intensity while producing the sequences. For the bi-syllabic sequences, one syllable of each token /pa/, /ta/, /ke/, and /pi/ was selected; the selection criterion consisted in the syllable that matched as closely as possible to the sequence rate of 3Hz (1 syllable / 333 ms). Stimuli sequences were then constructed using the Praat freeware (Boersma, 2002). For the mono-syllabic word and non-word sequences, one clearly articulated token /psə/, /səp/, /plā/ and /lāp/ of 333-ms duration was selected from the record and repeated 150 times in Praat. For the bi-syllabic word and non-word sequences, pairs of syllables were assembled and repeated 100 times. For all recordings the syllabic length was of 333 ms, leading to a repetition rate of 1.5 Hz in bi-syllabic sequences and 3 Hz in mono-syllabic sequences. All syllables had equalized sound-levels.

2.2.2. Procedure

The experiment was divided into 2 main experimental blocks. In the first MEG block (“spontaneous”, fig. 1a), the experiment started with a verbal transformation task (Basirat et al., 2012), in which participants reported *spontaneously* their perception while listening to the auditory sequence. Participants were told that they would hear a sequence consisted of repeated acoustic utterances and that their perception of the auditory utterance could change in the course of the sequence. During the sequence, participants were asked to keep pressing on a button that corresponded to the currently perceived utterance, and to switch button as soon as their perception had changed. Three button responses were assigned during the experience: two corresponded to the main reported speech utterances of each sequence (/kepi/ vs. /pike/; /pata/ vs. /tapa/; /psə/ vs. /səp/; /plā/ vs. /lāp/) and a third one was given to participants to report any other percept. All auditory sequences were presented in a block design according to the type of perceptual outcome (/kepi/ vs. /pike/; /pata/ vs. /tapa/; /psə/ vs. /səp/; /plā/ vs. /lāp/). The mono-syllabic word and non-word sequences were repeated twice (300 repetitions each); the bi-syllabic word and non-word sequences were repeated 3 times (200 repetition each). Blocks were presented in random order across individuals.

In the second experimental block (“volitional”, fig. 1b), participants were asked to maintain a given percept a long as possible while listening to a sequence. Four sequences were chosen: /kepi/, /pata/, /lāp/, and /səp/. Subjects listened twice to each sequence, and asked to maintain

either one or the other of the main auditory utterances. Subjects reported that they succeeded in maintaining the required percept by pressing the button ‘maintain’; if they failed to maintain the required percept, they were asked to report it by pressing the button corresponding to the other auditory utterance or the button ‘other’ would they perceive a third token.

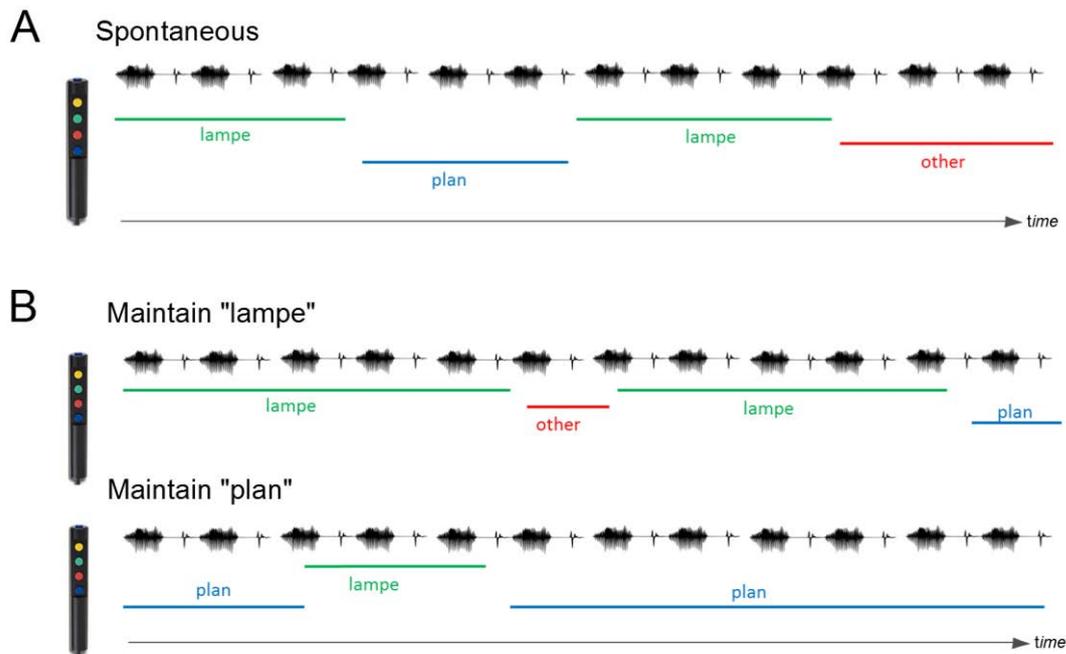


Figure 1: Example of behavioral response during (A) the spontaneous task and (B) the volitional task. While listening to the speech sequence, participants had to keep the button that was related to their perception pressed. Two buttons were associated with the main perceived speech utterances in the signal (for instance “plan” and “lampe”), the last button “other” was pressed if participants heard a different speech utterance. (A) During the spontaneous task, participants provided spontaneous perceptual changes. (B) The volitional task, participants heard twice the same sequence. In the first sequence they had to keep in mind as long as possible one of the main speech utterances (e.g. “lampe”). In the second sequence they had to maintain the other speech utterance (e.g. “plan”). If they failed to maintain the required percept, participants had to indicate it by pressing another response button. Only trials which concerned the successful maintenance of the required were kept for MEG analysis

2.3. Psychophysical data analysis

Subjects’ perception of bistable auditory sequences were evaluated during the first MEG block, i.e. the “spontaneous” verbal transformation task, and their ability to voluntarily control speech perception was assessed during the second block, i.e. the “volitional” task.

Two measures assessed perceptual bistability: (1) the proportion of time spent by participants in perceiving each utterance during the sequence, and (2) the durations of the alternating percepts. The first measure captured the average dominance of the perceived utterance when listening to one sequence and the other quantified the stability of the speech percepts, i.e. how often subjects' perception changed over the sequence.

For each duet of speech sequences (“kepi/piquer”, “pata/tapa”, “plan/lampe”, “pse/sep”), a repeated measures ANOVA was performed to assess the influence of the type of word being repeated on the proportion of perceived words. Specifically, the factors of interest were Sequence Order (2 levels: “repeating word1” / “repeating word2”), Percept (3 levels: “percept1”, “percept2”, “other”), and Subjects as random factor. Following a significant main effect, a Tukey-Kramer multiple comparisons procedure ($\alpha = 0.05$) was performed to assess significant differences between the different levels of the factor.

The histograms of durations of alternative bistable percepts follow a log-normal distribution in agreement with previous research (Pressnitzer and Hupé, 2006). To assess the influence of speech sequences on perceptual stability, repeated measure ANOVA were performed on log durations. The factors of interest were Speech Sequence (4 levels: “kepi/piquer”, “pata/tapa”, “plan/lampe”, “pse/sep”), Sequence Order (2 levels: “repeating word1” / “repeating word2”), and Sequence Type (2 levels: “word”, and “pseudo-word”, Speech Sequence nested in Sequence Type).

2.4. MEG analysis

2.4.1. MEG data acquisition

Brain magnetic fields were collected in a magnetically shielded room using the whole-head Elekta Neuromag Vector View 306 MEG system (Neuromag Elekta LTD, Helsinki) equipped with 102 triple-sensor elements (two orthogonal planar gradiometers and one magnetometer per location). Participants were seated in upright position. Participants' head position was measured before each block with four head position coils (HPI) placed over frontal and mastoid areas. MEG recordings were sampled at 1 kHz, band-pass filtered between 0.03 Hz and 330 Hz and used Maxshield. The electro-oculograms (EOG, horizontal and vertical eye movements) and electrocardiogram (ECG) were simultaneous recorded with MEG.

2.4.2. MEG data preprocessing

Signal Space Separation (SSS) method was applied to decrease the impact of external noise (Taulu et al., 2003). SSS correction, head movement compensation, and bad channel rejection was done using MaxFilter Software (Elekta Neuromag). Signal-space projection (SSP) were computed by principal component analysis (PCA) using Graph software (Elekta Neuromag) to correct for eye-blinks and cardiac artifacts (Uusitalo and Ilmoniemi, 1997). A rejection criterion for epochs was applied for gradiometers with amplitude exceeding $4000 \text{ e}^{-13}\text{T/m}$.

2.4.3. Evoked response analysis

MEG analysis was restricted to gradiometers channels. Data were segmented in epochs of 1.5s centered at the onset of the plosive burst in the speech signal. Epochs were filtered between 1Hz and 40 Hz. Epochs were averaged across left and right temporal gradiometer channels (see Fig. S1, black dots). SSP was estimated on evoked data covariance. The resulting spatial filter was used for channel averaging. This procedure was done to weight the sensors and privilege those which were strongly modulated by the evoked component of the signal, and to alleviate sensor cancellation due to opposite signal polarities.

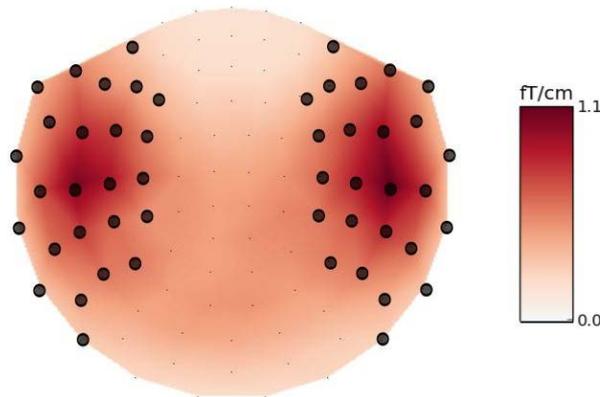


Fig S1: Topography of Steady State Responses (3 Hz SSRs) and channel selection. Grand-averaged topography of SSRs (norms of gradiometers). Black dots represent the sensors of interest delimiting left and right hemispheres. These regions of interest were considered for further analysis.

The comparisons of evoked responses between conditions were computed using a non-parametric permutation test. Correction for multiple comparisons was performed with cluster level statistics using as base statistic Student t-test computed at each time sample (Maris and Oostenveld, 2007). Only temporal clusters with corrected p-value ≤ 0.05 are reported.

2.4.4. Frequency analysis

The MEG signals were divided into epochs of 8.2 s in order to compute the Power Spectrum Density (PSD) for each sequence condition in each hemisphere. To extract the frequency peaks of neural entrainment to the acoustic utterances, the $1/f$ component was removed by subtracting at each frequency bin the mean power of the neighboring frequency values (4 frequency values were: $[f_0 - 0.24\text{Hz} ; f_0 - 0.12\text{ Hz}; f_0 + 0.12\text{ Hz}; f_0 + 0.24\text{ Hz}]$ (Nozaradan et al., 2011). The $1/f$ subtraction procedure relies on the assumption that the power at a specific frequency bin should be similar to the power of its surrounding frequencies.

After subtraction, one-sample t-tests were performed at the entrainment frequency (3Hz), its sub-harmonic (1.5 Hz) and each of its harmonics (4.5, 6, 7.5 and 9 Hz) to examine which frequency significantly entrained to the acoustic utterance – i.e. had a power significantly higher than zero. Repeated-measures ANOVA were then performed to assess the contributions of each target frequency according to hemisphere, speech sequence and perceived word.

2.4.5. Phase analysis

The phase variations of the 3 Hz neural entrainment and its sub-harmonic (1.5 Hz) were tested. To do so, single trial data were convolved with a 3-cycle Morlet at 1.5 Hz frequency for the computation of 1.5 Hz preferential phase. Convolution between single trial data and a 4-cycles Morlet at 3 Hz frequency was performed to compute 3 Hz preferential phase.

The Phase-Locking Value (PLV) (Lachaux et al., 1999) was computed as:

$$PLV(t) = \frac{1}{K} \left| \sum_{k=1}^K e^{j\theta(t,k)} \right|$$

where K is the number of trials, and $\theta(t,k)$ is the instantaneous phase at time t and trial k . PLVs were computed to assess intra-subject variability in the preferential phase. In all sequences, PLV and instantaneous phase are computed at the onset of the plosive burst.

To assess statistical significance of phase shifts between the 2 verbal transformation percepts during one speech sequence (e.g. “plan” and “lampe”), the difference in the preferential phase of entrainment was computed on a per individual basis. Bootstrap measures of 95% confidence intervals were used on the phase distribution of the differences (Fisher, 1995). Phase distributions were statistically different if the mean of the difference was statistically different from zero, i.e. if zero lied outside the measured confidence interval ($p \leq 0.05$).

3. RESULTS

3.1. Psychophysics

As predicted, participants perceived verbal transformations while listening to each speech sequence during the spontaneous task. Participants mainly perceived the two proposed bistable speech percepts, with a few “other” responses (Fig. 2b, Table S1). Participants dominantly perceived the two proposed speech utterances in all conditions (main effect of Percept: “kepi” / “piquer” sequences, $F[2,28] = 5.8$; $p = 0.008$. “pata” / “tapa”, $F[2,28] = 10.1$, $p < 0.001$. “plan” / “lampe”, $F[2,28] = 14.5$; $p < 0.001$. “pse” / “sep”, $F[2,28] = 17.2$, $p < 0.001$). Post-hoc analyses revealed that the proportion of the two main percepts was significantly greater than the “other” percepts. Additionally, participants more often reported the percepts “piquer” in the “kepi” / “piquer” sequences (Fig. 1b, top panel) and the “sep” in the “pse” / “sep” sequences (Fig. 2b, bottom panel). The dominance of “piquer” and “sep” percepts was surprising considering that previous reports showed biases in the other direction, namely “pse” was more dominant in “sep” / “pse” verbal transformation tasks (Sato et al., 2007). Also contrary to previous reports (Sato et al., 2007; Basirat et al., 2012), the word being repeated in the sequence did not affect the proportions of perceived words, except for “pse” / “sep” sequences (Fig. 2a, $F[1,14] = 2.76$, $p = 0.02$). Even for this condition, the mono-syllabic

sequences elicited more verbal transformations than previously reported (Sato et al., 2007; Basirat et al., 2012).

Overall, the reports in the spontaneous block confirmed that all speech sequences successfully elicited verbal transformations. It should be noted that, as expected, the main verbal transformations were related to changes in the parsing of the acoustic signal at syllabic or bi-syllabic level.

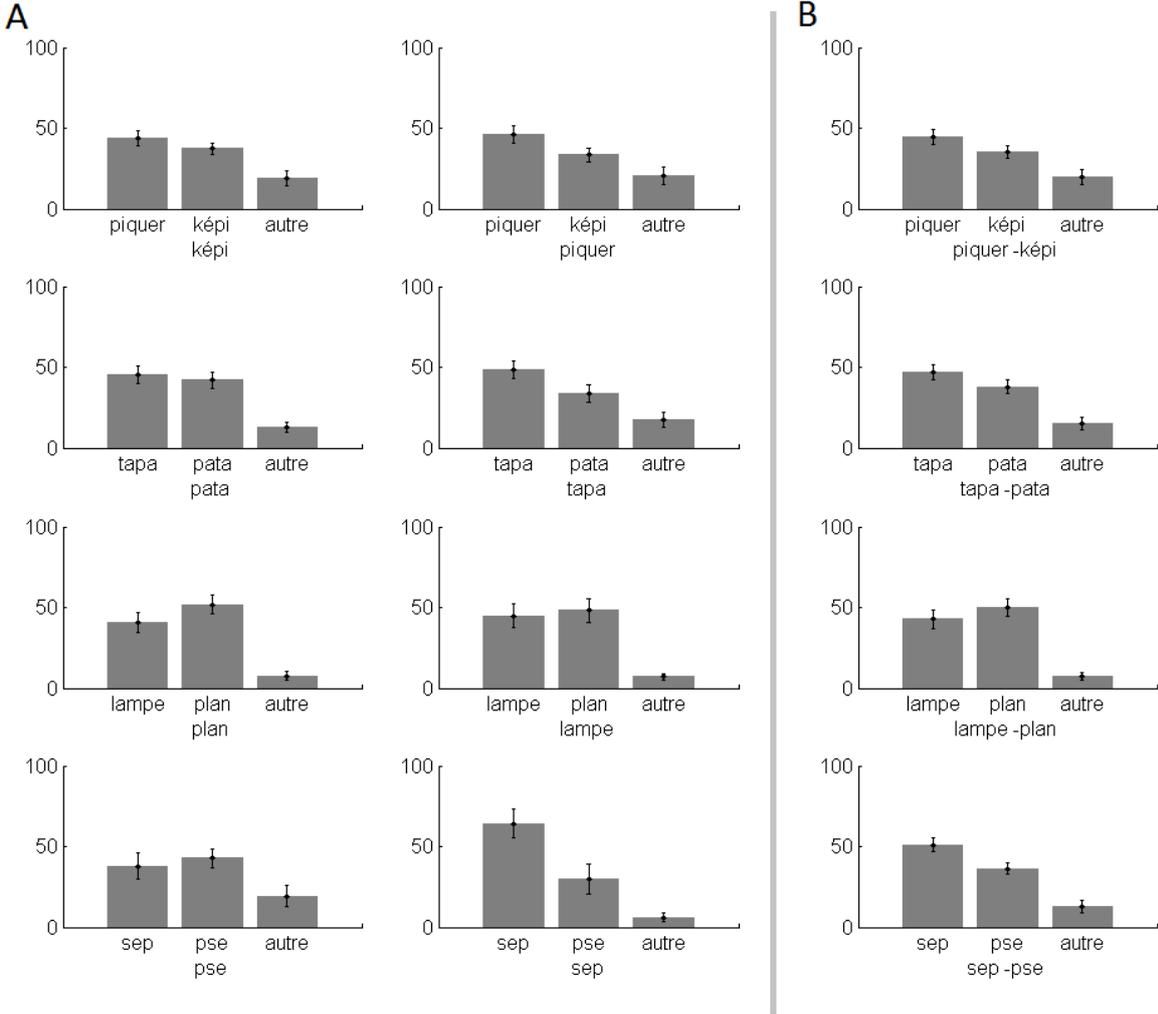


Figure 2: Proportion of responses in the spontaneous verbal transformation block. (A) Responses according to the repeated word in the sequence. (B) Grand average responses across sequences. Error bars denote s.e.m.

	Phonetic modifications	Lexical/Semantic modifications	Segmentation modifications	Streaming
Képi-Piquier	. /keli/ . /ike/ . /epi/ . /pie/	. Mickey . Gouter . t'es où		
Pata-Tapa	. /tafa/ . /ata/ . /pia/	. lapin . tape	. /terata/ . /taota/ . /pa/-/ta/	. /p/-/t/ /a/-/a/
Plan-Lampe	. /ã/	. nan . lent		. /pl/ -/ã/
Pse-Sep	. /pia/	. tout ça		. /p/-/sə/ . /ə/-/ə/

Table S1: List of “other” percepts. Words accounting for less than 15% of the total percepts were spontaneously reported by participants during debriefing.

We then assessed the mean duration of percepts which differed across auditory sequences. Figure 3 reports the distributions of durations for each sequences “kepi”-“piquer” (A), “pata”-“tapa” (B), “plan/lampe” (C) and “pse”-“sep” (D). Distributions were highly asymmetric and followed a log normal distribution (Figure 3e-h, Kolmogorov-Smirnov test for normality of each distribution of log(duration): $p > 0.05$). Surprisingly, the durations of the word percepts were shorter than for pseudo-words ($F[1,14]=9.2$; $p=0.009$). No difference were perceived between “kepi” and “plan” sequences, nor between “pse” and “pata” sequences ($F[2,28] = 1.1$, $p = 0.35$). This suggests that participants switched more their perception when listening to ambiguous word sequences than pseudo-words sequences. At first glance, these results seem to contradict previous reports (Pitt and Shoaf, 2002; Shoaf and Pitt, 2002), showing that words are less sensitive to verbal transformations than non-words. In these experiments, verbal transformations competed in a speech sequence could either elicit the perception of a real word or a non-word. Participants perceived in proportion more real words than non-words. These findings indicated that lexical processing influenced the verbal transformation task, and suggested that lexical representation could serve the stabilization of speech perception (Shoaf and Pitt, 2002). However, in our experiment, the two utterances in competition were either both words or both non-words. We thus argue that our results are in line with the previous verbal transformation reports, and that the shortening of percept duration in word sequences could reflect lexical competition. In agreement with this view, reports showed that speech word parsing heavily relied on lexical effects (Elman and

McClelland, 1988; Vroomen et al., 1996; Billig et al., 2013), suggesting that speech processing is inherently predictive (Gagnepain et al., 2012).

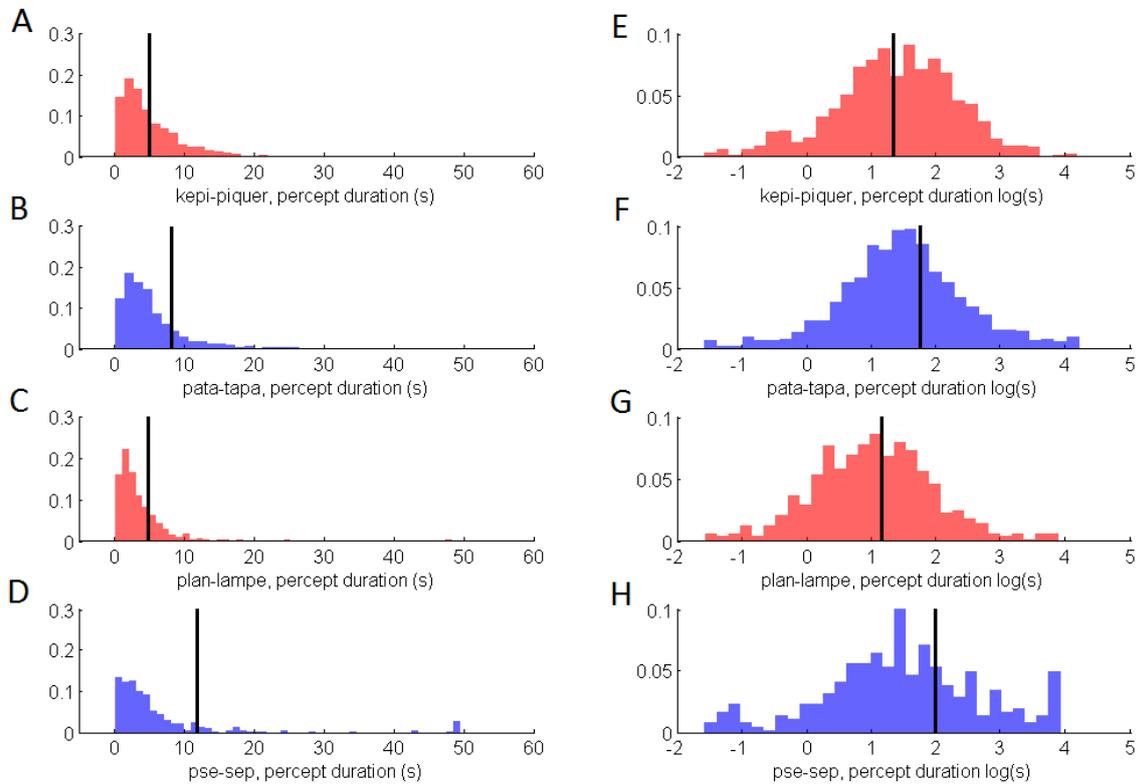


Figure 3: Normalized histograms of percept durations in the spontaneous verbal transformation task.

Histograms on the left represent percept durations on a linear scale for (A) “kepi” / “piquer” sequences, (B) “pata” / “tapa” sequences, (C) “plan” / “lampe” sequences, (D) “pse” / “sep” sequences. Histograms on right represent the same data on a logarithmic scale. The vertical bars represent the averaged median of the distributions across subjects. Word percepts (red) were less stable over time than pseudo-words (blue) percepts.

During the maintenance task, participants were successful in maintaining the required stimulus in all conditions ($F[2,28] = 32.0, p < 0.001$) (fig. 4). However; multiple comparison procedure showed that the percept “pse” was significantly harder to maintain than the other percepts (59% maintenance, against 85% to 96% in the other conditions); nevertheless, this percept was still dominant compared to the alternative percept “sep” (37 %). The duration of a given percept was also significantly longer when participants were asked to maintain than when they were asked to ignore it (fig. 5). As the ignored distribution of residuals did not follow a normal distribution (significant Kolmogorov-Smirnov test) a Kruskal-Wallis test was performed: $\text{Chi}^2[1,14] = 16.4, p < 0.001$.

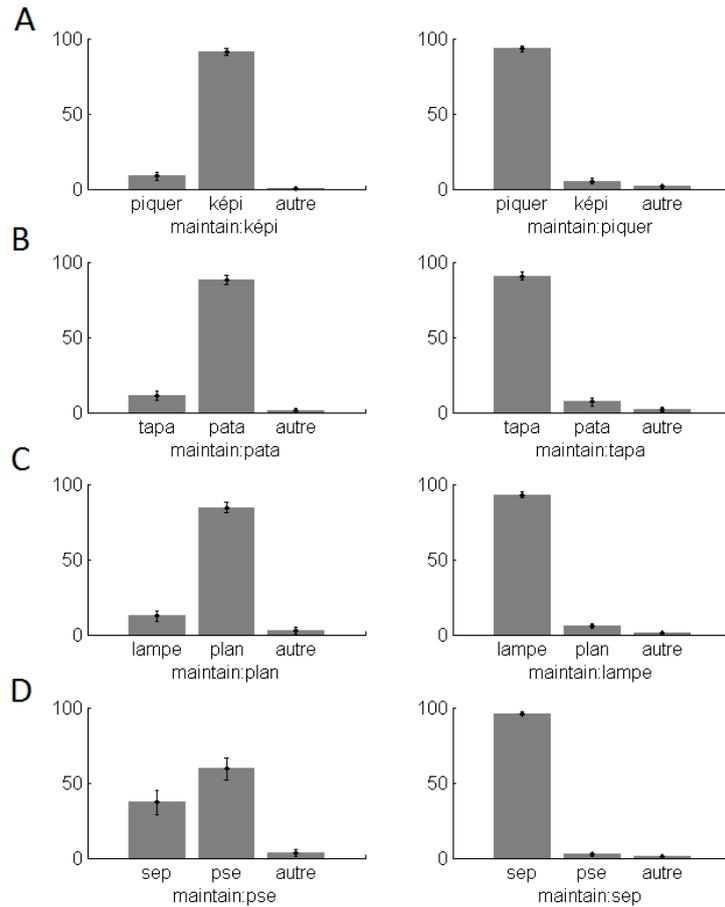


Figure 4: Proportion of perceived words during the volition task. Errors bars denote s.e.m. Each sequence (A) “képi”, (B) “pata”, (C) “lampe”, (D) “sep” was repeated twice. Subjects were asked to maintain the perception of one bistable word in one sequence (left panels), and to maintain the other bistable word in the second sequence (right panels). On average subjects performed well the task.

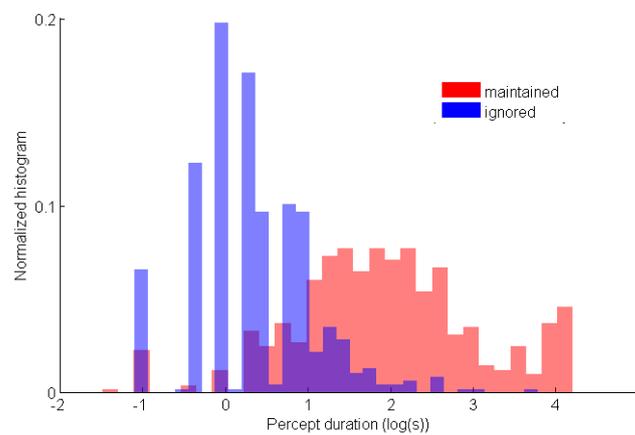


Figure 5: Percept duration during the volition task. Duration distributions of percepts if maintained (red) or ignored (blue) are displayed in log scale. Maintained percepts were more stable than ignored percepts.

Importantly, percepts were more stable in the “volitional” than in the “spontaneous” task: the average percept duration during spontaneous reports was 4.5 s and 8.6 s in the volition task.

In light of this successful experimental manipulation, we decided to test the main hypothesis on the MEG data recorded during the volitional task.

3.2. Different frequency profiles according to speech sequence

Oscillations in auditory cortex are known to track the temporal regularities of external inputs (Regan, 1966; Hari, 1989; Besle et al., 2011) and hypothesized to play a crucial role in parsing when listening to speech (Ghitza, 2011; Giraud and Poeppel, 2012). Hence, neural oscillations in auditory cortex should reflect how the acoustic signal is being chunked into perceptual units. Here, we specifically hypothesized that verbal transformations were caused by changes in auditory speech tracking mechanisms.

To test this, we first checked whether the presentation of different speech sequences elicited neural entrainment in auditory cortices. Power Spectra Density (PSD) analysis was performed in the left and right hemispheres (see Methods). Detrended power analysis was performed to remove $1/f$ endogenous neural component (Nozaradan et al., 2011) allowing to assess whether MEG signals followed the 3 Hz rhythmic input, and whether its harmonics were present (see Methods). Significant steady-states component are reported in Figure 6. As expected, auditory cortices showed a significant 3 Hz rhythm in all conditions as well as its first harmonic at 6 Hz. Interestingly, listening to bisyllabic words induced a 1.5 Hz frequency response in auditory cortex. The 1.5 Hz peak was clearly visible in MEG data frequency spectrum during “kepi” sequence listening (Fig. 6a) as well during “pata” sequences in right auditory cortex (Fig. 6b). Crucially, this frequency component was not present when listening to the repetition of mono-syllabic words (Fig. 6 c-d).

As bi-syllabic sequences consisted in the repetition of acoustic patterns at 1.5 Hz, the acoustic signal may have already contained a frequency peak at 1.5 Hz to which auditory cortex would entrain; as such, the 1.5 Hz observed in the PSD may just reflect a bottom-up signal frequency tagging. To test this, PSD was computed on the envelope of the acoustic sequences (Fig. S2): while the “kepi” auditory sequence PSD showed a peak at 1.5 Hz, no peak was detectable in the “pata” auditory sequence PSD. Hence, the 1.5 Hz component that is present in MEG

signal when listening to the “pata” sequence cannot reflect pure acoustic signal tracking, and may be a consequence of endogenous perceptual parsing: the 1.5 Hz oscillation could have been recruited due to the temporal contingencies of the task (the relevant parsing window occurs at a rate of 1.5 Hz for bi-syllabic sequences).

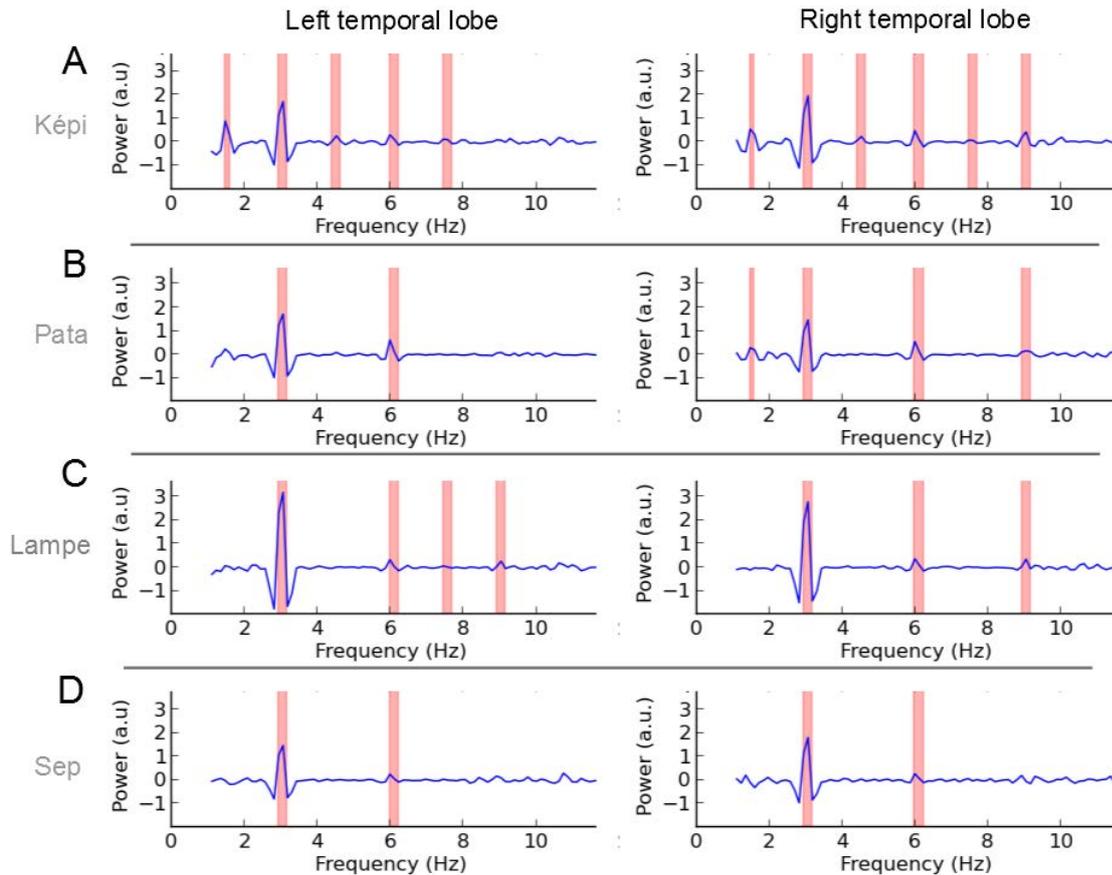


Figure 6: Detrended frequency power spectra. (A) Listening to the repetition of “kepi” words entrained auditory cortices at 1.5 Hz and direct harmonics: 3 Hz, 4.5 Hz, 6 Hz, 7.5 Hz and 9 Hz (B) “pata” sequences elicited entrainment at 3 Hz and 6 Hz in the left hemispheric sensors, and at 1.5 Hz, 3 Hz, 6 Hz and 9 Hz in the right hemisphere sensors. (C) “lampe” sequences elicited neural entrainment at 3 Hz, 6 Hz, 7.5 Hz and 9 Hz in left temporal lobe, and at 3 Hz, 6 Hz and 9 Hz in right temporal lobe. (D) “sep” sequences elicited entrainment at 3 Hz and 6 Hz.

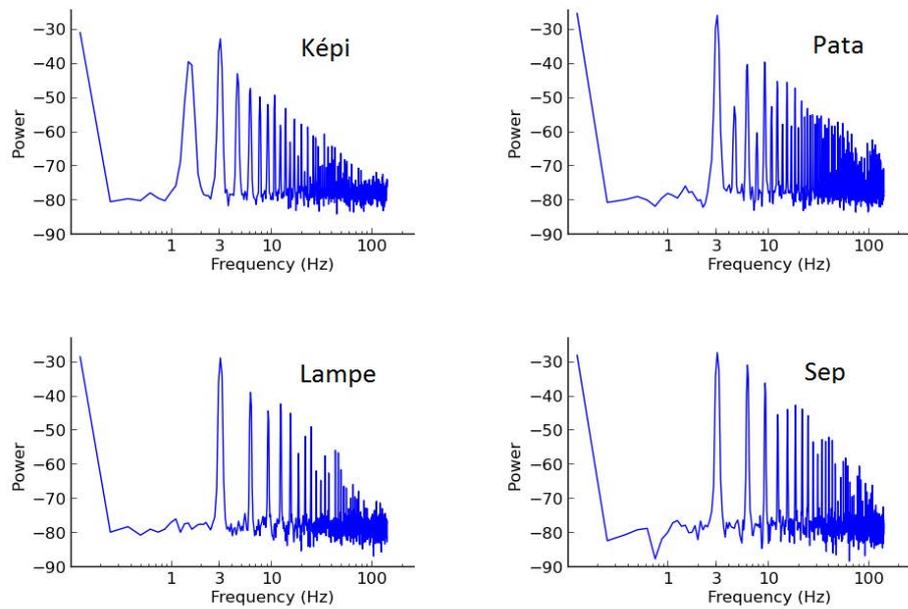


Figure S2: Frequency power spectra of the acoustic stimuli. The “kepi” sequence is the only acoustic signal that presents a peak at 1.5 Hz in its frequency power spectrum.

In the next analysis, we tested whether these steady-state components were modulated according to the perceived word and if there were hemispheric asymmetries. In accordance with the detrend fluctuation analysis, 1.5 Hz power was higher in bi- than in mono-syllabic sequences ($F[3,42] = 5.3$, $p = 0.004$; post-hoc analysis: significant difference between [kepi-pata] and [lampe-sep] sequences). 1.5 Hz power was however not indicative of perceived word ($F[1,14] < 1$), and did not differ between hemispheres ($F[1,14] < 1$).

The 3 Hz component was also modulated by the speech sequence ($F[3,42] = 4.7$, $p = 0.006$). Post-hoc analysis revealed that the peak in “lampe” sequences were higher than in the other speech sequences conditions. 3 Hz power was not modulated by percept ($F[1,14] < 1$). Again, no hemispheric bias in 3 Hz power was detected ($F[1,14] < 1$). The absence of asymmetry is surprising considering the previous reports of a rightward bias of slow oscillation neural tracking (Boemio et al., 2005; Giraud et al., 2007).

The 6 Hz component was also modulated by speech sequence ($F[3,42] = 3.28$, $p = 0.03$). The 6 Hz peak represents a harmonic of the rhythm of entrainment (3 Hz) and is present for all speech signals, post-hoc test suggest however that 6 Hz power is higher for bi-syllabic

sequences than mono-syllabic sequences. 6Hz power changes were not observed depending on percept ($F[1,14]<1$) with no hemispheric bias ($F[1,14]<1$).

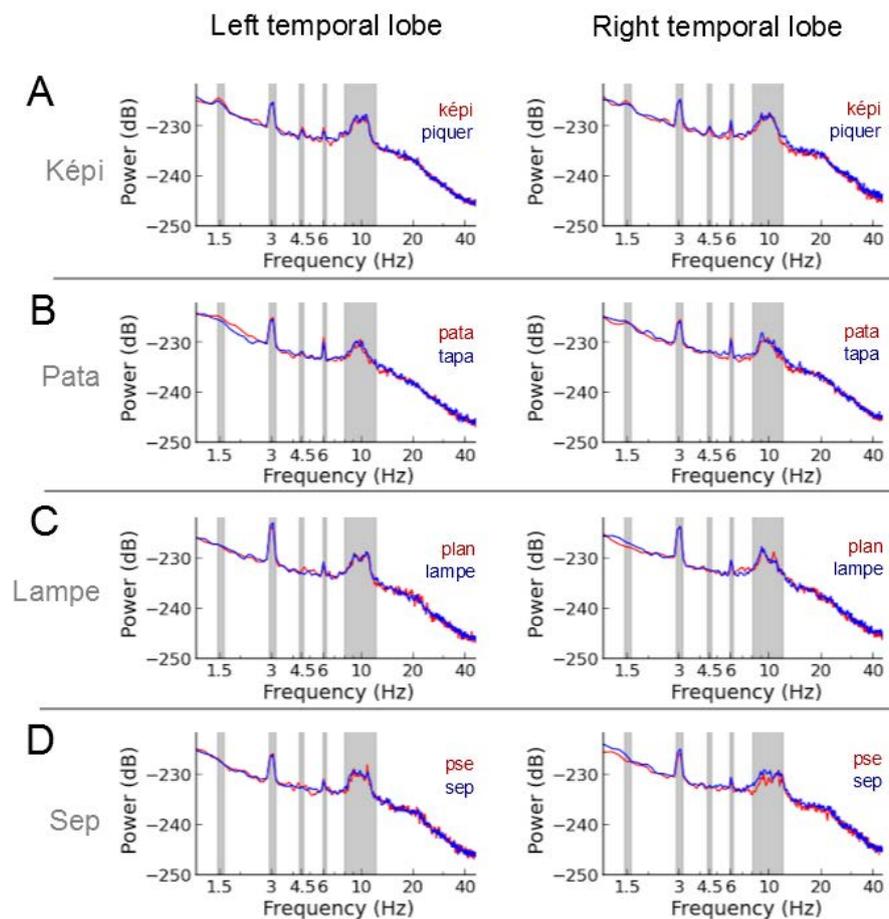


Figure 7: Frequency power spectra across conditions. Red and blue lines correspond to the frequency spectra of the main speech utterances in (A) “kepi” (B) “pata” (C) “lampe” and (D) “plan” sequences. The grey areas highlight the main peaks in frequency power, e.g. 1.5 Hz, 3 Hz, 4.5 Hz, 6 Hz, and alpha (8 Hz-12 Hz).

In addition to steady- state components a prominent peak was observe at alpha frequency (8-12 Hz) (fig. 7). Alpha power was no modulated by speech sequence ($F[3,42] = 2.1, p = 0.12$), and no difference were observed between hemispheres ($F[1,14]<1$). Results however show a modulation of alpha power according to percept ($F[1,14]=9.4, p=0.008$). More alpha power was observed for percepts “piquer” than “kepi”, “tapa” than “pata”, “lampe” than “plan” and “sep” than “sep”. Crucially, for the bi-syllabic sequences, the percepts associated with higher alpha power correspond to the words that were perceived in higher proportion (Fig. 2a-b). For

mono-syllabic sequences, it corresponds to the repeated word in the sequence, i.e. the easiest word to maintain. Hence, we interpret the change of alpha power as an attentional effect in the volitional task: more attention is required to perceive the “less” easy word, which entails a global decrease in alpha power.

3.3. Verbal transformations elicited phase shifts in 1.5 Hz and 3 Hz oscillations

Here, we examined the hypothesis that auditory cortex oscillations provide a mechanistic means for the brain to parse the acoustic signal into relevant perceptual units (Ghitza, 2011; Giraud and Poeppel, 2012). According to this view, each cycle of the oscillation should correspond to the parsing of one linguistic unit. Hence, any temporal shift in the oscillatory cycle according to the acoustic signal should lead to a change in the perceived speech unit. In our task, speech utterances were either repeated at 1.5 Hz or 3 Hz frequencies; we thus hypothesized that the reported verbal transformations would be elicited by phase shifts of neural oscillations at 1.5 Hz and 3 Hz, respectively.

The 3 Hz preferential phase was computed for each condition at the onset of the plosive burst. As expected, listening to the rhythmic sequences elicited strong 3 Hz phase locking (table 1 and 2). Interestingly in the left temporal lobe, Phase-Locking Values (PLVs) tended to be higher when the perceived word in the utterance started with a plosive, but this trend was only significant when contrasting “pse”/ “sep” sequences (table 1). No PLV difference was observed between the different perceived in the right hemisphere.

Perceived word	PLV	Perceived word	PLV	Difference (t, pval)
Lampe	0.62	Plan	0.65	t = 2.1, p = 0.055
Sep	0.55	Pse	0.62	t = 2.2, p = 0.03
Képi	0.49	Piquer	0.46	t = 1.1, p = 0.3
Tapa	0.50	Pata	0.55	t = 1.9, p = 0.08

Table 1: Phase Locking Values (PLVs) at 3 Hz for left hemisphere sensors.

Perceived word	PLV	Perceived word	PLV	Difference (t,pval)
Lampe	0.62	Plan	0.63	t = 0.4, p = 0.69
Sep	0.57	Pse	0.56	t = -0.05; p = 0.95
Képi	0.46	Piquer	0.45	t = -0.37, p = 0.72
Tapa	0.49	Pata	0.55	t = 2.1; p = 0.055

Table 2: Phase Locking Values (PLVs) at 3 Hz for right hemisphere sensors.

Crucially, the 3 Hz preferential phase was significantly modulated according to the perceived word: significant phase differences of -8° and -9° distinguishing the perceived percepts were obtained for mono-syllabic sequences in the right temporal hemisphere sensors (Fig. 8c-d). These phase differences signified that, compared to the default state in which “lampe” was perceived, perceiving “plan” was associated with a shift of 8° of the 3 Hz component. This phase shift corresponds to a temporal delay of 7 ms of the 3 Hz oscillation and is consistent with a latency shift observed in the evoked steady-state response (Fig. 8c-d). However, no significant effects were seen in the ERF analysis suggesting that oscillatory phase analysis may provide a more sensitive measure of neural latency than classic evoked response analysis.

Additionally, significant differences in the phase of the 3 Hz response were observed between perceiving “pata” and perceiving “tapa”: for this speech sequence, the difference in 3 Hz phase was accompanied with significant changes in the ERF amplitude (Fig 8b). No significant changes were observed between the different percepts of the “kepi” sequence (fig. 1a).

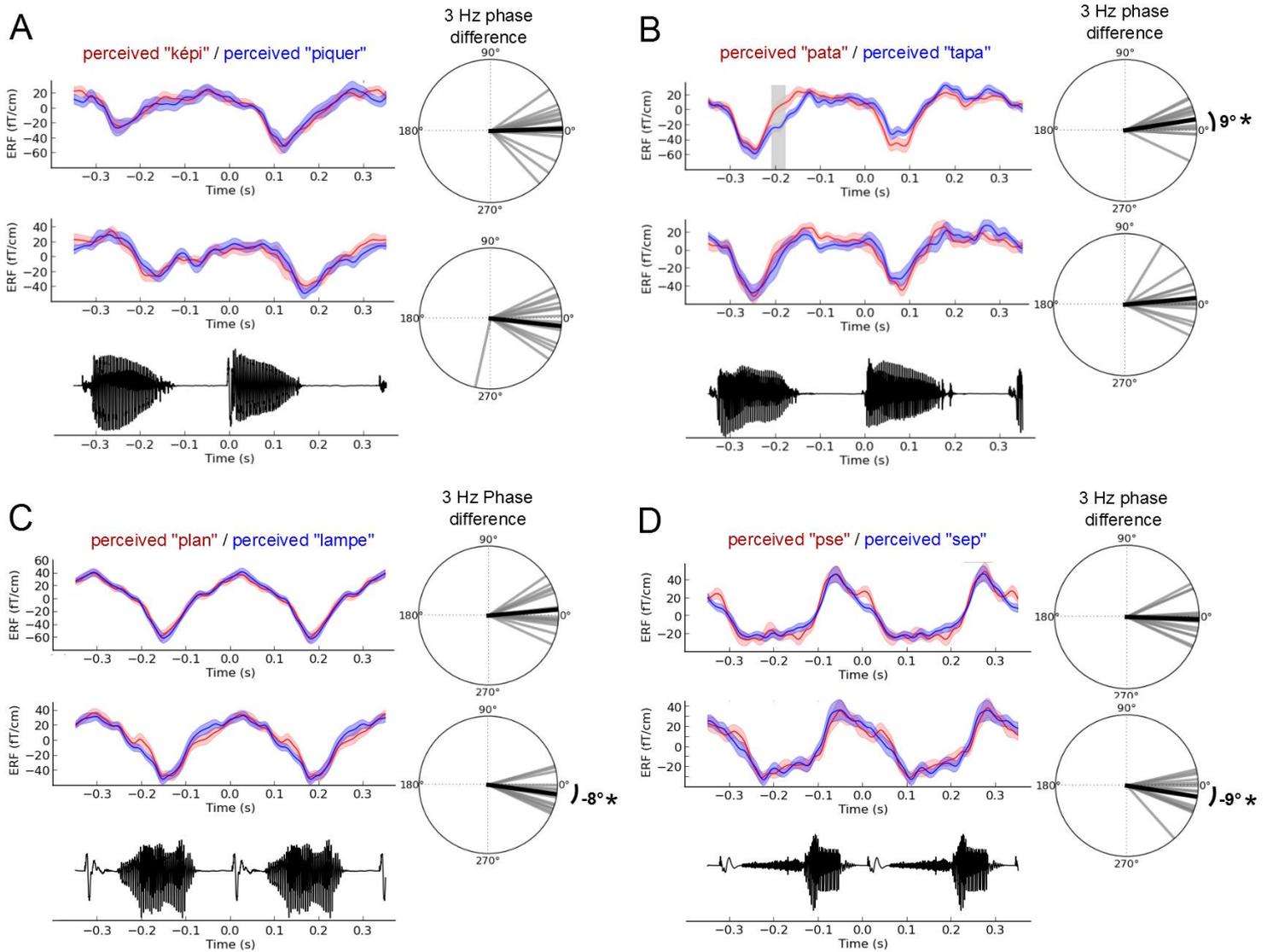


Fig. 8: Event-Related Fields (ERFs) and 3 Hz phase shifts in (A) “kepi” (B) “pata” (C) “lampe” and “sep” sequences. ERFs and phase distributions in top panels correspond to left temporal lobe, middle panels correspond to right temporal lobe. Bottom panels depict the speech signal for each sequence condition. Polar plots denote the 3 Hz phase difference between the 2 percept conditions: each grey bar denotes the phase difference between the two percepts conditions for one subject, the black bar corresponds to the average phase shift across subjects.

Similarly, the perceptual changes in the “kepi” and “pata” sequences were expected to be accompanied by modulations of the 1.5 Hz oscillatory component. Significant changes in the PLV and in the preferential phases were observed in the left hemisphere sensors between the perceived words in “kepi” sequences (Table 3, Fig. 9a). No differences were observed in the right temporal sensors, and in the non-word sequence “pata” (Table 3, 4, Fig. 9)

Perceived word	PLV	Perceived word	PLV	Difference (t, pval)
Képi	0.34	Piquer	0.27	t = 2.2, p = 0.04
Tapa	0.25	Pata	0.25	t = 0.02, p = 0.98

Table 3: Phase Locking Values (PLVs) at 1.5 Hz for left temporal sensors.

Perceived word	PLV	Perceived word	PLV	Difference (t, pval)
Képi	0.33	Piquer	0.24	t = 1.9, p = 0.08
Tapa	0.21	Pata	0.21	t = -0.03, p = 0.97

Table 4: Phase Locking Values (PLVs) at 1.5 Hz for right temporal sensors.

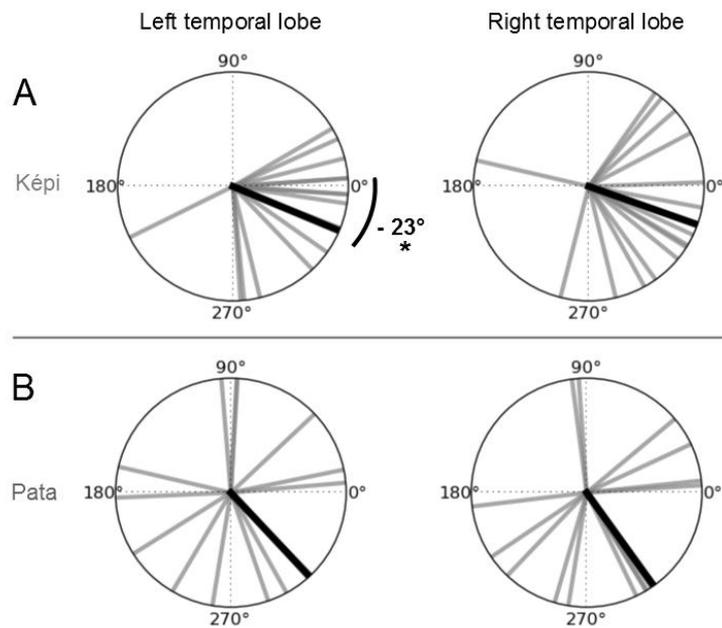


Figure 9: Differences in preferential phase between percept conditions at 1.5 Hz frequency. Each grey bar denotes the phase difference between the two percepts conditions for one subject; the black bar corresponds to the average phase shift across subjects. (A) During the listening of “kepi” sequences, change in perceived word was associated with significant phase shift in 1.5 Hz oscillations in left temporal lobe (mean phase shift of -23° (-63ms), CI = $[-52^\circ, -2^\circ]$). Phase shifts of the same amplitude were observed in right temporal lobe, but were not significant (mean phase shift of -20° , CI = $[-51, 15^\circ]$). (B) No constant 1.5 Hz phase shifts were observed during “pata” sequences listening (mean phase shift in left temporal lobe: -47° (-131 ms), CI = $[-151^\circ, 158^\circ]$, mean phase shift in right temporal lobe: -54° (-151 ms) $[-123^\circ, 44^\circ]$).

4. DISCUSSION

4.1. Modulation of slow neural oscillations reflect endogenous parsing

Using a verbal transformation task, we tested the hypothesis that cortical oscillations in auditory cortices reflect an endogenous speech parsing mechanism constrained by internal linguistic representational units (Poeppel, 2003; Poeppel et al., 2008). More precisely, we asked whether changes in cortical oscillations could inform us on *when* the speech signal was parsed into such representational units. A growing body of evidence shows that neural oscillations can track the temporal regularities of auditory speech signals and influence intelligibility (Ahissar et al., 2001; Luo and Poeppel, 2007, 2012; Gross et al., 2013; Peelle et al., 2013; Doelling et al., 2014) yet their causal role with regards to speech comprehension remains debated (Howard and Poeppel, 2010; Obleser et al., 2012). In particular, oscillatory tracking may principally reflect the temporal statistics of the speech signals – i.e. following or frequency-tagged responses - from which linguistic processing would be derived. To the contrary, our results suggest that low-frequency neural oscillations do not simply follow the temporal dynamics of the speech signal; rather, the reported changes in auditory response latencies appear to be commensurate with the observed changes in conscious perceptual parsing.

Specifically, the phase of neural oscillations following syllabic rate (3 Hz) was modulated in both monosyllabic and bisyllabic verbal transformations. Of interest here, the phase differences observed in the right hemisphere could indicate the perceived word in monosyllabic sequences, i.e. the “lampe” and “sep” sequences. The observed right bias in the encoding of slow speech dynamics is consistent with previous reports of an asymmetry in speech sampling mechanisms across hemispheres (Poeppel, 2003; Boemio et al., 2005; Giraud et al., 2007). Additionally, the sub-harmonic (1.5 Hz) oscillation was present only in bi-syllabic sequences suggesting that, when needed, supplementary brain oscillators are recruited for speech parsing. The PLV and phase of 1.5 Hz oscillations could also discriminate perceived word in the “kepi” word sequence, suggesting that 1.5 Hz dynamics could reflect parsing mechanisms. In agreement with previous reports, these results point to an important role of slow dynamics in speech perception (Luo and Poeppel, 2007; Peelle et al., 2013; Zion Golumbic et al., 2013; Doelling et al., 2014). Consistently, rhythms in the 1-3 Hz (delta) range are associated with the prosodic dynamics in the speech signal (Poeppel,

2003). These slow fluctuations give strong cues for speech parsing (Greenberg et al., 2003), that are reliable even in the presence of noise (Ding and Simon, 2012).

We also found a correlation between perceived speech utterance and alpha power. There is evidence that alpha power suppression is associated with intelligibility (Obleser and Weisz, 2012; Doelling et al., 2014). In our study, if low alpha power maximizes intelligibility then it would have to be associated with the perception of the veridical repeated utterance in the monosyllabic sequences. Hence, less alpha power should be linked with the percept “lampe” in the “lampe” sequence, and “sep” in the “sep” sequence. Yet, the opposite is observed: the perception of the veridical utterance was correlated with higher alpha power compared to the perception of the transformed utterance (here, “plan” or “pse”). For the bi-syllabic sequences, the sequences were elaborated through the concatenation of syllables; as such there is no “veridical” bi-syllabic speech utterance in the signal: for example, despite the starting syllables both “kepi” and “piquer” acoustic sequences are similar. Yet, we know from the participant’s spontaneous reports that one word is more easily perceivable in these sequences (fig. 1). Then, the word that is more often reported during the spontaneous sequence should also be easier to maintain during the volitional task. Interestingly, more alpha power is observed during the maintenance of these words, linking alpha power to the difficulty of the volitional task. Therefore, we rather interpret changes in alpha power as an attentional effect. Alpha power decreases when the volitional task is harder, i.e. when the word to maintain is less spontaneously perceived within the speech sequence. This is in agreement with prior reports demonstrating that decrease in alpha power is observed at the locus of attention (Klimesch et al., 1998; Herrmann and Knight, 2001; Sauseng et al., 2005).

4.2. Are slow oscillations parsing the acoustic signal?

While we report differences in 1.5 Hz and 3 Hz phase according to speech percept, the phase differences appear too small to confirm the brain oscillatory parsing hypothesis (Ghitza, 2011; Giraud and Poeppel, 2012). For mono-syllabic sequences, phase shifts in the 3 Hz neural oscillations were of 8-9° which corresponded to changes in neural latency of 7-8 ms yet the temporal distance between the two consecutive consonants in the mono-syllabic sequences “lampe” and “sep” were of 80-100 ms. Hence, if 3 Hz oscillations parsed the signal into perceptual word chunks during the verbal transformation task, a phase difference of 80-100 ms (~ 90-110°) would have been expected. Similarly, phase differences at 1.5 Hz should have

been radically bigger: out-of phase differences were expected if 1.5 Hz parsed the signal into bi-syllabic units.

These results seem to question the imputed role of brain oscillations as parsers of the acoustic signals. Yet, due to the poor spatial resolution of our MEG analysis, it is possible that our findings reflect the combined activity of distinct brain areas that serve dissociable mechanisms. Previous studies have indeed suggested that distinct networks reflect pure bottom-up processing and lexical and phonemic parsing. In particular, the tracking of fine-grained acoustic features is hypothesized to be restricted to primary auditory areas (Kubaneck et al., 2013), while envelope tracking in higher order regions (as STS or Broca's region) and might be specialized for speech processing (Boemio et al., 2005; Kubaneck et al., 2013; Zion Golumbic et al., 2013), and for attentional selection (Besle et al., 2011; Zion Golumbic et al., 2013). In other words, our MEG data might capture analogous stimulus-tracking mechanisms in primary auditory cortices and brain oscillators for perceptual speech parsing in higher order areas.

Alternatively, our results might suggest that during speech listening, neural low-frequency dynamics mostly reflect stimulus tracking and that delta-theta oscillation phase is weakly modulated by endogenous factors. The discrepancy between our small effects and the strong out-of phase modulations reported during audiovisual attentional selection (Lakatos et al., 2008; Besle et al., 2011; Gomez-Ramirez et al., 2011) and cocktail party effects (Zion Golumbic et al., 2013) might be explained by *the presence of competing sensory inputs*. In these experiments, two distinct rhythmic inputs were competing for attentional selection, and the phase of slow oscillations was reflecting the dynamics of the selected sensory input (Lakatos et al., 2008; Besle et al., 2011; Gomez-Ramirez et al., 2011). Even if these selective mechanisms are drastically modulated depending on the attended stimulus (Zion Golumbic et al., 2013) their dynamics are always based of *existing* external temporal information. It is thus possible that the slow dynamics reflect primarily the gain relevant sensory information, rather than endogenous temporal parsing. In our experiment, there is only one stream of information, as such the contribution of gain mechanisms might be small as there are no competing sensory inputs. To account for changes in perceptual parsing, we now explore a new hypothesis that the phase of coupling between the entrained syllabic oscillation and higher frequencies might encode perceptual parsing windows. This hypothesis is based on our previous findings and on previous evidence (Kayser et al., 2012; Ng et al., 2013) which suggest that low frequency oscillations provide reliable temporal metrics for sensory processing. Hence the position of

gamma oscillations (which is hypothesized to reflect phonemic encoding) according to the theta cycle may be a crucial cue for segmenting speech into syllables. In other words, we now test whether the modulation of PAC between the entrained slow oscillations and gamma power predicts perceived speech parsing.

4.3. Volitional and spontaneous speech perception

The present MEG report concerns the voluntary control of speech perception during the presentation of the ambiguous speech sequence. The volitional session was particularly interesting to analyze, considering that speech percepts were more stable in this session than during the spontaneous task. Yet, it is possible that distinct mechanisms are recruited during these two sessions. In particular, we expect that the reported alpha power modulations are restricted to the volitional task, but that slow oscillations phase predicts the perceived word in the sequence in the spontaneous task as well. Ongoing work is investigating the link between spontaneous and voluntary report of speech segmentation.

ACKNOWLEDGMENTS

This work was supported by an ERC-YStG-263584, an ANR10JCJC-1904 and an IRG-249222 to V.vW. We are grateful to the NeuroSpin infrastructure groups, for their support in participant recruitment. We thank Anahita Basirat for the elaboration of the acoustic sequences, Leila Azizi for her help on data acquisition, and Alexandre Gramfort for his help on data analysis.

REFERENCES

- Ahissar E, Nagarajan S, Ahissar M, Protopapas A, Mahncke H, Merzenich MM (2001) Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proc Natl Acad Sci U S A* 98:13367–13372.
- Akam T, Kullmann DM (2014) Oscillatory multiplexing of population codes for selective communication in the mammalian brain. *Nat Rev Neurosci* 15:111–122.

- Basirat A, Schwartz J-L, Sato M (2012) Perceptuo-motor interactions in the perceptual organization of speech: evidence from the verbal transformation effect. *Philos Trans R Soc Lond B Biol Sci* 367:965–976.
- Besle J, Schevon CA, Mehta AD, Lakatos P, Goodman RR, McKhann GM, Emerson RG, Schroeder CE (2011) Tuning of the human neocortex to the temporal dynamics of attended events. *J Neurosci* 31:3176–3185.
- Billig AJ, Davis MH, Deeks JM, Monstrey J, Carlyon RP (2013) Lexical influences on auditory streaming. *Curr Biol* 23:1585–1589.
- Boemio A, Fromm S, Braun A, Poeppel D (2005) Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nat Neurosci* 8:389–395.
- Boersma P (2002) Praat, a system for doing phonetics by computer. *Glott Int* 5:341–345.
- Canolty RT, Edwards E, Dalal SS, Soltani M, Nagarajan SS, Berger MS, Barbaro NM, Knight RT (2006) High gamma power is phase-locked to theta oscillations in human neocortex. *Science* 313:1626–1628.
- Canolty RT, Knight RT (2010) The functional role of cross-frequency coupling. *Trends Cogn Sci* 14:506–515.
- Ding N, Simon JZ (2012) Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc Natl Acad Sci U S A* 109:11854–11859.
- Doelling KB, Arnal LH, Ghitza O, Poeppel D (2014) Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage* 85 Pt 2:761–768.
- Elman JL, McClelland JL (1988) Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *J Mem Lang* 27:143–165.
- Fisher NI (1995) *Statistical Analysis of Circular Data*. Cambridge: Cambridge University Press.
- Gagnepain P, Henson RN, Davis MH (2012) Temporal predictive codes for spoken words in auditory cortex. *Curr Biol* 22:615–621.
- Ghitza O (2011) Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. *Front Psychol* 2:130.
- Giraud A-L, Kleinschmidt A, Poeppel D, Lund TE, Frackowiak RSJ, Laufs H (2007) Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron* 56:1127–1134.
- Giraud A-L, Poeppel D (2012) Cortical oscillations and speech processing: emerging computational principles and operations. *Nat Neurosci* 15:511–517.

- Gomez-Ramirez M, Kelly SP, Molholm S, Sehatpour P, Schwartz TH, Foxe JJ (2011) Oscillatory Sensory Selection Mechanisms during Intersensory Attention to Rhythmic Auditory and Visual Inputs: A Human Electrographic Investigation. *J Neurosci* 31:18556–18567.
- Greenberg S, Carvey H, Hitchcock L, Chang S (2003) Temporal properties of spontaneous speech—a syllable-centric perspective. *J Phon* 31:465–485.
- Gross J, Hoogenboom N, Thut G, Schyns P, Panzeri S, Belin P, Garrod S (2013) Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biol* 11:e1001752.
- Hari R (1989) Neuromagnetic steady-state responses to auditory stimuli. *J Acoust Soc Am* 86:1033.
- Herrmann CS, Knight RT (2001) Mechanisms of human attention: event-related potentials and oscillations. *Neurosci Biobehav Rev* 25:465–476.
- Hertrich I, Dietrich S, Ackermann H (2013) Tracking the speech signal--time-locked MEG signals during perception of ultra-fast and moderately fast speech in blind and in sighted listeners. *Brain Lang* 124:9–21.
- Howard MF, Poeppel D (2010) Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. *J Neurophysiol* 104:2500–2511.
- Kayser C, Ince RAA, Panzeri S (2012) Analysis of slow (theta) oscillations as a potential temporal reference frame for information coding in sensory cortices. *PLoS Comput Biol* 8:e1002717.
- Klimesch W, Doppelmayr M, Russegger H, Pachinger T, Schwaiger J (1998) Induced alpha band power changes in the human EEG and attention. *Neurosci Lett* 244:73–76.
- Kubanek J, Brunner P, Gunduz A, Poeppel D, Schalk G (2013) The tracking of speech envelope in the human cortex. Rodriguez-Fornells A, ed. *PLoS One* 8:e53398.
- Lachaux JP, Rodriguez E, Martinerie J, Varela FJ (1999) Measuring phase synchrony in brain signals. *Hum Brain Mapp* 8:194–208.
- Lakatos P, Karmos G, Mehta AD, Ulbert I, Schroeder CE (2008) Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science* 320:110–113.
- Lakatos P, Shah AS, Knuth KH, Ulbert I, Karmos G, Schroeder CE (2005) An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *J Neurophysiol* 94:1904–1911.
- Luo H, Poeppel D (2007) Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54:1001–1010.

- Luo H, Poeppel D (2012) Cortical oscillations in auditory perception and speech: evidence for two temporal windows in human auditory cortex. *Front Psychol* 3:170.
- Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods* 164:177–190.
- Ng B, Logothetis N, Kayser C (2013) EEG phase patterns reflect the selectivity of neural firing. *Cereb Cortex* 23:389–398.
- Nozaradan S, Peretz I, Missal M, Mouraux A (2011) Tagging the neuronal entrainment to beat and meter. *J Neurosci* 31:10234–10240.
- Obleser J, Herrmann B, Henry MJ (2012) Neural Oscillations in Speech: Don't be Enslaved by the Envelope. *Front Hum Neurosci* 6:250.
- Obleser J, Weisz N (2012) Suppressed alpha oscillations predict intelligibility of speech and its acoustic details. *Cereb Cortex* 22:2466–2477.
- Peelle JE, Gross J, Davis MH (2013) Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cereb Cortex* 23:1378–1387.
- Pitt M, Shoaf L (2002) Linking verbal transformations to their causes. *J Exp Psychol Hum Percept Perform* 28:150.
- Poeppel D (2003) The analysis of speech in different temporal integration windows: cerebral lateralization as “asymmetric sampling in time.” *Speech Commun* 41:245–255.
- Poeppel D, Idsardi WJ, van Wassenhove V (2008) Speech perception at the interface of neurobiology and linguistics. *Philos Trans R Soc Lond B Biol Sci* 363:1071–1086.
- Pressnitzer D, Hupé J-M (2006) Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. *Curr Biol* 16:1351–1357.
- Regan D (1966) Some characteristics of average steady-state and transient responses evoked by modulated light. *Electroencephalogr Clin Neurophysiol* 20:238–248.
- Sato M, Basirat A, Schwartz J-L (2007) Visual contribution to the multistable perception of speech. *Percept Psychophys* 69:1360–1372.
- Sato M, Schwartz J-L, Abry C, Cathiard M-A, Loevenbruck H (2006) Multistable syllables as enacted percepts: a source of an asymmetric bias in the verbal transformation effect. *Percept Psychophys* 68:458–474.
- Sauseng P, Klimesch W, Stadler W, Schabus M, Doppelmayr M, Hanslmayr S, Gruber WR, Birbaumer N (2005) A shift of visual spatial attention is selectively associated with human EEG alpha activity. *Eur J Neurosci* 22:2917–2926.
- Schroeder CE, Lakatos P (2009) Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci* 32:9–18.

- Schroeder CE, Lakatos P, Kajikawa Y, Partan S, Puce A (2008) Neuronal oscillations and visual amplification of speech. *Trends Cogn Sci* 12:106–113.
- Shoaf LC, Pitt MA (2002) Does node stability underlie the verbal transformation effect? A test of node structure theory. *Percept Psychophys* 64:795–803.
- Stefanics G, Hangya B, Hernadi I, Winkler I, Lakatos P, Ulbert I (2010) Phase entrainment of human delta oscillations can mediate the effects of expectation on reaction speed. *J Neurosci* 30:13578–13585.
- Taulu S, Kajola M, Simola J (2003) Suppression of Interference and Artifacts by the Signal Space Separation Method. *Brain Topogr* 16:269–275.
- Uusitalo MA, Ilmoniemi RJ (1997) Signal-space projection method for separating MEG or EEG into components. *Med Biol Eng Comput* 35:135–140.
- Vroomen J, Zon M, Gelder B (1996) Cues to speech segmentation: Evidence from juncture misperceptions and word spotting. *Mem Cognit* 24:744–755.
- Warren RM (1968) Verbal transformation effect and auditory perceptual mechanisms. *Psychol Bull* 70:261.
- Zion Golumbic EM, Ding N, Bickel S, Lakatos P, Schevon CA, McKhann GM, Goodman RR, Emerson R, Mehta AD, Simon JZ, Poeppel D, Schroeder CE (2013) Mechanisms Underlying Selective Neuronal Tracking of Attended Speech at a “Cocktail Party.” *Neuron* 77:980–991.

GENERAL DISCUSSION

5.1. SUMMARY OF THE FINDINGS

As outlined in the introduction, neural oscillations are ideal candidates to encode perceptual timing. First, they represent periodic fluctuations in time thereby providing reliable temporal metrics to construct a brain-centric reference for neural timing (Panzeri et al., 2010; Kayser et al., 2012). Second, the brain likely uses oscillations as a communication system across sensory cortices, and oscillations are hypothesized to play a crucial role in temporal binding (Engel and Singer, 2001a; Fries, 2005). Third, neural oscillations are modulated stimulus-driven rhythms. Oscillations could then manifest the tuning of brain activity to external temporal contingencies (Schroeder and Lakatos, 2009; Stefanics et al., 2010a). Finally, oscillations are also modulated by top-down control (Lakatos et al., 2008; Besle et al., 2011). Actually, large scale dynamics as captured through oscillatory activity even constitute a form of top down modulation of the signal. “Top-down” is usually seen as feedback influence of higher hierarchy to local networks activity. Yet, top-down processing could also concern any other modulation of the sensory signal which manifests past experience (Engel et al., 2001; Arnal and Giraud, 2012). Hence, oscillatory behavior could both explain the reliance of perceptual timing to external information and the departure of perceptual timing from veridical timing.

The distinct studies of this thesis provide evidence in support to the role of brain oscillations in the encoding of event timing for perception. In the Chapter 2, we showed that the shift in the phase of low-frequency neural oscillation in auditory cortices could predict the subjective perception of audiovisual timing. Specifically, we showed that the auditory and visual cortices could tune neural activity to the temporal regularities of

sensory inputs. The variation of this tuning in auditory cortices over time was commensurate with variations of perceptual timing. We interpret these results as the emergence of a “pacemaker” oscillation through neural entrainment. The phase of the pacemaker oscillation would in particular serve the tagging of event timing for perception. In other words, we argue that entrained oscillations provide an internal reference of timing for perception and that the modulation of this reference explains distortions of perceptual timing. However, phase encoding of timing may be specific to entrainment situations. When no external temporal information was given prior to temporal judgments, no clear pacemaker oscillation was observed in auditory or in visual cortices. The present findings in chapter 2 point to a specific role of neural entrainment in multisensory temporal correspondence, but its characterization in terms of frequency was fixed at a slow rate, leaving unanswered the possible role of higher frequency oscillations in audiovisual perceptual timing. The impact of the frequency of neural entrainment in the building up of multisensory temporal expectations was reported in Chapter 3. Results show that optimal temporal binding occurs at relatively low stimulation temporal rates. This suggests that the mechanisms underlying audiovisual temporal communication systems are bounded in the delta range (1-2 Hz). Finally, we tried to replicate these findings in chapter 4 in a task where event timing was implicitly measured. To do so we used a speech comprehension task. We tested whether changes in the timing of perceptual parsing of the acoustic signal was commensurate with shifts in auditory cortex oscillations. The present results show that changes in low frequency oscillation entrainment could predict perceptual speech parsing. Yet the observed phase shifts were too small to account for the perceptual temporal shifts in the speech parsing window. These results could hint at an important role of the coupling between the phase of the low frequency oscillations and higher frequencies.

5.2. TWO ROLES FOR THE PHASE OF SLOW-OSCILLATIONS IN EEG/MEG RECORDINGS?

5.2.1. *Absolute and relative phase*

Our results, in line with previous reports, provide evidence that cortical rhythms serve as temporal reference frames for events processing. Specifically, results in the first experiment in chapter 2 have prompted the notion that the *relative* position of the evoked response within the entrained oscillation cycle provides information on perceived stimulus timing. These findings point to a special role of relative phase in neural encoding schemes. Yet, the phase of low-frequency oscillations in the LFP signal was first found to be associated with an excitability state of the recorded neural network (Lakatos et al., 2005; Canolty et al., 2006; Womelsdorf et al., 2007). It is suggested that there is an absolute relationship between low-frequency phase and neural excitability, such as maximal firing occurs at the trough of the low-frequency oscillation (Lakatos et al., 2005). As such, this mechanism is expected to suppress neural activity at specific time periods, and should have major consequences on perception over time. Human M/EEG studies consistently find a link between the phase of neural oscillation and perceptual performance in time (Monto et al., 2008; Busch et al., 2009; Mathewson et al., 2009; Chakravarthi and Vanrullen, 2012; Henry and Obleser, 2012; Neuling et al., 2012b; Fiebelkorn et al., 2013). In these reports, a specific phase of the slow oscillation is associated with maximal perceptual performance (in stimulus detection or discrimination), while the opposite phase is associated with minimal performance.

In most of these studies, the phase linked with optimal performance was not fixed among subjects. Variation was reported in the preferential phase of maximal excitability (Busch et al., 2009; Chakravarthi and Vanrullen, 2012; Ng et al., 2012; Fiebelkorn et al., 2013) or in the phase of neural entrainment (Besle et al., 2011; Henry and Obleser, 2012). Phase normalization procedures were used in order to quantify phase effects at the inter-subject level (Busch et al., 2009; Chakravarthi and Vanrullen, 2012; Ng et al., 2012; Fiebelkorn et al., 2013). The variability in phase across individuals is striking: it shows that best performance does not correspond to the same phase across subjects thus raising concerns with regards to the optimal excitability in the phase of low frequency oscillations. Specifically, if we consider that the absolute phase carries excitability characteristics (i.e. oscillatory peaks correspond to periods of inhibition), and that best performance is associated to maximal excitability of the recruited sensory system, the phase of the neural oscillation should be consistent across individuals. In line with these counterintuitive findings, we also report in our studies a large variability in the preferential phases of the auditory and visual entrainment (fig. 5.1). Here again, a small inter-subject variability

was expected, considering that entrainment should reflect alignment of neural excitability to the dynamics of external stimulation.

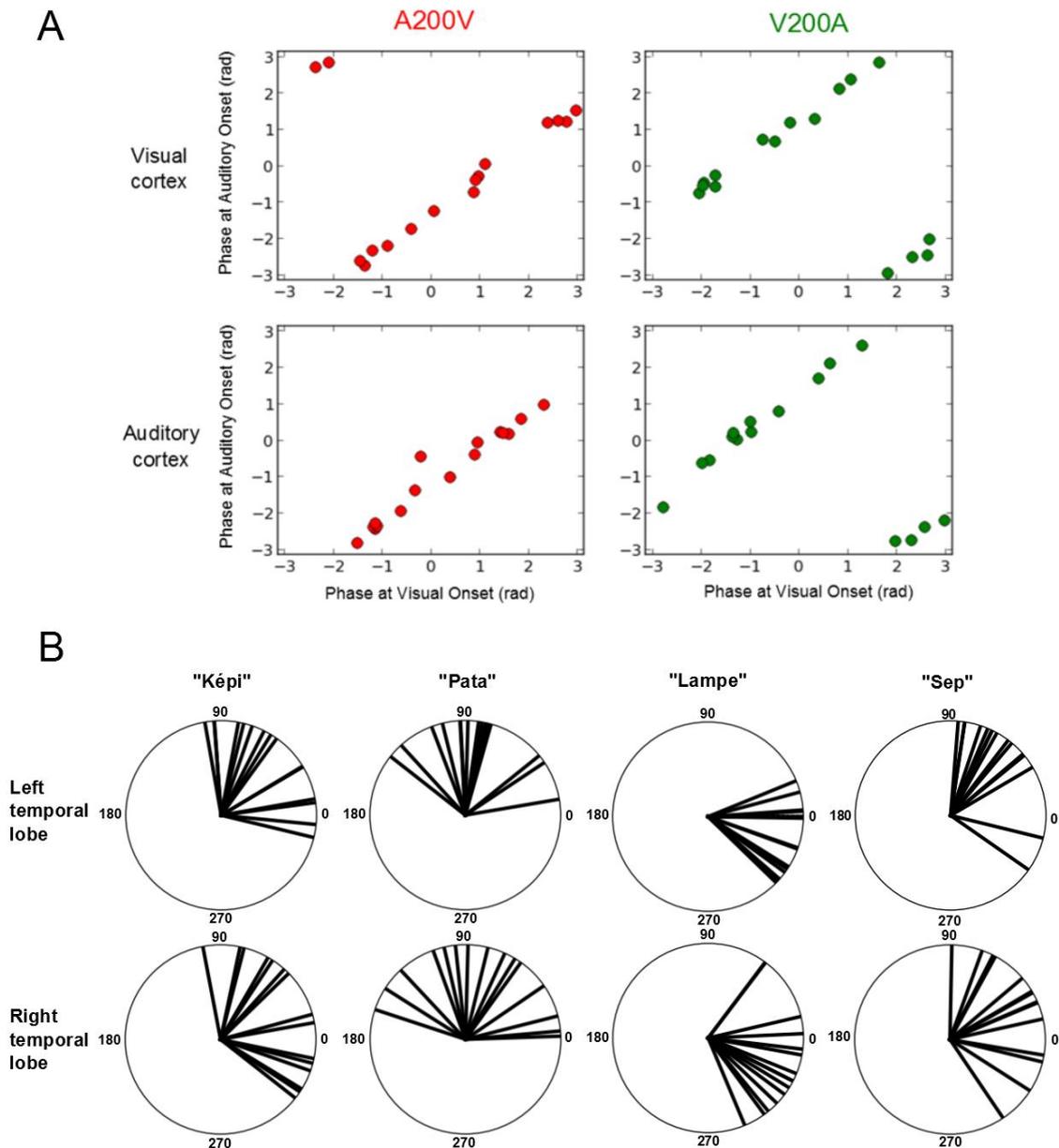


Figure 5.1.: Inter-subject variability in the absolute phase of entrainment. A. Correlation between the preferential phase computed at auditory onset and the phase computed at visual onset at the end of the adaptation period in the study presented in chapter 2. Each dot corresponds to one subject. The phase computed at auditory onset is highly correlated with the phase at visual onset across conditions. Here, auditory and visual onsets are separated by 200 ms, hence the reported correlations are a marker of oscillatory entrainment. Importantly, the variability across individuals is very high. B. Preferential phase distributions for each individual in the speech parsing experiment presented in chapter 4. Each black bar corresponds to one subject. Again inter-subject variability in entrainment is observed.

The normalization procedure in M/EEG studies are justified with the fact that absolute polarity of EEG and MEG signals are hard to quantify. Contrary to LFP studies, the polarity of electrical activity as recorded with EEG and MEG techniques is undetermined. In EEG, the absolute value depends on the choice of the reference. The electric and magnetic fields could be recorded from distinct sensors with opposite polarities. As such it is not possible to know this exactitude whether a peak in a M/EEG oscillation corresponds to a through in neural excitability. It is therefore difficult to interpret absolute phase values with these techniques. But polarity effects cannot explain the reported subject variability, as polarity effects would lead to strongly oriented bipolar distributions, and the distributions that we report have a different shape (fig. 5.1.).

The ambiguity between absolute and relative phase coding is also present in neurophysiological and neuroimaging studies that report neural synchronization between distinct brain regions. Indeed, reported “oscillation synchrony” can refer to two distinct observations. First, oscillation synchrony can relate to the true synchronization of neuronal activity within and between neural assemblies. True synchrony was reported in seminal neurophysiological studies, which showed gamma oscillations in the cross-correlation between neural signals from neural assemblies. Crucially, the peak of the cross-correlogram was centered around 0 ms, suggesting that the two neural assemblies were fluctuating synchronously at gamma rate (Gray et al., 1989; Engel and Kreiter, 1991; Engel et al., 1991d) (fig. 1.11). In comparison, it is much harder to detect true synchrony of local neural ensembles when using EEG or MEG neuroimaging techniques, because the spatial resolution is much lower than with intracranial recordings. An augmentation in local synchronization could still be observed through an increase in oscillation power: if a network resonates at a specific frequency, this should cause a global increase in oscillation power that could be captured with EEG and MEG sensors (von Stein and Sarnthein, 2000). The second observation of oscillation synchrony refers to the *coherence* of neural activity across distinct brain regions. In this particular scenario, synchronization is observed when two regions present similar temporal patterns of activity, albeit some neural delays. In other words, neural coherence analysis captures brain regions that present comparable temporal fluctuations in oscillatory activity but do not necessarily fire synchronously. 0 ms-coherence can be problematic in neuroimaging data, as it reflects artifactual synchrony due to volume conduction spread-out of neural activity (i.e. that sensors do not capture independent signatures of the neural signal)

(Tognoli and Kelso, 2009). As such 0 ms-coherence synchrony is often discarded from the results (Lachaux et al., 1999; von Stein and Sarnthein, 2000; Palva and Palva, 2011). Hence, a lot of evidence for oscillatory synchronization may reflect delayed coherence between brains regions.

The observed phase shifts in neural coherence could possibly reflect the transmission delay of information between the two communicating networks (Gregoriou et al., 2009). However, we show in this thesis that relative phase shifts in oscillations can be modified with experience. They may reflect more than direct transmission delays and can give insightful information about perceptual content. As a result, we suspect that the relative delays in neural activity_ especially at low frequencies _ could provide a valuable code for coordinating cell assemblies in time in the context of temporal perception.

5.2.2. Temporal integration and segregation

Neural oscillations have been associated with two distinct temporal mechanisms: temporal integration and temporal ordering. Temporal integration (or temporal sampling) relates to the fusion of sensory information over a certain period of time. A consequence of this fusion is that all the information of the sub-elements that are fused should be lost. By analogy, it is as if the brain digitalizes sensory information: when a picture or a sound is digitalized, the information that was contained within a pixel or a sound sample cannot be retrieved anymore. Yet, several studies suggest to the contrary that these elements are not lost, but hierarchically organized with banks of brain oscillators (Lakatos et al., 2005; Schroeder et al., 2008; Canolty and Knight, 2010; Ghitza, 2011; Giraud and Poeppel, 2012).

This hierarchical oscillatory code is particularly interesting as it would permit to multiplex multiple cognitive processing via different oscillations (Akam and Kullmann, 2010). Hence distinct processes could be recruited depending on the task and temporal demands. The parallelization of the different oscillators should lead to distinct temporal windows of interest. When performing a certain task, the relevant information could be selected by filtering the neural signal at the good frequency (Akam and Kullmann, 2010, 2014). The resulting signal would have a temporal resolution that is directly imposed by

the recruited oscillation. Hence, the temporal level at which the sensory signal will be fused into a unified percept depends on the cognitive task.

Following this idea, brain mechanisms for temporal integration and for segregation should be hierarchically organized. Based on our results and on previous reports, we argue that integration/ sampling mechanisms recruit essentially slow oscillation mechanisms, while temporal coding of sub-elements are preserved in nested higher oscillations (Akam and Kullmann, 2010, 2014; Buzsáki, 2010). First, previous distinctions between temporal window of encoding and temporal window of integration suggest that integration windows are of a coarser time scale than encoding windows (Theunissen and Miller, 1995; van Wassenhove, 2013). The temporal window of encoding reflects the minimal temporal window below which the two sensory inputs cannot be distinguished. Conversely, the temporal window of integration includes the temporal window of encoding plus some temporal tolerance to timing noise (Theunissen and Miller, 1995). In other words, the temporal windows of integration should always be bigger than the temporal window of encoding.

Temporal integration mechanisms are usually known to be of different temporal scales between sensory modalities (Hirsh and Sherrick Jr., 1961; Shimojo, 2001) and stimulus complexity (van Wassenhove et al., 2007; van Wassenhove, 2009, 2013). Interestingly, perceptual reports in sensory binding usually have a coarser temporal scale for congruent audiovisual features than incongruent features (van Wassenhove et al., 2007). This suggests that possible integration between features increases tolerance to temporal delay. Furthermore, dissociations can be made between the different types of timing reports. There is emerging consensus that perceptual reports of sensory synchrony and temporal order operate via different mechanisms (Van Eijk et al., 2008; Vatakis et al., 2008; Ipser et al., 2012; Love et al., 2013). Numerous studies have found dissociations between Temporal Order Judgments (TOJ) and synchrony judgments (SJ), as well as between TOJ and audiovisual binding measures, such as the proportion of McGurk illusion (Ipser et al., 2012; Freeman et al., 2013). To the contrary, SJ and McGurk temporal windows seem to correlate positively (Martin et al., 2013). Interestingly, perceptual synchrony reflects the perception of temporal unity between two distinct features, and is close to that sense to sensory integration. To the contrary, temporal order judgments rely on the accurate temporal dissociation between two sensory events. Hence it is plausible that

temporal order judgments require finer temporal resolution than synchrony or sensory integration mechanisms. Temporal order judgments may thus recruit oscillatory mechanisms on a higher temporal scale than simultaneity judgments and temporal integration.

Yet, a direct consequence of the parallelization of oscillatory codes is that cognitive processes that work at slow and fast dynamics are interdependent. Of interest our results in chapter 2 and 3 suggest that audiovisual temporal order and integration both recruit slow oscillations in the same temporal scale (~1 Hz). However, they seem to take advantage of different information that is conveyed through the slow rhythm. We showed in Chapter 3 that automatic audiovisual binding has a coarse temporal window, suggesting that audiovisual perceptual units might be computed in one cycle of the slow oscillation. Further studies reported consistent effects, and suggest that ~1 Hz oscillations play a crucial role in audiovisual integration (Lakatos et al., 2008; Fiebelkorn et al., 2011, 2013; Gomez-Ramirez et al., 2011). Yet audiovisual integration mechanisms do not alleviate event timing information, which could be retrieved through the relative phase of the slow oscillation cycle. Hence, we propose that the dissociation between the “what” and “when” of sensory information could be explained by the same oscillatory mechanism. However, the “what” of sensory information would correspond to the information arriving within the time window delimited by one cycle of the recruited oscillation; while the “when” would be encoded within the phase of the slower oscillation.

5.3. MULTI- AND AR- RHYTHMIC BRAIN MECHANISMS OF EVENT TIMING

5.3.1. Multiplexing of temporal information with multiple brain oscillators

If oscillations provide metrics of timing, it is possible that they do not only concern one carrier frequency. The brain should indeed capitalize on multiple oscillators to accurately tune to the external rhythmicity of the world. This characteristic was exemplified in

particular in chapter 3. There is emerging consensus that speech listening mechanisms involve brain oscillations at phrasal, syllabic and phonemic temporal rates (Arnal and Giraud, 2012; Luo and Poeppel, 2012; Doelling et al., 2014). Consistent with this, as presented in chapter 4, listening to mono-syllabic stimuli engendered the emergence of entrained oscillation at syllabic rate, but sub-harmonic oscillations were representative of perceived words when bi-syllabic. These findings suggest that the brain creates rhythms that are not physically present in the signal to process it with better accuracy. Accordingly, the perception of meter in music perception also relies on brain oscillatory mechanisms (Nozaradan et al., 2011). Specifically, when listening to an isochronous acoustic rhythm, individuals tend to naturally group these sequences by two or more elements (Jones, 1987; Grahn, 2012; Grahn and Rowe, 2013). This grouping refers to meter in music. Asking musicians to imagine that these sequences were elaborated from a binary (groups of 2 crotchets) or ternary (groups of 3 crotchets) meter, makes the auditory cortex fluctuate at this imagined rhythm (Nozaradan et al., 2011). In a similar vein, perceiving complex rhythmic structures generates rhythmic activity in auditory cortex that are consistent with both perceived beat and metrical structure (Nozaradan et al., 2012).

These findings are in line with more complex forms of the dynamic attending theory (DAT) (Jones, 1976; Large and Jones, 1999). The DAT suggests that temporal expectations of subsequent sensory information could be built based on the temporal structure of previous stimulation (Jones, 1976). While most evidence for the DAT stands for the influence of simple rhythmic sequences (Jones and Boltz, 1989; Barnes and Jones, 2000; Jones et al., 2002; Mathewson et al., 2010), it is expected that complex rhythmic sequences also entails temporal predictions. Hence, brain tuning when listening to music could produce expectation at sub-tempo level (Mcauley, 1995). Similarly, finger tapping in synchrony with complex sounds is affected by both beat rate and subdivisions of rhythms (Repp, 2008). Further work is needed to test whether these behavioral effects emerge from the parallel phase tuning of multiple neural oscillators.

The existence of multiple oscillators provides an additional line of direction for the interpretation and analysis of the MEG data presented in chapter 4. In this chapter, we focused on the role of low frequency oscillations in the perceptual parsing of syllabic and bi-syllabic words within acoustic sequences. Yet other neural oscillations in higher

frequency bands are suggested to take part in speech comprehension mechanisms; in particular low gamma oscillations could represent phonemic encoding (Poeppel, 2003; Boemio et al., 2005; Lehongre et al., 2011; Giraud and Poeppel, 2012; Luo and Poeppel, 2012), and beta oscillations could serve the parsing of dyads (i.e. the acoustic boundary between phones) (Ghitza, 2011). In the Tempo model introduced by Ghitza (2011), it is the relative phase of these higher frequency oscillators according to the theta reference that determines the temporal parsing window of speech. Hence, slow oscillations may principally track the temporal regularities of the acoustic signal, and then give the “tempo” to higher frequency oscillators that decode the content of speech. This model offers new perspective and refines our proposal of how slow oscillations could constitute an internal temporal reference for speech perception.

5.3.2. Entrainment vs. arrhythmic stimulation

Here, the results presented in Chapter 2 and 4 highlight the role of cortical oscillations in the presence of external rhythmic stimulation. Yet, stimulus presentation can be highly arrhythmic. Furthermore, detecting the transience of events in time is actually an important cue for attentional selection and accurate perception.

The data presented in chapter 2 may suggest that distinct mechanisms are involved in the encoding of time in the presence or absence of external temporal regularities. As discussed in chapter 2, while neural entrainment could provide temporal metrics for timing perception, no clear endogenous pacemaker oscillation was observed when no temporal information was given prior to temporal order judgments. In particular, we either expected the involvement of oscillations in the delta range in the construction of subjective audiovisual timing. This was based on the results of the entrained experiment, and also as delta-frequency oscillations are known to affect audiovisual binding even with arrhythmic stimulation (Fiebelkorn et al., 2011, 2013). The participation of strong endogenous oscillatory components was also emitted as an alternative hypothesis. In particular alpha phase in visual cortex was previously shown to affect visual synchrony and temporal order judgments (Varela et al., 1981; Gho and Varela, 1988; Chakravarthi and Vanrullen, 2012). From the lack of significant results in our experiment we cannot conclude whether endogenous oscillations in sensory cortices take part in subjective

timing processes. Based on the big inter-subject variability in our preliminary results, we suspect that if endogenous mechanisms in sensory cortices serve the encoding of time in this absence of entrainment, they should recruit oscillators of distinct frequencies that are subject-specific. We consider also the possibility that, in the absence of stimulation, temporal pacemakers originate from other brain regions, in particular frontal regions as previously suggested (Matell and Meck, 2004; Buhusi and Meck, 2005).

In addition, the lack of endogenous pacemakers for audiovisual timing may relate to the apparent asymmetry in temporal processing between visual and auditory/ audiovisual processing: evidence of visual sampling through alpha band oscillations have been put forward. While sound processing is also hypothesized to rely on oscillatory parsing mechanisms in the auditory cortex (Schroeder et al., 2008; Ghitza, 2011; Giraud and Poeppel, 2012), alpha oscillations behavior does not seem to have the same impact in auditory and in visual cortices (İlhan and VanRullen, 2012). In general, no specific frequency oscillation has been found to sample auditory information in the absence of external temporal regularities (Thorne and Debener, 2014). Periodic sampling is an efficient way to extract sensory information, because it entails a reduction in sensory information complexity (all the detailed fluctuations within the sample are reduced to one data point), and is energetically advantageous (neural processing is shunt periodically). Visual processing could indeed benefit from such system, as all the elements in the visual scene can be captured within a very short amount of time. To the contrary, auditory information is by essence transmitted via temporal changes or air pressures in the signal. As such, the precise capture of temporal information is crucial to accurately retrieve the sound content. It seems obvious that sampling sounds periodically with no knowledge of the acoustic signal would have damaging consequences on perception. Yet, auditory processing could capitalize on external rhythms to build temporal expectations on the arrival of next relevant events. Interestingly, the most common rhythmic stimulations in the external world _ speech, music, and movement namely _ are in the auditory and motor domain. Hence it seems reasonable to think that brain tuning to external regularities has more consequences for auditory information than visual information (Thorne and Debener, 2014). In conclusion, while visual processing could rely on non-informed *sampling* mechanisms to efficiently process information in time, this strategy cannot be applied in the auditory domain due to the necessity of fine temporal accuracy. Yet, audition could still take advantage of the temporal cues given by

the environment to parse the signal into perceptually meaningful units, while preserving the finer temporal resolution of information for accurate content extraction.

5.4. CONCLUSION

Watching a movie without noticing the succession of images and the asynchrony of dubbing, thinking that the day never ends when reading a boring text, all these situations signal that perceptual timing is an inaccurate construct of reality. The departure of perceptual timing from physical timing was thought to originate from the dynamics of sensory processing. In accordance with this view, our work provides evidence that the perceptual timing of events relies on an internal reference frame that is generated from brain rhythms. Specifically, we show that low-frequency oscillations (in 1-3 Hz range) in auditory cortices make the link between the physical timing of events and its distorted perception. We discussed previously possible explanations for the specific impact of low-frequency oscillations in multisensory temporal binding and the inherent asymmetry between visual and auditory processes in the encoding of timing. Yet, these explanations are still speculative and remain to be tested. Additionally, it would be interesting to test if the reported timing phase code extends to other sensory modalities and to sensorimotor timing. Overall, this thesis exposes the initial elements of a phase code model of perceptual timing. Further research is needed to understand the exact contributions of cortical oscillations in the emergence of time awareness.

REFERENCES

- Ahissar E, Nagarajan S, Ahissar M, Protopapas A, Mahncke H, Merzenich MM (2001) Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proc Natl Acad Sci U S A* 98:13367–13372.
- Akam T, Kullmann DM (2010) Oscillations and filtering networks support flexible routing of information. *Neuron* 67:308–320.
- Akam T, Kullmann DM (2014) Oscillatory multiplexing of population codes for selective communication in the mammalian brain. *Nat Rev Neurosci* 15:111–122.
- Alais D, Blake R, Lee SH (1998) Visual features that vary together over time group together over space. *Nat Neurosci* 1:160–164.
- Andrews T, Coppola D (1999) Idiosyncratic characteristics of saccadic eye movements when viewing different visual environments. *Vision Res* 39:2947–2953.
- Arnal LH, Giraud A-L (2012) Cortical oscillations and sensory predictions. *Trends Cogn Sci* 16:390–398.
- Arnold DH, Clifford CW., Wenderoth P (2001) Asynchronous processing in vision. *Curr Biol* 11:596–600.
- Arnold DH, Durant S, Johnston A (2003) Latency differences and the flash-lag effect. *Vision Res* 43:1829–1835.
- Bair W, Koch C (1996) Temporal Precision of Spike Trains in Extrastriate Cortex of the Behaving Macaque Monkey. *Neural Comput* 8:1185–1202.
- Barnes R, Jones MR (2000) Expectancy, Attention, and Time. *Cogn Psychol* 41:254–311.
- Benevento LA, Fallon J, Davis BJ, Rezak M (1977) Auditory-visual interaction in single cells in the cortex of the superior temporal sulcus and the orbital frontal cortex of the macaque monkey. *Exp Neurol* 57:849–872.
- Benjamins J, Smagt M van der, Verstraten F (2008) Matching auditory and visual signals: Is sensory modality just another feature? *Perception* 37:848.
- Besle J, Schevon CA, Mehta AD, Lakatos P, Goodman RR, McKhann GM, Emerson RG, Schroeder CE (2011) Tuning of the human neocortex to the temporal dynamics of attended events. *J Neurosci* 31:3176–3185.
- Bishop G (1932) Cyclic changes in excitability of the optic pathway of the rabbit. *Am J Physiol Content* 103:213–224.

- Blake R, Lee S-H (2005) The role of temporal structure in human vision. *Behav Cogn Neurosci Rev* 4:21–42.
- Boemio A, Fromm S, Braun A, Poeppel D (2005) Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nat Neurosci* 8:389–395.
- Bonnefond M, Jensen O (2012) Alpha oscillations serve to protect working memory maintenance against anticipated distracters. *Curr Biol* 22:1969–1974.
- Bragin A, Jandó G, Nádasdy Z (1995) Gamma (40-100 Hz) oscillation in the hippocampus of the behaving rat. *J Neurosci* 15:47–60.
- Brasselet R, Panzeri S, Logothetis NK, Kayser C (2012) Neurons with stereotyped and rapid responses provide a reference frame for relative temporal coding in primate auditory cortex. *J Neurosci* 32:2998–3008.
- Bruce V, Green P, Georgeson M (1996) *Visual Perception: Physiology, Psychology, and Ecology*. Laurence Erlbaum Assoc Hove.
- Buhusi C V, Meck WH (2005) What makes us tick? Functional and neural mechanisms of interval timing. *Nat Rev Neurosci* 6:755–765.
- Buhusi C V, Meck WH (2009) Relativity theory and time perception: single or multiple clocks? Burr DC, ed. *PLoS One* 4:e6268.
- Busch N a, Dubois J, VanRullen R (2009) The phase of ongoing EEG oscillations predicts visual perception. *J Neurosci* 29:7869–7876.
- Bushara KO, Grafman J, Hallett M (2001) Neural Correlates of Auditory-Visual Stimulus Onset Asynchrony Detection. *J Neurosci* 21:300–304.
- Buzsáki G (2004) Large-scale recording of neuronal ensembles. *Nat Neurosci* 7:446–451.
- Buzsáki G (2010) Neural syntax: cell assemblies, synapsembles, and readers. *Neuron* 68:362–385.
- Buzsáki G, Draguhn A (2004) Neuronal oscillations in cortical networks. *Science* 304:1926–1929.
- Buzsáki G, Logothetis N, Singer W (2013) Scaling Brain Size, Keeping Timing: Evolutionary Preservation of Brain Rhythms. *Neuron* 80:751–764.
- Cai M, Stetson C, Eagleman DM (2012) A neural model for temporal order judgments and their active recalibration: a common mechanism for space and time? *Front Psychol* 3:470.
- Canolty RT, Edwards E, Dalal SS, Soltani M, Nagarajan SS, Berger MS, Barbaro NM, Knight RT (2006) High gamma power is phase-locked to theta oscillations in human neocortex. *Science* 313:1626–1628.
- Canolty RT, Knight RT (2010) The functional role of cross-frequency coupling. *Trends Cogn Sci* 14:506–515.

- Capilla A, Pazo-Alvarez P, Darriba A, Campo P, Gross J (2011) Steady-state visual evoked potentials can be explained by temporal superposition of transient event-related responses. Valdes-Sosa PA, ed. *PLoS One* 6:e14543.
- Celesia GG (1976) Organization of auditory cortical areas in man. *Brain* 99:403–414.
- Chakravarthi R, Cavanagh P (2007) Temporal properties of the polarity advantage effect in crowding. *J Vis* 7:11.1–13.
- Chakravarthi R, Vanrullen R (2012) Conscious updating is a rhythmic process. *Proc Natl Acad Sci U S A* 109:10599–10604.
- Chase SM, Young ED (2007) First-spike latency information in single neurons increases when referenced to population onset. *Proc Natl Acad Sci U S A* 104:5175–5180.
- Church RM (1984) Properties of the Internal Clock. *Ann N Y Acad Sci* 423:566–582.
- Coles M, Rugg M (1995) Event-related brain potentials: An introduction. In: *Electrophysiology of Mind: Event-Related Brain Potentials and Cognition* (Coles M, Rugg M, eds), pp 1–26. London: Oxford University Press.
- Coull J, Frith C, Büchel C, Nobre A (2000) Orienting attention in time: behavioural and neuroanatomical distinction between exogenous and endogenous shifts. *Neuropsychologia* 38:808–819.
- Coull JT, Nobre AC (1998) Where and When to Pay Attention: The Neural Systems for Directing Attention to Spatial Locations and to Time Intervals as Revealed by Both PET and fMRI. *J Neurosci* 18:7426–7435.
- Coull JT, Vidal F, Nazarian B, Macar F (2004) Functional anatomy of the attentional modulation of time estimation. *Science* 303:1506–1508.
- Craig A (2009) Emotional moments across time: a possible neural basis for time perception in the anterior insula. *Philos Trans R Soc Lond B Biol Sci* 364:1933.
- Cravo AM, Rohenkohl G, Wyart V, Nobre AC (2013) Temporal expectation enhances contrast sensitivity by phase entrainment of low-frequency oscillations in visual cortex. *J Neurosci* 33:4002–4010.
- Deisz R, Prince D (1989) Frequency-dependent depression of inhibition in guinea-pig neocortex in vitro by GABAB receptor feed-back on GABA release. *J Physiol* 412:513–541.
- Destexhe A, Rudolph M, Paré D (2003) The high-conductance state of neocortical neurons in vivo. *Nat Rev Neurosci* 4:739–751.
- Dhamala M, Assisi CG, Jirsa VK, Steinberg FL, Kelso JAS (2007) Multisensory integration for timing engages different brain networks. *Neuroimage* 34:764–773.
- Di Lollo V (1980) Temporal integration in visual memory. *J Exp Psychol Gen* 109:75.
- Di Luca M, Machulla TK, Ernst MO (2009) Recalibration of multisensory simultaneity: Cross-modal transfer coincides with a change in perceptual latency. *J Vis* 9:1–16.

- Doelling KB, Arnal LH, Ghitza O, Poeppel D (2014) Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage* 85 Pt 2:761–768.
- Eagleman D (2010) How does the timing of neural signals map onto the timing of perception? *Sp Time Percept Action*:216–231.
- Eagleman DM, Sejnowski TJ (2000) Motion Integration and Postdiction in Visual Awareness. *Science* 287:2036–2038.
- Eagleman DM, Tse PU, Buonomano D, Janssen P, Nobre AC, Holcombe AO (2005) Time and the brain: how subjective time relates to neural time. *J Neurosci* 25:10369–10371.
- Eckhorn R, Bauer R, Jordan W, Brosch M, Kruse W, Munk M, Reitboeck HJ (1988) Coherent oscillations: A mechanism of feature linking in the visual cortex? *Biol Cybern* 60:121–130.
- Engel A, König P, Singer W (1991a) Direct physiological evidence for scene segmentation by temporal coding. *Proc Natl Acad Sci U S A* 88:9136–9140.
- Engel A, König P, Kreiter A, Gray C, Singer W (1991b) Temporal coding by coherent oscillations as a potential solution to the binding problem: physiological evidence. *Nonlinear Dyn neural networks*:149–163.
- Engel A, König P, Kreiter A, Singer W (1991c) Interhemispheric synchronization of oscillatory neuronal responses in cat visual cortex. *Science* 252:1177–1179.
- Engel A, König P, Kreiter A, Singer W (1991d) Interhemispheric synchronization of oscillatory neuronal responses in cat visual cortex. *Science* 252:1177–1179.
- Engel A, Kreiter A (1991) Synchronization of oscillatory neuronal responses between striate and extrastriate visual cortical areas of the cat. *Proc Natl Acad Sci U S A* 88:6048–6052.
- Engel A, Singer W (2001a) Temporal binding and the neural correlates of sensory awareness. *Trends Cogn Sci*.
- Engel AK, Fries P, König P, Brecht M, Singer W (1999) Does time help to understand consciousness? *Conscious Cogn* 8:260–268.
- Engel AK, Fries P, Singer W (2001) Dynamic predictions: oscillations and synchrony in top-down processing. *Nat Rev Neurosci* 2:704–716.
- Engel AK, Singer W (2001b) Temporal binding and the neural correlates of sensory awareness. *Trends Cogn Sci* 5:16–25.
- Fahle M, Koch C (1995) Spatial displacement, but not temporal asynchrony, destroys figural binding. *Vision Res* 35:491–494.
- Felleman DJ, Van Essen DC (1991) Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cereb Cortex* 1:1–47.
- Fiebelkorn IC, Foxe JJ, Butler JS, Mercier MR, Snyder AC, Molholm S (2011) Ready, set, reset: stimulus-locked periodicity in behavioral performance demonstrates the consequences of cross-sensory phase reset. *J Neurosci* 31:9971–9981.

- Fiebelkorn IC, Snyder AC, Mercier MR, Butler JS, Molholm S, Foxe JJ (2013) Cortical cross-frequency coupling predicts perceptual outcomes. *Neuroimage* 69:126–137.
- Fox MD, Raichle ME (2007) Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nat Rev Neurosci* 8:700–711.
- Freeman ED, Ipser A, Palmbaha A, Paunoiu D, Brown P, Lambert C, Leff A, Driver J (2013) Sight and sound out of synch: Fragmentation and renormalisation of audiovisual integration and subjective timing. *Cortex* 49:2875–2887.
- Fries P (2005) A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn Sci* 9:474–480.
- Fries P (2009) Neuronal gamma-band synchronization as a fundamental process in cortical computation. *Annu Rev Neurosci* 32:209–224.
- Fujisaki W, Nishida S (2005) Temporal frequency characteristics of synchrony–asynchrony discrimination of audio-visual signals. *Exp Brain Res* 166:455–464.
- Fujisaki W, Nishida S (2010) A common perceptual temporal limit of binding synchronous inputs across different sensory attributes and modalities. *Proc Biol Sci* 277:2281–2290.
- Fujisaki W, Shimojo S, Kashino M, Nishida S (2004) Recalibration of audiovisual simultaneity. *Nat Neurosci* 7:773–778.
- Gawne TJ, Kjaer TW, Richmond BJ (1996) Latency: another potential code for feature binding in striate cortex. *J Neurophysiol* 76:1356–1360.
- Gerstein GL, Kiang NY-S (1960) An Approach to the Quantitative Analysis of Electrophysiological Data from Single Neurons. *Biophys J* 1:15–28.
- Ghitza O (2011) Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. *Front Psychol* 2:130.
- Gho M, Varela FJ (1988) A quantitative assessment of the dependency of the visual temporal frame upon the cortical rhythm. *J Physiol (Paris)* 83:95–101.
- Giard MH, Peronnet F (1999) Auditory-Visual Integration during Multimodal Object Recognition in Humans: A Behavioral and Electrophysiological Study. *J Cogn Neurosci* 11:473–490.
- Gibson JJ (1975) *The Study of Time II* (Fraser JT, Lawrence N, eds). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Giraud A-L, Poeppel D (2012) Cortical oscillations and speech processing: emerging computational principles and operations. *Nat Neurosci* 15:511–517.
- Gomez-Ramirez M, Kelly SP, Molholm S, Sehatpour P, Schwartz TH, Foxe JJ (2011) Oscillatory Sensory Selection Mechanisms during Intersensory Attention to Rhythmic Auditory and Visual Inputs: A Human Electroencephalographic Investigation. *J Neurosci* 31:18556–18567.

- Graaf T de, Gross J, Paterson G, Rusch T (2013) Alpha-Band Rhythms in Visual Task Performance: Phase-Locking by Rhythmic Sensory Stimulation. *PLoS One*.
- Grahn JA (2012) Neural mechanisms of rhythm perception: current findings and future perspectives. *Top Cogn Sci* 4:585–606.
- Grahn JA, Rowe JB (2013) Finding and feeling the musical beat: striatal dissociations between detection and prediction of regularity. *Cereb Cortex* 23:913–921.
- Gramfort A, Luessi M, Larson E, Engemann D, Strohmeier D, Brodbeck C, Goj R, Jas M, Brooks T, Parkkonen L, Hämäläinen M (2013) MEG and EEG data analysis with MNE-Python. *Front Neurosci* 7:267.
- Gray C, König P, Engel A, Singer W (1989) Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature* 338:334–337.
- Green DM (1971) Temporal auditory acuity. *Psychol Rev* 78:540–551.
- Gregoriou GG, Gotts SJ, Zhou H, Desimone R (2009) High-frequency, long-range coupling between prefrontal and visual cortex during attention. *Science* 324:1207–1210.
- Gross J, Baillet S, Barnes GR, Henson RN, Hillebrand A, Jensen O, Jerbi K, Litvak V, Maess B, Oostenveld R, Parkkonen L, Taylor JR, van Wassenhove V, Wibrals M, Schoffelen J-M (2013) Good practice for conducting and reporting MEG research. *Neuroimage* 65:349–363.
- Haegens S, Nacher V, Luna R, Romo R, Jensen O (2011) α -Oscillations in the monkey sensorimotor network influence discrimination performance by rhythmical inhibition of neuronal spiking. *Proc Natl Acad Sci U S A* 108:19377–19382.
- Hafting T, Fyhn M, Bonnevie T, Moser M-B, Moser EI (2008) Hippocampus-independent phase precession in entorhinal grid cells. *Nature* 453:1248–1252.
- Hari R (1989) Neuromagnetic steady-state responses to auditory stimuli. *J Acoust Soc Am* 86:1033.
- Harrington DL, Haaland KY, Knight RT (1998) Cortical Networks Underlying Mechanisms of Time Perception. *J Neurosci* 18:1085–1095.
- Harris L, Harrar V, Jaekl P, Kopinska A (2010) Mechanisms of simultaneity constancy. *Sp Time Percept Action*:232–253.
- Healy K, McNally L, Ruxton GD, Cooper N, Jackson AL (2013) Metabolic rate and body size are linked with perception of temporal information. *Anim Behav* 86:685–696.
- Heil P (2004) First-spike latency of auditory neurons revisited. *Curr Opin Neurobiol* 14:461–467.
- Heil P, Irvine DRF (1997) First-Spike Timing of Auditory-Nerve Fibers and Comparison With Auditory Cortex. *J Neurophysiol* 78:2438–2454.
- Henry MJ, Obleser J (2012) Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proc Natl Acad Sci U S A* 109:20095–20100.

- Henze D., Buzsáki G (2001) Action potential threshold of hippocampal pyramidal cells in vivo is increased by recent spiking activity. *Neuroscience* 105:121–130.
- Heron J, Aaen-Stockdale C, Hotchkiss J, Roach NW, McGraw P V, Whitaker D (2012) Duration channels mediate human time perception. *Proc Biol Sci* 279:690–698.
- Heron J, Roach NW, Whitaker D, Hanson JVM (2010) Attention regulates the plasticity of multisensory timing. *Eur J Neurosci* 31:1755–1762.
- Hipp JF, Engel AK, Siegel M (2011) Oscillatory Synchronization in Large-Scale Cortical Networks Predicts Perception. *Neuron* 69:387–396.
- Hirsh IJ, Sherrick Jr. CE (1961) Perceived order in different sense modalities. *J Exp Psychol* 62:423–432.
- Howard MF, Poeppel D (2010) Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. *J Neurophysiol* 104:2500–2511.
- Hsu Y-F, Hämäläinen JA, Waszak F (2013) Temporal expectation and spectral expectation operate in distinct fashion on neuronal populations. *Neuropsychologia* 51:2548–2555.
- Ilhan B, VanRullen R (2012) No counterpart of visual perceptual echoes in the auditory system. *PLoS One* 7:e49287.
- Ipser A, Paunoiu D, Freeman ED (2012) Telling the time with audiovisual speech and non-speech: Does the brain use multiple clocks? *Seeing Perceiving* 25:14–15.
- Ivry R (1996) The representation of temporal information in perception and motor control. *Curr Opin Neurobiol* 6:851–857.
- Ivry RB, Schlerf JE (2008) Dedicated and intrinsic models of time perception. *Trends Cogn Sci* 12:273–280.
- James W (1886) The Perception of Time. *J Specul Philos* 20:374–407.
- Jensen O, Bonnefond M, VanRullen R (2012) An oscillatory mechanism for prioritizing salient unattended stimuli. *Trends Cogn Sci* 16:200–206.
- Jensen O, Lisman JE (1996) Hippocampal CA3 region predicts memory sequences: accounting for the phase precession of place cells. *Learn Mem* 3:279–287.
- Jensen O, Lisman JE (2000) Position Reconstruction From an Ensemble of Hippocampal Place Cells: Contribution of Theta Phase Coding. *J Neurophysiol* 83:2602–2609.
- Jensen O, Mazaheri A (2010) Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Front Hum Neurosci* 4:186.
- Johnston A, Nishida S (2001) Time perception: brain time or event time? *Curr Biol* 11:R427–30.
- Jones MR (1976) Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychol Rev* 83:323.

- Jones MR (1987) Dynamic pattern structure in music: Recent theory and research. *Percept Psychophys* 41:621–634.
- Jones MR, Boltz M (1989) Dynamic attending and responses to time. *Psychol Rev* 96:459.
- Jones MR, Moynihan H, MacKenzie N, Puente J (2002) Temporal Aspects of Stimulus-Driven Attending in Dynamic Arrays. *Psychol Sci* 13:313–319.
- Jones MW, Wilson MA (2005) Phase precession of medial prefrontal cortical activity relative to the hippocampal theta rhythm. *Hippocampus* 15:867–873.
- Joundi R, Jenkinson N, Brittain J, Aziz T, Brown P (2012) Driving oscillatory activity in the human cortex enhances motor performance. *Curr Biol* 22:405–407.
- Kayser C, Ince RAA, Panzeri S (2012) Analysis of slow (theta) oscillations as a potential temporal reference frame for information coding in sensory cortices. *PLoS Comput Biol* 8:e1002717.
- Kayser C, Montemurro MA, Logothetis NK, Panzeri S (2009) Spike-Phase Coding Boosts and Stabilizes Information Carried by Spatial and Temporal Spike Patterns. *Neuron* 61:597–608.
- Keetels M, Vroomen J (2012) Perception of synchrony between the senses. *Front neural basis multisensory Process*:147–177.
- Koch C, Rapp M, Segev I (1996) A Brief History of Time (Constants). *Cereb Cortex* 6:93–101.
- Kopinska A, Harris LR (2004) Simultaneity constancy. *Perception* 33:1049–1060.
- Kraus BJ, Robinson RJ, White JA, Eichenbaum H, Hasselmo ME (2013) Hippocampal “Time Cells”: Time versus Path Integration. *Neuron* 78:1090–1101.
- Kuffler S (1953) Discharge patterns and functional organization of mammalian retina. *J Neurophysiol*.
- Lachaux JP, Rodriguez E, Martinerie J, Varela FJ (1999) Measuring phase synchrony in brain signals. *Hum Brain Mapp* 8:194–208.
- Lakatos P, Karmos G, Mehta AD, Ulbert I, Schroeder CE (2008) Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science* 320:110–113.
- Lakatos P, Shah AS, Knuth KH, Ulbert I, Karmos G, Schroeder CE (2005) An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *J Neurophysiol* 94:1904–1911.
- Large EW, Jones MR (1999) The dynamics of attending: How people track time-varying events. *Psychol Rev* 106:119.
- Laufs H, Krakow K, Sterzer P, Eger E, Beyerle A, Salek-Haddadi A, Kleinschmidt A (2003) Electroencephalographic signatures of attentional and cognitive default modes in spontaneous brain activity fluctuations at rest. *Proc Natl Acad Sci U S A* 100:11053–11058.
- Laughlin SB, Sejnowski TJ (2003) Communication in neuronal networks. *Science* 301:1870–1874.

- Laughlin SB, Weckström M (1993) Fast and slow photoreceptors ? a comparative study of the functional diversity of coding and conductances in the Diptera. *J Comp Physiol A* 172:593–609.
- Lehongre K, Ramus F, Villiermet N, Schwartz D, Giraud A-L (2011) Altered low- γ sampling in auditory cortex accounts for the three main facets of dyslexia. *Neuron* 72:1080–1090.
- Lewis P, Miall R (2003) Distinct systems for automatic and cognitively controlled time measurement: evidence from neuroimaging. *Curr Opin Neurobiol* 13:250–255.
- Lewis P, Miall R (2006) Remembering the time: a continuous clock. *Trends Cogn Sci* 10:401–406.
- Linkenkaer-Hansen K (2004) Prestimulus oscillations enhance psychophysical performance in humans. *J Neurosci* 24:10186–10190.
- Lisman J (2005) The theta/gamma discrete phase code occurring during the hippocampal phase precession may be a more general brain coding scheme. *Hippocampus* 15:913–922.
- Lisman JE, Jensen O (2013) The Theta-Gamma Neural Code. *Neuron* 77:1002–1016.
- Livingstone M, Hubel D (1988) Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science* 240:740–749.
- Love SA, Petrini K, Cheng A, Pollick FE (2013) A psychophysical investigation of differences between synchrony and temporal order judgments. Suzuki S, ed. *PLoS One* 8:e54798.
- Luo H, Liu Z, Poeppel D (2010) Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS Biol* 8:e1000445.
- Luo H, Poeppel D (2012) Cortical oscillations in auditory perception and speech: evidence for two temporal windows in human auditory cortex. *Front Psychol* 3:170.
- MacDonald CJ, Lepage KQ, Eden UT, Eichenbaum H (2011) Hippocampal “Time Cells” Bridge the Gap in Memory for Discontiguous Events. *Neuron* 71:737–749.
- Mainen Z, Sejnowski T (1995) Reliability of spike timing in neocortical neurons. *Science* 268:1503–1506.
- Martin B, Giersch A, Huron C, van Wassenhove V (2013) Temporal event structure and timing in schizophrenia: preserved binding in a longer “now”. *Neuropsychologia* 51:358–371.
- Matell M, Meck W (2004) Cortico-striatal circuits and interval timing: coincidence detection of oscillatory processes. *Cogn Brain Res* 21:139–170.
- Mathewson KE, Fabiani M, Gratton G, Beck DM, Lleras A (2010) Rescuing stimuli from invisibility: Inducing a momentary release from visual masking with pre-target entrainment. *Cognition* 115:186–191.
- Mathewson KE, Gratton G, Fabiani M, Beck DM, Ro T (2009) To see or not to see: prestimulus alpha phase predicts visual awareness. *J Neurosci* 29:2725–2732.

- Mathewson KE, Lleras A, Beck DM, Fabiani M, Ro T, Gratton G (2011) Pulsed out of awareness: EEG alpha oscillations represent a pulsed-inhibition of ongoing cortical processing. *Front Psychol* 2:99.
- Mathewson KE, Prudhomme C, Fabiani M, Beck DM, Lleras A, Gratton G (2012) Making waves in the stream of consciousness: entraining oscillations in EEG alpha and fluctuations in visual awareness with rhythmic visual stimulation. *J Cogn Neurosci* 24:2321–2333.
- Maunsell JHR, Ghose GM, Assad JA, McAdams CJ, Boudreau CE, Noerager BD (1999) Visual response latencies of magnocellular and parvocellular LGN neurons in macaque monkeys. *Vis Neurosci* 16:1–14.
- Mcauley JD (1995) On the Perception of Time as Phase: Toward an Adaptive Oscillator Model of Rhythm.
- McDonald JJ, Teder-Sälejärvi WA, Di Russo F, Hillyard SA (2005) Neural basis of auditory-induced shifts in visual time-order perception. *Nat Neurosci* 8:1197–1202.
- Meck W (1996) Neuropharmacology of timing and time perception. *Cogn Brain Res* 3:227–242.
- Meredith M, Nemitz J, Stein B (1987) Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *J Neurosci* 7:3215–3229.
- Michon J (1985) The complete time experimenter. *Time, mind, Behav*:20–52.
- Michon J (1990) Implicit and explicit representations of time. *Cogn Model Psychol time*:37–58.
- Mickus T, Jung H, Spruston N (1999) Properties of slow, cumulative sodium channel inactivation in rat hippocampal CA1 pyramidal neurons. *Biophys J* 76:846–860.
- Miyazaki M, Yamamoto S, Uchida S, Kitazawa S (2006) Bayesian calibration of simultaneity in tactile temporal order judgment. *Nat Neurosci* 9:875–877.
- Montemurro MA, Rasch MJ, Murayama Y, Logothetis NK, Panzeri S (2008) Phase-of-Firing Coding of Natural Visual Stimuli in Primary Visual Cortex. *Curr Biol* 18:375–380.
- Monto S, Palva S, Voipio J, Palva JM (2008) Very slow EEG fluctuations predict the dynamics of stimulus detection and oscillation amplitudes in humans. *J Neurosci* 28:8268–8272.
- Monyer H, Sprengel R, Schoepfer R, Herb A (1992) Heteromeric NMDA receptors: molecular and functional distinction of subtypes. *Science* 256:1217–1221.
- Morillon B, Kell CA, Giraud A-L (2009) Three stages and four neural systems in time estimation. *J Neurosci* 29:14803–14811.
- Moutoussis K, Zeki S (1997) A direct demonstration of perceptual asynchrony in vision. *Proc Biol Sci* 264:393–399.
- Nadasdy Z (2010) Binding by asynchrony: the neuronal phase code. *Front Neurosci* 4.
- Nakanishi S (1994) Metabotropic glutamate receptors: synaptic transmission, modulation, and plasticity. *Neuron* 13:1031–1037.

- Neuling T, Rach S, Wagner S, Wolters CH, Herrmann CS (2012a) Good vibrations: Oscillatory phase shapes perception. *Neuroimage* 63:771–778.
- Neuling T, Rach S, Wagner S, Wolters CH, Herrmann CS (2012b) Good vibrations: Oscillatory phase shapes perception. *Neuroimage* 63:771–778.
- Ng B, Logothetis N, Kayser C (2013) EEG phase patterns reflect the selectivity of neural firing. *Cereb Cortex* 23:389–398.
- Ng BSW, Schroeder T, Kayser C (2012) A precluding but not ensuring role of entrained low-frequency oscillations for auditory perception. *J Neurosci* 32:12268–12276.
- Nijhawan R (2002) Neural delays, visual motion and the flash-lag effect. *Trends Cogn Sci* 6:387–393.
- Nobre A, Correa A, Coull J (2007) The hazards of time. *Curr Opin Neurobiol* 17:465–470.
- Nobre AC (2001) Orienting attention to instants in time. *Neuropsychologia* 39:1317–1328.
- Noesselt T, Rieger JW, Schoenfeld MA, Kanowski M, Hinrichs H, Heinze H-J, Driver J (2007) Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. *J Neurosci* 27:11431–11441.
- Nozaradan S, Peretz I, Missal M, Mouraux A (2011) Tagging the neuronal entrainment to beat and meter. *J Neurosci* 31:10234–10240.
- Nozaradan S, Peretz I, Mouraux A (2012) Selective neuronal entrainment to the beat and meter embedded in a musical rhythm. *J Neurosci* 32:17572–17581.
- O’Keefe J, Recce ML (1993) Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus* 3:317–330.
- Olson I, Chun M (2001) Temporal contextual cuing of visual attention. *J Exp Psychol Learn Mem Cogn* 27:1299.
- Palva S, Palva J (2011) Functional roles of alpha-band phase synchronization in local and large-scale cortical networks. *Front Psychol* 2.
- Panzeri S, Brunel N, Logothetis NK, Kayser C (2010) Sensory neural codes using multiplexed temporal scales. *Trends Neurosci* 33:111–120.
- Panzeri S, Petersen RS, Schultz SR, Lebedev M, Diamond ME (2001) The Role of Spike Timing in the Coding of Stimulus Location in Rat Somatosensory Cortex. *Neuron* 29:769–777.
- Pastalkova E, Itskov V, Amarasingham A, Buzsáki G (2008) Internally generated cell assembly sequences in the rat hippocampus. *Science* 321:1322–1327.
- Patel SS (2000) Flash-Lag Effect: Differential Latency, Not Postdiction. *Science* 290:1051a–1051.
- Plomp R (1964) Rate of Decay of Auditory Sensation. *J Acoust Soc Am* 36:277.
- Poeppel D (2003) The analysis of speech in different temporal integration windows: cerebral lateralization as “asymmetric sampling in time.” *Speech Commun* 41:245–255.

- Pöppel E (1997) A hierarchical model of temporal perception. *Trends Cogn Sci* 1:56–61.
- Pöppel E (2009) Pre-semantically defined temporal windows for cognitive processing. *Philos Trans R Soc Lond B Biol Sci* 364:1887–1896.
- Pöppel E, Schill K, Steinbüchel N (1990) Sensory integration within temporally neutral systems states: A hypothesis. *Naturwissenschaften* 77:89–91.
- Praamstra P, Turgeon M, Hesse CW, Wing a M, Perryer L (2003) Neurophysiological correlates of error correction in sensorimotor-synchronization. *Neuroimage* 20:1283–1297.
- Purushothaman G, Patel SS, Bedell HE, Ogmen H (1998) Moving ahead through differential visual latency. *Nature* 396:424.
- Purves D (1996) The wagon wheel illusion in movies and reality. *Proc Natl Acad Sci U S A* 93:3693–3697.
- Raiguel SE, Xiao D-K, Marcar VL, Orban GA (1999) Response Latency of Macaque Area MT/V5 Neurons and Its Relationship to Stimulus Parameters. *J Neurophysiol* 82:1944–1956.
- Raij T, Ahveninen J, Lin F (2010) Onset timing of cross-sensory activations and multisensory interactions in auditory and visual sensory cortices. *Eur J Neurosci* 31:1772–1782.
- Rees A, Green GGR, Kay RH (1986) Steady-state evoked responses to sinusoidally amplitude-modulated sounds recorded in man. *Hear Res* 23:123–133.
- Regan D (1966) Some characteristics of average steady-state and transient responses evoked by modulated light. *Electroencephalogr Clin Neurophysiol* 20:238–248.
- Repp BH (2008) Multiple temporal references in sensorimotor synchronization with metrical auditory sequences. *Psychol Res* 72:79–98.
- Roach NW, Heron J, Whitaker D, McGraw P V (2011) Asynchrony adaptation reveals neural population code for audio-visual timing. *Proc Biol Sci* 278:1314–1322.
- Roache R (1999) Mellor and Dennett on the perception of temporal order. *Philos Q*.
- Roberts TP, Ferrari P, Stufflebeam SM, Poeppel D (2000) Latency of the auditory evoked neuromagnetic field components: stimulus dependence and insights toward perception. *J Clin Neurophysiol* 17:114–129.
- Rodiek R (1998) *Retinal Organization. First Steps Seeing* Sinauer Assoc:195–209.
- Rodriguez E, George N, Lachaux J (1999) Perception's shadow: long-distance synchronization of human brain activity. *Nature* 397:430–433.
- Roelfsema P, Engel A, Konig P, Singer W (1997) Visuomotor integration is associated with zero time-lag synchronization among cortical areas. *Nature* 385:157–161.
- Romei V, Gross J, Thut G (2012) Sounds reset rhythms of visual cortex and corresponding human visual perception. *Curr Biol* 22:807–813.

- Romei V, Murray MM, Cappe C, Thut G (2009) Preperceptual and stimulus-selective enhancement of low-level human visual cortex excitability by sounds. *Curr Biol* 19:1799–1805.
- Roopun AK, Kramer MA, Carracedo LM, Kaiser M, Davies CH, Traub RD, Kopell NJ, Whittington MA (2008) Temporal Interactions between Cortical Rhythms. *Front Neurosci* 2:145–154.
- Roseboom W, Arnold DH (2011) Twice upon a time: multiple concurrent temporal recalibrations of audiovisual speech. *Psychol Sci* 22:872–877.
- Russell IJ, Sellick PM (1978) Intracellular studies of hair cells in the mammalian cochlea. *J Physiol* 284:261–290.
- Sakurai Y (1999) How do cell assemblies encode information in the brain? *Neurosci Biobehav Rev* 23:785–796.
- Sato M, Schwartz J-L, Abry C, Cathiard M-A, Loevenbruck H (2006) Multistable syllables as enacted percepts: a source of an asymmetric bias in the verbal transformation effect. *Percept Psychophys* 68:458–474.
- Scharnowski F, Rees G, Walsh V (2013) Time and the brain: neurorelativity: The chronoarchitecture of the brain from the neuronal rather than the observer's perspective. *Trends Cogn Sci* 17:51–52.
- Schmolesky MT, Wang Y, Hanes DP, Thompson KG, Leutgeb S, Schall JD, Leventhal AG (1998) Signal Timing Across the Macaque Visual System. *J Neurophysiol* 79:3272–3278.
- Schroeder CE, Lakatos P (2009) Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci* 32:9–18.
- Schroeder CE, Lakatos P, Kajikawa Y, Partan S, Puce A (2008) Neuronal oscillations and visual amplification of speech. *Trends Cogn Sci* 12:106–113.
- Schwartz M, Kotz SA (2013) A dual-pathway neural architecture for specific temporal prediction. *Neurosci Biobehav Rev* 37:2587–2596.
- Sejnowski TJ, Paulsen O (2006) Network oscillations: emerging computational principles. *J Neurosci* 26:1673–1676.
- Senkowski D, Schneider TR, Foxe JJ, Engel AK (2008) Crossmodal binding through neural coherence: implications for multisensory processing. *Trends Neurosci* 31:401–409.
- Shimojo S (2001) Sensory modalities are not separate modalities: plasticity and interactions. *Curr Opin Neurobiol* 11:505–509.
- Siegel M, Donner TH, Engel AK (2012) Spectral fingerprints of large-scale neuronal interactions. *Nat Rev Neurosci* 13:121–134.
- Sirota A, Montgomery S, Fujisawa S, Isomura Y, Zugaro M, Buzsáki G (2008) Entrainment of Neocortical Neurons and Gamma Oscillations by the Hippocampal Theta Rhythm. *Neuron* 60:683–697.

- Skaggs WE, McNaughton BL, Wilson MA, Barnes CA (1996) Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus* 6:149–172.
- Slutsky DA, Recanzone GH (2001) Temporal and spatial dependency of the ventriloquism effect. *Neuroreport* 12:7–10.
- Sokolik R, VanRullen R (2013) The Flickering Wheel Illusion: When α Rhythms Make a Static Wheel Flicker. *J Neurosci* 33:13498–13504.
- Spence C, Parise C (2010) Prior-entry: A review. *Conscious Cogn* 19:364–379.
- St. Augustine (1900) *The Confessions of Saint Augustine*. EP Dutt.
- Stefanics G, Hangya B, Hernádi I, Winkler I, Lakatos P, Ulbert I (2010a) Phase entrainment of human delta oscillations can mediate the effects of expectation on reaction speed. *J Neurosci* 30:13578–13585.
- Stefanics G, Hangya B, Hernádi I, Winkler I, Lakatos P, Ulbert I (2010b) Phase entrainment of human delta oscillations can mediate the effects of expectation on reaction speed. *J Neurosci* 30:13578–13585.
- Stein BE, Meredith MA (1993) *The merging of the senses*. Cognitive neuroscience. The MIT Press.
- Stone J V, Hunkin NM, Porrill J, Wood R, Keeler V, Beanland M, Port M, Porter NR (2001) When is now? Perception of simultaneity. *Proc Biol Sci* 268:31–38.
- Stroud JM (1967) The fine structure of psychological time. *Ann N Y Acad Sci* 138:623–631.
- Sugita Y, Suzuki Y (2003) Implicit estimation of sound-arrival time. *Nature* 421.
- Tallon-Baudry C, Bertrand O (1999) Oscillatory gamma activity in humans and its role in object representation. *Trends Cogn Sci* 3:151–162.
- Tallon-Baudry C, Bertrand O, Delpuech C, Pernier J (1996) Stimulus Specificity of Phase-Locked and Non-Phase-Locked 40 Hz Visual Responses in Human. *J Neurosci* 16:4240–4249.
- Talsma D, Senkowski D, Soto-Faraco S, Woldorff MG (2010) The multifaceted interplay between attention and multisensory integration. *Trends Cogn Sci* 14:400–410.
- Tepas DI, Armington JC (1962) Properties of evoked visual potentials. *Vision Res* 2:449–461.
- Theunissen F, Miller JP (1995) Temporal encoding in nervous systems: a rigorous definition. *J Comput Neurosci* 2:149–162.
- Thorne JD, De Vos M, Viola FC, Debener S (2011) Cross-modal phase reset predicts auditory task performance in humans. *J Neurosci* 31:3853–3861.
- Thorne JD, Debener S (2014) Look now and hear what's coming: On the functional role of cross-modal phase reset. *Hear Res* 307:144–152.
- Titchener E (1908) *Lectures on the elementary psychology of feeling and attention*. Macmillan.

- Tognoli E, Kelso J a S (2009) Brain coordination dynamics: true and false faces of phase synchrony and metastability. *Prog Neurobiol* 87:31–40.
- Treisman A (1996) The binding problem. *Curr Opin Neurobiol* 6:171–178.
- Treisman M (1963) Temporal discrimination and the indifference interval: Implications for a model of the “internal clock”. *Psychol Monogr Gen Appl* 77:1.
- Treisman M (1984) Temporal Rhythms and Cerebral Rhythms. *Ann N Y Acad Sci* 423:542–565.
- Treisman M, Faulkner A, Naish PL, Brogan D (1990) The internal clock: Evidence for a temporal oscillator underlying time perception with some estimates of its characteristic frequency. *Perception* 19:705–743.
- Treisman M, Faulkner A, Naish PLN (1992) On the Relation Between Time Perception and the Timing of Motor Action: Evidence for a Temporal Oscillator Controlling the Timing of Movement. *Q J Exp Psychol Sect A* 45:235–263.
- Treisman M, Howarth C (1959) Changes in threshold level produced by a signal preceding or following the threshold stimulus. *Q J Exp Psychol* 11:129–142.
- Tyler CW (1985) Analysis of visual modulation sensitivity. II. Peripheral retina and the role of photoreceptor dimensions. *J Opt Soc Am A* 2:393.
- Van de Grind W (2002) Physical, Neural, and Mental Timing. *Conscious Cogn* 11:241–264.
- Van der Burg E, Olivers CNL, Bronkhorst AW, Theeuwes J (2008) Pip and pop: Nonspatial auditory signals improve spatial visual search. *J Exp Psychol Hum Percept Perform* 34:1053.
- Van der Burg E, Talsma D, Olivers CN, Hickey C, Theeuwes J (2011) Early multisensory interactions affect the competition among multiple visual objects. *Neuroimage* 55:1208–1218.
- Van Dijk H, Schoffelen J-M, Oostenveld R, Jensen O (2008) Prestimulus oscillatory activity in the alpha band predicts visual discrimination ability. *J Neurosci* 28:1816–1823.
- Van Eijk RLJ, Kolhtrausch A, Juola JF, Van de Par S (2008) Audiovisual synchrony and temporal order judgments: Effects of experimental method and stimulus type. *Percept Psychophys* 70:955–968.
- Van Wassenhove V (2009) Minding time in an amodal representational space. *Philos Trans R Soc Lond B Biol Sci* 364:1815–1830.
- Van Wassenhove V (2013) Speech through ears and eyes: interfacing the senses with the supramodal brain. *Front Psychol* 4:388.
- Van Wassenhove V, Grant K, Poeppel D (2007) Temporal window of integration in auditory-visual speech perception. *Neuropsychologia* 45:598–607.
- Van Wassenhove V, Grant KW, Poeppel D (2005) Visual speech speeds up the neural processing of auditory speech. *Proc Natl Acad Sci U S A* 102:1181–1186.

- Van Wassenhove V, Nagarajan SS (2007) Auditory cortical plasticity in learning to discriminate modulation rate. *J Neurosci* 27:2663–2672.
- VanRullen R, Guyonneau R, Thorpe SJ (2005a) Spike times make sense. *Trends Neurosci* 28:1–4.
- VanRullen R, Koch C (2003) Is perception discrete or continuous? *Trends Cogn Sci* 7:207–213.
- VanRullen R, Macdonald J (2012) Perceptual echoes at 10 Hz in the human brain. *Curr Biol*.
- VanRullen R, Reddy L, Koch C (2005b) Attention-dependent discrete sampling of motion perception. *J Vis* 5:237–237.
- VanRullen R, Reddy L, Koch C (2006) The continuous wagon wheel illusion is associated with changes in electroencephalogram power at approximately 13 Hz. *J Neurosci* 26:502–507.
- Varela F, Lachaux J (2001) The brainweb: phase synchronization and large-scale integration. *Nat Rev Neurosci* 2:229–239.
- Varela F, Toro A, John ER, Schwartz E (1981) Perceptual framing and cortical alpha rhythm. *Neuropsychologia* 19:675–686.
- Vatakis A, Navarra J, Soto-Faraco S, Spence C (2008) Audiovisual temporal adaptation of speech: temporal order versus simultaneity judgments. *Exp brain Res* 185:521–529.
- Verstraten F, Cavanagh P, Labianca A (2000) Limits of attentive tracking reveal temporal properties of attention. *Vision Res* 40:3651–3664.
- Vibell J, Klinge C, Zampini M, Spence C, Nobre AC (2007) Temporal order is coded temporally in the brain: early event-related potential latency shifts underlying prior entry in a cross-modal temporal order judgment task. *J Cogn Neurosci* 19:109–120.
- Victor JD (2000) How the brain uses time to represent and process visual information. Published on the World Wide Web on 16 August 2000. *Brain Res* 886:33–46.
- Von Stein A, Sarnthein J (2000) Different frequencies for different scales of cortical integration: from local gamma to long range alpha/theta synchronization. *Int J Psychophysiol* 38:301–313.
- Vroomen J, Keetels M (2010) Perception of intersensory synchrony: a tutorial review. *Atten Percept Psychophys* 72:871–884.
- Vroomen J, Keetels M, de Gelder B, Bertelson P (2004) Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Brain Res Cogn Brain Res* 22:32–35.
- Wang X (2010) Neurophysiological and computational principles of cortical rhythms in cognition. *Phys Rev* 90:1195–1268.
- Whitney D, Murakami I (1998) Latency difference, not spatial extrapolation. *Nat Neurosci* 1:656–657.
- Wickelgren W (1999) Webs, cell assemblies, and chunking in neural nets: Introduction. *Can J Exp Psychol Can Psychol expérimentale* 53:118.

- Wittmann M (2009) The inner experience of time. *Philos Trans R Soc Lond B Biol Sci* 364:1955–1967.
- Wittmann M (2013) The inner sense of time: how the brain creates a representation of duration. *Nat Rev Neurosci* 14:217–223.
- Wittmann M, Paulus M (2008) Decision making, impulsivity and time perception. *Trends Cogn Sci* 12:7–12.
- Womelsdorf T, Fries P, Mitra PP, Desimone R (2006) Gamma-band synchronization in visual cortex predicts speed of change detection. *Nature* 439:733–736.
- Womelsdorf T, Schoffelen J-M, Oostenveld R, Singer W, Desimone R, Engel AK, Fries P (2007) Modulation of neuronal interactions through neuronal synchronization. *Science* 316:1609–1612.
- Yamamoto S, Miyazaki M, Iwano T, Kitazawa S (2012) Bayesian calibration of simultaneity in audiovisual temporal order judgments. *PLoS One* 7:e40379.
- Zampini M, Guest S, Shore DI, Spence C (2005) Audio-visual simultaneity judgments. *Percept Psychophys* 67:531–544.
- Zeki S (2001) Localization and globalization in conscious vision. *Annu Rev Neurosci* 24:57–86.
- Zeki S, Bartels A (1998) The asynchrony of consciousness. *Proc Biol Sci* 265:1583–1585.
- Zeki SM (1978) Uniformity and diversity of structure and function in rhesus monkey prestriate visual cortex. *J Physiol* 277:273–290.
- Zion Golumbic EM, Ding N, Bickel S, Lakatos P, Schevon CA, McKhann GM, Goodman RR, Emerson R, Mehta AD, Simon JZ, Poeppel D, Schroeder CE (2013) Mechanisms Underlying Selective Neuronal Tracking of Attended Speech at a “Cocktail Party.” *Neuron* 77:980–991.