Phase entrainment and perceptual cycles in audition and vision

Benedikt Zoefel

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

Présentée et soutenue par :
Benedikt ZOEFEL

Le mardi 8 décembre 2015

Titre :

Phase entrainment and perceptual cycles in audition and vision
(Entraînement de phase et cycles perceptifs dans l'audition et la vision)

ED CLESCO : Neurosciences, comportement et cognition

Unité de recherche :

Centre de Recherche Cerveau et Cognition, CNRS UMR 5549

Directeur(s) de Thèse :

Rufin VANRULLEN

Rapporteurs :

Anne-Lise GIRAUD, Department of Neuroscience, University of Geneva
Ole JENSEN, Donders Institute for Brain, Cognition, and Behavior, Nijmegen

Autre(s) membre(s) du jury :

Peter LAKATOS, Nathan Kline Institute for Psychiatric Research, Orangeburg, NY

ขอบคุณครับเพชรน้อยของผม เธอเป็นคนที่สำคัญที่สุดในชีวิตของผม
ผมเป็นของเธอตลอดไป

ACKNOWLEDGEMENTS

Thank you *Rufin*, for competently and perseveringly answering all my questions. I have learned so much under your guidance. I could not imagine a more capable and qualified supervisor.

Thank you *Peter*, I could not have been happier being able to stay in your lab and learning all these things from you. I feel honored to know such a great researcher and such a great person.

Thank you *Doug*, *Marina*, *Biao*, *Jaya* and *Edd*, for so many shared moments, shared thoughts, and shared friendship.

Thank you *Diego*, for being such a pleasant, relaxed, and entertaining officemate.

Thank you *Manu*, for extremely component translations into French, and for joining the lab *before* I left it.

Thank you *Marie*, *Damien*, *Sasskia*, *Christelle*, *Rasa*, *Séb*, *Grace*, *Samy*, *Anne-Claire*, *Céline*, *Laia*, *Mehdi*, *Tracy*, *Jake*, *Lola*, *Adri*, *Felipe*, *Marlene* and *Gab*, for making the CerCo such a lively, happy place.

Thank you *Noelle* and *Annie*, for sharing office at the NKI. Without you my time there would not have been the same.

Thank you *Jordi*, for so many discussions, for shared work, and for being a friend. Soon we'll meet again.

Thank you *Emmanuelle*, *Arthur*, *Natalia* and *Arnaud*, for your friendship, for help with work, and for many rounds of Settlers.

Merci *Claire* et *Zoé*, pour une organisation du labo qui est extrêmement compétente et agréable en même temps.

Grazie *Marcello*, non dimenticherò mai le nostre conversazioni italiane.

Danke *Rodi*, für viel gute Laune und Unterhaltung im Büro und überhaupt, dass wir uns so gut verstehen. Bald sehen wir uns wieder: Das Leben scheint uns ja immer in dieselben Ecken zu führen...

Danke liebe *Eltern*! Für eine wunderschöne, Wissensdurst weckende Kindheit (so etwas wie das hier wird dann eben daraus) und für großartige Unterstützung im darauf folgenden Leben.

Danke *Mama, Papa*, *Molch*, *Rainer*, *Friederike*, *Pauline*, *Frederik*, *Hannah*, *Sophia* und *Peter*, einfach nur, dass wir alle zusammen eine Familie sind.

Danke *Eugen*, für deine Unterstützung, ohne die ich das hier (und alles andere auch) niemals hinbekommen hätte.

ขอบคุณครอบครัวจากประเทศไทยที่รับผมเข้าไปอยู่ในครอบครัว
ผมภูมิใจที่ได้เป็นสมาชิกของครอบครัวนี้ครับ

ABSTRACT

Phase entrainment and perceptual cycles in audition and vision

Recent research indicates fundamental differences between the auditory and visual systems: Whereas the visual system seems to sample its environment, cycling between "snapshots" at discrete moments in time (creating *perceptual cycles*), most attempts at discovering discrete perception in the auditory system failed. Here, we show in two psychophysical experiments that subsampling the very input to the visual and auditory systems is indeed more disruptive for audition; however, the existence of perceptual cycles in the auditory system is possible if they operate on a relatively *high level* of auditory processing. Moreover, we suggest that the auditory system, due to the rapidly fluctuating nature of its input, might rely to a particularly strong degree on *phase entrainment*, the alignment between neural activity and the rhythmic structure of its input: By using the low and high excitability phases of *neural oscillations*, the auditory system might *actively control* the timing of its "snapshots" and thereby amplify relevant information whereas irrelevant events are suppressed. Not only do our results suggest that the oscillatory phase has important consequences on how simultaneous auditory inputs are perceived; additionally, we can show that phase entrainment to speech sound does entail an active high-level mechanism. We do so by using specifically constructed speech/noise sounds in which fluctuations in low-level features (amplitude and spectral content) of speech have been removed, but intelligibility and high-level features (including, but not restricted to phonetic information) have been conserved. We demonstrate, in several experiments, that the auditory system can entrain to these stimuli, as both perception (the detection of a click embedded in the speech/noise stimuli) and neural oscillations (measured with electroencephalography, EEG, and in intracranial recordings in primary auditory cortex of the monkey) follow the conserved "high-level" rhythm of speech. Taken together, the results presented here suggest that, not only in vision, but also in audition, neural oscillations are an important tool for the discretization and processing of the brain's input. However, there seem to be fundamental differences between the two systems: In contrast to the visual system, it is critical for the auditory system to adapt (via *phase entrainment*) to its environment, and input subsampling is done most likely on a *hierarchically high level* of stimulus processing.

Keywords: neural oscillations – perceptual cycles – phase – entrainment – audition – vision – electroencephalography – current-source density – subsampling – speech – noise – high-level – intelligibility

RESUME

Entraînement de phase et cycles perceptifs dans l'audition et la vision

Des travaux récents indiquent qu'il existe des différences fondamentales entre les systèmes visuel et auditif: tandis que le premier semble échantillonner le flux d'information en provenance de l'environnement, en passant d'un "instantané" à un autre (créant ainsi des *cycles perceptifs*), la plupart des expériences destinées à examiner ce phénomène de discrétisation dans le système auditif ont mené à des résultats mitigés. Dans cette thèse, au travers de deux expériences de psychophysique, nous montrons que le sous-échantillonnage de l'information à l'entrée des systèmes perceptifs est en effet plus destructif pour l'audition que pour la vision. Cependant, nous révélons que des cycles perceptifs dans le système auditif pourraient exister à *un niveau élevé* du traitement de l'information. En outre, nos résultats suggèrent que du fait des fluctuations rapides du flot des sons en provenance de l'environnement, le système auditif tend à avoir son activité alignée sur la structure rythmique de ce flux. En synchronisant la phase des oscillations neuronales, elles-mêmes correspondant à différents états d'excitabilité, le système auditif pourrait *optimiser activement* le moment d'arrivée de ses "instantanés" et ainsi favoriser le traitement des informations pertinentes par rapport aux événements de moindre importance. Non seulement nos résultats montrent que cet *entrainement de la phase* des oscillations neuronales a des conséquences importantes sur la façon dont sont perçus deux flux auditifs présentés simultanément ; mais de plus, ils démontrent que l'*entraînement de phase* par un flux langagier inclut des mécanismes de haut niveau. Dans ce but, nous avons créé des stimuli *parole/bruit* dans lesquels les fluctuations de l'amplitude et du contenu spectral de la parole ont été enlevés, tout en conservant l'information phonétique et l'intelligibilité. Leur utilisation nous a permis de démontrer, au travers de plusieurs expériences, que le système auditif se synchronise à ces stimuli. Plus précisément, la perception, estimée par la détection d'un clic intégré dans les stimuli *parole/bruit*, et les oscillations neuronales, mesurées par Electroencéphalographie chez l'humain et à l'aide d'enregistrements intracrâniens dans le cortex auditif chez le singe, suivent la rythmique "de haut niveau" liée à la parole. En résumé, les résultats présentés ici suggèrent que les oscillations neuronales sont un mécanisme important pour la discrétisation des informations en provenance de l'environnement en vue de leur traitement par le cerveau, non seulement dans la vision, mais aussi dans l'audition. Pourtant, il semble exister des différences fondamentales entre les deux systèmes: contrairement au système visuel, il est essentiel pour le système auditif de se synchroniser (par *entraînement de phase*) à son environnement, avec un échantillonnage du flux des informations vraisemblablement réalisé *à un niveau hiérarchique élevé*.

Mots clés: oscillations neuronales – cycles perceptifs – phase – entraînement – audition – vision – electroencéphalographie – current-source density – sous-échantillonnage – parole – bruit – haut niveau – intelligibilité

RESUME SUBSTANTIEL

Il est possible que la perception humaine ne fonctionne pas de façon continue, mais plutôt de manière *discrète*, à la manière d'une caméra vidéo. De nombreuses études suggèrent que le système visuel échantillonne son environnement, en extrayant des "instantanés" qui correspondent à des moments distincts dans le temps, créant ainsi des moments optimums et d'autres défavorables pour le traitement de l'input visuel (par exemple, Busch et al., 2009; VanRullen and Macdonald, 2012). Curieusement, les tentatives pour mettre en évidence une discrétisation perceptive par le système auditif se sont révélées infructueuses (İlhan and VanRullen, 2012; Zoefel and Heil, 2013). Ceci pourrait refléter une différence cruciale entre le système visuel et le système auditif : en raison des constantes fluctuations temporelles du flot auditif, un sous-échantillonnage brut pourrait se révéler déstructurant, dû à la perte d'informations essentielles. Dans une première expérience de psychophysique, nous avons sous-échantillonné temporellement des stimuli auditifs (des extraits auditifs de discours) et des stimuli visuels (des vidéos de discours en langue des signes) et nous avons testé l'impact de la fréquence de sous-échantillonnage sur les performances de reconnaissance auprès de sujets humains. Ainsi, nous avons pu montrer que la discrétisation des stimuli à l'entrée des systèmes visuel et auditif a un effet plus perturbateur pour le système auditif (VanRullen et al., 2014). En principe, ce résultat pourrait indiquer qu'un traitement discontinu mène à une perte d'informations trop importante pour le système auditif empêchant une correcte extraction des caractéristiques pertinentes pour la compréhension. Il existe cependant une alternative: le sous-échantillonnage auditif pourrait être réalisé à un *niveau hiérarchiquement supérieur* du traitement de l'information, c'est à dire après que l'extraction de certaines caractéristiques auditives soit achevée. Pour tester cette hypothèse, nous avons construit des

stimuli langagier ayant été sous-échantillonnés à des fréquences différentes, soit directement dans le domaine temporel (directement sur l'onde d'entrée) ou après extraction de caractéristiques auditives (obtenu par un vocodeur utilisant un codage prédictif linéaire). Dans une deuxième expérience psychophysique, nous avons montré que la reconnaissance auditive est plus résistante au sous-échantillonnage réalisé à un niveau élevé de traitement de l'information auditive en comparaison d'un sous-échantillonnage réalisé dans le domaine temporel (Zoefel et al., 2015). Bien que ces résultats ne prouvent pas que le système auditif procède à une discrétisation du flux auditif, ils (1) montrent qu'un sous-échantillonnage est possible dans une certaine mesure sans pertes majeures et (2) suggèrent que, s'il existe une discrétisation, elle devrait être opérée à un niveau relativement élevé du traitement auditif.

Il existe aussi une seconde possibilité, qui permettrait le sous-échantillonnage temporel dans le domaine auditif, et qui n'est pas mutuellement exclusive de la première: les effets délétères du sous-échantillonnage pourraient être réduits activement si le système auditif pouvait *décider* du moment où les «instantanés» de l'environnement sont extraits. Basé sur des recherches récentes, un mécanisme physiologique important, les oscillations neuronales, pourrait s'avérer crucial pour cette alternative. Les oscillations neuronales reflètent des changements cycliques dans l'excitabilité des groupes de neurones (Buzsaki and Draguhn, 2004). Comme des études précédentes l'ont montré (Schroeder and Lakatos, 2009; Schroeder et al., 2010), en jouant sur les phases d'excitabilité de ses oscillations, le cerveau peut *contrôler activement* quelle partie de l'information entrante est amplifiée (l'information coïncidant avec une phase de plus grande excitabilité) et quelle partie est négligée (l'information coïncidant avec une phase de faible excitabilité). Ce phénomène, la synchronisation d'un système oscillant avec un système externe, a été appelé *l'entraînement de phase*. Dans une expérience d'enregistrement électrophysiologique du cortex auditif

primaire chez le singe, nous avons confirmé que dans une scène auditive ambiguë, la phase des oscillations neuronales influence si deux inputs auditifs simultanés sont regroupés en une seule entité ou séparés en deux séquences distinctes. Ce résultat indique que la phase des oscillations neuronales a des conséquences sur le traitement et la perception des stimuli, et, que l'alignement actif de cette phase serait un outil important pour contrôler ou filtrer l'entrée des stimuli et leur traitement ultérieur. En outre, il a été suggéré que cet alignement (c'est-à-dire l'*entraînement de phase*) entre la parole et les oscillations neuronales - sans doute le stimulus rythmique qui est le plus important dans l'environnement auditif humain - pourrait améliorer la compréhension de la parole (Luo and Poeppel, 2007). Toutefois, la parole est une construction complexe, et il n'est toujours pas clair à quelles caractéristiques de la parole le cerveau s'ajuste effectivement. Par exemple, dans un discours standard, l'amplitude du son et de son contenu spectral (la distribution de l'énergie dans les différentes fréquences), ci-après définis comme des caractéristiques de bas niveau, fluctuent rythmiquement, et pourraient être "suivis" par le cerveau. Ainsi, lorsque l'amplitude du signal est large ou lorsqu'il a un contenu spectral "riche", le discours sera perçu clairement et les caractéristiques de haut niveau (notamment la quantité d'informations que l'auditeur peut extraire, c'est-à-dire "l'information phonétique") seront élevées. En revanche, lors de silence entre les mots ou des syllabes, l'amplitude du son, mais aussi l'information phonétique est faible. En d'autres termes, les variations de l'amplitude du son ou de son contenu spectral sont corrélées avec les changements de l'information phonétique et le cerveau pourrait s'y adapter. Toutefois, si l'on pouvait montrer que l'entraînement de phase à la parole persiste même en l'absence de fluctuations systématiques de l'amplitude et du contenu spectral, cela indiquerait un processus plus "élaboré" du cerveau: un "suivi" de l'information phonétique elle-même. Dans ce but, nous avons développé des stimuli *parole/bruit*, sans fluctuations systématiques de

l'amplitude du son et du contenu spectral, et qui conservent suffisamment les caractéristiques de la parole pour permettre la compréhension. En utilisant ces stimuli, nous avons montré, dans plusieurs expériences, que le cerveau peut en effet s'adapter à ces stimuli *parole/bruit*. Dans une expérience psychophysique (Zoefel and VanRullen, 2015), des clics ont été intégrés dans les stimuli *parole/bruit* et les sujets devaient appuyer sur un bouton à chaque fois qu'ils les détectaient. Singulièrement, la détection de ces clics était modulée par le "rythme de haut niveau" (la discrétisation inhérente au langage) ayant été conservé dans les stimuli parole/bruit, ceci indique que la perception s'aligne aux aspects de haut niveau de la parole. Ce résultat était lié aux caractéristiques linguistiques de la parole, dans la mesure où il a été aboli lorsque les stimuli étaient présentés à l'envers. Dans une expérience utilisant l'électroencéphalographie (EEG) avec des sujets humains (Zoefel and VanRullen, in press), nous avons pu montrer que les oscillations neuronales (mesurées à l'aide de l'EEG) s'alignent non seulement aux stimuli classiques mais aussi à nos stimuli qui ne conservent que les caractéristiques "de haut niveau", indiquant que l'entraînement de phase neuronal est à la base des effets perceptifs mesurés précédemment. Aussi il est intéressant de noter que les oscillations neuronales se synchronisent aux caractéristiques de haut niveau de la parole même en l'absence d'informations linguistiques (lorsque les stimuli étaient présentés à l'envers), ce qui indique que la compréhension de la parole est impliquée dans un processus sophistiqué à l'interface entre l'entraînement neuronal et le comportement. Dans une troisième expérience, les stimuli *parole/bruit* et des stimuli parole simple étaient présentés pendant des enregistrements intracrâniens dans le cortex auditif du singe. Comme dans l'expérience EEG, les deux types de stimuli utilisés induisaient un *entraînement de phase*, qui était de plus ici couplé à une modulation de l'activité neuronale dans les hautes fréquences. Cependant pour les stimuli parole/bruit (pour lesquels les fluctuations de bas niveau de la

parole étaient absentes) la synchronisation de l'activité neuronale s'établissait à un autre moment de la phase des oscillations neuronales et le couplage s'établissait avec des fréquences moins élevées. Ces résultats ont des implications importantes pour la théorie d'*entraînement de phase*, dans la mesure où nous avons démontré, pour la première fois, que celui-ci ne reflète pas seulement la réponse neurale aux fluctuations d'amplitude: il consiste également à un alignement à l'information phonétique.

Tous les résultats décrits ici peuvent être combinés pour formuler une conclusion générale: Alors que l'idée du traitement discrétisé dans le système visuel est déjà relativement établie (VanRullen and Koch, 2003), pour le système auditif les résultats négatifs reportés dans la littérature semblaient être – à première vue – surprenants. Nos résultats démontrent que cette absence apparente de sous-échantillonnage dans le domaine auditif n'indique pas qu'il n'y a pas de discrétisation. Mais plutôt, et contrairement au système visuel, qu'il y a pour le system auditif, une nécessité d'utiliser des mécanismes neurophysiologiques pour éviter les effets destructifs du sous-échantillonnage. Ainsi, l'*entraînement* des oscillations neuronales aux stimuli rythmiques présents dans l'environnement pourrait être un processus fondamental pour le système auditif, le sous-échantillonnage étant probablement réalisé à un niveau hiérarchique élevé du traitement des stimuli. En utilisant ces deux processus, le sous-échantillonnage temporel de l'entrée auditive pourrait être possible sans perte des informations essentielles.

Table of Contents

The perception, if not the enjoyment, of musical cadences and of rhythm is probably common to all animals, and no doubt depends on the common physiological nature of their nervous system.

Charles Darwin, 1871

GENERAL INTRODUCTION

Every movie that is shown in cinema or television is a series of images, yet it appears smooth and continuous – a classical example for the fact that *subsampling* a continuous stream of information can be parsimonious without affecting its perceived integrity and continuity. Is it possible that the human brain uses a similar mechanism: Subsampling the vast incoming flow of information into "chunks" or "snapshots" of input, being at the same time parsimonious yet without losing essential information? This question can be seen as the central theme of this thesis and we will follow it throughout this work. We will encounter phenomena that potentially represent reflections of the brain's subsampling procedure and meet difficulties that the brain has to face on its way to an efficient stimulus processing. We will realize that these difficulties might be of particular importance for the auditory system, and that it might have developed a clever "tool" to face them. It will become clear that *periodicity* is an important feature of this tool, a feature that might be *actively* used by the brain for an optimal stimulus selection. All these points will be discussed in the following, chapter by chapter, with the focus on the comparison of arguably the two most important sensory systems for humans: Vision and audition. I will start by providing a short summary of the most important "periodic tool" for the brain, *neural oscillations*, and show how they were discovered and can be measured and analyzed. I will also shortly summarize the structure of the visual and auditory systems, and show that we can find neural oscillations in (almost) all of their hierarchical stages. The rest (and main part) of the thesis is composed of eight chapters, each of them based on a manuscript that is either published, submitted, or ready to submit. In the first chapters, we will meet the idea that neural oscillations are ultimately related to the creation of *perceptual cycles,* periodic fluctuations in perception that can be seen as evidence for the

brain's discrete way of stimulus processing. These cyclic processes can be used to structure the system's input, and individual elements of this influx are either grouped or segregated, depending on where they fall with respect to the oscillatory cycle. Moreover, we will see that, although neural oscillations play an important role for both visual and auditory systems, and mechanisms of discretization are plausible in both of them, the two systems do differ in their ways of stimulus processing and selection. We will come to the conclusion that the adaption of neural oscillations to the stimulus influx is of particular importance for the auditory system, due to its vulnerability to a loss of information if the related subsampling is done "blindly". To the characterization of this adaption, also called *phase entrainment,* are dedicated the following chapters. They will mainly concentrate on speech sound, arguably one of the most important rhythmic stimuli in the auditory environment. We will see that the adaptation of neural oscillations to speech sound includes mechanisms that take place on a relatively high level of stimulus processing. The characterization of these high-level mechanisms, based on results obtained in this thesis as well as in other studies coming to a similar conclusion, is then summarized and discussed in a separate chapter, in the form of a review article. In the final chapter, a general discussion is provided, also in form of a scientific article. In contrast to the preceding chapter, which focuses on phase entrainment to speech sound, more general insights obtained in this thesis and elsewhere are summarized, explaining how neural oscillations are used by the visual and auditory system, resulting in an efficient way of environmental subsampling, stimulus selection and processing. Both similarities and differences between the two systems are illustrated.

Neural oscillations and how they are recorded and analyzed

It is relatively unclear who described neural oscillations for the first time. It is speculated that in the 1870's, almost 150 years ago, Caton observed, but did not record, spontaneous oscillations in the potential of cortical gray matter (Bremer, 1958). Certainly one of the most famous "discoverer" of oscillations is Hans Berger, who connected a string galvanometer to electrodes at the skull of patients and relatives and reported potential changes with surprising regularity (for a review, see Herrmann et al., 2015). Although already described in mammals in 1925 by Práwdicz-Neminski, it is Hans Berger who is generally accepted to be the inventor of the electroencephalogram (EEG; Berger, 1929). Indeed, the most prominent and "famous" neural oscillation, the alpha-rhythm, was named by him in 1930. Berger's discovery was met with skepticism first, but soon oscillations at other frequencies were discovered (e.g., Dietsch, 1932; Rohracher, 1935; Walter, 1936; Jaspers and Andrews, 1938). Already at that time, Berger (among others) seemed to have realized that the measured signals can be decomposed into different oscillatory frequency bands by the use of certain signal processing algorithms, such as Fourier transformation (Rohracher, 1935; Herrmann et al., 2015). Although the definition of these bands have changed somewhat (Figure 1), this notion is still kept up nowadays, and researchers have tried to assign different functions to the different frequency bands (for reviews, see, e.g., Başar et al., 2001; Lopes da Silva, 2013; Herrmann et al., 2015).

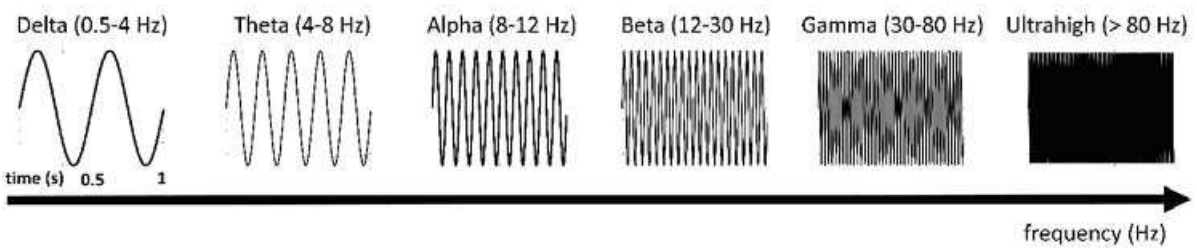


Figure 1. Neural oscillations can be found at various frequency bands in the brain.

Given that they are evolutionary preserved (Buzsáki et al., 2013), it seems to be surprisingly difficult to find different roles for neural oscillations at different frequency bands. Slow oscillations (~0.5 – 4 Hz), also called delta-oscillations, seemed to be increased during sleep (Steriade et al., 1993), but they also adjust to rhythmic stimulations during wake, as we will see in detail in several chapters of this thesis. Theta-oscillations (~4-8 Hz) in the hippocampal gyrus of the brain are important for memory processes (Pavlides et al., 1988; Lisman and Jensen, 2013), but cortical theta oscillations might fulfill a different role, especially for audition (Luo and Poeppel, 2007; Strauß et al., 2014a). Alpha oscillations (~8-12 Hz) were long considered as the brain's idle rhythm (Pfurtscheller et al., 1996), but it becomes increasingly clear that their role as an inhibitory rhythm is of particular importance (Klimesch et al., 2007; see also chapter 8). Beta oscillations (~12-30 Hz) might be important for the maintenance of the current cortical state (Engel and Fries, 2010), but they also seem to be involved in motor planning (Picazio et al., 2014) and even memory (Salazar et al., 2012) or audiovisual integration (Keil et al., 2014). Gamma oscillations (~30-70 Hz) are a relatively local phenomenon in the brain, potentially involved in binding or integration of stimulus features (Fries, 2009), but also in several other tasks (Schepers et al., 2012; Lee et al., 2014). Recently, evidence accumulated that even activity at higher frequencies might play an important role in the processing of information in the cortex (Lachaux et al., 2012).

Another point should be mentioned here, as it is essential for the reasoning behind many results reported in this thesis. Already in the time of Berger, Bishop (1932) found that the excitability in the optic pathway in the rabbit changes in a cyclic manner. This result might not sound exciting at first glance, but it might have important consequences for our understanding of neural oscillations today: It is now commonly accepted that neural oscillations might underlie – or be a reflection of – these rhythmic changes in neuronal excitability. This is

important, because it creates a different role for these oscillations, besides the cognitive functions mentioned above: They can be used to control the input to the brain, with “windows of opportunity” (Buzsáki and Draguhn, 2004) for the input to be processed when arriving at a phase of high excitability, and unfavorable moments for stimulus processing at a phase of low excitability. This notion is a central idea for this work and will be taken up in many chapters of the thesis.

Apart from their role, a question of similar importance but also ambiguity is where these oscillations actually come from. Although the membrane potential of single neurons already shows regular fluctuations (Llinás, 1988), it seems to be the interplay between many different neurons that produces the oscillations that are relevant for perception and behavior (Buzsaki and Draguhn, 2004): Neural oscillations, as they are commonly measured (see below), stem from synchronized activity of neuronal populations (Uhlhaas et al., 2009). Interestingly, different neuronal populations might have different preferred frequencies in which they oscillate (Hutcheon and Yarom, 2000). Also, synchronized populations do not necessarily have to be located in the same area: The famous alpha-rhythm, for example, might be produced by complex thalamo-cortical (Lopes da Silva, 1991) or cortico-cortical loops (Bollimunta et al., 2008).

Neural oscillations can be recorded at several levels (Figure 2) and some of them will appear in this thesis. Apart from their assumed reflection in psychophysical data (Chapters 1, 2 and 4), they are visible in intracranial recordings (as local field potentials, LFP, or their second spatial derivative current source density, CSD; Chapters 3 and 6), on the cortical surface (as ECoG) or as recorded from the scalp (with EEG or magnetoencephalogram, MEG; Chapter 5).

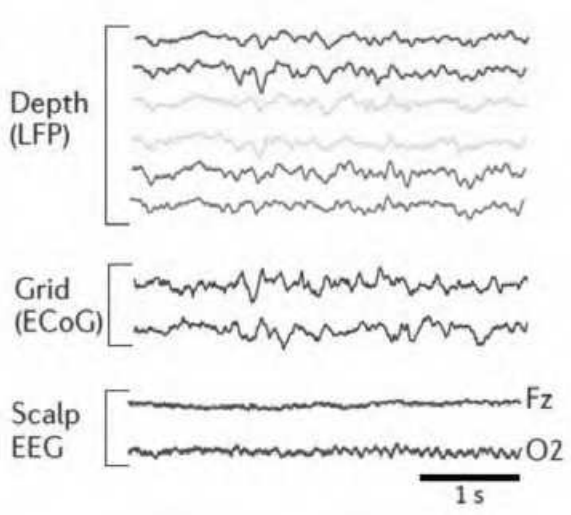


Figure 2. Neural oscillations can be measured at different scales. Here, exemplary signals from LFP, ECoG, and EEG recordings are shown. The spatial resolution decreases in this order. Modified from Buzsaki et al., 2012.

Studies have shown that oscillations measured at different levels are related: For instance, EEG oscillations can be associated with neuronal firing (Whittingstall and Logothetis, 2009; Ng et al., 2013). Nevertheless, there are differences between the levels of recording: Of course, the spatial resolution decreases from LFP/CSD via ECoG to EEG. At the level of LFP/CSD, laminar profiles can be recorded (Figure 3), and oscillatory activity can be differentiated between layers. Moreover, the current flow can be spatially estimated, resulting in sinks and sources of neural activity (Schroeder et al., 1998). However, simultaneous recordings are usually restricted to one or two brain regions, and often change site from measurement to measurement, making both comparison across experiments and a complete overview of brain activity difficult. Also, LFP/CSD are, for obvious reasons, rare for human subjects and, consequently, insights for the human brain have to be inferred from animal experiments. At the level of the EEG, at each electrode, the measured activity is a transformed mix from many different cortical sources (Lopes da Silva, 2013), and methods of source analysis have only

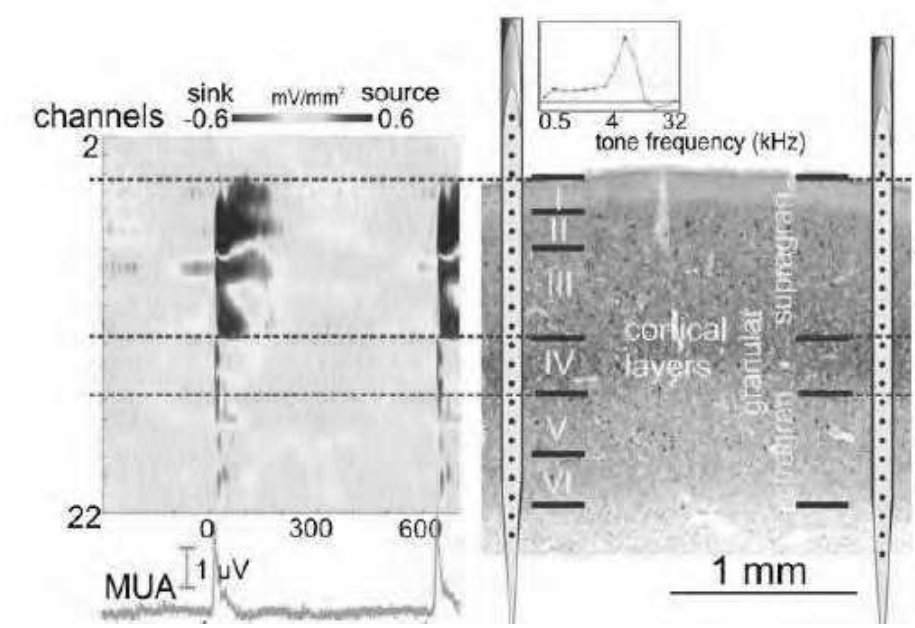


Figure 3. Example for a laminar profile recorded intracranially in monkey primary auditory cortex. The current flow (left) across cortical layers (right) can be estimated and results in sinks (red) and sources (blue). The recording site usually has a sound frequency it is tuned to (green curve in inset). From Lakatos et al., 2013a.

partly met with success (Michel et al., 2004). EEG can only capture the activity perpendicular to the cortical surface, mostly from pyramidal cells (Lopes da Silva, 2013). However, EEG is cheap, can be easily applied in human recordings (Figure 4) and is an appropriate tool for research projects where spatial origins are not relevant.

We now turn towards the analysis of neural oscillations. As mentioned above, the idea of spectral decomposition was introduced relatively early but is still up-to-date, as we will see throughout this thesis. Most studies concerning the functional role or clinical relevance of brain oscillations focus on the amplitude of the oscillations in particular frequency bands. However, as shown in Figure 5, an oscillation is not only defined by frequency and amplitude, but also by its phase. Indeed, the phase of neural oscillations will be the most important oscillatory parameter in this work. The phase of an oscillation can be defined as the fraction of a whole period that has elapsed with respect to an arbitrary reference (Ballou, 2005). Most commonly, a cosine wave is used as a reference. In Figure 5, two 1 Hz waves are shown, which

Figure 4. EEG is recorded from the scalp, commonly using between 16 and 256 electrodes. From Herrmann et al., 2015.

are shifted by $\frac{\pi}{2}$. Due to the phase shift, they differ in their phase state for each arbitrary point in time. Alternatively, an oscillatory signal can be thought of as a rotating vector in the complex plane, where the radius *r* (or the absolute value) reflects the amplitude and the angle reflects the phase of the oscillation (Figure 5B). Thus, every rotation of a full 360° cycle (or 2π) by the vector represents a single cycle of the oscillation. As an example, the state of the two signals (blue and red) at time t = 0.5 s in Figure 5A is shown again in Figure 5B, this time in the complex plane. It can be seen that the two signals do not differ in amplitude (r = 1) or frequency (1 Hz), but only in their phase. There are several ways of estimating amplitude (or power, which is proportional to the squared amplitude) or phase of an oscillation. Most of the standard methods rely on the same idea: Waveforms with different frequency, phase and/or amplitude are compared with the signal of interest and the goodness of fit between waveform and signal is taken as an indicator of how much of this "reference" waveform is "hidden" within the signal. This waveform can be a sine or cosine wave, as in the case of the Fourier transform, or of a more complex shape, as in the case of Wavelet Transformation. Whatever the method,

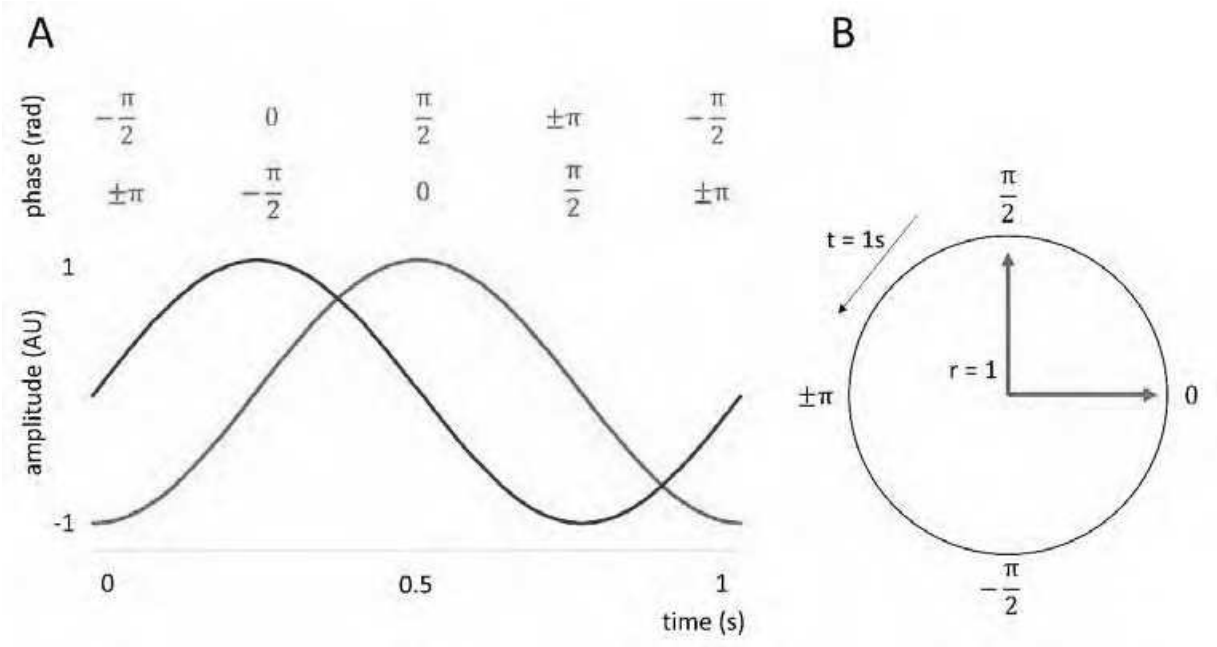


Figure 5. A. Two phase-shifted oscillations (blue and red; frequency 1 Hz, amplitude 1 AU) differ in their phase state (top, in radians) at each arbitrary moment in time. B. Oscillatory signals can be displayed as a rotating vector in the complex plane. The radius *r* reflects the amplitude and the angle reflects the phase of the oscillation. The signals that are shown here correspond to the blue and red waves at time t = 0.5 s in A. Note, that their phase shift is now visible as the angle between them.

the ultimate goal is a decomposition of the complex recorded signal into different frequency bands (Figure 1), each of which defined by an amplitude and phase. This can potentially be done as a function of time, as in the case of time-frequency transformations such as the Wavelet Transformation (Herrmann et al., 2014).

The visual system and its oscillations

Light at wavelengths of ~350-750 nm stimulates receptors (rods and cones) in the retina. The signal is then transmitted to retinal ganglion cells, the first cells in the visual system that fire action potentials (Bear et al., 2006). Surprisingly (or not, keeping in mind that the retina is considered part of the brain; Bear et al., 2006), these cells already exhibit oscillations whose frequency seems to differ between spontaneous (1-5 Hz and 30 Hz) and evoked activity (~70-90 Hz; Neuenschwander et al., 1999; Arai et al., 2004). These oscillations can reach and affect neural activity in cortical areas (Koepsell et al., 2009). From retinal ganglion cells, the neural

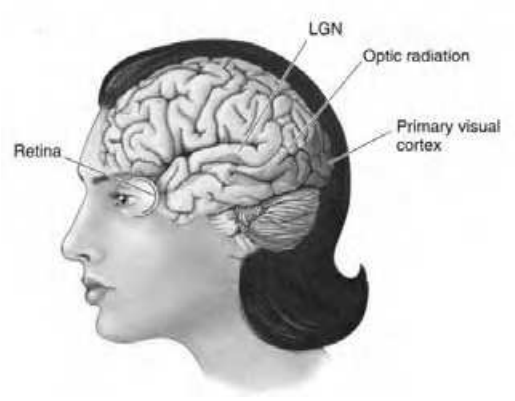


Figure 6. Schematic view of the visual pathway. There is only one relay station between retina and primary visual cortex: The LGN. From Bear et al., 2006.

signal is transmitted to the brain via the optic nerve. There is only one relay station[1] between the optic nerve and primary visual cortex (V1): the lateral geniculate nucleus (LGN) of the thalamus (Figure 6). The LGN is a six-layered structure with dense connections to higher cortical regions. These connections seem to be important for neural oscillations, especially for alpha and gamma frequency bands: The latter can be observed both in LGN and in the cortex and thalamocortical connections seem to be necessary for their generation (Wall et al., 1979; Deschênes et al., 1984; Usrey and Reid, 1999; Lorincz et al., 2009; Minlebaev et al., 2011; although a matter of debate for gamma oscillations, Bastos et al., 2014). Neural oscillations are abundant in visual cortex, both in primary and higher-order regions. As already described by Hans Berger and mentioned above, the alpha rhythm is the most prominent rhythm in the visual system (this topic is also treated extensively in the General Discussion), and this is particularly true for V1. Although alpha waves can readily be measured in and above V1, it is somewhat surprising that their origins are still debated and can be assigned to thalamocortical or corticocortical connections (or both; Steriade et al., 1990; Lopes da Silva, 1991; Jones, 2002;

[1]For both visual and auditory systems, only the principal ascending pathway is described here. Of course, many alternate pathways and connections exist, and there are extensive feedback connections from higher-level regions back to earlier stages of the pathways. Also, for both visual and auditory systems, the cortical areas seem to consist of “core” and “belt” regions and the ascending cortical pathway splits into a ventral “what” and a dorsal “when”-stream (Rauschecker, 2015). These are not treated further in this chapter.

Bollimunta et al., 2008). The different regions of the cortical visual hierarchy seem to communicate by the use of neural oscillations in different frequency bands. Again, alpha and gamma oscillations seem to be of particular importance, with gamma band activity mainly associated with feedforward processing in supragranular layers whereas alpha band activity seems to be important for feedback processing in deeper cortical layers (Maier et al., 2010; Buffalo et al., 2011; Spaak et al., 2012; van Kerkoerle et al., 2014). There is also evidence that the alpha phase is coupled to the power of the faster gamma oscillations (Jensen et al., 2014). Finally, there are several regions associated with the processing of visual information that seem to be involved in oscillatory processes, but whose role clearly needs further investigation. One example is the lateral pulvinar, which is, again, tightly linked with alpha oscillations (Purushothaman et al., 2012), cortical processing and cognitive functions (Saalmann et al., 2012).

The auditory system and its oscillations

Only counting the number of relay stations between sensory organ and cortex, the auditory system is more complex than the visual one (Figure 7). Moreover, the characterization of neural oscillations in the auditory system is complicated by the fact that its input is already oscillatory: The ear gathers variations in air pressure (these variations can have a frequency between 15 and 20000 Hz to evoke neural activity in the auditory pathway) which is transformed into motion of fluid and finally into electrical (neural) signals, both of which in the inner ear, and most importantly, by the cochlea. It is interesting to note that the cochlea has frequency sensitivity (due to variations in mechanical properties: Different parts of the cochlea are activated, depending on the frequency of the sound input) and therefore already breaks down the input into its spectral components.

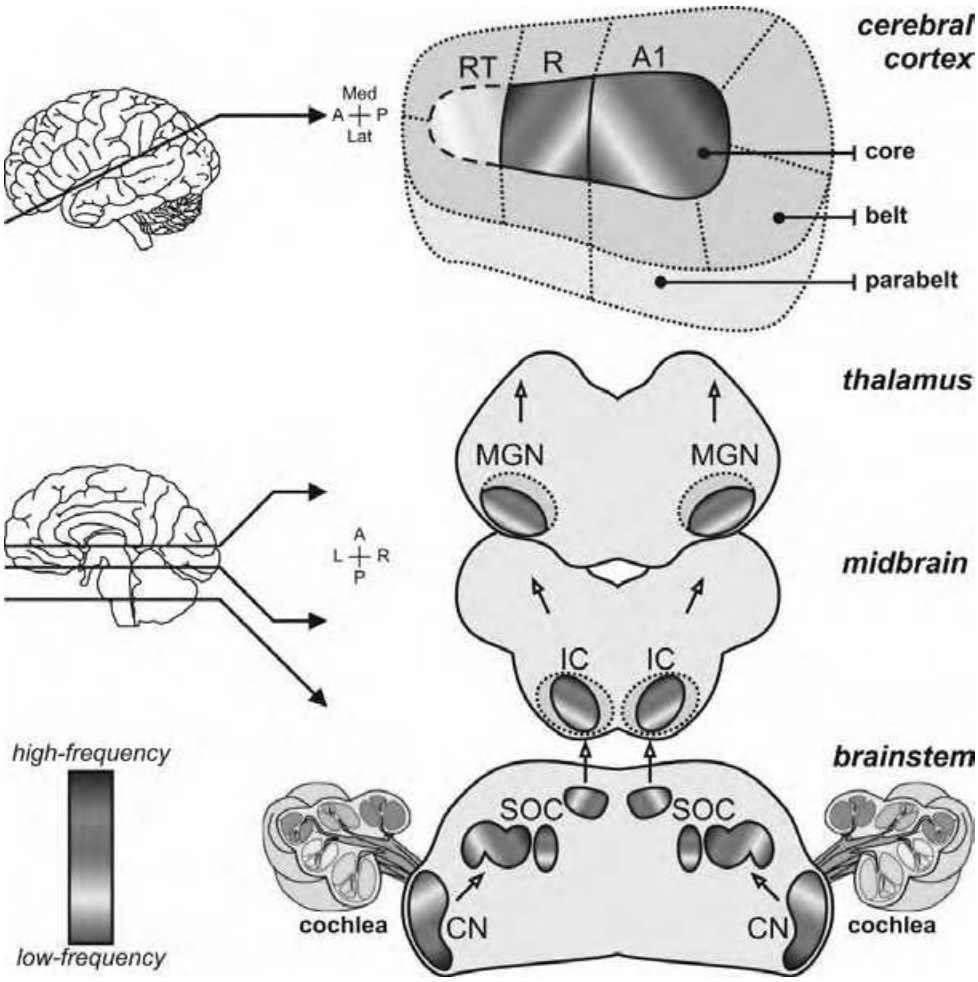


Figure 7. Overview of the principal auditory pathway. CN: Cochlear Nucleus. SOC: Superior Olivary Complex. IC: Inferior Colliculus. MGN: Medial Geniculate Nucleus. A1: Primary Auditory Cortex. R: Rostral field. RT: Rostrotemporal field. Note that areas R and RT are mostly described for non-human primates, but there is evidence that there are corresponding regions in the human brain with respect to their histological and electrophysiological properties. The tonotopical organization throughout the auditory system is shown as color gradients. From Saenz and Langers, 2014.

This "tonotopy[2]" of the cochlea is conserved up to the level of primary auditory cortex (A1; Figures 3 and 7; Saenz and Langers, 2014; Rauschecker, 2015), and beyond (although the tuning becomes broader at later stages of the auditory pathway). As explained above, the hair cells in the cochlea necessarily oscillate: They can "follow" (i.e. be in phase with) their rhythmic input (i.e. the sound) up to frequencies of 4 kHz (Palmer and Russell, 1986; Köppl, 1997; Heil

[2] Note that the tonotopic organization of the auditory system is equivalent to the retinotopic organization of the visual system (Rauschecker, 2015). Therefore, the time domain can be considered of similar importance for the auditory system as the spatial domain for the visual system (Kubovy, 1988) and sweeps of frequency-modulated tones for the auditory system might be equivalent to moving bars of light for the visual system (Mendelson and Cynader, 1985).

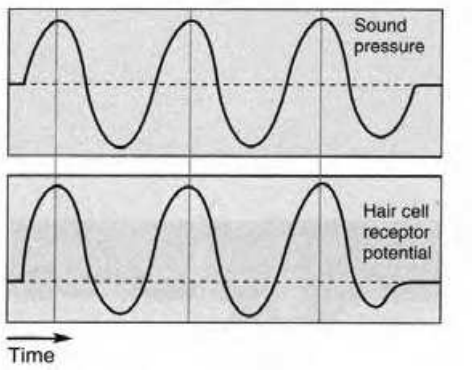


Figure 8. Sound is an oscillatory stimulus (top). The potentials recorded in the cochlea can "follow" this rhythm up to 4 kHz (bottom). From Bear et al., 2006.

and Peterson, 2015; of course, beyond those frequencies the sound is still processed, but cells cannot stay in phase anymore). Although many neurons seem to follow the sound *envelope* (rather than the fine structure) at later stages of the auditory system, it is clear that these *stimulus-evoked* oscillations make it difficult to decide when neurons merely follow their input (Figure 8), and when "real" (*endogenous*) oscillations are "at work". This problem is existent through the whole auditory system and several chapters are dedicated to it (Chapters 4-7).

In the first relay station of the auditory system, the cochlear nucleus (CN), high-frequency oscillations (0.4 – 1 kHz) are abundant and it is likely that they play a role for pitch processing (Wiegrebe and Winter, 2001; McMahon and Patuzzi, 2002). Not much is known about slower oscillations at this level, although relatively complex processes, such as auditory stream formation, might already take place here (Pressnitzer et al., 2008). The situation is similar for the superior olive (SOC), an early auditory structure associated with sound localization (Grothe, 2000): High-frequency oscillations seem to be present (Goldwyn et al., 2014), but slower oscillations are rare. Again, essentially nothing is known about neural oscillations at the level of the inferior colliculus (IC), although one study (Langner and Schreiner, 1988) reports "intrinsic oscillations" of very high frequency (periods of "integer multiples of 0.4 ms"). This absence of slower neural oscillations changes somewhat at the level of the thalamus, in

the medial geniculate body (MGN): Gamma oscillations seem to be the most prominent oscillations here (Brett et al., 1996). However, these oscillations also might be an important feature of cortical auditory processing and the MGN might merely modulate the gamma activity in auditory cortex (Barth and MacDonald, 1996; Smith et al., 2007). Oscillations at many frequencies can be found at the level of A1 and beyond: Delta (Lakatos et al., 2008; Stefanics et al., 2010), theta (Henry and Obleser, 2012; Ghitza, 2013), alpha (Frey et al., 2014; Strauß et al., 2014b), and gamma (Lakatos et al., 2005; Fontolan et al., 2014) frequencies seem to play a role for cortical auditory processing. Similar to the visual system, the phase of slower oscillations is coupled to the power of faster oscillations (Lakatos et al., 2005; Fontolan et al., 2014), and gamma activity is associated with feedforward processing between primary and higher auditory areas, whereas slower oscillations might be responsible for feedback processing (Fontolan et al., 2014). There seems to be a general phenomenon for the auditory (and potentially also for the visual) system: The higher in the pathway of stimulus processing, the slower the frequency of the oscillations (Edwards and Chang, 2013).

As it is apparent from this summary, neural oscillations are abundant in both visual and auditory systems, especially in cortical regions. Also, there are clear differences between the systems: For instance, in contrast to vision, the input to the auditory system is already oscillatory[3] and can easily be followed. The spectral characteristics of neural oscillations in the auditory system seem to be more variable than in the visual one, where alpha oscillations are dominant. We will discover these differences – in particular their meaning for stimulus processing and selection – in the next seven chapters. The results described in these chapters

[3] It is acknowledged here that light – as the principal visual stimulus – can also be considered a wave and thus oscillatory. However, as retinal cells cannot follow the frequency of light, this fact is disregarded here.

(and elsewhere) lead to a final summary and conclusion in the form of the eighth chapter, the General Discussion.

What this thesis is about

Having introduced neural oscillations and their abundance in the visual and auditory systems, we can now come back to the question that was presented at the beginning of this Introduction. Does the brain subsample its environment or: Is perception discrete or continuous (VanRullen and Koch, 2003)? Sudden changes are rare in nature. If perception were indeed discrete (and not continuous), instead of taking one "snapshot" at a particular moment in time followed by an absence of stimulus processing, one would rather expect a smooth, *periodic*, or *cyclic* change from one "snapshot" to the other, with most incoming information being processed at the time of the "snapshot", and the least likelihood of input being perceived between two subsequent "snapshots" (i.e. after half the period of the perceptual cycle with respect to the moment of a "snapshot"). Following this notion, and combined with the information summarized in this chapter, it becomes clear that neural oscillations are the optimal "candidate" for a reflection of these *perceptual cycles*. Perceptual cycles[4] create "windows of opportunity" for incoming information to be perceived or processed, with these windows gradually cycling between being open and closed. If neural oscillations reflected *perceptual* cycles, one would expect them to be linked to perception as well – and, importantly, one would expect the *phase* of neural oscillations to be related to the instantaneous state of perception. Indeed, this is what has been found in the last few years: Busch et al. (2009) showed that the pre-stimulus phase of alpha oscillations in the EEG is

[4] It is important to distinguish perceptual cycles from a mere "low-pass" filtering of information, such as when the pace of information is too fast for a system to follow (and is therefore lost), as the latter does not show cyclic properties.

predictable for the perception of a visual stimulus at threshold. Dugué et al. (2011) reported that the likelihood of perceiving a phosphene induced by transcranial magnetic stimulation (TMS) depends on the alpha phase as well. Finally, it has been shown a presented luminance sequence (with an equal amount of power at all frequencies within a given range) "reverberates" in the brain for more than one second at a frequency of 10 Hz (VanRullen and Macdonald, 2012). Another evidence for perceptual cycles stems from a phenomenon called "continuous wagon wheel illusion": If a signal (here: the wagon wheel) is a moving periodic visual pattern, and the information processing system is taking temporally discrete samples with a sampling rate lower than a critical limit, then the system's representation of the signal is inaccurate (this effect can be described by the term "aliasing"). This inaccuracy can be experienced as a pattern that seems to move in the wrong direction. Interestingly, the critical limit frequency has been found to be located at approximately 13 Hz and thus, again, lying in the alpha band (Purves et al., 1996; VanRullen et al., 2005). The perception of the illusion goes along with a peak in the power spectrum of the EEG at the same frequency (VanRullen et al., 2006; Piantoni et al., 2010) and is thus reflected on a neural level as well.

These findings strongly suggest that perceptual cycles do exist in the human brain and that they are intimately linked to neural oscillations (in particular in the alpha range). However, it is intriguing to see that all of the above mentioned studies report findings in the visual domain. Can we find perceptual cycles in the auditory domain as well? I tried to answer this question in this PhD thesis and – given the prominence of neural oscillations in the auditory system – it might be surprising to find a (at first glance) negative answer, as presented in Chapter 1: Perceptual cycles seemed to bring about too many drawbacks for the auditory system. However, this answer brought up more questions: What can the auditory system do, in order to avoid these drawbacks and afford perceptual cycles? Two solutions are presented in this

thesis. First, as shown in Chapter 2, perceptual cycles in the auditory system might be located on a hierarchically higher level, where auditory stimuli are temporally more stable. Second, the auditory system might adjust (*entrain*) to its (dominantly rhythmic) input, and thereby *decide* which information is processed (at the moment of the "snapshot") and which information is lost (between "snapshots"). The entrainment of the phase of neural oscillations to rhythmic stimulation, in particular for auditory input, has often been demonstrated (e.g., Lakatos et al., 2008; Stefanics et al., 2010; Henry and Obleser, 2012). However, considering phase entrainment as the auditory system's "tool" for an efficient way of stimulus selection (as it is necessary to avoid disruptive effects of perceptual cycles), it needs to be shown that this tool can be used *actively.* In Chapter 3, it is shown that, in an ambiguous auditory scene, it is the neural phase that "decides" whether two simultaneous auditory inputs are grouped into a single stream or segregated into two separate streams. For Chapters 4-7, speech/noise stimuli have been constructed without systematic fluctuations in amplitude and spectral content – in order to entrain to those stimuli, the auditory system would have to actively disentangle speech and noise features (because it cannot "follow" the fluctuations in amplitude anymore). Phase entrainment to the constructed speech/noise stimuli is shown in a psychophysical experiment (Chapter 4), using EEG recordings (Chapter 5) and intracranial recordings in monkey A1 (Chapter 6) and an overview of phase entrainment to high-level features of speech sound is given (Chapter 7). Having described all these results, we can come back to the apparent discrepancies in the use of neural oscillations for stimulus processing and selection (including subsampling and perceptual cycles) between the visual and auditory system. On the basis of the obtained results, a general framework to explain these differences is presented in Chapter 8, the General Discussion.

CHAPTER 1: ON THE CYCLIC NATURE OF PERCEPTION IN VISION VERSUS AUDITION

In the previous chapter, we have seen that neural oscillations are ubiquitous in the brain and that they reflect cyclic changes between more and less favorable moments for stimuli to be processed, due to underlying rhythmic changes in neuronal excitability. We will now make a step onto a more abstract level and try to see these oscillations as a cyclic reflection of discrete processing in the brain (and ultimately, perception).

As summarized above and reviewed in the following article, there is plenty of evidence for periodic changes in perception, called *perceptual cycles*: For instance, the detection of a visual target at threshold depends on the EEG phase at 7 Hz (Busch et al., 2009) and a given luminance sequence reverberates in the brain for up to one second at a frequency of 10 Hz (VanRullen and Macdonald, 2012). Perceptual cycles would represent clear support for a discretization of information in the brain, both on a neuronal and perceptual level. However, evidence for perceptual cycles is abundant for vision, but remains sparse for audition. Most of the above-mentioned findings have been failed to be replicated in the auditory domain: The perception of an auditory click in silence seems to be independent of EEG phase (Zoefel and Heil, 2013); No “neural reverberation” of auditory stimulation could be found (İlhan and VanRullen, 2012); The attempt to create an illusion similar to the continuous wagon wheel illusion (as explained in the preceding chapter), by measuring the perceived motion direction of a spatially periodic sound source, failed (unpublished). This chapter is dedicated to shed some light on this controversy. The current state-of-the-art is described and several psychophysical experiments are presented, all of them speaking against perceptual cycles in audition.

As mentioned above, measuring the perceived motion direction of a spatially periodic sound source did not succeed in revealing an auditory "wagon wheel" illusion. However, it might be possible that sound frequency or "pitch", rather than spatial position, may be the proper equivalent to the spatial location of visual objects. Thus, periodic stimuli were designed that moved in particular directions in the frequency domain—so-called Shepard or Risset sequences (Shepard, 1964). Participants were presented with sequences of each temporal frequency in randomized order, moving up or down in frequency space, and asked to report their perceived motion direction (up/down). Note that an auditory "wagon wheel" illusion introduced by those stimuli would not merely result in performance at chance level, but, critically, in performance *below* chance level, as perception is "biased" towards the "wrong direction". It was found that the direction judgments were only accurate up to 3–4 Hz, and deteriorated rapidly at higher temporal frequencies. This low-pass sensitivity function critically limits the possibility of measuring a "wagon wheel" illusion in the auditory domain: If perceptual performance is already at chance at the hypothesized frequency of the illusion, then this illusion will simply not be observed—whether the perceptual process relies on periodic sampling or not. In other words, if auditory perceptual cycles exist, then it must be at a sampling rate above 3–4 Hz (if the illusion occurred at frequencies below 3-4 Hz, then it would have been possible for our subjects to perceive it, and they would have systematically reported reversed motion).

Why is it so difficult to observe auditory perceptual cycles? In this chapter, we present the following hypothesis: In contrast to vision, the auditory system heavily relies on temporal, continuously changing, information. Whereas a visual scene might be stable for a relatively long time, acoustic stimuli are fluctuating by nature (as they can be described by their spectral content, a variable that can only be defined in time). Consequently, whereas "blind"

sampling of the environment (i.e., taking “snapshots” that are unrelated to the stimulus input to the system) might not disrupt the integrity and perceived continuity of visual information, the same sampling mechanism could be destructive for auditory processing. This notion was confirmed in a psychophysical experiment. Participants watched and listened to sequences of 3-s long video and audio snippets (respectively) in different blocks while performing a two-back task (responding to a repeat of the penultimate snippet). The visual and auditory inputs were subsampled in a representation space roughly equivalent to the first sensory stage of each system: the retina (with the entire image representing a subsampling “frame”) and the cochlea (with the instantaneous complex frequency spectrum resulting from a wavelet decomposition of the audio signal as a subsampling “frame”), respectively. That is, for both sensory systems, the consequences of the most severe possible temporal subsampling strategy were evaluated, by subsampling the very input to the system. It was found that visual performance only started to deteriorate below 2.5 frames per second, but auditory performance suffered for all subsampling frequencies below 32 Hz. Thus, vision is indeed an order of magnitude more robust to temporal subsampling than audition, arguing against the feasibility (or even the existence) of auditory perceptual cycles.

Nevertheless, is it possible that we cannot discover auditory perceptual cycles in the same way as we do for vision? And if so, why is this the case? In the following article, several explanations are presented and suggestions are made that can “keep alive” the notion of perceptual cycles in the auditory domain. These suggestions are then taken up again and developed further in Chapter 2.

Article:

VanRullen R, Zoefel B, Ilhan B (2014) On the cyclic nature of perception in vision versus audition. Philos Trans R Soc Lond B Biol Sci 369:20130214.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

On the cyclic nature of perception in vision versus audition

Rufin VanRullen, Benedikt Zoefel and Barkin Ilhan

Phil. Trans. R. Soc. B 2014 **369**, 20130214, published 17 March 2014

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

rstb.royalsocietypublishing.org

Research

Cite this article: VanRullen R, Zoefel B, Ilhan B. 2014 On the cyclic nature of perception in vision versus audition. *Phil. Trans. R. Soc. B* **369**: 20130214.
http://dx.doi.org/10.1098/rstb.2013.0214

One contribution of 13 to a Theme Issue 'Understanding perceptual awareness and its neural basis'.

Subject Areas:
cognition

Keywords:
perceptual cycles, discrete perception, brain rhythms, perceptual sampling, sampling rate, speech processing

Author for correspondence:
Rufin VanRullen
e-mail: rufin.vanrullen@cerco.ups-tlse.fr

On the cyclic nature of perception in vision versus audition

Rufin VanRullen[1,2], Benedikt Zoefel[1,2] and Barkin Ilhan[3]

[1]Université de Toulouse, France
[2]CNRS-CerCo, UMR 5549, CHU de Purpan, Toulouse, France
[3]Meram Medical Faculty, Konya University, Konya, Turkey

Does our perceptual awareness consist of a continuous stream, or a discrete sequence of perceptual cycles, possibly associated with the rhythmic structure of brain activity? This has been a long-standing question in neuroscience. We review recent psychophysical and electrophysiological studies indicating that part of our visual awareness proceeds in approximately 7–13 Hz cycles rather than continuously. On the other hand, experimental attempts at applying similar tools to demonstrate the discreteness of auditory awareness have been largely unsuccessful. We argue and demonstrate experimentally that visual and auditory perception are not equally affected by temporal subsampling of their respective input streams: video sequences remain intelligible at sampling rates of two to three frames per second, whereas audio inputs lose their fine temporal structure, and thus all significance, below 20–30 samples per second. This does not mean, however, that our auditory perception must proceed continuously. Instead, we propose that audition could still involve perceptual cycles, but the periodic sampling should happen only after the stage of auditory feature extraction. In addition, although visual perceptual cycles can follow one another at a spontaneous pace largely independent of the visual input, auditory cycles may need to sample the input stream more flexibly, by adapting to the temporal structure of the auditory inputs.

1. Introduction: perceptual awareness, a discrete process?

Our conscious perception of the world appears smooth and continuous. A moving object is not seen to disappear here and reappear there, but as successively occupying all positions in between. Similarly, the sound of an approaching car seems to steadily loom closer, without being interrupted by brief recurring moments of silence. And yet, it is not at all certain that the brain mechanisms supporting our sensory perception are themselves continuous; rather, visual and auditory perception may well be intrinsically discrete or cyclic [1–5]. We are not referring here to the discreteness of individual neuronal events (action potentials, synaptic release) but to the potentially discrete nature of perceptual experience itself. In this case, the continuity of our inner experience would merely be an illusion, a temporal 'filling-in' created by our brain to hide its recurring (albeit brief) moments of blindness and deafness, perhaps in the same way as the 'blind spot' of the retina is hidden from our consciousness by spatial filling-in mechanisms [6].

The notion of perceptual 'snapshots', 'moments' or 'cycles', once popular [1–4] but later discarded without ever being firmly disproved [7] has regained momentum in recent years owing to a number of converging experimental studies. As we shall see in the following sections, however, most of this new experimental evidence concerns the periodicity of visual perception; it has been more challenging, it seems, to uncover similar signatures of auditory 'snapshots'. After reviewing the corresponding findings, we consider the major differences between the visual and auditory modalities, both in terms of cerebral organization and information processing demands, that could justify a difference in temporal perceptual organization. Finally, we speculate that both systems, under the influence of neuronal oscillations, may, indeed, represent

sensory information as a sequence of perceptual cycles, but we will argue that the properties of these cycles must be vastly different for vision versus audition.

2. Perceptual cycles in vision

The notion of discrete perception was a prevalent idea after World War II and until at least the 1960s [1–4]. Even though some authors considered this discreteness to be an intrinsic property of all sensory modalities [1], most of the available experimental evidence came from studies of visual perception [3,4]. One possible reason for such a bias is the fact that the hypothesis of discrete perception was always strongly tied to the observation of large-amplitude 'alpha' (8–13 Hz) oscillations in electroencephalographic (EEG) recordings [8]. Because these alpha rhythms were found to be more heavily modulated by visual [9] than by auditory inputs, scientists naturally focused on the visual modality. Most of these previous studies have been reviewed elsewhere [5]; for various reasons, they failed to convince the larger scientific community, and the notion of discrete perception was gradually proscribed. In the past 10 years, however, significant experimental advances have occurred that somewhat restored the option of a discrete perceptual organization in the visual domain. These recent advances are reviewed in the following.

(a) The continuous wagon wheel illusion

In engineering, the term 'aliasing' refers to a potential artefact occurring when a signal is sampled by a discrete or periodic information processing system: if the sampling rate is lower than a critical limit (the Nyquist frequency), then the system's representation of the signal is inaccurate. A special case of aliasing occurs when the signal is a moving periodic visual pattern, and the information processing system is taking temporally discrete samples; in this case, the resulting aliasing has been termed the 'wagon wheel illusion', and is vividly experienced as the pattern seems to move in the wrong direction. This illusion is most commonly observed in movies or on television, owing to the periodic sampling of video cameras (generally around 24 frames or snapshots per second). But it is also possible to experience a similar effect under continuous conditions of illumination, such as in daylight [10–12]. This must imply that aliasing can also take place within the visual system itself. Thus, this continuous version of the wagon wheel illusion (or c-WWI) has been taken as evidence that the visual system samples motion information periodically [11–14].

This 'discrete' interpretation of the c-WWI is supported by several arguments. First, the illusory reversed motion is perceived only over a specific range of stimulus temporal frequencies, and this range is compatible with a sampling rate (the number of 'snapshots' per second) of approximately 13 Hz [11–13]. Second, the critical frequency range for the c-WWI was found to be largely independent of the spatial frequency of the stimulus [12,13] and of the type of motion presented (e.g. rotation versus translation, first-order versus second-order motion) [12]. Such an aliasing determined exclusively by the temporal properties of the stimulus is precisely what would be expected from a discrete sampling perceptual system. Third, during the c-WWI, there is only one frequency band of the EEG oscillatory spectrum that changes significantly, right in the same frequency range of approximately 13 Hz [15,16]. Altogether, these experimental findings converge towards the conclusion that the motion perception system (or at least part of it) samples information periodically, at a rate of approximately 13 samples per second.

Alternative interpretations of the c-WWI have also been put forward which do not rely on temporal subsampling and aliasing. Although all authors agree that the illusion is a bistable phenomenon, coming and going with stochastic dynamics as a result of a competition between neural signals supporting the veridical and the erroneous motion directions [17], most of the disagreement is now focused on the origin of the erroneous signals. While we assume that they arise from periodic sampling and aliasing, other authors have argued that they originate instead from spurious activation of low-level motion detectors [18,19] or from motion adaptation signals that would temporarily prevail over the veridical input [20,21]. We have argued, however, that this alternative account is incompatible with the available evidence. First, the c-WWI is maximal at around the same temporal frequency for first- and second-order motion patterns, whereas motion detectors in the brain have widely different temporal frequency response properties for the two types of motion [22]. Second, focused attention was found to be necessary for the c-WWI to occur [12]; furthermore, attention modulated not only the magnitude, but also the spatial extent and even the optimal temporal frequency of the c-WWI [23,24]. Although the absolute amount of motion adaptation could be assumed to vary with attentional load [25,26], there is no evidence to date that the frequency-tuning of motion adaptation (or of low-level motion detectors) can also be modified by attention. Third, motion adaptation can be strictly dissociated from the c-WWI: by varying stimulus contrast or eccentricity, it is possible to increase the amount of motion adaptation (as measured by both the static and the dynamic motion aftereffects) while decreasing the c-WWI, and vice versa [27]. Lastly, there is converging evidence that the neural correlates of the c-WWI primarily involve the right parietal lobe [15,28,29]; if the illusion was due to adaptation of low-level motion detectors, then its correlates would probably not be expected in such a high-level hierarchical region.

In conclusion, we believe that reversed motion signals in the c-WWI originate from attention-based motion perception systems that sample inputs periodically at approximately 13 Hz, thereby producing aliasing. At the same time, other motion perception systems (e.g. the low-level or 'first-order' system) continue to encode the veridical motion direction; it is the ensuing competition between these opposite signals that explains the bistability of the illusion.

(b) Ongoing electroencephalographic signatures of perceptual cycles

The c-WWI implies that a certain part of the visual sensory input (namely motion signals) can be sampled discretely or 'periodically'. One might predict, therefore, that it would be possible to record neural signatures of this sampling process in the form of a brain signal that waxes and wanes with every sample. Neuronal oscillations in various frequency bands are a natural candidate of choice for such a signature. Recently, our group and others have tested this prediction by assessing the influence of the phase of ongoing EEG oscillations (even

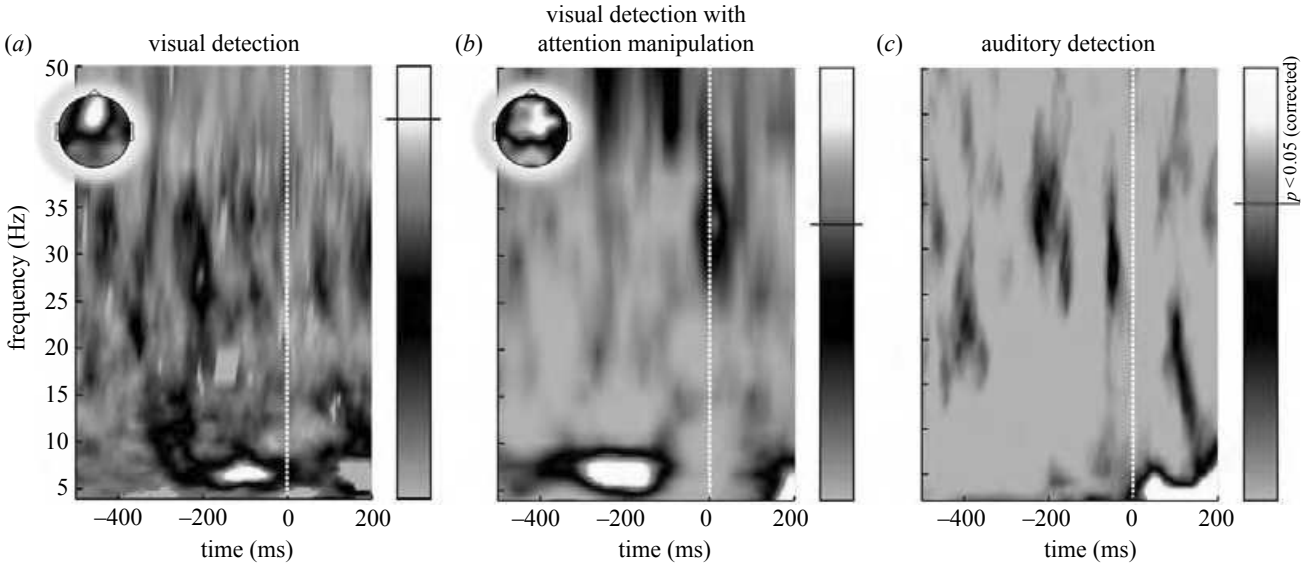


Figure 1. Pre-stimulus EEG signatures of visual, but not auditory perceptual cycles. The time–frequency maps represent the significance of 'phase opposition' between target-perceived and target-unperceived trials. This measure is determined by comparing the average phase-locking (or intertrial coherence, ITC) of each trial group with surrogate phase-locking values obtained over trial groups of the same size but randomly drawn among perceived and unperceived trials. A significant phase opposition at a particular time and frequency indicates that perceived and unperceived trials are associated with different phase values. (*a*) A pre-stimulus phase opposition was found at approximately 7 Hz in a visual experiment where subjects ($n = 12$) were free to attend to the target location (which was known in advance). The inset on the top-left represents the scalp topography of the pre-stimulus phase-opposition at this frequency, with a maximum around frontocentral electrodes. (*b*) The same result was replicated in a visual experiment with explicit attentional manipulation, but only for targets appearing on the attended side ($n = 13$). (*c*) However, no pre-stimulus phase opposition was found in an auditory experiment in which subjects ($n = 21$) were required to detect auditory clicks in a silent background. (No scalp topography is shown here as there were no significant pre-stimulus time–frequency points). (Online version in colour.)

before any stimulus is presented) on the subsequent perception of a visual stimulus (for a detailed review, see [30]).

In our first study [31], we presented dim flashes (6 ms long) in the visual periphery, with the luminance set around perceptual threshold. That is, only half of those flashes were perceived by the observers, while the other half remained unnoticed. The pre-stimulus phase locking on frontocentral electrodes was found to increase just before flash onset, for both the perceived and the unperceived trials (figure 1*a*). This effect occurred specifically for an EEG frequency band around 7 Hz, meaning that certain approximately 7 Hz pre-stimulus phase values facilitated the conscious perception of the flash, whereas other phase values impaired it. Indeed, when considering the phase of the 7 Hz band-pass-filtered EEG just before stimulus onset on each trial, we could predict the subsequent percept of the subject well above chance. In the same year, Mathewson *et al.* [32] also reported that the pre-stimulus phase of low-frequency oscillations (around 10 Hz) predicted the trial-by-trial perception of masked stimuli. Such a relationship between visual perception and the phase of spontaneous oscillations implies that visual inputs are not processed equally at all times, but periodically sampled by the visual system.

In our next study, we sought to determine the role of top-down attentional factors in this periodic sampling [33]. Because the target location was known in advance, we reasoned that subjects may have covertly attended to that location in order to improve their perceptual performance. Would ongoing oscillations still modulate target perception at an *unattended* location? In this new experiment, therefore, there were two possible target locations, and a central cue indicated before each trial the location at which subjects should pay attention. When the target appeared at that attended location, everything happened exactly as in the previous experiment, and indeed, we confirmed our previous results in this condition, with a strong impact of approximately 7 Hz pre-stimulus EEG phase on target perception (figure 1*b*). When the target appeared on the other, unattended side, however, the phase of ongoing oscillations had no effect on perception (data not shown here). In other words, ongoing EEG phase was related to visual perception solely by the implication of attention. We thus hypothesized that attention samples visual information periodically, and that each approximately 7 Hz ongoing EEG cycle is the signature of a new attentional sample [33]. This conclusion is well in line with another body of recent experimental work that will be reviewed in §2(*d*).

We have also applied the same generic method, identifying pre-stimulus EEG phase opposition between the different outcomes of a given cognitive process, to perceptual tasks other than the mere detection of a peripheral flash. For example, we recently showed that the phase of ongoing EEG oscillations at approximately 10 Hz can also predict the perception of a transcranial magnetic stimulation (TMS) phosphene, i.e. an illusory visual percept that follows the administration of a TMS pulse [34]. Similarly, we showed that saccadic reaction times to a peripheral target differed for different pre-stimulus 10–15 Hz EEG oscillatory phases [35]. The likelihood of identifying a target in a difficult search array (a T among Ls) was also found to depend on pre-stimulus oscillatory phase, this time at a slower frequency of approximately 6 Hz [36].

All these studies together seem to imply that there is an ongoing succession of 'good' and 'bad' phases for visual perception and attention, i.e. that perception and attention are intrinsically periodic or cyclic phenomena. As such, these studies constitute a solid initial body of evidence for the

notion of discrete perception. It might be argued, however, that a proper demonstration of discrete perception should involve more than just a cyclic fluctuation of sensory excitability. A truly discrete system, just as in the epitomic example of the video camera, should also exhibit a periodicity in the fine-grained perception of time itself, a so-called temporal 'framing'—meaning that two events separated by a given time interval would be perceived as occurring simultaneously or sequentially depending on whether they happened to fall within the same or distinct perceptual cycles. None of the experiments mentioned above can speak to this question, because they did not directly probe time perception. One such experiment on temporal framing was, in fact, published by Varela *et al.* [37] (see also [38]). They reported that the perception of two flashes separated by approximately 60–80 ms changed drastically as a function of the phase of the alpha rhythm (7–13 Hz) at which the first flash was presented; at one phase, they would be perceived as simultaneous, at the opposite phase as sequential. Unfortunately, this result has never been replicated, despite several attempts by our group and at least one other (D. Eagleman 2003, personal communication). Critically, however, one of our more recent experiments can also address this issue, albeit indirectly [39]. We examined the 'flash-lag' effect, a common illusion in which a steadily moving object is incorrectly perceived ahead of its true location at the moment of a flash [40]. The perceptual lag is generally accepted to reflect the time necessary for updating the conscious representation of the world after the 'flash' signal [40,41]. We showed that the trial-to-trial magnitude of this flash-lag effect systematically varied along with pre-stimulus 7–15 Hz EEG phase. That is, the oscillatory phase at (or just before) the moment of the flash determined whether an earlier or a later part of the ongoing motion sequence would be temporally grouped (or 'framed') with the flash. This may be the only solid evidence to date for a periodicity affecting not only sensory excitability, but also the fine-grained perception of time.

(c) Perceptual echoes

The various EEG experiments described in §2*b* indicate that ongoing brain oscillations create 'perceptual cycles' in which visual inputs are processed periodically. As a result, we were naturally led to ask the following questions. First, could these perceptual cycles be recorded not just before the time of stimulus presentation (i.e. in the ongoing EEG brain signals) but also afterwards, during stimulus processing itself (i.e. in the evoked EEG brain activity)? Second, for a visual event occurring at a particular instant, how many subsequent cycles would actually process the corresponding visual information? Do the perceptual cycles begin anew with each new sample, or do they also integrate the contents of past cycles, and if so, for how many successive cycles? We designed a simple experiment to answer both of these questions [42]. We presented a 'white noise' visual stimulus to our observers while recording their EEG activity. The stimulus was a static disc whose luminance varied randomly at each screen refresh. This random sequence of luminance intensity values had equal energy at all temporal frequencies (between 0 and 80 Hz, only limited by the 160 Hz refresh of the computer screen). We then cross-correlated the recorded EEG activity with the stimulus sequence on every trial, and averaged the results to obtain a cross-correlation function, describing the strength of correlation between the stimulus and the brain response recorded after a certain lag, for all successive values of the lag. One might have expected this cross-correlation function to resemble a classic visual-evoked potential (VEP) [43], a sequence of positive and negative deflections lasting about 300–500 ms [44,45]. Instead, we found a much longer-lasting response in the cross-correlation functions, which took the form of an approximately 10 Hz oscillation that extended, in many subjects, for 10 or more successive cycles (figure 4*a*). This oscillatory cross-correlation response implies that visual events in the world are represented cyclically in the brain, and that this periodicity is also visible in post-stimulus EEG activity. Furthermore, it indicates that a given instant in the world is not merely represented at one instant (or in one 'cycle') in the brain, but in several successive cycles. Arguably, this property could provide a significant contribution to the apparent continuity of our subjective experience.

(d) Periodic attentional sampling

Many of the perceptual periodicities described in §2*a–c* are tightly linked to visual attention. For example, the temporal sampling causing the wagon wheel illusion in continuous light (c-WWI) only occurs when attention is focused on the moving pattern [12]. Similarly, the phase of ongoing EEG oscillations only modulates the probability of detection for *attended* stimuli [33]. The ongoing EEG phase can also predict the likelihood of detecting a target in a difficult search array [36], an archetypal attentional function. The 10 Hz perceptual echoes were also shown to be enhanced by focused attention [42]. In other words, ongoing perceptual cycles in the brain could be attentional by nature. Is attention a cyclic process? This question becomes particularly interesting when multiple attentional targets must be monitored: in this case, does covert attention periodically sample the targets, just like our gaze, often dubbed 'overt' attention, would? Or does the attentional system process all of the targets in parallel? This is a question that has been vastly debated in the past few decades [46,47]. We have recently argued that discrete versus continuous perception and sequential versus parallel attention are but two facets of the same debate [48]. The cornerstone of this theory is that attention is intrinsically periodic (figure 2): when a single attentional target is present, this periodicity is expressed as a sequence of successive discrete samples of the unique target; when multiple targets are present, this periodicity naturally provides attention with a means to scan the targets in a sequential manner.

There are many recent pieces of experimental evidence in support of this notion. The idea that rhythmic attentional sampling could occur not just in the presence of multiple potential targets (a classic form of 'switching spotlight' [49–51]), but also for a single attended object (a notion we called 'blinking spotlight') originated in a 2007 study in which we modelled the effect of set size on psychometric functions for target detection as a function of target duration [52]. To summarize, we contrasted different models of attention and found this 'blinking spotlight' to explain human performance better than either the 'switching spotlight' or the 'parallel attention' models. The intrinsic sampling rate of attention was estimated around 7 Hz (in agreement with several subsequent EEG experiments, such as those illustrated in figure 1).

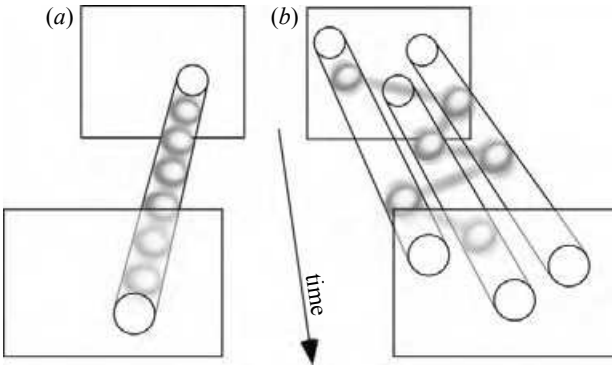


Figure 2. Discrete perception and sequential attention could reflect a unique periodic sampling mechanism. (*a*) A sensory process that samples a single visual input periodically illustrates the concept of discrete perception. (*b*) A sensory process that serially samples three simultaneously presented visual stimuli demonstrates the classic notion of a sequential or 'switching' attention spotlight. Because many of our findings implicate attention in the periodic sampling processes displayed in panel (*a*), we proposed that both types of periodic operations (*a*,*b*) actually reflect a common oscillatory neuronal process. According to this view, the spotlight of attention is intrinsically rhythmic, which gives it both the ability to rapidly scan multiple objects (as in the classic 'switching' spotlight), and to discretely sample a single source. This dual behaviour is what we refer to as a 'blinking' spotlight. (The yellow balls linked by red lines illustrate successive attentional samples). (Online version in colour.)

Recently, Landau & Fries [53] used another psychophysical paradigm in which they drew attention using a salient cue at one of two possible target locations. The observers reported target detection (a contrast decrement in a continuously moving pattern) at either of the two locations. After the salient cueing event, detection performance was found to oscillate at both locations, but in counter-phase such that optimal performance at one location coincided with minimal performance at the other. In other words, it again seemed that attention periodically and sequentially sampled the two locations, with an intrinsic sampling rate of about 7–10 Hz.

We have also used the c-WWI effect to address this question [24]. We varied the set size (number of simultaneously presented moving wheels) and the wheel(s) rotation frequency while asking observers to report any occurrence of reversed motion. As previously (see §2*a*), we found that reversals were most likely to happen in a specific range of temporal frequencies. For a single target wheel, the effect was compatible with aliasing caused by attentional sampling at approximately 13 Hz, exactly as in our previous studies. But when set size increased, the effective sampling frequency systematically decreased. When four wheels were present, illusory reversals still happened, but they were now compatible with each wheel being sampled at only approximately 7 Hz. One interpretation, in line with the idea of a 'blinking spotlight', is that the successive attentional samples, instead of repeatedly sampling the same wheel, were now sequentially exploring the different wheels (or a subset of them); as a result, each wheel experienced aliasing at a lower frequency.

(e) Conclusion: discreteness in visual perception, attention and awareness

It is becoming more and more evident that, in the visual domain, neural oscillations in the 7–13 Hz range have direct perceptual consequences that can be described as perceptual 'cycles'. This does not mean, of course, that higher-frequency oscillations, e.g. in the gamma range (30–80 Hz), do not influence perception, but these inherently more local oscillatory signals are less easily accessible to our EEG surface-recording methods. It is important to insist that it is not only sensory excitability that fluctuates cyclically at 7–13 Hz, but also higher-level perceptual representations involving visual attention, and possibly even visual awareness. There is, indeed, a tight relationship between these perceptual cycles and attentional processes, as reviewed in §2*d*. Attention is often considered as the gateway to consciousness [54,55], and it follows that if the gate opens periodically, the contents of awareness will also update periodically. Furthermore, we have described at least one instance in which the conscious perception of temporal simultaneity (i.e. which events in the world are experienced as a single 'group', a 'snapshot' or a mental 'frame') is constrained by the phase of ongoing oscillations [39]. This type of temporal framing is a hallmark of discrete perception, of the successive 'moments' of awareness [2].

3. Perceptual cycles in audition

After having reviewed the available evidence for discrete perception in vision, we now turn to the auditory system. It might be expected that the same experimental paradigms that helped uncover visual perceptual cycles could be similarly applied to audition to reveal its intrinsic discreteness. We may anticipate such auditory cycles to occur in the same frequency range as in vision (7–13 Hz), but this is not mandatory. In particular, because the frequency of visual cycles coincides roughly with the maximal range for steady-state visual-evoked responses (SSVEP) [56–58], one might predict that auditory cycles would occur instead around 40 Hz, which is the optimal frequency for auditory steady-state responses (ASSR) [59,60]. Another possibility, supported by certain theories of speech processing [61–65], could be that periodic auditory samples are taken at the same rate at which the relevant phonemic or syllabic events are expressed in normal speech, roughly between 2 and 8 Hz. Unfortunately, as we shall see, this straightforward approach of adapting our experimental paradigms to the auditory domain has not met with overwhelming success.

(a) No auditory wagon wheel illusion

In an initial attempt at directly translating the c-WWI paradigm (see §2*a*) to the auditory modality, we sought to measure the perceived motion direction of a spatially periodic sound source, such as a sound rotating around the listener through a circular array of speakers. By analogy to the illusion in the visual domain, we hoped to observe decreased perceptual performance, or even reversed motion perception, within a narrow range of temporal frequencies of the sound movement. Identifying this frequency of aliasing would then allow us to determine the intrinsic sampling frequency of the auditory system. It turned out, however, that such perceptual judgements of auditory sound motion can only be performed accurately at low temporal frequencies of sound movement, less than approximately 2–3 Hz (in agreement with previous reports [66,67]). If perceptual performance is already at chance at the hypothesized

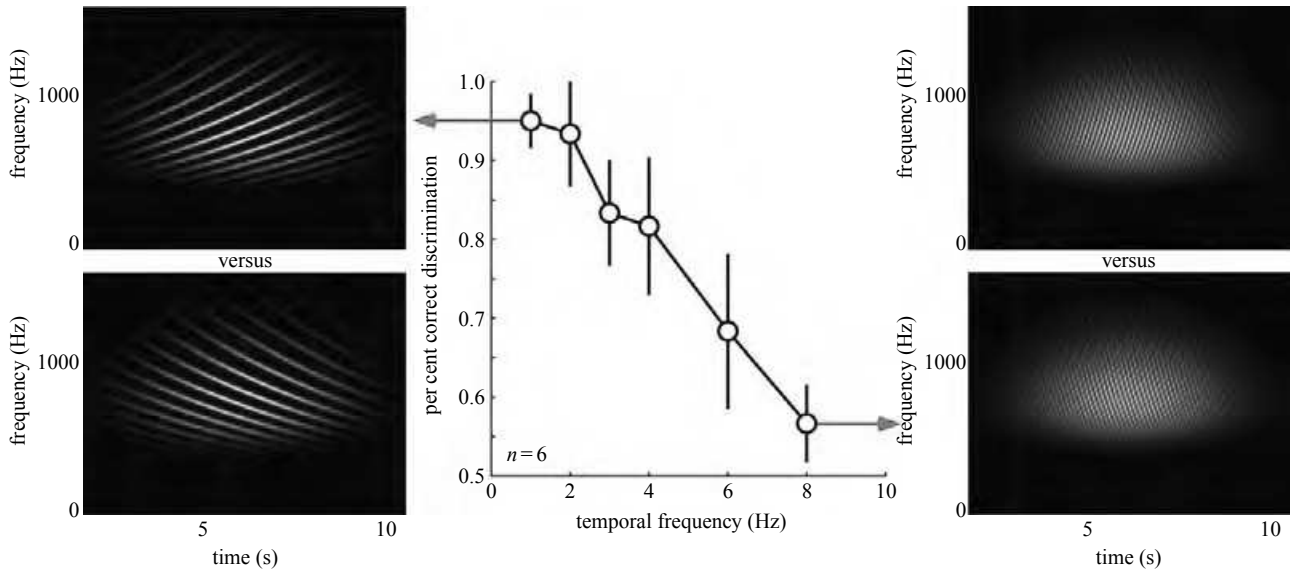


Figure 3. Auditory sensitivity for the direction of periodic sounds in the frequency domain. Shepard sequences [69] were created as a superposition of pure tone sweeps, increasing or decreasing in frequency over time. The sweep speed varied for different sequences: each tone increased or decreased logarithmically between the boundaries of audible space (set between 80 and 11025 Hz) over a fixed duration *T*; a new tone was inserted into the sequence (and an old tone disappeared) every *T*/40, such that at every instant 40 sweeping tones were simultaneously present. The temporal frequency of this periodic sound motion, therefore, was defined as TF = 40/*T*. In different trials, this temporal frequency was varied between 1 and 8 Hz. Each sequence was 12 s long, with a Gaussian amplitude profile in both time and frequency space, to limit the perception of artefacts linked to sequence onset/offset and tone insertion/disappearance, respectively. Examples of upward and downward sequences for TF = 1 Hz and TF = 8 Hz are presented as spectrograms in the figure (colour map indicates stimulus energy at each time and frequency), and the corresponding sound files can be downloaded from http://www.cerco.ups-tlse.fr/~rufin/audiovisual/. Participants ($n = 6$) were presented with 10 sequences of each temporal frequency in randomized order, moving up or down in frequency space (randomly determined with 50% probability), and asked to report their perceived motion direction (up/down) by pressing arrow keys on the keyboard. To limit the possibility of relying on perceived pitch differences between upward and downward sequences, the frequency-domain Gaussian amplitude envelope (s.d. 0.25 log units) was centred at one of three frequencies (700, 800 or 900 Hz), randomly chosen for each trial. The direction judgements were only accurate up to 3–4 Hz (*t*-test against 0.5 = chance-level, $p < 0.05$), and deteriorated rapidly at higher temporal frequencies. This low-pass sensitivity function critically limits the possibility of measuring a c-WWI effect in the auditory domain. Error bars represent standard error of the mean across subjects. (Online version in colour.)

frequency of aliasing, then this aliasing will simply not be observed—whether the perceptual process relies on periodic sampling or not. In other words, the only conclusion that can be drawn from this attempt is that, if auditory perceptual cycles exist, then they must occur at a rate faster than 3 Hz—hardly a revealing conclusion.

We then reasoned that sound frequency or 'pitch', rather than spatial position, may be the proper equivalent to the spatial location of visual objects. Indeed, the 'retinotopic' neuronal organization of early visual cortex is not found in the auditory system, where neurons are instead organized in a 'tonotopic' manner [68]. Thus, we designed periodic stimuli that moved in particular directions in the frequency domain—so-called Shepard or Risset sequences [69]. Again, we were disappointed to find that the direction of these periodic frequency sweeps could not be reliably identified when the temporal frequency of presentation was increased beyond 3–4 Hz (figure 3; no temporal aliasing is visible, i.e. no performance below chance or local minimum in performance).

In sum, although temporal aliasing (as measured in the c-WWI) is, in principle, a choice paradigm to probe the rhythms of perception, our attempts so far at applying this technique to the auditory domain have been foiled by the strict temporal limits of auditory perception. Of course, the auditory system is widely regarded as a temporally precise one, but this precision observed for specific auditory features (discrimination of nearby pitch frequencies, interaural time delays) does not extend to periodic sound motion, either in the spatial or in the frequency domains. This limitation precludes using the wagon wheel phenomenon to determine the sampling rate of audition or possibly, the absence of discrete auditory sampling. What we can safely conclude is that, if discrete sampling exists in audition, then it must be at a sampling rate above 3–4 Hz (if aliasing occurred at frequencies below 3 Hz, then it would have been possible for our subjects to perceive it, and they would have systematically reported reversed motion).

(b) No ongoing electroencephalographic signatures of auditory perceptual cycles

One major and undisputable piece of evidence in favour of ongoing perceptual cycles in vision is the finding that the conscious detection of a flash at luminance threshold fluctuates along with the phase of ongoing EEG oscillations (see §2*b*). Similarly, a dependence of auditory detection on ongoing EEG phase would indicate the existence of ongoing auditory perceptual cycles. We attempted to measure this relation by presenting threshold-intensity 'clicks' (0.5 ms square wave pulses) in a silent environment, and asking participants ($N = 21$) to report their perception via button-presses. Upon applying the same time-frequency phase opposition analysis techniques (figure 1*a*,*b*) as in our previous visual experiments [30,31,33], we were unable to reveal any systematic relationship between pre-stimulus phase and auditory perception in any frequency band (figure 1*c*). A similar negative report was independently published by Zoefel & Heil [70].

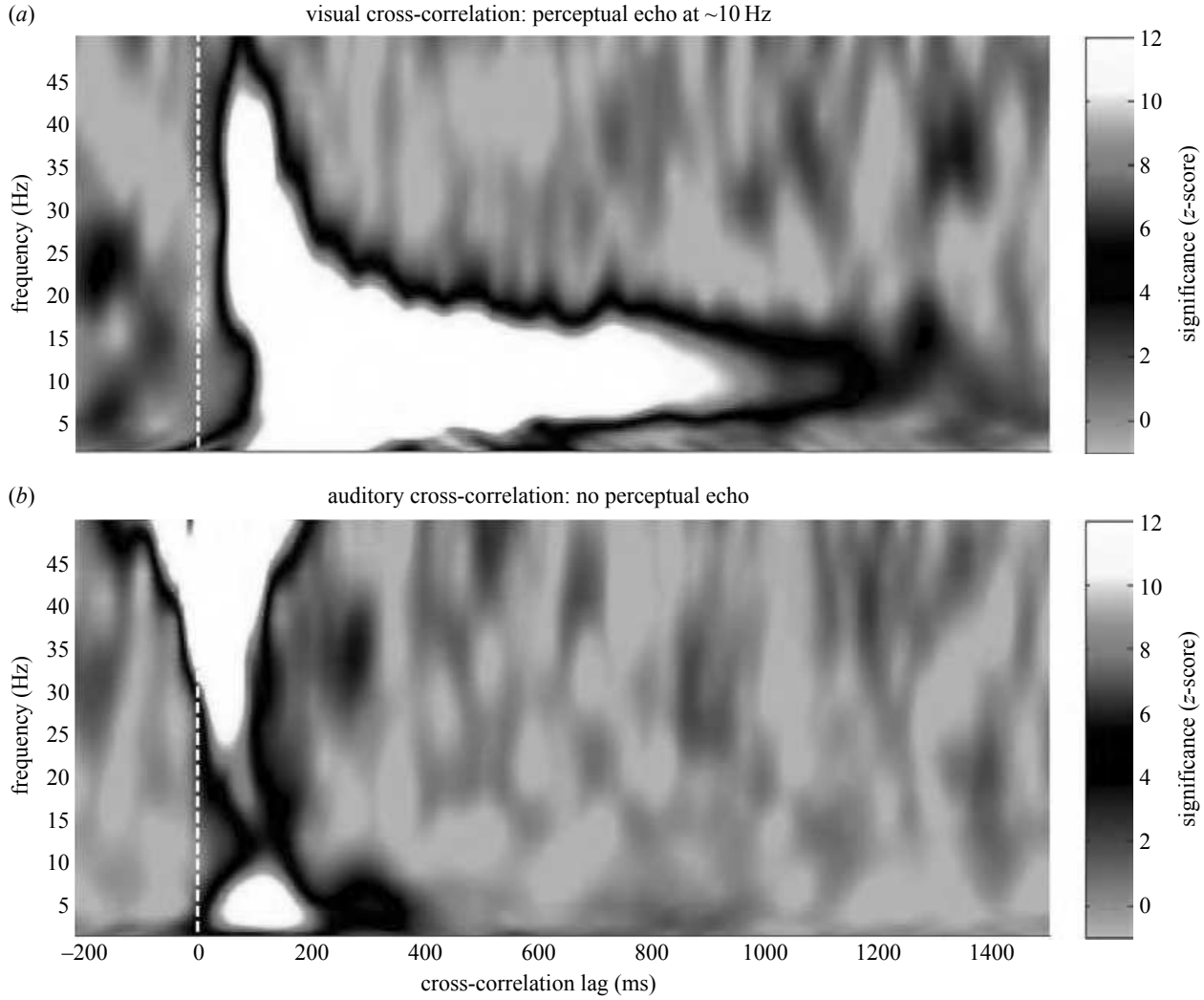


Figure 4. Perceptual echoes in the visual, but not in the auditory modality. Each panel is a time–frequency representation of the cross-correlation function between a white noise stimulus sequence and the simultaneously recorded EEG response. The cross-correlation is computed for several lags between the stimulus and EEG signals; then, a time-frequency transform is applied separately for each subject ($n = 12$); the grand-average results are expressed as a *z*-score (comparison against surrogate cross-correlation functions obtained by randomizing stimulus–EEG pairings). (*a*) When the white noise stimulus sequence reflects the changing luminance of a disc in the visual field, after a transient broadband response for time lags below 200 ms, a long-lasting reverberation (up to lags of approx. 1 s) is observed, peaking at approx. 10 Hz. (*b*) When the white noise stimulus sequence encodes the changing loudness of an auditory pure tone (1000 Hz carrier frequency), the transient broadband response is present, but no subsequent reverberation is observed in any frequency range. The increasing width of the transient response above 30 Hz is likely due to the auditory middle latency response (MLR) [90], a short-lived auditory potential (<50 ms) which appears smeared in time owing to our wavelet time–frequency transform (using an eight-cycle window length at 50 Hz). The same 12 subjects participated in the visual and auditory experiments. (Online version in colour.)

It appears that auditory perceptual cycles, if they exist, cannot be detected with the very same experimental method that has successfully and repeatedly allowed us to reveal periodicities in visual perception. One critical aspect of this method was the presentation of auditory targets in a silent environment—the auditory equivalent of a flash in the dark. In fact, Ng and colleagues recently reported that auditory perception does vary with the phase of 2–6 Hz (theta-band) EEG oscillations when the target sound is embedded in an ongoing sequence made up of several superimposed naturalistic background noises [71]. However, the use of background noise in this study also implies that the relevant EEG oscillations cannot be considered as spontaneous or ongoing signals any more, but are instead driven or entrained by the background noise [72,73].

Phase entrainment to auditory streams has been demonstrated in many previous studies using rhythmic background sounds at delta (1–4 Hz) and theta (2–8 Hz) frequencies, and auditory detection performance was found to covary with the entrained oscillatory phase [74–77]. This phase entrainment mechanism is thought to serve a critical role in speech perception [64,78–85] by aligning the optimal oscillatory phase to the peaks of the speech envelope (which also recur at a frequency roughly between 2 and 6 Hz), and thereby enhancing speech intelligibility [65,78,86,87]. However, it is difficult in such entrainment studies (even the one by Ng *et al*. [71] in which the entraining background noise contained energy in several frequency bands, including theta), to tease apart the contribution of low-level physical differences in the entraining sound to the perceptual changes recorded at different theta phases. Because the EEG is entrained by (or 'phase-locked' to) the background stimulus [88], different EEG phases directly correspond to different moments in the background entraining sound, with

systematic differences in auditory properties (such as loudness and pitch); in turn, these varying physical properties can conceivably affect target detection probability (e.g. through masking or contextual enhancement phenomena). In this way, a rhythmic background sound can both entrain EEG oscillations, and modulate auditory detection in a periodic fashion. Yet, the perceptual modulation in this case is not *intrinsically* periodic: should the stimulus amplitude profile resemble, say, the outline of Mount Everest or the New York City skyline, so would the listener's perceptual performance. In other words, the existing evidence so far is insufficient to decide whether the frequently observed theta-band periodic fluctuations of auditory performance reflect an intrinsic periodicity of the auditory system (i.e. true perceptual cycles) or an intrinsic periodicity of the auditory environment (or both).

To conclude, contrary to what we have observed in the visual modality, it would appear that the presence of an entraining (and ideally, rhythmic) auditory background stimulus may be a necessary condition to observe rhythmic fluctuations in auditory perception [73,76,89]. Even then, owing to the possibility of low-level confounds, it is not yet evident that such entrained rhythmic fluctuations can be considered as a signature of 'entrained' auditory perceptual cycles. It is likely, on the other hand, that purely ongoing or spontaneous oscillations (i.e. those recorded in silence) do not reflect an ongoing auditory sampling process, as they do in vision.

(c) No auditory perceptual echoes

The cross-correlation paradigm that allowed us to reveal perceptual echoes in vision (§2*c*) could prove a useful tool to test the hypothesis of 'entrained' (in opposition to 'ongoing') auditory perceptual cycles. Indeed, this paradigm is designed to reveal the resonance properties of a sensory system, that is, whether it presents a frequency-specific response (an 'echo', which is also a form of phase entrainment) during a white noise stimulation sequence. In the visual system, this echo was found around 10 Hz, and lasted for up to 10 cycles (figure 4*a*). If one assumes that auditory perceptual cycles exist, but are only active when they can be entrained by a background sound (an assumption suggested by the data reviewed in §3*b*), then they may be expected to show up as an auditory echo in this cross-correlation paradigm. More precisely, one might predict observing a resonance in the theta-frequency range, in accordance with the numerous theta-phase entrainment results described in §3*b* (and in particular the strong theta-frequency periodicity of human speech signals and human speech processing mechanisms). Another (non-exclusive) hypothesis could be that auditory echoes occur in the gamma-frequency band, around 40 Hz: indeed, while alpha (approx. 10 Hz) is the optimal visual stimulation frequency to produce an SSVEP [56–58], gamma (approx. 40 Hz) is the optimal frequency for ASSR [59,60]. A direct auditory equivalent to our approximately 10 Hz visual echoes could thus also be expected around 40 Hz.

Unfortunately, no significant auditory perceptual echo was detected in our experiments [91], either in the theta nor in the gamma range, or in any other frequency band (figure 4*b*). While definite, this absence still does not disprove the existence of perceptual cycles in the auditory system for at least two reasons. First, although echoes were associated with perceptual cycles in vision, this association is not mandatory: a reverberation and integration of sensory information over several cycles is likely to be detrimental to auditory perception, so audition may instead rely on cycles that are more temporally independent (i.e. 'short-lived' echoes). In this case, figure 4*b* (and much of the existing literature [62,64,80]) suggests that the cycles may occur in the theta (2–8 Hz) and/or gamma (30–80 Hz) frequency ranges. Second, the absence of long-lasting auditory echoes in our experiment merely indicates that perceptual sampling and reverberation do not affect the processing of auditory loudness (the sensory feature that varied in our white noise sequences), even though they affect the visual equivalent, luminance perception. It is still possible, however, that perceptual sampling and reverberation could involve higher hierarchical levels of representation, after the extraction of basic auditory features. In accordance with this idea, oscillations have been repeatedly shown to contribute to speech perception by temporally framing the input stream according to the speech envelope [64,65,78–87]. This suggests that auditory echoes, absent with low-level stimuli such as amplitude-modulated pure tones, may still be observed with stimuli having more complex semantic content, such as speech. In that case, we predict that they should be visible around theta or gamma frequencies.

4. Different sensory inputs, different rhythmic sampling strategies

So far, all of the experimental paradigms that have succeeded in demonstrating visual perceptual cycles have also failed at revealing the auditory equivalent. Short of embracing the conclusion that perceptual cycles simply do not exist in the auditory domain, we must contemplate the possibility that these cycles could be implemented in very different ways in the two systems and thus may not be responsive to the same experimental approaches. In particular, it might prove useful to consider the different computational requirements with which each sensory modality is faced in terms of statistical properties and temporal structure of their respective sensory inputs, as well as their respective anatomical and functional architectures. This could help us explain why a processing strategy that is efficient for visual inputs may not be directly applicable to auditory inputs. In particular, we suggested above (§3*c*) that in audition, contrary to vision, perceptual sampling and reverberation could be restricted to higher hierarchical levels of representation, after the extraction of basic auditory features. There are two arguments to support this hypothesis: first, directly subsampling an auditory input stream (after conversion to the wavelet domain) has much more devastating consequences than the equivalent temporal subsampling of a visual input stream; second, a great deal of auditory feature extraction takes place subcortically, whereas visual processing is predominantly a cortical phenomenon. These arguments are developed in the following sections.

(a) Perceptual effects of visual versus auditory subsampling

In the visual environment, important events and changes tend to occur on a relatively slow time-scale. If one were to take two pictures of the same scene, separated by 150 ms

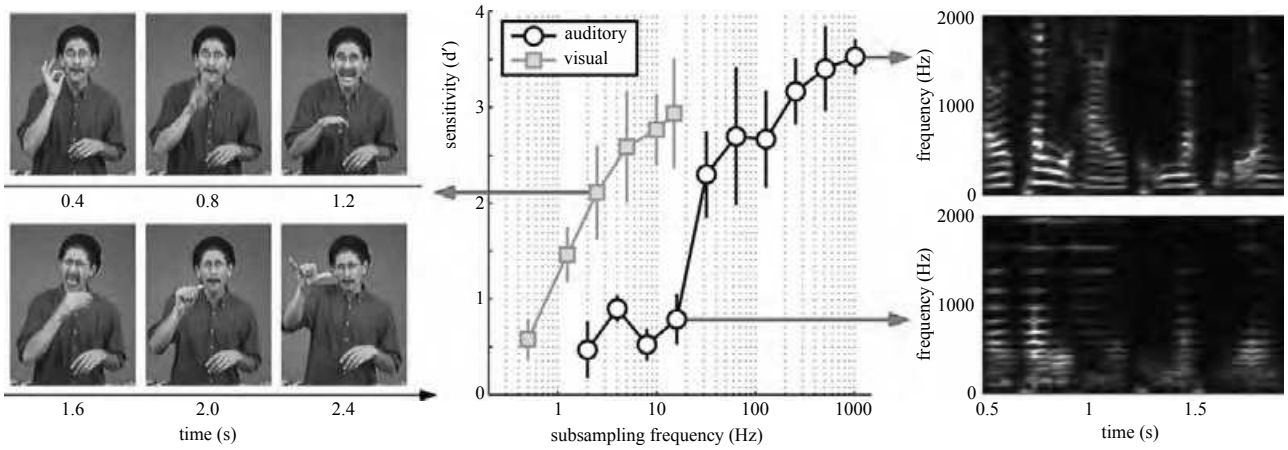


Figure 5. Auditory and visual vulnerability to input stream subsampling. The same participants ($n = 4$) watched and listened to sequences of 3 s long video and audio snippets (respectively) in different blocks while performing a two-back task (responding to a repeat of the penultimate snippet). Video snippets depicted an actor reading a children's book in American sign language, whereas audio snippets were recordings of a speaker reading an English literary classic. In each block, all snippets were temporally subsampled to the same frequency. A 2 s long excerpt from a video snippet at 2.5 frames per second is illustrated on the left, and the spectrograms of a single 2 s long excerpt from an audio stream subsampled at 1024 Hz (top) and 16 Hz (bottom) are shown on the right. The corresponding video/sound files can be downloaded from http://www.cerco.ups-tlse.fr/~rufin/audiovisual/. The two-back recognition task performance is expressed in terms of sensitivity (d', corresponding to the z-scored difference between hit rates—correctly detecting a two-back repeat—and false alarm rates—incorrectly reporting a two-back repeat). It is an order of magnitude more robust to temporal subsampling for vision than for audition. Error bars represent standard error of the mean across subjects. (Online version in colour.)

(about one-seventh of a second), most if not all of the scene would likely remain unchanged between them. Movement on a biological time-scale (e.g. human actions and displays of emotions, the displacement of preys or predators) will result in only minor differences between the two pictures; furthermore, these inconsistencies can easily be recovered by temporal 'interpolation' (and indeed, the so-called apparent motion mechanisms in the brain seem to excel at this task [92]). Only the rapid movement of a spatially periodic stimulus (such as a wheel) could conceivably create a difficult 'correspondence problem' between the two images [93], but this is admittedly a rare situation (this situation describes, in fact, the temporal aliasing discussed in §2*a*). In other words, our visual system may be fairly robust to temporal subsampling of the visual environment. By contrast, auditory stimuli are defined mainly as temporal fluctuations: vocal or musical pitch, speech phoneme distinction or speech recognition all require processing fine-grained temporal information in different frequency ranges. This has moved certain authors to propose that the time dimension in audition could be equivalent to the spatial dimension in vision [94]. A periodic sampling or sensory reverberation of the auditory input stream could therefore dramatically alter signal intelligibility.

There are a number of existing studies reporting subjective judgements of video quality at different sampling rates [95,96], converging to the conclusion that frame rates above approximately 5 Hz are generally deemed acceptable. To the best of our knowledge, there is, however, no equivalent data on the perceptual effects of temporally subsampling the auditory input stream, and no direct comparison of the two modalities using the same task in the same subjects. We therefore implemented such a comparison in a new experiment (figure 5). Our comparison approach was voluntarily naive. We subsampled the visual and auditory inputs in a representation space roughly equivalent to the first sensory stage of each system: the retina (with the entire image representing a subsampling 'frame') and the cochlea (with the instantaneous complex frequency spectrum resulting from a wavelet decomposition of the audio signal as a subsampling 'frame'), respectively. That is, for both sensory systems, we evaluated the consequences of the most severe possible temporal subsampling strategy, by subsampling the very input to the system; then we simply asked 'will the system be able to cope?' Of course, a positive answer does not imply that perceptual cycles actually occur at this frequency, but a negative answer casts serious doubt on this idea. In addition, any difference in sensitivity between the two sensory modalities can inform us about viable strategies for each system.

We hasten to mention that past studies have investigated the influence of temporal distortions on auditory perception and more particularly on speech processing [63,65,81,97–103], converging on the notion that audition can withhold temporal degrading of speech envelopes down to 16 Hz or even lower (approx. 4 Hz in [99]). But none of the distortion methods used was equivalent to a strict temporal subsampling of audio inputs. For example, the now classic 'Shannon' method [98] consists of low-pass filtering the audio signal envelope, and does it independently for several separate spectral bands.

In our experiment, one original 10 min video and one original 10 min audio sequence were used as the primary stimuli. The audio sequence was a 8000 Hz recording of a male native English speaker reading aloud an English literary classic. The video sequence was a 30 frames s^{-1} (silent) recording of a male actor reading a children's book in American sign language, shot from a static camera angle. Both audio and video recordings were cut into 3 s long 'snippets'. The snippets were contiguous excerpts that did not take into account the structure of the story. Although some snippets were certainly more informative than others, before temporal subsampling they were all intelligible or visually distinct (the participants had no prior experience with

American sign language). These 200 snippets were presented (separated by 1 s blank intervals) in a randomized order to four human observers and listeners who were instructed to perform a two-back task: indicate by a button press any snippet that matched the one presented two snippets ago. These two-back repeats occurred randomly with a probability of 33%.

Audio and video snippets were presented in separate blocks of 30 snippets, and in each block a different temporal subsampling was applied. For video subsampling at frequency TF, we selected a subset of frames (one frame every 30/TF, rounded to the nearest frame), and simply played the videos with a frame rate set to TF. For every snippet and subsampling frequency, two subsampled versions were created by starting the frame subset selection either on the first frame, or on the nearest frame to $1 + 30/TF/2$. Whenever a two-back repeat occurred in the sequence, it was always between distinct subsampled versions (this was done to prevent the use of static information for recognition). For audio subsampling at frequency TF, we first converted the snippets into the wavelet domain to approximate cochlear transduction (continuous Morlet wavelet transform of order 6). Discrete samples were taken every 8000/TF point, and all points between the samples were replaced with a linear interpolation of the two surrounding samples. Both amplitude and phase of the complex wavelet coefficients were interpolated to avoid artefacts created by discrete phase transitions. As for video subsampling, we created two subsampled versions of each audio snippet by starting the samples on the first data point, or the nearest point to $1 + 8000/TF/2$. Finally, we converted the signals back to the time domain via the inverse wavelet transform.

As expected, we found a dramatic difference between the two modalities' sensitivity to temporal subsampling (figure 5). While visual performance only started to deteriorate below 2.5 frames per second, auditory performance suffered for all subsampling frequencies below 32 Hz. That is, audition was about an order of magnitude more vulnerable to this subsampling than vision.

The observed difference in temporal robustness may explain, in part, why the auditory system does not sample incoming information as the visual system does. In the visual system, we have suggested that ongoing sampling induced by brain oscillations could take place at frequencies between 7 and 13 Hz (§2). As can be appreciated from figure 5, little information is lost by directly subsampling visual inputs in this frequency range. On the other hand, directly subsampling the auditory inputs in the same frequency range has dramatic consequences: the fine temporal structure is irremediably lost, and the signals cannot be recovered (even through temporal interpolation, which was an integral part of our auditory subsampling procedure in the wavelet domain). This may be an argument for the notion that auditory sampling involves higher oscillatory frequencies, for example in the gamma range [59,62,80,104–106]. Yet our results do not imply that brain oscillations at lower frequencies have no bearing on auditory perception. As mentioned before, there are still two possible (and non-exclusive) oscillatory sampling strategies involving lower frequencies that could remain compatible with these data: first, by sampling auditory representations not in an 'ongoing' manner (a regular succession of samples, blind to the temporal structure of the inputs) but in a more flexible manner, 'entrained' by the temporal structure; second, by sampling auditory representations not at the input level (e.g. cochlea or subcortical nuclei) but at a higher hierarchical level (e.g. auditory cortex).

In an attempt to address the former possibility, we repeated the above auditory subsampling experiment, this time comparing two modes of audio input subsampling: ongoing or 'blind' subsampling, as before, and entrained or 'flexible' subsampling. To create these 'flexible' subsampling stimuli, we first extracted the speech envelope of each snippet (weighted average of instantaneous signal energy across frequencies weighted by the average human cochlear sensitivity). Instead of selecting regular sampling points throughout the snippet ('blind' sampling), we distributed the same number of sampling points at the peaks and troughs of the 2–8 Hz band-pass-filtered speech envelope (starting with the highest peak and its immediately preceding trough; adding peak/trough pairs in decreasing order of peak amplitude; in case more sampling points were available than the number of peaks and troughs in the speech envelope, the remaining points were assigned so as to minimize the maximal sampling interval duration). In sum, this flexible subsampling kept the same average sampling rate as for blind sampling, but concentrated the samples at those moments where phonetic information was maximal. Yet we found no significant difference in the sensitivity of human listeners ($n = 7$) between the 'blind' and the 'flexible' subsampling of the input stream at frequencies between 8 and 64 Hz (two-way ANOVA with factors 'frequency' = (8,16,32,64 Hz) and 'sampling type' = [blind, flexible]; main effect of frequency $F_{3,48} = 17.45$, $p < 0.0001$, no main effect of sampling type or interaction, $p > 0.5$; data not shown). That is, audition remains an order of magnitude more vulnerable to temporal subsampling of its inputs than vision, even for a 'flexible' auditory subsampling. This finding definitely rules out the possibility that sampling at lower frequencies (less than 30 Hz) could occur early in auditory processing, since neither ongoing ('blind') nor entrained ('flexible') subsampling applied directly to the input stream would leave enough temporal information for further processing. In addition, it is worth noting that early subcortical auditory structures can display exquisite temporal resolution (greater than 100 Hz) that seems incompatible with temporal subsampling [107].

The last remaining option to rescue the notion of auditory perceptual cycles is, therefore, that they could sample auditory representations at a higher hierarchical level, after the stage of auditory feature extraction: such representations are more stable temporally, and would suffer less from a moderate loss of temporal resolution. This strategy is, in fact, the one used in modern speech compression techniques or 'vocoders' (e.g. LPC, MELP or CELP [108]) that extract phonetic features from high temporal resolution signals, but can then transmit the features in (lower resolution) temporally discrete packets or 'frames'. In future work, it may be interesting to apply temporal subsampling (either 'blind' or 'flexible' subsampling) to the output of one of these vocoders and test human auditory recognition in the same way as above: we predict that the auditory system may prove significantly more robust to this subsampling of higher-level representations.

(b) Differences in hierarchical organization

As mentioned previously, there are important architectural differences between the auditory and visual processing

hierarchies. Without going too deep into anatomical details, the most relevant discrepancy for our purposes can be summarized as follows: visual perception (even for salient low-level features such as luminance and spatial localization) depends, in great part, on cortical activity, whereas auditory stimuli reach primary auditory cortex after an already extensive processing by subcortical structures [109,110].

Consequently, applying an *architecturally* similar perceptual sampling strategy in the two systems (perceptual cycles that sample sensory representations at a similar *cortical* level, possibly under the influence of attention) could then have very different *functional* consequences, compatible with what we have observed experimentally. Apparently simple visual tasks (e.g. flash detection in the dark; figure 1) would suffer periodic fluctuations in performance, but equivalent low-level auditory tasks (e.g. click detection in silence) would appear continuous, because their outcome can be determined on the basis of subcortical representations, prior to any perceptual sampling. Perceptual cycles would only be observed with higher-level auditory stimuli such as music or speech that are not differentially processed at a subcortical level and thus require cortical activation for efficient discrimination. In agreement with this idea, a recent study demonstrated that arbitrary white noise auditory stimuli could elicit theta-band phase entrainment, but only after sufficient exposure, presumably turning the meaningless patterns into meaningful auditory objects [111]. Note finally that this reasoning remains compatible with the postulated role for attention in perceptual cycles (§2*d*), because attention is primarily a cortical function in both visual and auditory modalities [112–114]. Until direct anatomo-functional evidence is uncovered, we prefer not to speculate on whether primary sensory areas (both auditory and visual) participate or not in this 'high-level' periodic perceptual sampling.

(c) Conclusion: ongoing visual attentional cycles, entrained auditory attentional cycles?

To recapitulate, we can now critically evaluate the possible existence of perceptual cycles in the two modalities on the basis of experimental evidence reviewed in previous sections. We organize this evidence for perceptual cycles along two dimensions of interest, that is (i) whether they sample hierarchically 'early' or 'late' representations, and (ii) whether they sample in an 'ongoing' (blind, stimulus-independent) manner or in an 'entrained' (flexible, stimulus-dependent) manner.

Visual perception samples sensory representations at approximately 7–13 Hz (§2). Vision could afford to sample at a hierarchically early level, because it is robust to input subsampling down to at least 5 Hz (§4*a*). But it probably does not. Indeed, we have seen that visual perceptual cycles are mainly an attentional phenomenon (§2*d*) and this is more in line with a high-level (or at least a cortical) sampling (§4*b*). Vision samples in an ongoing manner, mostly blind to the stimulus content (§2*b* and §4*a*). Yet it also has the capability to entrain to and resonate with stimuli that contain an appropriate rhythmic structure (§2*c*; [56,57]). The periodicity of visual perception, in sum, is best described as an ongoing series of attentional cycles at approximately 7–13 Hz.

Audition does not sample at a hierarchically early level in either an ongoing way (§3*b* and §4*a*) or an entrained (flexible) way (§3*c* and §4*a*). If it does sample, then it must be at a higher level, that is, a *cortical* level (§4*b*). Many reports of attentional control of phase entrainment [74,75,115] suggest that the sampling may also be attentional, as in the visual system. Can this high-level sampling be an ongoing process as in vision, or must it be entrained by the temporal structure of auditory inputs? The lack of ongoing EEG phase influence in audition (§3*b*) as well as the finding that phase entrainment strongly facilitates intelligibility [65,78,86] compels us to favour the latter alternative, just like other authors have recently argued [73,76,89].

To sum up, *if* perceptual cycles exist in audition, then they must be a relatively high-level or attentional phenomenon (as in vision), and they must proceed by stimulus entrainment (contrary to vision). Based on numerous studies of rhythmic entrainment and speech processing, we believe that the cycles are most likely to be observed in the theta-frequency range (though gamma-frequency sampling cannot be categorically ruled out). But the big 'if' lingers. Definite evidence for auditory perceptual cycles is still lacking.

Acknowledgements. We are grateful to Daniel Pressnitzer and Leila Reddy for providing useful comments on the manuscript.

Funding statement. This work was supported by a EURYI Award to R.V. and a Studienstiftung des deutschen Volkes scholarship to B.Z.

References

1. Pitts W, McCulloch WS. 1947 How we know universals: the perception of auditory and visual forms. *Bull. Math. Biophys.* **9**, 127–147. (doi:10.1007/BF02478291)
2. Stroud JM. 1956 The fine structure of psychological time. In *Information theory in psychology* (ed. H Quastler), pp. 174–205. Chicago, IL: Free Press.
3. Harter MR. 1967 Excitability cycles and cortical scanning: a review of two hypotheses of central intermittency in perception. *Psychol. Bull.* **68**, 47–58. (doi:10.1037/h0024725)
4. Allport DA. 1968 Phenomenal simutaneity and the perceptual moment hypothesis. *Br. J. Psychol.* **59**, 395–406. (doi:10.1111/j.2044-8295.1968.tb01154.x)
5. VanRullen R, Koch C. 2003 Is perception discrete or continuous? *Trends Cogn. Sci.* **7**, 207–213. (doi:10.1016/S1364-6613(03)00095-0)
6. Durgin FH, Tripathy SP, Levi DM. 1995 On the filling in of the visual blind spot: some rules of thumb. *Perception* **24**, 827–840. (doi:10.1068/p240827)
7. Di Lollo V, Wilson AE. 1978 Iconic persistence and perceptual moment as determinants of temporal integration in vision. *Vision Res.* **18**, 1607–1610. (doi:10.1016/0042-6989(78)90251-1)
8. Berger H. 1929 Über das Elektroenkephalogramm des Menschen. *Archiv. Psychiatr. Nervenkrankheiten* **87**, 527–570. (doi:10.1007/BF01797193)
9. Chapman RM, Shelburne Jr SA, Bragdon HR. 1970 EEG alpha activity influenced by visual input and not by eye position. *Electroencephalogr. Clin. Neurophysiol.* **28**, 183–189. (doi:10.1016/0013-4694(70)90186-0)
10. Schouten JF. 1967 Subjective stroboscopy and a model of visual movement detectors. In *Models for the perception of speech and visual form* (ed. I Wathen-Dunn), pp. 44–45. Cambridge, MA: MIT Press.

11. Purves D, Paydarfar JA, Andrews TJ. 1996 The wagon wheel illusion in movies and reality. *Proc. Natl Acad. Sci. USA* **93**, 3693–3697. (doi:10.1073/pnas.93.8.3693)
12. VanRullen R, Reddy L, Koch C. 2005 Attention-driven discrete sampling of motion perception. *Proc. Natl Acad. Sci. USA* **102**, 5291–5296. (doi:10.1073/pnas.0409172102)
13. Simpson WA, Shahani U, Manahilov V. 2005 Illusory percepts of moving patterns due to discrete temporal sampling. *Neurosci. Lett.* **375**, 23–27. (doi:10.1016/j.neulet.2004.10.059)
14. Andrews T, Purves D. 2005 The wagon-wheel illusion in continuous light. *Trends Cogn. Sci.* **9**, 261–263. (doi:10.1016/j.tics.2005.04.004)
15. 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. (doi:10.1523/JNEUROSCI.4654-05.2006)
16. Piantoni G, Kline KA, Eagleman DM. 2010 Beta oscillations correlate with the probability of perceiving rivalrous visual stimuli. *J. Vis.* **10**, 18. (doi:10.1167/10.13.18)
17. Blake R, Logothetis NK. 2002 Visual competition. *Nat. Rev. Neurosci.* **3**, 13–21. (doi:10.1038/nrn701)
18. Kline K, Holcombe AO, Eagleman DM. 2004 Illusory motion reversal is caused by rivalry, not by perceptual snapshots of the visual field. *Vision Res.* **44**, 2653–2658. (doi:10.1016/j.visres.2004.05.030)
19. Holcombe AO, Clifford CW, Eagleman DM, Pakarian P. 2005 Illusory motion reversal in tune with motion detectors. *Trends Cogn. Sci.* **9**, 559–560. (doi:10.1016/j.tics.2005.10.009)
20. Holcombe AO, Seizova-Cajic T. 2008 Illusory motion reversals from unambiguous motion with visual, proprioceptive, and tactile stimuli. *Vision Res.* **48**, 1743–1757. (doi:10.1016/j.visres.2008.05.019)
21. Kline K, Eagleman DM. 2008 Evidence against the temporal subsampling account of illusory motion reversal. *J. Vis.* **8**, **13** 11–15.
22. Hutchinson CV, Ledgeway T. 2006 Sensitivity to spatial and temporal modulations of first-order and second-order motion. *Vision Res.* **46**, 324–335. (doi:10.1016/j.visres.2005.03.002)
23. VanRullen R. 2006 The continuous wagon wheel illusion is object-based. *Vision Res.* **46**, 4091–4095. (doi:10.1016/j.visres.2006.07.030)
24. Macdonald JSP, Cavanagh P, VanRullen R. 2013 Attentional sampling of multiple wagon wheels. *Attent. Percept. Psychophys.* **46**, 4091–4095.
25. Chaudhuri A. 1990 Modulation of the motion aftereffect by selective attention. *Nature* **344**, 60–62. (doi:10.1038/344060a0)
26. Rezec A, Krekelberg B, Dobkins KR. 2004 Attention enhances adaptability: evidence from motion adaptation experiments. *Vision Res.* **44**, 3035–3044. (doi:10.1016/j.visres.2004.07.020)
27. VanRullen R. 2007 The continuous wagon wheel illusion depends on, but is not identical to neuronal adaptation. *Vision Res.* **47**, 2143–2149. (doi:10.1016/j.visres.2007.03.019)
28. VanRullen R, Pascual-Leone A, Battelli L. 2008 The continuous wagon wheel illusion and the 'when' pathway of the right parietal lobe: a repetitive transcranial magnetic stimulation study. *PLoS ONE* **3**, e2911. (doi:10.1371/journal.pone.0002911)
29. Reddy L, Remy F, Vayssiere N, VanRullen R. 2011 Neural correlates of the continuous wagon wheel illusion: a functional MRI study. *Hum. Brain Mapp.* **32**, 163–170. (doi:10.1002/hbm.21007)
30. VanRullen R, Busch NA, Drewes J, Dubois J. 2011 Ongoing EEG phase as a trial-by-trial predictor of perceptual and attentional variability. *Front. Percept. Sci.* **2**, 1–9.
31. Busch NA, Dubois J, VanRullen R. 2009 The phase of ongoing EEG oscillations predicts visual perception. *J. Neurosci.* **29**, 7869–7876. (doi:10.1523/JNEUROSCI.0113-09.2009)
32. 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. (doi:10.1523/JNEUROSCI.3963-08.2009)
33. Busch NA, VanRullen R. 2010 Spontaneous EEG oscillations reveal periodic sampling of visual attention. *Proc. Natl Acad. Sci. USA* **107**, 16 048–16 053. (doi:10.1073/pnas.1004801107)
34. Dugue L, Marque P, VanRullen R. 2011 The phase of ongoing oscillations mediates the causal relation between brain excitation and visual perception. *J. Neurosci.* **31**, 11 889–11 893. (doi:10.1523/JNEUROSCI.1161-11.2011)
35. Drewes J, VanRullen R. 2011 This is the rhythm of your eyes: the phase of ongoing electroencephalogram oscillations modulates saccadic reaction time. *J. Neurosci.* **31**, 4698–4708. (doi:10.1523/JNEUROSCI.4795-10.2011)
36. Dugue L, Marque P, VanRullen R. Submitted. Theta oscillations modulate attentional search performance periodically. *J. Neurosci.*
37. Varela FJ, Toro A, John ER, Schwartz EL. 1981 Perceptual framing and cortical alpha rhythm. *Neuropsychologia* **19**, 675–686. (doi:10.1016/0028-3932(81)90005-1)
38. Gho M, Varela FJ. 1988 A quantitative assessment of the dependency of the visual temporal frame upon the cortical rhythm. *J. Physiol.* **83**, 95–101.
39. Chakravarthi R, VanRullen R. 2012 Conscious updating is a rhythmic process. *Proc. Natl Acad. Sci. USA* **109**, 10 599–10 604. (doi:10.1073/pnas.1121622109)
40. Nijhawan R. 1994 Motion extrapolation in catching. *Nature* **370**, 256–257. (doi:10.1038/370256b0)
41. Eagleman DM, Sejnowski TJ. 2000 Motion integration and postdiction in visual awareness. *Science* **287**, 2036–2038. (doi:10.1126/science.287.5460.2036)
42. VanRullen R, Macdonald JSP. 2012 Perceptual echoes at 10 Hz in the human brain. *Curr. Biol.* **22**, 995–999. (doi:10.1016/j.cub.2012.03.050)
43. Lalor EC, Pearlmutter BA, Reilly RB, McDarby G, Foxe JJ. 2006 The VESPA: a method for the rapid estimation of a visual evoked potential. *Neuroimage* **32**, 1549–1561. (doi:10.1016/j.neuroimage.2006.05.054)
44. Ciganek L. 1969 Variability of the human visual evoked potential: normative data. *Electroencephalogr. Clin. Neurophysiol.* **27**, 35–42. (doi:10.1016/0013-4694(69)90106-0)
45. Hillyard SA, Teder-Salejarvi WA, Munte TF. 1998 Temporal dynamics of early perceptual processing. *Curr. Opin. Neurobiol.* **8**, 202–210. (doi:10.1016/S0959-4388(98)80141-4)
46. Townsend J. 1990 Serial vs. parallel processing: sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. *Psychol. Sci.* **1**, 46–54. (doi:10.1111/j.1467-9280.1990.tb00067.x)
47. Jans B, Peters JC, De Weerd P. 2010 Visual spatial attention to multiple locations at once: the jury is still out. *Psychol. Rev.* **117**, 637–684. (doi:10.1037/a0019082)
48. VanRullen R, Dubois J. 2011 The psychophysics of brain rhythms. *Front. Psychol.* **2**, 203. (doi:10.3389/fpsyg.2011.00203)
49. Eriksen CW, Spencer T. 1969 Rate of information processing in visual perception: some results and methodological considerations. *J. Exp. Psychol.* **79**, 1–16. (doi:10.1037/h0026873)
50. Treisman A. 1969 Strategies and models of selective attention. *Psychol. Rev.* **76**, 282–299. (doi:10.1037/h0027242)
51. Kahneman D. 1973 *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.
52. VanRullen R, Carlson T, Cavanagh P. 2007 The blinking spotlight of attention. *Proc. Natl Acad. Sci. USA* **104**, 19 204–19 209. (doi:10.1073/pnas.0707316104)
53. Landau AN, Fries P. 2012 Attention samples stimuli rhythmically. *Curr. Biol.* **22**, 1000–1004. (doi:10.1016/j.cub.2012.03.054)
54. Posner MI. 1994 Attention: the mechanisms of consciousness. *Proc. Natl Acad. Sci. USA* **91**, 7398–7403. (doi:10.1073/pnas.91.16.7398)
55. Mack A, Rock I. 1998 *Inattentional blindness*. Cambridge, MA: MIT Press.
56. Regan D. 1977 Steady-state evoked potentials. *J. Opt. Soc. Am.* **67**, 1475–1489. (doi:10.1364/JOSA.67.001475)
57. Herrmann CS. 2001 Human EEG responses to 1–100 Hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. *Exp. Brain Res.* **137**, 346–353. (doi:10.1007/s002210100682)
58. Rager G, Singer W. 1998 The response of cat visual cortex to flicker stimuli of variable frequency. *Eur. J. Neurosci.* **10**, 1856–1877. (doi:10.1046/j.1460-9568.1998.00197.x)
59. Galambos R, Makeig S, Talmachoff PJ. 1981 A 40-Hz auditory potential recorded from the human scalp. *Proc. Natl Acad. Sci. USA* **78**, 2643–2647. (doi:10.1073/pnas.78.4.2643)
60. Stapells DR, Linden D, Suffield JB, Hamel G, Picton TW. 1984 Human auditory steady state potentials. *Ear Hear.* **5**, 105–113. (doi:10.1097/00003446-198403000-00009)
61. Greenberg S. 1998 A syllable-centric framework for the evolution of spoken language. *Behav. Brain Sci.* **21**, 518. (doi:10.1017/S0140525X98301260)

62. Poeppel D. 2003 The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Commun.* **41**, 245–255. (doi:10.1016/S0167-6393(02)00107-3)
63. Ghitza O, Greenberg S. 2009 On the possible role of brain rhythms in speech perception: intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica* **66**, 113–126. (doi:10.1159/000208934).
64. Giraud AL, Poeppel D. 2012 Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* **15**, 511–517. (doi:10.1038/nn.3063)
65. Peelle JE, Davis MH. 2012 Neural oscillations carry speech rhythm through to comprehension. *Front. Psychol.* **3**, 320. (doi:10.3389/fpsyg.2012.00320)
66. Lakatos S, Shepard RN. 1997 Constraints common to apparent motion in visual, tactile, and auditory space. *J. Exp. Psychol. Hum. Percept. Perform.* **23**, 1050–1060. (doi:10.1037/0096-1523.23.4.1050)
67. Feron FX, Frissen I, Boissinot J, Guastavino C. 2010 Upper limits of auditory rotational motion perception. *J. Acoust. Soc. Am.* **128**, 3703–3714. (doi:10.1121/1.3502456)
68. Saenz M, Langers DR. 2013 Tonotopic mapping of human auditory cortex. *Hear. Res.* **307**, 42–52. (doi:10.1016/j.heares.2013.07.016)
69. Shepard RN. 1964 Circularity in judgments of relative pitch. *J. Acoust. Soc. Am.* **36**, 2346–2353. (doi:10.1121/1.1919362)
70. Zoefel B, Heil P. 2013 Detection of near-threshold sounds is independent of EEG phase in common frequency bands. *Front. Psychol.* **4**, 262. (doi:10.3389/fpsyg.2013.00262)
71. Ng BS, Schroeder T, Kayser C. 2012 A precluding but not ensuring role of entrained low-frequency oscillations for auditory perception. *J. Neurosci.* **32**, 12 268–12 276. (doi:10.1523/JNEUROSCI.1877-12.2012)
72. Vanrullen R, McLelland D. 2013 What goes up must come down: EEG phase modulates auditory perception in both directions. *Front. Psychol.* **4**, 16. (doi:10.3389/fpsyg.2013.00016)
73. Henry MJ, Herrmann B. 2012 A precluding role of low-frequency oscillations for auditory perception in a continuous processing mode. *J. Neurosci.* **32**, 17 525–17 527. (doi:10.1523/JNEUROSCI.4456-12.2012)
74. Large EW, Jones MR. 1999 The dynamics of attending: how people track time-varying events. *Psychol. Rev.* **106**, 119–159. (doi:10.1037/0033-295X.106.1.119)
75. 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. (doi:10.1126/science.1154735)
76. Schroeder CE, Lakatos P. 2009 Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci.* **32**, 9–18. (doi:10.1016/j.tins.2008.09.012)
77. Henry MJ, Obleser J. 2012 Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proc. Natl Acad. Sci. USA* **109**, 20 095–20 100. (doi:10.1073/pnas.1213390109)
78. Luo H, Poeppel D. 2007 Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* **54**, 1001–1010. (doi:10.1016/j.neuron.2007.06.004)
79. Kerlin JR, Shahin AJ, Miller LM. 2010 Attentional gain control of ongoing cortical speech representations in a 'cocktail party'. *J. Neurosci.* **30**, 620–628. (doi:10.1523/JNEUROSCI.3631-09.2010)
80. Ghitza O. 2011 Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. *Front. Psychol.* **2**, 130. (doi:10.3389/fpsyg.2011.00130)
81. Ghitza O. 2012 On the role of theta-driven syllabic parsing in decoding speech: intelligibility of speech with a manipulated modulation spectrum. *Front. Psychol.* **3**, 238. (doi:10.3389/fpsyg.2012.00238)
82. Ding N, Simon JZ. 2012 Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc. Natl Acad. Sci. USA* **109**, 11 854–11 859. (doi:10.1073/pnas.1205381109)
83. Ding N, Simon JZ. 2013 Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. *J. Neurosci.* **33**, 5728–5735. (doi:10.1523/JNEUROSCI.5297-12.2013)
84. Zion-Golumbic EM, Poeppel D, Schroeder CE. 2012 Temporal context in speech processing and attentional stream selection: a behavioral and neural perspective. *Brain Lang.* **122**, 151–161. (doi:10.1016/j.bandl.2011.12.010)
85. Zion-Golumbic EM *et al*. 2013 Mechanisms underlying selective neuronal tracking of attended speech at a 'cocktail party'. *Neuron* **77**, 980–991. (doi:10.1016/j.neuron.2012.12.037)
86. Doelling KB, Arnal LH, Ghitza O, Poeppel D. 2013 Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage* **85**, 761–768. (doi:10.1016/j.neuroimage.2013.06.035)
87. 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. (doi:10.1093/cercor/bhs118)
88. Llinas RR, Pare D. 1991 Of dreaming and wakefulness. *Neuroscience* **44**, 521–535. (doi:10.1016/0306-4522(91)90075-Y)
89. Thorne JD, Debener S. 2013 Look now and hear what's coming: on the functional role of cross-modal phase reset. *Hear. Res.* **307**, 144–152. (doi:10.1016/j.heares.2013.07.002)
90. Kavanagh KT, Domico WD. 1986 High-pass digital filtration of the 40 Hz response and its relationship to the spectral content of the middle latency and 40 Hz responses. *Ear Hear.* **7**, 93–99. (doi:10.1097/00003446-198604000-00007)
91. İlhan B, VanRullen R. 2012 No counterpart of visual perceptual echoes in the auditory system. *PLoS ONE* **7**, e49287. (doi:10.1371/journal.pone.0049287)
92. Wertheimer M. 1912 Experimentelle Studien uber das Sehen von Bewegung. *Zeit. Psychol.* **61**, 161–265.
93. van Santen JP, Sperling G. 1985 Elaborated Reichardt detectors. *J. Opt. Soc. Am. A* **2**, 300–321. (doi:10.1364/JOSAA.2.000300)
94. Kubovy M. 1988 Should we resist the seductiveness of the space:time::vision:audition analogy? *J. Exp. Psychol. Hum. Percept. Perform.* **14**, 318–320. (doi:10.1037/0096-1523.14.2.318)
95. Apteker RT, Fisher JA, Kisimov VS, Neishlos H. 1995 Video acceptability and frame rate. *IEEE Multimedia* **2**, 32–40. (doi:10.1109/93.410510)
96. McCarthy JD, Sasse MA, Miras D. 2004 Sharp or smooth? Comparing the effects of quantization vs. frame rate for streamed video. In *Proc. SIGCHI Conference on Human Factors in Computing Systems, Vienna, Austria, 24–29 Apr 2004.* pp. 535–542. Aarhus, Denmark: Interaction Design Foundation.
97. Miller GA, Licklider JCR. 1950 The intelligibility of interrupted speech. *J. Acoust. Soc. Am.* **22**, 167. (doi:10.1121/1.1906584)
98. Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. 1995 Speech recognition with primarily temporal cues. *Science* **270**, 303–304. (doi:10.1126/science.270.5234.303)
99. Obleser J, Eisner F, Kotz SA. 2008 Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *J. Neurosci.* **28**, 8116–8123. (doi:10.1523/JNEUROSCI.1290-08.2008)
100. Saberi K, Perrott DR. 1999 Cognitive restoration of reversed speech. *Nature* **398**, 760. (doi:10.1038/19652)
101. Stilp CE, Kiefte M, Alexander JM, Kluender KR. 2010 Cochlea-scaled spectral entropy predicts rate-invariant intelligibility of temporally distorted sentences. *J. Acoust. Soc. Am.* **128**, 2112–2126. (doi:10.1121/1.3483719)
102. Drullman R, Festen JM, Plomp R. 1994 Effect of reducing slow temporal modulations on speech reception. *J. Acoust. Soc. Am.* **95**, 2670–2680. (doi:10.1121/1.409836)
103. Drullman R, Festen JM, Plomp R. 1994 Effect of temporal envelope smearing on speech reception. *J. Acoust. Soc. Am.* **95**, 1053–1064. (doi:10.1121/1.408467)
104. Joliot M, Ribary U, Llinas R. 1994 Human oscillatory brain activity near 40 Hz coexists with cognitive temporal binding. *Proc. Natl Acad. Sci. USA* **91**, 11 748–11 751. (doi:10.1073/pnas.91.24.11748)
105. Hirsh IJ, Sherrick CEJ. 1961 Perceived order in different sense modalities. *J. Exp. Psychol.* **62**, 423–432. (doi:10.1037/h0045283)
106. Poppel E. 1997 A hierarchical model of temporal perception. *Trends Cogn. Sci.* **1**, 56–61. (doi:10.1016/S1364-6613(97)01008-5)
107. Frisina RD. 2001 Subcortical neural coding mechanisms for auditory temporal processing. *Hear. Res.* **158**, 1–27. (doi:10.1016/S0378-5955(01)00296-9)
108. Kohler MA. 1997 A comparison of the new 2400 bps MELP Federal Standard with other standard coders. In *ICASSP-97: IEEE Intl. Conf. Acoustics,*

Speech, and Signal Processing, Munich, Germany, 21–24 April 1997, vol. 2, pp. 1587–1590. (doi:10.1109/ICASSP.1997.596256)

109. Funke K, Kisvarday F, Volgushev M, Worgotter F. 2002 The visual cortex. In *Models of neural networks IV: early vision and attention* (eds JL van Hemmen, JD Cowan, E Domany), pp. 131–160. New York: Springer.
110. Nelken I, Las L, Ulanovsky N, Farkas D. 2005 Are speech, pitch, and space early or late in the auditory system? In *The auditory cortex: a synthesis of human and animal research* (eds R Konig, P Heil, E BudingerH Scheich). Mahwah, NJ: Lawrence Erlbaum Associates.
111. Luo H, Tian X, Song K, Zhou K, Poeppel D. 2013 Neural response phase tracks how listeners learn new acoustic representations. *Curr. Biol.* **23**, 968–974. (doi:10.1016/j.cub.2013.04.031)
112. Petkov CI, Kang X, Alho K, Bertrand O, Yund EW, Woods DL. 2004 Attentional modulation of human auditory cortex. *Nat. Neurosci.* **7**, 658–663. (doi:10.1038/nn1256)
113. Rinne T, Stecker GC, Kang X, Yund EW, Herron TJ, Woods DL. 2007 Attention modulates sound processing in human auditory cortex but not the inferior colliculus. *Neuroreport* **18**, 1311–1314. (doi:10.1097/WNR.0b013e32826fb3bb)
114. Buffalo EA, Fries P, Landman R, Liang H, Desimone R. 2010 A backward progression of attentional effects in the ventral stream. *Proc. Natl Acad. Sci. USA* **107**, 361–365. (doi:10.1073/pnas.0907658106)
115. Lakatos P, O'Connell MN, Barczak A, Mills A, Javitt DC, Schroeder CE. 2009 The leading sense: supramodal control of neurophysiological context by attention. *Neuron* **64**, 419–430. (doi:10.1016/j.neuron.2009.10.014)

CHAPTER 2: THE ABILITY OF THE AUDITORY SYSTEM TO COPE WITH TEMPORAL SUBSAMPLING DEPENDS ON THE HIERARCHICAL LEVEL OF PROCESSING

The name says it all: In the previous chapter, reasons for the apparent lack of perceptual cycles in audition were discussed. It was concluded that direct subsampling of the auditory environment might prove detrimental for the brain, as essential information for the extraction of auditory features would be lost. However, different solutions were provided that can "keep alive" the notion of perceptual cycles in the auditory system. Two solutions are shortly summarized in the following:

1.) Auditory input is mostly rhythmic (speech, music, animal calls etc.). Note that "rhythmic" implies "predictable", making it possible to adjust auditory perceptual cycles such that "snapshots" are centered on particular(ly relevant) moments in time without loss of information. This mechanism can be called *phase entrainment* (e.g., Schroeder and Lakatos, 2009) and is treated extensively in Chapters 4-7.

2.) Perceptual cycles are a high-level phenomenon in audition: As shown in Chapter 1, taking "snapshots" of the actual input to the auditory system might be detrimental. However, it might still be possible to take "snapshots" after the input has been processed. For instance, after the stage of auditory feature extraction, stimulus representations are more stable temporally, and would suffer less from a moderate loss of temporal resolution.

This chapter concentrates on the second hypothesis which will be underlined with psychophysical findings. Whereas consequences of temporal subsampling have been compared for vision and audition in Chapter 1, this chapter concentrates on the auditory

system. Again, speech stimuli were constructed, subsampled at different frequencies. Temporal subsampling was applied on two "simulated" levels of the auditory system: At the cochlear level (simulated by subsampling the very input to the system, in the Wavelet domain[1]) and at the (potentially cortical) level of auditory features (simulated by a vocoder using linear predictive coding, LPC). We then tested auditory recognition of our stimuli by randomly presenting them to subjects, asking them to indicate any snippet that matched the one presented two snippets ago (2-back task). We can show that auditory recognition is more robust to subsampling on a relatively high level of auditory processing than to subsampling in the input domain. Although our results do not prove discrete perception in audition, they (1) show that subsampling (i.e. perceptual cycles) is possible without critically disrupting temporal information and (2) suggest that, if subsampling exists, it should operate on a relatively high level of auditory processing: Perceptual cycles on a higher level of auditory processing can reduce harmful effects of discretization and enable discrete auditory perception.

Article:

Zoefel B, Reddy Pasham N, Brüers S, VanRullen R (2015) The ability of the auditory system to cope with temporal subsampling depends on the hierarchical level of processing. Neuroreport 26:773-778.

[1] Note that cochlear processing can be mimicked by the Wavelet Transformation, as both filter their input signal into many narrow, logarithmically-spaced frequency bands.

The ability of the auditory system to cope with temporal subsampling depends on the hierarchical level of processing

Benedikt Zoefel[a,b], Naveen Reddy Pasham[c], Sasskia Brüers[a,b] and Rufin VanRullen[a,b]

Evidence for rhythmic or 'discrete' sensory processing is abundant for the visual system, but sparse and inconsistent for the auditory system. Fundamental differences in the nature of visual and auditory inputs might account for this discrepancy: whereas the visual system mainly relies on spatial information, time might be the most important factor for the auditory system. In contrast to vision, temporal subsampling (i.e. taking 'snapshots') of the auditory input stream might thus prove detrimental for the brain as essential information would be lost. Rather than embracing the view of a continuous auditory processing, we recently proposed that discrete 'perceptual cycles' might exist in the auditory system, but on a hierarchically higher level of processing, involving temporally more stable features. This proposal leads to the prediction that the auditory system would be more robust to temporal subsampling when applied on a 'high-level' decomposition of auditory signals. To test this prediction, we constructed speech stimuli that were subsampled at different frequencies, either at the input level (following a wavelet transform) or at the level of auditory features (on the basis of LPC vocoding), and presented them to human listeners. Auditory recognition was significantly more robust to subsampling in the latter case, that is on a relatively high level of auditory processing. Although our results do not directly demonstrate perceptual cycles in the auditory domain, they (a) show that their existence is possible without disrupting temporal information to a critical extent and (b) confirm our proposal that, if they do exist, they should operate on a higher level of auditory processing. *NeuroReport* **26:773–778**

NeuroReport 2015, **26:**773–778

Keywords: auditory, high level, perceptual cycles, robustness, sampling

[a]Université Paul Sabatier, Toulouse, France, [b]Centre de Recherche Cerveau et Cognition (CerCo), CNRS, UMR5549, Pavillon Baudot CHU Purpan, Toulouse Cedex, France and [c]Indian Institute of Technology, Bhubaneswar, Odisha, India

Correspondence to Benedikt Zoefel, Centre de Recherche Cerveau et Cognition (CerCo), CNRS, UMR5549, Pavillon Baudot CHU Purpan, BP 25202, 31052 Toulouse Cedex, France
Tel: +33 562 746 131; fax: +33 562 172 809; e-mail: zoefel@cerco.ups-tlse.fr

Received 9 June 2015 **accepted** 15 June 2015

Introduction

Recent research suggests that the visual system does not continuously monitor the environment, but rather samples it, cycling between 'snapshots' at discrete moments in time (perceptual cycles; for a review, see VanRullen *et al.* [1]). Interestingly, most attempts at discovering analogous perceptual cycles in the auditory system failed [2,3], indicating crucial differences between the visual and the auditory systems. A reason for this becomes evident when comparing the temporal structure of visual and auditory stimuli: whereas visual scenes are relatively stable over time, auditory input changes rapidly over time. In fact, whereas the visual system might rely particularly on the spatial dimension, time might be the most important factor for the auditory system [4] – and thus, subsampling auditory input in the time domain might destroy essential information [1]. Does this mean that perceptual cycles cannot be found in the auditory domain because it is impossible to subsample the auditory stream without losing important information? In this article, we argue that this is not necessarily the case – rather, it is possible that subsampling does take place in the auditory system, but on a relatively 'high' level of auditory processing: auditory information might be more temporally stable after a certain amount of feature extraction, enabling auditory subsampling without a significant loss of information. Thus, in this study, temporal subsampling was not only applied to the direct input to the auditory system, but we also subsampled the auditory stream on a higher-level representation (i.e. on the output level of a vocoder extracting auditory features by the use of linear predictive coding, LPC [5]). We predicted that the auditory system may prove significantly more robust to subsampling on the level of auditory features than when a similar subsampling was applied on the input level. We tested auditory vulnerability in a two-back recognition task (see Methods). An improved performance in this task for stimuli subsampled on a higher-level representation than for those subsampled at the input level would support the possibility that auditory perceptual cycles operate on a hierarchically high level of auditory processing.

Methods

Participants

Seven participants (four women, mean age 26.2 years), all fluent in English, volunteered to participate in the

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website (*www.neuroreport.com*).

DOI: 10.1097/WNR.0000000000000422

experiment. All participants provided written informed consent, reported normal hearing, and received compensation for their time. The experimental protocol was approved by the relevant ethical committee at Centre National de la Recherche Scientifique (CNRS).

Stimulus construction

One original 10-min audio sequence [sampling rate (SR) = 44 100 Hz], a recording of a male native English speaker reading parts of a classic novel, was used as the primary stimulus in our experiment. The audio recording was cut into 200 3-s long 'snippets'. These snippets were then subsampled, either at the input level ('input condition'; i.e. at the level of the very input to the auditory system, such as in the cochlea; Fig. 1, top) or at the level of auditory features ('feature condition'; i.e. at a level beyond cochlear processing; Fig. 1, bottom). 'Subsampling' a given input stream does not necessarily mean 'forgetting' or 'ignoring' information. It might just be that the temporal order of information is lost, while the information itself is preserved. Thus, in our study, for both conditions (input and feature), we simulated 'subsampling' of the auditory system by shuffling auditory samples within a given time interval: for a SR of 4 Hz, for instance, all samples within a 250-ms window were shuffled. Of course, the larger this interval, the more difficult for the system to restore the exact (order of) information. However, we hypothesized that this restoration would be easier if the subsampling takes place in the auditory feature domain than when the input is subsampled at the input level as the former is temporally more stable. For every snippet, to prevent the use of static information for recognition, two subsampled versions were created by starting the shuffling interval either on the first sample or on the nearest sample to 1 + SR/SF/2. Sample sound files are available for both conditions as Supplemental digital content, 1–18 (*http://links.lww.com/WNR/A322*, *http://links.lww.com/WNR/A323*, *http://links.lww.com/WNR/A324*, *http://links.lww.com/WNR/A325*, *http://links.lww.com/WNR/A326*, *http://links.lww.com/WNR/A327*, *http://links.lww.com/WNR/A328*, *http://links.lww.com/WNR/A329*, *http://links.lww.com/WNR/A330*, *http://links.lww.com/WNR/A331*, *http://links.lww.com/WNR/A332*, *http://links.lww.com/WNR/A333*, *http://links.lww.com/WNR/A334*, *http://links.lww.com/WNR/A335*, *http://links.lww.com/WNR/A336*, *http://links.lww.com/WNR/A337* *http://links.lww.com/WNR/A338*, *http://links.lww.com/WNR/A339*).

Subsampling at the input level

For the input condition (Fig. 1, top), snippets were converted into the wavelet domain to approximate cochlear transduction (continuous Morlet wavelet transform of order 6). Snippets were divided into intervals, with the reciprocal of this interval corresponding to the desired subsampling frequency (SF). The amplitudes of the complex wavelet coefficients within the respective interval were shuffled. The phase information at the first sample of each interval was preserved and interpolated to avoid artifacts created by discrete phase transitions. After shuffling, final snippets for the time condition were obtained by applying the inverse wavelet transform.

Subsampling at the feature level

For the feature condition (Fig. 1, bottom), auditory features for each snippet were extracted using an LPC vocoder [5]. More precisely, linear prediction coefficients α_k were constructed such that each auditory sample s at time t can be seen as a linear combination of past p samples (p is the order of prediction):

$$s'(t) = -\sum_{k=1}^{p} a_k \times s(t-k), \tag{1}$$

where $s'(t)$ is the predicted auditory sample. A pre-emphasis filter [6] was applied on s before $s'(t)$ was calculated. 11 α_k (among which the first is unity) were calculated for each frame of 30 ms, with 20 ms between centers of subsequent frames (resulting in an overlap of 10 ms between frames). For each frame fr, after a Hamming window was applied, α_k were constructed using the method of least squares, that is the following total prediction error E was minimized:

$$E(fr) = \sum_{t=-\infty}^{\infty} e^2(fr,t), \tag{2}$$

where

$$\begin{aligned} e(fr,t) &= s(fr,t) - s'(fr,t) \\ &= s(fr,t) + \sum_{k=1}^{p} a_k(fr) \times s(fr,t-k). \end{aligned} \tag{3}$$

This was done using the Levinson–Durbin algorithm, which we do not explain in detail here, but which is described thoroughly in the relevant literature [7,8].

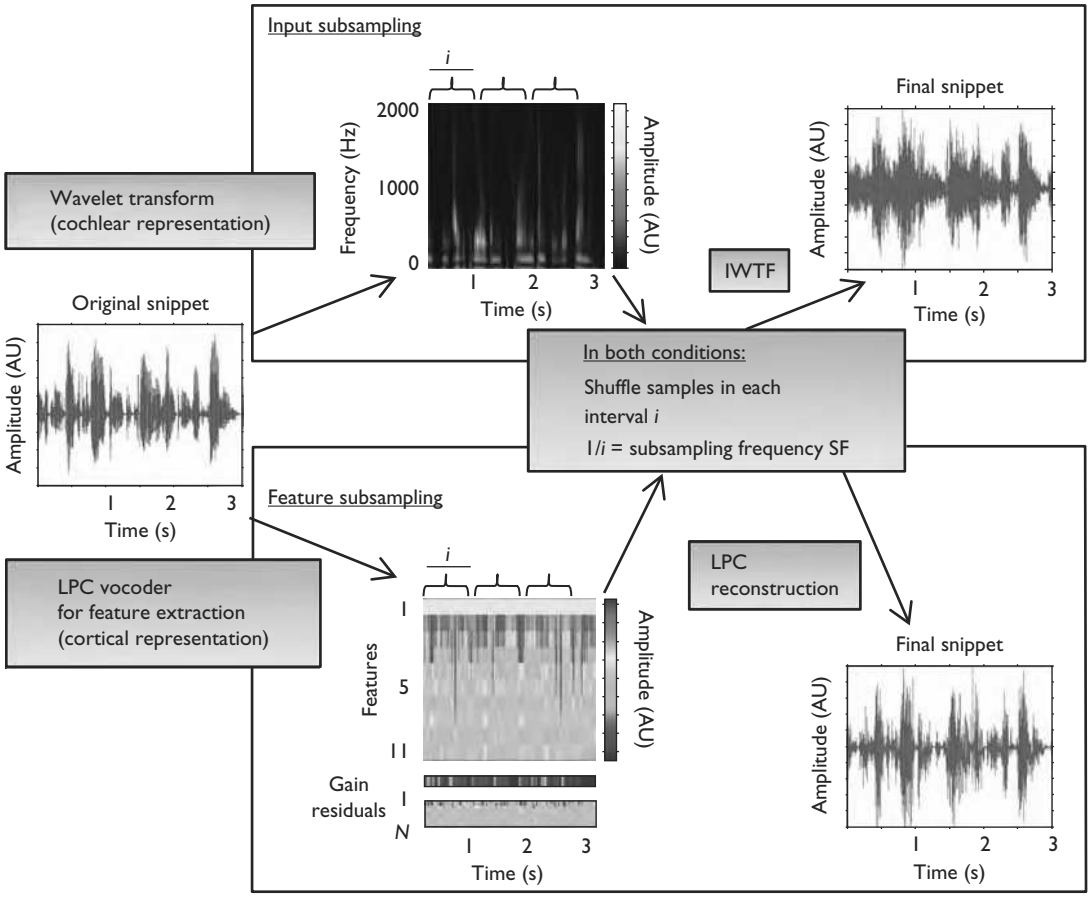
Two more parameters were extracted for each frame: the gain g (defined as the power of the speech signal in each frame) and the residual r:

$$r(fr,t) = \frac{e(fr,t)}{g(fr)}. \tag{4}$$

Discrete cosine transform (DCT) was applied to the residual of each frame; all except the first 50 DCT coefficients were discarded (as energy of the speech signal is concentrated in those 50 coefficients) and the inverse DCT was applied to obtain residuals with an improved signal-to-noise ratio [9]. Gaussian noise was added to the residuals (signal–noise ratio ~1:1) to improve the final sound quality of the reconstructed speech snippets.

For subsampling, α_k, gain, and residual were always shuffled together (i.e. α_k, gain and residual for a given frame were never separated). The step size of 20 ms

Fig. 1



Overview of the experimental approach. Original snippets were subsampled, either at the input level (input condition; top) or at the level of auditory features (feature condition; bottom). In both conditions, subsampling was realized by shuffling auditory samples in a certain time interval, with the reciprocal of this interval corresponding to the SF in the respective snippet. This shuffling was performed in the wavelet domain (corresponding to a cochlear representation of the sound) for the input condition and in the feature domain (obtained by an LPC vocoder; corresponding to a cortical representation of the sound) for the feature condition (see Methods). Note that subsampling in the input domain corresponds to simulating perceptual cycles on the level of the cochlea, whereas subsampling in the feature domain simulates perceptual cycles on a higher level (beyond cochlear processing) of the auditory pathway.

between frames restricted our maximal SF to 50 Hz, without any additional subsampling/shuffling. After shuffling, final snippets for the feature condition were obtained by filtering, in each frame, the residual, multiplied by the gain in the respective frame, by the obtained α_k:

$$s_{\text{final}}(fr,t) = \sum_{k=1}^{p} a_k(fr) \times r_g(fr, t-k), \qquad (5)$$

where $r_g(fr,t)$ is $r(fr,t) \times g(fr)$.

Experimental paradigm

For both conditions, snippets were presented (separated by 1 s blank intervals) in a randomized order to our participants, who were instructed to perform a two-back task: they were asked to indicate by a button press any snippet that matched the one presented two snippets ago. These two-back repeats occurred randomly with a probability of 33%. Whenever a two-back repeat occurred in the sequence, it was always between nonidentical subsampled versions (with the same subsampling interval durations, but differing in the exact delay at which subsampling intervals were applied; see above). Stimuli were presented in separate blocks of 30 snippets. In each block, a different SF was applied. Participants completed 60 trials for each SF and condition. The two conditions ('input' and 'feature') were tested on separate days.

Data analyses

We hypothesized that the auditory system is more robust to temporal subsampling at the level of auditory features (feature condition) than at the input level (input condition). This robustness was tested in an auditory recognition task for snippets of different SF. Of course, with decreasing SF, performance will decline in both conditions. However, if

our hypothesis is true, the precise SF where auditory recognition starts to decline will be lower for the feature than for the input condition.

We defined auditory recognition as d', the sensitivity of our participants in the two-back task. d' takes into account both correct responses (participants' response 'repeat' when there was actually a repeat) and false alarms (participants' response 'repeat' when there was no repeat):

$$d' = z(\text{hits}) - z(\text{false alarms}),$$

where $z(p)$, $p \epsilon [0,1]$, is the inverse of the cumulative Gaussian distribution [10]. Performance in these signal detection tasks usually results in psychometric curves that have sigmoidal shapes [10]. We thus defined the lowest sustainable SF as the inflection point of those psychometric curves in both conditions. To test whether perception was significantly more robust against temporal subsampling in the feature condition, for each participant, we fitted a sigmoidal curve to the performance in both conditions and calculated its inflection point (in Hz). Inflection points were then compared across conditions using Student's t-test to test whether the robustness of auditory perception against temporal subsampling differs between the two conditions.

Results

In this study, participants were presented with speech stimuli that were subsampled (at different temporal SF) either at the input level (input condition) or at the level of auditory features (feature condition). Using a two-back task (see Methods), we tested the robustness of the auditory system to this subsampling – if perceptual cycles do exist in the auditory system, they can only occur at frequencies that prove to be robust against temporal subsampling. Of course, positive results would not imply that they actually do occur, but our approach can inform us whether sampling on a hierarchically high level of auditory processing (i.e. in the feature condition) can reduce the detrimental effects of environmental sampling (i.e. loss of information) and thus 'keep alive' the notion of perceptual cycles in the auditory system.

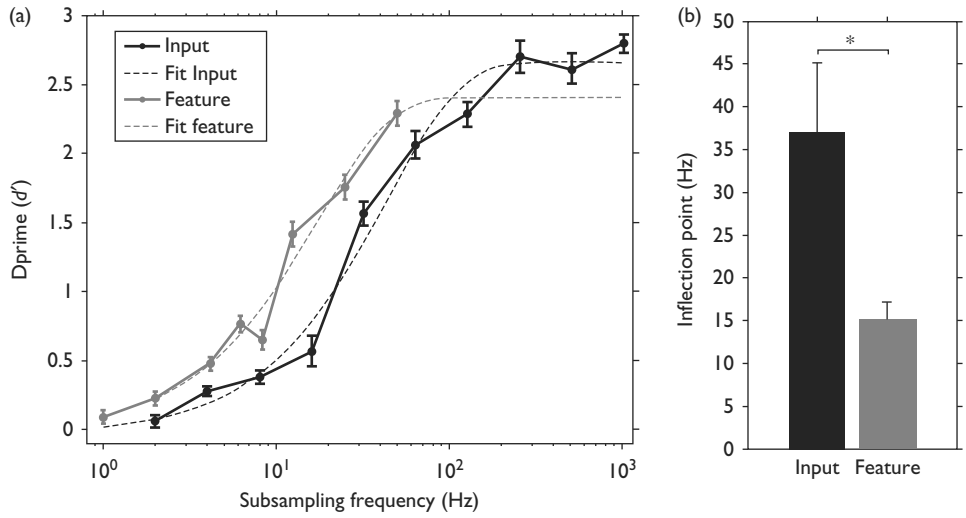
The performance (measured in d'; see Methods) of our participants ($N = 7$) in the two-back task for both conditions is shown in Fig. 2. For both conditions, of course, performance increased with increasing SF, and both curves resemble sigmoid psychometric functions. However, this curve is shifted toward lower SF for the feature condition, indicating that the auditory system is more robust against subsampling at the level of auditory features than at the input level (Fig. 2a). When we define the lowest sustainable SF as the inflection point of those psychometric curves under both conditions (Fig. 2b; see Methods), this SF is significantly lower [$t(6) = 2.61$, $P = 0.023$] for the feature condition (15.1 ± 5.3 Hz; mean and SD across participants) than for the time condition (37.0 ± 21.5 Hz). Thus, subsampling on a hierarchically higher level of auditory processing would occur with a clear advantage for the auditory system as the SF could be reduced without a significant loss of information.

Discussion

In a previous study, we showed that subsampling the auditory stream in the input domain has detrimental effects on stimulus processing already at SF below 32 Hz [1]. This result suggests that, if the auditory environment is indeed monitored in a discrete manner, this sampling cannot take place on the very input to the system as important information would be lost (or the SF would have to be so high that subsampling would be useless). Instead of rejecting the idea of perceptual cycles in audition, we proposed an alternative idea, which is supported by experimental findings in the present study: subsampling in the auditory system is less detrimental to performance if it takes place on a hierarchically higher level of auditory processing – on the level of auditory features. Our results are in line with the work published by Suied *et al.* [11]: although they used short vocal sounds instead of speech sound, they were able to show that participants could still recognize those sounds even though they were reduced to a few perceptually important features, with the number of features (10 features/s) similar to our perceptual 'threshold'. We extend their findings by systematically testing different SF and by comparing perceptual consequences for subsampling at the level of the cochlea with those for a hierarchically higher stage of auditory processing.

Whereas there is plenty of (psychophysical and electrophysiological) evidence for perceptual cycles in vision [12–15], equivalent experiments consistently fail for the auditory domain ([2,3]; reviewed in VanRullen *et al.* [1]) or remain debated [16–19]. Thus, it is possible that the reported experimental approaches for the visual system are not appropriate for the investigation of the auditory system. Our study directly contributes toward resolving this discrepancy by providing an answer to why this could be the case: perceptual cycles might operate on different hierarchical levels in the two systems and thus might not be captured with the same experimental methods (although this does not necessarily mean that the visual system is not influenced by high-level factors; for instance, some evidence for perceptual cycles in vision is modulated by attention [15]). The visual world is relatively stable over time and subsampling does not disrupt essential information, even when the very input to the system is discretized. This was verified in our previous study [1], where, in a two-back recognition paradigm similar to the one used here, human observers could robustly recognize visual inputs at subsampling rates below 5 Hz. In contrast, the fluctuating nature of auditory stimuli makes it necessary to extract features that are

Fig. 2



Auditory perception is more robust against temporal subsampling at the level of auditory features than when the same subsampling is applied at the input level. (a) Performance in a two-back task in the input (black) and feature condition (gray). Note that the sigmoidal curve (a fit is shown using dashed lines) is shifted toward lower SF for the feature condition, indicating that, for a given SF, performance was better (and perception more robust) in the latter condition. (b) Inflection points for both conditions, averaged across sigmoidal curves fitted on the data of individual participants. The inflection point is significantly higher for the input condition than for the feature condition, indicating that subsampling has more detrimental effects when it is performed at the level of the cochlea (i.e. at the input level) than at a hierarchically higher level of auditory processing (i.e. in the feature domain). SEM across participants is shown by error bars.

both relevant and more stable before subsampling can be applied (see Results [1,20]). Indeed, whereas early auditory representations of speech seem to entail all acoustic details, representations at later hierarchical stages are rather categorical (i.e. relatively independent of acoustic information) and thus more stable in time [21], and therefore, more robust to subsampling. This property of an increasing abstraction of auditory representation along the pathway begins beyond the primary auditory cortex and is particularly outstanding in the anterior temporal cortex (ventral stream) [21], making it, although speculatively, a good candidate for perceptual cycles in the auditory system: 'auditory objects' are 'built' within this stream [22], transforming spectrotemporal (i.e. time-resolved) properties of the input stream into more abstract 'identities' (i.e. relatively independent of the time domain). The superior temporal sulcus – in which certain neurons respond more strongly to speech than to other sounds [23] – is part of this stream.

Auditory perception in our study did not prove as robust to temporal subsampling as observed previously for vision. This does not necessarily imply that auditory perceptual cycles, if they exist, must be faster than visual ones. Instead, it may just be that our 'feature' decomposition did not fully capture the complexity of the auditory representation at which subsampling occurs. The higher in the hierarchy of the auditory pathway, the slower are the 'preferred' frequencies of the auditory system [24]. It thus remains to be shown in future studies whether subsampling an even more complex decomposition of the auditory signal can result in performance that equals that obtained in vision.

More studies are necessary to find an appropriate experimental approach for perceptual cycles in audition and to characterize them with respect to their location in the auditory pathway. One step forward toward auditory perceptual cycles has been published recently by our group [25]: in that study, specifically constructed noise was mixed with speech sound to counterbalance fluctuations in low-level features of the latter (i.e. fluctuations in amplitude and spectral content). Importantly, these mixture speech/noise stimuli remained intelligible, indicating that high-level features of speech (including phonetic information) were preserved and fluctuated rhythmically. We could show that the detection of tone pips is modulated by this 'high-level rhythm' and, consequently, that phase entrainment, the brain's adjustment to regular stimulation, indeed involves a high-level component. This finding is in line with the present study, suggesting a periodic mechanism on a high level of auditory processing.

To conclude, our data suggest that, even in the auditory world of continuous, rapid temporal fluctuations, the idea of discrete perceptual processing can be kept alive: discretization on a hierarchically high level of auditory processing is possible without disrupting essential information. Of course, our experiment does not prove that perceptual cycles do exist in audition; however, we

conclude that (a) there is a possibility that they exist and (b) if so, they are a high-level phenomenon.

Acknowledgements

The authors are grateful to Daniel Pressnitzer for helpful comments and discussions. This study was supported by a Studienstiftung des deutschen Volkes (German National Academic Foundation) scholarship to B.Z. and an ERC grant P-CYCLES number 614244 to R.V.

Conflicts of interest

There are no conflicts of interest.

References

1 VanRullen R, Zoefel B, Ilhan B. On the cyclic nature of perception in vision versus audition. *Philos Trans R Soc Lond B Biol Sci* 2014; **369**:20130214.
2 İlhan B, VanRullen R. No counterpart of visual perceptual echoes in the auditory system. *PloS One* 2012; **7**:e49287.
3 Zoefel B, Heil P. Detection of near-threshold sounds is independent of EEG phase in common frequency bands. *Front Psychol* 2013; **4**:262.
4 Kubovy M. Should we resist the seductiveness of the space:time::vision: audition analogy? *J Exp Psychol Hum Percept Perform* 1988; **14**:318–320.
5 Kinnunen T, Li H. An overview of text-independent speaker recognition: from features to supervectors. *Speech Commun* 2010; **52**:12–40.
6 Chung K, McKibben N. Microphone directionality, pre-emphasis filter, and wind noise in cochlear implants. *J Am Acad Audiol* 2011; **22**:586–600.
7 Levinson N. The Wiener RMS error criterion in filter design and prediction. *J Math Phys* 1947; **25**:261–278.
8 Durbin J. The fitting of time series models. *Rev Inst Int Stat* 1960; **28**:233–243.
9 Soon IY, Koh SN, Yeo CK. Noisy speech enhancement using discrete cosine transform. *Speech Commun* 1998; **24**:249–257.
10 Macmillan NA, Creelman CD. *Detection theory: a user's guide*. Mahwah, New Jersey: Lawrence Erlbaum Associates; 2004.
11 Suied C, Drémeau A, Pressnitzer D, Daudet L. Auditory sketches: sparse representations of sounds based on perceptual models. In: Aramaki M, Barthet M, Kronland-Martinet R, Ystad S, editors. *From sounds to music and emotions*. Berlin Heidelberg: Springer; 2013. pp. 154–170.
12 VanRullen R, Reddy L, Koch C. The continuous wagon wheel illusion is associated with changes in electroencephalogram power at approximately 13 Hz. *J Neurosci* 2006; **26**:502–507.
13 Busch NA, Dubois J, VanRullen R. The phase of ongoing EEG oscillations predicts visual perception. *J Neurosci* 2009; **29**:7869–7876.
14 Mathewson KE, Gratton G, Fabiani M, Beck DM, Ro T. To see or not to see: prestimulus alpha phase predicts visual awareness. *J Neurosci* 2009; **29**:2725–2732.
15 Busch NA, VanRullen R. Spontaneous EEG oscillations reveal periodic sampling of visual attention. *Proc Natl Acad Sci U S A* 2010; **107**:16048–16053.
16 Henry MJ, Herrmann B. A precluding role of low-frequency oscillations for auditory perception in a continuous processing mode. *J Neurosci* 2012; **32**:17525–17527.
17 Henry MJ, Obleser J. Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proc Natl Acad Sci U S A* 2012; **109**:20095–20100.
18 Ng BSW, Schroeder T, Kayser C. A precluding but not ensuring role of entrained low-frequency oscillations for auditory perception. *J Neurosci* 2012; **32**:12268–12276.
19 Vanrullen R, McLelland D. What goes up must come down: EEG phase modulates auditory perception in both directions. *Front Psychol* 2013; **4**:16.
20 Thorne JD, Debener S. Look now and hear what's coming: on the functional role of cross-modal phase reset. *Hear Res* 2014; **307**:144–152.
21 Davis MH, Johnsrude IS. Hearing speech sounds: top-down influences on the interface between audition and speech perception. *Hear Res* 2007; **229**:132–147.
22 Bizley JK, Cohen YE. The what, where and how of auditory-object perception. *Nat Rev Neurosci* 2013; **14**:693–707.
23 Overath T, McDermott JH, Zarate JM, Poeppel D. The cortical analysis of speech-specific temporal structure revealed by responses to sound quilts. *Nat Neurosci* 2015; **18**:903–911.
24 Edwards E, Chang EF. Syllabic (~2-5 Hz) and fluctuation (~1-10 Hz) ranges in speech and auditory processing. *Hear Res* 2013; **305**:113–134.
25 Zoefel B, VanRullen R. Selective perceptual phase entrainment to speech rhythm in the absence of spectral energy fluctuations. *J Neurosci* 2015; **35**:1954–1964.

CHAPTER 3: THE PHASE OF NEURAL OSCILLATIONS ACTS AS A TOOL FOR THE SEGREGATION AND INTEGRATION OF AUDITORY INPUTS

In the previous chapter, we have seen that perceptual cycles in audition might rely on a hierarchically higher level of processing. However, as discussed in Chapter 1 (and summarized in the Introduction of Chapter 2), there is a second (and not mutually exclusive) possibility how the auditory system might be able to avoid disruptive effects of subsampling: By an *active control* of its neural oscillations. These oscillations reflect perceptual cycles; by shifting their phase such that the "snapshot phase" coincides with the "input of interest" and the phase between "snapshots" is aligned with irrelevant information, the loss of information (caused by temporal subsampling) can be reduced: Indeed, information is still lost – but only the irrelevant one. Of course, this mechanism only works if the input is predictable – but rhythmicity is an inherent feature of the auditory environment, and rhythmicity implies predictability. In the following article, evidence for an active control of the oscillatory phase in the auditory system is presented. In an auditory scene, single events are not perceived as individual elements, but rather grouped to form a single "auditory object" or segregated to form separate auditory stream (Bregman, 1994). The idea underlying this chapter is that perceptual cycles (i.e. neural oscillations) might play an important role for auditory stream formation: Individual events might be grouped or segregated, depending on the oscillatory phase they fall into. Clearly, this would provide the auditory system with an important control mechanism: As mentioned above, by changing the phase of its oscillations, it could decide on stream integration and segregation. Indeed, using a popular paradigm for the investigation of auditory stream formation ("ABA

paradigm") and recording in the primary auditory cortex (A1) in the monkey, we show that the neural phase can be used as a tool for stream integration and segregation. Tone triplets (ABA) were presented that can be heard as either a single stream or two separated streams, depending on the frequency separation between A and B tones. The monkey listened passively to the tone sequences while current source density (CSD) profiles and single/multiunit activity (see General Introduction) was recorded. In line with our expectation, we observed a phase shift of neural oscillations in A1 when the assumed perception changed from segregation (for large frequency separations between A and B tones) to integration (for small frequency separations), thereby shifting the perceptual cycles hypothesized to enable stream integration and segregation. Importantly, the same phase shift was apparent when comparing moments of identical stimulation but different "stimulation history" that might have influenced the monkey's percept, suggesting that the observed phase difference reflects a change in the state of endogenous oscillations. Moreover, neural phases were more consistent for assumedly integrated than for segregated tone triplets, indicating the presence of one integrated stream in the auditory environment rather than two for the segregated percept. These effects were largest in the supragranular layers of A1, in line with previous reports of strongest adaption to auditory stimulation in those layers (e.g., Lakatos et al., 2008). Thus, our findings support the idea that perceptual cycles can be controlled by the auditory system, for stimulus selection and potentially for an efficient subsampling of its environment.

Article:

Zoefel B, O'Connell N, Barczak A, VanRullen R, Lakatos P (in preparation) The phase of neural oscillations acts as a tool for the segregation and integration of auditory inputs.

The phase of neural oscillations acts as a tool for the segregation and integration of auditory inputs

Authors: Benedikt Zoefel[a,b,c*], Noelle O'Connell[c], Annemarie Barczak[c] , Rufin VanRullen[a,b] and Peter Lakatos[c]

Affiliations: [a] Université Paul Sabatier, Toulouse, France

[b] Centre de Recherche Cerveau et Cognition (CerCo), CNRS, UMR5549, Pavillon Baudot CHU Purpan, BP 25202, 31052 Toulouse Cedex, France

[c] Nathan Kline Institute for Psychiatric Research, Orangeburg, NY, United States

*Corresponding author: Benedikt Zoefel
Centre de Recherche Cerveau et Cognition (CerCo)
Pavillon Baudot CHU Purpan, BP 25202
31052 Toulouse Cedex
France

Phone: +33 562 746 131
Fax: +33 562 172 809
Email: zoefel@cerco.ups-tlse.fr

Number of pages: 35
Number of figures: 7
Number of words:
Abstract: 244
Introduction: 1263
Discussion: 2019

Key words: stream integration, stream segregation, oscillation, phase, auditory

Running title: Oscillatory mechanisms for stream integration and segregation

Acknowledgements: This study was supported by a Studienstiftung des deutschen Volkes (German National Academic Foundation) scholarship to BZ, a EURYI Award as well as an ERC Consolidator grant P-CYCLES under grant agreement 614244 to RV, and NIH R01DC012947 to PL.

Conflict of Interest: The authors declare no competing financial interests.

Abstract

Due to a continuous influx coming from multiple sources, the integration and segregation of individual streams is a key feature for the auditory system. It would be thus very beneficial to possess a tool that can simultaneously be used to segregate relevant streams, and group or parse their elements. Recently, it has been hypothesized that the phase of neural oscillation might be highly relevant for this task: Whereas the low excitability phase might represent the "parser", stimuli falling into the high excitability phase might be grouped to form a single auditory object and thereby get segregated from the background in the auditory scene. We investigated this hypothesis by presenting one monkey with tone triplets (ABA paradigm) that can be heard as either a single stream or two segregated streams, depending on the frequency separation between A and B tones. The monkey listened passively to the tone sequences while neuronal activity was recorded intracortically in primary auditory cortex (A1). In line with our expectation, we observed a phase shift of neural oscillations when the assumed perception changed from segregation to integration. Importantly, the same phase shift was apparent when comparing moments of identical stimulation but different "stimulation history" that might have influenced the monkey's percept, suggesting that the observed phase difference reflects a change in the state of endogenous oscillations. Our results indicate that neuronal oscillations are used to perform multiple simultaneous operations for the meaningful interpretation of the auditory environment: segregation, parsing and integration.

Introduction

Auditory events are rarely perceived as individual, isolated elements. Rather, the brain continuously groups and parses them, forming different "auditory objects" (Griffiths and Warren, 2004; Denham and Winkler, 2014; Nelken et al., 2014) in order to interpret and evaluate its environment. The "fate" of the individual auditory element (i.e. to which auditory object it is assigned to) depends on various spectrotemporal properties. For instance, auditory elements with similar frequency (Bregman and Campbell, 1971; Van Noorden, 1975) or similar timing (Elhilali et al., 2009; Shamma et al., 2011) are usually combined and assigned to the same auditory object (for a review, see Denham and Winkler, 2014). The formation of auditory objects is assumed to underlie the segregation and integration of auditory streams in the presence of multiple auditory sources, such that auditory elements assigned to the same auditory object are integrated to form the perception of a single auditory stream, whereas auditory elements assigned to different auditory objects are segregated to be perceived as separated auditory streams (e.g., Nelken et al., 2014). One popular paradigm to test auditory stream segregation and integration is the so-called ABA paradigm (Van Noorden, 1975). Here, a triplet ("ABA") of tones is presented, with different tone frequencies for A and B tones. The presentation rate of B tones is half that of A tones, because every second B tone is "omitted" (with respect to a regular ABAB presentation scheme). The spectrotemporal properties of A and B tones leading to stream segregation and integration have been described extensively in psychophysical experiments (for a review, see Moore and Gockel, 2012). For instance, the ABA triplet can be perceived as one integrated stream when the tone frequencies for A and B tones are similar, and as two segregated streams (one stream of A tones and another stream of B tones with half the presentation rate) when the tone frequencies for A and B

tones are far apart. However, despite the vast amount of psychophysical literature, the neural mechanisms underlying stream segregation and integration remain mostly unclear.

There is increasing evidence that neural oscillations, reflecting cyclic changes in the excitability of neural populations (e.g., Buzsáki and Draguhn, 2004), play an important role for the structuring of auditory input. For instance, auditory inputs that coincide with the high excitability phase of neuronal ensemble oscillations are amplified whereas stimuli coinciding with the low excitability phase are suppressed (e.g., Schroeder and Lakatos, 2009). Based on this property, if properly aligned to the input stream via oscillatory entrainment, the oscillatory phase might be used by the auditory system to parse its input (e.g., Giraud and Poeppel, 2012; Ghitza, 2013): The high excitability part of the oscillatory cycle might act as a "window of integration" (e.g., Dehaene, 1993; van Wassenhove, 2009) and stimuli falling within one high excitability phase become united to form the percept of one integrated item, while items falling within successive high excitability phases might be perceived as being part of the same stream (i.e. integrated). Thus, by examining the relationship of neuronal oscillatory phases to the timing of auditory stimuli in the environment, theoretically we can devise how the brain is integrating and segregating the available auditory inputs.

We tested this hypothesis by presenting monkeys with the ABA paradigm while neural activity was recorded intracortically in A1. We expected a change in the phase of neural oscillations in relation to the timing of ABA stimuli when the assumed percept changed from segregation to integration (1 vs. 2 and 3 vs. 4 in Figure 1). In two experiments, the frequencies of A and B tones were systematically varied such that the assumed percept continuously changed from segregation to integration, and vice versa. In these experiments,

we were able to confirm our expectations of a neural phase shift associated with a change in the assumed percept. These results were confirmed when moments of identical stimulation were compared and the monkey's percept was inferred based on the "stimulation history" (a "perceptual bias" by "stimulation history" was verified in human subjects). Our results provide important evidence for the neural mechanism of simultaneous stream segregation and integration and for an important role of neural oscillations for auditory object formation.

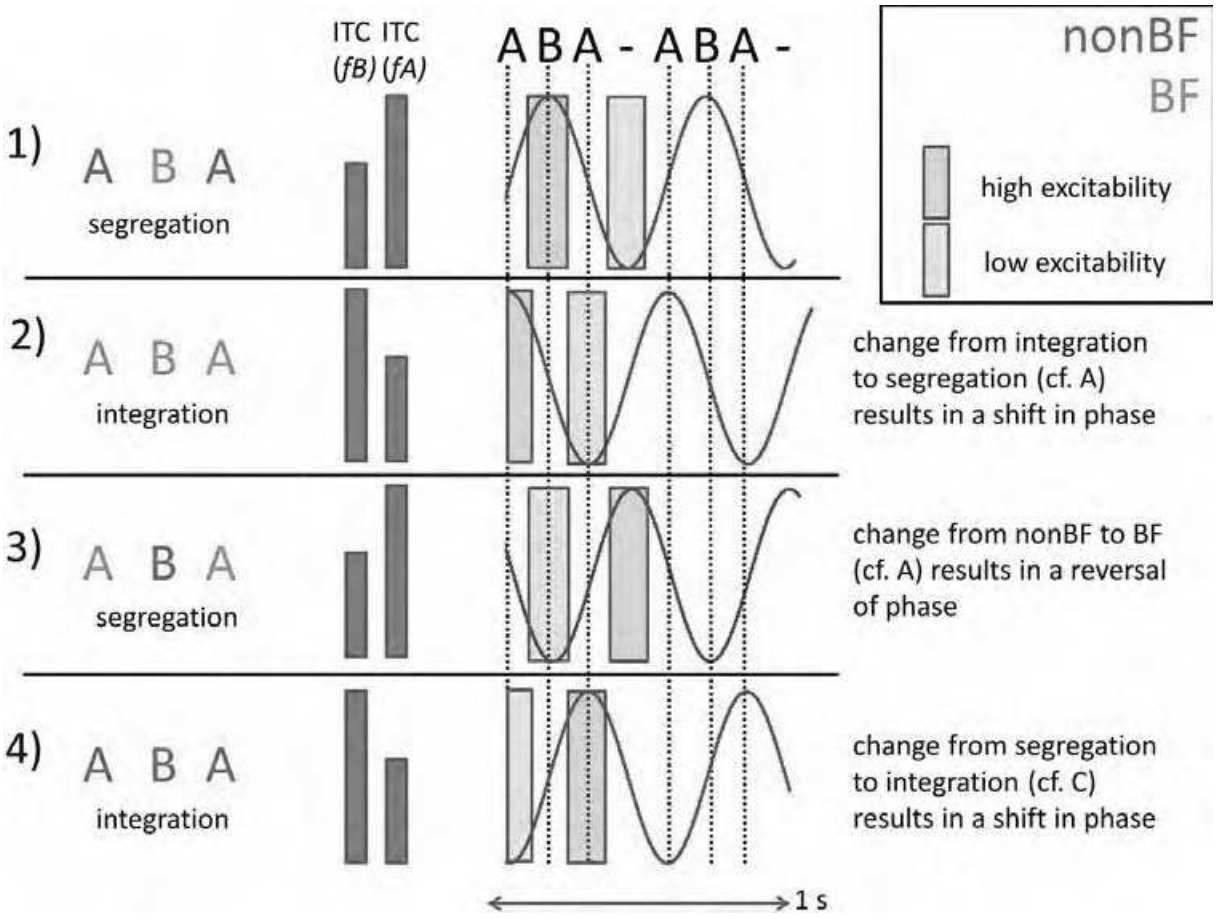


Figure 1: Predictions in the current study. Sequences of ABA triplets were presented that can either be perceived as two segregated steams of A and B tones (1,3) or as a single, integrated stream of triplets (2,4), depending on the frequency separation between A and B tones. We hypothesized that a change in neural phase (at a frequency corresponding to the presentation rate of B tones and triplets) might underlie a change in percept. We assumed that the high excitability phase (shaded red) might be centered on the B tones when A and B streams are perceptually segregated, and that this phase might be shifted – now centered on the first A tone – in the case of stream integration. Note that this description is true only when the frequency of B tones corresponds to the BF of the recorded site in A1 (1,2) – based on previous results (Lakatos et al., 2013; O'Connell et al., 2014) a phase reversal is expected when this frequency changes to a non-BF (3,4). Moreover, as stimulus processing is focused on the triplet (presented at fB) in the case of stream integration but can focus on either of the two separated streams (presented at fB and fA) in the case of stream segregation, we assumed a high phase consistency at fB for the former, and a lower/higher phase consistency at fB/fA for the latter, respectively (shown schematically as bars). This assumption was used for "channel picking" (see Materials and Methods).

Materials and Methods

Subjects

In the present study, we analyzed the electrophysiological data recorded during 10 penetrations of area A1 of the auditory cortex of one female rhesus macaque weighing 6 kg, who had been prepared surgically for chronic awake electrophysiological recordings. Recordings in another subject are ongoing. 2 penetrations were excluded due to excessive movement or electrical artefacts. Data for Experiment 1 was recorded in 7 of the remaining 8 penetrations, and data for Experiment 2 was recorded in 7 penetrations (see below for the description of the two experiments). Before surgery, the animal was adapted to a custom-fitted primate chair and to the recording chamber. All procedures were approved in advance by the Animal Care and Use Committee of the Nathan Kline Institute.

Surgery

Preparation of subjects for chronic awake intracortical recording was performed using aseptic techniques, under general anesthesia, as described previously (Schroeder et al., 1998). The tissue overlying the calvarium was resected and appropriate portions of the cranium were removed. The neocortex and overlying dura were left intact. To provide access to the brain and to promote an orderly pattern of sampling across the surface of the auditory areas, plastic recording chambers (Crist Instrument) were positioned normal to the cortical surface of the superior temporal plane for orthogonal penetration of area A1, as determined by preimplant MRI. Together with socketed Plexiglas bars (to permit painless head restraint), they were secured to the skull with orthopedic screws and embedded in dental acrylic. A recovery time of 6 weeks was allowed before we began data collection.

Electrophysiology

During the experiments, the animal sat in a primate chair in a dark, isolated, electrically shielded, sound-attenuated chamber with head fixed in position, and was monitored with infrared cameras. Neuroelectric activity was obtained using linear array multicontact electrodes (23 contacts, 100 um intercontact spacing, Plexon). The multielectrodes were inserted acutely through guide tube grid inserts, lowered through the dura into the brain, and positioned such that the electrode channels would span all layers of the cortex, which was determined by inspecting the laminar response profile to binaural broadband noise bursts. Neuroelectric signals were impedance matched with a preamplifier (10X gain, bandpass dc 10 kHz) situated on the electrode, and after further amplification (500X) they were recorded continuously with a 0.01–8000 Hz bandpass digitized with a sampling rate of 40 kHz and precision of 16 bits using custom-made software in Labview. The signal was split into the local field potential (LFP; 0.1–300 Hz) and multiunit activity (MUA; 300–5000 Hz) range by zero phase shift digital filtering. MUA data were also rectified to improve the estimation of firing of the local neuronal ensemble (Legatt et al., 1980). One-dimensional CSD profiles were calculated from LFP profiles using a three-point formula for the calculation of the second spatial derivative of voltage (Freeman and Nicholson, 1975). The advantage of CSD profiles is that they are not affected by volume conduction like the LFP, and they also provide a more direct index of the location, direction, and density of the net transmembrane current flow (Mitzdorf, 1985; Schroeder et al., 1998). At the beginning of each experimental session, after refining the electrode position in the neocortex, we established the “best frequency” (BF) of the recording site using a “suprathreshold” method (Steinschneider et al., 1995; Lakatos et al., 2005). The method entails presentation of a stimulus train consisting of 100 random order occurrences of a broadband noise burst and pure tone stimuli with

frequencies ranging from 353.5 Hz to 32 kHz in half-octave steps (duration: 100 ms, r/f time: 5 ms; inter-stimulus interval, ISI: 624.5 ms). Auditory stimuli for tonotopy and for the behavioural task were generated at 100 kHz sampling rate in Labview using a multifunction data acquisition device (National Instruments DAQ USB-6259), and presented through SA1 stereo amplifiers coupled to FF1 free field speakers (Tucker-Davis Technologies). Loudness was calibrated using measurements made with an ACO Pacific PS9200/4012 calibrated microphone system.

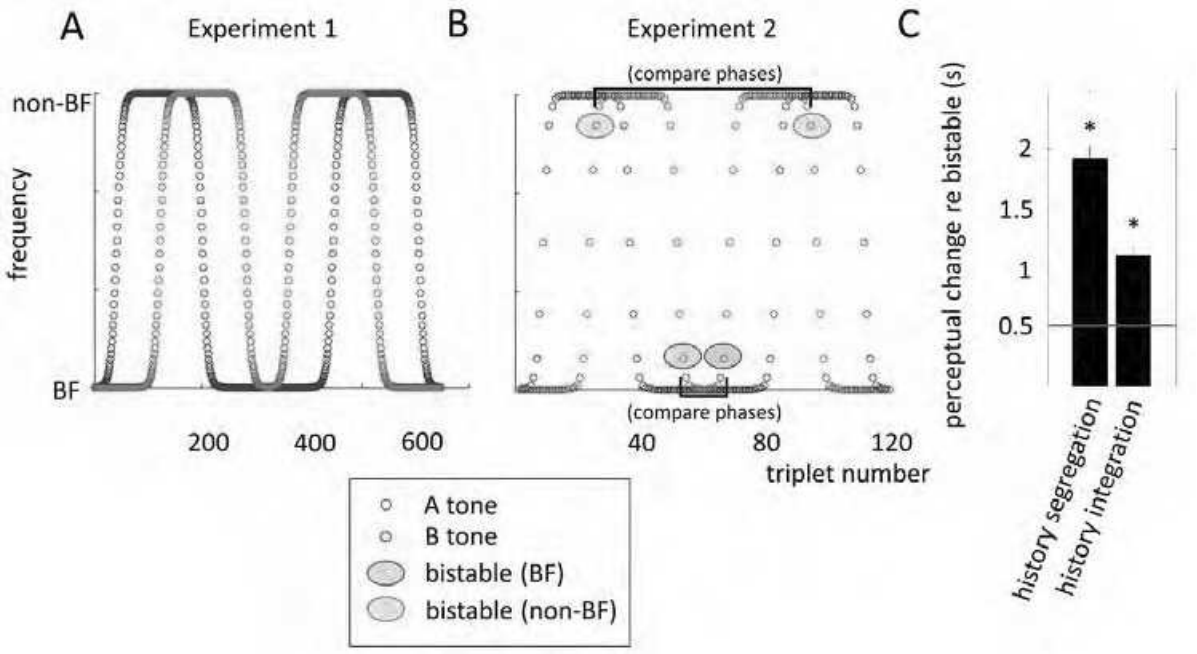


Figure 2. Stimulation protocol in Experiment 1 (A) and Experiment 2 (B). The frequencies of A tones (blue) and B tones (red) were varied systematically between BF and non-BF of the recorded site, resulting in a continuous change in the assumed percept. This change was slower for Experiment 1 (changing from "complete" integration – i.e. A and B tones having the same frequency – to "complete" segregation – i.e. either A or B tone at BF and the other at non-BF – within 81 triplets) than for Experiment 2 (changing from "complete" integration to "complete" segregation within 15 triplets). The pattern shown in A for Experiment 1 was repeated 4 times and that shown in B for Experiment 2 was repeated 15 times. Note that, for each pattern, every constellation between A and B tones appears twice, but the two occurrences in each pair differ by their "stimulation history" (one preceded by integration, the other by segregation). The phases for pairs at a bistable constellation (shown in green or brown in B) could be statistically compared in Experiment 2, since here they occurred numerous (15) times each. This gave us the possibility to compare neuronal activity related to physically identical stimuli with potentially differing percepts ("biased" by the respective "stimulation history"). C. The "perceptual bias" by "stimulation history" was verified in (4) human subjects. Subjects had to indicate perceptual changes in the stimulation protocol shown in B by a button press. The time of button press (averaged across subjects; the median across trials was used for within-subject analysis in order to take into account the commonly skewed distributions of reaction time) with respect to the bistable constellations shown in B are shown, for different "stimulation histories". Button presses occurred well after the bistable

constellations (and significantly later than a common reaction time, shown in red) for both "stimulation histories", indicating that, indeed, perception was biased by "stimulation history".

Experimental design

Neural mechanisms of stream segregation and integration were investigated in two experiments using the ABA paradigm (Van Noorden, 1975). In these experiments, the frequencies of A and B tones were varied systematically (Figure 2), cycling between BF and a non-BF, the latter chosen, for each penetration, as the frequency two octaves above or below the BF of the respective recording site. Two octaves were chosen since psychophysical experiments show that when the frequency of A and B tones is 2 octaves apart, the percept is always segregation of A and B streams (e.g., Deike et al., 2012). Either A or B tones gradually changed their frequency, but never both at the same time. Thus, the assumed percept gradually changed from integration (when the frequency of A and B tones were similar or equal) to segregation (when the frequencies were far apart) and vice versa. The speed of this change was the main difference between the two experiments: The tone frequencies changed from BF to non-BF (or vice versa) within 81 triplets in Experiment 1 (Figure 2A), and within 15 triplets in Experiment 2 (Figure 2B). As Figure 2 demonstrates, A and B tones changed their frequencies such that each possible constellation of tones occurred, with respect to tone frequency and direction of change. The pattern shown in this graph (comprising 648 triplets for Experiment 1 and 120 triplets for Experiment 2) was repeated 4 times for Experiment 1 (resulting in a total number of 2592 triplets) and the respective pattern was repeated 15 times for Experiment 2 (resulting in a total number of 1800 triplets). For one of the recordings in Experiment 2, only the first 60 (of 120) triplets of the pattern was presented, but 25 times (resulting in a total number of 1500 triplets).

For all experiments, tones had a duration of 25 ms (including 5 ms rise/fall time). All stimuli were presented at 40 dB SPL. An ISI of 121 ms between tone onsets was used (similar to Micheyl et al., 2005), resulting in a presentation rate of A tones of 4.13 Hz ("*fA"*) and in a presentation rate of B tones of 2.07 Hz ("*fB"*).

Data analyses

Several predictions can be made based on previous findings on "the rules of entrainment" and our main hypothesis, and all of them are summarized in Figure 1. As explained above, all of them are centered on the idea that the neural phase should co-vary with the assumed percept and therefore behave differently when the latter changes from stream segregation to stream integration or vice versa. We expected this change in phase to be maximal at *fB*, as it reflects both the presentation rate of B tones and that of the ABA triplet. We furthermore expected – in the case of B tones having a frequency that corresponds to the BF of the recording site – the neural high excitability phase to be centered on B tones in the case of stream segregation (1 in Figure 1), and to be centered on the beginning of the triplet (i.e. on the first A tone) in the case of stream integration (2 in Figure). A reversal of phase was expected in the case of B tones having a frequency that corresponds to the non-BF of the recording site (3 and 4 in Figure 1; O'Connell et al., 2011). When phase differences between trials of stream segregation and integration are mentioned in the following, we always mean the circular difference "integration minus segregation related phases".

As a first step, A or B tones were labeled "BF" if their frequency was BF ± 6 semitones of the recording site or "non-BF" if it was non-BF ± 6 semitones. Based on the results of previous psychophysical experiments in humans (e.g., Deike et al., 2012; Moore and Gockel, 2012), the percept "segregated" was assumed if A tones were BF and B tones were non-BF, or vice

versa (1 and 3 in Figure 1). The percept "integrated" was assumed if A and B tones were either both BF or both non-BF (2 and 4 in Figure 1). Note that for both percepts, two possible constellations of A and B tones exist (the constellation number in the following corresponds to the respective number in Figure 1): For segregation, A tones could be non-BF and B tones BF (constellation 1), or A tones could be BF and B tones non-BF (constellation 3). For integration, both A and B tones could be BF (constellation 2), or both could be non-BF (constellation 4). Thus, two possibilities exist to compare trials of segregation and integration: 1 vs. 4 and 3 vs. 2, but also 1 vs. 2 and 3 vs. 4. Note that for these comparisons, one tone (A or B) is always fixed to BF or non-BF, respectively, and the other is BF for stream segregation but non-BF for stream integration (or vice versa). For the first possibility (1 vs. 4 and 3 vs. 2), A tones are fixed to BF or non-BF, respectively, and B tones are either BF or non-BF, depending on the assumed percept. However, this comparison can be contaminated, as the change of B tones from BF to non-BF (for a given comparison) can affect the phase at *fB*, as the latter reflects both the presentation rate of B tones (resulting in rhythmic evoked potentials at a rate of *fB only* if B is BF) and that of the ABA triplet. This is not the case for the second possibility (1 vs. 2 and 3 vs. 4): The presentation rate of A tones (*fA*) is not equal to the rate of triplets, making a contamination less likely. We therefore chose the second possibility for our comparison of stream segregation and integration (i.e. 1 vs. 2 and 3 vs. 4 in Figure 1) and focused the analysis on neural phases at the onset of B tones. In order to estimate oscillatory phases, data were downsampled to 500 Hz and instantaneous phase in single trials was extracted by wavelet decomposition (Morlet wavelet of order 6). In order to minimize an artificial bias of phase estimation (Zoefel and Heil, 2013), a linear interpolation was applied to the single trials before wavelet analysis in the 0–100 ms time interval (with

respect to B tone onset) which in the case of most BF tones contained evoked-type activation (Lakatos et al., 2013).

All analyses were done separately for supragranular, granular, and infragranular layers. For each layer, one channel was chosen based on two different criteria: First, we used the laminar profile in response to the BF as obtained in the "suprathreshold method" which normally results in clear sinks and sources that can readily be assigned to the different layers (e.g., Lakatos et al., 2013). Second, we based channel selection on the following idea: We expected neural phases at *fB* to be more stable for stream integration than for segregation. This is because, in the case of integration, the oscillatory cycle is assumed to consistently and invariably group the ABA triplet, whereas, in the case of segregation, the attentional focus has to be divided between two streams (A stream and B stream), resulting in oscillations that sometimes track one stream (at *fB*) or the other (at *fA*) – and inevitably in phases that are more variable and less consistent. Based on Lachaux et al. (1999; their PLV), this phase consistency can be quantified as the inter-trial coherence (ITC), ranging between 0 (no phase consistency across trials) and 1 (perfect phase consistency across trials), and calculated as follows:

$$ITC(t,f) = |\frac{1}{N}\sum_{n=1}^{N} e^{i(\varphi_n(t,f))}|$$

where $\varphi_n(t,f)$ is the phase of trial *n* at time *t* and frequency *f*, and *N* is the number trials. We expected the ITC at *fB* to be larger for integrated trials than for segregated trials. Thus, we calculated an "ITC profile" *pr* for each recording as follows

$$pr\ (chan) = ITC(t,f,chan,integration) - ITC(t,f,chan,segregation)$$

where *t* and *f* are set to the onset of B tones and to *fB*, respectively. We then assigned those channels to the three layers (supragranular, granular, infragranular) that showed a peak in this profile.

We focused on *fB* for our procedure of "channel picking", in order to keep it as simple as possible. Logically, though, we also expected the ITC at *fA* to be larger for segregated trials than for integrated trials, as only in the case of stream segregation, A tones are perceived as an independent stream (otherwise they are fused into an integrated stream at half the frequency). Therefore, it is possible to characterize the response of the chosen channels as the ratio between the ITC at *fB* and the ITC at *fA.* As shown in Figure 3, this ratio is higher for integrated than for segregated trials (significant for supragranular layer, $t(26) = 2.62$, $p = 0.015$, and for infragranular layer, $t(26) = 2.80$, $p = 0.010$, but not for granular layer, $t(26) = 2.03$, $p = 0.053$; Student's t-Test). This might indicate that neural oscillations followed the ABA triplet in a more consistent manner when stream integration occurred (as phase consistency was higher at the frequency of the triplets than at the frequency of the A tones) than when the two streams were assumed to be perceptually segregated (for integrated trials, the ratio between ITC at *fB* and ITC at *fA* is significantly larger than 1 for supragranular layer, $t(13) = 2.62$, $p = 0.021$ and for granular, $t(13) = 2.34$, $p = 0.036$, but not for infragranular layer, $t(13) = 1.97$, $p = 0.070$; Student's t-Test*)*. In the case of stream segregation, the ITC ratio shifted towards a more counterbalanced relation between ITC at *fB* and ITC at *fA*, indicating a balanced tracking of the two simultaneous streams (for segregated trials, the ratio between ITC at *fB* and ITC at *fA* is not significantly different from 1 for supragranular layer, $t(13) = -0.59$, $p = 0.563$ and for granular, $t(13) = 0.13$, $p = 0.900$, and even significantly smaller than 1 for infragranular layer, $t(13) = -2.22$, $p = 0.045$; Student's t-Test. However, we note that this neural behavior – at least partly – depends on

our method of "channel picking": We therefore can only describe it, but not treat it as an actual result.

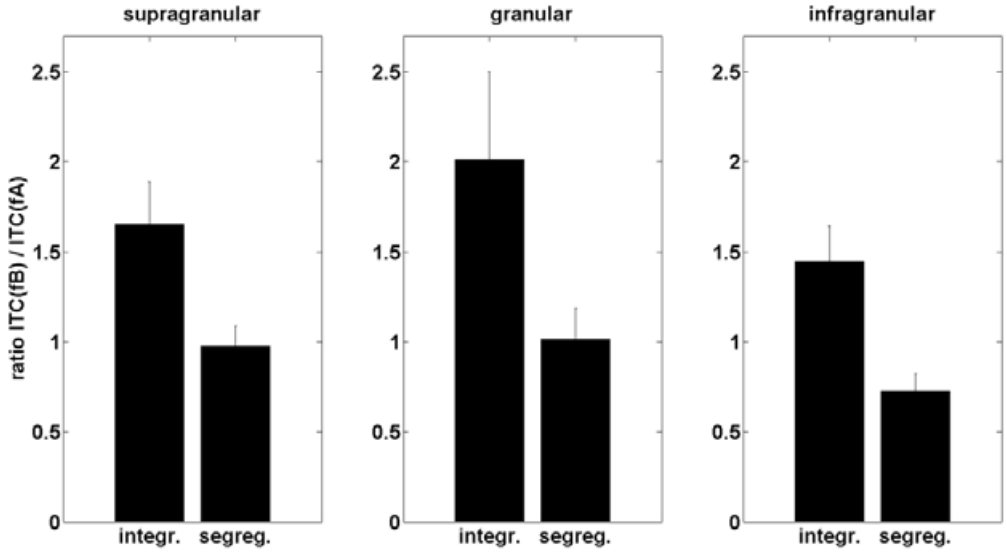


Figure 3. Ratio of phase consistency (measured as inter-trial coherence, ITC) at fB and fA for the different layers in A1, averaged across recordings (pooled across experiments; thus, there are 14 recordings).For all layers, this ratio is higher for an assumed stream integration than for stream segregation, indicating that the neural phase mainly followed the ABA triplet in the case of stream integration and was divided between two separated streams (one presented at fB, the other at fA) in the case of stream segregation. Standard error of the mean across recordings is shown as error bars.

Note that this difference in ITC (ratio) is independent from the question whether the phases for segregation and integration differ. We tested this hypothesis as follows: For each recording, CSDs, downsampled to 125 Hz and epoched in windows of -250 ms to +250 ms around B onset, were sorted into four "categories" (segregation/BF, integration/BF, segregation/non-BF, integration/non-BF; Figure 1) and averaged across trials within each category. Example average signals for each category from a supragranular channel in one recording are shown in Figure 5A. A sine wave with frequency *fB* was fitted to each average (i.e. to the signals shown in Figure 5A, and the corresponding signals for other recordings) and its phase was extracted for each category and recording. We expected a difference in fitted phase for segregation vs. integration. In order to test for significance of this phase difference, we developed an analysis based on the following logic: Assuming different

phases for integration and segregation trials, phases should be more coherent for both segregation and integration trials separately than phases across all trials.

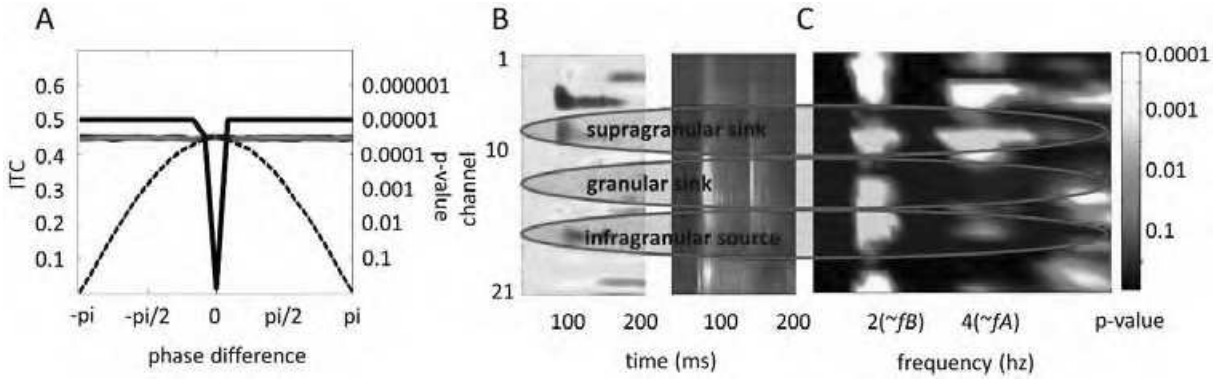


Figure 4. A. When two phase distributions with different circular mean are shuffled, the ITC of that shuffled distribution (dashed black line) is lower than the ITC calculated separately for the two distributions (red and blue lines). This mathematical fact underlies our analysis, in which the ITC, calculated separately for stream segregation and integration, is compared with a surrogate distribution in which trials of stream segregation and integration were shuffled. In our simulation for A, even small phase differences between the distributions result in high p-values (continuous black line). An example plot for significance values found in the "real" data is shown in C. Note that the significance "blobs" seem to be spatially and spectrally specific, as (1) they seem to correspond to the different layers in A1 (their profile shown in B), and (2) they are only visible for frequencies involved in the paradigm: fB and fA.

In other words, under the hypothesis of a phase difference between segregation and integration at time point t and frequency f,

$$\frac{ITC_{segregation(t,f)} + ITC_{integration(t,f)}}{2} > \frac{ITC_{all1(t,f)} + ITC_{all2(t,f)}}{2}$$

whereas the null hypothesis predicts

$$\frac{ITC_{segregation(t,f)} + ITC_{integration(t,f)}}{2} = \frac{ITC_{all1(t,f)} + ITC_{all2(t,f)}}{2}$$

where $ITC_{all1}$and ITC_{all2} reflect the ITC across the first and second half of all trials, respectively, but independently of the assumed percept. The split into two halves of trials assures that both sides of the equations above consist of the same number of trials. Our notion was underlined by simulations, as illustrated in Figure 4A: Here, several pairs of phase distributions were constructed, with a certain (circular) mean phase difference between

each pair. When the ITC was calculated for each phase distribution separately, the ITC was ~0.45 (blue and red lines in Figure 4A), of course irrespectively of the phase difference between the pairs. However, when the phases of each pair were shuffled and the ITC was re-calculated, the ITC of this shuffled distribution was well below 0.45 (black dashed line), even for very small phase differences between the two distributions. For our analysis, we thus calculated the ITC, separately for segregation and integration, and compared the average with a surrogate distribution (reflecting ITC_{all}), obtained by a permutation procedure. Here, data for segregation and integration were shuffled and split into two halves ($ITC_{all1}$and ITC_{all2}) before re-calculating the average ITC across the two halves. This procedure was repeated 1,000,000 times. Thus, it was possible to obtain a range of ITC values under the null hypothesis of no phase difference between segregation and integration. P-values were obtained for the recorded data by comparing "real" ITC values (averaged across recordings and BF/nonBF categories) with the surrogate distributions (averaged likewise). For our simulation in Figure 4A, those p-values are shown as a continuous black line, showing that, even for very small phase differences, significant p-values are obtained and confirming the validity of this analysis. For the actual analysis, p-values were corrected for multiple comparisons using FDR (Benjamini and Hochberg, 1995). An example plot of the outcome, for one representative recording is shown in Figure 4C. Several "blobs" of significance are visible – importantly, these "blobs" are spatially and spectrally specific, as (1) they seem to correspond to the different layers in A1, as shown in Figure 4B, and (2) effects are found for *fB* and *fA*, with no significant effects at other frequencies.

For all analyses described so far, the frequency of stimuli (i.e. the frequencies of A and B tones) differed between the assumed percepts (i.e. stream segregation and integration), making them difficult to compare. We provided a solution for this in Experiment 2: As visible

in Figure 2, in our stimulation protocol, every constellation between A and B tones appeared twice, but the two occurrences in each pair differ by their "stimulation history": One occurrence is preceded by an assumed percept of segregation, the other by a percept of integration. As the change from stream segregation to integration (and vice versa) was relatively rapid in Experiment 2, we assumed that, at bistable constellations (two pairs of which are present in our protocol, shown in green or brown in Figure 2B), the percept might be influenced by this "history". Thus, this experimental protocol had the advantage of potentially differing percepts ("biased" by the respective "stimulation history") despite identical stimulation. To test this, we compared neural phases at these pairs of bistable constellations (± 3 trials, in order to increase the number of trials), expecting different phases for the two occurrences in each pair, depending on the "stimulation history". For each pair and each of the two occurrences, the circular mean across repetitions and trials was calculated: As phases were extracted at the respective bistable constellation ± 3 trials and there were 15 repetitions per recording, this means that, for each occurrence, the circular mean was determined across 7*15=105 phases. For each of the two pairs, the circular difference between the two mean phases (for each pair, there is one mean phase for "stimulation history" segregation, and one mean phase for "stimulation history" integration) was calculated and pooled across pairs and recordings. Significance of these phase differences was calculated by means of a permutation test as explained above. For one of the seven recordings in Experiment 2, this analysis could not be performed, as the pattern shown in Figure 2B was shortened to 60 (instead of 120) triplets.

All analyses were performed in MATLAB, using the toolbox for circular statistics (Berens, 2009) where appropriate.

Results

We designed experiments in order to test neural mechanisms underlying auditory stream segregation and integration. One monkey was presented with triplets of tones ("ABA") that can either be perceived as two separated streams (A stream and B stream) or as a single stream of triplets. We hypothesized that the neural phase at a frequency corresponding to the presentation rate of both ABA triplets and B tones (*fB)* reflects this percept. We further assumed that the high excitability part of neural oscillations is centered on the B tone in the case of segregated streams (if the frequency of B tones corresponds to the "best frequency", BF, of the recording site; 1 in Figure 1). We expected a phase shift in the case of integrated streams, such that the high excitability part is centered on the beginning of the triplet (i.e. on the first A tone; 2 in Figure 1). Moreover, based on previous research (e.g., O'Connell et al., 2011), we expected a reversal of phase for both segregated and integrated percepts when oscillations are recorded in regions of A1 tuned to frequencies not corresponding to the frequency of B tones (non-BF; 1 and 2 vs. 3 and 4 in Figure 1). Thus, our experimental protocol resulted in four trial "categories": segregation/BF, segregation/non-BF, integration/BF, integration/non-BF (see Materials and Methods and Figure 1). Our expectations were tested in two experiments, where the frequencies of A and B tones were systematically varied such that the monkey's percept could be biased based on the frequency separation between A and B tones. Note that in these experiments, the exact stimulation (i.e. the frequencies of A and B tones) might differ between percepts (i.e.

segregation and integration). Thus, we additionally compared phases extracted under identical stimulation but with differing "stimulation history". In short, our results suggest that, in all experiments and even in the case of invariant stimulation, the phase of neural oscillations reflects whether two simultaneous auditory input streams are perceived as two segregated or one integrated stream(s).

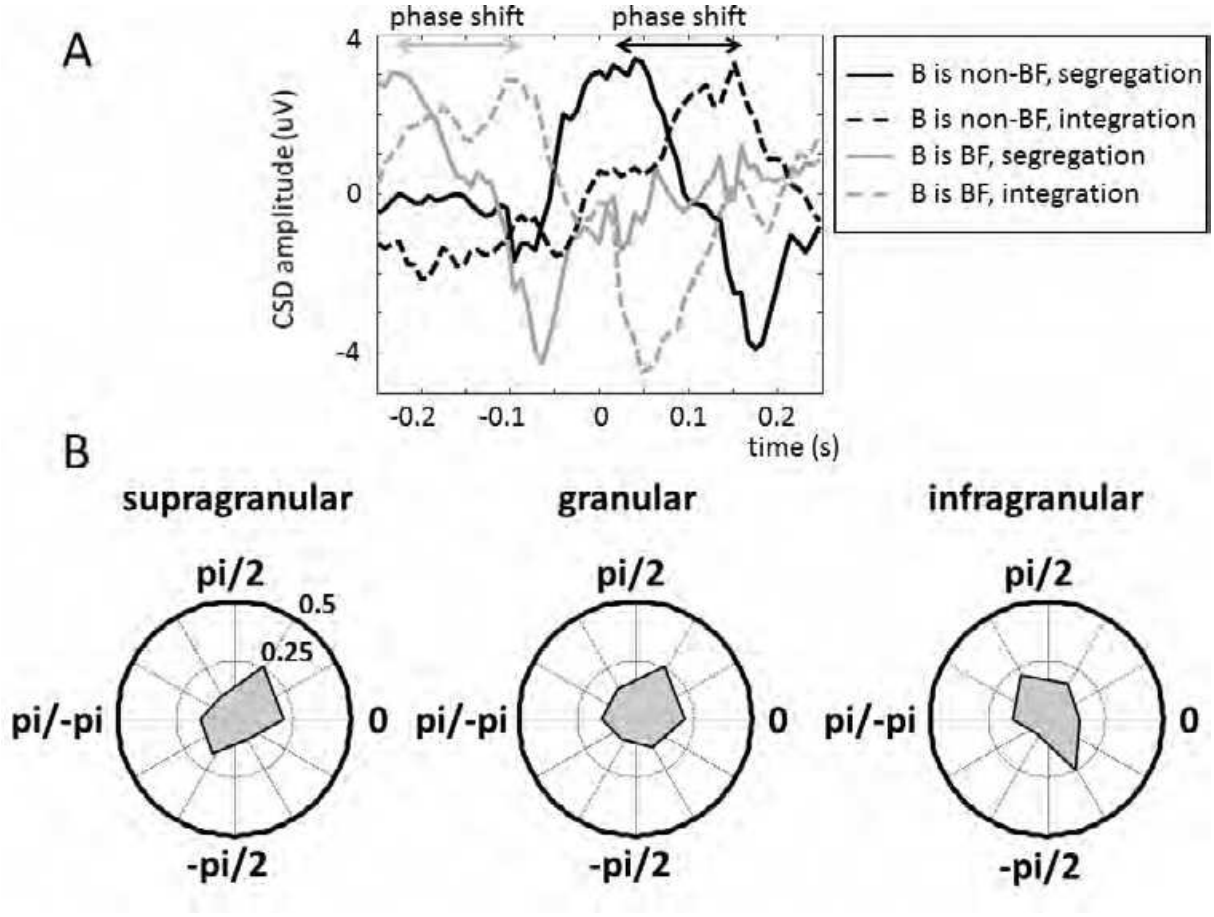


Figure 5. A change in the assumed percept goes along with a change in neural phase. A. CSD signals, averaged across trials within each category for a supragranular channel in a single recording. For each category, an oscillation is apparent whose frequency corresponds to fB (~2 Hz). This oscillation exhibits a phase difference when the assumed percept changes (compare the two black and the two gray lines). Time 0 corresponds to the onset of B tones. B. Circular phase differences of sine waves (frequency fB) fitted to the signals shown in A (divided into seven bins and depicted as the percentage of phase differences falling into the respective bin), pooled across experiments, recordings and BF/non-BF categories. For all three layers, a phase shift is visible that approximates $\frac{\pi}{2}$.

Percept related changes in the phase of neural entrainment

We found that a change in the assumed percept from stream segregation to stream integration resulted in a change in phase consistency, with a larger ratio between the ITC at

fB and that at *fA* for an integrated than a segregated percept (Figure 3). This result partly was due to our method of "channel picking" (see Materials and Methods). However, it does not allow to draw any conclusions about the exact neural phases underlying this change in percept: It might be that, although phase consistency is affected, the phases themselves do not differ between stream segregation and integration (as hypothesized in Figure 1). For each recording, we therefore fitted sine waves at *fB* to the averaged CSD signals for each trial category (see Materials and Methods) and compared the fitted phases across percepts. As an example, the averaged CSD signals from the supragranular layer in one recording are shown in Figure 5A. Importantly, a phase shift can be seen for each pair of signals corresponding to a change in percept (segregation/BF vs. integration/BF, and segregation/non-BF vs. integration/non-BF). Results are shown in Figure 5B when circular phase differences (at the onset of B tones) are pooled across recordings, across experiments, and across BF/non-BF categories. Although the effect is rather weak, the phase difference for segregation vs. integration that was already observed in Figure 5A is visible now as a phase shift that seems to approximate $\frac{\pi}{2}$, with stream segregation lagging behind stream integration. Interestingly, this phase shift would correspond to the interval between A and B tones in the ABA triplet. Assuming the oscillatory high excitability phase to be centered on the B tone for stream segregation, in contrast to our expectation (2 in Figure 1), this result would indicate that it is centered on the second A tone (i.e. on the triplet offset) when the stimulus is perceived as a single stream of ABA triplets.

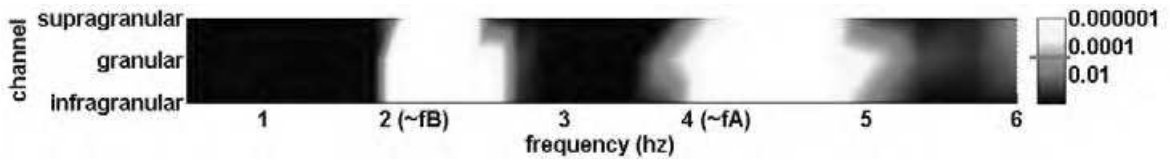


Figure 6. P-values (color-coded), as obtained by a permutation procedure, for the neural phase differences between stream segregation and integration shown in Figure 5 (corresponding to fB) and at other neural frequencies. Significant phase differences at fB and fA are obtained for all layers (but for no other frequencies). Note that only three channels (one for each

layer) are shown; data between channels are interpolated. The FDR-corrected significance threshold is marked in red on the colorbar.

In order to test the significance of this effect, we designed a permutation test (see Material and Methods and Figure 4). Results of this test (averaged across experiments and across BF/non-BF categories) are shown in Figure 6 for the different layers of A1. As it can be seen, for all layers, phases at *fB* significantly differ between percepts. Note that a significant phase difference is also found for *fA*. However, we refrain from interpreting this difference as a functional role for oscillations at *fA* during stream integration is difficult to assign and contamination by evoked potentials cannot be ruled out. Importantly, we do not find an effect for any other frequency but *fB* and *fA*, indicating that our observed effect is specific for frequencies involved in our stimulation protocol.

Although in evidence for the single recording shown in Figure 5A, in contrary to our expectation (1 and 2 vs. 3 and 4 in Figure 1), we did not observe any significant phase difference (e.g., a reversal in phase) between oscillatory signals for BF vs. non-BF stimuli when pooled across experiments and recordings (data not shown). This finding is in contrast to previous studies (O'Connell et al., 2011; Lakatos et al., 2013) and merits further investigation in future experiments.

Taken together, our results indicate an important role of the phase of neural oscillations for stream segregation and integration and ultimately, for the formation of auditory objects. However, in the analyses described thus far, the actual stimulation (i.e. the frequencies of A and B tones) differed between the assumed percepts (i.e. stream segregation and integration), making them challenging to compare due to possibly differing bias related to evoked event related components. Thus, we designed an additional analysis of data

recorded in Experiment 2, where phases observed under identical stimulation but potentially differing percept were compared.

Stimulation history related neural phase shift

In the course of the experimental protocol applied in Experiment 2, certain supposed bistable constellations of A and B tones were present: For each recording, there were two pairs of bistability (green and brown patches in Figure 2B). Importantly, for each pair, one occurrence of bistability was directly preceded by stream segregation, whereas the other was preceded by stream integration (compare the "stimulation history" of the patches with identical color in Figure 2B). As the experimental protocol in Experiment 2 changed very rapidly from segregation to integration and vice versa, it might thus be that the "history" of stimulation affects perception at the time of bistability: towards the percept of the preceding trials. We verified this hypothesis in a psychophysical experiment on human subjects (N = 4) in which the experimental protocol of Experiment 2 (Figure 2B) was presented. We asked subjects to press a button immediately after their percept switched from stream segregation to integration, or vice versa. Indeed, we found that perception was biased by preceding trials: For both "stimulation histories", the indicated perceptual change occurred significantly later than 500 ms after the bistable constellations (history segregation: $t(3) = 9.63$, $p = 0.002$; history integration: $t(3) = 5.26$, $p = 0.013$). The "threshold" of 500 ms was chosen conservatively, because it is a commonly observed reaction time in response to near-threshold sounds that are, by definition and in contrast to stimuli applied here, difficult to detect (Zoefel and Heil, 2013). Thus, our finding indicates that, at the bistable constellations shown in Figure 2B (shaded patches), the percept at each pair of identical

stimulation is indeed biased towards the percept evoked by the preceding stimulation (at least in human subjects).

Thus, for each pair and recording in our monkey experiment, we compared the circular mean phase at *fB* across trials where the preceding trials were characterized by stream segregation with those with a "history" of stream integration (note that, for each pair, there are 105 trials per "stimulation history"; see Materials and Methods). Those phase differences, pooled across pairs and recordings, are shown in Figure 7A: Again, a phase difference of $\sim\frac{\pi}{2}$ is visible, reflecting a phase shift corresponding to the interval between A and B tones in the ABA triplet. P-values of this phase shift were determined in a permutation procedure similar to the one above (Figure 6), for all three layers and for phases at several neural frequencies. Results of this test are shown in Figure 7B: Importantly, the observed phase shift shown in Figure 7A (at *fB*) is significant for all layers – and this effect is frequency-specific, in that significance is only obtained for neural frequencies corresponding to our rhythm of stimulation (i.e. *fB* and *fA*; the effect at frequencies higher than *fA* is probably due to frequency smearing). Thus, under invariant stimulation, differences in "stimulation history", assumed to reflect differences in the monkey's percept, go along with a significant shift in the phase of neural oscillations. Due to the identical stimulation at the time of phase extraction, the observed phase difference must reflect an *endogenous* phase shift of neural oscillations, with the oscillatory cycle either segregating the triplet into two separate auditory streams (1 and 3 in Figure 1), or grouping the ABA triplet into an integrated auditory object (2 and 4 in Figure 1).

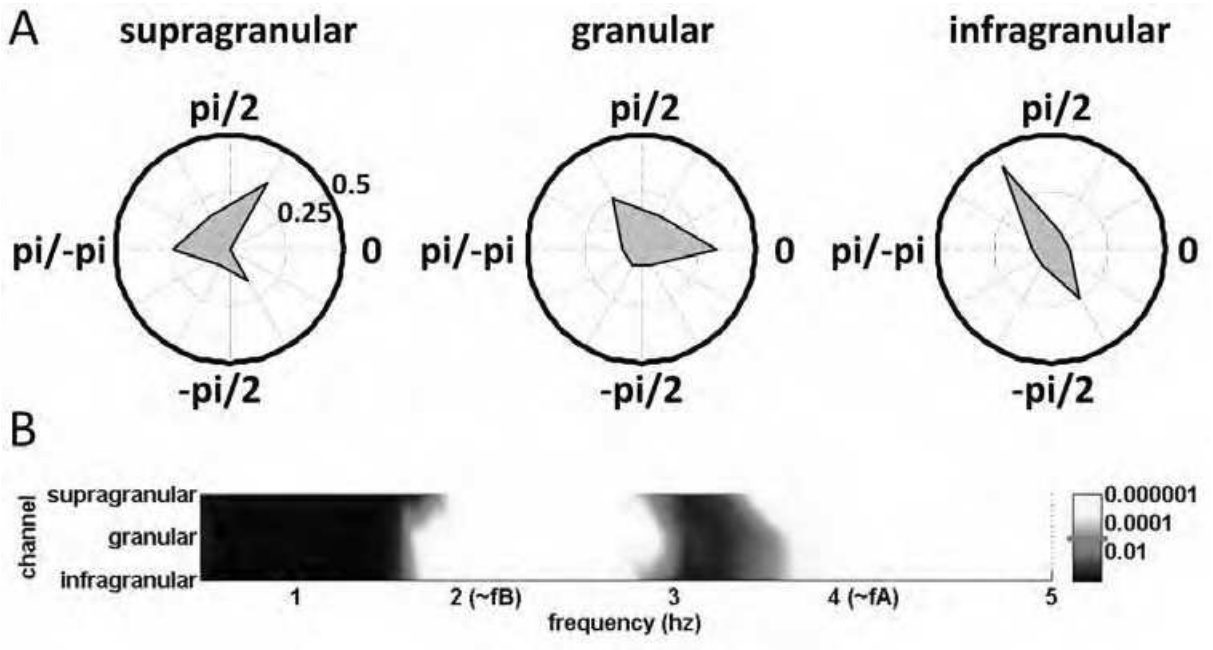


Figure 7. Circular phase differences and corresponding MUA responses between trials with different "stimulation history". A. Phases at fB were extracted from bistable ABA constellations in Experiment 2. These constellations had identical stimulation but were preceded either by stream segregation or stream integration (colored patches in Figure 2B). The circular phase differences between trials of different "stimulation history" is shown, pooled across pairs of identical stimulation (green and brown in Figure 2B) and across recordings. Phase differences are divided into seven bins and depicted as the percentage of phase differences falling into the respective bin. B. P-values (color-coded), obtained by a permutation test, for the phase shift shown in A (corresponding to fB) and at other neural frequencies. Significant phase differences at fB and fA are obtained for all layers (but for no other frequencies; the significant effect at frequencies higher than fA is probably due to frequency smearing). Note that only three channels (one for each layer) are shown; data between channels are interpolated. The FDR-corrected significance threshold is marked in red on the colorbar.

Discussion

Already more than 80 years ago, it has been recognized that the phase of neural oscillations reflects rhythmic changes in neural excitability (Bishop, 1932). This property makes the neural phase an ideal candidate for the parsing of an input stream: The oscillation's high excitability phase might represent the "focus" of input processing within each oscillatory cycle, whereas the low excitability phase might represent the "parser" between different cycles.

In the auditory world, the brain continuously has to group individual auditory elements and segregate the streams these form. Single auditory elements that are segregated, based on their spectrotemporal properties, are assigned to different auditory objects; auditory elements that are grouped are perceived as one auditory object (e.g., Nelken et al., 2014). It seems to be a logical assumption that the phase of neural oscillations might play an important role for the segregation and integration of auditory streams. It is all the more surprising that – although the formation of auditory objects are characterized thoroughly on a psychophysical basis (Bregman, 1994; Griffiths and Warren, 2004; Rose and Moore, 2005; Moore and Gockel, 2012) – the underlying neural mechanisms are poorly understood and only a few studies have built the link between the segregation and integration of auditory streams and the phase of neural oscillations.

The "cocktail party effect", describing a multi-speaker scenario in which auditory streams of single speakers have to be segregated from others (Cherry, 1953), is an example for the importance of stream integration and segregation in the auditory environment. Whereas A and B streams can be considered the two "auditory objects" in our study, the stream of each individual speaker fulfills all criteria of an auditory object in the case of the "cocktail party"

(Simon, 2015). Indeed, for human subjects, it has been found recently that the phase of neural oscillations "tracks" the speech envelope of several speakers in this scenario (Ding and Simon, 2012a; Zion Golumbic et al., 2013). When the intensity of one of the speakers is changed, only the neural representation of that speaker (but not those of others) is enhanced, indicating that the neural phase can encode independent auditory objects (Ding and Simon, 2012b; Simon, 2015). Moreover, the "tracking" of the stream in the attentional focus is enhanced (Ding and Simon, 2012a; Zion Golumbic et al., 2013). Similar effects have been described for non-human primates: In the presence of both an auditory and a visual stream, the high excitability phase of these oscillations is aligned with the attended stream (Lakatos et al., 2008). These results underline an important role for neural oscillations for stream integration and segregation, and for their phase as a tool for input amplification and attentional selection (Schroeder and Lakatos, 2009). Ghitza (2011, 2013) and several other authors (Giraud and Poeppel, 2012; Jensen et al., 2012; Peelle and Davis, 2012; Zion Golumbic et al., 2012) have discussed the role of neural oscillations as a "parsing rhythm" and their considerations have been underlined with psychophysical findings (e.g., Ghitza and Greenberg, 2009; Ghitza, 2012, 2014). Moreover, the existence of a "window of integration" on a perceptual level has often been speculated (for a review, see van Wassenhove, 2009). In our study, results provide compelling evidence that one reflection of this "window of integration" is the phase of neural oscillations and that it can be shifted based on the length of the to be integrated auditory information: This was possible, as the applied "ABA" rhythm can be perceived as being part of the same stream, and does so when hypothetically falling into the oscillatory "window of integration. Indeed, we were able to show that an assumed change from stream segregation to stream integration (or vice versa) goes along with a change in neural phase that corresponds to the interval between A and B tones in the ABA

stimulation. The latter finding is well in line with the notion that the high excitability phase is centered on the individual (B) tone in the case of stream segregation (1 in Figure 1) but otherwise groups the triplet into a single stream (2 in Figure 1). Our findings are in line with results from previous studies: In Kösem et al. (2014), subjects were presented with visual and auditory events with certain delays between them. It was found that the phase of neural oscillations (measured with magnetoencephalogram, MEG) at the end of an experimental block with a certain audiovisual asynchrony is shifted when compared to the beginning of the block, and that this phase shift is correlated with the perceived audiovisual simultaneity as tested after the respective block. Although perception and phase shift have not been tested in the same trials and the link between them is therefore relatively indirect, this result is further evidence for the neural phase being related to the integration of (in this case: auditory and visual) information. Although we concentrated on the auditory system in this study, a similar idea – linking the phase of neural oscillations and feature integration – has been formulated recently for vision (Wutz and Melcher, 2014). It has been argued that visual information is processed over time, and that features belonging to one visual object are processed within one cycle of the alpha band (~7-13 Hz). Indeed, the successful separation of visual targets (separated by less than one alpha cycle) has been associated with a stronger phase-reset of alpha oscillations than for unsuccessful trials, indicating an important role of the oscillatory phase for temporal integration and segregation also for the visual domain (Wutz et al., 2014).

Although the effect was rather weak and more experiments are necessary to substantiate our findings, our results seem to indicate that the phase difference of neural oscillations between trials of stream segregation and integration corresponds to about one tone of the ABA triplet, and that during segregation, oscillations consistently lag behind those during

integration. Assuming that the high excitability phase of those oscillations is centered on B tones during segregation (1 in Figure 1), this would mean that it is centered on the second A tone during integration. We can only speculate why this is the case. First, it is important to note that, although we chose the oscillation for our analysis whose frequency corresponds to the presentation rate of ABA triplets (and B tones), the high excitability part of this oscillation might be too short to include the complete triplet. Therefore, the high excitability part might be centered on the most informative part of the triplet instead. Both the onset and offset of the triplet might be informative, as they are preceded and followed by a longer silence, respectively, and therefore signal a change in the input stream. Humans tend to automatically group sounds and accentuate one (or more) beat within the group (resulting in meter; Essens and Povel, 1985); this meter can even be seen in electrophysiological data (Nozaradan et al., 2012). Which beat is accentuated seems to depend on experience (Iversen et al., 2008), indicating that it is not necessarily the first beat of the group that is considered most important (and on which the high excitability phase would be centered). Indeed, when three physically identical tones are presented in a row, the first *and* the third tone are perceived as most important (Povel and Okkerman, 1981). Although monkeys seem to be able to perceive and follow rhythms in principle (Merchant et al., 2015), first results suggest that they do not perceive meter as humans do (Honing et al., 2012; Fitch, 2013). However, this result does not have to mean that they do not prioritize parts of the ABA triplet when it is grouped into a single stream. Clearly, research is in its infancy here and more work is needed to characterize rhythmic processing and grouping in animals. Finally, although the relation between the phase of neural oscillations and neuronal excitability is well established (e.g., Buzsáki and Draguhn, 2004), the detailed characteristics, linking oscillatory phase and neuronal activity, are currently relatively unclear: Therefore, it might well be that the phase

of lowest excitability (i.e. the hyperpolarizing phase) happens directly after the most ideal high excitability phase of ongoing oscillations. Thus, if an oscillation's high excitability phase has to encompass the whole auditory triplet, the most ideal phase would be shifted to the end of the triplet, in order to avoid a coincidence between low excitability phase and triplet.

Of course, in this study, we did not measure the monkey's percept, as it would have been possible in the case of human subjects. However, evidence based on earlier research suggests that mechanisms of stream segregation for non-human primates operate in a very similar way as for humans, as laid out, for instance, by Fishman et al. (2001). Nevertheless, the availability of psychophysical data for non-human primates would be an important goal for future studies. Moreover, it would be interesting to see whether the results (i.e. the phase shift) observed here are directly transferable to (EEG/MEG) data recorded superficially in human subjects, in which subjects are actually able to report their percept.

Independent of the role of neural oscillations, potential neural mechanisms of stream segregation and integration have been postulated before (reviewed, e.g., in Denham and Winkler, 2014). Bistable constellations of the ABA paradigm have played an important role: Researchers were interested in neural mechanisms that determine whether the perceived rhythm switches from segregation to integration or vice versa. Here, neural adaptation seems to be a hypothesis favored by the multitude of studies. For instance, using single-cell recordings in monkey A1, Micheyl et al. (2005) were able to show a general decrease in neural response magnitude within the temporal course of the experiment. Building on human psychophysical results, showing a tendency to report stream segregation after a sufficient exposure of the (bistable) ABA paradigm (Bregman, 1978; but see Deike et al., 2012), the authors suggest that a "one stream" percept is mainly evoked when the response

to both A and B tones exceeds a certain threshold; due to neural adaptation, after a certain presentation time, this threshold is not reached anymore, explaining the tendency to hear two segregated streams. Data reported by several studies suggest that forward suppression (in the sense of the A tone suppressing the response to the B tone) is an associated mechanism: For instance, Fishman et al. (2001) found that, using ABAB sequences, responses to B tones in monkey A1 were suppressed when the presentation rate increased, the latter associated with a tendency to hear stream segregation (Van Noorden, 1975). Similar results were obtained by Bee and Klump (2004) in the starling's forebrain. These effects can also be observed in humans: Somewhat surprisingly, when subjects reported stream segregation, Gutschalk et al. (2007) found an increased activity in auditory cortex, using fMRI, and Snyder et al. (2006) found an increased P1 component in response to B tones, using MEG. Our results extend the findings described in this paragraph by suggesting that an additional mechanism – besides neural adaptation and forward suppression – might underlie stream integration and segregation. The important link is the finding that the neural phase is associated with a change in neuronal activity (Jacobs et al., 2007; Whittingstall and Logothetis, 2009). Thus, the pattern of neural responses to A and B tones, as an important factor that influences the way that simultaneous auditory streams are processed or perceived (as suggested by the studies described above), can easily be changed by the brain by shifting the phase of neural oscillations, as demonstrated in the present study. Again, human data would be interesting at this point: If a modulation of the oscillatory phase, e.g. by transcranial alternating current stimulation (tACS; Herrmann et al., 2013), would influence the reported percept, e.g. during a bistable ABA paradigm, this would indicate a causal role of entrained brain oscillations for stream segregation and integration.

To conclude, we were able to demonstrate that the phase of neural oscillations exhibits a shift when the assumed percept in a simple auditory scene changes from stream segregation to integration. This finding supports previous, and mostly theoretical, considerations about the neural phase integrating auditory input into auditory streams and reveals an important role of neural oscillations for the formation of auditory objects.

References

Bee MA, Klump GM (2004) Primitive auditory stream segregation: a neurophysiological study in the songbird forebrain. J Neurophysiol 92:1088–1104.

Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B Methodol 57:289–300.

Berens P (2009) CircStat: A Matlab Toolbox for Circular Statistics. J Stat Softw 31:1–31.

Bishop GH (1932) Cyclic changes in excitability of the optic pathway of the rabbit. Am J Physiol 103:213–224.

Bregman AS (1978) Auditory streaming is cumulative. J Exp Psychol Hum Percept Perform 4:380–387.

Bregman AS (1994) Auditory Scene Analysis: The Perceptual Organization of Sound, Revised. Cambridge, MA: MIT Press.

Bregman AS, Campbell J (1971) Primary auditory stream segregation and perception of order in rapid sequences of tones. J Exp Psychol 89:244–249.

Buzsáki G, Draguhn A (2004) Neuronal oscillations in cortical networks. Science 304:1926–1929.

Cherry CE (1953) Some experiments on the recognition of speech, with one and with two ears. J Acoust Soc Am 25:975–979.

Dehaene S (1993) Temporal Oscillations in Human Perception. Psychol Sci 4:264–270.

Deike S, Heil P, Böckmann-Barthel M, Brechmann A (2012) The Build-up of Auditory Stream Segregation: A Different Perspective. Front Psychol 3:461.

Denham SL, Winkler I (2014) Auditory perceptual organization. In: Oxford Handbook of Perceptual Organization, Johan Wagemans. Oxford University Press.

Ding N, Simon JZ (2012a) Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. J Neurophysiol 107:78–89.

Ding N, Simon JZ (2012b) Emergence of neural encoding of auditory objects while listening to competing speakers. Proc Natl Acad Sci U S A 109:11854–11859.

Elhilali M, Ma L, Micheyl C, Oxenham AJ, Shamma SA (2009) Temporal coherence in the perceptual organization and cortical representation of auditory scenes. Neuron 61:317–329.

Essens PJ, Povel DJ (1985) Metrical and nonmetrical representations of temporal patterns. Percept Psychophys 37:1–7.

Fishman YI, Reser DH, Arezzo JC, Steinschneider M (2001) Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. Hear Res 151:167–187.

Fitch WT (2013) Rhythmic cognition in humans and animals: distinguishing meter and pulse perception. Front Syst Neurosci 7:68.

Freeman JA, Nicholson C (1975) Experimental optimization of current source-density technique for anuran cerebellum. J Neurophysiol 38:369–382.

Ghitza O (2011) Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. Front Psychol 2:130.

Ghitza O (2012) On the role of theta-driven syllabic parsing in decoding speech: intelligibility of speech with a manipulated modulation spectrum. Front Psychol 3:238.

Ghitza O (2013) The theta-syllable: a unit of speech information defined by cortical function. Front Psychol 4:138.

Ghitza O (2014) Behavioral evidence for the role of cortical θ oscillations in determining auditory channel capacity for speech. Front Psychol 5:652.

Ghitza O, Greenberg S (2009) On the possible role of brain rhythms in speech perception: intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. Phonetica 66:113–126.

Giraud A-L, Poeppel D (2012) Cortical oscillations and speech processing: emerging computational principles and operations. Nat Neurosci 15:511–517.

Griffiths TD, Warren JD (2004) What is an auditory object? Nat Rev Neurosci 5:887–892.

Gutschalk A, Oxenham AJ, Micheyl C, Wilson EC, Melcher JR (2007) Human cortical activity during streaming without spectral cues suggests a general neural substrate for auditory stream segregation. J Neurosci 27:13074–13081.

Herrmann CS, Rach S, Neuling T, Strüber D (2013) Transcranial alternating current stimulation: a review of the underlying mechanisms and modulation of cognitive processes. Front Hum Neurosci 7:279.

Honing H, Merchant H, Háden GP, Prado L, Bartolo R (2012) Rhesus monkeys (Macaca mulatta) detect rhythmic groups in music, but not the beat. PloS One 7:e51369.

Iversen JR, Patel AD, Ohgushi K (2008) Perception of rhythmic grouping depends on auditory experience. J Acoust Soc Am 124:2263–2271.

Jacobs J, Kahana MJ, Ekstrom AD, Fried I (2007) Brain oscillations control timing of single-neuron activity in humans. J Neurosci 27:3839–3844.

Jensen O, Bonnefond M, VanRullen R (2012) An oscillatory mechanism for prioritizing salient unattended stimuli. Trends Cogn Sci 16:200–206.

Kösem A, Gramfort A, van Wassenhove V (2014) Encoding of event timing in the phase of neural oscillations. NeuroImage 92:274–284.

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, Musacchia G, O'Connel MN, Falchier AY, Javitt DC, Schroeder CE (2013) The spectrotemporal filter mechanism of auditory selective attention. Neuron 77:750–761.

Lakatos P, Pincze Z, Fu K-MG, Javitt DC, Karmos G, Schroeder CE (2005) Timing of pure tone and noise-evoked responses in macaque auditory cortex. Neuroreport 16:933–937.

Legatt AD, Arezzo J, Vaughan HG (1980) Averaged multiple unit activity as an estimate of phasic changes in local neuronal activity: effects of volume-conducted potentials. J Neurosci Methods 2:203–217.

Merchant H, Grahn J, Trainor L, Rohrmeier M, Fitch WT (2015) Finding the beat: a neural perspective across humans and non-human primates. Philos Trans R Soc Lond B Biol Sci 370:20140093.

Micheyl C, Tian B, Carlyon RP, Rauschecker JP (2005) Perceptual organization of tone sequences in the auditory cortex of awake macaques. Neuron 48:139–148.

Mitzdorf U (1985) Current source-density method and application in cat cerebral cortex: investigation of evoked potentials and EEG phenomena. Physiol Rev 65:37–100.

Moore BCJ, Gockel HE (2012) Properties of auditory stream formation. Philos Trans R Soc Lond B Biol Sci 367:919–931.

Nelken I, Bizley J, Shamma SA, Wang X (2014) Auditory cortical processing in real-world listening: the auditory system going real. J Neurosci 34:15135–15138.

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'Connell MN, Falchier A, McGinnis T, Schroeder CE, Lakatos P (2011) Dual mechanism of neuronal ensemble inhibition in primary auditory cortex. Neuron 69:805–817.

O'Sullivan JA, Shamma SA, Lalor EC (2015) Evidence for Neural Computations of Temporal Coherence in an Auditory Scene and Their Enhancement during Active Listening. J Neurosci 35:7256–7263.

Peelle JE, Davis MH (2012) Neural Oscillations Carry Speech Rhythm through to Comprehension. Front Psychol 3:320.

Povel DJ, Okkerman H (1981) Accents in equitone sequences. Percept Psychophys 30:565–572.

Pressnitzer D, Sayles M, Micheyl C, Winter IM (2008) Perceptual organization of sound begins in the auditory periphery. Curr Biol 18:1124–1128.

Rose MM, Moore BCJ (2005) The relationship between stream segregation and frequency discrimination in normally hearing and hearing-impaired subjects. Hear Res 204:16–28.

Schroeder CE, Lakatos P (2009) Low-frequency neuronal oscillations as instruments of sensory selection. Trends Neurosci 32:9–18.

Schroeder CE, Mehta AD, Givre SJ (1998) A spatiotemporal profile of visual system activation revealed by current source density analysis in the awake macaque. Cereb Cortex 8:575–592.

Shamma SA, Elhilali M, Micheyl C (2011) Temporal coherence and attention in auditory scene analysis. Trends Neurosci 34:114–123.

Simon JZ (2015) The encoding of auditory objects in auditory cortex: insights from magnetoencephalography. Int J Psychophysiol 95:184–190.

Snyder JS, Alain C, Picton TW (2006) Effects of attention on neuroelectric correlates of auditory stream segregation. J Cogn Neurosci 18:1–13.

Steinschneider M, Reser D, Schroeder CE, Arezzo JC (1995) Tonotopic organization of responses reflecting stop consonant place of articulation in primary auditory cortex (A1) of the monkey. Brain Res 674:147–152.

Teki S, Chait M, Kumar S, Kriegstein K von, Griffiths TD (2011) Brain bases for auditory stimulus-driven figure-ground segregation. J Neurosci 31:164–171.

Van Noorden LPAS (1975) Temporal coherence in the perception of tone sequences. PhD thesis, Eindhoven University of Technology.

van Wassenhove V (2009) Minding time in an amodal representational space. Philos Trans R Soc Lond B Biol Sci 364:1815–1830.

Whittingstall K, Logothetis NK (2009) Frequency-band coupling in surface EEG reflects spiking activity in monkey visual cortex. Neuron 64:281–289.

Wutz A, Melcher D (2014) The temporal window of individuation limits visual capacity. Front Psychol 5:952.

Wutz A, Weisz N, Braun C, Melcher D (2014) Temporal windows in visual processing: “prestimulus brain state” and “poststimulus phase reset” segregate visual transients on different temporal scales. J Neurosci 34:1554–1565.

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.

Zion Golumbic EM, Poeppel D, Schroeder CE (2012) Temporal context in speech processing and attentional stream selection: a behavioral and neural perspective. Brain Lang 122:151–161.

Zoefel B, Heil P (2013) Detection of Near-Threshold Sounds is Independent of EEG Phase in Common Frequency Bands. Front Psychol 4:262.

CHAPTER 4: SELECTIVE PERCEPTUAL PHASE ENTRAINMENT TO SPEECH RHYTHM IN THE ABSENCE OF SPECTRAL ENERGY FLUCTUATIONS

In the previous chapter, we have seen that perceptual cycles can "decide" on the "fate" of auditory input: Whether they are grouped or not depends on the underlying oscillatory phase. However, this chapter did not answer the question yet whether these cycles are able to adapt (*entrain*) to rhythmic input. As explained in previous chapters, the entrainment of the phase of neural oscillations might be of particular importance for the auditory system: The alignment of "snapshots" with relevant events can avoid harmful effects of temporal subsampling.

Arguably the most salient rhythmic stimulus in the auditory environment is speech sound. Indeed, phase entrainment of neural oscillations to speech sound has been demonstrated repeatedly (e.g., Luo and Poeppel, 2007; Giraud and Poeppel, 2012; Peelle et al., 2013). However, speech sound consists of large fluctuations in amplitude and spectral content: Neural oscillations might merely "follow" these fluctuations, as moments of high amplitude in speech sound evoke an strong neural response, moments of low amplitude in speech sound evoke a weak neural response, and moments of high and low amplitude alternate regularly. Thus, by presenting everyday speech sound, we cannot answer the questions that appears so often in this thesis: Can neural oscillations *actively* align their phase with relevant moments? Can they select the input their "snapshot" is centered on? The answer to these questions has critical implications for the feasibility of perceptual cycles in the auditory system and the following four chapters are dedicated to their answers. In the following

article, the construction of speech/noise stimuli is presented. They were constructed in a way that rules out the mere "following" of fluctuations in amplitude or spectral content of speech – but does not disrupt the possibility to adjust to the high-level information of speech sound that usually co-varies with these fluctuations. To do so, original speech snippets were mixed with complementary noises, designed to counterbalance changes in spectral content: When the spectral content of the original speech was already rich, that of the added noise was poor and, consequently, the perception of speech predominated that of noise – and vice versa. In contrast to an adaption to fluctuations of sound amplitude or spectral content, which could happen on a sensory level, the ability of the brain to "track" changes in phonetic (high-level) information in our stimuli would indicate a process on a rather "high" brain level, as speech and noise (with, on average, the same spectral content) can only be distinguished after several stages of auditory processing. Perceptual entrainment to the constructed speech/noise stimuli was assessed by presenting clicks at auditory threshold at random moments during the stimuli, resulting in two predictions: (1) If there were entrainment to the speech sound, click detection should vary as a function of the remaining high-level information. (2) Based on previous results in monkeys (O'Connell et al., 2011), we expected the modulation of click detection by high-level information to change phase with the (sound) frequency of the click, when the latter changes from irrelevant (located beyond the spectral content of speech) to relevant (located within the principal spectral content of speech). Indeed, we show that the auditory threshold can be entrained by the rhythm of speech even when high-level information is not accompanied by fluctuations in input sound amplitude or spectral content. This entrainment is frequency-specific in that fluctuations in high-level information differently affected click detection depending on whether the frequency of the click was part of the principal frequency content

of the speech sound or not. Interestingly, perceptual phase entrainment was abolished when the speech/noise sound was presented in a time-reversed manner, indicating that linguistic information underlie the observed effects.

To conclude, our findings demonstrate that perceptual phase entrainment to rhythmically occurring sounds, potentially reflecting flexibly adjusted perceptual cycles in audition, does entail a high-level process.

Article:

Zoefel B, VanRullen, R (2015) Selective perceptual phase entrainment to speech rhythm in the absence of spectral energy fluctuations. J Neurosci 35:1954–1964.

Behavioral/Cognitive

Selective Perceptual Phase Entrainment to Speech Rhythm in the Absence of Spectral Energy Fluctuations

Benedikt Zoefel and Rufin VanRullen

Université Paul Sabatier, 31062 Toulouse Cedex 9, France, and Centre de Recherche Cerveau et Cognition, Centre National de la Recherche Scientifique, 31052 Toulouse Cedex, France

Perceptual phase entrainment improves speech intelligibility by phase-locking the brain's high-excitability and low-excitability phases to relevant or irrelevant events in the speech input. However, it remains unclear whether phase entrainment to speech can be explained by a passive "following" of rhythmic changes in sound amplitude and spectral content or whether entrainment entails an active tracking of higher-level cues: in everyday speech, rhythmic fluctuations in low-level and high-level features always covary. Here, we resolve this issue by constructing novel speech/noise stimuli with intelligible speech but without systematic changes in sound amplitude and spectral content. The probability of detecting a tone pip, presented to human listeners at random moments during our speech/noise stimuli, was significantly modulated by the rhythmic changes in high-level information. Thus, perception can entrain to the speech rhythm even without concurrent fluctuations in sound amplitude or spectral content. Strikingly, the actual entrainment phase depended on the tone-pip frequency, with tone pips within and beyond the principal frequency range of the speech sound modulated in opposite fashion. This result suggests that only those neural populations processing the actually presented frequencies are set to their high-excitability phase, whereas other populations are entrained to the opposite, low-excitability phase. Furthermore, we show that the perceptual entrainment is strongly reduced when speech intelligibility is abolished by presenting speech/noise stimuli in reverse, indicating that linguistic information plays an important role for the observed perceptual entrainment.

Key words: **entrainment; envelope; high-level; oscillation; phase; speech**

Introduction

Speech is intrinsically rhythmic. The brain makes use of this rhythmicity (Schroeder and Lakatos, 2009) by entraining its neural oscillations so their high-excitability phase matches informative features, increasing speech intelligibility (Ahissar et al., 2001; Luo and Poeppel, 2007; Kerlin et al., 2010; Ding and Simon, 2013), while the phase of low excitability is aligned with irrelevant information. It has been shown repeatedly that the neural phase is correlated with auditory perception (Henry and Obleser, 2012; Ng et al., 2012; but see Zoefel and Heil, 2013). Neural phase entrainment (between ~2 and 8 Hz) is an integral part of many current theories of speech perception (Poeppel, 2003; Ghitza, 2011, 2012, 2013; Giraud and Poeppel, 2012). However, in normal speech sounds (Fig. 1*A*, top), the ongoing rhythmic modulations simultaneously affect all aspects of the signal, from low-level acoustic features (sound amplitude, spectral content) to higher-level phonetic ones (Fig. 1*A*, bottom). Neural phase entrainment to speech, therefore, can take place at all stages of auditory processing, even the earliest ones, such as the cochlea, in which neurons respond selectively to certain sound frequencies. This raises the possibility that much of the previously reported speech entrainment phenomena could merely derive from the rhythmic activation of the very input to the auditory system. In other words, speech entrainment could well reflect a passive, low-level process having little to do with phonetic information processing, or speech per se. Indeed, it is well known that phase entrainment can happen with rather low-level stimuli, such as pure tones [the basic mechanism for the auditory steady-state response (ASSR); Galambos et al., 1981], and that intelligibility is not required for entrainment to speech sound (Howard and Poeppel, 2010; Stefanics et al., 2010; Luo and Poeppel, 2012; Peelle et al., 2013). Logically, rhythmic activation of the cochlea must thus be sufficient to induce at least some forms of phase entrainment to speech, a conclusion that we do not question in this study. The specific question we ask is, rather, whether such low-level rhythmic activation is a necessary condition for phase entrainment to speech.

One way to demonstrate the existence of genuine high-level phase entrainment to phonetic information would be to create artificial speech stimuli in which only phonetic information, but no other aspect of the signal, fluctuates rhythmically. This strategy appears problematic, however, particularly as phonetic information is a rather ill-defined notion. Here, we took the complementary approach of designing speech stimuli in which the lowest-level information (sound amplitude and spec-

Received Aug. 19, 2014; revised Oct. 28, 2014; accepted Nov. 23, 2014.

Author contributions: B.Z. and R.V. designed research; B.Z. performed research; B.Z. and R.V. contributed unpublished reagents/analytic tools; B.Z. and R.V. analyzed data; B.Z. and R.V. wrote the paper.

This work was supported by a Studienstiftung des deutschen Volkes (German National Academic Foundation) scholarship to B.Z. and a European Young Investigator Award to R.V. The authors thank Daniel Pressnitzer and Jesko Verhey for helpful comments and discussions.

The authors declare no competing financial interests.

Correspondence should be addressed to Benedikt Zoefel, Centre de Recherche Cerveau et Cognition (CerCo), Pavillon Baudot CHU Purpan, BP 25202, 31052 Toulouse Cedex, France. E-mail: zoefel@cerco.ups-tlse.fr.

DOI:10.1523/JNEUROSCI.3484-14.2015

 0270-6474/15/351954-11$15.00/0

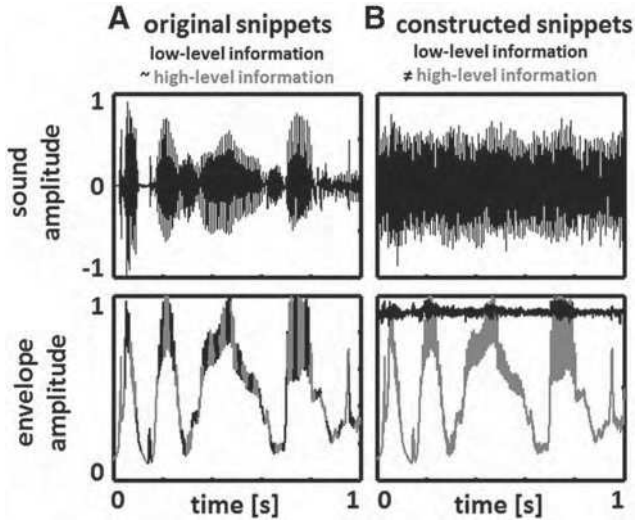


Figure 1. Overview of the experimental approach. ***A***, In normal speech sound (top), low-level features (such as slow amplitude fluctuations, here labeled as signal envelope and shown as the black signal, bottom) and fluctuations in high-level features (including, but not restricted to phonetic information; gray signal, bottom) covary. ***B***, We constructed speech/noise snippets in which spectral energy is comparable across phases of the original signal envelope (***A***, bottom, black signal). To do so, original speech and complementary noise were summed, with the spectral content of the noise matched to counterbalance spectral differences between phases of the signal envelope of the original speech (see Materials and Methods; Fig. 2). Consequently, a hypothetical phase entrainment to the constructed stimuli could not be due to a mere "following" of low-level stimulus properties as the global signal envelope of the constructed snippets (and the underlying spectral content) did not show systematic fluctuations anymore (***B***, bottom, black signal; Fig. 3). However, high-level features of the constructed snippets (***B***, bottom, gray signal) still fluctuated with the rhythm of the original signal envelope. We were thus able to disentangle phase entrainment to low-level (fluctuations in amplitude and spectral content) and high-level (e.g., fluctuations in phonetic information) stimulus properties. Note that we equalize original signal envelope with high-level features (***A***, bottom) for the sake of convenience and illustrational purposes (as no quantitative measure is available in the literature). For our paradigm, it is enough to assume that the two quantities are positively related.

tral content, corresponding to information processed in the cochlea) does not fluctuate along with the rhythm of speech (see Materials and Methods and Discussion). In short, we constructed speech/noise hybrid stimuli (Fig. 1*B*, top) with spectral content comparable across phases of the original signal envelope (see Fig. 3). Higher-level features still fluctuated rhythmically (Fig. 1*B*, bottom) and speech remained intelligible, providing potential means for (high-level) phase entrainment. Entrainment was assessed by presenting tone pips at auditory threshold at random moments during our stimuli. Can phase entrainment still happen without rhythmic fluctuations in the lowest-level acoustic signals? Answering this question has critical implications for current theories of speech perception.

Materials and Methods

Stimulus overview. As the heart of this study, we developed novel speech/noise stimuli without systematic fluctuations in amplitude and spectral content. Here we give a short overview, while the detailed procedure of stimulus construction is provided below. Original speech snippets were recorded (sampling rate, 44,100 Hz) of a male native-English speaker reading parts of a novel. Experimental stimuli were constructed by merging original speech with noise. Critically, spectral content and energy of the stimuli constructed in this way were designed to be statistically comparable at all phases of the original signal envelope. In normal speech sound, both spectral content and energy are not constant but fluctuate over time. Thus, when merging original speech snippets with noise, spectral content and energy of the noise had to be specifically adapted to the instantaneous characteristics (i.e., envelope phase; Fig. 1*A*, bottom) of the original speech. To do so, for each of the original speech snippets, a complementary noise was constructed: when spectral energy (the specific distribution of power across sound frequencies) of the original speech was high, that of the noise was low and vice versa. As spectral energy in the original speech fluctuated rhythmically (~2–8 Hz), the perception of speech and noise in the constructed snippets also alternated rhythmically by construction. Importantly, the perception of speech was driven by high-level features and not by changes in sound amplitude or spectral content, as the latter was kept statistically comparable by stimulus construction (Fig. 1*B*, bottom, black line).

Use of terms and definitions. To avoid ambiguity of some critical terms, we will shortly describe the usage of the most important expressions in the context of this study.

We defined "signal envelope" as the weighted sum of the instantaneous energy of the signal across frequencies. "Weighted" refers to the fact that frequency sensitivity is not constant in the auditory system, and energy at some frequencies is amplified more strongly by the system than energy at other frequencies (see Stimulus construction). Our choice of envelope definition was motivated by an observation from Ghitza (2001) stating that degraded envelope cues can be recovered by cochlear filtering. For instance, frequency-modulated stimuli with flat broadband (full-signal) envelopes can exhibit amplitude fluctuations at the cochlear output level (Ghitza et al., 2012). Importantly, here we aimed to avoid this confound by (1) extracting the signal envelope by wavelet transformation (which can be seen as a filtering of the speech signal into many narrow frequency bands); (2) spacing those "frequency bands" on a logarithmic scale (Ding and Simon, 2013), thus mimicking the spectral organization of the cochlea; and (3) weighing each of those frequency bands by the cochlear sensitivity at this particular frequency (see Stimulus construction). In fact, running original speech snippets through a bank of gammatone filters (a well known alternative method for simulating cochlear processing; Immerseel and Peeters, 2003) before calculating their envelope (as the sum of energy across frequencies) resulted in envelope phases that were strikingly similar to those obtained by our method (see Fig. 7*C*).

Low-level and high-level features of speech are not well defined in the literature. In this study, speech/noise stimuli were constructed so that they could not passively entrain the lowest levels of auditory processing (such as the cochlear level, where speech and noise would excite the same cells to the same extent). Based on this, we term as "low-level" those speech features that are equated in our stimuli: sound amplitude and spectral content. Logically, then, we term the remaining features as "high level," as they enable the listener to distinguish speech and noise on a processing level beyond the cochlea; and we emphasize that our high-level features include (but might not be restricted to) phonetic information.

When using the term "frequency," we always refer to sound frequency (i.e., to speech sound or click frequency; the latter explicitly stated as such) and not, for instance, to the frequency of a brain oscillation. When using the term "phase," we always refer to the (Hilbert) phase of the signal envelope (and not, for instance, to the spectral phase of the speech sound).

Participants. Eleven participants volunteered in Experiments 1 and 2 (see below). Two participants were excluded from further analyses, one because of poor speech comprehension in the speech-attended condition task (see Experimental paradigm; d' of -0.35), which was far below the average performance of the 11 subjects (0.71 ± 0.50), and the other because of missed button presses in 89% of the trials during the same task. Both reasons might have prevented proper phase entrainment. Nine subjects remained in the analyses of Experiments 1 and 2 (seven female; mean age, 27.8 years). Ten participants volunteered in Experiment 3 (four of whom had already participated in the other two experiments). One participant was excluded from further analyses due to an inability to differentiate forward and backward speech/noise sound, indicated by a poor performance in the associated perceptual task (see below; 51.7% correct) that was significantly worse ($p < 0.001$; Student's t test) than the mean performance of the other participants (on average, 81.5% correct). Nine subjects remained in the analyses of Experiment 3 (four female; mean age, 28.1 years). Participants of all experiments were fluent in English, reported normal hearing, and gave written informed consent.

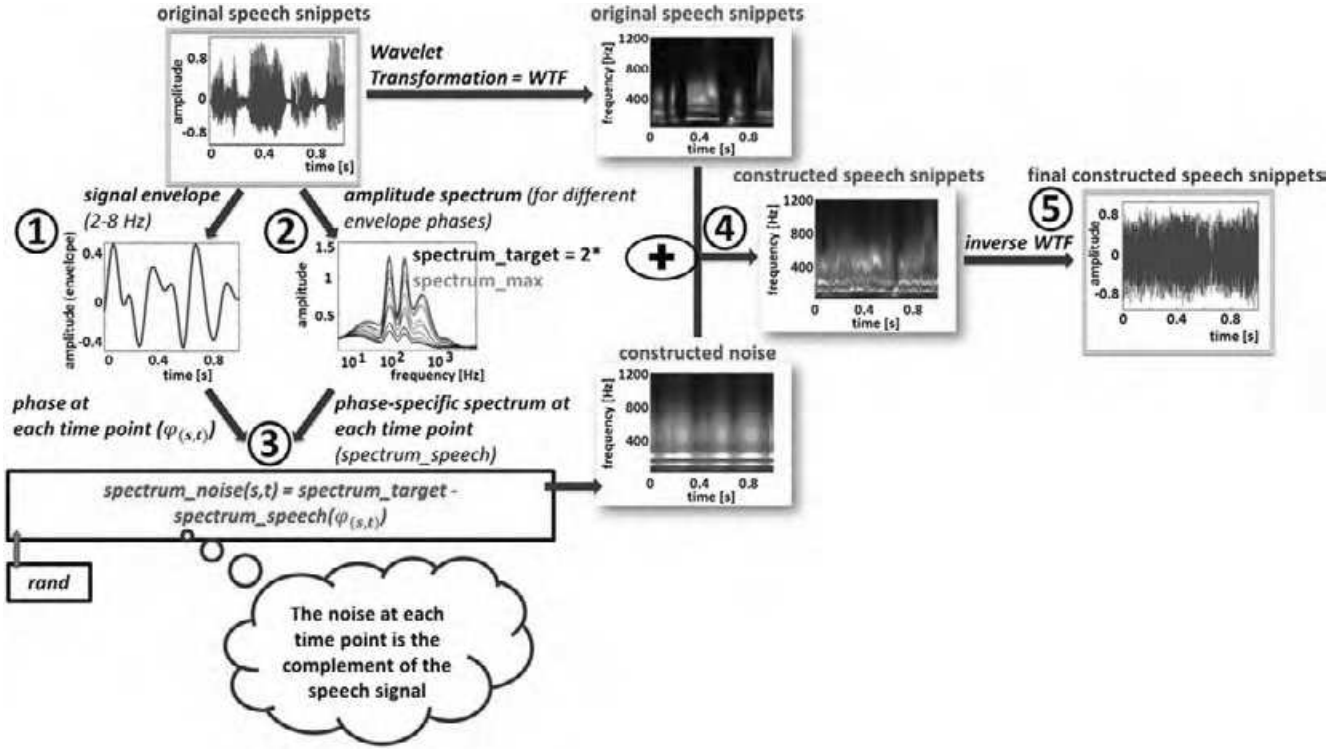


Figure 2. Construction of the speech/noise stimuli that were used in this study. Original speech snippets and complementary noise (based on the instantaneous envelope phase of the individual snippet and the average original spectral content at the respective phase; see Materials and Methods for details) were added, yielding speech snippets with conserved high-level features (fluctuating in the frequency range of ~2–8 Hz) but comparable spectral content across envelope phases. Parts processed in the real domain are specified in blue and as a signal in time. Parts processed in the wavelet domain are specified in red and are represented as a time–frequency signal.

The experimental protocol was approved by the relevant ethical committee at Centre National de la Recherche Scientifique.

Stimulus construction. The detailed procedure of stimulus construction is shown in Figure 2 and is described step by step as follows (the numbers correspond to those in Fig. 2): (1) The signal envelope *env* was extracted for each individual original snippet *s* as the sum of the instantaneous energy *e* (or amplitude; extracted by wavelet transformation for 304 logarithmically spaced frequencies in the range between 0.59 and 21,345 Hz) at each time point *t* of the signal across frequencies *F*, weighted by the cochlear sensitivity *w* (ISO 226 Equal-Loudness-Contour Signal for Matlab, J. Tackett) to correct for differences in frequency sensitivity in the auditory system according to the following equation:

$$env(s,\ t) = \frac{1}{F}\sum_{f=0}^{F} w(f) * e(s,\ f,\ t)$$

The envelope was then bandpass filtered between 2 and 8 Hz (second-order Butterworth filter). (2) A "target spectrum" ($spectrum_{target}$) was computed from all original snippets. Here, the average amplitude spectrum [$spectrum_{speech}(\varphi, f)$] was extracted separately for different phases φ (divided into 12 nonoverlapping phase bins of width $\pi/6$) of the original signal envelope [i.e., for each envelope phase bin, spectra ($spectrum_{inst}$) were averaged across all instances *I* of the respective phase bin and across all snippets *S*] as follows:

$$spectrum_{speech}(\varphi) = \frac{1}{S}\sum_{s=1}^{S}\frac{1}{I}\sum_{i=1}^{I} spectrum_{inst}(\mathrm{i},\ \mathrm{s},\ \varphi)$$

For each frequency *f*, $spectrum_{target}$ was twice the maximal amplitude across phase bins (this was necessary to leave enough "room" for additional noise at all time points of the original speech snippets) as expressed in the following equation: $spectrum_{target}\ (f) = 2 * \max_{-\pi \le \varphi \le \pi} spectrum_{speech}\ (\varphi, f)$. (3) To build the experimental stimuli, for each time point and each original snippet, the amplitude spectrum of the noise (individually generated for *s* and *t*) was constructed as the difference between $spectrum_{target}$ and $spectrum_{speech}$ (i.e., the average amplitude spectrum across original snippets) for phase φ of the individual signal envelope of snippet *s* at time point *t*. Note that this difference depends on φ. This is thus a critical step to render spectral content comparable across all instances of φ: $spectrum_{noise}\ (s,\ t) = spectrum_{target} - spectrum_{speech}\ (\varphi_{(s,\ t)})$, where $\varphi_{(s,\ t)} = angle(hilbert[env(s, t)])$. Note also that, for a given phase of the envelope, this difference is the same for all snippets, whereas the individual spectra at different instances of this phase might slightly differ from each other. This results in small random signal fluctuations that remain with each snippet and differ between individually constructed snippets (these random fluctuations can be seen, for instance, in panel 4 of Fig. 2, showing 1 s of a single constructed speech/noise snippet in the time–frequency domain). However, and importantly, these "residues" do not systematically depend on the phase of the original signal envelope, as they disappear upon averaging the spectral content for different envelope phases across snippets (Fig. 3). Spectrally matched noises were constructed by multiplying, for each time point, white noise (Fig. 2, rand) by $spectrum_{noise}\ (s,\ t)$ in the wavelet domain. (4) Original snippets and constructed noises were added in the wavelet domain (note that the original speech remains intact, it is only "hidden" by the noise in various degrees, depending on the instantaneous phase of the original signal envelope). (5) Finally, this sum of original speech and constructed noise was transferred back into the real domain by inverse wavelet transform. Due to the overcomplete nature of our wavelet time–frequency decomposition (i.e., the fact that wavelets at neighboring frequencies significantly overlap in the frequency domain), this step produced residual errors: when expressed back into the wavelet domain, the resulting stimuli differed slightly from the intended time–frequency signal. However, this difference could be minimized by applying the desired wavelet amplitude to each wavelet coefficient and transforming again into the real domain. We iterated this correction procedure 100 times, yielding the final constructed snippets.

The result of our stimulus construction procedure is shown in Figure 3: here, for both original and constructed speech snippets, the average amplitude spectrum is shown as a function of the envelope phase of the original speech snippets. As is clearly visible in Figure 3, different phases

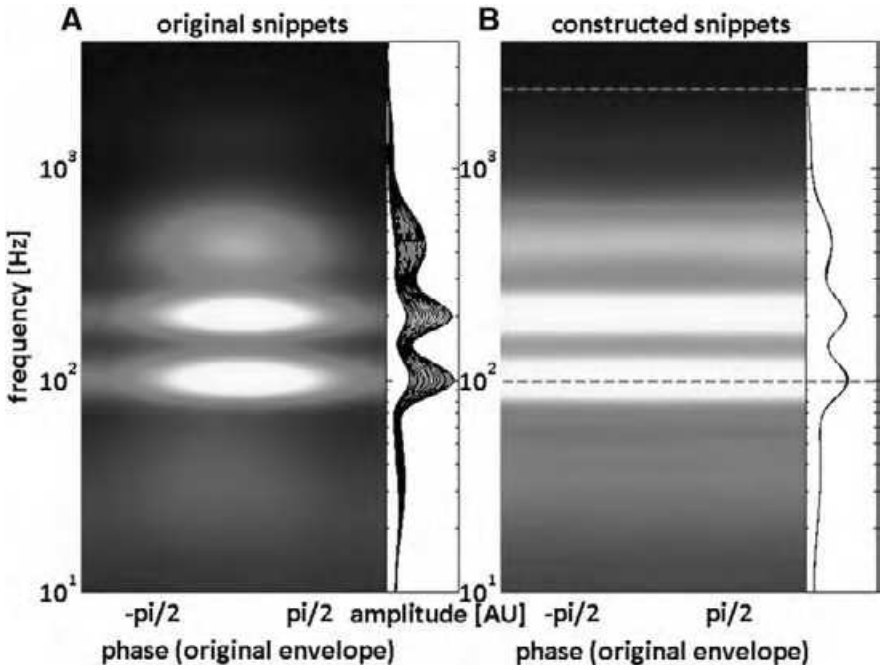


Figure 3. ***A***, ***B***, Amplitude spectra for original (***A***) and constructed snippets (***B***) as a function of different (binned) phases of the original signal envelope. Amplitudes are shown either as the color of a phase–frequency plot or as a family of curves where each trace represents the amplitude spectrum for one-phase bin. Note that amplitude spectra differ markedly across phase bins for the original snippets (***A***) whereas amplitude spectra are similar across phase bins for the constructed snippets (***B***). The spectral locations of the tone pips used in Experiments 1 (red) and 2 (gray) with respect to the entraining speech/noise stimuli are shown as dashed lines.

of the envelope differ markedly in both their spectral content and energy, as far as the original snippets are concerned. Yet, by processing the stimuli as described above, we were able to remove these fluctuations: both spectral content and energy of the constructed snippets are now, on average, comparable across phases of the original signal envelope. Thus, although systematic spectral energy fluctuations are removed by our stimulus processing (the circular correlation between envelope phases of original and constructed snippets is $r = -0.042$), speech sound is still intelligible and high-level features (e.g., phonetic information) still fluctuate rhythmically at ~2–8 Hz, providing potential means for oscillatory phase entrainment.

Samples of several stages of stimulus processing (original speech snippet, constructed noise, final constructed speech snippet, reversed constructed speech snippet) are presented in Movie 1.

Experimental paradigm. Phase entrainment to our constructed stimuli was assessed in three psychophysical experiments (tested on 3 separate days). For all experiments, one trial consisted of the presentation of a 6 s stimulus that was randomly chosen from all concatenated constructed snippets (total length, ~10 min). Signals between concatenated snippets were interpolated to avoid artificial clicks that could potentially have influenced the subjects' performance. In 90% of the trials, between one and three (equal probability) tone pips were presented at threshold level at random moments during our speech/noise snippets. The remaining 10% of the trials were catch trials, where no tone pips were presented. It has been shown before that phase entrainment is frequency-specific: neural oscillations were entrained to their high-excitability phase only in those parts of monkey primary auditory cortex actually processing the sound frequency of the entraining stimulus, whereas in remaining areas oscillations were entrained to their low-excitability phase (O'Connell et al., 2011; Lakatos et al., 2013). Thus, we "probed" entrainment at two different sound frequencies. In Experiments 1 and 3, tone pips had a duration of 2.9 ms and a carrier frequency of 2.4 kHz, which was beyond the principal frequency range of our speech stimuli. In Experiment 2, tone pips had a duration of 30 ms and a frequency of 100 Hz, which was within the principal frequency range of our speech stimuli. The spectral location of the tone pips with respect to the spectral content of the entraining speech/noise stimuli is shown as dashed lines in Figure 3. The minimum interval between tone pips was 1 s. Subjects were asked to press a button whenever they detected such a tone pip. A tone pip was considered as detected, and the response classified as a hit, if the subjects pressed the response button within 1 s following tone-pip onset, otherwise as a miss. The amplitude of the tone pip was adapted constantly (based on the performance of the preceding 100 trials) so that tone-pip detection had a mean probability of 50%.

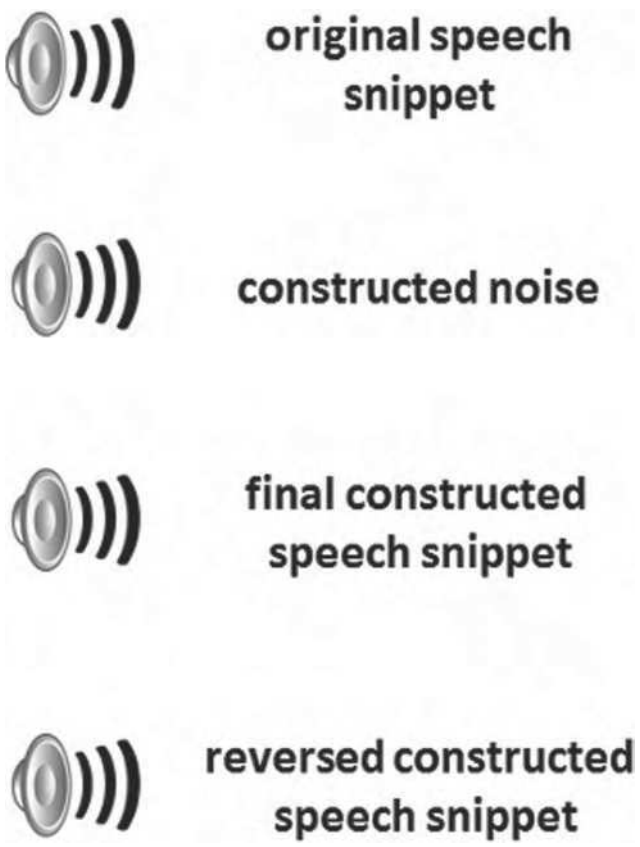
In Experiment 1, subjects were tested in two experimental conditions. In one ("speech-unattended") condition, subjects were asked to attend only the tone pips and to ignore the speech sound. In the other ("speech-attended") condition, subjects were asked to attend both tone pips and speech sound. In the latter condition, attention to speech was tested by an additional task in which subjects were presented with a word after each trial. Subjects had to indicate by a button press whether that word was or was not present in the previous trial (true on average in 63.7 ± 3.8% of the trials; mean and SD across subjects). The button press had to occur within 1 s after word presentation, increasing task difficulty. If subjects did not press any button during a given trial, their response was counted as a "word absent" press. Subjects completed 300 trials (on average 552 tone pips in the speech-attended condition and 537 tone pips in the speech-unattended condition) for both conditions.

In Experiment 2, subjects were asked to press a button for each detected tone pip, but to pay attention to both speech sound and tone pips. As we did not find differences between the two experimental conditions in the first experiment (data not shown), we did not include an attentional manipulation in this experiment. Subjects completed 600 trials (on average 1078 tone pips).

There is a current debate of whether intelligibility is important for phase entrainment to speech sound (Peelle and Davis, 2012; Ding and Simon, 2014), and there are studies supporting the view that speech comprehension is relevant (Gross et al., 2013; Peelle et al., 2013) and others indicating that it is not (Howard and Poeppel, 2010; Luo and Poeppel, 2012). In Experiment 3, we tackled this question by presenting our speech/noise snippets either forwards ("forward condition") or backwards ("backward condition"), randomly determined from trial to trial. Only 2400 Hz tone pips were presented and subjects were asked to press a button for each detected tone pip. Additionally, subjects were asked to indicate by a button press after each trial whether the speech sound was played forwards or backwards. This was done to ensure that subjects indeed perceived the modulation in intelligibility. Subjects completed 300 trials for each condition (on average 531 and 546 tone pips for forward and backward condition, respectively).

Stimuli were presented diotically via loudspeakers (Logitech Z130). Subjects were seated 55 cm away from the loudspeakers, with their head position fixed by a chin rest. Sound level was adjusted to 75 dB SPL at the position of the chin rest for all subjects. The Psychophysics Toolbox for Matlab (Brainard, 1997) was used for stimulus presentation.

Data analyses. To evaluate (perceptual) phase entrainment to our stimuli, the probability of detecting a tone pip was analyzed as a function of the phase of the original signal envelope of the snippets. Note that the signal envelope of the original snippet (computed using low-level features, i.e., instantaneous spectral energy) covaries with high-level features of both the original and the constructed speech sounds (e.g., phonetic information is highest at the peak of the envelope and lowest at its trough). The phase of the original signal envelope at each moment before and after the tone pip was divided into 40 overlapping π-wide



Movie 1. Speech/noise stimuli were constructed in this study by summing original speech snippets and their individually constructed, complementary noise, with the noise spectrally matched to counterbalance differences in spectral content across phases of the original signal envelope. Samples of several stages of stimulus processing (original speech snippet, constructed noise, final constructed speech snippet, reversed constructed speech snippet) are presented in this movie.

bins. We used overlapping bins to smooth the outcome of our analyses without affecting the actual results (Fiebelkorn et al., 2013). Results were similar with nonoverlapping bins and different bin widths, including those used for stimulus construction (i.e., 12 non-overlapping bins; data not shown).

If there were phase entrainment to our speech/noise stimuli, detection of tone pips would depend on the phase of the original signal envelope. Also, this dependence would not necessarily be restricted to the time of the tone pip. Therefore, we fitted a sine wave to the average performance (tone-pip detection as a function of phase) across subjects, separately for each time point (downsampled to 100 Hz) in a time window of 1.6 s centered on tone-pip onset. The amplitude of this fitted sine wave reflects the magnitude of performance modulation by the phase of the original signal envelope (or equivalently, by rhythmic fluctuations in high-level information in the constructed speech/noise stimuli). For Experiments 1 and 2, as we found phase entrainment for both experiments (Fig. 4*A*), the two amplitudes obtained for both experiments were averaged. To test whether this modulation was significant for any time point around tone-pip onset, we used a permutation test to construct surrogate distributions for each time point. Here, hits and misses were shuffled for each subject separately before recalculating the amplitude values described above. This procedure was repeated 1,000,000 times. Thus, it was possible to obtain a range of amplitude values that can be observed under the null hypothesis that the original signal envelope does not affect detection of tone pips. p values were obtained for the recorded data by comparing "real" amplitude values with the surrogate distribution for each respective time point. p values were corrected for multiple comparisons using the false discovery rate (FDR) procedure. Here, a significance threshold is computed that sets the expected rate of falsely rejected null hypotheses at 5% (Benjamini and Hochberg, 1995).

A similar permutation procedure was used to compare effects for the two (speech-attended and speech-unattended) conditions in Experiment 1. Here, instead of shuffling hits and misses, the assignment of trials to both conditions was shuffled separately for each subject before recalculating the amplitude values described above. This procedure was repeated 100,000 times to obtain surrogate distributions for each time point under the null hypothesis that sine fit amplitudes do not differ across conditions.

All analyses were performed in Matlab. The Circular Statistics Toolbox for Matlab (Berens, 2009) was used for circular statistics.

Results

We constructed speech/noise stimuli without systematic fluctuations in sound amplitude and spectral content (Fig. 3) while keeping high-level features (including phonetic information) intact, fluctuating at ~2–8 Hz (Fig. 1*B*, bottom). Stimuli were constructed by summing original speech snippets and their individually constructed, complementary noise, with the noise spectrally matched (i.e., having a phase-specific spectrum) to counterbalance differences in spectral content across phases of the original signal envelope (example sound files for stimulus construction are presented in Movie 1; see Materials and Methods). During certain phases of the original signal envelope (e.g., at its peak; Fig. 3*A*), when spectral energy of the original speech was already high, that of the added noise was low. Consequently, the perception of speech predominated that of noise. Vice versa, when spectral energy of the original speech was low (e.g., at its trough), that of the added noise was high and the perception of noise predominated that of speech (see Materials and Methods; Fig. 2). Phase entrainment to the generated speech/noise snippets was tested psychophysically by presenting tone pips at auditory threshold at random moments during the presentation of these stimuli ($N = 9$ subjects in all experiments). A dependence of tone-pip detection on the original signal envelope (which in our constructed speech/noise stimuli reflects high-level information) would indicate perceptual phase entrainment (note that we call this entrainment "high-level," as it cannot take place on the lowest level of auditory processing, the cochlear level; see Introduction, Materials and Methods, and Discussion). As previous studies showed frequency-specific entrainment (O'Connell et al., 2011; Lakatos et al., 2013; see Materials and Methods), we "probed" entrainment by tone pips at two different frequencies (Fig. 3, dashed lines): one beyond the principal frequency range of our speech/noise stimuli (tone pip at 2400 Hz; Experiment 1) and one within the principal frequency range of our speech/noise stimuli (tone pip at 100 Hz; Experiment 2). In Experiment 1, the role of attention was tested in two experimental conditions (see Materials and Methods): subjects were given an additional task related to the content of the speech stimulus (after each trial, subjects had to decide whether a given word was present in the previous trial or not) or were told to ignore the speech sound. For the analysis of phase entrainment, as results did not differ between conditions (data not shown), trials have been pooled across conditions. In Experiment 2, subjects were told to attend to both speech and tone pips, but no additional task was given. In Experiment 3, the role of intelligibility was tested by presenting speech/noise stimuli either forwards or backwards. As in the other two experiments, subjects had to indicate tone-pip detection (only 2400 Hz) by a button press. In addition, to assure that subjects indeed perceived the modulation in intelligibility, they had to indicate after each trial whether the speech was played backwards or not.

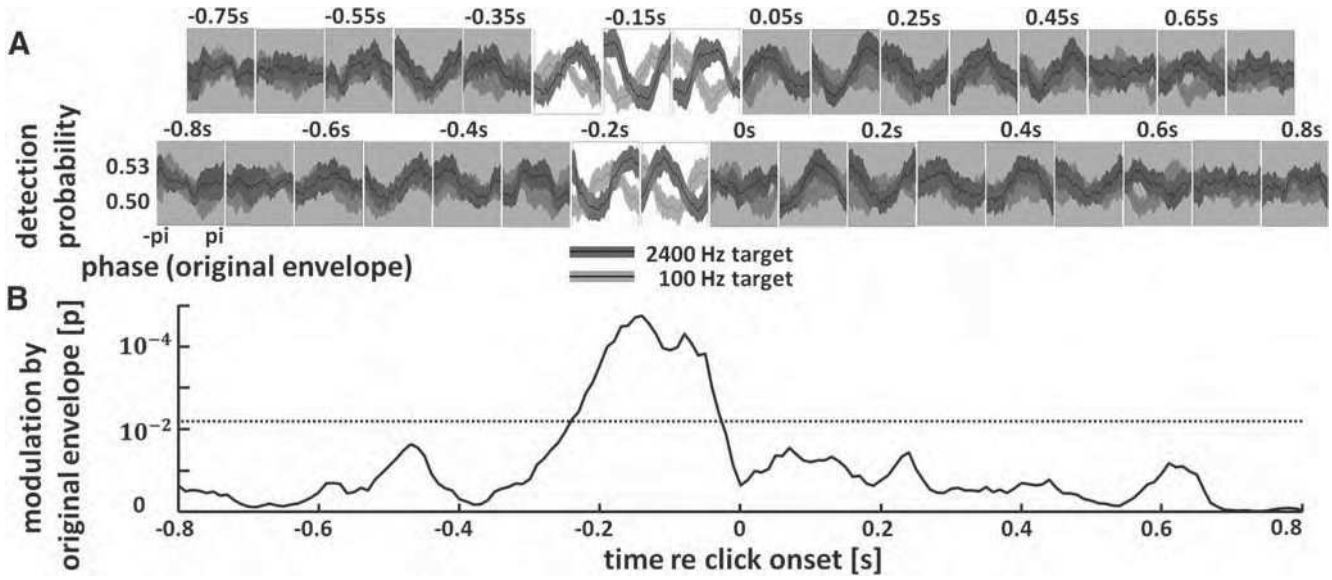


Figure 4. Perception entrains to speech rhythm in the absence of spectral energy fluctuations. ***A***, Tone-pip detection probability as a function of the original signal envelope at different time points (offset in 2 rows). Note the sinusoidal modulation of performance for ~250 ms before tone-pip onset. This modulation differs in phase between the two tone-pip frequencies (black: Experiment 1, tone-pip frequency 2400 Hz; gray: Experiment 2, tone-pip frequency 100 Hz), reflecting frequency-specific phase entrainment. Non-significant time windows of pooled data are shaded gray. SEM is shown by contours around the lines. ***B***, Statistical significance of the phase entrainment shown above, pooled across experiments. Significance threshold is FDR corrected and shown as a dotted line.

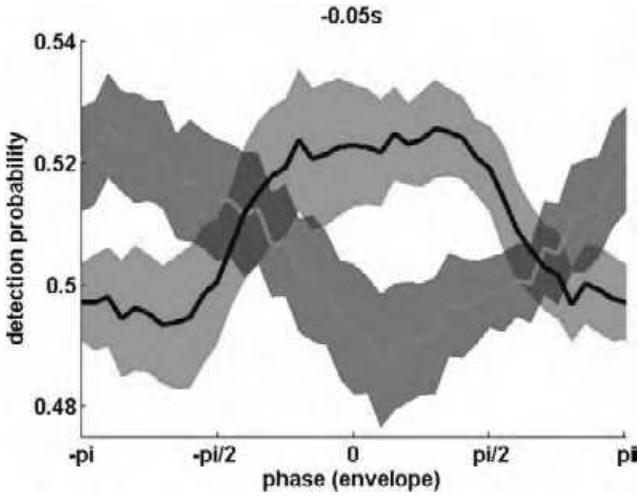
Behavioral results

As expected from our staircase procedure, subjects detected approximately half of the tone pips in all experiments (mean and SD: 51.8 ± 4.0% and 50.4 ± 0.8% in the speech-attended and speech-unattended conditions of Experiment 1, 50.7 ± 2.1% in Experiment 2, and 51.0 ± 1.6% in Experiment 3). In Experiment 1, both mean adjusted amplitude of the tone pips and median reaction time were slightly higher in the speech-attended (tone-pip amplitude: 0.154 ± 0.030 AU; reaction time: median, 0.53 s) than in the speech-unattended condition (tone-pip amplitude: 0.146 ± 0.029 AU; reaction time: median, 0.51 s), but not significantly different (tone-pip amplitude $p = 0.59$, Student's t test; reaction time $p = 0.23$, Kruskal-Wallis test). Median reaction time was 0.57 s in Experiment 2 and 0.58 s in Experiment 3. The adjusted tone-pip amplitude was higher in Experiment 2 than the mean of Experiments 1 and 3 (0.693 ± 0.351 vs 0.144 ± 0.034 AU; $p < 0.001$, Student's t test), most likely because the energy of the speech/noise signal was higher at that frequency. In all experiments, tone-pip amplitude did not differ between phase bins used for further analyses ($p > 0.97$; one-way ANOVA). False alarm probability (as the percentage of button presses in all non-overlapping 1 s windows not following a tone pip) was on average 4.3 ± 2.1% and 3.5 ± 2.1% in the speech-attended and speech-unattended condition of Experiment 1, 3.8 ± 1.6% in Experiment 2, and 5.7 ± 3.0% in Experiment 3. The additional speech comprehension task in the speech-attended condition of Experiment 1 (see Materials and Methods) was rather difficult, but above chance level: on average, subjects completed the task with a d' (a measurement of sensitivity to the task, combining both hits and false alarms; a d' of 0 reflects performance at chance level) of 0.82 ± 0.40, significantly above 0 ($p < 0.001$). The performance in the additional task of Experiment 3 (see Materials and Methods) indicated that subjects did perceive the modulation of intelligibility: on average, the speech sound was correctly classified (forwards vs backwards) in 81.5 ± 13.1% of the cases.

Perception selectively entrains to speech rhythm in the absence of spectral energy fluctuations

Mean probability of tone-pip detection as a function of the original signal envelope at different time points around tone-pip onset is shown in Figure 4*A* for Experiments 1 and 2. In both cases, a sinusoidal modulation of performance by the original signal envelope is visible before tone-pip onset. Note that the same phase bins were used in Figures 3 and 4*A*. Our modulatory effect thus cannot be influenced by low-level features (i.e., amplitude fluctuations in specific frequency bands) as they have been made comparable across phase bins (Fig. 3*B*). However, higher-level features were still present in the constructed snippets and still fluctuated rhythmically as a function of the phase of the original signal envelope (Fig. 1, bottom). Consequently, the dependence of tone-pip detection probability on the phase of the original signal envelope can only be explained by an entrainment to the preserved rhythmic fluctuations in high-level features of the presented speech/noise sounds.

Strikingly, the actual phase of modulation clearly differed between the two experiments: a peak of the original signal envelope (phase 0; when the least amount of noise was added) ~50 ms before tone-pip onset was favorable for detecting a 2400 Hz pip (Experiment 1), but, in stark contrast, was disadvantageous for detecting a 100 Hz pip (Experiment 2). A video of this frequency-specific entrainment effect, unfolding over time, is presented as Movie 2. To statistically evaluate phase entrainment to our stimuli, for each experiment a sine wave was fitted to the mean performance (detection of tone pip as a function of original signal envelope, averaged across subjects) at each time point around tone-pip onset, and its amplitude (averaged across the two experiments) was compared with surrogate distributions (see Materials and Methods). Results of this statistical test are shown in Figure 4*B*. Perceptual phase entrainment is significant between −250 and −30 ms with respect to click onset.



Movie 2. Average performance (mean tone-pip detection probability as a function of phase of the original signal envelope) in time (−800 to +800 ms with respect to tone-pip onset), separately for the two tone-pip frequencies (2400 Hz, Experiment 1: black line and dark gray contour as mean and SEM, respectively; 100 Hz, Experiment 2: dark gray line and light gray contour). The contours change color (2400 Hz: red; 100 Hz: blue) during all significant time points of the pooled data (Fig. 4*B*). Note the phase opposition of the modulatory effect shortly before tone-pip onset, reflecting frequency-specific phase entrainment.

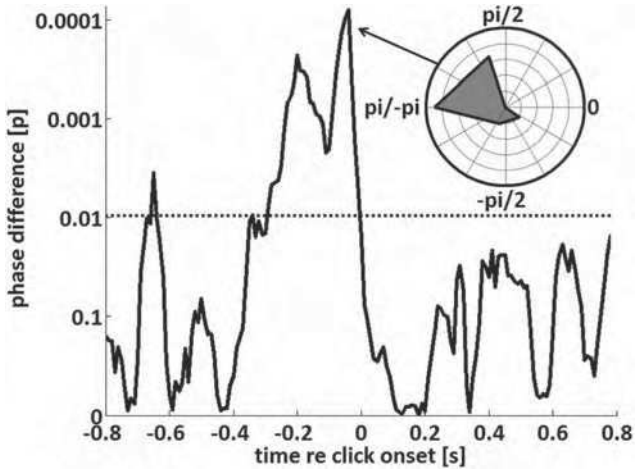


Figure 5. Statistical significance (Watson-William test) of the circular difference between modulatory phases for the two tone-pip frequencies. A significant test value would indicate a modulation of tone-pip detection by the phase of the original signal envelope whose direction (e.g., favorable or disadvantageous) depends on tone-pip frequency. Note the similarity of the time window of significance to Figure 4*B*. The distribution of phase differences across subjects is shown as an inset for the time point of maximal significance. Note that this phase distribution is strongly biased toward $\pm\pi$, indicating a phase opposition of the modulatory effect at that time. The FDR-corrected significance threshold is shown by a dotted line.

As mentioned above, we found perceptual phase entrainment in both experiments, but the actual phase of the modulatory effect seemed to depend on tone-pip frequency. Therefore, to quantify this effect, for each time point, we compared the phases of the sine waves that were fitted to the performance of single subjects (thus obtaining nine phases for each time point and tone-pip frequency; Watson-William test: tests whether two phase distributions have a different mean). The results of this test are shown in Figure 5, indicating that the modulatory phases of the two experiments significantly differed in a time window of ~220 ms before tone-pip onset. The distribution of phase differences (Fig. 5, inset) at the time of maximal significance (at ~−50 ms with respect to tone-pip onset) is strongly biased toward $\pm\pi$, indicating a phase opposition of the modulatory effect at that time (Fig. 4*A*, compare respective time panels).

Perceptual phase entrainment depends on speech intelligibility

It is still under debate whether speech comprehension is critical for phase entrainment to speech sound (Peelle and Davis, 2012; Ding and Simon, 2014). Thus, we ran a third experiment in order to test the role of intelligibility for the observed phase entrainment. Again, subjects were asked to detect (2400 Hz) tone pips. However, the entraining speech/noise sound was presented either forwards or backwards. Obviously, if intelligibility played an important role for phase entrainment, one would expect a modulation of tone-pip detection by the original envelope phase only in the forward (but not in the backward) condition. Indeed, we were able to replicate the perceptual modulation observed in Experiment 1 (Fig. 6). Note that *p* values (corresponding to Fig. 4*B*) are smaller for the forward condition of Experiment 3 (Fig. 6*A*, blue line) than those for Experiment 1 (Fig. 6*A*, black line) due to a higher number of tone pips in Experiment 1 (1089 vs 531). However, both the magnitude and phase of the actual perceptual modulation (as visible in Fig. 6*B* for the time point of −230 ms with respect to tone-pip onset; corresponding to Fig. 4*A*) are very similar. Importantly, the effect is abolished for the backward condition (Fig. 6, red line), suggesting that the observed perceptual phase entrainment depends on speech intelligibility.

In short, the presented findings demonstrate a frequency-specific high-level phase entrainment to speech sound that is tightly linked to intelligibility. These findings are as follows: (1) whether a tone pip is detected or not depends on the timing of high-level features before the tone pip is presented; (2) the direction of this dependence depends on the location of the tone pip with respect to the frequency content of the entraining stimuli (the detection of a tone pip located beyond or within the principal frequency range of the speech sound is modulated by preceding high-level features in an opposite fashion); and (3) the effect is abolished when speech/noise snippets are presented in reverse, indicating an important role of intelligibility for perceptual phase entrainment.

The observed effects are not due to small residual spectral differences during stimulus construction

One might argue that small residual spectral differences across original envelope phases in our stimuli could have been sufficient to produce the results presented here. Indeed, our stimulus construction method is based on iterative convergence toward a phase-equalized spectrum (see Materials and Methods), and small residual spectral differences can therefore not be totally excluded. To control for this, we designed noise stimuli displaying the average spectral content of our constructed speech/noise snippets either at the peak (Fig. 3*B*, phase 0) or the trough (Fig. 3*B*, phase $\pm\pi$) of the original signal envelope, without any rhythmic modulation. We had seven subjects listen to these steady noise stimuli. We asked them to report the detection of embedded tone pips (same tone-pip frequencies as in Experiments 1/3 and 2) presented at random moments. The amplitude of the tone pips was determined in a pre-experiment such that ~50% of the tone pips were detected when embedded in noise, presenting a spectrum equal to the average of the two spectra. During the control experiment, the tone-pip amplitude was kept constant. Importantly, detection of the tone pips did not differ between the

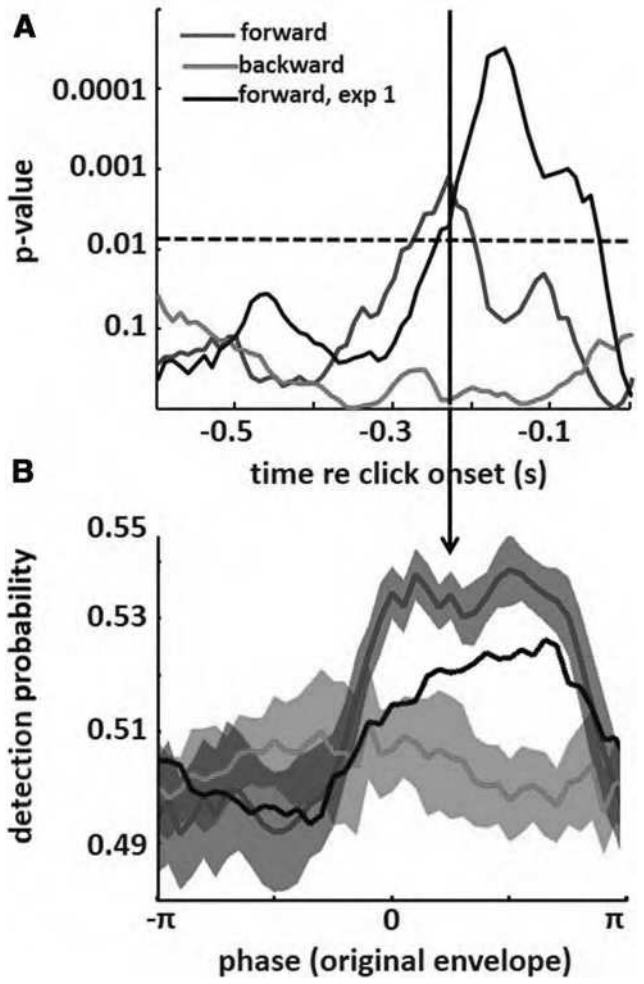


Figure 6. Perceptual phase entrainment depends on intelligibility. ***A***, Statistical significance of perceptual phase entrainment in Experiment 3, corresponding to Figure 4*B* (for ease of comparison, results from Experiment 1 are also shown). The modulation of click detection by original envelope phase is significant (between 280 and 200 ms before tone-pip onset) for the forward condition (blue line; replicating the effect from Experiment 1, shown in black), but not for the backward condition (red line), indicating that the observed perceptual phase entrainment depends on intelligibility. Significance threshold is FDR-corrected and shown as a dotted line. ***B***, Tone-pip detection probability as a function of the original signal envelope at −230 ms with respect to tone-pip onset. Note the sinusoidal modulation of performance for the forward condition, which is absent for the backward condition. This modulation has a similar degree and a similar phase as in Experiment 1, indicating that the effect could indeed be replicated. SEM is shown by contours around the lines.

two types of noise [on average, 52.2 vs 53.4% (2400 Hz pips; Fig. 7*A*) and 44.0 vs 45.0% (100 Hz pips; Fig. 7*B*) for peak vs trough, not significantly different: $p = 0.79$ (2400 Hz pips) and $p = 0.81$ (100 Hz pips), Student's *t* test]. This result suggests that our phase entrainment results cannot be explained by small residual spectral differences across original envelope phases. A related concern is whether momentary remaining fluctuations in the envelope of our constructed stimuli might influence perception as well (such momentary fluctuations are due to the fact that it is only the average spectral content that is comparable across original envelope phases). To test for this, we computed the modulation of tone-pip detection by the signal envelope of the constructed snippets (corresponding to Fig. 4*B*, which shows results from the same analysis, but with respect to the signal envelope of the original snippets). However, *p* values did not reach significance (same FDR-corrected significance threshold as for Fig. 4*B*), indicating that those remaining random fluctuations did not affect tone-pip detection (data not shown).

Discussion

Many studies showing phase entrainment to speech (Ahissar et al., 2001; Luo and Poeppel, 2007; Nourski et al., 2009; Kerlin et al., 2010; Horton et al., 2013; Millman et al., 2013; Zion Golumbic et al., 2013a, 2013b) did not fully disentangle low-level effects of fluctuations in sound amplitude or spectral content (corresponding to information processed in the cochlea) and covarying high-level effects (i.e., phase entrainment beyond the cochlear level). Theoretically, their findings could thus involve a passive phase locking of the auditory system to amplitude changes in the signal envelope (VanRullen et al., 2014). In particular, peaks and troughs in the signal envelope could systematically reset ongoing oscillations (Gross et al., 2013; Doelling et al., 2014), possibly resulting in regular repetitions of evoked potentials (i.e., ASSR)—and an apparent alignment between brain oscillations and speech sound. A conclusive answer to the question whether phase entrainment entails an active component was still lacking. Although the term "envelope tracking" is commonly used in the literature to describe the adjustment of oscillations to the regularity of speech sound (Giraud and Poeppel, 2012), the implied active mechanism has never been directly shown. Potential evidence for an active mechanism would be the demonstration that phase entrainment to speech entails a high-level process, such as phase locking to speech sound in the absence of systematic low-level fluctuations. Indeed, in this study, we provide evidence for an active, high-level process in phase entrainment to speech sound. We do so by controlling for differences in low-level properties (fluctuations in sound amplitude and spectral content) in speech sound while keeping both high-level information and intelligibility intact (Fig. 1*B*, bottom; Fig. 3).

An increasing number of recent studies have investigated the phenomenon of phase entrainment to speech, some of them highly relevant for the question mentioned above. In Doelling et al. (2014), each peak in the signal envelope was replaced by a peak of uniform height and shape (i.e., a pulse train was centered at envelope peaks, reflecting the stimulus rhythm), leaving only envelope cues but severely reducing high-level information. Although they found a drastic reduction of intelligibility, phase locking between brain oscillations and speech stimuli was not abolished, clearly underlining the importance of low-level features for phase entrainment. On the other hand, evidence for a high-level influence in phase entrainment to speech was reported by Peelle and colleagues (2013) and confirmed by Gross et al. (2013). They showed that, although neural oscillations did entrain to unintelligible sentences, phase locking was enhanced when intelligibility of speech was restored (Luo and Poeppel, 2007, their Fig. S2). Complementary to our study, Ding et al. (2013) reduced spectrotemporal fine structure of speech without strongly affecting the signal envelope. This manipulation resulted in an impaired phase locking to the signal envelope, indicating a role of high-level features in phase entrainment to speech. However, these reports of a high-level component in phase entrainment are still indirect, as they measure high-level modulations (e.g., by attention: Peelle et al., 2013; Zion Golumbic et al., 2013b; or intelligibility: Gross et al., 2013) of a phase entrainment that may still be primarily driven by low-level features of speech sound. Indeed, the existing literature remains compatible with the notion that low-level fluctuations are necessary for phase entrainment, and that this entrainment is merely upregulated or downregulated by high-level factors, such as

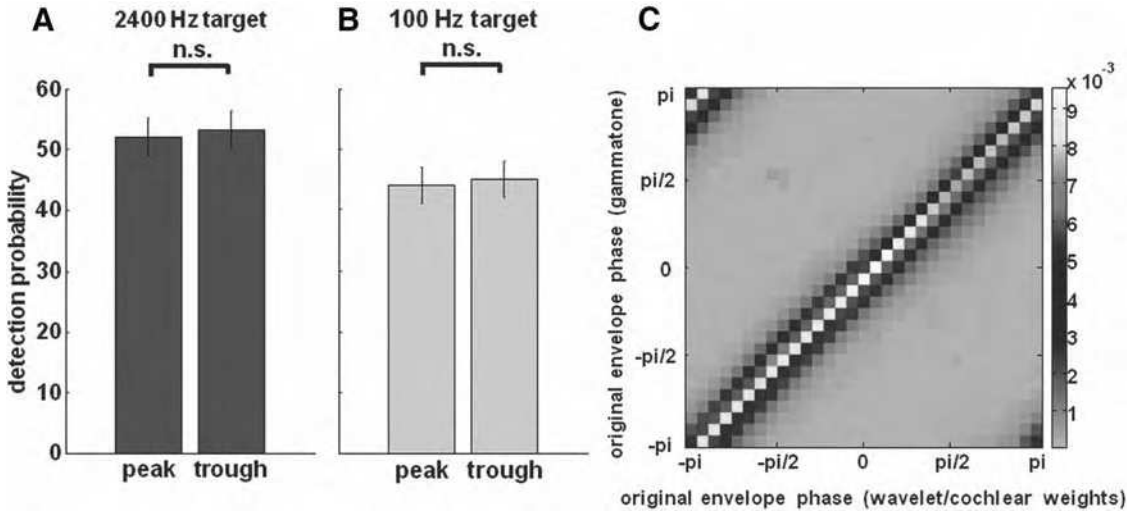


Figure 7. Control experiments/analyses to rule out alternative explanations for the observed high-level modulation of click detection. ***A***, ***B***, Results of a control experiment in which subjects were asked to detect (***A***, 2400 Hz; ***B***, 100 Hz) tone pips that were embedded in different noises. The noises were designed to have the average spectral content of the constructed speech/noise snippets either at the peak (Fig. 3*B*, phase 0) or the trough (Fig. 3*B*, phase $\pm\pi$) of the original signal envelope (~500 tone pips for each pip frequency and noise spectrum). As tone-pip detection probability did not differ between the two spectra, our effect cannot be explained by small residual spectral differences across original envelope phases in our constructed speech/noise stimuli. SEM across subjects is shown by error bars. ***C***, Correlation of original envelope phases extracted as the sum of power across frequencies, either obtained by wavelet transformation (using 304 logarithmically spaced frequencies in the range between 0.59 and 21,345 Hz) and weighted by the cochlear sensitivity (*x*-axis), or by gammatone filtering (*y*-axis; 64 gammatone filters in the frequency range between 80 and 8020 Hz). Note that phases were divided into bins. The color bar shows the proportion (in percentage) of phase pairs (i.e., the phase obtained by wavelet transformation and cochlear weights paired with the corresponding phase obtained by gammatone filtering) falling into each phase bin. The similarity between the two methods for phase extraction is evident in that essentially all phase pairs lie on the diagonal, indicating that our method (wavelet transformation and cochlear weights) is essentially equivalent to that using cochlear (gammatone) filtering (see Materials and Methods).

attention or intelligibility. By contrast, the technique presented here allowed us, for the first time, to isolate high-level fluctuations in speech sound: we could demonstrate that low-level features of speech are not necessary to induce phase entrainment, and thus that this entrainment truly entails a high-level process.

Furthermore, the role of intelligibility in phase entrainment is currently under debate: on the one hand, intelligibility is not required for entrainment to speech sound (Howard and Poeppel, 2010; Luo and Poeppel, 2012; Peelle et al., 2013); on the other hand, phase entrainment is enhanced in intelligible compared with nonintelligible sentences (Gross et al., 2013; Peelle et al., 2013). Here, we were able to show that high-level perceptual phase entrainment depends on intelligibility (Fig. 6), indicating that comprehension is indeed an important factor for the brain's adjustment to speech sound.

Again, we note that low-level and high-level features of speech are not well defined in the literature. For instance, phonemes are normally defined in terms of what is perceived than in terms of acoustic patterns, and auditory processing beyond the cochlear level is complex and not well understood (Moore, 2003; Nelken, 2008). Because all phases were equalized with respect to their amplitude and spectral content, the underlying rhythm cannot passively entrain the lowest (i.e., cochlear) level of auditory processing, where every rhythmic phase should normally excite the same cells to the same extent. We do not want to discount the possibility, however, that part of the remaining auditory fluctuations may entail "low-level" features, as this directly depends on one's specific definition of low-level versus high-level speech information. What we can argue, nonetheless, is that some of the most basic low-level properties of speech (fluctuations in sound amplitude and spectral content) are not strictly necessary for phase entrainment—something that had not been shown before.

It was recently demonstrated that phase entrainment of neural oscillations can be frequency specific: Lakatos et al. (2013) and O'Connell et al. (2011) presented regular trains of clicks while recording in different regions of monkey A1, whose response properties either matched the frequency of the click ["best frequency" (BF) regions] or did not match the frequency of the click (non-BF regions). They found that only BF regions entrained their high-excitability phase to the expected click onset whereas entrainment of low-excitability phase was found in non-BF regions, indicating a suppression of neuronal responses in regions not tuned to the respective frequencies. This property of phase entrainment can be described as a spectrotemporal amplifier–attenuator system: important events are aligned with periodically reoccurring "windows of opportunity" (Buzsáki and Draguhn, 2004) for stimulus processing, but only in brain regions processing the concerned frequencies, resulting in an alignment of the high-excitability phase with and an amplification of the expected events. Critically, these "windows" are closed in brain regions processing irrelevant frequencies, resulting in an alignment of low-excitability phase with and an attenuation of unexpected or irrelevant events (Lakatos et al., 2013). In our study, we show for the first time that frequency-specific entrainment can also be found in humans and that it directly affects perception. Whereas we found perceptual phase entrainment in both experiments, reflected by a modulation of tone-pip detection by the phase of the original envelope, the actual phase of the modulatory effect depended on the frequency of the tone pip and was different for tone pips beyond and within the principal frequency content of the speech stimuli, respectively.

Although a logical next step could be to combine these psychophysical data with electrophysiological recordings, for instance using electroencephalography (EEG), we argue that our present demonstration of phase entrainment at the perceptual (rather than the neural) level is, in this case, even more sensitive.

Our results must imply that at least some neuronal population is tracking the rhythm of high-level features. On the other hand, this population's activity may or may not be sufficiently high to be visible in EEG. Further studies are necessary to clarify this issue. Furthermore, in the present study, we constructed speech/noise stimuli such that the average activity profile in the cochlea does not differ across original signal envelope phases. However, it is possible that for some phases the cochlear activity deviates around the mean (across different envelope cycles) more than for others. Similarly, we did not equalize instantaneous spectral entropy across original envelope phases. Thus, as a next step, it would be interesting to use stimuli where not only the mean spectral content is made comparable across envelope phases (a first-order control), but also its temporal variability and/or spectral entropy (a second-order control).

In this study, entrainment to speech sound lasted only one cycle of the signal envelope (200–250 ms; Fig. 4). Based on previous reports in vision, this finding might seem surprising at first glance: it has been shown that visual stimuli influence the visual system for a much longer time (VanRullen and Macdonald, 2012). One reason for this discrepancy between visual and auditory processing might originate from the need for flexible sampling of the auditory system, which, in contrast to the visual system, relies heavily on temporal, continuously changing information (Thorne and Debener, 2014). Whereas a visual scene might be stable for a relatively long time, acoustic stimuli are fluctuating by nature. Consequently, temporal predictions (reflected by phase entrainment) about the "auditory world" might only be possible for the near future; predictions that reach too far ahead might easily turn out to be wrong and could even be disruptive for auditory processing (VanRullen et al., 2014). In line with our finding, Lalor and Foxe (2010) showed that the signal envelope of speech sound is reflected in EEG data for a period corresponding to ~1 envelope cycle (200–250 ms).

In conclusion, as it was demonstrated here, perceptual phase entrainment in the auditory system is possible in the absence of spectral energy fluctuations. Our results indicate that, even in response to nontrivially rhythmic stimuli (not containing any obvious rhythmic fluctuations in their lowest-level features), the brain actively generates predictions about upcoming input by using stimulus features on a relatively high cognitive level (which is necessary when differentiating speech from noise). These predictions depend on intelligibility of the underlying speech sound and have frequency-specific consequences on stimulus processing in the auditory system, opening "windows of opportunity" for relevant frequencies, but closing them for others.

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. CrossRef Medline

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Methodol 57:289–300.

Berens P (2009) CircStat: A Matlab toolbox for circular statistics. J Stat Softw 31:1–21.

Brainard DH (1997) The psychophysics toolbox. Spat Vis 10:433–436. CrossRef Medline

Buzsáki G, Draguhn A (2004) Neuronal oscillations in cortical networks. Science 304:1926–1929. CrossRef Medline

Ding N, Simon JZ (2013) Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. J Neurosci 33:5728–5735. CrossRef Medline

Ding N, Simon JZ (2014) Cortical entrainment to continuous speech: functional roles and interpretations. Front Hum Neurosci 8:311. CrossRef Medline

Ding N, Chatterjee M, Simon JZ (2013) Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. Neuroimage 88C:41–46. CrossRef Medline

Doelling KB, Arnal LH, Ghitza O, Poeppel D (2014) Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. Neuroimage 85:761–768. CrossRef Medline

Fiebelkorn IC, Snyder AC, Mercier MR, Butler JS, Molholm S, Foxe JJ (2013) Cortical cross-frequency coupling predicts perceptual outcomes. Neuroimage 69:126–137. CrossRef Medline

Galambos R, Makeig S, Talmachoff PJ (1981) A 40-Hz auditory potential recorded from the human scalp. Proc Natl Acad Sci U S A 78:2643–2647. CrossRef Medline

Ghitza O (2001) On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception. J Acoust Soc Am 110:1628–1640. CrossRef Medline

Ghitza O (2011) Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. Front Psychol 2:130. CrossRef Medline

Ghitza O (2012) On the role of theta-driven syllabic parsing in decoding speech: intelligibility of speech with a manipulated modulation spectrum. Front Psychol 3:238. CrossRef Medline

Ghitza O (2013) The theta-syllable: a unit of speech information defined by cortical function. Front Psychol 4:138. CrossRef Medline

Ghitza O, Giraud AL, Poeppel D (2012) Neuronal oscillations and speech perception: critical-band temporal envelopes are the essence. Front Hum Neurosci 6:340. CrossRef Medline

Giraud AL, Poeppel D (2012) Cortical oscillations and speech processing: emerging computational principles and operations. Nat Neurosci 15:511–517. CrossRef Medline

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. CrossRef Medline

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. CrossRef Medline

Horton C, D'Zmura M, Srinivasan R (2013) Suppression of competing speech through entrainment of cortical oscillations. J Neurophysiol 109:3082–3093. CrossRef Medline

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. CrossRef Medline

Immerseel LV, Peeters S (2003) Digital implementation of linear gammatone filters: comparison of design methods. Acoust Res Lett Online 4:59–64. CrossRef

Kerlin JR, Shahin AJ, Miller LM (2010) Attentional gain control of ongoing cortical speech representations in a "cocktail party." J Neurosci 30:620–628. CrossRef Medline

Lakatos P, Musacchia G, O'Connel MN, Falchier AY, Javitt DC, Schroeder CE (2013) The spectrotemporal filter mechanism of auditory selective attention. Neuron 77:750–761. CrossRef Medline

Lalor EC, Foxe JJ (2010) Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. Eur J Neurosci 31:189–193. CrossRef Medline

Luo H, Poeppel D (2007) Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron 54:1001–1010. CrossRef Medline

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. CrossRef Medline

Millman RE, Prendergast G, Hymers M, Green GG (2013) Representations of the temporal envelope of sounds in human auditory cortex: can the results from invasive intracortical "depth" electrode recordings be replicated using non-invasive MEG "virtual electrodes"? Neuroimage 64:185–196. CrossRef Medline

Moore BCJ (2003) An introduction to the psychology of hearing. Amsterdam: Academic.

Nelken I (2008) Processing of complex sounds in the auditory system. Curr Opin Neurobiol 18:413–417. CrossRef Medline

Ng BS, Schroeder T, Kayser C (2012) A precluding but not ensuring role of entrained low-frequency oscillations for auditory perception. J Neurosci 32:12268–12276. CrossRef Medline

Nourski KV, Reale RA, Oya H, Kawasaki H, Kovach CK, Chen H, Howard MA 3rd, Brugge JF (2009) Temporal envelope of time-compressed speech represented in the human auditory cortex. J Neurosci 29:15564–15574. CrossRef Medline

O'Connell MN, Falchier A, McGinnis T, Schroeder CE, Lakatos P (2011) Dual mechanism of neuronal ensemble inhibition in primary auditory cortex. Neuron 69:805–817. CrossRef Medline

Peelle JE, Davis MH (2012) Neural oscillations carry speech rhythm through to comprehension. Front Psychol 3:320. CrossRef Medline

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. CrossRef Medline

Poeppel D (2003) The analysis of speech in different temporal integration windows: cerebral lateralization as "asymmetric sampling in time." Speech Commun 41:245–255.

Schroeder CE, Lakatos P (2009) Low-frequency neuronal oscillations as instruments of sensory selection. Trends Neurosci 32:9–18. CrossRef Medline

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. CrossRef Medline

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. CrossRef Medline

VanRullen R, Macdonald JS (2012) Perceptual echoes at 10 Hz in the human brain. Curr Biol 22:995–999. CrossRef Medline

VanRullen R, Zoefel B, Ilhan B (2014) On the cyclic nature of perception in vision versus audition. Philos Trans R Soc Lond B Biol Sci 369:20130214. CrossRef Medline

Zion Golumbic E, Cogan GB, Schroeder CE, Poeppel D (2013a) Visual input enhances selective speech envelope tracking in auditory cortex at a "cocktail party." J Neurosci 33:1417–1426. CrossRef Medline

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 (2013b) Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party." Neuron 77:980–991. CrossRef Medline

Zoefel B, Heil P (2013) Detection of near-threshold sounds is independent of EEG phase in common frequency bands. Front Psychol 4:262. CrossRef Medline

CHAPTER 5: EEG OSCILLATIONS ENTRAIN THEIR PHASE TO HIGH-LEVEL FEATURES OF SPEECH SOUND

In the previous chapter, it was shown that perception can entrain to high-level features of speech sound even when not accompanied by fluctuations in amplitude or spectral content. However, it remained to be shown that this *perceptual* entrainment goes in line with *neural* entrainment. This is done in the following article, in which the EEG of human subjects was recorded while they listened to the constructed speech/noise stimuli. Neural phase entrainment was assessed by measuring the coherence of the phase of neural oscillations (as measured in the EEG) with the phase of the envelope of the presented speech sound. Indeed, we found that neural oscillations entrain to the rhythm of speech even when high-level information is not accompanied by changes in sound amplitude or spectral content. Moreover, this entrainment was not quantitatively different from that in response to everyday speech and was not abolished when linguistic information was removed (by time-reversing the stimuli). The latter result is in contrast to our perceptual findings (Chapter 4) and will be discussed in detail in Chapter 7. Additionally, cross-correlation analyses revealed two prominent time lags of coherence between original speech and neural signal at ~110 ms and 190 ms, whereas only the latter component was present for constructed speech/noise snippets. The former, earlier component reflects low-level processing, as it is only present in response to everyday speech (but not in response to the constructed speech/noise sound), and was found to be located in the gamma-band (~30-60 Hz). The latter, later component reflects high-level processing, as it is present in response to the constructed speech/noise sound as well, and was found to be located in the theta-band (~2-8 Hz). Finally, we found that, it is not the degree of entrainment (as mentioned above), but rather the entrained

phase that differs between entrainment in response to everyday speech (containing fluctuations in both low- and high-level features) and in response to the constructed speech/noise sound (containing fluctuations in high-level features only).

Thus, in the following article, it is shown that the entrainment of neural oscillations might underlie the perceptual entrainment observed in the previous chapter. This neural entrainment is characterized in detail with respect to its spectrotemporal properties. However, it is also demonstrated that perceptual and neural entrainment do differ in some aspects (e.g., in their dependence on intelligibility), and reasons for this discrepancy are discussed.

Article:

Zoefel B, VanRullen R (in press) EEG oscillations entrain their phase to high-level features of speech sound. NeuroImage.

NeuroImage 124 (2016) 16–23

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

EEG oscillations entrain their phase to high-level features of speech sound

Benedikt Zoefel *, Rufin VanRullen

Université Paul Sabatier, Toulouse, France
Centre de Recherche Cerveau et Cognition (CerCo), CNRS, UMR5549, Pavillon Baudot CHU Purpan, BP 25202, 31052 Toulouse Cedex, France

ARTICLE INFO

Article history:
Received 17 April 2015
Accepted 19 August 2015
Available online 1 September 2015

Keywords:
EEG
Oscillation
Phase entrainment
High-level
Speech
Auditory
Intelligibility

ABSTRACT

Phase entrainment of neural oscillations, the brain's adjustment to rhythmic stimulation, is a central component in recent theories of speech comprehension: the alignment between brain oscillations and speech sound improves speech intelligibility. However, phase entrainment to everyday speech sound could also be explained by oscillations passively following the low-level periodicities (e.g., in sound amplitude and spectral content) of auditory stimulation—and not by an adjustment to the speech rhythm *per se*. Recently, using novel speech/noise mixture stimuli, we have shown that behavioral performance can entrain to speech sound even when high-level features (including phonetic information) are not accompanied by fluctuations in sound amplitude and spectral content. In the present study, we report that neural phase entrainment might underlie our behavioral findings. We observed phase-locking between electroencephalogram (EEG) and speech sound in response not only to original (unprocessed) speech but also to our constructed "high-level" speech/noise mixture stimuli. Phase entrainment to original speech and speech/noise sound did not differ in the degree of entrainment, but rather in the actual phase difference between EEG signal and sound. Phase entrainment was not abolished when speech/noise stimuli were presented in reverse (which disrupts semantic processing), indicating that acoustic (rather than linguistic) high-level features play a major role in the observed neural entrainment. Our results provide further evidence for phase entrainment as a potential mechanism underlying speech processing and segmentation, and for the involvement of high-level processes in the adjustment to the rhythm of speech.

Introduction

The auditory environment is essentially rhythmic (e.g., music, speech, animal calls), and relevant information (e.g., phonemes, sounds) alternates with irrelevant input (such as silence in-between) in a regular fashion. Based on these environmental rhythms, the brain might have developed a clever tool for an efficient way of stimulus processing (Calderone et al., 2014; Schroeder and Lakatos, 2009): Neural oscillations could align their high excitability (i.e., amplifying) phase with regularly occurring important events, whereas their low excitability (i.e., suppressive) phase could coincide with irrelevant events.

This phenomenon has been called *phase entrainment* and has been shown to improve speech intelligibility (Ahissar et al., 2001; Kerlin et al., 2010; Luo and Poeppel, 2007). However, the presented stimuli in most experiments contain pronounced fluctuations in (sound) amplitude and may simply evoke a passive "amplitude following" of brain oscillations (i.e., auditory steady-state potentials, ASSR; Galambos et al., 1981). In other words, past reports of phase entrainment to speech might reflect an adjustment to fluctuations in low-level features and/or to co-varying high-level features[1] of speech sound. Critically, in the former case, phase entrainment would only reflect the periodicity of the auditory stimulation and could not be seen as an *active* "tool" for efficient stimulus processing (VanRullen et al., 2014). On the other hand, were one able to observe phase adjustment to (hypothetical) speech-like stimuli that retain a regular speech structure but that do not evoke ASSR at a purely sensory level of auditory processing (such as the cochlea), this would provide important evidence for the proposed active mechanism of stimulus processing (Giraud and Poeppel, 2012; Schroeder et al., 2010). Recently, we reported the construction of such stimuli (Zoefel and VanRullen, 2015)—speech/noise snippets with conserved patterns of high-level features, but without concomitant changes in sound amplitude or spectral content. We could show that

* Corresponding author at: Centre de Recherche Cerveau et Cognition (CerCo), Pavillon Baudot CHU Purpan, BP 25202, 31052 Toulouse Cedex, France. Fax: +33 562 172 809.
E-mail address: zoefel@cerco.ups-tlse.fr (B. Zoefel).

[1] The definition of "low-level" and "high-level" features of speech sound is difficult and often vague. In this paper, "low-level" features are defined as those equated in our stimuli: sound amplitude and spectral content. Speech features are considered "high-level" if they cannot passively entrain the lowest levels of auditory processing (such as the cochlea). Necessarily, these high-level features include (but might not be restricted to) phonetic information, and it is difficult to assign a particular level of auditory processing to them (see Discussion). This issue is discussed extensively in Zoefel and VanRullen (2015).

http://dx.doi.org/10.1016/j.neuroimage.2015.08.054
1053-8119/

auditory behavioral performance entrains to those stimuli, as detection of a tone pip was modulated by the phase of the preserved high-level rhythm. However, it remained to be tested whether this behavioral modulation also entails neural phase entrainment.

In addition, we focus on a highly relevant question recently brought up by Peelle and Davis (2012), based on the previously reported correlation between phase entrainment and intelligibility (Ahissar et al., 2001; Kerlin et al., 2010; Luo and Poeppel, 2007): Does speech intelligibility enhance phase entrainment, or does phase entrainment enhance speech intelligibility? If the latter is true, so they argue, phase entrainment should occur based on *acoustic* (e.g., voice gender, identity) and not *linguistic* (e.g., semantic) information. Still, so far, this question remains unsolved: Although *behavioral* phase entrainment does depend on linguistic cues (the observed phase adjustment for our speech/noise mixture stimuli did not occur for time-reversed stimuli; Zoefel and VanRullen, 2015), this does not have to be the case for the potentially underlying *neural* phase entrainment. Thus, we compared entrainment of EEG oscillations to *original* (unprocessed) speech snippets with that to our *constructed* speech/noise mixture stimuli but also to *reversed* speech/noise snippets (Fig. 1).

Materials and methods

Participants

Twelve participants volunteered after giving written informed consent (7 female; mean age: 27.6 years). All participants reported normal hearing and received compensation for their time. The experimental protocol was approved by the relevant ethical committee at Centre National de la Recherche Scientifique (CNRS).

Experimental stimuli

A detailed description of stimulus construction was given by Zoefel and VanRullen (2015). In short, phase-specific auditory noise was added to original snippets such that sound amplitude and spectral content of our constructed speech/noise mixture stimuli were statistically

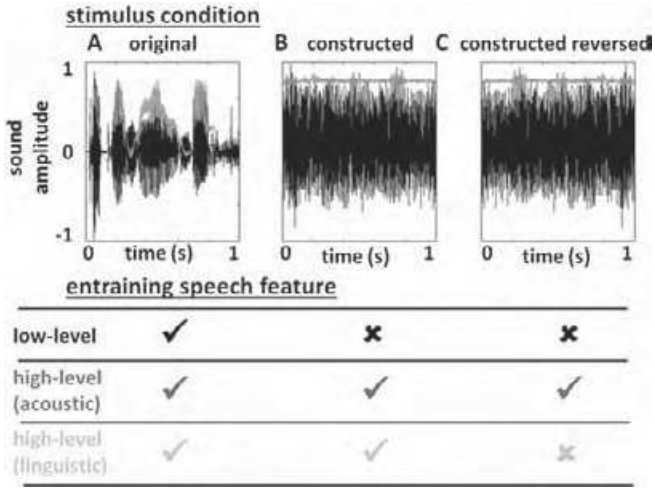


Fig. 1. The contribution of low- and high-level components of speech sound to phase entrainment was studied in three conditions (in the upper panel, for each condition, 1 s of an exemplary stimulus is shown in black, with its signal envelope in gray). Original speech snippets (A) were presented, entailing fluctuations in low-level (here defined as sound amplitude and spectral content) and both acoustic and linguistic high-level features of speech. Constructed speech/noise snippets (B; Zoefel and VanRullen, 2015) entailed both acoustic and linguistic high-level, but no systematic fluctuations in low-level features of speech. Finally, reversed constructed speech/noise snippets (C) were presented, entailing only acoustic high-level, but no linguistic or low-level fluctuations, designed in order to investigate the impact of intelligibility (i.e., linguistic information) on high-level phase entrainment.

comparable at all phases of the original signal envelope, φ_{env}. φ_{env} was first extracted for each individual original snippet s (a male native English speaker reading parts of a novel; sampling rate 44,100 Hz) as the sum of the instantaneous energy e (or amplitude; extracted by Wavelet Transformation for 304 logarithmicallyspaced frequencies in the range between 0.59 Hz and 21,345 Hz) at each time point t of the signal across frequencies F, weighted by the cochlear sensitivity w (ISO 226 equal-loudness contour signal for MATLAB, J. Tackett) in order to correct for differences in frequency sensitivity in the auditory system:

$$\varphi_{env}(s,t) = \frac{1}{F}\sum_{f=0}^{F} w(f) * e(s,f,t).$$

Then, speech/noise mixture stimuli were constructed by summing original speech snippets with a complementary, individually constructed noise: When spectral energy (the specific distribution of power across sound frequencies) of the original speech was high, that of the noise was low and vice versa. The spectral content of the noise was specific for each phase of the original signal envelope, resulting in constructed snippets whose mean spectral content did not differ across original envelope phases. Thus, systematic spectral energy fluctuations were removed by our stimulus processing and entrainment based on low-level properties of speech sound could thus be ruled out. However, speech sound was still intelligible and high-level features still fluctuated rhythmically at ~2–8 Hz (with the same timing as the original signal envelope, as low- and high-level cues in normal speech co-vary), providing potential means for oscillatory phase entrainment. Several sound samples for the different levels of stimulus construction (original speech snippet, constructed noise, final constructed speech/noise snippets) are available as Supplementary Material. Moreover, Supplementary Fig. 1 shows spectral energy as a function of original envelope phase for both original speech snippets and constructed speech/noise stimuli (reproduced from Zoefel and VanRullen, 2015). It can be seen that spectral energy is strongly concentrated at a certain envelope phase (phase 0; i.e., at the peak) for the original speech snippets. This imbalance of spectral energy can trivially and passively entrain the auditory system already at the level of the cochlea. Note that we corrected for this in our constructed speech/noise snippets: As spectral energy (but not other high-level features, such as phonetic information) is now equivalent across original envelope phases, neural entrainment in response to these stimuli is not trivial anymore and can be considered a high-level phenomenon.

Experimental paradigm

In this study, we were interested to determine low- and high-level components of neural phase entrainment. We thus designed three experimental conditions (Fig. 1) in order to dissociate the different components. Here, we made the distinction between acoustic high-level features of speech, cues that are specific to speech sound but are unrelated to speech comprehension (i.e., they are conserved even when the speech is reversed; for example, voice gender or identity), and linguistic high-level features of speech, cues that are specific to speech sound and important for speech comprehension (i.e., they are destroyed when the speech is reversed). In one condition ("original"), original speech snippets were presented, entailing rhythmic fluctuations in low-level and both acoustic and linguistic high-level features of speech (Fig. 1A). In another condition ("constructed"), our constructed speech/noise speech snippets (as described in the previous section) were presented, entailing rhythmic fluctuations in both acoustic and linguistic high-level features of speech (Fig. 1B). Finally, in the last condition ("constructed reversed"), we presented reversed constructed speech/noise speech snippets, entailing rhythmic fluctuations only in acoustic high-level information of speech (Fig. 1C). Note that, although intelligibility is removed by the reversal, some speech qualities are

preserved, enabling the listener to clearly distinguish noise and reversed speech (for instance, speakers can still be identified if speech is reversed, and time-reversed sentences can be discriminated based on neural phase information; Ding and Simon, 2014; Howard and Poeppel, 2010; Sheffert et al., 2002). Also, the essential properties of the signal envelope – i.e., the absence of systematic fluctuations in low-level features – remain unchanged compared to the constructed condition. For all conditions, one trial consisted of the presentation of a 10-s stimulus that was randomly chosen from all concatenated original or constructed snippets (total length about 10 min). Signals between concatenated snippets were interpolated to avoid artificial clicks that could potentially have influenced the subjects' EEG entrainment. Subjects listened to the stimuli while their EEG was recorded and completed 120 trials per conditions (in 3 blocks of 40 trials each; the block order was randomized, but such that the number of blocks per condition was always balanced during the experiment). In order to keep subjects engaged with the auditory stimulation, in each trial, between 3 and 5 (equal probability) tone pips were presented at threshold level at random moments during the trial. Tone pips had either a duration of 2.9 ms and a carrier frequency of 2.4 kHz, or a duration of 30 ms and a frequency of 100 Hz. The minimum interval between tone pips was 1 s. Subjects were asked to press a button whenever they detected a tone pip, with separate buttons for the two pip frequencies. Both tone pip frequencies could be presented in the same trial. The amplitude of the tone pip was adapted constantly (based on the performance of the preceding 100 trials) and separately for the two pip frequencies, so that tone pip detection had a mean probability of 50%. In this study, we did not focus on behavioral consequences of the entrainment (i.e., on a potential modulation of tone pip detection by the speech sound or EEG phase), for two reasons. Firstly, a behavioral modulation of tone pip detection by remaining high-level features of the constructed speech/noise stimuli was already reported in a previous study (Zoefel and VanRullen, 2015). Secondly, the number of tone pips per condition (<125 detected tone pips per subject, condition and tone pip frequency) was not sufficient to reliably separate trials as a function of phase (compared with ~500 detected tone pips per subject, condition and tone pip frequency in our previous study; see Discussion).

Stimuli were presented diotically via loudspeakers (Logitech Z130, Morges, Switzerland). The Psychophysics Toolbox for MATLAB (Brainard, 1997) was used for stimulus presentation.

EEG recordings

EEG was recorded continuously using a 64-channel ActiveTwo Biosemi system. Two additional electrodes (an active electrode, CMS, common mode sense, and a passive electrode, DRL, driven right leg) were used as reference and ground and to compose a feedback loop for amplifier reference. Horizontal and vertical electrooculograms were recorded by three additional electrodes around the subjects' eyes. Electrodes were placed according to the international 10–10 system. All signals were digitized at 1024 Hz, and highpass-filtered online above 0.16 Hz. Data were filtered (notch filters between 47 and 53 Hz to remove 50 Hz line noise, and between 80 and 90 Hz to remove electrical noise at the frequency of the screen's refresh rate, 85 Hz) and downsampled offline to 256 Hz and converted to an average reference. Trials were screened manually for eye or movement artifacts, and contaminated trials were rejected.

Triggers associated with the onset of each trial were sent to the EEG system via MATLAB using the parallel port as described in İlhan and VanRullen (2012). In short, a loud pulse followed by a jittered silence (0.75–1.25 s) was inserted before sound onset to serve as an analog trigger. The auditory signal was split into two cables, one connected to the speaker system (to be presented to the subject), and the other into the parallel port interface of the EEG system to be registered as a trigger along with the EEG stream. Correct detection of the trigger was ensured by its high amplitude (at least four times as high as the auditory stimulation). The silent interval between trigger and stimulus ensured that any ERP response to the click sound caused by the trigger had vanished at the start of the trial. The remainder of the sound sequence (10 s speech snippet in one of our 3 experimental conditions) never produced an erroneous detection of the trigger.

Data analyses

In the following analyses, whenever whole-trial signals were used, the first 500 ms of each trial were discarded, in order to avoid artificial phase-locking caused by evoked responses after sound onset.

All analyses were performed in MATLAB. The EEGLAB Toolbox (Delorme and Makeig, 2004) was used for pre-processing of EEG data.

Phase entrainment

Phase entrainment can be defined as the alignment between two rhythmic structures—in our study, we thus analyzed neural phase entrainment as the amount of phase-locking between EEG oscillations and the presented speech features. Note that low- and high-level features co-vary in normal speech sound: Slow amplitude fluctuations (here labeled as signal envelope) and the underlying fluctuations in spectral content (together defined as low-level features in this study) inevitably go along with fluctuations in high-level (acoustic and linguistic) features in everyday speech. Using the *same* analysis for all 3 conditions, we were thus able to evaluate phase entrainment to *different* features of speech sound: Phase-locking between original signal envelope and EEG reflects (1) both low- and high-level entrainment in the original condition, (2) only high-level entrainment, but based on both acoustic and linguistic information, in the constructed condition, and (3) high-level entrainment, but restricted to acoustic information, in the constructed reversed condition (Fig. 1).

According to Lachaux et al. (1999), the phase-locking value (PLV) between signal envelope and EEG was calculated, for each channel *ch*, as the norm of the difference between the phase of the filtered (in the theta-band, 2–8 Hz) original signal envelope (φ_{env}) and the phase of the correspondingly filtered EEG (φ_{eeg}), averaged in the complex domain across T time points, N trials, and S subjects:

$$PLV(ch) = \left| \frac{1}{S}\sum_{s=1}^{S}\frac{1}{N}\sum_{n=1}^{N}\frac{1}{T}\sum_{t=1}^{T} e^{i\left(\varphi_{env}(n,t)-\varphi_{eeg}(n,ch,t)\right)} \right|$$

φ_{env} and φ_{eeg} are defined as the phase angle of the Hilbert-transformed filtered original signal envelope and EEG, respectively. The PLV ranges between 0 (no phase-locking) and 1 (maximal phase-locking). Note that since our formula averages all phase vectors in the complex domain before the norm of the result is taken, the resulting PLV will be maximal when the phase angle difference between signal envelope and EEG is consistent across time, across trials, and across subjects. If the phase angle is computed instead of the norm, the phase difference between EEG and speech signal can be determined. We tested the significance of our results by comparing the observed PLVs, averaged across EEG channels, with surrogate distributions. Thus, we were able to determine (1) whether the PLV in any of the conditions significantly differs from 0 (reflecting significant phase entrainment to the speech or speech/noise stimuli) and (2) whether PLVs significantly differ across conditions. In order to test the significance of the obtained PLVs (1), a surrogate distribution was constructed by calculating PLVs as before, but with φ_{env} and φ_{eeg} drawn from different trials. In order to test whether the obtained PLVs differ across conditions (2), the difference in PLV was calculated for each possible combination of conditions (i.e., original vs. constructed, original vs. constructed reversed, constructed vs. constructed reversed). For each combination, a surrogate distribution was constructed by randomly assigning trials to the respective conditions and re-calculating the PLV difference. Both procedures (1 and 2) were repeated 1,000,000 times in order to obtain a range of PLVs

and PLV differences under the null hypotheses of no phase-locking between signal envelope and EEG signal and no difference in phase-locking between conditions, respectively. P-values were calculated for the recorded data by comparing "real" PLVs and PLV differences against the respective surrogate distributions. P-values were corrected for multiple comparisons across three conditions using the false discovery rate (FDR) procedure. Here, a significance threshold is computed which sets the expected rate of falsely rejected null hypotheses to 5% (Benjamini and Hochberg, 1995).

The PLV only indicates *overall* phase-locking between signal envelope and EEG, but no information can be obtained about its timing or the different frequency components involved. As an additional step, in order to evaluate spectro-temporal characteristics of the entrainment, we thus calculated the cross-correlation between signal envelope and EEG (Lalor et al., 2009; VanRullen and Macdonald, 2012), computed for time lags between −1 and 1 s:

$$cross-correlation\ (ch,t) = \sum_{T} env(T).eeg(ch,T+t)$$

where *env(T)* and *eeg(T)* denote the unfiltered standardized (z-scored) signal envelope and the corresponding standardized (z-scored) EEG response at time T and channel *ch*, respectively, and *t* denotes the time lag between envelope and EEG signal. Cross-correlations were averaged across trials and subjects, but separately for each channel, and time–frequency transforms of those cross-correlations were computed (using Fast Fourier Transformation (FFT) and Hanning window tapering; 128 linear-spaced frequencies from 1 Hz to 128 Hz; window size 0.5 s, zero-padded to 1 s). These time–frequency representations were then averaged across channels. Note that, due in part to the convolution theorem, this time–frequency analysis of the cross-correlation between signal envelope and EEG response is roughly equivalent to the sum of cross-correlations between narrow-band filtered versions of the signal envelope and EEG response.

In order to test the obtained results for significance (with the null hypothesis of no correlation between speech signal and brain response), EEG data from each trial were cross-correlated with the signal envelope from another trial and cross-correlations and their time–frequency representations were re-computed for this simulated set of data. By repeating this simulation (100,000 times), it was possible to obtain a range of time–frequency values that can be observed under the null hypothesis that speech and EEG signals are not correlated. P-values were calculated by comparing surrogate distribution and real data for each time–frequency point. P-values were again corrected for multiple comparisons using FDR.

In order to contrast cross-correlation effects across the different experimental conditions, a repeated-measurements one-way ANOVA was performed with condition as the independent variable (original, constructed, constructed reversed) and the standard deviation across electrodes of the cross-correlation values for a given time point as the dependent variable. Where necessary, p-values were corrected for non-sphericity using the Greenhouse–Geisser correction. Post-hoc tests were applied using paired t-tests and Bonferroni correction for multiple comparisons (threshold $p < 0.05$).

Results

We presented 12 subjects with speech/noise stimuli without systematic fluctuations in low-level features (here defined as sound amplitude and spectral content; see Zoefel and VanRullen (2015) for a detailed discussion of this definition), but with intact high-level features of speech sound, fluctuating at ~2 − 8 Hz ("constructed condition"). Additionally, those speech/noise snippets were presented in reverse ("constructed reversed condition"), thus potentially disentangling high-level features based on acoustic vs. linguistic information. We compared phase entrainment in those two conditions to that obtained in response to original speech snippets ("original condition"). Thus, we were, for the first time, able to dissociate 3 possible components of neural phase entrainment: Whereas systematic low-level feature changes were only present in the original condition, acoustic high-level information (independent of intelligibility; Peelle and Davis, 2012) was available in all three conditions, and linguistic high-level information was preserved in both the original and constructed conditions, but not in the constructed reversed condition (Fig. 1). Therefore, if neural phase entrainment were merely caused by ASSR to low-level features, it would happen only in the original condition; if it depended on the rhythmic structure of linguistic features, it should be seen in the original and constructed conditions but not in the constructed reversed condition; finally, if neural EEG phase mainly followed rhythmic fluctuations of acoustic high-level features, entrainment should occur in all three conditions. This latter result is what we observed, as detailed below.

Fig. 2A shows, for all conditions, average phase-locking (shown as bars) between the recorded EEG (filtered between 2 and 8 Hz) and the original signal envelope (filtered likewise; note that the *original* signal envelope reflects rhythmic fluctuations in both low- and high-level features in the original condition, in both acoustic and linguistic high-level features in the constructed condition, and only in acoustic high-level features in the constructed reversed condition; Fig. 1): Significant phase-locking, reflecting phase entrainment, is visible in all conditions. This phase entrainment does not significantly differ across conditions (original vs. constructed: $p = 0.120$; original vs. constructed reversed: $p = 0.199$; constructed vs. constructed reversed: $p = 0.052$; all p-values non-significant after FDR-correction), as determined by permutation tests (see Material and Methods). Topographies of PLVs are shown in Fig. 2B. A dipolar configuration appears in all conditions; this dipole is slightly shifted toward the right hemisphere for the original condition, in line with previous studies suggesting that slow amplitude fluctuations are preferentially processed in the right hemisphere (Abrams et al., 2008; Gross et al., 2013; Poeppel, 2003). The actual phase difference between EEG signal and original signal envelope is shown, separately for each EEG channel, in the topographies in Fig. 2C. Again, a dipolar configuration appears in all conditions; the polarity of this dipole seems to be inversed when comparing original and the two constructed conditions. Thus, whereas high-level features of speech sound can entrain EEG oscillations to a similar degree as unprocessed speech sound, the removal of systematic fluctuations in low-level features seems to be reflected in a change of the entrained phase.

Whereas Fig. 2 represents the overall amount and (phase) topographies of phase entrainment in the three conditions, precise temporal and spectral characteristics cannot be extracted. However, they might be necessary to explain the observed phase differences for the entrainment in the original and the two constructed conditions. We thus calculated the cross-correlation between EEG and original signal envelope, as described for example in VanRullen and Macdonald (2012). The outcome of this analysis (see Materials and Methods) provides an estimate of when (i.e., at which time lags) the stimulus (unfiltered original signal envelope) and response (unfiltered EEG signal) are related. Cross-correlations, averaged across trials and subjects, are shown in Fig. 3 (top panels) for all conditions, separately for all channels (black lines). Note that there are time lags at which many channels simultaneously deviate from their baseline, but with different polarities: The standard deviation across channels (shown in blue) can thus be used to quantify the magnitude of cross-correlation between overall EEG and the signal envelope at a given time lag. For the original condition, this standard deviation shows two peaks, one earlier component at ~110 ms, and another later component at ~190 ms with inversed polarity across the scalp (topographical maps for the peaks are shown as insets). Interestingly, only the later component is present in the constructed and constructed reversed conditions, potentially reflecting entrainment to (acoustic) high-level features of speech sound. Indeed, a one-way ANOVA on standard deviation values of single subjects for

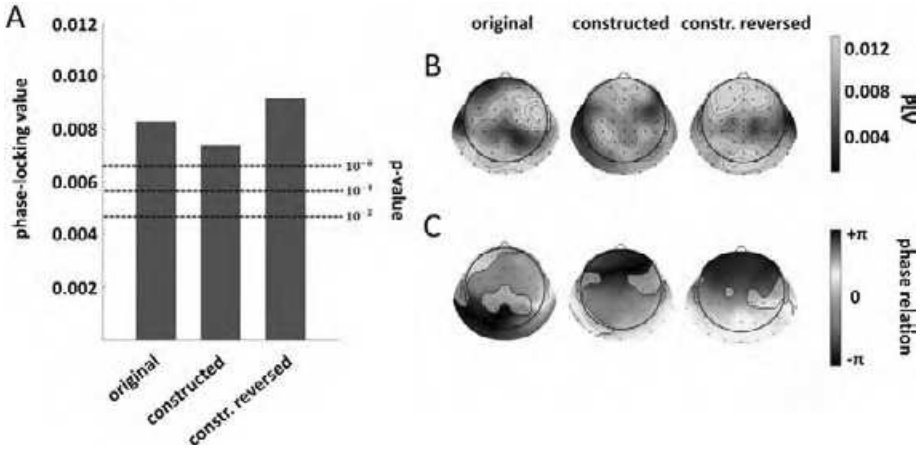


Fig. 2. Phase-locking between EEG signal and original signal envelope (i.e., phase entrainment) in the three conditions. A. The average phase-locking across channels is shown as bars. P-values of phase-locking (obtained by a permutation procedure; see Materials and Methods) are shown as dashed lines, indicating significant phase entrainment ($p < 10^{-6}$) in all conditions and thus a major role for high-level acoustic cues as the underlying entraining feature of speech sound. (Note that the p-value thresholds were obtained by independent permutation tests for the 3 experimental conditions, yet turned out near-identical.) B. Topographies corresponding to A, showing phase-locking values at each electrode location. A dipolar structure of phase entrainment is visible in all conditions. C. Topographies corresponding to A, representing the phase difference between EEG signal and original signal envelope at each electrode location. Electrodes without significant phase entrainment are shaded out. The dipolar structure of phase entrainment visible in B now shows an inversed polarity in the absence of low-level features of speech sound (constructed and constructed reversed condition).

the two time lags yields a significant effect of condition for the 110 ms time lag ($F(2) = 14.62$, $p = 0.002$), with post-hoc tests indicating a stronger cross-correlation for the original condition than for the two constructed conditions, but no significant effect of condition for the 190 ms time lag ($F(2) = 3.84$, $p = 0.057$). The inversed polarity between earlier low-level component (specific to the original condition) and later high-level component (present in all conditions) is reminiscent of the dipoles that were observed for the analysis of entrained phases (Fig. 2C) and showed an inversed polarity for the original and the two constructed conditions. This might suggest that the topography of entrainment phases is mainly driven by low-level components in the original condition, and by high-level components in the constructed conditions. Moreover, an early (~50 ms) cross-correlation component seems to be present in some conditions. Although a one-way ANOVA yields a main effect of condition at that time lag ($F(2) = 7.26$, $p = 0.004$), post-hoc tests reveal no significant difference between original and constructed reversed condition, a finding that rules out low-level entrainment involved in this peak (however, cross-correlation for both original and constructed reversed condition is significantly stronger than for the constructed condition at that time lag). In order to characterize spectral properties of the entrained responses, we computed a time–frequency transform of the cross-correlation signals (averaged across channels). We obtained significance values for each time–frequency point by comparing our cross-correlation results with surrogate distributions where EEG data from each trial was cross-correlated with the signal envelope from another trial (see Materials and Methods). Results are shown in the bottom panels of Fig. 3: Whereas high-level information common to all three conditions preferentially involves correlations in the theta-band, the low-level component at ~110 ms additionally entails gamma-band correlations (here ~20–50 Hz).

Thus, in summary, our results show (1) that phase entrainment of EEG oscillations is possible even when speech sound is not accompanied by fluctuations in low-level features, (2) that the removal of those

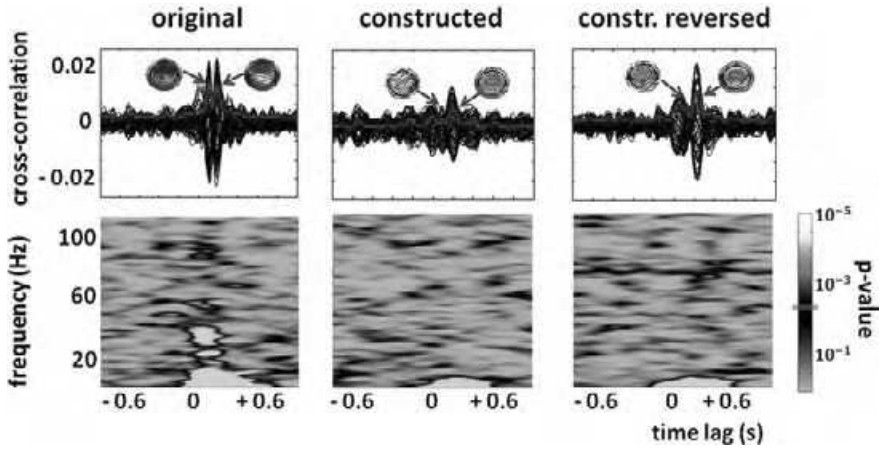


Fig. 3. Top panels: Cross-correlation between original signal envelope and EEG signal (both unfiltered) for all channels (black lines) and the standard deviation across channels (blue line). Only the original condition shows a peak in standard deviation at ~110 ms (time lag between speech and EEG), indicating an entrainment to low-level cues. A later peak (~190 ms) can be seen in all conditions, and an earlier, slightly weaker peak (~50 ms) that is most evident in the original and constructed reversed condition. Both peaks indicate an entrainment to acoustic high-level cues. The insets show the topographical distribution of cross-correlation, with respect to the timing of the two most pronounced peaks (110 ms and 190 ms) in the original condition. Bottom panels: Significance values of the time–frequency transform of cross-correlation functions (averaged across channels). Note that the 110-ms (low-level) component of cross-correlation involves significant correlations at higher frequencies including the gamma-range, whereas the other (high-level) components entail correlations restricted to the theta-range. FDR-corrected significance threshold, alpha < 0.05, is shown as a red line in the colorbar.

features results in a change of the entrained phase, (3) that linguistic information is not necessary for this high-level neural phase entrainment, and (4) that the entrainment to low- and high-level features occurs at different time lags between stimulus and EEG, with the entrainment to low-level features occurring earlier and in the gamma-range, whereas high-level entrainment occurs later (but with an additional, weaker peak occurring earlier than the low-level component; see Discussion) and in the theta-band.

Discussion

Phase entrainment of neural oscillations as a potential tool for efficient stimulus processing has been described repeatedly (Calderone et al., 2014; Lakatos et al., 2005, 2013; Schroeder et al., 2010; Schroeder and Lakatos, 2009) and is paramount in current theories of speech comprehension (Doelling et al., 2014; Ghitza, 2011, 2012, 2013, 2014; Giraud and Poeppel, 2012; Zion Golumbic et al., 2013). However, the underlying mechanisms are far from clear (Ding and Simon, 2014). Here, we disentangled the influence of low-level (i.e., sound amplitude and spectral content) and higher-level features of speech, by comparing neural entrainment to everyday speech sound to entrainment based only on high-level speech cues. Our results suggest that neural phase entrainment is not reduced when high-level features of speech are not accompanied by fluctuations in sound amplitude or spectral content (a complementary study, reaching a similar conclusion, has been presented by Ding et al., 2013). Instead, we observed a change in the phase difference between the entrained EEG oscillations and the speech sound. In line with a recent psychophysical study (Zoefel and VanRullen, 2015), this effect cannot be explained by a passive response to the periodic auditory stimulation at early stages of auditory processing (e.g., in the cochlea), and thus provides important evidence for phase entrainment as an active tool of efficient stimulus processing (e.g., Schroeder et al., 2010). Note a more elaborate discussion of definitions of low-level and high-level features of speech can be found in Zoefel and VanRullen (2015).

Based on the results obtained in our previous study, one would expect the detection of a tone pip, presented at random moments during stimulation, to depend on the phase of the entrained EEG oscillations (similar results have been reported in studies using non-speech sound as entraining stimulus: Henry and Obleser, 2012; Ng et al., 2012; note, however, that those results remain debated: Vanrullen and McLelland, 2013; Zoefel and Heil, 2013). However, in the present study, we did not attempt such behavioral analyses due to the reduced statistical power (each condition counted about one fourth of tone pip events compared to Zoefel and VanRullen, 2015; this was due to both the time-consuming preparation of EEG recordings and an increased number of experimental conditions). As expected, the examination of a potential modulation of tone pip detection by pre-stimulus EEG phase only showed negative results (data not shown). Thus, although we could not demonstrate behavioral consequences of the entrained EEG oscillations in the present study, we refer to our earlier psychophysical experiments where we could show (with sufficient statistical power) a modulation of perceptual behavior (i.e., tone pip detection) by the "high-level rhythm" in the same type of constructed speech/noise snippets (Zoefel and VanRullen, 2015). Further studies, using similar stimuli but with improved signal-to-noise ratio, are necessary in order to show simultaneous entrainment of behavior and electrophysiological markers.

The role of intelligibility in phase entrainment is currently debated. On the one hand, intelligibility is not required for entrainment to speech sound or other, simpler stimuli such as pure tones (Besle et al., 2011; Gross et al., 2013; Howard and Poeppel, 2010; Luo and Poeppel, 2012; O'Connell et al., 2011; Peelle et al., 2013; Stefanics et al., 2010; Zoefel and Heil, 2013); on the other hand, phase entrainment is sometimes enhanced in intelligible compared to non-intelligible sentences (Gross et al., 2013; Peelle et al., 2013). In our previous study, we found that behavioral phase entrainment to high-level speech cues is indeed reduced if speech intelligibility is abolished by reversing the stimuli (Zoefel and VanRullen, 2015), but this does not necessarily have to be the case for neural phase entrainment. In the present study, we compared entrainment to acoustic *and* linguistic high-level cues ("constructed condition") with that to acoustic high-level cues alone ("constructed reversed condition"). We found that the amount of phase entrainment did not differ between these two high-level conditions, indicating a principal role of acoustic (and not linguistic) features in the reported high-level phase entrainment of neural oscillations. These acoustic high-level features of speech might be the characteristic part of an intermediate step of speech analysis in the brain, prior to the actual linguistic processing (Hickok and Poeppel, 2007). Similar results (i.e., entrainment of neural oscillations by unintelligible speech) have been obtained by other groups (Howard and Poeppel, 2010; Millman et al., 2014). Furthermore, the observed results suggest an interesting mechanism, although very speculative, for the interaction between entrainment and intelligibility: Whereas we found neural (i.e., EEG) phase entrainment to both forward and time-reversed speech/noise sound, this neural entrainment only seemed to have perceptual consequences in behavioral measurements when the speech/noise sound was played forward (Zoefel and VanRullen, 2015). We can thus speculate that neural phase entrainment and "tone pip" stimulus detection might occur in different areas of the brain. In a recent study by Steinschneider et al. (2014), for instance, neuronal responses in temporal regions were modulated by the semantic context of sound, but did not predict behavioral outcome, which was only reflected in activity in prefrontal cortex. Moreover, it has been reported that phase entrainment to both attended and unattended speech can be observed in early cortical regions; however, the entrainment to unattended (but not attended) speech is "lost" in more frontal areas (Ding and Simon, 2012; Horton et al., 2013; Zion Golumbic et al., 2013). A similar effect might underlie our findings: Both intelligible and unintelligible speech might entrain early cortical regions, but only intelligible speech might entrain more frontal areas (and affect behavior). Thus, although intelligibility of speech might not directly affect the neural entrainment in regions of auditory processing, it might act as a crucial variable that determines whether the entrained neural activity affects decisions in frontal areas or not (or possibly, whether temporal and frontal areas are functionally connected; Weisz et al., 2014). Finally, Ding and Simon (2014) recently hypothesized that it might be necessary to differentiate entrainment to speech in the delta-range (1–4 Hz) from that in the theta-range (4–8 Hz), with the former adjusting to acoustic and the latter to phonetic information. As our stimulus construction was based on a signal envelope filtered between 2 and 8 Hz (comprising both delta- and theta-range), we were not able to separate our observed entrainment into those two frequency bands. It is possible that only theta-entrainment affected pip detection in our behavioral task and that this entrainment is indeed larger in the constructed condition than in the constructed reversed one. Clearly, further studies are necessary to determine under what circumstances linguistic cues are important for phase entrainment or not.

We acknowledge that, in the current study, we were only able to equalize speech features on a very early level of auditory processing (e.g., on the cochlear level), making it difficult to assign the observed entrainment to a particular level in the auditory pathway. Thus, we can speculate only based on the current literature: We presume that entrainment to low-level features of speech sound occurs relatively early in the auditory pathway (i.e., somewhere between cochlea and primary auditory cortex, including the latter; Davis and Johnsrude, 2003; Lakatos et al., 2005), whereas entrainment to high-level features occurs beyond primary auditory cortex (Uppenkamp et al., 2006). Important candidates are the supratemporal gyrus (more specifically, mid- and parietal STG) and sulcus (STS), which seem to be primarily involved in the analysis of phonetic features (Binder et al., 2000; DeWitt and Rauschecker, 2012; Hickok and Poeppel, 2007; Mesgarani et al.,

2014; Poeppel et al., 2012; Scott et al., 2000). To confirm these assumptions, it may thus be interesting to present our constructed speech/noise stimuli during intracranial recordings, which offer a spatial resolution vastly superior to that of EEG (Buzsáki et al., 2012).

Using a cross-correlation procedure, we were able to extract spectro-temporal characteristics of low- and high-level processing of speech sounds. Here, we observed an earlier (~110 ms) component reflecting low-level processing and involving the gamma-band, and a later (~190 ms) component that was spectrally restricted to the theta-band and potentially reflects high-level processing. Our results are consistent with the current literature, concerning both the observed timing, topography, and separation into low- and high-level components. For instance, Horton et al. (2013) reported very similar time lags and topographies when cross-correlating EEG with the envelope of normal speech sound. McMullan et al. (2013) presented subjects with first-order (change in energy) and higher-order (change in perceived pitch without change in overall energy) boundaries in the auditory scene and compared the responses measured in the EEG. Very similar to our study, they observed an earlier gamma-component in the response to first-order boundaries which was absent for the high-order stimuli. A later component in the theta-band was recorded for both types of boundaries. Krumbholz et al. (2003) compared magnetoencephalogram (MEG) responses to sound onset with those to a transition from noise to a discrete pitch without accompanying energy changes, and reported an earlier (~100 ms) component for the former, whereas perceived pitch produced a later (~150 ms) response. Finally, the frequencies of our observed cross-correlation components are in line with the currently emerging role of different oscillatory frequency bands (Bastos et al., 2014; Buffalo et al., 2011; Fontolan et al., 2014): Although more work needs to be done, there is accumulating evidence that faster frequency bands (e.g., the gamma-band) might reflect bottom-up mechanisms (i.e., processing of sensory information) whereas slower bands (e.g., the alpha-band) might be responsible for top-down mechanisms (i.e., processing of cognitive information, such as predictions about upcoming events). The two different frequency components (earlier "low-level gamma" and later "high-level theta") in our cross-correlation results support this idea and provide evidence for a similar mechanism in the auditory system. Note that the theta-band might have a similar role for the auditory system as the alpha-band for the visual system, as it seems to be related to higher-order cognitive functions, such as temporal predictions (Arnal and Giraud, 2012; Luo et al., 2013; Schroeder et al., 2010; Stefanics et al., 2010) or adjustment to musical rhythm (Nozaradan, 2014). Although not as easy to explain as the other two components, it is worth mentioning that an additional component of high-level processing appeared in our data: A peak around 50 ms was visible, entailing activity in the theta-band, whose amplitude did not statistically differ between original and constructed reversed conditions, suggestive of a high-level effect (however, we note that this peak was significantly less pronounced for the constructed condition, a finding that is not necessarily expected). We also point out that the seemingly early timing of the first high-level effect (~50 ms) does not contradict its being a high-level process. Indeed, the time lag of a cross-correlation between two quasi-rhythmic signals (signal envelope and brain activity) cannot be directly interpreted as the latency in response to a stimulus. For example, perfect phase synchronization (i.e., phase entrainment with no phase difference) between speech stimulus and entrained brain responses would result in a cross-correlation peak at time lag 0. Thus, a time lag of 50 ms does not necessarily mean that the stimulus is processed at relatively early latencies—it merely reflects the phase lag between stimulus and recorded signal.

In conclusion, by means of speech/noise stimuli without systematic fluctuations in sound amplitude or spectral content, we were able to dissociate low- and high-level components of neural phase entrainment to speech sound. We suggest that EEG phase entrainment includes an adjustment to high-level acoustic features, as neural oscillations phase-lock to these cues. We speculate that this entrainment to speech might only affect behavior when speech is intelligible, potentially mediated by an improved connectivity between temporal and frontal regions. Finally, low-level cues (e.g., large changes in energy) induce an additional response in the brain, differing from high-level EEG entrainment with respect to spectro-temporal characteristics, the entrained phase, and potentially anatomical location.

Conflict of Interest

The authors declare no competing financial interests.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2015.08.054.

Acknowledgements

The authors are grateful to Alain de Cheveigné and Daniel Pressnitzer for helpful comments and discussions. This study was supported by a Studienstiftung des deutschen Volkes (German National Academic Foundation) scholarship to BZ, and a EURYI Award as well as an ERC Consolidator grant P-CYCLES under grant agreement 614244 to RV.

References

Abrams, D.A., Nicol, T., Zecker, S., Kraus, N., 2008. Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. J. Neurosci. 28, 3958–3965. http://dx.doi.org/10.1523/JNEUROSCI.0187-08.2008.

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., Merzenich, M.M., 2001. Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. Proc. Natl. Acad. Sci. U. S. A. 98, 13367–13372. http://dx.doi.org/10.1073/pnas.201400998.

Arnal, L.H., Giraud, A.-L., 2012. Cortical oscillations and sensory predictions. Trends Cogn. Sci. 16, 390–398. http://dx.doi.org/10.1016/j.tics.2012.05.003.

Bastos, A.M., Briggs, F., Alitto, H.J., Mangun, G.R., Usrey, W.M., 2014. Simultaneous recordings from the primary visual cortex and lateral geniculate nucleus reveal rhythmic interactions and a cortical source for γ-band oscillations. J. Neurosci. 34, 7639–7644. http://dx.doi.org/10.1523/JNEUROSCI.4216-13.2014.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Methodol. 57, 289–300.

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. http://dx.doi.org/10.1523/JNEUROSCI.4518-10.2011.

Binder, J.R., Frost, J.A., Hammeke, T.A., Bellgowan, P.S., Springer, J.A., Kaufman, J.N., Possing, E.T., 2000. Human temporal lobe activation by speech and nonspeech sounds. Cereb. Cortex 10, 512–528.

Brainard, D.H., 1997. The psychophysics toolbox. Spat. Vis. 10, 433–436.

Buffalo, E.A., Fries, P., Landman, R., Buschman, T.J., Desimone, R., 2011. Laminar differences in gamma and alpha coherence in the ventral stream. Proc. Natl. Acad. Sci. U. S. A. 108, 11262–11267. http://dx.doi.org/10.1073/pnas.1011284108.

Buzsáki, G., Anastassiou, C.A., Koch, C., 2012. The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes. Nat. Rev. Neurosci. 13, 407–420. http://dx.doi.org/10.1038/nrn3241.

Calderone, D.J., Lakatos, P., Butler, P.D., Castellanos, F.X., 2014. Entrainment of neural oscillations as a modifiable substrate of attention. Trends Cogn. Sci. 18, 300–309. http://dx.doi.org/10.1016/j.tics.2014.02.005.

Davis, M.H., Johnsrude, I.S., 2003. Hierarchical processing in spoken language comprehension. J. Neurosci. 23, 3423–3431.

Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods 134, 9–21. http://dx.doi.org/10.1016/j.jneumeth.2003.10.009.

DeWitt, I., Rauschecker, J.P., 2012. Phoneme and word recognition in the auditory ventral stream. Proc. Natl. Acad. Sci. U. S. A. 109, E505–E514. http://dx.doi.org/10.1073/pnas.1113427109.

Ding, N., Chatterjee, M., Simon, J.Z., 2013. Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. NeuroImage 88C, 41–46. http://dx.doi.org/10.1016/j.neuroimage.2013.10.054.

Ding, N., Simon, J.Z., 2012. Emergence of neural encoding of auditory objects while listening to competing speakers. Proc. Natl. Acad. Sci. U. S. A. 109, 11854–11859. http://dx.doi.org/10.1073/pnas.1205381109.

Ding, N., Simon, J.Z., 2014. Cortical entrainment to continuous speech: functional roles and interpretations. Front. Hum. Neurosci. 8, 311. http://dx.doi.org/10.3389/fnhum.2014.00311.

Doelling, K.B., Arnal, L.H., 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. http://dx.doi.org/10.1016/j.neuroimage.2013.06.035.

Fontolan, L., Morillon, B., Liegeois-Chauvel, C., Giraud, A.-L., 2014. The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex. Nat. Commun. 5, 4694. http://dx.doi.org/10.1038/ncomms5694.
Galambos, R., Makeig, S., Talmachoff, P.J., 1981. A 40-Hz auditory potential recorded from the human scalp. Proc. Natl. Acad. Sci. U. S. A. 78, 2643–2647.
Ghitza, O., 2011. Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. Front. Psychol. 2, 130. http://dx.doi.org/10.3389/fpsyg.2011.00130.
Ghitza, O., 2012. On the role of theta-driven syllabic parsing in decoding speech: intelligibility of speech with a manipulated modulation spectrum. Front. Psychol. 3, 238. http://dx.doi.org/10.3389/fpsyg.2012.00238.
Ghitza, O., 2013. The theta-syllable: a unit of speech information defined by cortical function. Front. Psychol. 4, 138. http://dx.doi.org/10.3389/fpsyg.2013.00138.
Ghitza, O., 2014. Behavioral evidence for the role of cortical θ oscillations in determining auditory channel capacity for speech. Front. Psychol. 5, 652. http://dx.doi.org/10.3389/fpsyg.2014.00652.
Giraud, A.-L., Poeppel, D., 2012. Cortical oscillations and speech processing: emerging computational principles and operations. Nat. Neurosci. 15, 511–517. http://dx.doi.org/10.1038/nn.3063.
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. http://dx.doi.org/10.1371/journal.pbio.1001752.
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, 20095–20100. http://dx.doi.org/10.1073/pnas.1213390109.
Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. Nat. Rev. Neurosci. 8, 393–402. http://dx.doi.org/10.1038/nrn2113.
Horton, C., D'Zmura, M., Srinivasan, R., 2013. Suppression of competing speech through entrainment of cortical oscillations. J. Neurophysiol. 109, 3082–3093. http://dx.doi.org/10.1152/jn.01026.2012.
Howard, M.F., Poeppel, D., 2010. Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. J. Neurophysiol. 104, 2500–2511. http://dx.doi.org/10.1152/jn.00251.2010.
İlhan, B., VanRullen, R., 2012. No counterpart of visual perceptual echoes in the auditory system. PLoS One 7, e49287. http://dx.doi.org/10.1371/journal.pone.0049287.
Kerlin, J.R., Shahin, A.J., Miller, L.M., 2010. Attentional gain control of ongoing cortical speech representations in a "cocktail party.". J. Neurosci. 30, 620–628. http://dx.doi.org/10.1523/JNEUROSCI.3631-09.2010.
Krumbholz, K., Patterson, R.D., Seither-Preisler, A., Lammertmann, C., Lütkenhöner, B., 2003. Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. Cereb. Cortex 13, 765–772.
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., Shah, A.S., Knuth, K.H., Ulbert, I., Karmos, G., Schroeder, C.E., 2005. An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. J. Neurophysiol. 94, 1904–1911. http://dx.doi.org/10.1152/jn.00263.2005.
Lakatos, P., Musacchia, G., O'Connel, M.N., Falchier, A.Y., Javitt, D.C., Schroeder, C.E., 2013. The spectrotemporal filter mechanism of auditory selective attention. Neuron 77, 750–761. http://dx.doi.org/10.1016/j.neuron.2012.11.034.
Lalor, E.C., Power, A.J., Reilly, R.B., Foxe, J.J., 2009. Resolving precise temporal processing properties of the auditory system using continuous stimuli. J. Neurophysiol. 102, 349–359. http://dx.doi.org/10.1152/jn.90896.2008.
Luo, H., Poeppel, D., 2007. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron 54, 1001–1010. http://dx.doi.org/10.1016/j.neuron.2007.06.004.
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. http://dx.doi.org/10.3389/fpsyg.2012.00170.
Luo, H., Tian, X., Song, K., Zhou, K., Poeppel, D., 2013. Neural response phase tracks how listeners learn new acoustic representations. Curr. Biol. 23, 968–974. http://dx.doi.org/10.1016/j.cub.2013.04.031.
McMullan, A.R., Hambrook, D.A., Tata, M.S., 2013. Brain dynamics encode the spectrotemporal boundaries of auditory objects. Hear. Res. 304, 77–90. http://dx.doi.org/10.1016/j.heares.2013.06.009.
Mesgarani, N., Cheung, C., Johnson, K., Chang, E.F., 2014. Phonetic feature encoding in human superior temporal gyrus. Science 343, 1006–1010. http://dx.doi.org/10.1126/science.1245994.
Millman, R.E., Johnson, S.R., Prendergast, G., 2014. The role of phase-locking to the temporal envelope of speech in auditory perception and speech intelligibility. J. Cogn. Neurosci. 1–13. http://dx.doi.org/10.1162/jocn_a_00719.
Ng, B.S.W., Schroeder, T., Kayser, C., 2012. A precluding but not ensuring role of entrained low-frequency oscillations for auditory perception. J. Neurosci. 32, 12268–12276. http://dx.doi.org/10.1523/JNEUROSCI.1877-12.2012.
Nozaradan, S., 2014. Exploring how musical rhythm entrains brain activity with electroencephalogram frequency-tagging. Philos. Trans. R. Soc. Lond. B Biol. Sci. 369, 20130393. http://dx.doi.org/10.1098/rstb.2013.0393.
O'Connell, M.N., Falchier, A., McGinnis, T., Schroeder, C.E., Lakatos, P., 2011. Dual mechanism of neuronal ensemble inhibition in primary auditory cortex. Neuron 69, 805–817. http://dx.doi.org/10.1016/j.neuron.2011.01.012.
Peelle, J.E., Davis, M.H., 2012. Neural oscillations carry speech rhythm through to comprehension. Front. Psychol. 3, 320. http://dx.doi.org/10.3389/fpsyg.2012.00320.
Peelle, J.E., Gross, J., Davis, M.H., 2013. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. Cereb. Cortex 23, 1378–1387. http://dx.doi.org/10.1093/cercor/bhs118.
Poeppel, D., 2003. The analysis of speech in different temporal integration windows: cerebral lateralization as "asymmetric sampling in time.". Speech Comm. 41, 245–255. http://dx.doi.org/10.1016/S0167-6393(02)00107-3.
Poeppel, D., Emmorey, K., Hickok, G., Pylkkänen, L., 2012. Towards a new neurobiology of language. J. Neurosci. 32, 14125–14131. http://dx.doi.org/10.1523/JNEUROSCI.3244-12.2012.
Schroeder, C.E., Lakatos, P., 2009. Low-frequency neuronal oscillations as instruments of sensory selection. Trends Neurosci. 32, 9–18. http://dx.doi.org/10.1016/j.tins.2008.09.012.
Schroeder, C.E., Wilson, D.A., Radman, T., Scharfman, H., Lakatos, P., 2010. Dynamics of active sensing and perceptual selection. Curr. Opin. Neurobiol. 20, 172–176. http://dx.doi.org/10.1016/j.conb.2010.02.010.
Scott, S.K., Blank, C.C., Rosen, S., Wise, R.J., 2000. Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123 (Pt 12), 2400–2406.
Sheffert, S.M., Pisoni, D.B., Fellowes, J.M., Remez, R.E., 2002. Learning to recognize talkers from natural, sinewave, and reversed speech samples. J. Exp. Psychol. Hum. Percept. Perform. 28, 1447–1469.
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. http://dx.doi.org/10.1523/JNEUROSCI.0703-10.2010.
Steinschneider, M., Nourski, K.V., Rhone, A.E., Kawasaki, H., Oya, H., Howard, M.A., 2014. Differential activation of human core, non-core and auditory-related cortex during speech categorization tasks as revealed by intracranial recordings. Front. Neurosci. 8, 240. http://dx.doi.org/10.3389/fnins.2014.00240.
Uppenkamp, S., Johnsrude, I.S., Norris, D., Marslen-Wilson, W., Patterson, R.D., 2006. Locating the initial stages of speech-sound processing in human temporal cortex. NeuroImage 31, 1284–1296.
VanRullen, R., Macdonald, J.S.P., 2012. Perceptual echoes at 10 Hz in the human brain. Curr. Biol. 22, 995–999. http://dx.doi.org/10.1016/j.cub.2012.03.050.
Vanrullen, R., McLelland, D., 2013. What goes up must come down: EEG phase modulates auditory perception in both directions. Front. Psychol. 4, 16. http://dx.doi.org/10.3389/fpsyg.2013.00016.
VanRullen, R., Zoefel, B., Ilhan, B., 2014. On the cyclic nature of perception in vision versus audition. Philos. Trans. R. Soc. Lond. B Biol. Sci. 369, 20130214. http://dx.doi.org/10.1098/rstb.2013.0214.
Weisz, N., Wühle, A., Monittola, G., Demarchi, G., Frey, J., Popov, T., Braun, C., 2014. Prestimulus oscillatory power and connectivity patterns predispose conscious somatosensory perception. Proc. Natl. Acad. Sci. U. S. A. 111, E417–E425. http://dx.doi.org/10.1073/pnas.1317267111.
Zion Golumbic, E.M., Ding, N., Bickel, S., Lakatos, P., Schevon, C.A., McKhann, G.M., Goodman, R.R., Emerson, R., Mehta, A.D., Simon, J.Z., Poeppel, D., Schroeder, C.E., 2013. Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party.". Neuron 77, 980–991. http://dx.doi.org/10.1016/j.neuron.2012.12.037.
Zoefel, B., Heil, P., 2013. Detection of near-threshold sounds is independent of EEG phase in common frequency bands. Front. Psychol. 4, 262. http://dx.doi.org/10.3389/fpsyg.2013.00262.
Zoefel, B., VanRullen, R., 2015. Selective perceptual phase entrainment to speech rhythm in the absence of spectral energy fluctuations. J. Neurosci. 35, 1954–1964. http://dx.doi.org/10.1523/JNEUROSCI.3484-14.2015.

CHAPTER 6: CHARACTERIZATION OF PHASE ENTRAINMENT TO LOW- AND HIGH-LEVEL FEATURES OF SPEECH SOUND IN LAMINAR RECORDINGS IN MONKEY A1

In the previous chapter, it has been shown that EEG oscillations can entrain to high-level features of speech sound. However, as explained in the General Introduction, the spatial resolution of the EEG is relatively poor. The auditory system is organized tonotopically and oscillatory processes seem to depend on the cortical layer they are operating in, and on the interplay between layers. Thus, the precise mechanisms of the reported "high-level entrainment" might differ between cortical sites or even cortical layers, but this cannot be determined using EEG. A similar reasoning can be applied for the processing of everyday speech sound, which is extensively studies using superficial recordings, but not intracranially. For the following article, both everyday speech sound and speech/noise stimuli were presented to a monkey while neural oscillations were recorded in primary auditory cortex (A1). Phase entrainment has been shown repeatedly in monkey A1 (e.g., Lakatos et al., 2008). Together with the close genetic relation between monkeys and humans, the characterization of phase entrainment to rhythmic stimulation (including speech) in non-human primates is therefore a valid and valuable tool. Moreover, this setup made it possible to characterize neural mechanisms of phase entrainment with respect to the laminar profile in auditory cortex.

During the experiment described in the following article, one monkey listened passively to both everyday sound and constructed speech/noise stimuli while intracortical (CSD)

oscillations and MUA activity (reflecting neuronal firing) were recorded. In line with our findings for EEG recordings in humans (Chapter 5), we found that neural activity in the primary auditory cortex of the monkey significantly entrained to stimuli in all conditions, including their time-reversed versions. Again, it was rather the entrained phase and not the degree of entrainment that changed in the absence of systematic fluctuations in low-level features of speech. This finding demonstrates that not only humans, but also monkeys can entrain to high-level features of speech sound, and indicates that acoustic, but not necessarily linguistic features of speech sound play an important role. Finally, as we will see in the following article, neural oscillations entrain to speech sound in a clever way that processes the relevant information but prepares the system to the upcoming sound at the same time. This is possible because, due to its intrinsic structure, low- and high-frequency components of speech alternate. Thus, whenever low sound frequencies prevail in the input, tonotopic regions in A1 dominantly processing sound with these frequencies can reset other regions, responding primarily to high frequencies, to their low-excitability phase. This phase reset automatically results in a convergence of the high-excitability phase in these regions and their preferred stimulation (i.e. the dominance of higher sound frequencies in the input). This input then initiates the same mechanism again: This time, regions processing lower sound frequencies are set to their low-excitability phase. A similar mechanism has been shown before for non-speech stimuli (O'Connell et al., 2011) and discussed theoretically for human speech (O'Connell et al., submitted), but has never been demonstrated experimentally for human speech.

Thus, the results described in this chapter argue for phase entrainment being a general mechanism for an efficient processing of rhythmic stimulation across species, potentially

centering “snapshots” (or perceptual cycles) on relevant stimuli or even preparing them for upcoming, important events.

Article:

Zoefel B, Costa-Faidella J, Lakatos P, Schroeder CE, VanRullen R (in preparation) Characterization of phase entrainment to low- and high-level features of speech sound in laminar recordings in monkey A1.

Characterization of phase entrainment to low- and high-level features of speech sound in laminar recordings in monkey A1

Authors: Benedikt Zoefel[a,b,c*], Jordi Costa-Faidella[c,d,e], Peter Lakatos[c], Charles E. Schroeder[c], and Rufin VanRullen[a,b]

Affiliations: [a] Université Paul Sabatier, Toulouse, France

[b] Centre de Recherche Cerveau et Cognition (CerCo), CNRS, UMR5549, Pavillon Baudot CHU Purpan, BP 25202, 31052 Toulouse Cedex, France

[c] Nathan Kline Institute for Psychiatric Research, Orangeburg, NY, United States

[d] Institute for Brain, Cognition and Behavior (IR3C), University of Barcelona, Catalonia, Spain

[e] Cognitive Neuroscience Research Group, Department of Psychiatry and Clinical Psychobiology, University of Barcelona, Catalonia, Spain

*Corresponding author: Benedikt Zoefel
Centre de Recherche Cerveau et Cognition (CerCo)
Pavillon Baudot CHU Purpan, BP 25202
31052 Toulouse Cedex
France

Phone: +33 562 746 131
Fax: +33 562 172 809
Email: zoefel@cerco.ups-tlse.fr

Number of pages: 40
Number of figures: 5
Number of words:
Abstract: 368
Introduction: 859
Discussion: 2711

Key words: Neural oscillations, phase, entrainment, speech, A1, monkey, phase-amplitude coupling

Running title: Phase entrainment to speech in monkey A1

Acknowledgements: This study was supported by a Studienstiftung des deutschen Volkes (German National Academic Foundation) scholarship to BZ, a Marie Curie International Outgoing Fellowship within the 7th European Community Framework Programme (FP7-PEOPLE-2012-IOF; PIOF-GA-2012-331251) to JCF, NIH R01DC012947 to PL, NIH DC011490 to CES, and an ERC Consolidator grant P-CYCLES under grant agreement 614244 to RV.

Conflict of Interest: The authors declare no competing financial interests.

Abstract

Neural oscillations in the auditory system entrain to rhythmic stimulation, thereby aligning their phase of high excitability with relevant input. Of particular importance is the entrainment to speech sound: Correlated with speech comprehension, phase entrainment is omnipresent in current theories of speech processing. Nevertheless, the underlying neural mechanisms, in particular with respect to processing in cortical layers, are still largely unknown. This fact is possibly due to the (at first view, justified) assumption that entrainment to human speech can only be investigated in humans – and laminar recordings in human subjects are rare and problematic. However, rhythmic communication calls are not unique to humans, and phase entrainment has repeatedly been demonstrated in non-human primates. Thus, it is possible that laminar recordings in these species provide us with important insight into neural mechanisms underlying phase entrainment to speech sound. We presented one monkey with everyday speech sound and recorded neural (as current-source density, CSD) oscillations in different areas of primary auditory cortex (A1). We observed phase entrainment to the rhythm of speech in all areas; however, only those spectral components of speech corresponding to the "best frequency" (BF) of the recording site entrained neural oscillations to their high-excitability phase, whereas other components evoked an entrainment to the opposite, low-excitability phase. Together with the fact that low- and high-frequency components in speech alternate, our findings confirm previous speculation that phase entrainment reflects a particularly efficient way of stimulus processing that includes the preparation of the relevant neuronal populations to the upcoming input. Moreover, presenting speech/noise sound without systematic fluctuations in low-level features of speech and the time-reversed version of all stimuli, we found significant phase entrainment in all conditions and all cortical layers. The entrainment in the

speech/noise conditions was characterized by a change in the entrained phase (compared to everyday speech), and this phase was dominantly coupled to activity in a lower gamma-band (in contrast to coupling to a gamma-band of higher frequency in response to everyday speech). These results show that (1) phase entrainment in A1 includes a high-level process with specific characteristics, (2) this process is not unique to humans and (3) potentially destroying "acoustic edges" by time-reversal of speech does not diminish phase entrainment, as previously argued.

Introduction

Many stimuli in the auditory environment – such as speech sound – are rhythmic and alternate between important and less relevant events. Brain activity is rhythmic as well: Neural oscillations reflect changes between high and low excitability phases of neuronal populations (Buzsáki and Draguhn, 2004). It has been proposed that these oscillations can be seen as alternations between open and closed "windows of opportunity" for input to be processed (Jensen et al., 2012). Thus, it is a reasonable assumption that the auditory system tries to align these two rhythms – brain oscillations and rhythmic input – a phenomenon called phase entrainment (Lakatos et al., 2008; Calderone et al., 2014). Indeed, it has been shown that phase entrainment to speech sound can improve speech comprehension (Ahissar et al., 2001; Luo and Poeppel, 2007; Park et al., 2015).

Nevertheless, phase entrainment of neural oscillations to speech sound is a mechanism whose characterization is still incomplete: For instance, besides very few studies (Nourski et al., 2009; Fontolan et al., 2014), reports of phase entrainment are usually based on electrophysiological recordings with low spatial resolution (using, e.g., electroencephalogram, EEG). The characterization of phase entrainment to speech sound

with respect to laminar processing in auditory cortex would thus represent a step forward in our understanding of the brain's processing of speech sound and, more generally, of rhythmic input. For this purpose, recordings in monkey auditory cortex are an important tool (Rauschecker and Scott, 2009): First, monkeys and humans are genetically closely related. Second, in contrast to humans, laminar profiles can be easily recorded in monkeys. Third, phase entrainment of neural oscillations has been demonstrated repeatedly in monkeys (Lakatos et al., 2005b, 2008, 2013a), indicating that humans and monkeys share a common mechanism of adaptation to rhythmicity.

We are aware of only one other study showing neural responses in monkey A1 during the presentation of human speech (Steinschneider et al., 2013). However, this study was limited to the presentation of words and focused on neural responses to phonemes. In the current study, we presented long (five one-minute) sequences of everyday speech sound and measured the entrainment of neural oscillations in the different cortical layers of A1. A1 is tonotopically organized (Merzenich and Brugge, 1973) – we were therefore interested in how oscillations entrain to such a spectrally complex stimulus as speech sound. Two important points should be made here: First, it has been shown – using pure tones as stimuli – that only oscillations in the area of A1 tuned to the frequency of the stimulus ("best frequency", BF, region) entrain their high-excitability phase to the stimulus – oscillations in the rest of auditory cortex are set to their low-excitability phase (O'Connell et al., 2011; Lakatos et al., 2013a). Second, certain spectral components – possibly vowels and some consonants, such as the fricative /s/ – seem to alternate in speech sound. It has recently been suggested that the counterphase entrainment described above could reflect an efficient way of stimulus processing: If one region, responsible for the processing of vowels, sets the oscillations in another region, responsible for the processing of fricatives, to their

low-excitability phase, their high-excitability phase "arrives in time" for the upcoming fricatives to be processed (O'Connell et al., submitted). For speech sound, this clever mechanism has so far only been hypothesized, but experimental data are still lacking – we thus analyzed our data based on this notion.

Although everyday speech sound is an interesting stimulus, it contains large fluctuations in amplitude and spectral content (Fig. 1A, top; Fig. 1B, left) – any observed entrainment can thus be "biased", as it cannot be ruled out that it entails a passive "following" of these low-level fluctuations at very early levels of auditory processing (e.g., a "ringing" by the cochlea; VanRullen et al., 2014). Recently, we reported the construction of speech/noise stimuli without systematic fluctuations in amplitude and spectral content, but with conserved high-level features, including phonetic information[1] (Fig. 1A, bottom; Fig. 1B, right; Zoefel and VanRullen, 2015). Moreover, we reported that EEG oscillations entrain to this "high-level rhythm", but that the entrainment does not depend on linguistic features, as it is not disrupted when the speech/noise sound is reversed (Zoefel and VanRullen, in press). Based on this finding, it is possible that we can find "high-level entrainment" in monkey A1 as well. Thus, in addition, we presented our speech/noise stimuli and measured the entrainment of neural oscillations to them. We were thus able to quantify phase entrainment to speech sound, both in response to concomitant low- and high-level features (i.e. to everyday speech sound) and to isolated high-level features (i.e. to our constructed speech/noise stimuli).

[1] The definition of low- and high-level features of speech sound is often vague. Here, we follow Zoefel and VanRullen (2015) in that we define low-level features as those that are equated (across envelope phases) in our stimuli: amplitude and spectral content. At the same time, these features are those processed at a very early level of auditory processing: at the cochlear level. Thus, the remaining features, here defined as high-level features and including, but not restricted to phonetic information, are processed at a level that can at least be assumed to be located beyond the earliest level of the auditory hierarchy.

Materials and Methods

Subjects

In the present study, we analyzed the electrophysiological data recorded during seven penetrations of area A1 of the auditory cortex of one female rhesus macaque weighing ~9 kg, who had been prepared surgically for chronic awake electrophysiological recordings. Before surgery, the animal was adapted to a custom-fitted primate chair and to the recording chamber. All procedures were approved in advance by the Animal Care and Use Committee of the Nathan Kline Institute.

Surgery

Preparation of subjects for chronic awake intracortical recording was performed using aseptic techniques, under general anesthesia, as described previously (Schroeder et al., 1998). The tissue overlying the calvarium was resected and appropriate portions of the cranium were removed. The neocortex and overlying dura were left intact. To provide access to the brain and to promote an orderly pattern of sampling across the surface of the auditory areas, plastic recording chambers (Crist Instrument) were positioned normal to the cortical surface of the superior temporal plane for orthogonal penetration of area A1, as determined by preimplant MRI. Together with socketed Plexiglas bars (to permit painless head restraint), they were secured to the skull with orthopedic screws and embedded in dental acrylic. A recovery time of 6 weeks was allowed before we began data collection.

Electrophysiology

During the experiments, the animal sat in a primate chair in a dark, isolated, electrically shielded, sound-attenuated chamber with head fixed in position, and was monitored with

infrared cameras. Neuroelectric activity was obtained using linear array multicontact electrodes (23 contacts, 100 um intercontact spacing, Plexon). The multielectrodes were inserted acutely through guide tube grid inserts, lowered through the dura into the brain, and positioned such that the electrode channels would span all layers of the cortex, which was determined by inspecting the laminar response profile to binaural broadband noise bursts. Neuroelectric signals were impedance matched with a preamplifier (10X gain, bandpass dc 10 kHz) situated on the electrode, and after further amplification (500X) they were recorded continuously with a 0.01–8000 Hz bandpass digitized with a sampling rate of 20 kHz and precision of 16 bits using custom-made software in Labview. The signal was split into the local field potential (LFP; 0.1–300 Hz) and multiunit activity (MUA; 300–5000 Hz) range by zero phase shift digital filtering. Signals were downsampled to 2000 Hz and LFP data were notch-filtered between 59 and 61 Hz to remove electrical noise. MUA data were also rectified to improve the estimation of firing of the local neuronal ensemble (Legatt et al., 1980). One-dimensional CSD profiles were calculated from LFP profiles using a three-point formula for the calculation of the second spatial derivative of voltage (Freeman and Nicholson, 1975). The advantage of CSD profiles is that they are not affected by volume conduction like the LFP, and they also provide a more direct index of the location, direction, and density of the net transmembrane current flow (Mitzdorf, 1985; Schroeder et al., 1998). At the beginning of each experimental session, after refining the electrode position in the neocortex, we established the “best frequency” (BF) of the recording site using a “suprathreshold” method (Steinschneider et al., 1995; Lakatos et al., 2005a). The method entails presentation of a stimulus train consisting of 100 random order occurrences of a broadband noise burst and pure tone stimuli with frequencies ranging from 353.5 Hz to 16 kHz in half-octave steps (duration: 100 ms, r/f time: 5 ms; inter-stimulus interval, ISI: 624.5

ms). Auditory stimuli were produced using Matlab in-house scripts and delivered via Experiment Builder Software (SR Research Ltd., Mississauga, Ontario, Canada) at 50dB SPL coupled with MF-1 free-field speakers.

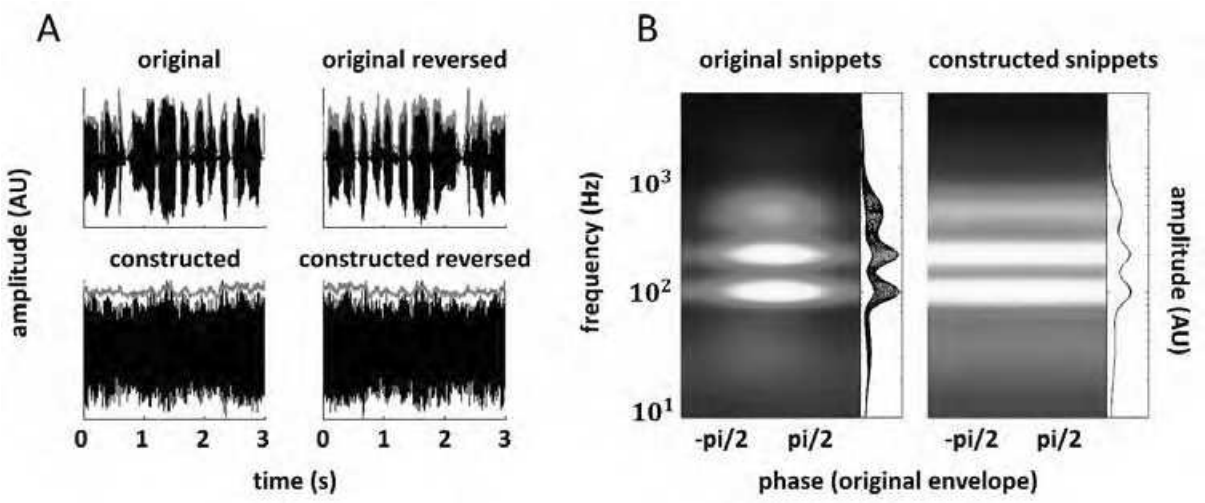


Figure 1. A. Three-second excerpts from the stimuli used for the four conditions in this study. For all conditions, the actual stimulus is shown in black, and the corresponding envelope in gray. In the first condition ("original"), everyday speech sound was presented. These stimuli were time-reversed for the second condition ("original reversed"). Speech/noise stimuli were used for the other two conditions – again, played forwards ("constructed" condition) and backwards (''constructed reversed" condition). The reasoning underlying the construction of these stimuli was described in detail in Zoefel and VanRullen (2015). In short, as shown in B, in everyday speech sound (left panel), the different phases of the speech envelope strongly differ in their spectral energy (color-coded in B) and might trivially entrain early levels (e.g., the cochlea) of the auditory system. We therefore constructed speech/noise stimuli (right panel) which do not exhibit these fluctuations in low-level properties – amplitude and spectral content – anymore. Entrainment to these stimuli thus requires a rather high-level process (located beyond the cochlear level), as speech and noise have to be distinguished based on their high-level properties (i.e. other features than amplitude or spectral content).

Experimental paradigm

Stimuli of four experimental conditions (Fig. 1A) were presented to the monkey – all of them were based on five one-minute snippets of an English speaker reading parts of a classic novel. In the first condition ("original"), these everyday speech sound snippets were presented. In the second condition ("original reversed"), these snippets were presented in reverse, a common procedure to test the influence of intelligibility on the observed results (e.g., Gross et al., 2013; Park et al., 2015). Although this test is not necessarily meaningful in our case of non-human primates, we were nonetheless interested in the outcome: The issue

has been raised that the reversal of speech *per se* might destroy "acoustic edges", leading to a decline in phase entrainment that has nothing to do with the decrease in intelligibility (Peelle and Davis, 2012; Millman et al., 2015). If this assumption were true, we should see a decrease in entrainment in this condition as well. In the third condition ("constructed"), we presented speech/noise sound in which systematic low-level fluctuations (i.e. fluctuations in amplitude and spectral content; those are shown in Fig. 1B, left) were removed. The detailed construction of these stimuli was described in Zoefel and VanRullen (2015). In short, the original speech sound was mixed with noise tailored to counterbalance fluctuations in amplitude and spectral content, so that – on average – the different phases of the speech envelope do not differ in these low-level properties anymore (Fig. 1B, right). Intelligibility – and therefore high-level features of speech, including phonetic information – was preserved; phase entrainment to these stimuli is thus possible, assuming that the distinct high-level features of speech and noise (which alternate rhythmically by construction) can be distinguished. Phase entrainment in this condition would thus indicate the involvement of a high-level process: Neural oscillations in monkey A1 could adjust their phase to high-level features of speech, a finding that has recently been reported for human subjects (Zoefel and VanRullen, 2015, in press). It is important to note that these remaining high-level fluctuations in the *constructed* speech/noise snippets are reflected in the *original* speech envelope (as in everyday speech, low- and high-level features co-vary). Finally, for completeness, in the fourth condition ("constructed reversed"), these constructed speech/noise snippets were presented in reverse. For each recording, the five one-minute snippets of all four conditions (i.e. 20 snippets) were presented in random order. The monkey listened passively to those stimuli. Snippets were separated by silence of 10 s.

Data analyses

All analyses were done separately for supragranular, granular, and infragranular layers. For supragranular and infragranular layers, two channels (one sink and one source) were chosen based on the laminar profile obtained in the "suprathreshold method". The latter normally results in clear sinks and sources that can readily be assigned to these layers (e.g., Lakatos et al., 2013a). Only one channel – the sink – was chosen for the granular layer, as the corresponding source is sometimes difficult to define. For each trial, CSD and MUA signals were baseline corrected by subtracting the average of the 1-s window recorded before the beginning of the respective trial. Moreover, the first and last 500 ms of each trial were rejected in order to avoid contamination by neural responses evoked by the onset or offset of the trial, respectively.

Phase entrainment: cross-correlation

Phase entrainment can be defined as an alignment between two oscillatory signals (here: speech or speech/noise sound and recorded signal). Note that, in order to investigate this alignment, the *original* speech envelope was used for all conditions – but it reflects fluctuations in both low- and high-level features of speech in the two original conditions ("original" and "original reversed"), and fluctuations in high-level features only in the two constructed conditions (as explained above; "constructed" and "constructed reversed"). In order to test phase entrainment to these features, the *broadband* speech envelope was used, and phase entrainment to this envelope was compared across conditions, as outlined below. However, an additional aim of this study was to investigate the entrainment of different tonotopically organized regions to the various spectral components of speech. Therefore, for the original condition (only), phase entrainment to the envelope of different *filtered* versions of the speech signal was also analyzed. As already mentioned above, we

expected differences in entrainment (e.g., in the phase relation between speech envelope and recorded signal) between regions whose BF correspond to different spectral components of speech (e.g., to vowels, whose spectral content is restricted to rather low frequencies, and to certain consonants, such as the fricative /s/, whose spectral content is restricted to rather high frequencies).

For this purpose, we calculated the cross-correlation between speech envelope and CSD/MUA signal, computed for time lags between -1 and 1s (Lalor et al., 2009; VanRullen and Macdonald, 2012):

$$cross-correlation\ (ch,t) = \sum_{T} env(T).signal(ch,T+t)$$

where *env(T)* and *signal(T)* denote the standardized (z-scored) speech envelope and the corresponding standardized (z-scored) CSD or MUA response at time T and channel *ch*, respectively, and *t* denotes the time lag between envelope and recorded signal. Cross-correlations were averaged across trials and recordings, but separately for each layer. Note that cross-correlation can also be used to determine the coherence between envelopes of different frequency bands of speech, if *signal* is replaced by *env2* in the above formula, and *env* and *env2* both represent envelopes of the speech signal, filtered into different frequency bands (cf. Fig. 4).

Phase entrainment: phase-dependent responses

We were able to use cross-correlation only to determine phase entrainment in response to the original speech sound: Although cross-correlation is an important tool for the estimation of phase entrainment, it was necessary to develop an additional analysis. This is because, during the construction of our speech/noise stimuli, amplitude and spectral content were

matched across binned phases of the original speech envelope. Cross-correlation analysis does not use phase bins, and might therefore be "contaminated" by spectral differences even in the case of the two constructed conditions ("constructed" and "constructed reversed"). We thus designed the following analysis: The original speech envelope (downsampled to 2000 Hz to match the sampling rate of the CSD/MUA signal and filtered between 2 and 8 Hz, the dominant frequency range of its spectrum) was divided into the same phase bins (i.e. 12 phase bins between –pi and pi) that were used for stimulus construction (Zoefel and VanRullen, 2015). Each data point of the speech envelope corresponded to one data point of CSD/MUA signal that was recorded at the same time: Thus, we were able to calculate the average CSD/MUA amplitude as a function of phase (bin) of the original speech envelope. The CSD signal was filtered between 2 and 8 Hz before this procedure in order to match the dominant frequency of the original speech envelope (a prerequisite for phase entrainment). If there were phase entrainment, the CSD/MUA signal should be influenced by the speech envelope – logically thus, it should fluctuate as a function of original envelope phase. In order to test this, we fitted a sine wave to the CSD/MUA amplitude as a function of envelope phase: The amplitude of this sine wave reflects the strength of entrainment and the phase of this sine wave reflects the phase relation between entrained signal and original speech envelope. Two different hypotheses could be tested based on the thus extracted amplitudes values: (1) whether there is significant overall phase entrainment in the different conditions and layers and (2) whether there are significant differences in phase entrainment across conditions or layers. For (1), significance of phase entrainment was tested by a permutation procedure. Here, the analysis was repeated 1,000,000 times, but speech envelope and recorded signal were drawn from different trials. Thus, it was possible to obtain a range (i.e. a surrogate distribution) of (sine

wave) amplitude values under the null hypothesis of no phase entrainment between original speech envelope and recorded signal. P-values were obtained for the recorded data by comparing “real” amplitude values (averaged across recordings and either across layers – to test entrainment in the different conditions – or across conditions – to test entrainment in the different layers) with the surrogate distribution (averaged likewise). P-values were corrected for multiple comparisons using FDR (Benjamini and Hochberg, 1995). Note that the obtained amplitude values are necessarily positive; therefore, even under the null hypothesis of no phase entrainment, amplitude values above 0 are likely. This results in the fact that a simple t-test (e.g., test whether the obtained amplitudes are significantly different from 0) would not represent a valid approach to test the null hypothesis and speak in favor of the permutation procedure as a more appropriate test for phase entrainment. For (2), the obtained amplitude values of the fitted sine waves were subjected to a two-factorial ANOVA (main factors condition and layer). We were also interested in whether the entrained phases, reflected by the phases of the fitted sine waves, differ across conditions. This question was analyzed on a single-trial level and the fitted phases from the original condition were used as a reference: For each trial, layer and recording, the circular difference between fitted phases in the original condition and any other condition was determined. Note that trials are independent from each other: Any possible combination of trials could thus be compared (e.g., trial 1 of the original condition with trial 1 of the constructed condition, trial 1 of the original condition and trial 2 of the constructed condition etc.). For the original condition, phases of different trials could be compared as well – but, of course, phases of the same trials were excluded from the comparison. Phase differences were determined separately for each layer and recording and then pooled across layers and recordings, leading to 7 (recordings) * 5 (channels) * 5*5 (trial combinations) = 875 phase differences for

original reversed, constructed and constructed reversed condition, and to 7 (recordings) * 5 (channels) * 5*4 (trial combination) = 700 phase differences for the original condition. For the original condition, this distribution of phase differences only serves control purposes and provides an estimation of the variability of the entrained phase across trials. For the other conditions, if there were a difference in entrained phase compared to the original condition, the mean phase difference would be different from 0. We analyzed these phase differences by means of two statistical tests. First, using Rayleigh's Test, we tested whether phase differences are non-uniformly distributed – only in this case it would be appropriate to interpret the circular mean of the respective distribution of phase differences. Second, using a circular test equivalent to Student's t-test with specified mean direction, we tested whether the circular mean of the respective distribution of phase differences is significantly different from 0. A significant value in this test would indicate a difference in entrained phase between the respective condition and the original condition.

Phase-amplitude coupling (PAC)

It has often been argued that the coupling of the phase of neural oscillations at frequencies corresponding to the dominant frequency range of the speech envelope (~2-8 Hz) with the amplitude of oscillations at higher frequencies (e.g., in the gamma-range, ~25-120 Hz) might play an important role for the "tracking" of speech sound, and rather theoretical argumentations (Poeppel, 2003; Ghitza, 2011, 2012; Giraud and Poeppel, 2012; Hyafil et al., 2015) have been underlined by practical findings in response to non-speech (Luo and Poeppel, 2012) and speech stimuli (Gross et al., 2013; Zion Golumbic et al., 2013; Fontolan et al., 2014; Park et al., 2015). As the specific roles of low-and high-level cues for phase-amplitude coupling (PAC) remained unclear, we were interested in differences in the

observed coupling for our experimental conditions. Moreover, it has been shown convincingly that artificial PAC can be produced for different reasons, such as imperfect sinusoidal shapes of the recorded signal (Dvorak and Fenton, 2014; Aru et al., 2015). However, this concern might be reduced somewhat if a coupling between phase and amplitude of different layers can be demonstrated, as it has been done before (Spaak et al., 2012). Thus, as a modified version of the phase-locking value proposed in Lachaux et al. (1999), we calculated PAC as follows:

$$PAC(l1, l2) = \frac{1}{R}\sum_{r=1}^{R} \left| \frac{1}{N}\sum_{n=1}^{N} \frac{1}{T}\sum_{t=1}^{T} e^{i\left(\varphi_{env\_csd}(t,n,r,l1) - \varphi_{csd}(t,n,r,l2)\right)} \right|$$

where T is the number of time points, N is the number of trials, R is the number of recordings, and l1 and l2 are the layers from which phase and amplitude were extracted (for the sake of simplification, only sinks were used for this analysis). PAC ranges between 0 (no coupling) and 1 (maximal coupling). φ_{csd} corresponds to the phase of the CSD signal, filtered between 2 and 8 Hz. φ_{env_csd} corresponds to the phase of the Hilbert envelope (filtered likewise) of the CSD signal, filtered at different gamma frequency bands. For the latter, center frequencies between 25 and 118 Hz were used, with the width of the band increasing with increased center frequency (smallest bandwidth 10 Hz, largest bandwidth 42 Hz). Note that by averaging across time and trials in the complex domain, PAC is only high when the absolute phase difference between gamma envelope and the slower CSD oscillation does not vary across trials. PAC values could then be averaged across layers or conditions (for comparison across conditions, PAC values were also averaged across the two original and the two constructed conditions, respectively; see Results). Again, two different hypotheses could be tested based on these PAC values: (1) whether there is significant overall PAC in the

different conditions and layers and (2) whether there are significant differences in PAC across conditions or layers. For (1), as described above for the sine fit amplitude values, we tested significance of the obtained PAC by assigning φ_{csd} and φ_{env_csd} of a given trial to different simulated trials and re-calculating PAC 1,000,000 times. This was done in order to obtain a range of PAC values under the null hypothesis of no PAC across layers or conditions. P-values were calculated for the recorded data by comparing "real" PAC values with the surrogate distribution. P-values were corrected for multiple comparisons using FDR. For (2), the obtained PAC values were subjected to a two-factorial ANOVA (main factors condition and layer). Here, in order to test potential interactions between main factors, for each gamma-frequency, the PAC difference between original and constructed conditions was tested for significance. This was done using another permutation test: A surrogate distribution was constructed for which the assignment "condition" was assigned randomly and, for each gamma-frequency, the PAC difference between ("pseudo"-)original and ("pseudo"-)constructed conditions was re-calculated 1,000,000 times. This was done in order to obtain a range of PAC difference values under the null hypothesis of no PAC difference between conditions for any given gamma-frequency. P-values were calculated for the recorded data by comparing "real" PAC difference values with the surrogate distribution. P-values were corrected for multiple comparisons using FDR.

All analyses were performed in MATLAB, using the toolbox for circular statistics (Berens, 2009) where appropriate.

Results

In this study, we characterized phase entrainment to human speech sound as observed in laminar recordings in primary auditory cortex (A1) of the monkey. As phase entrainment to

speech sound is an important phenomenon that is ubiquitous in current theories of speech comprehension (e.g., Ghitza, 2012; Giraud and Poeppel, 2012) but the detailed mechanisms with respect to processing in the different cortical layers are largely unknown, this study thus represents an important step forward towards our understanding of the brain's adaptation to complex rhythmic stimuli such as speech. We presented one monkey (seven recordings) with four different experimental categories (Fig. 1A; five one-minute snippets per condition in each recording): Original (everyday) speech sound ("original" condition), speech/noise sound, for which the original speech was mixed with noise to counterbalance low-level fluctuations in amplitude and spectral content ("constructed" condition; Fig. 1B; Zoefel and VanRullen, 2015), and the time-reversed version of both conditions ("original reversed" and "constructed reversed" condition). Note that both low- and high-level features (the former are defined as amplitude and spectral content in this study, the latter as the remaining features, including phonetic information; for a detailed definition, see Zoefel and VanRullen, 2015) exhibit systematic fluctuations in the two original conditions. In contrast, only high-level features systematically vary in the two constructed conditions. Moreover, all these fluctuations are reflected in the original speech envelope (see Materials and Methods; Zoefel and VanRullen, 2015). The layer- and frequency-specific adjustment (i.e. phase entrainment) to sound in these four conditions is described in the following.

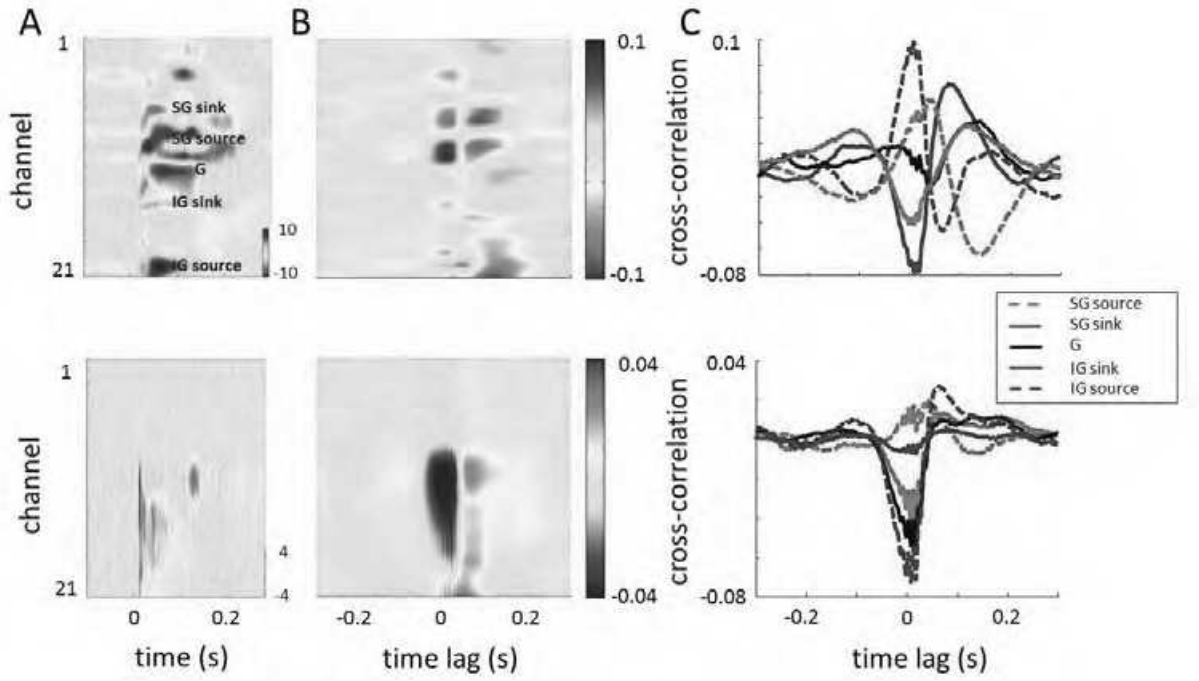


Figure 2. Layer-specific entrainment of CSD (top) and MUA (bottom) signals to the broadband envelope of speech sound in one exemplary recording (original condition). Using cross-correlation between speech envelope and recorded signals, a clear laminar pattern of entrainment can be revealed (B): The time lag and polarity of entrainment depends on the layer in which the signal is recorded (A). Choosing layers based on the laminar profile obtained in response to pure tones at the BF of the recording site (A) results in cross-correlations fluctuating at ~5 Hz (C, top), reflecting entrainment to the speech envelope (whose dominant frequency is in the same range). This phase entrainment goes along with a change in neural firing (reflected in MUA), as visible in C (bottom). Note that, for the MUA and to a smaller degree also for the CSD, during peaks and troughs of the cross-correlations, ripples at higher frequencies are visible that might reflect the phase-amplitude coupling of neural oscillations that is often described in the literature (cf. Fig. 6; e.g., Lakatos et al., 2005b). SG: supragranular, G: granular, IG: infragranular.

Phase entrainment to speech sound in monkey A1

We used cross-correlation to characterize the alignment between everyday speech sound and the recorded CSD and MUA signals. Exemplary results from one recording in the original condition are shown in Fig. 2B: Prominent peaks of cross-correlation at different time lags and polarity are visible. Importantly, these peaks correspond to the different cortical layers (Fig. 2A), indicating that they entrain to the speech sound with specific delays and polarity. This finding can be seen again in Fig. 2C, in which cross-correlation signals are shown for the different layers as chosen based on the laminar profile (obtained in response to pure tones; see Materials and Methods) in Fig. 2A. For the CSD (Fig. 2, top), in all layers, one positive and one negative peak of cross-correlation can be seen. Rather trivially, the polarity of these

peaks is opposite when comparing sinks and sources of current flow. Interestingly though, one complete "cycle" of cross-correlation (i.e. one positive peak followed by one negative peak or vice versa) has a duration of ~200 ms, which corresponds to one cycle of the speech envelope. This result reflects entrainment of the CSD signal to the envelope of speech sound (and contradicts the notion of the observed cross-correlation being a mere reflection of evoked responses; see Discussion), as expected. Phase entrainment can be seen in all layers, but is strongest in supragranular layers. Importantly, the entrainment goes along with a change in neuronal firing (reflected in MUA; Fig. 2, bottom): A strong decrease, followed by an increase, in neuronal firing can be seen in response to speech sound, and this effect is strongest in granular and extragranular layers. Note, however, that this pattern can be reversed, depending on the BF of the recording site, as we will see in the following paragraph.

It has been proposed before that phase entrainment is a remarkably efficient tool for the processing of speech sound (O'Connell et al., submitted), and this suggestion was based on two complementary facts: First, as shown by the use of pure tones (O'Connell et al., 2011; Lakatos et al., 2013a), regions tonotopically tuned to the frequency of the stimulus (BF regions) entrain the high-excitability phase of their oscillations to stimulus onset, but set oscillations in other (non-BF) regions to their low-excitability phase. Second, it is possible that phonemes with different spectral content (e.g., vowels and fricatives) alternate in speech sound. Combining these facts, non-BF stimuli for a given region would automatically prepare the oscillations for the arrival of BF stimuli half a cycle later – and vice versa. As this idea was presented mostly theoretically (O'Connell et al., submitted), we tried to show experimental evidence in this study. First, it has to be shown that speech sound (and not only pure tones) with a given spectral content can set brain oscillations to their high- and

low-excitability phase, depending on their BF. We thus filtered the speech signal into different frequency bands and cross-correlated their envelope and the recorded (unfiltered) signal in different tonotopical regions (only for the original condition). Cross-correlation results were compared (subtracted) between regions tuned to frequencies typical for the spectral content of vowels[2] (<= 1000 Hz; 4 recordings) and those tuned to frequencies associated with the spectral content of consonants (e.g., fricatives, such as /s/; >= 8000 Hz; 2 recordings). In Fig. 3, these differences are shown as a function of the frequency band of speech used for cross-correlation analysis. As it can be seen, there are pronounced differences in the phase (or "polarity", color-coded) of CSD entrainment (Fig. 3A) which, as expected, depend both on the stimulus frequency and on the polarity of the layer of interest (i.e. sink vs. source). We made sure that these differences resulted from opposite patterns (and not from a similar pattern that is more pronounced in one case) between sites of different BF (not shown). Regardless of layer, the entrainment to stimulus frequencies corresponding to the BF of the recording site always goes along with an increase in MUA (assumedly reflecting neuronal firing) at a time lag close to 0, and a decrease otherwise (Fig. 3B). This finding indicates that, indeed, oscillations entrain their high-excitability phase (associated with an increase in MUA) to speech sound only when the spectral content matches their BF, and their low-excitability phase otherwise. For sinks, the high-excitability phase corresponds to the trough of a CSD oscillation and the low-excitability phase corresponds to the peak, as it can be extracted from Fig. 3 (e.g., an increase in MUA, red in Fig. 3B, goes along with the oscillatory trough, blue in Fig. 3A). Second, it also has to be shown that phonemes of different spectral content alternate in speech sound. We therefore

[2] We acknowledge here that our spectral classification of vowels and consonants is oversimplified and probably insufficient for phonological purposes. However, we argue that it is sufficient for a demonstration of frequency-specific entrainment to speech sound.

filtered the speech signal around its fundamental frequency (here ~100 Hz) and cross-correlated its envelope with the envelopes of the speech signal filtered into other frequency bands. Results are shown in Fig. 4: Importantly, a negative cross-correlation at time lag 0 can be seen for all frequency bands above ~3000 Hz which becomes positive at a time lag of approximately ± one cycle of the broadband speech envelope (~±200 ms). This finding indicates that phonemes with lower frequencies (e.g., vowels) and phonemes with higher frequencies alternate in speech sound. Combining results shown in Figs. 3 and 4, these findings demonstrate what was already speculated by O'Connell et al. (submitted): Phase entrainment functions as a tool for an efficient processing of speech sound (Giraud and Poeppel, 2012), and can be considered as a spectrotemporal filter (Lakatos et al., 2013a).

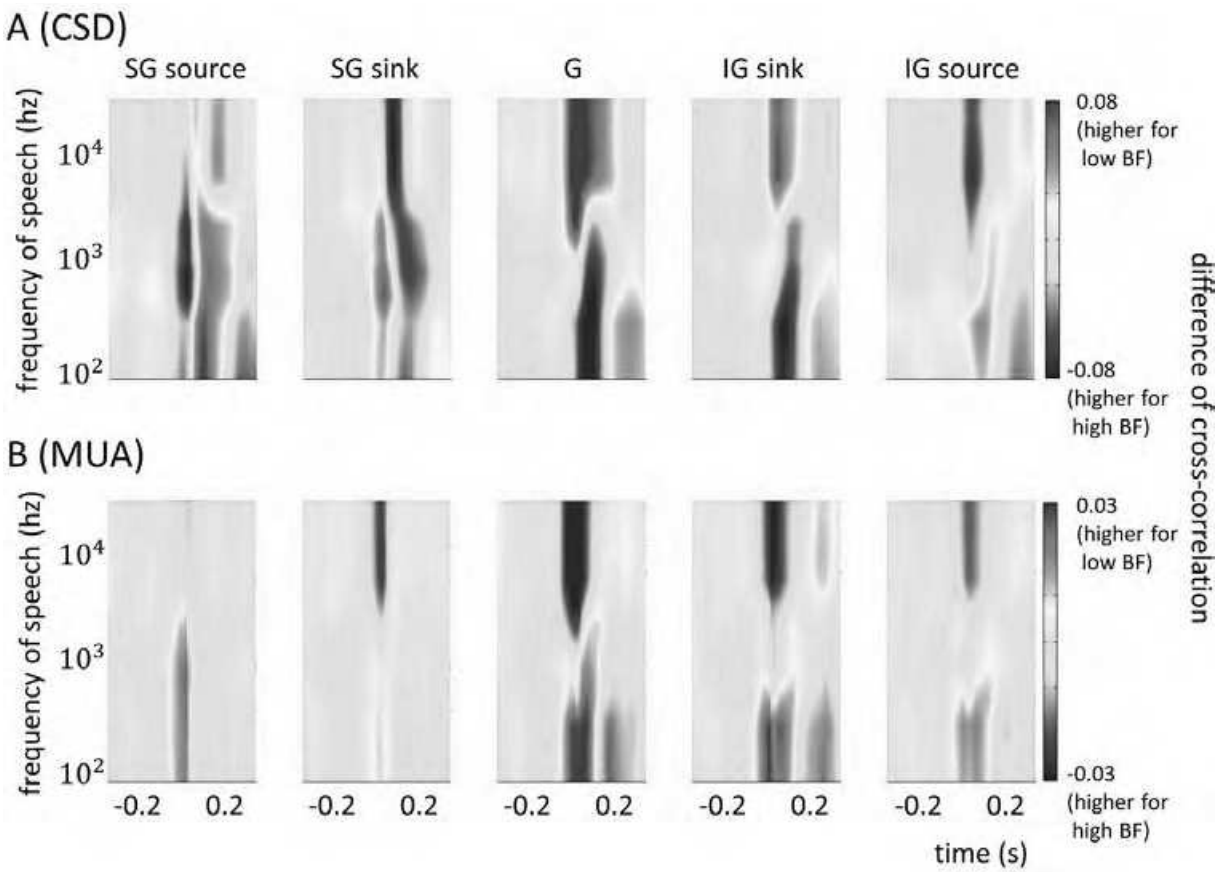


Figure 3. Difference of cross-correlation (recording sites tuned to frequencies <= 1000 Hz minus recording sites tuned to frequencies >= 8000 Hz) between CSD (A) or MUA (B) and the envelope of speech sound filtered into narrow frequency bands (original condition). Cross-correlation patterns – as those shown in Fig. 2 – are visible again. However, for CSD, their "polarity" (i.e. the entrained phase, reflected by the sign of cross-correlation) depends on both the BF of the recording site and on the polarity of the layer (i.e. sink vs. source). Regardless of the latter, entrainment in response to sound frequencies

corresponding to the BF/non-BF of the recording site always results in an increase/decrease (respectively) of MUA with a time lag close to 0. SG: supragranular, G: granular, IG: infragranular.

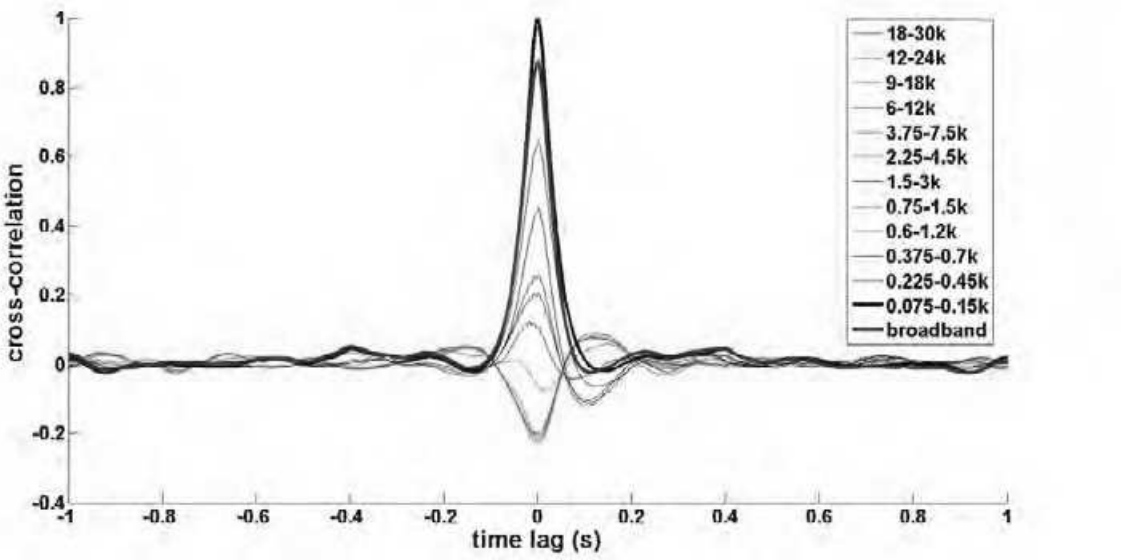


Figure 4. The original speech sound, used in this study, was filtered into different frequency bands and the envelope of these bands was computed. The envelope of speech sound filtered around its fundamental frequency (here 100 Hz; thick black line) was used as a reference and its coherence (using cross-correlation) with the envelopes of other frequency bands is shown here. At time lag 0, this cross-correlation is positive for most frequency bands of relatively low frequency (up to 2.25 kHz center frequency; black lines), including the broadband envelope (thick blue line). Importantly, this changes for higher frequencies, which are (at this time lag) negatively correlated with the envelope of speech sound around 100 Hz (red lines). Note also that these correlations change their sign at a time lag of ~±one cycle of the broadband speech envelope (~±200 ms). This finding indicates that low- and high-frequency components in speech (e.g., vowels and certain fricatives) alternate.

Phase entrainment to low- and high-level features of speech sound in monkey A1

We developed an additional analysis in order to compare phase entrainment across the four experimental conditions and the different layers, respectively (see Materials and Methods). In short, CSD (filtered between 2 and 8 Hz in order to match the dominant frequency in the speech envelope) and MUA amplitudes were binned as a function of original envelope phase (reflecting low- *and* high-level features of speech in original and original reversed condition, but only high-level features in constructed and constructed reversed condition). A modulation of CSD and/or MUA amplitude by envelope phase would indicate phase entrainment, and the amplitude of a sine wave fitted to this function (CSD/MUA amplitude as a function of original envelope phase) represents the strength of entrainment. Moreover,

this amplitude can be averaged across layers (in order to compare entrainment across conditions) or across conditions (in order to compare entrainment across layers) and the outcome is shown in Fig. 5. First, we tested whether there is significant overall phase entrainment in the different conditions and/or layers. This was done by means of a surrogate procedure. Here, the observed sine fit amplitude values were compared with a surrogate distribution, the latter obtained by repeating the extraction of those values multiple times, but with speech envelope and recorded signal drawn from different trials. For both CSD and MUA, we found that the amplitude values for all conditions (Fig. 5A) and layers (Fig. 5B) are significantly ($p < 10^{-6}$) different from the surrogate distributions (the significance thresholds for alpha = 0.05, corrected for multiple comparisons, are shown as dashed lines); thus, there was significant phase entrainment for all conditions and layers, indicating that neural oscillations entrain to both low- and high-level features of speech sound. Second, we used a two-factorial ANOVA in order to compare phase entrainment across conditions and layers, respectively. For CSD, this test revealed a main effect of condition ($F(3) = 3.37$, $p = 0.021$), but no main effect of layer ($F(4) = 1.49$, $p = 0.211$) and no interaction ($F(12) = 0.65$, $p = 0.795$). Post-hoc tests resulted in a significant difference in CSD entrainment between original and constructed condition, as expected (indeed, in the original condition the sound entails both low- and high-level features that could potentially entrain oscillations, whereas the constructed speech/noise sound only entails high-level features). For MUA, the ANOVA revealed a main effect of layer ($F(4) = 3.45$, $p = 0.010$), but no main effect of condition ($F(3) = 1.82$, $p = 0.147$) and no interaction ($F(12) = 0.49$, $p = 0.918$). Post-hoc tests resulted in a significant difference in MUA entrainment between granular layer and supragranular sink as well as infragranular source, as expected, due to the highest firing rates in the granular layer. Finally, for the CSD, we compared the phases of the

fitted sine waves (reflecting the phase relation between entrained CSD signal and original speech envelope) across conditions. This was done on a single-trial level by using phases from the original condition as a reference (see Materials and Methods) – in Fig. 5C, circular histograms show the observed phase differences between entrained phases in the original condition and entrained phases in any experimental condition (including those obtained in different trials for the original condition). All distributions of phase differences in Fig. 5C are significantly non-uniform (Rayleigh's Test; $p < 10^{-5}$); the mean direction of these distributions can thus be interpreted. As shown in Fig. 5C, phases were very consistent across trials in the original condition: The mean circular difference between phases from different trials is not significantly different from 0 and the confidence interval of the mean phase difference includes 0 ([-0.11 – 0.11]; shown in blue in Fig. 5C). Although the mean circular phase difference between original and original reversed condition is significantly different from 0, it is nevertheless obvious that phases are very similar between these conditions: The confidence interval of the mean phase difference, with the lower limit very close to 0 ([0.06 – 0.29]), suggests that they entrained to a very similar phase and that the observed phase difference might functionally not be relevant. In contrast, and interestingly, for both constructed conditions, there is a phase shift with respect to the original condition, indicating that the removal of systematic fluctuations in low-level features of speech also results in a change of the entrained phase of neural oscillations to the speech sound (confidence interval of mean phase difference, constructed condition: [-1.36 - -0.48], constructed reversed condition: [-2.14 - -1.28].

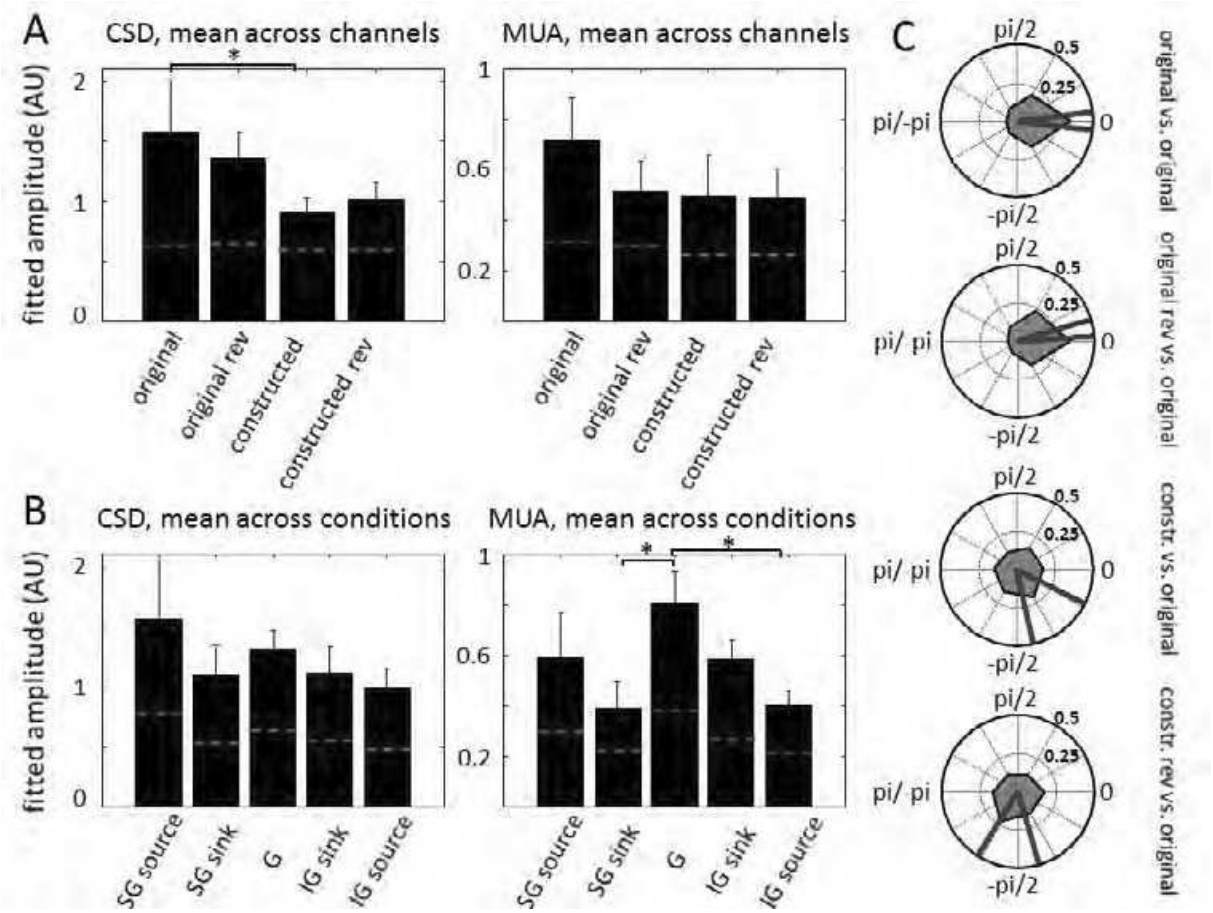


Figure 5. Amplitude values of sine waves fitted to the CSD or MUA amplitude as a function of envelope phase (separately for each recording) – reflecting the strength of the measured phase entrainment – and averaged across recordings, layers (A) or conditions (B), respectively. All amplitude values shown here are significantly larger than those obtained in a surrogate distribution, reflecting phase entrainment in all conditions and layers (significance thresholds are shown as dashed lines). CSD entrainment significantly differs between original and constructed condition, and MUA entrainment significantly differs between granular and supragranular sink as well as infragranular source. Standard error of mean is shown as errorbars. The circular differences between phases of the fitted sine waves are shown in C. They are shown using the original condition as a reference; circular histograms thus show, on a single-trial level (see Materials and Methods), phase distributions of differences in the entrained phase between original and any experimental condition (differences in the entrained phase of different trials in the case of original vs. original condition, topmost plot). As it can be seen, fitted (i.e. entrained) phases are very similar between different trials in the original condition and between the two original conditions. However, there is a pronounced phase difference between original and the two constructed conditions. Phase differences falling into a given phase bin are shown in percent; the confidence interval of the mean phase difference is shown as blue lines. SG: supragranular, G: granular, IG: infragranular.

Phase-amplitude coupling in response to low- and high-level features of speech sound in monkey A1

The coupling between the phase of relatively slow oscillations and the amplitude of faster ones has often been described (PAC; e.g., Lakatos et al., 2005b; Fell and Axmacher, 2011; Spaak et al., 2012; Lisman and Jensen, 2013), and a particular importance of this mechanism has been suggested for speech sound (Poeppel, 2003; Giraud and Poeppel, 2012; Hyafil et

al., 2015), as the frequency of both slower (delta/theta; ~2-8 Hz) and faster (low gamma; e.g., 25-40 Hz) oscillations seems to match the intrinsic structure of speech (e.g., the rate of syllabic and phonetic fluctuations, respectively). However, PAC in response to speech/noise sound without systematic fluctuations in low-level features has not been tested yet; also, results for PAC in monkey A1 in response to everyday speech sound are still lacking. Thus, we quantified PAC in our study (see Materials and Methods; only sinks have been used here) and compared results across conditions (Fig. 6A; here, the two original and the two constructed conditions have been averaged, respectively) or across combination of layers (Fig. 6B; i.e. PAC was tested with the phase taken from one layer and the amplitude taken from another). For these analyses, the CSD phase was computed in a band between 2 and 8 Hz, and the band for which the amplitude was extracted was varied between ~25 and 120 Hz ("gamma-frequencies"). For all conditions and gamma-frequencies, we found significant PAC (the significance threshold, corrected for multiple comparisons, is shown as dotted line in the respective color in Fig. 6A). However, even more interesting, for the original conditions, we found the strongest PAC between delta/theta phase (i.e. the phase that is entrained in this study; Fig. 5) and a band with a center frequency of ~90 Hz (Fig. 6A). This frequency ("high gamma") is likely to be related more or less directly to neuronal firing (Lachaux et al., 2012). For the constructed conditions, and in contrast to the original conditions, a strong coupling between delta/theta phase and a band with a center frequency of ~30 Hz can be seen, the latter close to the suggested rate of phonemes, and in line with current models of neural oscillations involved in speech processing (Poeppel, 2003; Ghitza, 2011; Giraud and Poeppel, 2012; Hyafil et al., 2015). A two-factorial ANOVA on these PAC values resulted in a main effect of gamma-frequency ($F(40) = 43.86$, $p < 0.0001$) but also in a significant interaction between condition and gamma-frequency ($F(40) = 19.54$, $p < 0.0001$). In order to

test this interaction, we determined, using a permutation test (see Materials and Methods), for which gamma-frequency there is a significant PAC difference between experimental conditions. We found a significantly higher PAC for the constructed than for the original conditions for gamma-frequencies in a range of ~31-35 Hz (area shaded red in Fig. 6A). When examining the laminar distribution of the observed PAC, similar peaks of PAC (at ~90 Hz and ~30 Hz) are visible when the coupling between supragranular or infragranular phase and granular amplitude are concerned (thick lines in Fig. 6B). For other combinations of layers, no clear peaks in PAC are visible (thin lines). These findings are well in line with our assumption of the 90-Hz peak reflecting neuronal firing (as firing is strongest in granular layers) as well as with the existing literature, showing the most prominent activity of slower oscillations in supra- and infragranular layers, and a dominance of gamma oscillations in the granular layer (Lakatos et al., 2005b; O'Connell et al., 2014). Note that, for both 90-Hz and 30-Hz peak, PAC is strongest for the combination of supragranular phase and granular amplitude. Thus, we selected this combination of layers; its PAC is also shown across conditions in Fig. 6A (dashed lines). A two-factorial ANOVA on PAC values for this combination of layers resulted in both a main effect of condition ($F(1) = 15.06$, $p = 0.0001$), a main effect of gamma-frequency ($F(40) = 49.86$, $p < 0.001$), and in a significant interaction between condition and gamma-frequency ($F(40) = 19.89$, $p < 0.001$). By means of another permutation test, we found a significantly higher PAC for the constructed than for the original conditions for gamma-frequencies in a range of ~31-35 Hz and ~115-118 Hz (areas shaded blue in Fig. 6A). Again, these results suggest a prominent role for PAC between delta/theta phase and a gamma frequency that might correspond to phonetic processing (~30 Hz) in the constructed conditions, whereas, although non-significant, PAC between

delta/theta phase and higher gamma (~90 Hz), potentially reflecting neuronal firing, seems to dominate in the original conditions.

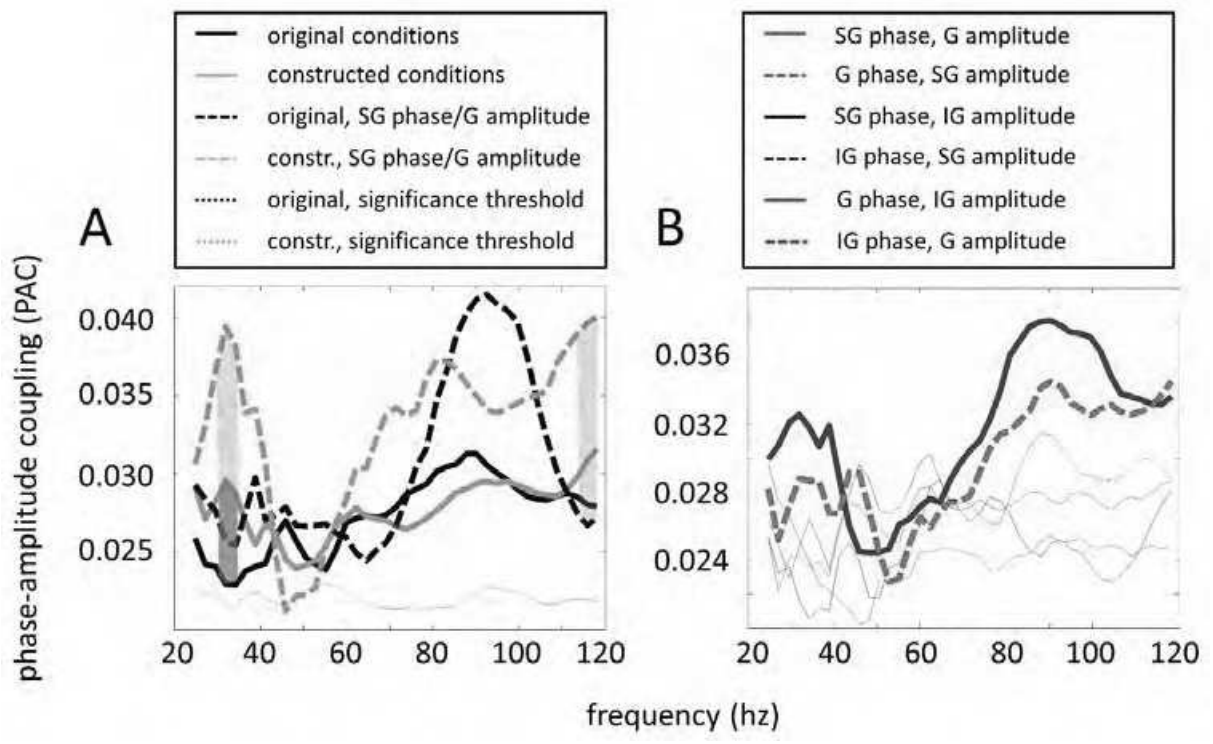


Figure 6. Phase-amplitude coupling (PAC) between the phase of slower CSD oscillations (delta/theta; filtered between 2 and 8 Hz) and the amplitude of oscillations at higher ("gamma") frequencies. Phase and amplitude are always taken from different layers (and only from sinks; resulting in six combinations of layers). A. Average PAC across layer combinations (and across the two original and constructed conditions, respectively) are shown as continuous lines. Significant PAC is found for all conditions and gamma-frequencies, the significance threshold (alpha = 0.05), corrected for multiple comparisons, is shown as dotted lines. The phase of delta/theta oscillations seems to be coupled most strongly to a frequency band around 90 Hz for the original conditions, and to a frequency band around 30 Hz for the constructed conditions. This effect is even more pronounced when a specific combination of layers (supragranular phase, granular amplitude) is chosen (dashed lines). Significant differences in PAC between conditions are shown as shaded areas. B. Average PAC across conditions. The two peaks (30 and 90 Hz) visible in A can be observed again, most strongly for PAC between supragranular (SG) phase and granular (G) amplitude as well as infragranular (IG) phase and granular amplitude (both shown as think lines; the remaining combinations are shown as thin lines).

Discussion

Brain activity is essentially rhythmic, a phenomenon that can be observed across species (Buzsáki et al., 2013). Importantly, this rhythm – reflected in neural oscillations – is used by the brain to adapt to rhythms in the environment (e.g., Schroeder and Lakatos, 2009). Within the last decade, there was mounting evidence that this mechanism – called phase

entrainment – is critical for the brain's functionality, as it is disturbed in many pathological conditions, such as schizophrenia or attention deficit hyperactivity disorder (Lakatos et al., 2013b; Calderone et al., 2014). Moreover, phase entrainment to speech sound is correlated with intelligibility (Ahissar et al., 2001; Luo and Poeppel, 2007; Peelle et al., 2013; Park et al., 2015). Based on these findings, it is all the more surprising that the detailed underlying mechanisms – in particular with respect to laminar processing and the tonopically organized structure of auditory cortex – are still unclear. In this study, we tried to fill this gap by presenting one monkey with speech or speech/noise sound entailing either both low- and high-level features of speech (in the form of everyday speech sound) or only high-level features (in the form of speech/noise sound without systematic fluctuations in sound amplitude or spectral content; Zoefel and VanRullen, 2015). In line with our previous study in humans (Zoefel and VanRullen, in press), we found that neural activity in the primary auditory cortex of the monkey significantly entrained to stimuli in all conditions, including their time-reversed versions. This finding has several implications, and they are discussed in the following paragraphs.

We could show that neural oscillations in all layers of the primary auditory cortex of non-human primates can entrain to speech sound even if it does not entail low-level features such as sound amplitude of spectral content. This result is interesting, because it shows that (1) phase entrainment to speech entails a high-level process and (2) that this process is not restricted to human subjects. Moreover, we were able to reveal specific characteristics of this "high-level entrainment": First, we observed a change in the entrained phase as compared to everyday speech sound (Fig. 5C). Interestingly, a similar phase shift has been observed when the same stimuli were presented to human subjects (Zoefel and VanRullen, in press). In the latter study, differences in phase entrainment (as measured with EEG)

between original and constructed condition were not characterized by a change in the degree of entrainment, but rather in a change of the entrained phase. Second, the entrained phase (dominantly in supra- and infragranular layers) was coupled to the amplitude of higher-frequency gamma-oscillations (dominantly in granular layers) in all conditions, a finding that is well in line with a multitude of previous studies (e.g., Lakatos et al., 2005b; Fell and Axmacher, 2011; Spaak et al., 2012; Lisman and Jensen, 2013). However, whereas this phase was coupled most prominently to oscillatory amplitudes around 90 Hz for the original conditions, it was dominantly coupled to amplitudes around 30 Hz for the constructed conditions. It has been shown before that CSD high-gamma activity (~50-150 Hz) is correlated with neuronal firing (and MUA; for a review, see Lachaux et al., 2012). Thus, we assume that the observed 90-Hz effect might reflect the modulation of neuronal activity by the phase of slow neural oscillations, a finding that has been described before (Whittingstall and Logothetis, 2009). It has been argued that theta-gamma coupling of neural oscillations is of particular importance for the processing of speech sound, because the frequency of both theta- and gamma-bands correspond to important characteristics of speech (e.g., to the syllabic and phonetic rhythm, respectively; Poeppel, 2003). In most theoretical models (Poeppel, 2003; Poeppel et al., 2008; Giraud and Poeppel, 2012; Hyafil et al., 2015), the frequency of this gamma-band is usually set to 25-40 Hz (sometimes also called beta; Ghitza, 2011). Thus, it is possible that the phase-amplitude coupling we observed in the constructed conditions is directly related to speech processing, and that it might be enhanced (as compared with the original conditions) due to an improved signal-to-noise ratio in the conditions where systematic fluctuations in low-level features of speech have been removed. Importantly, our findings are in line with two of the few studies examining phase entrainment to speech sound with high spatial resolution (by means of intracranial

recordings in humans; cortical layers could not be resolved in these data): Nourski et al. (2009) showed that high-gamma power (> 70 Hz) entrains to the envelope of speech sound in human auditory cortex and that this entrainment is not abolished when speech intelligibility is disrupted (by time-compression of the sound; see below). Fontolan et al. (2014) demonstrated phase entrainment and phase-amplitude coupling in response to speech sound and that it involves top-down and bottom-up processes at different neural frequencies. Strikingly, and partly in agreement with our assumptions made above (but see next paragraph), they observed peaks at 30 and 90 Hz as well and related them to top-down and bottom-up processes, respectively (cf. their Fig. 3).

These top-down and bottom-up processes lead us to the next point: For the first time, it was possible to investigate the processing of long sequences of human speech with a spatial resolution corresponding to the different cortical layers. In line with previous studies on entrainment of neural oscillations to trains of pure tones (e.g., Lakatos et al., 2005b, 2013a), we found that the CSD in supragranular layers entrains most strongly to the envelope of speech sound, whereas entrainment of MUA to this envelope is most pronounced in granular layers (Fig. 2C, 5B). Moreover, the entrained phase of CSD oscillations in extragranular layers was coupled to faster (gamma) oscillations in granular layers. Importantly, slower oscillations (such as the alpha band in the visual system or the theta band in the auditory system) are often associated with top-down processing in extragranular layers whereas faster oscillations might reflect bottom-up processing in granular layers (van Kerkoerle et al., 2014; Bastos et al., 2015; Bonnefond and Jensen, 2015; Jensen et al., 2015). Phase entrainment reflects predictions about upcoming events (Schroeder and Lakatos, 2009) and therefore top-down processes: Our findings are thus well in line with the current literature, in that they suggest that the entrainment of slower, supragranular CSD

oscillations is an important tool of the brain (in particular of the auditory system; VanRullen et al., 2014) to control (or "gate") feedforward stimulus processing (reflected in gamma activity) via top-down mechanisms. Moreover, as already discussed above, our data suggest a differentiation into low-level (~90 Hz) and high-level (~30 Hz) feedforward processes that differ in their oscillatory characteristics. However, based on the similarity of these spectral components to those obtained in Fontolan et al. (2014), it needs to be determined in future experiments whether our high-level component in the lower gamma range (~30 Hz) really reflects feedforward processing (as suggested by its presence in granular layers and by its coupling to slower oscillations in extragranular layers) or rather a top-down component as in their study.

Of note is the finding in our study that time-reversal of the stimuli (removing linguistic content that the monkey might have recognized due to its contact with humans and human speech) did not significantly decrease the entrainment. Again, this result supports the findings from our recent work on human subjects (Zoefel and VanRullen, in press). We found that reversing our constructed speech/noise stimuli did not disrupt the entrainment of EEG oscillations. Together with the results presented here, both studies reveal acoustic high-level features of speech (i.e. features that are specific to speech sound but are independent of linguistic content) as an important contributor to phase entrainment to speech sound. Nevertheless, it should be mentioned that some studies did report a decrease in phase entrainment when speech sound was rendered unintelligible by time-reversal (Gross et al., 2013; Park et al., 2015) or noise-vocoding (Peelle et al., 2013). One important suggestion was that the time-reversal *per se* might have destroyed acoustic edges, important landmarks for neural entrainment (Doelling et al., 2014), leading to a decrease in both phase entrainment and intelligibility without a direct link between the two (Peelle and Davis, 2012;

Millman et al., 2015). Our study directly argues against this point, as, if this assumption were true, even in monkey A1, a decrease in entrainment should have been observed. Nevertheless, the question why some studies report a correlation between phase entrainment and speech intelligibility, and some do not, is still open and further studies are necessary for its answer.

Steinschneider et al. (2013) recorded evoked neural responses to human speech in A1 of both humans and non-human primates and found similar neural patterns in response to changes in amplitude envelope, voice-onset time or fundamental frequency. Their conclusion is directly related to the one obtained in the current study: Neural processing of human speech in A1 of monkeys is similar to that in humans. Moreover, entrainment of neural activity in primary auditory cortex to animal vocalizations has been reported for several species (Wang et al., 1995; Schnupp et al., 2006; Grimsley et al., 2012). Also, the dominant frequency of the envelope of humans speech (~2-8 Hz) is similar to the rhythm of other animal calls (see, e.g., Fig. 1 in Wang et al., 1995), and A1 of different species (including humans) seems to be tuned to these relatively slow frequencies (Eggermont, 1998; Oshurkova et al., 2008; Edwards and Chang, 2013). Thus, it is likely that the adjustment to human speech, as it has been observed frequently (Ding and Simon, 2012, 2013; Peelle and Davis, 2012; Zion Golumbic et al., 2012, 2013; Gross et al., 2013; Peelle et al., 2013; Doelling et al., 2014; Park et al., 2015), is only one specific occurrence of phase entrainment to rhythmic stimuli (including vocalization calls) as a general mechanism of efficient stimulus processing across species.

An additional aim of this study was a detailed characterization of how the different cortical layers and tonotopically organized regions entrain to such a complex stimulus as speech

sound. We found that all regions of A1 entrain to (everyday) speech sound, independent of their BF – but the latter influences the phase relation between speech envelope and (CSD) oscillation (Fig. 3): Whereas the spectral components of speech sound corresponding to the BF of a given region entrain oscillations to their high-excitability phase, oscillations in other (non-BF) regions are set to their low-excitability phase (as visible in the differences in MUA in Fig. 3B, assumedly reflecting differences in neuronal firing). In line with findings by O'Connell et al. (2011), using pure tones, and theoretical considerations of the same group (O'Connell et al., submitted), these results demonstrates that phase entrainment acts like a spectrotemporal filter that can be easily applied to speech sound: Due to its intrinsic structure, low- and high-frequency components of speech alternate (Fig. 4). Whenever low sound frequencies prevail in the input, tonotopic regions in A1 dominantly processing (speech) sound with these frequencies reset other regions, responding primarily to high frequencies, to their low-excitability phase. This phase reset automatically results in a convergence of the high-excitability phase in these regions and their preferred stimulation (i.e. the dominance of higher sound frequencies in the input). This input then initiates the same mechanism again: This time, regions processing lower sound frequencies are set to their low-excitability phase.

It is currently still debated whether phase entrainment is more than a simple "regular repetition of evoked potentials" (Capilla et al., 2011; Lakatos et al., 2013a; Zoefel and Heil, 2013; Keitel et al., 2014; VanRullen et al., 2014). We acknowledge that, although the removal of systematic fluctuations in amplitude and spectral content of the stimulus is certainly a first step towards an answer on this question, our "high-level stimuli" used here cannot fully resolve this issue. One reason is that different regions in auditory cortices can respond to different auditory features: There are regions that prefer noise to pure tones or

vice versa (Wessinger et al., 2001; Wang et al., 2005). Thus, it cannot be ruled out that the entrainment we observed in response to our speech/noise stimuli is not merely a succession of "high-level" evoked responses. However, two important points should be made here. First, there is evidence for entrainment being more than a reactive process: Entrained oscillations can be observed even after the offset of the entraining stimulus (Lakatos et al., 2013a) and in response to auditory stimuli that are too weak to be perceived (and therefore do not evoke neural responses; Zoefel and Heil, 2013). In our results, cross-correlations revealed a periodicity that was related to the dominant frequency of the speech envelope and extended to negative time lags (i.e., an increment of stimulus amplitude was *preceded* by an increment of CSD signal, reflecting the predictable nature of speech envelopes and the predictive nature of neural entrainment; Fig. 2C, top); both findings would not be expected for a mere reflection of evoked responses. Second, even when assuming a fully reactive mechanism, in principal, the spectrotemporal filter mechanism as described in the previous paragraph would not be rendered unfeasible: It is likely that evoked responses entail a phase-reset of neural oscillations (Makeig et al., 2002; Sauseng et al., 2007). As this phase-reset is the only prerequisite to "prepare" oscillations for the upcoming stimulation (e.g., their "preferred" spectral components of speech), even the absence of an active entrainment mechanism (although unlikely, see first point) would not contradict the mechanism of spectrotemporal filtering as described above.

Of course, it needs to be mentioned that, so far, all results, as described in this study, hold true only for primary auditory cortex. In the auditory system, an extensive amount of stimulus processing is already done subcortically (e.g., Nelken, 2008). Thus, it might be no surprise that certain high-level processes (those necessary to distinguish speech and noise with identical spectral content, a pre-requisite to entrain to the constructed speech/noise

stimuli in our study) already take place in primary cortical regions. Nevertheless, it would be interesting to test entrainment in auditory regions beyond A1 which process more abstract, rather categorical representations of auditory input (Davis and Johnsrude, 2003). Humans and non-human primates share the division of the auditory pathway into two streams (the postero-dorsal "where" stream and the antero-ventral "what" stream; Rauschecker and Scott, 2009; Rauschecker, 2015). In both species, regions specifically processing communication sounds can be found in the "what" stream (Tian et al., 2001; Petkov et al., 2008; Rauschecker and Scott, 2009). Thus, it is possible that our absence of differences between experimental conditions changes in these higher auditory regions. For instance, as auditory regions beyond A1 in both humans and non-human primates often prefer noise bursts to pure tones (Wessinger et al., 2001), it is possible that speech/noise sound has an advantage over everyday speech in these regions. Moreover, as long as both neural oscillations and tonotopical organization are prevalent on a given level of the auditory system, it is reasonable to assume the "spectrotemporal filter" mechanism as described in the previous paragraph. Whereas evidence of neural oscillations in the frequency range corresponding to the dominant rhythm of speech on subcortical levels remains sparse or even absent, the tonotopical organization beyond A1 becomes increasingly complex (e.g., less selective or broader in tuning bandwidth; Moerel et al., 2012). Thus, it might be that A1, as a hub between early and late auditory processing (Nelken, 2008), is indeed the most promising location for phase entrainment as "spectrotemporal filter". Further studies are necessary, systematically comparing mechanisms of phase entrainment to speech sound at different levels of the auditory hierarchy.

We conclude that phase entrainment of neural oscillations to rhythmic input is a highly efficient "filter" mechanism that is preserved across species and helps to reliably process

events of high relevance within a continuous stream of input. This mechanism includes but is not restricted to oscillations in the human brain as "filter" and human speech as the to-be-filtered input. Did phase entrainment of neural oscillations evolve as an adaptation to the rhythmic structure of the auditory environment (including communication sounds) – or was it evolutionary successful to adapt communication sounds to the rhythmic structure of the brain (indeed, oscillations are present even in species that do not exhibit rhythmic calls; Buzsáki et al., 2013)? Although this study cannot answer this question, its exciting answer might help us understand the brain' adjustment to rhythm and, ultimately, the origin of human speech.

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.

Aru J, Aru J, Priesemann V, Wibral M, Lana L, Pipa G, Singer W, Vicente R (2015) Untangling cross-frequency coupling in neuroscience. Curr Opin Neurobiol 31:51–61.

Bastos AM, Vezoli J, Bosman CA, Schoffelen J-M, Oostenveld R, Dowdall JR, De Weerd P, Kennedy H, Fries P (2015) Visual areas exert feedforward and feedback influences through distinct frequency channels. Neuron 85:390–401.

Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B Methodol 57:289–300.

Berens P (2009) CircStat: A Matlab Toolbox for Circular Statistics. J Stat Softw 31:1–31.

Bonnefond M, Jensen O (2015) Gamma activity coupled to alpha phase as a mechanism for top-down controlled gating. PloS One 10:e0128667.

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.

Calderone DJ, Lakatos P, Butler PD, Castellanos FX (2014) Entrainment of neural oscillations as a modifiable substrate of attention. Trends Cogn Sci 18:300–309.

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:e14543.

Davis MH, Johnsrude IS (2003) Hierarchical processing in spoken language comprehension. J Neurosci 23:3423–3431.

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.

Ding N, Simon JZ (2013) Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. J Neurosci 33:5728–5735.

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.

Dvorak D, Fenton AA (2014) Toward a proper estimation of phase-amplitude coupling in neural oscillations. J Neurosci Methods 225:42–56.

Edwards E, Chang EF (2013) Syllabic (∼2-5 Hz) and fluctuation (∼1-10 Hz) ranges in speech and auditory processing. Hear Res 305:113–134.

Eggermont JJ (1998) Representation of spectral and temporal sound features in three cortical fields of the cat. Similarities outweigh differences. J Neurophysiol 80:2743–2764.

Fell J, Axmacher N (2011) The role of phase synchronization in memory processes. Nat Rev Neurosci 12:105–118.

Fontolan L, Morillon B, Liegeois-Chauvel C, Giraud A-L (2014) The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex. Nat Commun 5:4694.

Freeman JA, Nicholson C (1975) Experimental optimization of current source-density technique for anuran cerebellum. J Neurophysiol 38:369–382.

Ghitza O (2011) Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. Front Psychol 2:130.

Ghitza O (2012) On the role of theta-driven syllabic parsing in decoding speech: intelligibility of speech with a manipulated modulation spectrum. Front Psychol 3:238.

Giraud A-L, Poeppel D (2012) Cortical oscillations and speech processing: emerging computational principles and operations. Nat Neurosci 15:511–517.

Grimsley JMS, Shanbhag SJ, Palmer AR, Wallace MN (2012) Processing of communication calls in Guinea pig auditory cortex. PloS One 7:e51646.

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.

Hyafil A, Fontolan L, Kabdebon C, Gutkin B, Giraud A-L (2015) Speech encoding by coupled cortical theta and gamma oscillations. eLife 4.

Jensen O, Bonnefond M, Marshall TR, Tiesinga P (2015) Oscillatory mechanisms of feedforward and feedback visual processing. Trends Neurosci 38:192–194.

Jensen O, Bonnefond M, VanRullen R (2012) An oscillatory mechanism for prioritizing salient unattended stimuli. Trends Cogn Sci 16:200–206.

Keitel C, Quigley C, Ruhnau P (2014) Stimulus-driven brain oscillations in the alpha range: entrainment of intrinsic rhythms or frequency-following response? J Neurosci 34:10137–10140.

Lachaux J-P, Axmacher N, Mormann F, Halgren E, Crone NE (2012) High-frequency neural activity and human cognition: past, present and possible future of intracranial EEG research. Prog Neurobiol 98:279–301.

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, Musacchia G, O'Connel MN, Falchier AY, Javitt DC, Schroeder CE (2013a) The spectrotemporal filter mechanism of auditory selective attention. Neuron 77:750–761.

Lakatos P, Pincze Z, Fu K-MG, Javitt DC, Karmos G, Schroeder CE (2005a) Timing of pure tone and noise-evoked responses in macaque auditory cortex. Neuroreport 16:933–937.

Lakatos P, Schroeder CE, Leitman DI, Javitt DC (2013b) Predictive suppression of cortical excitability and its deficit in schizophrenia. J Neurosci 33:11692–11702.

Lakatos P, Shah AS, Knuth KH, Ulbert I, Karmos G, Schroeder CE (2005b) An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. J Neurophysiol 94:1904–1911.

Lalor EC, Power AJ, Reilly RB, Foxe JJ (2009) Resolving precise temporal processing properties of the auditory system using continuous stimuli. J Neurophysiol 102:349–359.

Legatt AD, Arezzo J, Vaughan HG (1980) Averaged multiple unit activity as an estimate of phasic changes in local neuronal activity: effects of volume-conducted potentials. J Neurosci Methods 2:203–217.

Lisman JE, Jensen O (2013) The θ-γ neural code. Neuron 77:1002–1016.

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.

Makeig S, Westerfield M, Jung TP, Enghoff S, Townsend J, Courchesne E, Sejnowski TJ (2002) Dynamic brain sources of visual evoked responses. Science 295:690–694.

Merzenich MM, Brugge JF (1973) Representation of the cochlear partition of the superior temporal plane of the macaque monkey. Brain Res 50:275–296.

Millman RE, Johnson SR, Prendergast G (2015) The Role of Phase-locking to the Temporal Envelope of Speech in Auditory Perception and Speech Intelligibility. J Cogn Neurosci 27:533–545.

Mitzdorf U (1985) Current source-density method and application in cat cerebral cortex: investigation of evoked potentials and EEG phenomena. Physiol Rev 65:37–100.

Moerel M, De Martino F, Formisano E (2012) Processing of natural sounds in human auditory cortex: tonotopy, spectral tuning, and relation to voice sensitivity. J Neurosci 32:14205–14216.

Nelken I (2008) Processing of complex sounds in the auditory system. Curr Opin Neurobiol 18:413–417.

Nourski KV, Reale RA, Oya H, Kawasaki H, Kovach CK, Chen H, Howard MA 3rd, Brugge JF (2009) Temporal envelope of time-compressed speech represented in the human auditory cortex. J Neurosci 29:15564–15574.

O'Connell MN, Barczak A, Ross D, McGinnis T, Schroeder CE, Lakatos P (submitted) Multi-scale entrainment of coupled neuronal oscillations in primary auditory cortex.

O'Connell MN, Barczak A, Schroeder CE, Lakatos P (2014) Layer specific sharpening of frequency tuning by selective attention in primary auditory cortex. J Neurosci 34:16496–16508.

O'Connell MN, Falchier A, McGinnis T, Schroeder CE, Lakatos P (2011) Dual mechanism of neuronal ensemble inhibition in primary auditory cortex. Neuron 69:805–817.

Oshurkova E, Scheich H, Brosch M (2008) Click train encoding in primary and non-primary auditory cortex of anesthetized macaque monkeys. Neuroscience 153:1289–1299.

Park H, Ince RAA, Schyns PG, Thut G, Gross J (2015) Frontal Top-Down Signals Increase Coupling of Auditory Low-Frequency Oscillations to Continuous Speech in Human Listeners. Curr Biol 25:1649–1653.

Peelle JE, Davis MH (2012) Neural Oscillations Carry Speech Rhythm through to Comprehension. Front Psychol 3:320.

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.

Petkov CI, Kayser C, Steudel T, Whittingstall K, Augath M, Logothetis NK (2008) A voice region in the monkey brain. Nat Neurosci 11:367–374.

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.

Rauschecker JP (2015) Auditory and visual cortex of primates: a comparison of two sensory systems. Eur J Neurosci 41:579–585.

Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat Neurosci 12:718–724.

Sauseng P, Klimesch W, Gruber WR, Hanslmayr S, Freunberger R, Doppelmayr M (2007) Are event-related potential components generated by phase resetting of brain oscillations? A critical discussion. Neuroscience 146:1435–1444.

Schnupp JWH, Hall TM, Kokelaar RF, Ahmed B (2006) Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex. J Neurosci 26:4785–4795.

Schroeder CE, Lakatos P (2009) Low-frequency neuronal oscillations as instruments of sensory selection. Trends Neurosci 32:9–18.

Schroeder CE, Mehta AD, Givre SJ (1998) A spatiotemporal profile of visual system activation revealed by current source density analysis in the awake macaque. Cereb Cortex 8:575–592.

Spaak E, Bonnefond M, Maier A, Leopold DA, Jensen O (2012) Layer-specific entrainment of γ-band neural activity by the α rhythm in monkey visual cortex. Curr Biol 22:2313–2318.

Steinschneider M, Nourski KV, Fishman YI (2013) Representation of speech in human auditory cortex: is it special? Hear Res 305:57–73.

Steinschneider M, Reser D, Schroeder CE, Arezzo JC (1995) Tonotopic organization of responses reflecting stop consonant place of articulation in primary auditory cortex (A1) of the monkey. Brain Res 674:147–152.

Tian B, Reser D, Durham A, Kustov A, Rauschecker JP (2001) Functional specialization in rhesus monkey auditory cortex. Science 292:290–293.

van Kerkoerle T, Self MW, Dagnino B, Gariel-Mathis M-A, Poort J, van der Togt C, Roelfsema PR (2014) Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. Proc Natl Acad Sci U S A 111:14332–14341.

VanRullen R, Macdonald JSP (2012) Perceptual echoes at 10 Hz in the human brain. Curr Biol 22:995–999.

VanRullen R, Zoefel B, Ilhan B (2014) On the cyclic nature of perception in vision versus audition. Philos Trans R Soc Lond B Biol Sci 369:20130214.

Wang X, Lu T, Snider RK, Liang L (2005) Sustained firing in auditory cortex evoked by preferred stimuli. Nature 435:341–346.

Wang X, Merzenich MM, Beitel R, Schreiner CE (1995) Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. J Neurophysiol 74:2685–2706.

Wessinger CM, VanMeter J, Tian B, Van Lare J, Pekar J, Rauschecker JP (2001) Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. J Cogn Neurosci 13:1–7.

Whittingstall K, Logothetis NK (2009) Frequency-band coupling in surface EEG reflects spiking activity in monkey visual cortex. Neuron 64:281–289.

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.

Zion Golumbic EM, Poeppel D, Schroeder CE (2012) Temporal context in speech processing and attentional stream selection: a behavioral and neural perspective. Brain Lang 122:151–161.

Zoefel B, Heil P (2013) Detection of Near-Threshold Sounds is Independent of EEG Phase in Common Frequency Bands. Front Psychol 4:262.

Zoefel B, VanRullen R (in press) EEG oscillations entrain their phase to high-level features of speech sound. NeuroImage.

Zoefel B, VanRullen R (2015) Selective perceptual phase entrainment to speech rhythm in the absence of spectral energy fluctuations. J Neurosci 35:1954–1964.

CHAPTER 7: THE ROLE OF HIGH-LEVEL PROCESSES FOR OSCILLATORY PHASE ENTRAINMENT TO SPEECH SOUND

In the previous chapters, it was demonstrated that phase entrainment of neural oscillations to speech sound entails a high-level component. This finding is important, as it shows that neural oscillations – or perceptual cycles – can be controlled in an active way by the auditory system. In the form of a review article, this chapter summarizes the results obtained in this thesis, showing that a high-level process is involved in the brain's adjustment to speech sound, and embeds them in the current literature.

In this article, we differentiate between (1) high-level modulations of phase entrainment and (2) high-level entrainment: In (1), phase entrainment can be produced as a following response to a low-level rhythmic stimulus sequence (potentially in early brain areas, as early as the cochlea; this notion has been brought up throughout previous chapters, see for instance the Introduction of Chapter 4); however, the entrainment is modulated by high-level processes that include attention or predictions. In contrast to (1), high-level entrainment (2) represents phase entrainment that can be observed even in the absence of systematic fluctuations of low-level properties. In this case, a simple "following" of stimulus amplitude is not possible anymore: Thus, it is the process of phase entrainment itself that operates on a higher level, as a certain level of processing is required in order to adjust to the rhythm of high-level features. Promising results have been obtained for both types of high-level processes, and both are addressed in the following article.

Concerning (1), it has been shown repeatedly that phase entrainment to speech sound in a multi-speaker scenario depends on attention: Only the speech rhythm of the attended (but not that of the unattended) speaker is "tracked" (e.g., Ding and Simon, 2012a; Horton et al., 2013; Zion Golumbic et al., 2013b). Moreover, visual information that usually precedes the auditory input in speech can adjust oscillations in auditory cortex so that they optimally align with the expected speech sound (e.g., Arnal et al., 2009, 2011; Zion Golumbic et al., 2013a; Perrodin et al., 2015). The latter finding shows that not only attention, but also predictions – both high-level processes – are involved in phase entrainment to speech sound.

Most evidence for (2) has been obtained in this thesis: Phase entrainment to speech sound is not only a reflection of fluctuations in low-level features of speech sound, but entails an adjustment to phonetic (high-level) information – and thus a genuine high-level process. This result has been demonstrated on a perceptual (Chapter 4) and a neural level (Chapter 5) and confirmed in intracortical recordings in non-human primates (Chapter 6). The latter result is not discussed in the review article itself, but separately in the same chapter of this thesis. Two complementary studies have been conducted that are also discussed in the review article: Ding et al. (2013) showed that the reduction of spectro-temporal fine structure (which can be considered as high-level features) of speech results in a decline in phase entrainment as compared to that in response to natural speech sound. Rimmele et al. (2015) showed that this effect depends on attention and is therefore modulated by high-level processes.

We conclude this review with a section dedicated to the role of intelligibility for phase entrainment to speech sound, as the influence of semantic information on the brain's adjustment to speech is currently a strongly debated topic. This debate is based on

controversial findings: Some studies show an enhanced phase entrainment of neural oscillations in response to intelligible (as compared to unintelligible) speech sound (e.g., Luo and Poeppel, 2007; Ding et al., 2013; Doelling et al., 2014; Park et al., 2015) whereas others do not (Howard and Poeppel, 2010; Millman et al., 2015; Chapters 5 and 6). Moreover, phase entrainment can be observed even in response to simple stimuli such as pure tones (e.g., Besle et al., 2011; Gomez-Ramirez et al., 2011; Lakatos et al., 2005; Zoefel and Heil, 2013). In the following article, we propose a model that can reconcile these findings: In short, we argue that intelligibility does not directly influence entrainment in temporal (auditory) regions of the brain; rather, it modulates the connectivity between temporal (entrained) and more frontal regions (important for behavioral outcome), synchronizing them in the case of intelligible speech (resulting in a periodic modulation of behavior by the entraining stimulus) and de-synchronizing them in the case of unintelligible speech (resulting in an independence of behavior from the entraining stimulus). Logically, intelligibility does not influence the connectivity between these regions if the input is not identified as speech sound. To sum up, the results summarized and discussed in this review show that the brain might constantly predict the timing of relevant and irrelevant events of speech sound, including high-level features, and actively align neural oscillations so that they efficiently boost the current locus of attention. Linguistic features, reflecting intelligibility, might play a modulatory, and speech-specific, role by determining the behavioral consequences of phase entrainment to speech sound.

Article:

Zoefel B, VanRullen R (submitted) The role of high-level processes for oscillatory phase entrainment to speech sound.

frontiers

The role of high-level processes for oscillatory phase entrainment to speech sound

Benedikt Zoefel[1, 2*], **Rufin VanRullen**[1, 2]

[1]Centre de Recherche Cerveau et Cognition (CerCo), CNRS, France, [2]Université Paul Sabatier, France

Submitted to Journal:
Frontiers in Human Neuroscience

Article type:
Review Article

Manuscript ID:
163585

Received on:
30 Jul 2015

Frontiers website link:
www.frontiersin.org

The role of high-level processes for oscillatory phase entrainment to speech sound

Authors: Benedikt Zoefel[a,b,*] and Rufin VanRullen[a,b]

Affiliations: [a] Université Paul Sabatier, Toulouse, France

[b] Centre de Recherche Cerveau et Cognition (CerCo), CNRS, UMR5549, Pavillon Baudot CHU Purpan, BP 25202, 31052 Toulouse Cedex, France

*Corresponding author: Benedikt Zoefel
Centre de Recherche Cerveau et Cognition (CerCo)
Pavillon Baudot CHU Purpan, BP 25202
31052 Toulouse Cedex
France

Phone: +33 562 746 131
Fax: +33 562 172 809
Email: zoefel@cerco.ups-tlse.fr

Number of pages: 17
Number of figures: 5
Number of words (total): 4902
Number of words (abstract): 223

Key words: EEG, oscillation, phase, entrainment, high-level, speech, auditory, intelligibility

Running title: High-level processes in phase entrainment

Acknowledgements: The authors thank Mitchell Steinschneider, Peter Lakatos, Daniel Pressnitzer, Alain de Cheveigné and Jesko Verhey for helpful comments and dicussions. This study was supported by a Studienstiftung des deutschen Volkes (German National Academic Foundation) scholarship to BZ, and an ERC grant ("P-Cycles", number 614244) to RV.

Conflict of Interest: The authors declare no competing financial interests.

Abstract

Constantly bombarded with input, the brain has the need to filter out relevant information while ignoring the irrelevant rest. A powerful tool may be represented by neural oscillations which entrain their high-excitability phase to important input while their low-excitability phase attenuates irrelevant information. Indeed, the alignment between brain oscillations and speech improves intelligibility and helps dissociating speakers during a "cocktail party". Although well-investigated, the contribution of low- and high-level processes to phase entrainment to speech sound has only recently begun to be understood. Here, we review those findings, and concentrate on three main results: (1) Phase entrainment to speech sound is modulated by attention or predictions, likely supported by top-down signals from other brain regions and indicating higher-level processes involved in the brain's adjustment to speech. (2) As phase entrainment to speech can be observed without systematic fluctuations in sound amplitude or spectral content, it does not only reflect a passive steady-state "ringing" of the cochlea, but entails a higher-level process. (3) The role of intelligibility for phase entrainment is debated. Recent results suggest a modulatory role at the interface between entrainment and behavioral consequences. We conclude that phase entrainment to speech reflects a sophisticated mechanism: Several high-level processes interact to optimally align neural oscillations with predicted events of high relevance, even when they are hidden in a continuous stream of background noise.

1. Phase entrainment as a tool for input gating

In virtually every situation of our life, the brain has to cope with an enormous amount of incoming information, only a fraction of which is essential for the scene's interpretation or resulting behavior. Clearly, the brain must have evolved strategies to deal with this vast influx, and both amplification of relevant input and suppression of irrelevant information will be critical for survival. Based on recent research, one prominent tool for the described purpose are neural oscillations, assumed to reflect cyclic changes in the excitability of groups of neurons (Mazzoni et al., 2010; Buzsáki and Draguhn, 2004; Rajkai et al., 2008). These *endogenous* fluctuations in neural excitability *per se* might seem useless at first glance, as long as they are *passive a*nd unrelated to the environment (Figure 1A). However, as previous studies showed, both on a theoretical (Schroeder et al., 2008; Schroeder and Lakatos, 2009; Schroeder et al., 2010; Ghitza, 2011; Morillon et al., 2015) and experimental level (Lakatos et al., 2005; Besle et al., 2011; Lakatos et al., 2008, 2013; O'Connell et al., 2014; Stefanics et al., 2010; Henry et al., 2014; Henry and Obleser, 2012; Nozaradan, 2014; Arnal et al., 2014; Morillon et al., 2014; Park et al., 2015), these oscillations might become an interesting tool when introducing the possibility that they can be *controlled* by the brain. By using the low and high excitability phases of those oscillations, the brain might *actively* "decide" what part of the incoming information should be amplified (the information coinciding with the oscillation's high excitability phase) and what part should be suppressed (the information coinciding with the oscillation's low excitability phase) (Figure 1B). This phenomenon, the synchronization of an oscillatory system with external input has been termed *phase entrainment* (Calderone et al., 2014). Of course, this kind of "input gating" only makes sense if the input is (1) rhythmic (i.e. predictable), (2) has a frequency that the brain can cope with and is relatively stable and (3) alternates between low and high informational content. Interestingly, one of the most salient stimuli in everyday life fulfills these criteria: speech sound. Although only considered "pseudo-rhythmic" (Cummins, 2012; but see Ghitza, 2013), the frequency of the speech envelope (roughly defined as the sum of energy across sound frequencies at a given point in time; shown as gray line in Figure 1) is relatively stable between 2 and 8 Hz and phases of low phonetic information (e.g., the silence between syllables) rhythmically alternate with phases of high phonetic information.

Indeed, the number of studies reporting an adaptation of neural oscillations to the envelope of speech sound is increasing continuously (Ding and Simon, 2014; Ding et al., 2013; Ding and Simon, 2012b, 2013, 2012a; Zion Golumbic et al., 2013b, 2012; Peelle et al., 2013; Peelle and Davis, 2012; Horton et al., 2013; Steinschneider et al., 2013; Power et al., 2013; Gross et al., 2013; Doelling et al., 2014; Park et al., 2015; Millman et al., 2015). But not only speech sound is able to evoke an entrainment of neural oscillations, even simple stimuli, such as pure tones, have been found to produce phase entrainment (Besle et al., 2011; Gomez-Ramirez et al., 2011; Stefanics et al., 2010; Zoefel and Heil, 2013). Furthermore, rhythmic fluctuations in stimulus amplitude (which are present in both trains of pure tones and speech sound) introduce fluctuations at a level of auditory processing as low as the cochlea, a notion that is obviously not compatible with phase entrainment as an active or "high-level" process. Similar concerns have been raised by several authors in the last years (Obleser et al., 2012; Ding and Simon, 2014; Ding et al., 2013; Peelle et al., 2013; Zion Golumbic et al., 2012). Thus, the role of high-level processes for phase entrainment to speech sound is far from clear. Nevertheless, significant progress has been made within the last decade, and the aim of this review is to summarize the obtained results in a systematic way. The

scope of this review is not a summary of existing literature showing an alignment between brain oscillations and speech sound, as comprehensive reviews have been published recently (Peelle and Davis, 2012; Ding and Simon, 2014; Zion Golumbic et al., 2012). Rather, we will focus on high-level processes that can modulate or even underlie this alignment. Critically, it is necessary to differentiate between (i) high-level *modulations* of phase entrainment in response to stimuli that comprise properties that evoke systematic fluctuations on a low level of auditory processing (such as the cochlea) and (ii) high-level *entrainment*, i.e. phase entrainment that can be observed even in the absence of systematic fluctuations of low-level properties. In this review, the latter are defined as stimulus amplitude and spectral content, as those two properties can passively entrain the lowest level of auditory processing and evoke steady-state-potential-like (ASSR; Galambos et al., 1981) fluctuations in the cochlea. Convincing results have been obtained in the last years for both types of high-level processes, and we will address them in separate sections. We conclude this review with a section dedicated to the role of intelligibility for phase entrainment to speech sound, as the influence of semantic information on the brain's adjustment to speech is currently a strongly debated topic.

2. High-level modulations of phase entrainment to speech sound

Certain cognitive processes, such as attention, expectation or interpretation, are often considered "high-level" functions of the human brain, as they require, for instance, evaluation, selection, and the comparison of the actual stimulation with experience (Lamme and Spekreijse, 2000; Peelen and Kastner, 2014; Gilbert and Li, 2013). A modulation of phase entrainment to speech sound by those cognitive processes would argue for phase entrainment being a process that is not restricted to a purely sensory mechanism, but rather the active gating mechanism (or "active sensing"; Schroeder et al., 2010) that was explained above. Indeed, there is accumulating evidence for phase entrainment critically relying on attentional processes: One example is based on the so-called "cocktail party effect" (Cherry, 1953), describing a situation of several competing speakers, one of which has to be selected within the "noise" of the other, potentially distracting, speakers.

Several recent studies have shown a relation between the "cocktail party effect" and phase entrainment (the theoretical background is shown in Figure 2A and underlined by experimental results in Figure 2B): In Kerlin et al. (2010), two different speech streams were presented to the participants, one to each ear, and they were asked to selectively attend one of those two competing streams. They found that the representation of the attended speech stream in the delta/theta range (~2-8 Hz; the dominant frequency range of the speech envelope) of the electroencephalogram (EEG) signal was enhanced compared to that of the unattended stream. In other words, phase-locking between the EEG signal and the speech envelope of the attended stream was stronger than that between the EEG signal and the unattended stream. A similar paradigm was used in the studies by Ding and Simon (2012a), Horton et al. (2013) and Zion Golumbic et al. (2013b) for magnetoencephalogram (MEG), EEG and intracranial recordings in human subjects, respectively. All studies confirmed the finding that the phase of delta/theta brain oscillations "tracks" the envelope of speech sound, and that this "tracking" is enhanced when the speech is attended in a multi-speaker scenario. Interestingly, all studies reported that even the unattended speech signal is still represented (albeit weakly) in lower-level auditory cortices (i.e. regions closely related to sensory processing). However, as shown in the work by Zion Golumbic et al. (2013b), this unattended signal is "lost" in higher-level regions (e.g., frontal regions). Ding and Simon (2012a)

demonstrated that only the representation of the attended (and not the unattended) speech envelope varies as a function of stimulus intensity. This finding is important, because it suggests that attended and unattended inputs are processed separately in the brain, and that the alignment between neural phase and speech rhythm is used to form individual "auditory objects" (for a review on this notion, see Simon, 2015). In line with the notion of phase entrainment as an "amplifier-attenuator mechanism" (see first section), Horton et al. (2013) reported cross-correlations between speech envelope and EEG signal for both attended and unattended streams, but with opposite signs, suggesting that phase entrainment is indeed used to amplify one stream while the other is attenuated. Finally, it has been shown in several studies that the speech envelope can be reconstructed (i.e., it can be identified which stimulus the listener is attending) in multi-speaker (Ding and Simon, 2012a; Zion Golumbic et al., 2013b; O'Sullivan et al., 2015) or noisy environments (Ding and Simon, 2013) by using the delta/theta phase of neural oscillations (but also their gamma power; Mesgarani and Chang, 2012; Zion Golumbic et al., 2013b). It is possible that in those kind of situations, where one speech stream has to be actively extracted from a noisy environment, attention is of particular importance for phase entrainment to speech sound, whereas clear speech can be processed largely independently of attention (Wild et al., 2012).

Not only attention can be considered a high-level process: Predictions reflect a comparison between present and previous experiences and its projection to the future and must therefore involve high-level functions of the brain (Friston, 2005; Arnal and Giraud, 2012). Indeed, it has been shown that predictions do influence phase entrainment to speech sound: For instance, in the "cocktail party" scenario described above, Zion Golumbic et al. (2013a) paired the auditory speech input with the speaker's face and found that phase entrainment to the speech envelope was significantly enhanced by this visual input. Similar results were obtained by Arnal et al. (2011) using congruent and incongruent audiovisual stimuli (syllables) and by Luo et al. (2010) when subjects were watching audiovisual movies. A common interpretation of these findings is that, due to the slight delay between visual and auditory components of a conversation (the visual input preceding the auditory one), the former can be used to predict the timing of speech sound, thus enabling a better alignment between the oscillatory phase and speech envelope (Arnal et al., 2009, 2011; Zion Golumbic et al., 2013a; Perrodin et al., 2015; for a review, summarizing several existing theories, see Peelle and Sommers, 2015). Recent research suggests that not only the visual, but also the motor system plays a critical role for an efficient adjustment of excitability fluctuations in auditory cortex to expected upcoming events (Arnal et al., 2014; Arnal and Giraud, 2012; Morillon et al., 2014, 2015; Fujioka et al., 2012; Morillon and Schroeder, 2015).

Not an experimental, but rather an analytical proof of high-level processes involved in phase entrainment was provided by two recent studies (Fontolan et al., 2014; Park et al., 2015). Fontolan et al. (2014) used Granger causality (Granger, 1969), applied on data recorded intracranially in human subjects, to demonstrate that information reflected in the phase of low-frequency oscillations in response to speech sound travels in top-down direction from higher-order auditory to primary auditory regions, where it modulates the power of (gamma) oscillations at higher frequencies. Park et al. (2015) analyzed their data, recorded with MEG, using transfer entropy measures (Schreiber, 2000). They were able to show that frontal and motor areas can modulate the phase of delta/theta oscillations in auditory cortex (note that the spatial resolution in this study was lower than for intracranial recordings. It is thus unclear whether these delta/theta oscillations correspond to those in higher-order auditory or primary auditory cortices described in Fontolan et

al., 2014). Importantly, these top-down signals were correlated with an enhanced phase entrainment to speech sound when tracking of forward vs. backward speech was compared, indicating that higher-level processes can directly control the alignment between neural oscillations and speech sound.

The results described in this section strongly support the view that phase entrainment is a tool for attentional selection (Schroeder and Lakatos, 2009), filtering out irrelevant input and enhancing the representation of the attended stimulus in the brain. Predictions, potentially reflected by top-down mechanisms, help "designing" this filter by providing the timing for the alignment of "good" and "bad" phases of the oscillation to predicted relevant and irrelevant stimuli, respectively. This mechanism would not only help selecting relevant input in a noisy background, but also parse the speech signal at the same time: Here, one cycle of the aligned oscillation would represent one segment of information (or "chunk"; Ghitza, 2011, 2014, 2013; Doelling et al., 2014) that is analyzed by means of faster oscillations (Giraud and Poeppel, 2012; Luo and Poeppel, 2012; for reviews, see Peelle and Davis, 2012; Ding and Simon, 2014). Thus, phase entrainment could function as a means of discretization (equivalent ideas are mentioned by Peelle and Davis, 2012; Zion Golumbic et al., 2012), similar to "perceptual cycles" commonly observed in vision (VanRullen et al., 2014).

3. Phase entrainment to high-level features of speech sound

In the previous section, we have seen that high-level mechanisms of the brain, related to attention or prediction, clearly contribute to phase entrainment to speech sound. However, it should be noted that this contribution may just be *modulatory*: high-level mechanisms could merely *influence* a process, namely phase entrainment, that *itself* might rely on purely low-level processes. Indeed, speech sound consists of large fluctuations in low-level properties (i.e. stimulus amplitude and spectral content) that might evoke systematic fluctuations in neural activity already at the earliest level of auditory processing: the cochlea. These fluctuations in neural activity accompanying changes in the speech envelope would be indistinguishable from an active entrainment response. It is therefore necessary to construct stimuli without systematic fluctuations in those low-level properties in order to prove genuine high-level entrainment. In a recent publication (Zoefel and VanRullen, 2015), we were able to construct such stimuli and we review the most important findings in this section. Figure 3 shows the idea underlying stimulus construction in this study: In everyday speech sound (Figure 3A), spectral energy (color-coded) clearly differs between different phases of the speech envelope. In the view of a single cochlear cell, this sound would periodically alternate between weak (e.g., at phase ± pi, which is the trough of the speech envelope) and strong excitation (e.g., at phase 0, which is the peak of the speech envelope). Consequently, at a larger scale, we would measure an oscillatory pattern of neural activity that strongly depends on envelope phase. This pattern, however, would only reflect the periodicity of the stimulation. Therefore, we constructed noise sound whose spectral energy was tailored to counterbalance spectral differences as a function of envelope phase of the original speech sound (for details of stimulus construction, see Zoefel and VanRullen, 2015). This noise was mixed with the original speech and resulted in speech/noise sound that did, on average, not show those systematic differences in spectral content anymore (Figure 3B). Critically, as those stimuli remain intelligible, high-level features of speech (such as, but not restricted to, phonetic information) are still present and enable the listener to entrain to the speech sound that is now "hidden" inside the noise (note that the degree to which the

speech is "hidden" in noise depends on the original envelope phase, with speech perceptually dominant at the original envelope peak, and noise perceptually dominant at the original envelope trough). We applied those stimuli in two studies: In the first (Zoefel and VanRullen, 2015), a psychophysical study, we found that the detection of a short tone pip was significantly modulated (p-values shown in Figure 4A) by the remaining high-level features: Performance (Figure 4B) depended on the original envelope phase and thus differed between periods of dominant speech and noise. Note that speech and noise were spectrally matched; differences in performance could thus not be due to spectral differences between speech and noise, but rather due to the remaining high-level features that enable the listener to differentiate speech and noise. In the second study (Zoefel and VanRullen, in press), those stimuli were presented to listeners while their EEG was recorded. We found that EEG oscillations phase-lock to those high-level features of speech sound (Figure 4C), and the degree of entrainment (but not the phase relation between speech and EEG signal; see insets in Figure 4C) was similar to when the original everyday speech was presented. These results suggest an entrainment of neural oscillations as the mechanism underlying our perceptual findings.

It is not only interesting to investigate phase entrainment to speech stimuli *without* potentially entraining low-level features, but also to speech stimuli *only* containing the latter. This was done in a study by Ding et al. (2013) that might be seen as complementary to the other two described in this section. In their study, noise vocoding (Green et al., 2002) was used in order to design stimuli where spectro-temporal fine structure (which can be considered as high-level features) was strongly reduced, but the speech envelope was essentially unchanged. Those stimuli were presented either in noise or in quiet, and the MEG was recorded in parallel. Ding et al. (2013) showed that, indeed, reducing spectro-temporal fine structure also reduces the observed phase entrainment to speech sound. It is important to mention that the effect was only observed in noise (and not in quiet), underlining the idea that separating speech and noise might be one of the main functions of phase entrainment to speech sound (see sections 1 and 2). Rimmele et al. (2015) both extend the findings by Ding et al. (2013) and build a bridge to section 2 by showing that the enhanced "envelope tracking" for natural compared to noise-vocoded speech is only present when the speech is attended. They interpret their results as evidence for a high-level mechanism ("linguistic processing") that is only possible when the speech is in the focus of the listener's attention.

Taken together, the results reported in this section suggest that phase entrainment to speech sound is not only a reflection of fluctuations in low-level features of speech sound, but entails an adaption to phonetic information – and thus a genuine high-level process.

4. The role of intelligibility for phase entrainment to speech sound

Of course, the ultimate goal of every conversation is to transmit information, and without intelligibility, this goal cannot be achieved. Thus, it is all the more surprising that the role of intelligibility for phase entrainment to speech is currently strongly debated. This controversy is due to seemingly contradictory results that have recently been published: On the one hand, both Ahissar et al. (2001) and Luo and Poeppel (2007) found a correlation between phase entrainment (i.e. alignment of delta/theta oscillations and speech envelope) and speech intelligibility, a finding that has been confirmed by recent studies (Doelling et al., 2014; Ding et al., 2013; Park et al.,

2015). On the other hand, phase entrainment is not a phenomenon that is unique to speech sound and can also be found in response to much simpler stimuli, such as pure tones (Besle et al., 2011; Gomez-Ramirez et al., 2011; Stefanics et al., 2010; Zoefel and Heil, 2013; Lakatos et al., 2005, 2008). Also, the manipulation of speech intelligibility might destroy acoustic (i.e. non-semantic) properties of the sound that the brain actually entrains to (such as acoustic "edges"; Doelling et al., 2014), leading to a decline in phase entrainment and speech intelligibility at the same time, but without any relation between the two (Peelle and Davis, 2012; Millman et al., 2015). Moreover, several studies showed phase entrainment of neural oscillations to reversed (i.e. unintelligible) speech sound (Howard and Poeppel, 2010; Millman et al., 2015; Peelle et al., 2013) suggesting that phase entrainment does not necessarily depend on intelligibility. The whole picture gets even more complicated, as, although phase entrainment to speech sound is possible even when the speech is unintelligible, is seems to be enhanced by intelligible speech in some (but not all) studies (Peelle et al., 2013; Gross et al., 2013; Park et al., 2015) and attention seems to be important for this enhancement (Rimmele et al., 2015). Further evidence that the role of intelligibility for phase entrainment is not trivial was reported in two of the studies described in the previous section: In Zoefel and VanRullen (2015), it was found that perceptual entrainment to high-level features of speech sound is disrupted when the speech/noise sound is reversed (Figure 4A,B; red line) and this result was interpreted as a critical role of intelligibility for perceptual phase entrainment. On the other hand, in Zoefel and VanRullen (in press), using the same reversed speech/noise stimuli, the observed EEG phase entrainment was similar to that obtained in response to everyday speech and to (forward) speech/noise sound, seemingly in contradiction to the behavioral results obtained in Zoefel and VanRullen (2015).

How can we reconcile these studies, some of them clearly arguing against, and some for an important role of intelligibility for phase entrainment? Based on the current state of research, it is important to avoid overhasty conclusions and our interpretations have to remain speculative. Overall, phase entrainment seems to be a necessary, but not sufficient condition for speech comprehension: Speech intelligibility might not be possible without phase-locking, as we are not aware of any study reporting intelligible stimuli without oscillations (or perception) aligned to critical (low- and high-level) features of the speech sound. On the other hand, neural oscillations entrain to rhythmic structures (including reversed speech) even in the absence of intelligibility. It is clear that phase entrainment is a much more general phenomenon, and the brain might continuously scan its input for rhythmic patterns (indeed, popularity for auditory rhythms can be found in all cultures across the world and synchronization with rhythms – e.g., by clapping or dancing – is a general reaction to them). Once a rhythmic pattern has been detected, neural oscillations will align their phase to it (operating in the "rhythmic mode" described in Schroeder and Lakatos, 2009; see also Zoefel and Heil, 2013). Based on this notion, neural oscillations might always align to sound, as long as a rhythmic pattern can be detected (note that even the reversed speech/noise sound used in Zoefel and VanRullen, 2015, in press, contains a rhythmic pattern, as speech and noise can perceptually be differentiated). But what is the role of intelligibility? It is important to find a model that is at the same time parsimonious and can explain most results described in the literature. These findings are shortly summarized in the following:

1. Rhythmic non-speech stimuli, such as trains of pure tones, entrain neural oscillations (Besle et al., 2011; Gomez-Ramirez et al., 2011; Zoefel and Heil, 2013; Lakatos et al.,

2005) and modulate behavior (Lakatos et al., 2008; Stefanics et al., 2010; Hickok et al., 2015; Thut et al., 2012).

2. Speech stimuli, both intelligible and unintelligible, entrain neural oscillations (Ahissar et al., 2001; Luo and Poeppel, 2007; Zoefel and VanRullen, in press; Doelling et al., 2014; Ding et al., 2013; Ding and Simon, 2012a; Park et al., 2015; Peelle et al., 2013; Rimmele et al., 2015; Howard and Poeppel, 2010; Millman et al., 2015; Zion Golumbic et al., 2013b).
3. The rhythm of speech only modulates behavior when speech is intelligible (Zoefel and VanRullen, 2015).
4. Neural entrainment to intelligible speech might be increased when compared to unintelligible speech (Luo and Poeppel, 2007; Peelle et al., 2013; Rimmele et al., 2015; Park et al., 2015; Doelling et al., 2014). However, not all studies can confirm this result (Howard and Poeppel, 2010; Zoefel and VanRullen, in press; Millman et al., 2015).

One model that can potentially reconcile these findings is presented in Figure 5, and the different parts and implications of this model are discussed in the following. However, we acknowledge that it is only one out of possibly several candidate models to explain the data available in the literature. Nevertheless, in our view, this model is currently the most parsimonious explanation for existing findings and we therefore focus our review on it. The first implication of our model is that different regions in the brain are "responsible" for different processes: Phase entrainment might be found throughout the whole auditory system, but most studies emphasize primary auditory cortex (A1; Lakatos et al., 2005; O'Connell et al., 2014; Lakatos et al., 2013) or early temporal regions (Gomez-Ramirez et al., 2011; Ding and Simon, 2012b; Zion Golumbic et al., 2013b). An influence of intelligibility is commonly related to regions specifically processing speech sound (Binder et al., 2000; Scott et al., 2000; Mesgarani et al., 2014; Hickok and Poeppel, 2007; DeWitt and Rauschecker, 2012; Poeppel et al., 2012). Finally, frontal regions are a likely candidate for behavioral outcome (Krawczyk, 2002; Coutlee and Huettel, 2012; Rushworth et al., 2012; Romo and de Lafuente, 2013). In order to satisfy point (1), we assume that the entrainment in temporal regions can directly influence behavior as determined in frontal regions, as long as the entrainment is introduced by non-speech stimuli (Figure 5A). This results in a periodic modulation of performance as often described (Spaak et al., 2014; Landau and Fries, 2012; Vanrullen and Dubois, 2011; Zoefel and Sokoliuk, 2014; Song et al., 2014; Hickok et al., 2015; Thut et al., 2012; Fiebelkorn et al., 2011). But not only non-speech stimuli can entrain temporal regions, the same is true for speech sound, irrespective of its intelligibility (point 2). However, speech intelligibility affects high-order auditory regions and they might directly influence the impact of temporal on frontal regions (Figure 5B). This notion is based on the increasing number of studies supporting the idea that the state of connectivity (or synchronization) between two (potentially distant) brain regions is crucial for perceptual outcome (Ruhnau et al., 2014; Weisz et al., 2014; Fries, 2005). Thus, speech intelligibility might modulate the state of connectivity between temporal and frontal regions. We hypothesize that speech-specific regions are only activated if the input contains *acoustic* high-level (i.e. speech-specific) features of speech; otherwise these regions remain inactive and do not exhibit any modulatory effect on other regions or their connectivity. However, once the input is identified as speech (based on these acoustic features), *linguistic* features determine whether the modulatory effect is negative (desynchronizing temporal and frontal regions, resulting in no behavioral effect of the entrainment; in case of unintelligible speech) or positive (synchronizing temporal and frontal regions, resulting in a behavioral effect of the

entrainment; in case of intelligible speech). This assumption satisfies point (3). In contrast to unintelligible speech, intelligible speech might result in an entrainment that also includes high-order (speech-specific) auditory regions: They might have to entrain to the speech sound in order to be able to synchronize temporal and frontal regions. That might be the reason that some studies show an increased entrainment for intelligible as compared to unintelligible speech whereas others do not (point 4): They might have captured the entrainment in those higher-level auditory regions – something which, due to the low spatial resolution in most EEG/MEG studies, is difficult to determine but could be resolved in future studies. More research is clearly needed: What are those behavioral variables that are differentially affected by intelligible and unintelligible speech? Where exactly are those brain regions hypothesized to be responsible for (or affected by) phase entrainment, for behavioral decisions and for the modulation of their relation by speech intelligibility? What are the mechanisms connecting these functional networks? Answering these questions has critical implications for our understanding of the brain's processing of human speech and rhythmic input in general.

5. Conclusions

Recently, phase entrainment has attracted researchers' attention as a potential reflection of the brain's mechanism to efficiently allocate attentional resources in time. Nevertheless, the periodicity of the stimulation itself complicates this interpretation, as the brain might simply follow the rhythm of its input. In this review, we presented an increasing amount of evidence that speaks against a merely passive role of neural oscillations for phase entrainment to speech sound. Instead, the brain might constantly predict the timing of relevant and irrelevant events of speech sound, including acoustic high-level features, and actively align neural oscillations so that they efficiently boost the current locus of attention in a noisy background. Linguistic high-level features, reflecting intelligibility, might play a modulatory, and speech-specific, role by determining the behavioral consequences of phase entrainment to speech sound.

6. References

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., and Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proc. Natl. Acad. Sci. U. S. A.* 98, 13367–13372. doi:10.1073/pnas.201400998.

Arnal, L. H., Doelling, K. B., and Poeppel, D. (2014). Delta-Beta Coupled Oscillations Underlie Temporal Prediction Accuracy. *Cereb. Cortex* doi:10.1093/cercor/bhu103.

Arnal, L. H., and Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. *Trends Cogn. Sci.* 16, 390–398. doi:10.1016/j.tics.2012.05.003.

Arnal, L. H., Wyart, V., and Giraud, A.-L. (2011). Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. *Nat. Neurosci.* 14, 797–801. doi:10.1038/nn.2810.

Besle, J., Schevon, C. A., Mehta, A. D., Lakatos, P., Goodman, R. R., McKhann, G. M., Emerson, R. G., and Schroeder, C. E. (2011). Tuning of the human neocortex to the temporal dynamics of attended events. *J. Neurosci.* 31, 3176–3185. doi:10.1523/JNEUROSCI.4518-10.2011.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., and Possing, E. T. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cereb. Cortex* 10, 512–528.

Buzsáki, G., and Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science* 304, 1926–1929. doi:10.1126/science.1099745.

Calderone, D. J., Lakatos, P., Butler, P. D., and Castellanos, F. X. (2014). Entrainment of neural oscillations as a modifiable substrate of attention. *Trends Cogn. Sci.* 18, 300–309. doi:10.1016/j.tics.2014.02.005.

Cherry, E. C. (1953). Some Experiments on the Recognition of Speech, with One and with Two Ears. *J. Acoust. Soc. Am.* 25, 975–979. doi:10.1121/1.1907229.

Coutlee, C. G., and Huettel, S. A. (2012). The functional neuroanatomy of decision making: prefrontal control of thought and action. *Brain Res.* 1428, 3–12. doi:10.1016/j.brainres.2011.05.053.

Cummins, F. (2012). Oscillators and syllables: a cautionary note. *Front. Psychol.* 3, 364. doi:10.3389/fpsyg.2012.00364.

DeWitt, I., and Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream. *Proc. Natl. Acad. Sci. U. S. A.* 109, E505–514. doi:10.1073/pnas.1113427109.

Ding, N., Chatterjee, M., and Simon, J. Z. (2013). Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. *NeuroImage* 88C, 41–46. doi:10.1016/j.neuroimage.2013.10.054.

Ding, N., and Simon, J. Z. (2013). Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. *J. Neurosci.* 33, 5728–5735. doi:10.1523/JNEUROSCI.5297-12.2013.

Ding, N., and Simon, J. Z. (2014). Cortical entrainment to continuous speech: functional roles and interpretations. *Front. Hum. Neurosci.* 8, 311. doi:10.3389/fnhum.2014.00311.

Ding, N., and Simon, J. Z. (2012a). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc. Natl. Acad. Sci. U. S. A.* 109, 11854–11859. doi:10.1073/pnas.1205381109.

Ding, N., and Simon, J. Z. (2012b). Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *J. Neurophysiol.* 107, 78–89. doi:10.1152/jn.00297.2011.

Doelling, K. B., Arnal, L. H., Ghitza, O., and Poeppel, D. (2014). Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage* 85 Pt 2, 761–768. doi:10.1016/j.neuroimage.2013.06.035.

Fiebelkorn, I. C., Foxe, J. J., Butler, J. S., Mercier, M. R., Snyder, A. C., and 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. doi:10.1523/JNEUROSCI.1338-11.2011.

Fontolan, L., Morillon, B., Liegeois-Chauvel, C., and Giraud, A.-L. (2014). The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex. *Nat. Commun.* 5, 4694. doi:10.1038/ncomms5694.

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn. Sci.* 9, 474–480. doi:10.1016/j.tics.2005.08.011.

Friston, K. (2005). A theory of cortical responses. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 360, 815–836. doi:10.1098/rstb.2005.1622.

Fujioka, T., Trainor, L. J., Large, E. W., and Ross, B. (2012). Internalized timing of isochronous sounds is represented in neuromagnetic β oscillations. *J. Neurosci.* 32, 1791–1802. doi:10.1523/JNEUROSCI.4107-11.2012.

Galambos, R., Makeig, S., and Talmachoff, P. J. (1981). A 40-Hz auditory potential recorded from the human scalp. *Proc. Natl. Acad. Sci. U. S. A.* 78, 2643–2647.

Ghitza, O. (2014). Behavioral evidence for the role of cortical θ oscillations in determining auditory channel capacity for speech. *Front. Psychol.* 5, 652. doi:10.3389/fpsyg.2014.00652.

Ghitza, O. (2011). Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. *Front. Psychol.* 2, 130. doi:10.3389/fpsyg.2011.00130.

Ghitza, O. (2013). The theta-syllable: a unit of speech information defined by cortical function. *Front. Psychol.* 4, 138. doi:10.3389/fpsyg.2013.00138.

Gilbert, C. D., and Li, W. (2013). Top-down influences on visual processing. *Nat. Rev. Neurosci.* 14, 350–363. doi:10.1038/nrn3476.

Giraud, A.-L., and Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* 15, 511–517. doi:10.1038/nn.3063.

Gomez-Ramirez, M., Kelly, S. P., Molholm, S., Sehatpour, P., Schwartz, T. H., and Foxe, J. J. (2011). Oscillatory sensory selection mechanisms during intersensory attention to rhythmic auditory and visual inputs: a human electrocorticographic investigation. *J. Neurosci.* 31, 18556–18567. doi:10.1523/JNEUROSCI.2164-11.2011.

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica* 37, 424–438. doi:10.2307/1912791.

Green, T., Faulkner, A., and Rosen, S. (2002). Spectral and temporal cues to pitch in noise-excited vocoder simulations of continuous-interleaved-sampling cochlear implants. *J. Acoust. Soc. Am.* 112, 2155–2164.

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., and Garrod, S. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biol.* 11, e1001752. doi:10.1371/journal.pbio.1001752.

Henry, M. J., Herrmann, B., and Obleser, J. (2014). Entrained neural oscillations in multiple frequency bands comodulate behavior. *Proc. Natl. Acad. Sci. U. S. A.* 111, 14935–14940. doi:10.1073/pnas.1408741111.

Henry, M. J., and Obleser, J. (2012). Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proc. Natl. Acad. Sci. U. S. A.* 109, 20095–20100. doi:10.1073/pnas.1213390109.

Hickok, G., Farahbod, H., and Saberi, K. (2015). The Rhythm of Perception: Entrainment to Acoustic Rhythms Induces Subsequent Perceptual Oscillation. *Psychol. Sci.* doi:10.1177/0956797615576533.

Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402. doi:10.1038/nrn2113.

Horton, C., Zmura, M. D', and Srinivasan, R. (2013). Suppression of competing speech through entrainment of cortical oscillations. *J. Neurophysiol.* 109, 3082–3093. doi:10.1152/jn.01026.2012.

Howard, M. F., and Poeppel, D. (2010). Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. *J. Neurophysiol.* 104, 2500–2511. doi:10.1152/jn.00251.2010.

Kerlin, J. R., Shahin, A. J., and Miller, L. M. (2010). Attentional gain control of ongoing cortical speech representations in a "cocktail party." *J. Neurosci.* 30, 620–628. doi:10.1523/JNEUROSCI.3631-09.2010.

Krawczyk, D. C. (2002). Contributions of the prefrontal cortex to the neural basis of human decision making. *Neurosci. Biobehav. Rev.* 26, 631–664.

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., and Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science* 320, 110–113. doi:10.1126/science.1154735.

Lakatos, P., Musacchia, G., O'Connel, M. N., Falchier, A. Y., Javitt, D. C., and Schroeder, C. E. (2013). The spectrotemporal filter mechanism of auditory selective attention. *Neuron* 77, 750–761. doi:10.1016/j.neuron.2012.11.034.

Lakatos, P., Shah, A. S., Knuth, K. H., Ulbert, I., Karmos, G., and Schroeder, C. E. (2005). An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *J. Neurophysiol.* 94, 1904–1911. doi:10.1152/jn.00263.2005.

Lamme, V. A., and Spekreijse, H. (2000). Modulations of primary visual cortex activity representing attentive and conscious scene perception. *Front. Biosci.* 5, D232–243.

Landau, A. N., and Fries, P. (2012). Attention samples stimuli rhythmically. *Curr. Biol.* 22, 1000–1004. doi:10.1016/j.cub.2012.03.054.

Luo, H., Liu, Z., and Poeppel, D. (2010). Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS Biol.* 8, e1000445. doi:10.1371/journal.pbio.1000445.

Luo, H., and Poeppel, D. (2012). Cortical oscillations in auditory perception and speech: evidence for two temporal windows in human auditory cortex. *Front. Psychol.* 3, 170. doi:10.3389/fpsyg.2012.00170.

Luo, H., and Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54, 1001–1010. doi:10.1016/j.neuron.2007.06.004.

Mazzoni, A., Whittingstall, K., Brunel, N., Logothetis, N. K., and Panzeri, S. (2010). Understanding the relationships between spike rate and delta/gamma frequency bands of LFPs and EEGs using a local cortical network model. *NeuroImage* 52, 956–972. doi:10.1016/j.neuroimage.2009.12.040.

Mesgarani, N., and Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature* 485, 233–236. doi:10.1038/nature11020.

Mesgarani, N., Cheung, C., Johnson, K., and Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science* 343, 1006–1010. doi:10.1126/science.1245994.

Millman, R. E., Johnson, S. R., and Prendergast, G. (2015). The Role of Phase-locking to the Temporal Envelope of Speech in Auditory Perception and Speech Intelligibility. *J. Cogn. Neurosci.* 27, 533–545. doi:10.1162/jocn_a_00719.

Morillon, B., Hackett, T. A., Kajikawa, Y., and Schroeder, C. E. (2015). Predictive motor control of sensory dynamics in auditory active sensing. *Curr. Opin. Neurobiol.* 31C, 230–238. doi:10.1016/j.conb.2014.12.005.

Morillon, B., and Schroeder, C. E. (2015). Neuronal oscillations as a mechanistic substrate of auditory temporal prediction. *Ann. N. Y. Acad. Sci.* 1337, 26–31. doi:10.1111/nyas.12629.
Morillon, B., Schroeder, C. E., and Wyart, V. (2014). Motor contributions to the temporal precision of auditory attention. *Nat. Commun.* 5, 5255. doi:10.1038/ncomms6255.
Nozaradan, S. (2014). Exploring how musical rhythm entrains brain activity with electroencephalogram frequency-tagging. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 369, 20130393. doi:10.1098/rstb.2013.0393.
Obleser, J., Herrmann, B., and Henry, M. J. (2012). Neural Oscillations in Speech: Don't be Enslaved by the Envelope. *Front. Hum. Neurosci.* 6, 250. doi:10.3389/fnhum.2012.00250.
O'Connell, M. N., Barczak, A., Schroeder, C. E., and Lakatos, P. (2014). Layer specific sharpening of frequency tuning by selective attention in primary auditory cortex. *J. Neurosci.* 34, 16496–16508. doi:10.1523/JNEUROSCI.2055-14.2014.
O'Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., Slaney, M., Shamma, S. A., and Lalor, E. C. (2015). Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. *Cereb. Cortex* 25, 1697–1706. doi:10.1093/cercor/bht355.
Park, H., Ince, R. A. A., Schyns, P. G., Thut, G., and Gross, J. (2015). Frontal Top-Down Signals Increase Coupling of Auditory Low-Frequency Oscillations to Continuous Speech in Human Listeners. *Curr. Biol.* 25, 1649–1653. doi:10.1016/j.cub.2015.04.049.
Peelen, M. V., and Kastner, S. (2014). Attention in the real world: toward understanding its neural basis. *Trends Cogn. Sci.* 18, 242–250. doi:10.1016/j.tics.2014.02.004.
Peelle, J. E., and Davis, M. H. (2012). Neural Oscillations Carry Speech Rhythm through to Comprehension. *Front. Psychol.* 3, 320. doi:10.3389/fpsyg.2012.00320.
Peelle, J. E., Gross, J., and Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cereb. Cortex* 23, 1378–1387. doi:10.1093/cercor/bhs118.
Poeppel, D., Emmorey, K., Hickok, G., and Pylkkänen, L. (2012). Towards a new neurobiology of language. *J. Neurosci.* 32, 14125–14131. doi:10.1523/JNEUROSCI.3244-12.2012.
Power, A. J., Mead, N., Barnes, L., and Goswami, U. (2013). Neural entrainment to rhythmic speech in children with developmental dyslexia. *Front. Hum. Neurosci.* 7, 777. doi:10.3389/fnhum.2013.00777.
Rajkai, C., Lakatos, P., Chen, C.-M., Pincze, Z., Karmos, G., and Schroeder, C. E. (2008). Transient cortical excitation at the onset of visual fixation. *Cereb. Cortex* 18, 200–209. doi:10.1093/cercor/bhm046.
Rimmele, J. M., Zion Golumbic, E., Schröger, E., and Poeppel, D. (2015). The effects of selective attention and speech acoustics on neural speech-tracking in a multi-talker scene. *Cortex* doi:10.1016/j.cortex.2014.12.014.
Romo, R., and de Lafuente, V. (2013). Conversion of sensory signals into perceptual decisions. *Prog. Neurobiol.* 103, 41–75. doi:10.1016/j.pneurobio.2012.03.007.
Ruhnau, P., Hauswald, A., and Weisz, N. (2014). Investigating ongoing brain oscillations and their influence on conscious perception - network states and the window to consciousness. *Front. Psychol.* 5, 1230. doi:10.3389/fpsyg.2014.01230.

Rushworth, M. F. S., Kolling, N., Sallet, J., and Mars, R. B. (2012). Valuation and decision-making in frontal cortex: one or many serial or parallel systems? *Curr. Opin. Neurobiol.* 22, 946–955. doi:10.1016/j.conb.2012.04.011.
Schreiber, T. (2000). Measuring information transfer. *Phys. Rev. Lett.* 85, 461–464.
Schroeder, C. E., and Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci.* 32, 9–18. doi:10.1016/j.tins.2008.09.012.
Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., and Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends Cogn. Sci.* 12, 106–113. doi:10.1016/j.tics.2008.01.002.
Schroeder, C. E., Wilson, D. A., Radman, T., Scharfman, H., and Lakatos, P. (2010). Dynamics of Active Sensing and perceptual selection. *Curr. Opin. Neurobiol.* 20, 172–176. doi:10.1016/j.conb.2010.02.010.
Scott, S. K., Blank, C. C., Rosen, S., and Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123, 2400–2406.
Simon, J. Z. (2015). The encoding of auditory objects in auditory cortex: insights from magnetoencephalography. *Int. J. Psychophysiol.* 95, 184–190. doi:10.1016/j.ijpsycho.2014.05.005.
Song, K., Meng, M., Chen, L., Zhou, K., and Luo, H. (2014). Behavioral oscillations in attention: rhythmic α pulses mediated through θ band. *J. Neurosci.* 34, 4837–4844. doi:10.1523/JNEUROSCI.4856-13.2014.
Spaak, E., de Lange, F. P., and Jensen, O. (2014). Local entrainment of α oscillations by visual stimuli causes cyclic modulation of perception. *J. Neurosci.* 34, 3536–3544. doi:10.1523/JNEUROSCI.4385-13.2014.
Stefanics, G., Hangya, B., Hernádi, I., Winkler, I., Lakatos, P., and Ulbert, I. (2010). Phase entrainment of human delta oscillations can mediate the effects of expectation on reaction speed. *J. Neurosci.* 30, 13578–13585. doi:10.1523/JNEUROSCI.0703-10.2010.
Steinschneider, M., Nourski, K. V., and Fishman, Y. I. (2013). Representation of speech in human auditory cortex: is it special? *Hear. Res.* 305, 57–73. doi:10.1016/j.heares.2013.05.013.
Thut, G., Miniussi, C., and Gross, J. (2012). The functional importance of rhythmic activity in the brain. *Curr. Biol.* 22, R658–663. doi:10.1016/j.cub.2012.06.061.
Vanrullen, R., and Dubois, J. (2011). The psychophysics of brain rhythms. *Front. Psychol.* 2, 203. doi:10.3389/fpsyg.2011.00203.
VanRullen, R., Zoefel, B., and Ilhan, B. (2014). On the cyclic nature of perception in vision versus audition. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 369, 20130214. doi:10.1098/rstb.2013.0214.
Weisz, N., Wühle, A., Monittola, G., Demarchi, G., Frey, J., Popov, T., and Braun, C. (2014). Prestimulus oscillatory power and connectivity patterns predispose conscious somatosensory perception. *Proc. Natl. Acad. Sci. U. S. A.* 111, E417–425. doi:10.1073/pnas.1317267111.
Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., and Johnsrude, I. S. (2012). Effortful listening: the processing of degraded speech depends critically on attention. *J. Neurosci.* 32, 14010–14021. doi:10.1523/JNEUROSCI.1528-12.2012.

Zion Golumbic, E., Cogan, G. B., Schroeder, C. E., and Poeppel, D. (2013a). Visual input enhances selective speech envelope tracking in auditory cortex at a "cocktail party." *J. Neurosci.* 33, 1417–1426. doi:10.1523/JNEUROSCI.3675-12.2013.

Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., Goodman, R. R., Emerson, R., Mehta, A. D., Simon, J. Z., et al. (2013b). Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party." *Neuron* 77, 980–991. doi:10.1016/j.neuron.2012.12.037.

Zion Golumbic, E. M., Poeppel, D., and Schroeder, C. E. (2012). Temporal context in speech processing and attentional stream selection: a behavioral and neural perspective. *Brain Lang.* 122, 151–161. doi:10.1016/j.bandl.2011.12.010.

Zoefel, B., and Heil, P. (2013). Detection of Near-Threshold Sounds is Independent of EEG Phase in Common Frequency Bands. *Front. Psychol.* 4, 262. doi:10.3389/fpsyg.2013.00262.

Zoefel, B., and Sokoliuk, R. (2014). Investigating the rhythm of attention on a fine-grained scale: evidence from reaction times. *J. Neurosci.* 34, 12619–12621. doi:10.1523/JNEUROSCI.2134-14.2014.

Zoefel, B., and VanRullen, R. (in press). EEG oscillations entrain their phase to high-level features of speech sound. *NeuroImage*.

Zoefel, B., and VanRullen, R. (2015). Selective perceptual phase entrainment to speech rhythm in the absence of spectral energy fluctuations. *J. Neurosci.* 35, 1954–1964. doi:10.1523/JNEUROSCI.3484-14.2015.

Figure captions

Figure 1. Entrainment as a tool for input gating. A. Brain oscillations (red) are unrelated to the stimulus input, here a segment of speech sound. Note that both oscillation and speech sound are rhythmic (~ 4 Hz) and that the speech input consists of phases of high (*) and low (#) informational content. Both phase and frequency (the latter to a certain extent; Ghitza, 2013, 2014) of the oscillations can be adjusted to match the input rhythm (red arrow), a phenomenon called phase entrainment. B. Phase entrainment results in an alignment of the oscillation's high and low excitability phases (blue) with the input's high and low informational content. It can thus be used as a tool for input gating.

Figure 2. Neural oscillations as a tool for attentional selection during a "cocktail party". A. Theoretical background (modified with permission from Zion Golumbic et al., 2012). Recorded neural activity in the delta/theta band (right column) aligns with the speech envelope (left column) of the respective speaker (red and blue), when presented separately. In a multi-speaker scenario ("cocktail party"), the recorded data will reflect the attended, but not necessarily (or to a smaller degree) the unattended speech envelope. B. Actual data (modified with permission from Zion Golumbic et al., 2013b) confirms the theoretical background. The speech envelope reconstructed from the recorded data (grey: single subject; red: averaged across subjects) strongly resembles the speech envelope of the attended, but not the unattended speaker.

Figure 3. Everyday speech sound (A) contains pronounced fluctuations in spectral energy (color-coded or shown as a family of curves; one curve for each phase bin of the speech envelope) that depend on the phase of the speech envelope. These (low-level) rhythmic fluctuations in energy *per*

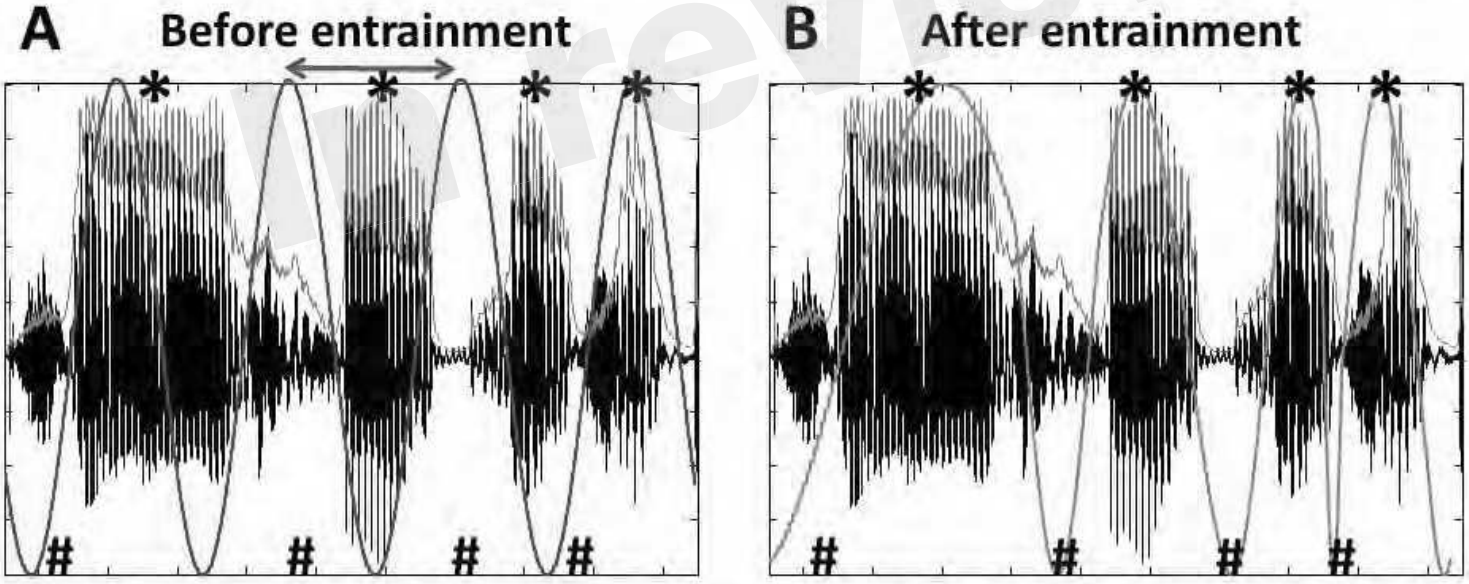
se might result in an apparent alignment between neural activity and speech envelope, as strong neural excitation (here at phase 0, due to high spectral energy) periodically alternates with weak neural excitation (here at phase ± pi, due to low spectral energy). Genuine high-level phase entrainment requires stimuli without those systematic fluctuations in spectral energy, as shown in B. The construction of those stimuli has recently been reported (Zoefel and VanRullen, 2015), and results obtained there are described in this review. Reproduced with permission from Zoefel and VanRullen (2015).

Figure 4. Perception and neural oscillations entrain to high-level features of speech sound. Speech/noise stimuli without systematic fluctuations in amplitude or spectral content were constructed, whose high-level features are conserved and reflected by the original speech envelope (cf. Figure 3). In a psychophysical study (A,B), the detection of a tone pip was significantly modulated by those high-level features (black; in this plot, results obtained in the original experiment and a follow-up replication have been combined; both experiments are described in Zoefel and VanRullen, 2015). The significance of this modulation is shown in panel A for different time lags relative to target onset, whereas the actual performance (at the time lag indicated by the vertical arrow in A) is shown in panel B. This effect was abolished when the speech/noise sound was reversed (red), indicating an important role of linguistic features (i.e. intelligibility) for behavioral consequences of the entrainment. In A, the significance threshold is shown as a dashed line (corrected for multiple comparisons). In B, standard error of mean (SEM) is shown by contours around the lines. When the same stimuli (and their original version of everyday speech) were presented in an EEG experiment (C), significant phase-locking between original speech envelope and EEG signal could be observed in all conditions (original, speech/noise sound and reversed speech/noise sound), suggesting that high-level features can entrain the phase of EEG oscillations, and do so even if the speech is unintelligible (note that acoustic high-level features remain present in the speech/noise sound, even when it is reversed, as the listener can still differentiate speech and noise). Bars show the average phase-locking across EEG channels, whereas the actual phase differences between EEG signal and original speech envelope, separately for each channel, are shown as insets above the bars (channels without significant entrainment are shaded out). P-values of phase entrainment, obtained by permutation tests, are shown as dashed lines. Note that, in contrast to the degree of entrainment which is comparable in all 3 conditions, the entrained phase does differ between everyday speech sound (original condition) and speech/noise sound in which systematic fluctuations in low-level features have been removed (constructed and constructed reversed conditions). Modified with permission from Zoefel and VanRullen (2015) (A,B) and Zoefel and VanRullen (in press) (C).

Figure 5. Intelligibility at the interface between phase entrainment and behavior. A. Non-speech stimuli do not activate speech-specific ("intelligibility") regions. Thus, entrainment in temporal regions can directly influence behavior – determined in frontal regions – in a periodic fashion, without an additional modulation by speech-specific regions. B. Acoustic high-level (speech-specific) features of speech activate speech-specific regions. This activation results in a modulation of the connectivity between temporal and frontal regions: If linguistic high-level features are present in the input (i.e. if the speech is intelligible), temporal and frontal regions are synchronized and entrainment in temporal regions can affect activity in frontal regions (and modulate behavior periodically, such as in A). If these features are not present (i.e. if the speech is unintelligible), temporal and frontal regions are desynchronized and entrainment in temporal regions cannot affect

frontal regions and behavior. Thus, (only) if the input is recognized as speech, intelligibility can act as a "switch", determining the influence of entrained oscillations on behavioral outcome.

Figure 1.TIF



High speech information * meets random excitatory phases

Low speech information # (silence) meets random excitatory phases

High speech information * meets high excitatory phases

Low speech information # (silence) meets low excitatory phases

Figure 2.TIF

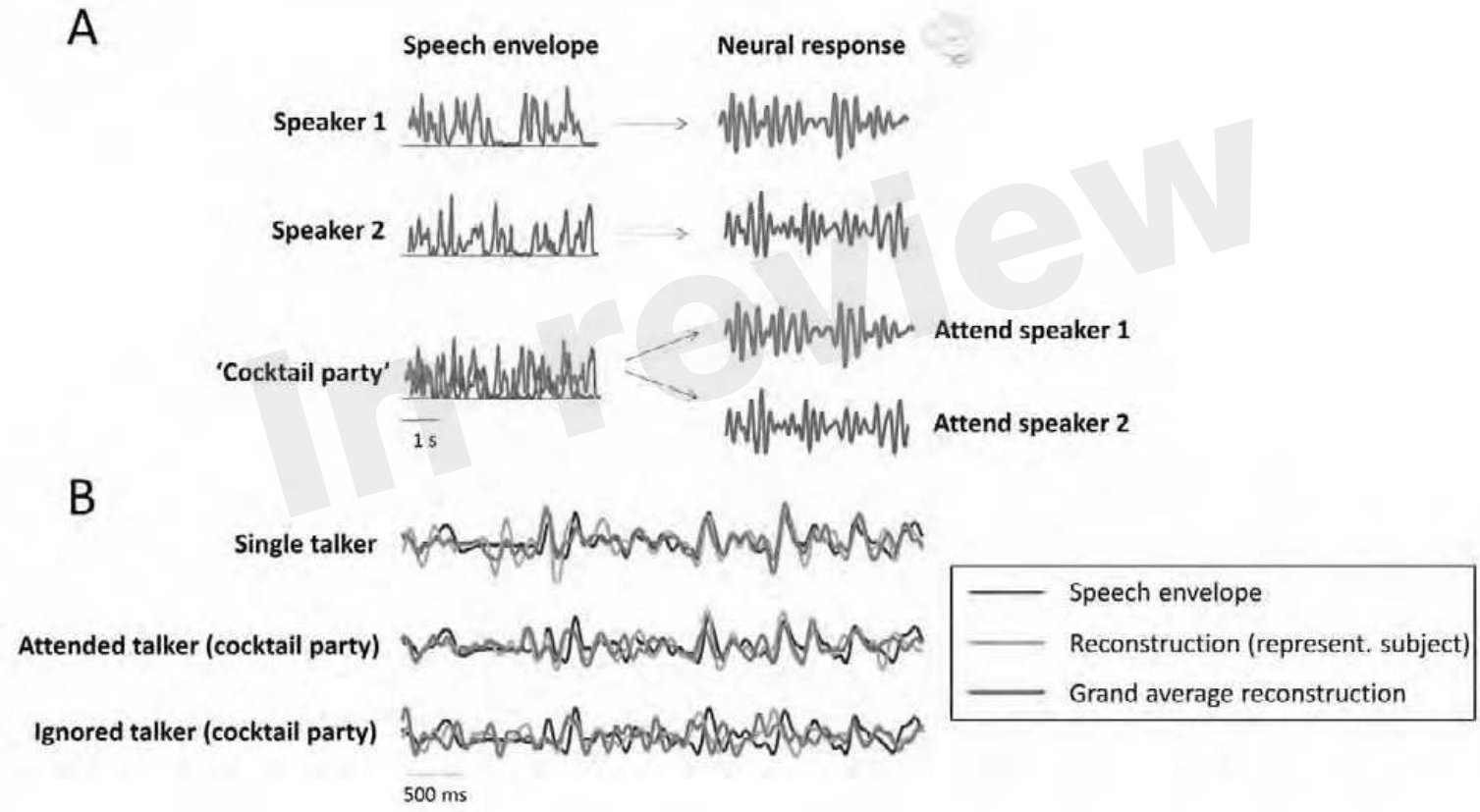


Figure 3.TIF

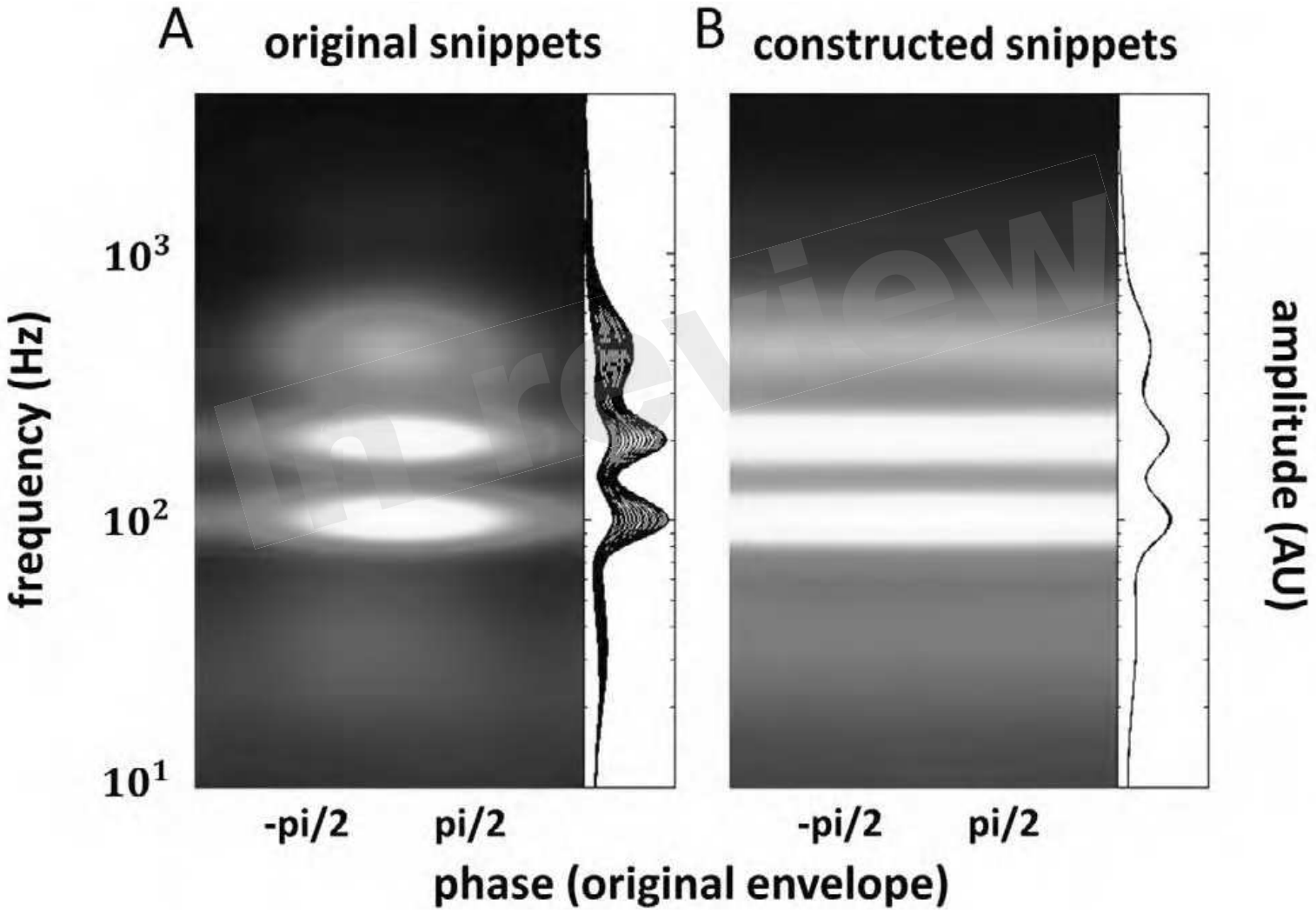


Figure 4.TIF

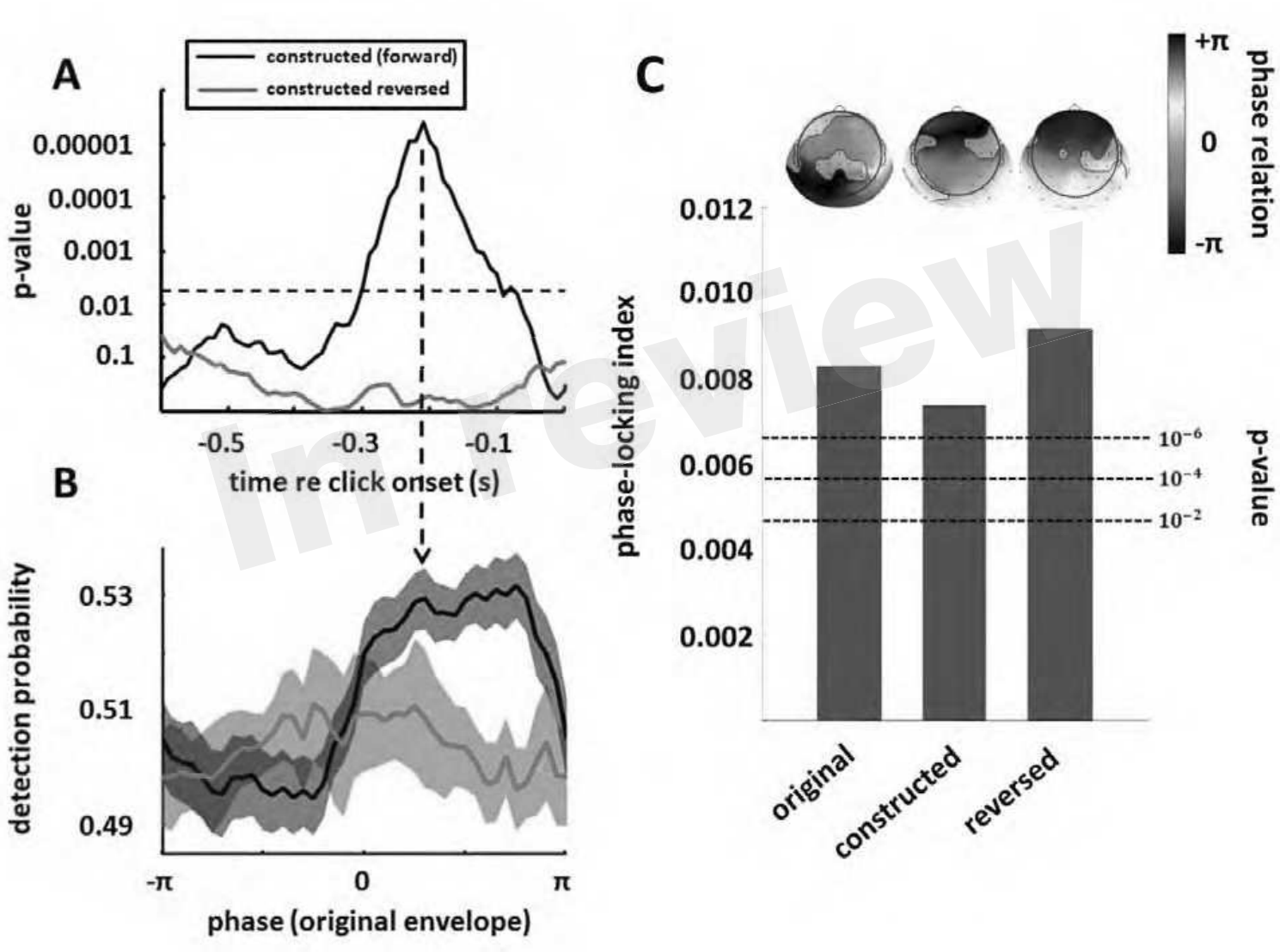
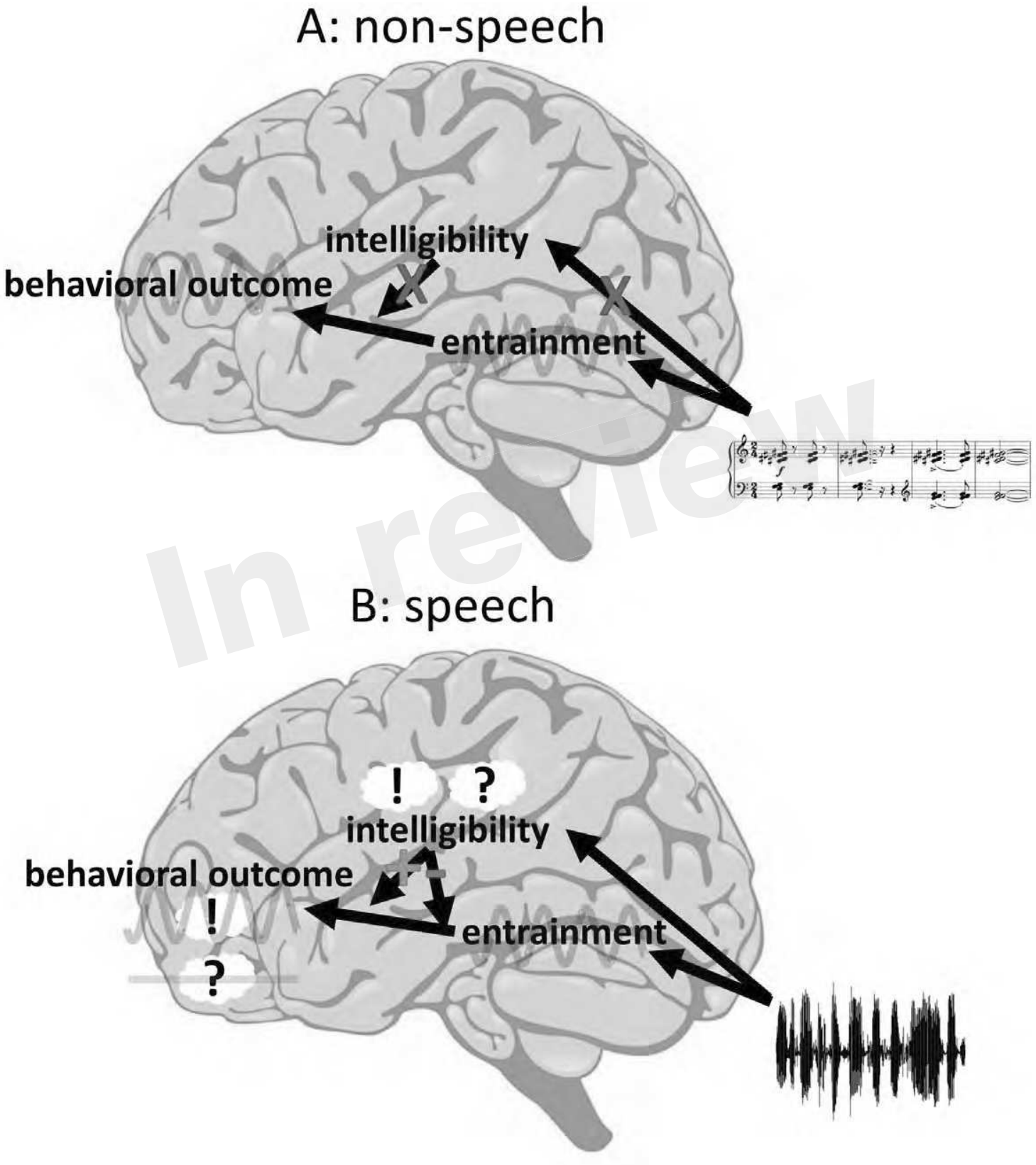


Figure 5.TIF



In the preceding review article, it was argued that phase entrainment of neural oscillations to human speech sound does entail different high-level mechanisms: These include high-level *modulations* of phase entrainment to speech, and entrainment to high-level features of speech *per se*. Two important additional findings should be mentioned here:

First, the described high-level mechanisms are not restricted to human subjects. The attentional modulation of phase entrainment to rhythmic sequences has been shown repeatedly for non-human primates: Lakatos et al. (2013) demonstrated that phase entrainment only functions as a "spectrotemporal filter" (described extensively in Chapter 6) in monkey A1 if the rhythmic stream is attended (but not if unattended). Lakatos et al. (2008) reported that the entrained phase in primary visual cortex depends on the attended modality, with a reversal of phase when visual or auditory rhythmic streams are attended, respectively. Moreover, as demonstrated in Chapter 6, neural oscillations in primary auditory cortex of the monkey are able to entrain to high-level features of human speech: This was shown by presenting one monkey with speech/noise stimuli without systematic fluctuations in spectral content or amplitude (their construction was presented in Chapter 4). We found that oscillations in monkey A1 significantly entrain to the remaining high-level features. Importantly, this "high-level entrainment" went along with a change in the entrained phase when compared with entrainment to regular speech sound, and this result was strikingly reminiscent of a similar change in the entrained phase measured in the human EEG (Chapter 5). The rhythm of human speech (~2-8 Hz, mainly reflecting the syllabic structure) is very similar to communication calls of other animals (Wang et al., 1995; Figure 1), and perception of both human and non-human species seems to be tuned to these modulation frequencies (Eggermont, 1998; Oshurkova et al., 2008; Edwards and Chang, 2013). Thus, it seems that, in describing these results, we might tap into a mechanism that is

much broader than previously thought and not restricted to human subjects, nor to human speech. Clearly, more work needs to be done: For instance, it would be interesting to apply our procedure of equalizing spectral content across different phases of the signal envelope to other communication calls and present those "high-level stimuli" to the respective species. By recording at different sites within the auditory stream, an even more thorough characterization of the brain's adaptation to rhythmic communication sounds would thus be possible, and certainly an interesting topic for future studies.

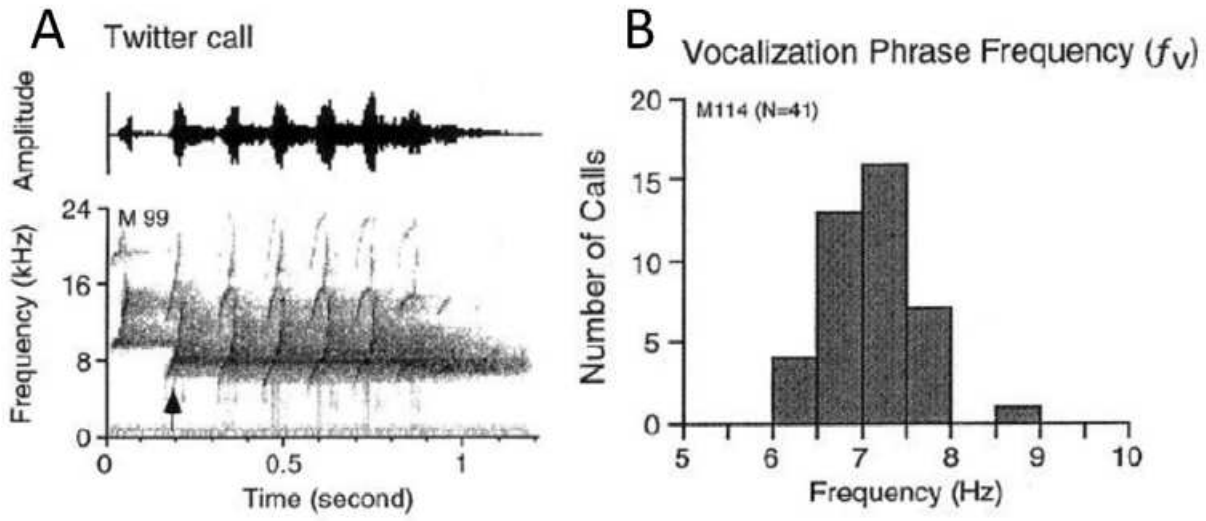


Figure 1. A. The call of a marmoset monkey (waveform: top; spectrum: bottom) is highly rhythmic. B. The amplitude spectrum of the envelope of this call (i.e. the repetition rate of "phrases") is very similar to that for the human speech envelope, showing a dominant frequency range between 6 and 8 Hz. Modified from Wang et al., 1995.

The second point again involves the constructed speech/noise snippets – but this time, we are going back to the beginning of this thesis, where perceptual cycles were introduced in Chapter 1. In the latter chapter, it was shown that *perceptual echoes*, an oscillatory reverberation of information in the brain for up to one second, could be found for the visual, but not the auditory system (İlhan and VanRullen, 2012; VanRullen and Macdonald, 2012; VanRullen et al., 2014). Due to their periodic nature, perceptual echoes can be taken as evidence for perceptual cycles, and their absence in the auditory system can be interpreted

as a lack of perceptual cycles in the auditory domain. However, in a previous study (İlhan and VanRullen, 2012), perceptual echoes have been tested in the auditory system by the use of pure tones as stimuli: If perceptual cycles indeed operated on a relatively high level of processing, as it was argued in this thesis (Chapters 1-2), it might not be possible to reveal these echoes with such "low-level" stimuli – instead, speech sound as a stimulus that is processed on a relatively late stage within the auditory system (Davis and Johnsrude, 2003) might be more appropriate. Previously, though, it was not possible to test auditory perceptual echoes using speech sound: The spectral content of the stimulus must be stable over time, as otherwise "echoes" can be introduced simply by changes in the spectral properties of the stimulus. However, having constructed speech/noise sound without systematic changes in spectral content, these stimuli can now be used in order to test whether an auditory perceptual echo can be revealed in response to speech sound. This is a project for future studies, but a glimpse into potentially interesting preliminary results can be offered here. It should be mentioned that, if there were an auditory perceptual echo, it is expected to be much shorter than the visual one: As auditory stimuli fluctuate rapidly over time (in contrast to visual ones), the auditory system needs to remain flexible and process input as fast as possible in order to avoid losing track of the rapidly changing input stream. Indeed, the duration of the influence of a given moment of speech sound on neural oscillations or perception, as described in many chapters of this thesis, was consistently restricted to one cycle of the speech envelope (see, e.g., Figure 4 in the article of Chapter 4). Additionally, the autocorrelation of a common speech envelope (thick black line in Figure 4 in the article of Chapter 6) revealed that speech sound is not correlated with sound of the same input stream if the latter precedes (or succeeds) the former by more than one cycle of the speech envelope. Based on this notion, it might not be optimal to compare a potential

auditory perceptual echo with the one observed in vision. Rather, it might be more appropriate to compare it with another auditory stimulus which is known not to introduce these echoes: pure tones. The reasoning underlying experiments of perceptual echoes is shortly summarized in the following: The brain can be seen as a filter that processes its input and generates its output (the latter measured, for instance, as an EEG signal), and our goal is to estimate this filter. However, if the input contains a dominant frequency in its power spectrum, the output is biased towards this frequency, independent of the characteristics of the filter. Similarly, if the input does not contain a specific frequency in its spectrum, it is impossible to find it in the output, even if the filter naturally operates in this frequency range. Thus, only when the input that is used for filter estimation contains equal power at all frequencies of interest, the "brain filter" can be reliably estimated. Thus, both a pure tone (440 Hz) and the constructed speech/noise stimuli were amplitude-modulated with amplitude value sequences which were constructed to have equal power at all frequencies up to 80 Hz. The EEG in response to these stimuli (amplitude-modulated pure tone or amplitude-modulated speech/noise stimuli) was then cross-correlated with the amplitude values that were used for modulation (and not with the stimuli *per se*[1]). Therefore, the stimuli (sine tone vs. speech/noise) could not bias the estimated "brain filter" – rather, they provided a different "background" or "context" for stimulus processing. We were thus able to see whether the estimated "brain filter" – reflecting the perceptual echo – differs with respect to the provided stimulus context (low-level vs. high-level). Preliminary results are shown for one subject in Figure 2. The response in both conditions (pure tone and speech/noise) is shown as the standard deviation of cross-correlation across channels (as in

[1] Note that, even though the stimuli themselves were not used for cross-correlation, they nevertheless need to exhibit a stable spectral content over time, as otherwise the amplitude values used for modulation would be biased.

Figure 3 in the article of Chapter 5) in Figure 2A. Interestingly, a later peak is visible for the speech/noise condition that is completely absent for the pure tone condition – a difference between the two standard deviations is shown in Figure 2B. A time-frequency transformation of the difference of cross-correlation between the two conditions (Figure 2C), averaged across channels, reveals that this peak is spectrally located in the theta-band (~4 Hz). The timing and frequency of this peak is strongly reminiscent of the "high-level component" of entrainment that we found in the EEG in response to the constructed speech/noise stimuli (Chapter 5). In that study, we found an early gamma-component that was restricted to everyday speech, and a later theta-component that was present for both everyday speech and speech/noise conditions in which systematic low-level features had been removed. Thus, although speculatively, first results are promising and might suggest that high-level stimuli are able to evoke a short auditory perceptual echo that is primarily located in the theta-band. In a future version of this study, several things need to be investigated: First, results shown in Figure 2 should be replicated using more subjects. Second, in Figure 2B,C, it is unclear whether an additional effect is present in the lower gamma-band (here ~25 Hz) – for instance, pure tones might have evoked only one "gamma-cycle", whereas speech/noise stimuli might have resulted in more than one cycle. Finally, it would also be interesting to use a stimulus that is even more comparable with the constructed speech/noise sound: For instance, an auditory perceptual echo could be tested in response to pure noise having the same spectral content (but not containing the high-level features that are specific for speech) as the constructed speech/noise sound.

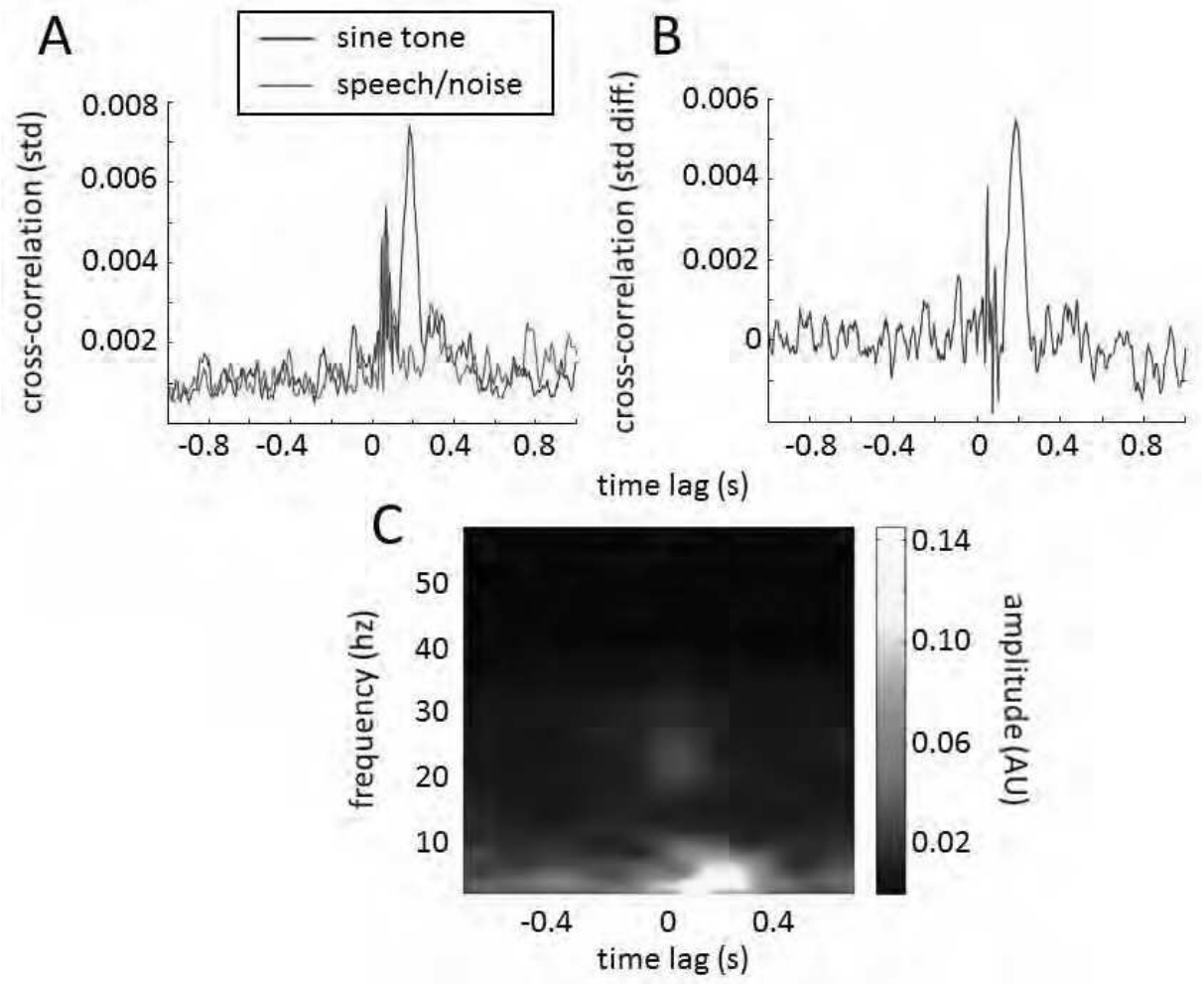


Figure 2. Cross-correlations between recorded EEG signal (separately for each channel) and random amplitude modulations of a sine tone (red) or speech/noise sound (blue) were calculated. A. The standard deviation across channels shows only an early peak for the sine tone condition, but an additional later peak for the speech/noise condition which is absent for the former. B. The difference between these standard deviations again reveals the later peak, but also suggests that the early peak could be longer for the speech/noise sound than for the pure tone. C. A time-frequency transformation of the difference of cross-correlations between conditions (averaged across channels) reveals a pronounced difference that is located in the theta-band (~4 Hz) and might reflect an auditory perceptual echo that can be introduced by a "high-level context". Note that this time-frequency representation also shows a weak difference in the lower gamma band (~25 Hz, at time lag ~80 ms) that needs further investigation.

GENERAL DISCUSSION

This thesis is concluded by a general view of oscillatory mechanisms of stimulus processing and selection. The following article combines all results obtained in this thesis and integrates them into existing literature and into a more global view. This chapter "closes the circle": We are coming back to the question of similarities and differences between audition and vision, in particular in the view of periodicities in brain and behavior. Compared to the General Introduction, we will go one step further and formulate more general questions:

- (a) What is the "default" frequency of stimulus processing (or: of perceptual cycles) in vision and audition?
- (b) Can the two systems adapt to their environment – and what happens when the input is unpredictable?
- (c) Does the mechanism of input selection (or: of subsampling) depend on the input itself or is it relatively inflexible or independent?
- (d) Where in the hierarchical organization of the two systems is input selected, prioritized or suppressed? Where in the hierarchical organization do perceptual cycles operate?

Based on results obtained in the first two chapters of this thesis, it is argued that oscillatory mechanisms in the auditory system should operate on a hierarchically elevated level of processing (d). Based on results obtained in Chapters 3 to 7, it is argued that the entrainment of neural oscillations (reflecting an adjustment of perceptual cycles) is a particularly fundamental mechanism of the auditory system (b). Answers to the other questions are partly based on the current literature and partly speculative. More precisely, it is argued that the alpha band (8-12 Hz), reflecting perceptual cycles in vision (Chapter 1), is

the dominant frequency of oscillatory processing for the visual system whereas that for the auditory system varies according to the rhythm of its input (a). If the input is unpredictable, the visual system can afford a "blind" subsampling of its input at a frequency corresponding to the alpha band (Chapter 1), whereas the auditory system might either operate in a continuous mode where oscillatory processes are suppressed or – speculatively – switch to a state of internal attention that operates at alpha frequency, but decoupled from stimulus input (Lakatos et al., submitted; b,c). In contrast to the auditory system (at least in the case of non-rhythmic input), the visual system could relatively easily afford oscillatory processing on a hierarchically low level, but results remain unclear (d). The following opinion article also provides experimental approaches for future studies that logically follow from both the questions asked above and from results obtained in this thesis.

In sum, this chapter – and this thesis overall – should hopefully leave the reader with the impression of a particularly important role of neural oscillations for stimulus processing and selection, tightly linked to periodicities in our environment, behavior, and perception: perceptual cycles.

Article:

Zoefel B, Lakatos P, VanRullen R (in preparation) Oscillatory mechanisms of stimulus processing and selection in the visual and auditory systems: A comparison.

Oscillatory mechanisms of stimulus processing and selection in the visual and auditory systems: A comparison

Authors: Benedikt Zoefel[a,b,c*] , Peter Lakatos[c], and Rufin VanRullen[a,b]

Affiliations: [a] Université Paul Sabatier, Toulouse, France

[b] Centre de Recherche Cerveau et Cognition (CerCo), CNRS, UMR5549, Pavillon Baudot CHU Purpan, BP 25202, 31052 Toulouse Cedex, France

[c] Nathan Kline Institute for Psychiatric Research, Orangeburg, NY, United States

*Corresponding author: Benedikt Zoefel
Centre de Recherche Cerveau et Cognition (CerCo)
Pavillon Baudot CHU Purpan, BP 25202
31052 Toulouse Cedex
France

Phone: +33 562 746 131
Fax: +33 562 172 809
Email: zoefel@cerco.ups-tlse.fr

Number of pages: 27
Number of figures: 1

Key words: oscillation, attention, vision, audition, perception, alpha, entrainment
Running title: Oscillations in vision and audition

Acknowledgements: This study was supported by a Studienstiftung des deutschen Volkes (German National Academic Foundation) scholarship to BZ, NIH R01DC012947 to PL, and a EURYI Award as well as an ERC Consolidator grant P-CYCLES under grant agreement 614244 to RV.

Conflict of Interest: The authors declare no competing financial interests.

Introduction

Imagine looking for someone in a crowd, trying to keep the person's characteristics in mind while suppressing other, potentially distracting events: Constantly bombarded with a continuous stream of influx, our brain needs to select, filter and prioritize, and the use of top-down processes for this task might be indispensable. Recent research suggests that *neural oscillations*, rhythmic fluctuations in the excitability of neural populations, are the brain's key feature in this process: Events that coincide with the oscillation's high excitability phase are amplified whereas events occurring during the low excitability phase are suppressed [1]. The possibility of the brain to *control* these oscillations, aligning high and low excitability phases with relevant and irrelevant events, respectively, makes them a powerful tool to gate and filter input [1]. This mechanism can also be seen as a way of environmental subsampling: "Snapshots" of the environment are taken at a rate that corresponds to the frequency of the respective oscillation and the moment of the "snapshot" might be optimized by an alignment of neural oscillations with external events (for a review, see [2]). Moreover, the oscillatory power can impact the overall responsiveness of a given brain region, a mechanism that has been associated with a modulation of the neural firing rate [3,4]. An important role of neural oscillations for attentional selection and stimulus processing has been shown across modalities: For the visual [5], auditory [6], somatosensory [3], motor [7] and olfactory system [8]. Whereas the basic mechanisms, common across modalities, are relatively well understood [1,9,10], only recently differences between modalities began to emerge [2,11]. In this opinion paper, we will focus on the two modalities that are arguably the most important ones for the human brain: vision and audition. We will summarize the most important findings of neural oscillations

involved in stimulus selection and processing, and highlight the differences between the two systems, probably arising from fundamental differences in the nature of the input that is sampled and processed in the two systems. We will concentrate on several questions and answer them in separate sections, in direct comparison of the two modalities: What is the "default" frequency of stimulus processing in vision and audition? Can the two systems adapt to their environment – and what happens when the input is unpredictable? Does the mechanism of input selection depend on the input itself or is it relatively inflexible or independent? Where in the hierarchical organization of the two systems is input selected, prioritized or suppressed? Answering these questions has critical implications for our understanding of neural oscillations involved in attention and stimulus selection. As we will see, significant progress has been made in the last years, but new questions arise with the increased knowledge. Those questions are also addressed in this paper. Several hypothetical answers are provided that are partly based on previous findings and partly, as we emphasize here, on speculation. Experimental approaches are discussed that are necessary to investigate the proposed hypotheses.

Frequency of stimulus processing

There is overwhelming evidence for the alpha band (7-13 Hz) as the principal frequency range of stimulus processing in the visual system (Fig. 1A). This observation was already published by Hans Berger in 1929 [12] who reported a dependence of alpha power on the visual input: Alpha power in the EEG increases when subjects close their eyes. Since then, both theoretical and experimental approaches provided convincing evidence that the alpha band is related to an

inhibition (or disengagement) of brain regions [13–16]: For instance, alpha power increases in the hemisphere that is ipsilateral to an attended stimulus [17,18], or in brain regions not involved in the current task [19]. Moreover, it has been shown that visual perception is directly related to the alpha band: Both detection of a visual target and the likelihood of the perception of a phosphene during transcranial magnetic stimulation (TMS) depend on the EEG alpha phase [20–22] (Fig. 1Ab) and power [23,24] (Fig. 1Aa) and random visual input seems to reverberate in the brain at a frequency corresponding to the alpha band [25] (Fig. 1Ac). Similarly, the strongest neural resonance in response to stimulation by both rhythmic visual input [26,27] and electric current [28,29] is observed in the alpha band, indicating that the intrinsic frequency of neurons [30] in the visual system is indeed located predominantly in the alpha band. Finally, both the probability of detecting a visual stimulus after a cue (Fig. 1Ad) and the following reaction time fluctuate periodically [31,32]. These fluctuations have been found at a frequency of 4 Hz *per visual hemifield*, indicating an overall rhythmicity of 8 Hz, thus lying within the alpha band [33]. Although neural activity in the gamma band (~30-70 Hz) has often been reported in the visual system, gamma-band activity might be tightly linked ("coupled") to the alpha band, with the phase of the latter controlling the instantaneous amount of gamma power [34–36]. One important distinction should be made here: Whereas some studies report effects in the alpha band around 10 Hz, linked to a topographical distribution that is centered on the occipital lobe (e.g., Fig. 1Aa,c), the peak frequency of the effect described in other studies seems to be somewhat lower and located in more frontal regions (7-8 Hz; e.g., Fig. 1Ab,d). It is not unlikely that the two types of effects stem from different generators of oscillatory processing, and they will play a role in the following paragraphs. Also, it is unclear whether a frequency of 7-8 Hz can

be assumed to reflect "textbook alpha" (or whether it is rather part of the theta band) – nevertheless, for the sake of simplicity, in the following, we will designate both bands as "alpha", but differentiate between an "occipital alpha" (~10 Hz) and "frontal alpha" (~7-8 Hz).

The dominant frequency of stimulus processing in the auditory system is less clear than in the visual one: On the one hand, many studies describe an alignment between neural oscillations in the delta/theta band (~1-8 Hz) and rhythmic stimulation [1,5,6] (Fig. 1Bb) and this alignment can decrease reaction time [6], increase efficiency of stimulus processing [37] and seems to be present even after stimulus offset [38,39] (or when subjects do not consciously perceive the stimulus, ruling out contamination by evoked potentials; [40]; Fig. 1Bb). On the other hand, the alpha band seems to be important as well [41–43]: Alpha power can be modulated by auditory attention [44,45], similar as in the visual system, speech intelligibility co-varies with alpha power [46,47], and the phase of the alpha band modulates auditory stimulus detection if entrained by transcranial alternating current stimulation (tACS; [48]). Moreover, the power of the gamma band can be coupled to any of these bands [49,50]. Although the auditory system seems to "resonate" most strongly in the 40-Hz (i.e. gamma) range [51], several studies suggest that similar phenomena can be found in lower frequency bands as well [e.g., 48] and human auditory perception is most sensitive to amplitude fluctuations and frequency modulations at a frequency of ~4 Hz [53]. Thus, it is difficult to determine a distinct frequency of stimulus processing in the auditory system. Instead, the auditory system might utilize different frequencies for different purposes, and the reported results have to be seen in relation with the respective stimulation protocol, as argued in the following section.

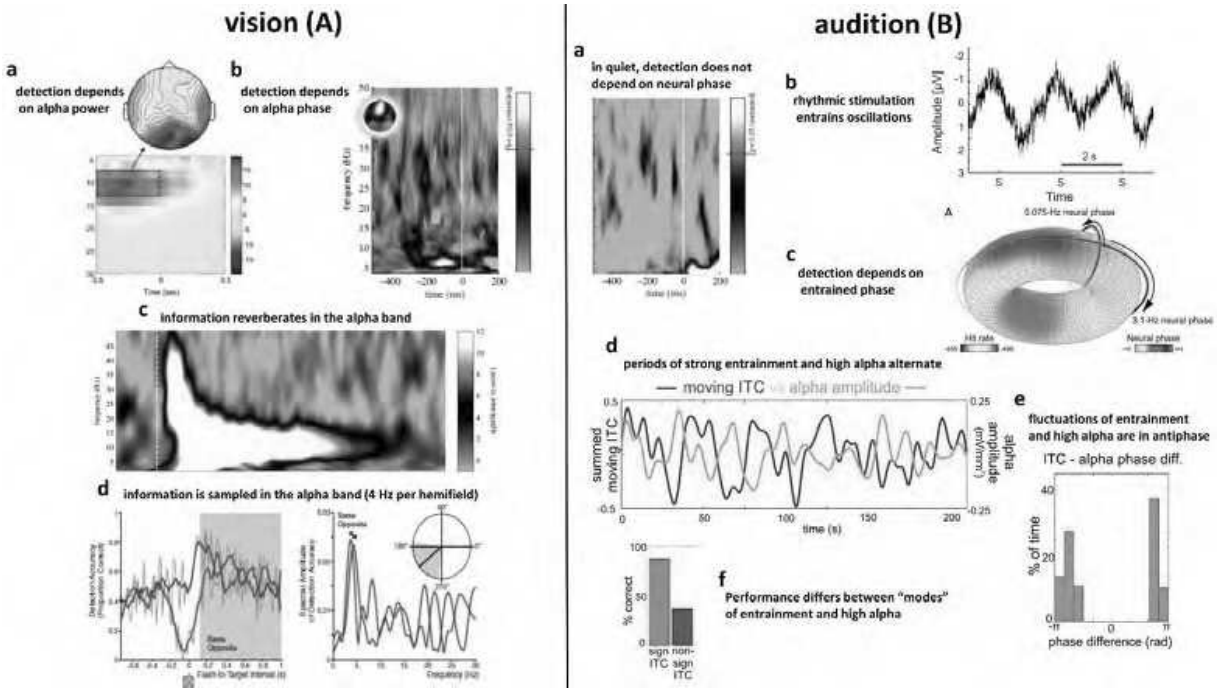


Figure 1. Overview of some fundamental results underlying the proposed role of neural oscillations for stimulus selection and processing. The alpha band seems to be the *oscillation for vision (A). The difference in EEG power (color-coded) around target onset is shown between subjects that did not perceive near-threshold visual targets and those that did (Aa; from* [24]*). Results indicate that visual detection depends on alpha power, with lower power leading to an improved detection. The detection of a weak visual target also depends on the phase of the alpha band, as measured in the EEG (Ab; from* [2]*, the original data is presented in* [20]*). The strength of modulation of target detection by the EEG phase in the respective frequency band is color-coded; the significance threshold is marked on the colorbar. When a random luminance sequence is presented to human subjects and their EEG is recorded in parallel, a reverberation ("perceptual echo") of this visual information can be found in the electrophysiological signal for up to one second (using cross-correlation between luminance sequence and EEG), but only in the alpha band (Ac; from* [2]*, the original data is presented in* [25]*). Moreover, after a visual stimulus cues attention to one visual hemifield, the probability of detecting a succeeding target fluctuates rhythmically, and in counterphase depending on whether the target occurred in the same or opposite hemifield (Ad, left; from* [31]*). This "visual rhythm" fluctuates at 4 Hz per visual hemifield (Ad, right), indicating an overall sampling rhythm of 8 Hz, thus lying within the alpha band. Note that some effects (a,c) seem to have a somewhat lower frequency than others (b,d), leading to the distinction between an "occipital alpha" (~10 Hz) and a "frontal alpha" (~7-8 Hz) in this paper. In audition (B), detection of a near-threshold target is independent of the EEG phase when presented in quiet (Ba; from* [2]*; the color-code corresponds to that in Ab). However, it is a widespread phenomenon that oscillations entrain to rhythmic stimulation. Shown exemplarily is the data from a study in which a train of pure tones, with a repetition rate of 0.5 Hz, has been presented to human subjects, and the EEG was recorded in parallel (Bb; from* [40]*). The amplitude of the tones was set to a near-threshold level and subjects had to press a button whenever a tone was detected; the plot shows EEG data, averaged across subjects, in response to three subsequently missed targets (denoted "S"). An oscillatory signal, entrained to the rhythmic stimulation, is apparent – as subjects did not consciously perceive the stimulation, a potential contamination by evoked potentials introduced by the stimulation can be ruled out. When subjects have to detect targets (gaps) that are embedded in an entraining stimulus (here: simultaneously modulated in frequency and amplitude), detection depends on the phase of the thus entrained oscillations (Bc; from* [54]*). Detection probability of the target is color-coded, whereas the outer and inner circles of the toroid represent the neural phases at the two entrained frequencies. As it can be seen, there is a combination of neural phases where detection is best, and one where detection is worst. Moreover, the auditory system seems to be able to switch between a "rhythmic mode", in which processing is determined by oscillations corresponding to the input rate of the entraining stimulus, and an "alpha mode", in which alpha oscillations dominate the processing. During rhythmic stimulation, large fluctuations in the amount of phase entrainment (indicated by the amount of phase-locking across trials, shown in black) and alpha power (light blue) exist (Bd; from* [55]*). Importantly, periods of pronounced entrainment and of high alpha power alternate (Be; from* [55]*), suggested by a phase opposition between the two functions shown in Bd. This finding was interpreted as alternating periods of external and internal attention. During periods of high phase-locking across trials,*

performance in a target detection task is improved compared to periods of low-phase locking (Bf; from [55]*), indicating an important role of the two "modes" ("entrainment mode" and "alpha mode") for stimulus processing. However, note that this effect has only been shown for rhythmic stimulation: In this paper, we hypothesize that processing in the "alpha mode" might be generalized to input in which no regular structure can be detected, and this speculation requires further experiments (cf. box 1).*

Relation to the system's input

It has recently been shown that a discrete sampling of the visual environment might not be disruptive to stimulus processing even when it is done independently of the input: "Blindly" subsampling (i.e. "snapshots" are taken independently of the input's content) videos of sign language on a level that corresponds to the very input of the visual system (i.e. on a frame level) is not particularly harmful to visual recognition performance, even at low subsampling frequencies (< 10 Hz), and much less disruptive for performance than a corresponding subsampling procedure for the auditory system [2]. Thus, the visual system might maintain its rhythm of stimulus processing even when it cannot be adjusted, such as during an unpredictable sequence of events. The alpha band seems to be the dominant frequency of stimulus processing in the visual system both in the presence [26,27] and absence [12,20,25] of rhythmicity in the environment. Alpha oscillations in the visual system have been found to adjust when the onset or spatial location of expected upcoming events is known, but no external rhythm is present: For instance, the alpha lateralization effect described above is influenced by the predictability of the spatial location of the target, indicating an active adjustment of alpha power based on anticipatory spatial attention [56–58]. Moreover, Bonnefond and Jensen [59] showed an adjustment of both alpha power and phase prior to the expected onset of a distractor in a visual working memory task, and Samaha and colleagues [60] demonstrated an improvement in performance in a visual discrimination task when the alpha phase was adjusted to the expected target onset. In the absence of regular stimulus timing

(indeed, stimulus timing was predictable, but not rhythmic[1] in [56–60]), there is not much evidence of other frequency bands adjusted to expected events or location, indicating that the alpha band is indeed the preferred frequency of stimulus processing for the visual system. It is of note that – of course – rhythmic stimuli (such as visual flicker) introduce a rhythmic component in the recorded signal whose frequency corresponds to the stimulation frequency (i.e. steady-state evoked potentials; [27]) and phase entrainment has been demonstrated for the visual system [5,61,62]. However, evidence for phase entrainment at frequencies beyond the alpha band remains sparse – for instance, steady-state potentials obtained in response to flicker show a prominent peak at 10 Hz [27] – and is often paired with auditory stimulation. Moreover, in contrast to the auditory system, visual events are rarely cyclic (indeed, flickering stimuli are rare in a natural visual environment), but rather restricted to a specific moment in time. Based on this notion, we suggest that, instead of entraining, the visual system mostly *adjusts* its oscillations to upcoming events. We acknowledge that this notion is speculative; it is therefore discussed in more detail in box 1 and in the final section of this article. A phase-reset prior to or at the moment of the expected event might be an important tool for this adjustment [63]. Another possibility is that the visual system does not prioritize adaptation to stimulation in time, but rather in the spatial domain. It might thus be more important for the visual system to precisely localize its oscillations (for instance by changing the speed of a traveling alpha wave; [35]) rather than to change their frequency, as the latter is, by definition, a temporal

[1] Note that this differentiation is kept up throughout the manuscript: Whereas *adjustment* to a stimulus is defined as an adaption of oscillatory parameters to the timing of anticipated event, *entrainment* involves an (additional) inherent regularity of the stimulus to which the oscillation can be aligned.

parameter. Thus, whereas phase entrainment might be an important and highly developed tool for the auditory system (as outlined below), this might not be the case for the visual one.

In contrast to the visual system, time is one of the most important features for the auditory system [2,64]. The need of the auditory system to adapt to the temporal structure of its input might thus be greater than for the visual one[2]. As shown in psychophysical experiments [2], "blind" subsampling of the environment might not be possible for the auditory system, as the temporal structure of the input might be destroyed. Due to this increased demand of temporal flexibility, the auditory system might make use of the different temporal scales provided by the brain: Neural oscillations cover a wide temporal range [66,67], cycling at intervals between seconds (infraslow, 0.1 Hz) and several milliseconds (high gamma range, > 60 Hz). Moreover, auditory stimuli are often rhythmic, making neural oscillations a valuable and convenient tool for synchronization with the environment [1]. This notion might explain the variety of findings described in the previous section: In contrast to the visual system, the frequency of operation might strongly depend on the input to the system in the auditory case. Many environmental sounds, including speech sound, contain amplitude fluctuations in the range of the delta/theta band. It is possible that one of the "preferred" rhythms of the auditory system includes this

[2] We note here that, for the visual system, saccades introduce "chunks" of input arriving at a frequency of ~2-3 Hz [65] that could be considered "snapshots" of the environment and result in a temporal structuring of the visual input as well. However, we emphasize that saccades are *initiated* by the brain: The timing of incoming information is thus known in advance – e.g., via feedback from the motor system. Therefore, we argue that, in the visual system, it might not be necessary to adapt stimulus processing to the input *per se*, but rather to the (rather irregular) scanning of the environment introduced by eye movements. Moreover, as the visual input "changes" every ~300-500 ms (induced by a saccade) but is rather stable within this time interval, it is not essential to process (or sample) the input at the moment of the saccade (it can be processed anytime within the ~300-500 ms interval). At the same time, this might be a reason why the visual "sampling rhythm" (assumed here as ~10 Hz), is faster than the saccadic rate: In this case, even "blind" sampling would not result in a loss of information (i.e. in the loss of one of the 300-500-ms "chunks"). Finally, we note that discrete sampling (via neural oscillations) in the visual system might even have evolved as a "shortcut" to generate "snapshots" of the environment without the metabolic costs of eye movements.

frequency range [53], explaining the multitude of studies reporting an alignment of delta/theta oscillations with environmental rhythms. In a multi-speaker scenario or when speech sound is mixed with noise, the alignment between these oscillations and the envelope of speech sound is increased for attended speech, suggesting a mechanism of auditory stream selection [68,69]. Thus, phase entrainment might be one of the key features of stimulus selection in the auditory system. If phase entrainment is impossible, due to a non-rhythmic stimulation (or due to an absence of attention, see below), it is possible that the auditory system switches to a "mode" where low-frequency oscillations, important for the adaptation to rhythmic stimulation, are suppressed [1,45]. Instead, alpha oscillations might become the dominant frequency of stimulus processing: Evidence for this was reported by Lakatos and colleagues [55], showing that in monkey primary auditory cortex, periods of phase entrainment alternate with periods of high alpha power (Fig. 1Bd,e). Both gamma-power and multi-unit activity (an index of neuronal firing) were coupled to the dominant oscillation: To the entrained phase when phase entrainment was strong, and to the alpha phase when alpha power was high, but entrainment was weak. Detection of deviants in an auditory sequence was significantly better in the state of strong phase entrainment than in an assumed "alpha-mode" (Fig. 1Bf), indicating that the auditory system might be "decoupled" from external input whenever alpha power is high. Indeed, in contrast to the visual system, where target detection depends on the alpha phase [20], auditory detection is independent of the oscillatory phase in quiet [40] (Fig. 1Ba), but this effect can be introduced when the auditory background or electrical stimulation is rhythmic [47,67,68] (Fig. 1Bc). If no regular temporal structure is present but the onset of an expected auditory target is known, some studies reported an adjustment of alpha power to this target

[41,72,73], such as described for the visual system, although evidence remains sparse and most paradigms focused on multimodal or (audio)spatial attention (reviewed in [15]) (the adjustment of phase has not been shown for the auditory system yet). Moreover, a change in alpha power in the auditory system can lead to illusionary phenomena, such as tinnitus [74]. Thus, the auditory system might be able to "switch" between a "mode" of processing that is tuned to the temporal structure of the input (with a bias for lower frequencies, due to their dominance in the auditory environment) and another, more internally oriented "mode" where alpha oscillations represent the dominant processing frequency. Interestingly, these two "modes" resemble two cortical states of primary auditory cortex that have recently been described [75]: A "synchronized state" that is relatively independent of sensory input (corresponding to the described "alpha-mode") and a "desynchronized state", where the processing of input sounds is precise and reliable (corresponding to the "entrainment" or "rhythmic mode").

Is this duality of oscillatory stimulus processing "modes" unique to the auditory system? Here, we argue that this is not the case, a notion that leads us back to the differentiation between "occipital alpha" and "frontal alpha" for the visual system. It has been argued before that the "classical" (occipital 10-Hz) alpha might serve the purpose of "saliency detection" [4]: The higher the alpha amplitude, the lower overall neuronal excitability, and the more difficult for a stimulus to reach consciousness. Thus, in an *unattended* visual scene (which leads to an increased alpha amplitude, as outlined above), occipital alpha might at the same time enable functional deactivation, but, given that an unattended stimulus is salient enough, also enable the system to switch attention to a potentially important event. This "occipital alpha" mode might be similar to the auditory "alpha mode" described in the previous paragraph: In the

absence of attention, both systems might switch to a mode that is relatively independent from stimulation, and this switch can be reversed by an event that is salient enough to overcome the inhibitory effect of an increased alpha amplitude. In contrast, attention seems to be a prerequisite for a modulation of performance by "frontal alpha" in vision: Visual detection only depends on the EEG phase at 7 Hz (Fig. 1Ab) if the stimulus is attended [76], and the observed periodicity in reaction time after a cue (Fig. 1Ad) depends on the attended visual hemifield [31]. Therefore, only stimuli that are located in the focus of visual attention seem to be sampled at a frequency of 7-8 Hz, and this sampling frequency is independent of stimulus input. As developed in the preceding section, this is a clear contrast to the auditory system, where – in the presence of attention – the adaption (i.e. phase entrainment) to the frequency of stimulation seems to be a prerequisite for an efficient stimulus processing. In the case of the auditory system, it is unclear how the dominant frequency of stimulus processing changes if no regular structure is present in the input but attention is focused on the auditory environment. We emphasize that the switch between "entrainment-mode" and "alpha-mode", as described above [55], has so far only been demonstrated during rhythmic stimulation. It was speculated that the "alpha-mode" can be activated – despite the regular stimulation – due to lapses in attention to external stimuli, leading to an increase of internal attention. However, in principle, a dominance of the alpha band when external input is (supposedly) ignored – and therefore virtually "*absent*" for the brain – might also mean that the alpha band dominates in the "true" *absence* of regular input. Thus, one possibility would be that, in the auditory system, the switch from "entrainment-mode" to "alpha-mode" can be generalized to a larger scheme and corresponds to a switch in processing mode for regular vs. irregular stimulation. Another

possibility would that the auditory system changes to a mode of continuous processing in which sampling mechanisms of neural oscillations are suppressed. We acknowledge that this notion is speculative and therefore discuss it in more detail in box 1.

Hierarchical level of processing

Subsampling the visual stream at a rhythm that corresponds to the alpha band would not result in a significant loss of information even when it is done at the earliest level of visual processing [2]. Thus, in principle, the visual system could afford oscillatory processes at low levels of the visual hierarchy. Recent research associated the alpha rhythm with top-down ("high-level") processing [13]; however, it is unclear which stages of visual processing are affected by this top-down control. There is accumulating evidence for feedback from second-order visual areas to primary visual cortex that is transmitted in the alpha band [77,78]. It remains to be shown whether subcortical areas, such as the thalamus [79], are also affected. Moreover, it has been hypothesized that the pulvinar plays an important role in synchronizing cortical regions at the alpha frequency [80]. The detailed interplay between these regions has yet to be investigated.

As mentioned above, the alpha band might be the dominant frequency of stimulus processing for the auditory system in the absence of attention (or in the absence of rhythmic stimulation), but "blind" subsampling of the environment might be destructive for the system. Thus, it is possible that auditory input is processed in the alpha band only after reaching a certain hierarchical stage of the auditory system. Whereas early auditory representations seem to entail all acoustic details, later representations are categorical [81], thus more stable, and might be more robust against harmful effects of "alpha-subsampling". Indeed, it has been shown that

recognition of subsampled speech sound is improved when the subsampling is done on the level of auditory features (i.e. on a cortical representation) and not on the very input to the system (i.e. on a cochlear level) [82]. The anterior temporal cortex seems to be a likely candidate for processing in the “alpha-mode”, due to its relatively long window of temporal integration, achieving greater abstraction (i.e. independence) from the acoustic input than other auditory regions [81]. However, we note that it has been speculated that alpha might already play an important role at the level of primary auditory cortex [41]. If the auditory stimulation is rhythmic and predictable (and attended), environmental subsampling is not “blind” anymore: Neural oscillations are aligned to the rhythmicity, efficiently selecting relevant moments (which are aligned with the high excitability phase) whereas irrelevant information (aligned with the low excitability phase) is suppressed [1]. This mechanism avoids a loss of relevant information and could therefore take place on a hierarchically lower level of stimulus processing [2]. Indeed, subcortical regions, such as cerebellum and basal ganglia, seem to be involved in the synchronization between brain activity and rhythmic input [83,84]. However, the described “bias” for slow frequencies during entrainment might be restricted to non-primary cortical areas, whereas lower areas might process stimuli at a fast rate [53]. Moreover, it has been shown that neural oscillations can entrain their phase to high-level features of speech sound, indicating that phase entrainment does entail a component of high-level processing [85–87]. Clearly, more studies are necessary to determine the respective levels of oscillatory processing in the auditory system.

Table 1. Summary of mechanisms of stimulus selection and processing in the visual and auditory systems, including the hypotheses made in this article.

	VISUAL SYSTEM	AUDITORY SYSTEM
Dominant frequency of processing	• **Alpha band (7-13 Hz): Differentiation into occipital alpha (~10 Hz) and frontal alpha (~7-8 Hz) is likely.** • **If stimulation is rhythmic and attended: Frequency of stimulation, but bias for occipital alpha** • **HYPOTHESIS: If attention is absent or directed internally: Occipital alpha** • **HYPOTHESIS: If stimulation is non-rhythmic and attended: Frontal alpha**	• **Changes with respect to stimulation** • **If stimulation is rhythmic and attended: Frequency of stimulation, but bias for slower frequencies (~1-8 Hz), as they are most prominent in natural stimuli** • **HYPOTHESIS: If attention is absent or directed internally: Alpha band** • **HYPOTHESIS: If stimulation is non-rhythmic and attended: Alpha band or non-oscillatory ("continuous") processing**
Adjustment to environment	• **HYPOTHESIS: Yes, but might be adjustment rather than entrainment**	• **Yes, alignment of oscillatory phase with the rhythmic stimulus (phase entrainment)** • **Unclear whether oscillations adjust if stimuli are predictable but non-rhythmic**
Level of processing on which oscillatory mechanisms take place	• **Oscillatory sampling on a low hierarchical level is affordable, but has not been demonstrated yet**	• **If stimulation is rhythmic: Oscillatory sampling on a low hierarchical level is affordable, but has not been demonstrated yet** • **If stimulation is non-rhythmic: Oscillatory sampling on a relatively high hierarchical level is necessary, as otherwise important information might be lost**

Summary and Conclusion

In table 1, the oscillatory mechanisms involved in stimulus processing and selection are directly compared between the visual and auditory system, as proposed in this opinion paper. Some properties might be common across all systems: Neural oscillations can be used as a mechanism of attentional selection, and both oscillatory power and phase can be used to gate stimulus input. Changes in power might reflect a rather tonic suppression of processing (e.g., in a region that is currently not involved in stimulus processing) and/or change the effectiveness of the phase of an oscillation, cycling between moments of amplification and suppression. In the absence of attention, an "alpha-mode" ("occipital alpha" in the visual system) seems to be present in both systems, and is associated with a state that is decoupled from external stimulation and in which only very salient events can overcome the increased alpha amplitude and reach consciousness. However, there are differences between the visual and auditory system: In the presence of attention, stimulus processing in the visual system might be focused on the ("frontal") alpha band, irrespective of the stimulation, whereas the dominant frequency of processing adapts to that of the environment in the auditory system. Speculatively, if stimulation is non-rhythmic, the auditory system might operate in the alpha rhythm as well, but – in contrast to the visual system – this mechanism would have to operate independently of the input to the system, as otherwise important information might be lost. An alternative would be a "continuous mode" of stimulus processing in which most oscillatory sampling mechanisms are suppressed. The oscillatory entrainment to rhythmic stimulation seems to be a fundamental feature of the auditory system, probably evolved due the rhythmic nature of the auditory

environment. Indeed, the tendency to synchronize with auditory rhythms is ubiquitous: We sing, we dance, we clap in response to music or even a simple beat [88]. Importantly, this phenomenon is much less pronounced for the visual system: We rarely synchronize with another person's walking rhythm and the urge to dance is significantly lowered when watching someone dancing without the corresponding sound. Thus, although in principle the visual system seems to be able to entrain as well, the adjustment of power and phase might be a more important feature in this system – visual stimuli are often predictable, but rarely rhythmic. Interestingly, and in line with this notion, it has been shown that the auditory system is superior to the visual one when movement has to be synchronized with a rhythmic sequence in either or both modalities [83,89,90] and auditory rhythmicity can influence the perceived flicker rate of a visual stimulus but not vice versa [91,92]. Task-irrelevant information in the auditory system impairs visual processing more strongly than vice versa if this information is of temporal nature [93]. Thus, although visual stimuli can in principle influence auditory processing and perception (potentially using alpha oscillations [94,95]) and do so even more prominently if rhythm is involved [96], a multitude of findings indicates that, indeed, the auditory system dominates the visual one in the time domain (an extensive summary of the literature on this conclusion is provided in [97]). Finally, a simple cue (*without* rhythmic component involved) is sufficient to introduce the mentioned periodic fluctuations in visual performance [31–33] – similar results have been reported for the auditory system, but so far only after the offset of a *rhythmic* stimulus [38]. For both vision and audition, it is relatively unclear on which level of the hierarchical system the oscillatory mechanisms operate. The visual system could afford sampling its input on a relatively low level, as it is relatively robust

against loss of information. The same is true for the auditory system as long as stimulation is rhythmic and important events are predictable; otherwise, the level of oscillatory selection must be relatively high in order to prevent a significant loss of information. The auditory alpha band might reflect internal processes [55], underlining the notion that oscillatory processes in the auditory system are confined to a hierarchically higher level of processing and potentially even “decoupled” from its input if the latter is unpredictable.

BOX 1: OPEN QUESTIONS, HYPOTHESES, AND EXPERIMENTAL APPROACHES

- Adjustment vs. Entrainment:
    - Phase entrainment has been demonstrated in both visual and auditory system, but adjustment might be a more important mechanism for the visual system, whereas entrainment might be more important for the auditory one. This suggestion remains mostly speculative and should be supported by experimental evidence. Here, it is critical to find a way to differentiate "true" entrainment (i.e. an oscillatory mechanism that includes predictions about the *rhythm* of the upcoming stimulation), "adjustment" (also including predictions, but rather about a single event without inherent rhythm) and a mere regular repetition of evoked neural activity by the rhythmic stimulation. One way to disentangle entrainment from the other two variations would be a demonstration of the alignment of neural oscillations to (or a modulation of behavior by) the expected rhythm after stimulus offset. Indeed, some studies already provided first promising results [38,62]. However, it needs to be shown additionally that oscillatory signals or behavior measured after stimulus onset are not simply a reverberation introduced by a phase-reset of brain oscillations by the last stimulus: Indeed, in particular in the visual domain, periodic fluctuations of performance can already be observed in response to a simple cue [31,32].
    - Findings suggesting that the auditory system is able to adjust to upcoming stimuli in the absence of rhythmic stimulation remain sparse, although alpha power seems to be the most promising candidate. Studies are necessary that systematically test the impact on neural oscillations in the two systems when rhythmic stimuli (evoking entrainment) or non-rhythmic, but predictable stimuli (evoking adjustment) are presented, potentially combining electrophysiological and behavioral measurements. It would also be interesting to see the outcome when visual and auditory stimuli are combined.
    - Although beyond the scope of this paper, auditory stimuli affect activity in the visual system, and vice versa [93-95,98,99]. Indeed, visual stimulation improves phase entrainment to speech sound [100] – interestingly, it has not been shown yet that speech sound can entrain visual cortices in turn. The oscillatory mechanisms involved in these cross-modal processes represent another exciting field of research – for instance, it needs to be determined whether stimuli of another modality can merely phase-reset (i.e. *adjust*) oscillations in primary cortical regions of a given modality, or whether "true" phase entrainment is involved

BOX 1: OPEN QUESTIONS, HYPOTHESES, AND EXPERIMENTAL APPROACHES

- "Occipital Alpha vs. Frontal Alpha" in the visual system:
    - As described throughout this article, there is relatively clear evidence of a distinction between a faster occipital, and a slower frontal alpha. However, both their functional role and their origins are poorly understood. It needs to be determined (1) whether these rhythms can co-exist, (2) how and where they are generated, (3) whether the term "frontal alpha" is justified or whether "frontal theta" would be more appropriate (and if yes, why). Experimental paradigms are needed in which the subjects' attentional resources can be modulated in a controlled way: According to our hypothesis, occipital alpha would play a most pronounced role in regions or tasks in which external attention is weak, and frontal theta would affect behavior most strongly in tasks on which visual attention is focused.

- "Entrainment vs. Alpha" in the auditory system:
    - The alternation between dominant moments of entrainment and those where alpha dominates has so far only been shown during rhythmic stimulation [55]. It is a plausible suggestion that the "alpha-mode" in the auditory system might represent lapses of external alertness and represent – similar to the "occipital alpha" in the visual system – the "default mode" of stimulus processing the absence of attention. However, it is possible that the "alpha-mode" might be reflection of a more general mode of processing that is always activated when no rhythm can be detected in the auditory environment. Another alternative would be a suppression of most oscillatory sampling mechanisms when auditory attention is focused on a non-rhythmic stimulation. Both speculations must be underlined with experimental evidence. For instance, similar analyses as in Lakatos et al. [55] might be applied in an experimental paradigm in which no regular structure is present at the input level. Intracranial recordings might be appropriate in this case, as activity in auditory cortices is, due to their nestled structure in the lateral sulcus, difficult to measure using superficial methods, such as EEG. An increase in alpha or entrained activity for irregular vs. regular stimulation, respectively, might be taken as evidence for the "alpha vs. entrainment" hypothesis described here. Another interesting approach would be the replication of previous experiments on the dependence of auditory stimulus detection in quiet on the phase of neural oscillations that so far resulted in negative results [2,40] (Fig. 1Ba), combined with an independent visual task on which the attention of the subjects is focused. The latter experimental manipulation would result in an absence of attention for the auditory stimulation. According to the hypothesis presented here, this lack of attention might provoke an increase of alpha activity in the auditory system, and result in a dependence of auditory detection on the phase of the alpha band.

BOX 1: OPEN QUESTIONS, HYPOTHESES, AND EXPERIMENTAL APPROACHES

- As mentioned above, the brain seems to be able to switch into its "alpha-mode" even though rhythmic stimulation is present. It has been speculated that this switch might reflect a change from external to internal attention [55], but evidence for this suggestion is lacking. It needs to be determined why this is the case, and what might be a trigger for this switch. Furthermore, it needs to be clarified whether the two "modes" operate on different hierarchical levels of processing (see next point).

- Hierarchical level of processing:

  - Both the visual and auditory system (the latter in the case of rhythmic stimulation) could theoretically afford oscillatory mechanisms of stimulus selection at a relatively low hierarchical stage of processing. However, this has not been demonstrated yet. Studies are necessary that systematically test the involvement of different brain regions in these processes, their preferred frequency of stimulus processing, and their interplay.

References

1 Schroeder, C.E. and Lakatos, P. (2009) Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci.* 32, 9–18

2 VanRullen, R. *et al.* (2014) On the cyclic nature of perception in vision versus audition. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 369, 20130214

3 Haegens, S. *et al.* (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

4 Jensen, O. *et al.* (2012) An oscillatory mechanism for prioritizing salient unattended stimuli. *Trends Cogn. Sci.* 16, 200–206

5 Lakatos, P. *et al.* (2008) Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science* 320, 110–113

6 Stefanics, G. *et al.* (2010) Phase entrainment of human delta oscillations can mediate the effects of expectation on reaction speed. *J. Neurosci.* 30, 13578–13585

7 Arnal, L.H. (2012) Predicting "When" Using the Motor System's Beta-Band Oscillations. *Front. Hum. Neurosci.* 6, 225

8 Kay, L.M. (2014) Circuit oscillations in odor perception and memory. *Prog. Brain Res.* 208, 223–251

9 Calderone, D.J. *et al.* (2014) Entrainment of neural oscillations as a modifiable substrate of attention. *Trends Cogn. Sci.* 18, 300–309

10 Arnal, L.H. and Giraud, A.-L. (2012) Cortical oscillations and sensory predictions. *Trends Cogn. Sci.* 16, 390–398

11 Thorne, J.D. and Debener, S. (2014) Look now and hear what's coming: on the functional role of cross-modal phase reset. *Hear. Res.* 307, 144–152

12 Berger, H. Ueber das Elektroenkephalogram des Menschen. *Arch F Psychiat* 87, 527–570

13 Klimesch, W. *et al.* (2007) EEG alpha oscillations: the inhibition-timing hypothesis. *Brain Res. Rev.* 53, 63–88

14 Jensen, O. and Mazaheri, A. (2010) Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Front. Hum. Neurosci.* 4, 186

15 Foxe, J.J. and Snyder, A.C. (2011) The Role of Alpha-Band Brain Oscillations as a Sensory Suppression Mechanism during Selective Attention. *Front. Psychol.* 2, 154

16 Klimesch, W. *et al.* (2007) EEG alpha oscillations: the inhibition-timing hypothesis. *Brain Res. Rev.* 53, 63–88

17 Thut, G. *et al.* (2006) Alpha-band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. *J. Neurosci.* 26, 9494–9502

18 Sauseng, P. *et al.* (2009) Brain oscillatory substrates of visual short-term memory capacity. *Curr. Biol.* 19, 1846–1852

19 Zumer, J.M. *et al.* (2014) Occipital alpha activity during stimulus processing gates the information flow to object-selective cortex. *PLoS Biol.* 12, e1001965

20 Busch, N.A. *et al.* (2009) The phase of ongoing EEG oscillations predicts visual perception. *J. Neurosci.* 29, 7869–7876

21 Mathewson, K.E. *et al.* (2009) To see or not to see: prestimulus alpha phase predicts visual awareness. *J. Neurosci.* 29, 2725–2732

22 Dugué, L. *et al.* (2011) The phase of ongoing oscillations mediates the causal relation between brain excitation and visual perception. *J. Neurosci.* 31, 11889–11893

23 Romei, V. *et al.* (2008) Spontaneous fluctuations in posterior alpha-band EEG activity reflect variability in excitability of human visual areas. *Cereb. Cortex* 18, 2010–2018

24 Hanslmayr, S. *et al.* (2007) Prestimulus oscillations predict visual perception performance between and within subjects. *NeuroImage* 37, 1465–1473

25 VanRullen, R. and Macdonald, J.S.P. (2012) Perceptual echoes at 10 Hz in the human brain. *Curr. Biol.* 22, 995–999

26 de Graaf, T.A. *et al.* (2013) Alpha-band rhythms in visual task performance: phase-locking by rhythmic sensory stimulation. *PloS One* 8, e60035

27 Herrmann, C.S. (2001) Human EEG responses to 1-100 Hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. *Exp. Brain Res.* 137, 346–353

28 Zaehle, T. *et al.* (2010) Transcranial alternating current stimulation enhances individual alpha activity in human EEG. *PloS One* 5, e13766

29 Thut, G. and Miniussi, C. (2009) New insights into rhythmic brain activity from TMS-EEG studies. *Trends Cogn. Sci.* 13, 182–189

30 Hutcheon, B. and Yarom, Y. (2000) Resonance, oscillation and the intrinsic frequency preferences of neurons. *Trends Neurosci.* 23, 216–222

31 Landau, A.N. and Fries, P. (2012) Attention samples stimuli rhythmically. *Curr. Biol.* 22, 1000–1004

32 Song, K. *et al.* (2014) Behavioral oscillations in attention: rhythmic α pulses mediated through θ band. *J. Neurosci.* 34, 4837–4844

33 Zoefel, B. and Sokoliuk, R. (2014) Investigating the rhythm of attention on a fine-grained scale: evidence from reaction times. *J. Neurosci.* 34, 12619–12621

34 Roux, F. *et al.* (2013) The phase of thalamic alpha activity modulates cortical gamma-band activity: evidence from resting-state MEG recordings. *J. Neurosci. 3*3, 17827–17835

35 Bahramisharif, A. *et al.* (2013) Propagating neocortical gamma bursts are coordinated by traveling alpha waves. *J. Neurosci.* 33, 18849–18854

36 Jensen, O. *et al.* (2014) Temporal coding organized by coupled alpha and gamma oscillations prioritize visual processing. *Trends Neurosci.* 37, 357–369

37 Cravo, A.M. *et al.* (2013) Temporal expectation enhances contrast sensitivity by phase entrainment of low-frequency oscillations in visual cortex. *J. Neurosci.* 33, 4002–4010

38 Hickok, G. *et al.* (2015) The Rhythm of Perception: Entrainment to Acoustic Rhythms Induces Subsequent Perceptual Oscillation. *Psychol. Sci.* 26, 1006-1013.

39 Lakatos, P. *et al.* (2013) The spectrotemporal filter mechanism of auditory selective attention. *Neuron* 77, 750–761

40 Zoefel, B. and Heil, P. (2013) Detection of Near-Threshold Sounds is Independent of EEG Phase in Common Frequency Bands. *Front. Psychol.* 4, 262

41 Strauß, A. *et al.* (2014) Cortical alpha oscillations as a tool for auditory selective inhibition. *Front. Hum. Neurosci.* 8, 350

42 Obleser, J. *et al.* (2012) Neural Oscillations in Speech: Don't be Enslaved by the Envelope. *Front. Hum. Neurosci.* 6, 250

43 Weisz, N. and Obleser, J. (2014) Synchronisation signatures in the listening brain: a perspective from non-invasive neuroelectrophysiology. *Hear. Res.* 307, 16–28

44 Kerlin, J.R. *et al.* (2010) Attentional gain control of ongoing cortical speech representations in a "cocktail party." *J. Neurosci.* 30, 620–628

45 Frey, J.N. *et al.* (2015) Not so different after all: The same oscillatory processes support different types of attention. *Brain Res.* DOI: 10.1016/j.brainres.2015.02.017

46 Obleser, J. and Weisz, N. (2012) Suppressed alpha oscillations predict intelligibility of speech and its acoustic details. *Cereb. Cortex* 22, 2466–2477

47 Wöstmann, M. *et al.* (2015) Neural alpha dynamics in younger and older listeners reflect acoustic challenges and predictive benefits. *J. Neurosci.* 35, 1458–1467

48 Neuling, T. *et al.* (2012) Good vibrations: oscillatory phase shapes perception. *NeuroImage* 63, 771–778

49 Lakatos, P. *et al.* (2005) An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *J. Neurophysiol.* 94, 1904–1911

50 Fontolan, L. *et al.* (2014) The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex. *Nat. Commun.* 5, 4694

51 Galambos, R. *et al.* (1981) A 40-Hz auditory potential recorded from the human scalp. *Proc. Natl. Acad. Sci. U. S. A.* 78, 2643–2647

52 Liégeois-Chauvel, C. *et al.* (2004) Temporal envelope processing in the human left and right auditory cortices. *Cereb. Cortex* 14, 731–740

53 Edwards, E. and Chang, E.F. (2013) Syllabic (~2-5 Hz) and fluctuation (~1-10 Hz) ranges in speech and auditory processing. *Hear. Res.* 305, 113–134

54 Henry, M.J. *et al.* (2014) Entrained neural oscillations in multiple frequency bands comodulate behavior. *Proc. Natl. Acad. Sci. U. S. A.* 111, 14935–14940

55 Lakatos, P. *et al.* (submitted) Global temporal dynamics of attention and its distinct neuronal signatures.

56 Haegens, S. *et al.* (2011) Top-down controlled alpha band activity in somatosensory areas determines behavioral performance in a discrimination task. *J. Neurosci.* 31, 5197–5204

57 Gould, I.C. *et al.* (2011) Indexing the graded allocation of visuospatial attention using anticipatory alpha oscillations. *J. Neurophysiol.* 105, 1318–1326

58 Horschig, J.M. *et al.* (2014) Alpha activity reflects individual abilities to adapt to the environment. *NeuroImage* 89, 235–243

59 Bonnefond, M. and Jensen, O. (2012) Alpha oscillations serve to protect working memory maintenance against anticipated distracters. *Curr. Biol.* 22, 1969–1974

60 Samaha, J. *et al.* (2015) Top-down control of the phase of alpha-band oscillations as a mechanism for temporal prediction. *Proc. Natl. Acad. Sci. U. S. A.* 112, 8439-8444

61 Spaak, E. *et al.* (2014) Local entrainment of α oscillations by visual stimuli causes cyclic modulation of perception. *J. Neurosci.* 34, 3536–3544

62 Gray, M.J. *et al.* (2015) Oscillatory recruitment of bilateral visual cortex during spatial attention to competing rhythmic inputs. *J. Neurosci.* 35, 5489–5503

63 Canavier, C.C. (2015) Phase-resetting as a tool of information transmission. *Curr. Opin. Neurobiol.* 31, 206–213

64 Kubovy, M. (1988) Should we resist the seductiveness of the space:time::vision:audition analogy? *J Exp Psychol Hum Percept Perform* 14, 318–320

65 Otero-Millan, J. *et al.* (2008) Saccades and microsaccades during visual fixation, exploration, and search: foundations for a common saccadic generator. *J. Vis.* 8, 21.1–18

66 Lopes da Silva, F. (2013) EEG and MEG: relevance to neuroscience. *Neuron* 80, 1112–1128

67 Buzsáki, G. and Draguhn, A. (2004) Neuronal oscillations in cortical networks. *Science* 304, 1926–1929

68 Ding, N. and Simon, J.Z. (2013) Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. *J. Neurosci.* 33, 5728–5735

69 Zion Golumbic, E.M. *et al.* (2013) Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party." *Neuron* 77, 980–991

70 Henry, M.J. and Obleser, J. (2012) Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proc. Natl. Acad. Sci. U. S. A.* 109, 20095–20100

71 Ng, B.S.W. *et al.* (2012) A precluding but not ensuring role of entrained low-frequency oscillations for auditory perception. *J. Neurosci.* 32, 12268–12276

72 Müller, N. and Weisz, N. (2012) Lateralized auditory cortical alpha band activity and interregional connectivity pattern reflect anticipation of target sounds. *Cereb. Cortex* 22, 1604–1613

73 Wilsch, A. *et al.* (2014) Alpha Oscillatory Dynamics Index Temporal Expectation Benefits in Working Memory. *Cereb. Cortex* 25, 1938-1946

74 Weisz, N. *et al.* (2005) Tinnitus perception and distress is related to abnormal spontaneous brain activity as measured by magnetoencephalography. *PLoS Med.* 2, e153

75 Pachitariu, M. *et al.* (2015) State-dependent population coding in primary auditory cortex. *J. Neurosci.* 35, 2058–2073

76 Busch, N.A. and VanRullen, R. (2010) Spontaneous EEG oscillations reveal periodic sampling of visual attention. *Proc. Natl. Acad. Sci. U. S. A.* 107, 16048–16053

77 Bastos, A.M. *et al.* (2015) A DCM study of spectral asymmetries in feedforward and feedback connections between visual areas V1 and V4 in the monkey. *NeuroImage* 108, 460–475

78 van Kerkoerle, T. *et al.* (2014) Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. *Proc. Natl. Acad. Sci. U. S. A.* 111, 14332–14341

79 Lopes da Silva, F.H. *et al.* (1997) Alpha rhythms: noise, dynamics and models. *Int. J. Psychophysiol.* 26, 237–249

80 Ketz, N.A. *et al.* (2015) Thalamic pathways underlying prefrontal cortex-medial temporal lobe oscillatory interactions. *Trends Neurosci.* 38, 3–12

81 Davis, M.H. and Johnsrude, I.S. (2007) Hearing speech sounds: top-down influences on the interface between audition and speech perception. *Hear. Res.* 229, 132–147

82 Zoefel, B. *et al.* (2015) The ability of the auditory system to cope with temporal subsampling depends on the hierarchical level of processing. *Neuroreport* 26, 773-778

83 Merchant, H. *et al.* (2015) Finding the beat: a neural perspective across humans and non-human primates. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 370, 20140093

84 Kotz, S.A. and Schmidt-Kassow, M. (2015) Basal ganglia contribution to rule expectancy and temporal predictability in speech. *Cortex* 68, 48-60

85 Zoefel, B. and VanRullen, R. (2015) Selective perceptual phase entrainment to speech rhythm in the absence of spectral energy fluctuations. *J. Neurosci.* 35, 1954–1964

86 Zoefel, B. and VanRullen, R. (in press) EEG oscillations entrain their phase to high-level features of speech sound. *NeuroImage*.

87 Zoefel, B. and VanRullen, R. (submitted) The role of high-level processes for oscillatory phase entrainment to speech sound.

88 Nozaradan, S. (2014) Exploring how musical rhythm entrains brain activity with electroencephalogram frequency-tagging. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 369, 20130393

89 Patel, A.D. *et al.* (2005) The influence of metricality and modality on synchronization with a beat. *Exp. Brain Res.* 163, 226–238

90 Repp, B.H. and Penel, A. (2002) Auditory dominance in temporal processing: New evidence from synchronization with simultaneous visual and auditory sequences. *J. Exp. Psychol. Hum. Percept. Perform.* 28, 1085–1099

91 Shipley, T. (1964) Auditory flutter-driving of visual flicker. *Science* 145, 1328–1330

92 Herrmann, C.S. *et al.* (2015) EEG oscillations: From correlation to causality. *Int. J. Psychophysiol.* DOI: 10.1016/j.ijpsycho.2015.02.003

93 Guttman, S.E. *et al.* (2005) Hearing what the eyes see: auditory encoding of visual temporal sequences. *Psychol. Sci.* 16, 228–235

94 Thorne, J.D. *et al.* (2011) Cross-modal phase reset predicts auditory task performance in humans. *J. Neurosci.* 31, 3853–3861

95 van Wassenhove, V. and Grzeczkowski, L. (2015) Visual-induced expectations modulate auditory cortical responses. *Front. Neurosci.* 9, 11

96 Oever, S. ten *et al.* (2014) Rhythmicity and cross-modal temporal cues facilitate detection. *Neuropsychologia* 63, 43–50

97 Grahn, J.A. (2012) See what I hear? Beat perception in auditory and visual rhythms. *Exp. Brain Res.* 220, 51–61

98 Lakatos, P. *et al.* (2009) The leading sense: supramodal control of neurophysiological context by attention. *Neuron* 64, 419–430

99 Romei, V. *et al.* (2012) Sounds reset rhythms of visual cortex and corresponding human visual perception. *Curr. Biol.* 22, 807–813

100 Zion Golumbic, E. *et al.* (2013) Visual input enhances selective speech envelope tracking in auditory cortex at a "cocktail party." *J. Neurosci. 3*3, 1417–1426

ANNEX: INVESTIGATING THE RHYTHM OF ATTENTION ON A FINE-GRAINED SCALE: EVIDENCE FROM REACTION TIMES

The following article is part of the *Journal Club* that enables PhD students and post-doctoral researchers to express their opinion on recent articles in *The Journal of Neuroscience*.

Article:

Zoefel B, Sokoliuk R (2014) Investigating the rhythm of attention on a fine-grained scale: evidence from reaction time. J Neurosci 17:12619-12621.

Journal Club

Editor's Note: These short, critical reviews of recent papers in the *Journal*, written exclusively by graduate students or postdoctoral fellows, are intended to summarize the important findings of the paper and provide additional insight and commentary. For more information on the format and purpose of the Journal Club, please see http://www.jneurosci.org/misc/ifa_features.shtml.

Investigating the Rhythm of Attention on a Fine-Grained Scale: Evidence from Reaction Times

Benedikt Zoefel[1,2] **and Rodika Sokoliuk**[1,2]

[1]Université Paul Sabatier, 31062 Toulouse Cedex 9, France and [2]Centre de Recherche Cerveau et Cognition (CerCo), CNRS, UMR5549, Pavillon Baudot CHU Purpan, 31052 Toulouse Cedex, France

Review of Song et al.

There is increasing evidence that the human brain does not continuously sample the environment; instead, perception is inherently rhythmic, alternating between phases of high and low receptiveness for stimulus input (Schroeder and Lakatos, 2009). This periodic modulation of neural processing might be seen as a rhythmic opening and closing of "windows of opportunity" for a stimulus to be perceived or processed (Buzsáki and Draguhn, 2004). Since perception and attention are tightly linked phenomena, these findings naturally raise the question whether attention is an inherently cyclic process as well. In addition to the fact that most evidence suggests that the periodic nature of perception relies on attentional processes (Schroeder and Lakatos, 2009), Landau and Fries (2012) directly demonstrated that attention indeed entails a rhythmic process. In their study, attention was cued to one of two possible positions and subjects had to detect a visual target that was presented at either of the two locations. Interestingly, detection probability was found to oscillate (at 4 Hz) at both locations, but, critically, in an antiphasic fashion: when the "window of opportunity" was open for one location, it was closed for the other, and vice versa, indicating that attention shifted at a frequency of 8 Hz (4 Hz per location).

In a recent publication in *The Journal of Neuroscience*, Song et al. (2014) used a paradigm similar to the one described in Landau and Fries (2012) and provided further evidence for a rhythmic component in visual attention. Again, subjects were presented with a cue that drew attention to one of two possible locations, then a visual target presented at one of the two locations had to be detected as fast as possible. Critically, the time between cue and target onset (SOA) was variable—thus, under the hypotheses of (1) rhythmic attentional sampling and (2) a reset of this sampling by the presented cue, the difficulty of target detection should not be constant with respect to time, but depend on—or covary with—the respective SOA.

The study of Song et al. (2014) is of particular interest, because it differs from the study by Landau and Fries (2012) in two points. First, instead of using the probability of target detection to investigate attention as a function of time, the relation between SOA and reaction time (RT) was examined here. Second, and most importantly, the authors present an elegant approach to characterize the rhythmicity of attention: by using time–frequency analyses, the authors were able to detect much finer changes in oscillatory variables than when using conventional analyses in the frequency domain [such as fast Fourier transformation (FFT)] that assume stationary oscillatory signals. Thus, Song et al. (2014) might have been able to detect properties of attentional sampling that remained hidden in earlier studies.

Indeed, not only did the authors corroborate the rhythmic sampling of attention, originally demonstrated by Landau and Fries (2012), they also reported an additional effect that directly depended on the phase of the sampling rhythm: the RT distribution showed an oscillatory pattern (again, around 4 Hz) that was in antiphase at cued and uncued locations and, strikingly, the power of the RT distribution in a broad frequency band (around 5–25 Hz) was phase-locked to this rhythm (i.e., RT power was highest at a certain phase and lowest at the opposite phase). The authors interpret their findings as an attentional sampling in the theta-band (3–5 Hz) that is coupled to changes in alpha-power (5–25 Hz), in line with phase-amplitude coupling of neural oscillations that is commonly found in electrophysiological recordings (Tort et al., 2010).

Although we share the excitement with which Song et al. (2014) present their work, we would like to raise two important points. First, the rhythmic fluctuations of RT were found for both cued and uncued locations, but in an antiphasic fashion. Thus, when a given SOA was favorable for detecting a target at one loca-

Received May 26, 2014; revised Aug. 6, 2014; accepted Aug. 9, 2014.

This study was supported by a Studienstiftung des deutschen Volkes (German National Academic Foundation) scholarship to B.Z.

The authors declare no competing financial interests.

Correspondence should be addressed to Benedikt Zoefel, Centre de Recherche Cerveau et Cognition (CerCo), Pavillon Baudot CHU Purpan, BP 25202, 31052 Toulouse Cedex, France. E-mail: zoefel@cerco.ups-tlse.fr.

DOI:10.1523/JNEUROSCI.2134-14.2014

 0270-6474/14/3412619-03$15.00/0

tion (reflected by a relatively low RT), it was disadvantageous for the other. Consequently, and similar to the conclusion drawn by Landau and Fries (2012), an attentional sampling of 4 Hz at one of the two locations would indicate an overall sampling rhythm at a frequency of 8 Hz. This finding is in accordance with an increasing amount of evidence for periodicity in perception that is tightly linked to brain rhythms in this frequency range for the visual system (Thut et al., 2012; Romei et al., 2012). Second, the power effect described by Song et al. (2014) covers a wide frequency range (5–25 Hz). Since, by definition, an oscillation is restricted to a narrow frequency range (Luck, 2005), we consider it problematic to assume an oscillatory component to underlie this finding, in particular because only one behavioral variable is investigated and conclusions about neural oscillations are indirect.

We propose a different scenario that could explain the data shown in Song et al. (2014). Assuming a "true" underlying attentional sampling at 4 Hz per location, it might be possible that the instantaneous state of this sampling affects the variability of responses (reflected in RT) in a cyclic manner. For instance, variability might be low when participants were attentive (i.e., at a phase of high receptiveness for visual input) and focused on the task, resulting in systematic responses with a low amount of RT fluctuations across trials. Conversely, variability might be high when participants were inattentive (i.e., at a phase of low receptiveness for visual input), resulting in unsystematic responses and a high amount of RT fluctuations across trials.

We illustrate our speculation with a (admittedly simplistic) simulation: four random signals (corresponding to the number of trials for all but one SOA in Song et al., 2014) were averaged before being transformed into the time–frequency domain. Critically, the signals were constructed such that their variance covaried with the phase of a hypothetical underlying 4 Hz wave (Fig. 1*A*) and thus, by construction, alternated between states of high and low variance. Two sets (i.e., 2 × 4) of signals were constructed, with the underlying 4 Hz wave in antiphase between the two sets, reflecting the two conditions ("valid" and "invalid") in Song et al. (2014). The average RT time courses (across "subjects"; see below) for both "conditions" (corresponding to Fig. 1*B*, top, in Song et al., 2014) are displayed in Figure 1*B*. Other properties of the simulated signals (e.g., the average) did not differ between states of high and low variance and the detailed parameters of the time–frequency analysis were as described in Song et al. (2014) (continuous complex Gaussian wavelet of order 4 for frequencies between 1 and 25 Hz in steps of 2 Hz, sampling rate 50 Hz). This procedure was repeated 49 times (corresponding to the number of subjects in Song et al., 2014) and the time–frequency transforms were averaged across "subjects," separately for each "condition." This analysis yields the RT power profiles corresponding to Figure 2*B* in Song et al. (2014). The result of our simulation is shown in Figure 1*C*, representing the difference of the simulated power profiles between the two conditions (corresponding to Fig. 2*A* in Song et al., 2014). Strikingly, this power difference exhibits regular fluctuations between positive and negative values at many frequencies, similar to the spectrotemporal profile shown by Song and colleagues (2014, their Fig. 2*A*). Thus, rhythmic changes in a broad frequency range of a given spectrotemporal profile can theoretically be explained by regular fluctuations in the variability of the underlying signals.

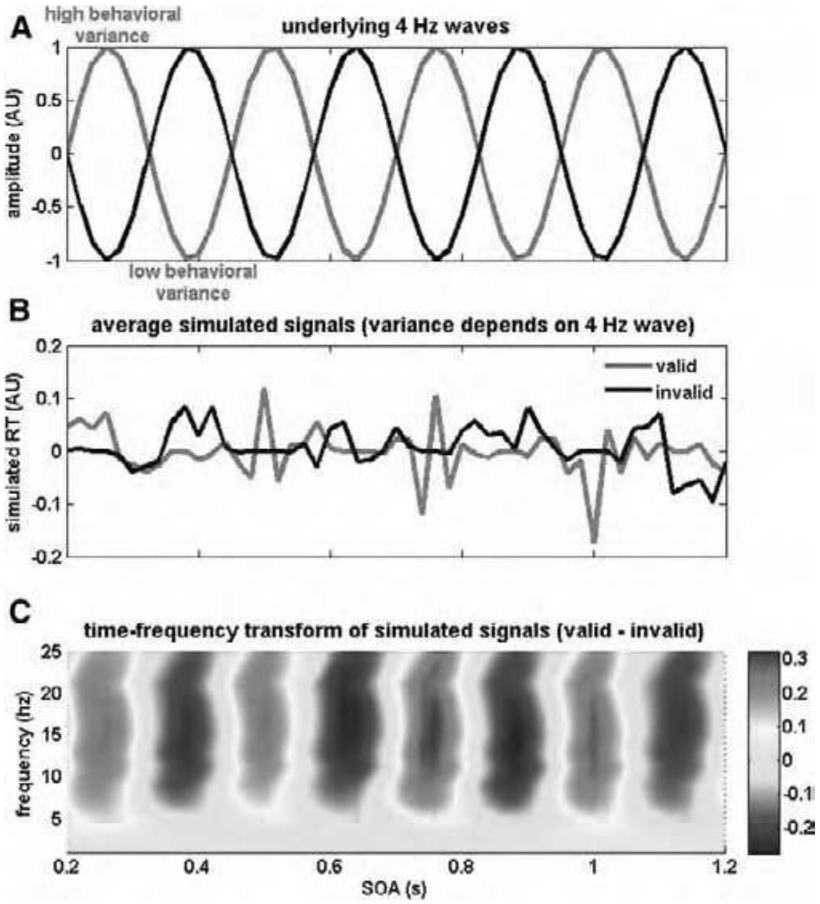


Figure 1. Rhythmic power changes can be explained by regular fluctuations in variance of a signal. ***A***, Simulated signals (2 × 4; "valid" and "invalid" condition in red and black, respectively) were constructed whose variance depends on the phase of an underlying 4 Hz oscillation (highest variance at the peak of the 4 Hz oscillation, lowest variance at the trough). Note that the 4 Hz oscillations underlying the two conditions are in antiphase and are assumed to be reset at time 0. ***B***, This reset results in states of high and low variance that are consistent across signals and therefore do not cancel each other out after averaging the four signals. However, these states are in opposition between the two conditions (when variance is high in one condition, it is low in the other, and vice versa). ***C***, Due to those rhythmic changes in variance, both the individual power profile of the two averaged signals (data not shown) and the average spectrotemporal profile of their difference show rhythmic changes in power (difference; color-coded). Note the similarity of ***B*** and the raw RT shown in Song et al. (2014; their Fig. 1*B*, top), and the similarity of ***C*** and the spectrotemporal profiles shown in Song et al. (2014; Fig. 2*A*).

In short, we believe that the study by Song et al. (2014) nicely underlines already reported evidence for perception and attention as cyclic processes, in line with electrophysiological results (Thut et al., 2012; Romei et al., 2012). However, the frequency range of the described alpha-power effect is too broad to reflect an oscillatory component and might reflect mere changes in RT variability that depend on the instantaneous state of in-

put gating (i.e., high and low behavioral variability when attention is low and high, respectively).

References

Buzsáki G, Draguhn A (2004) Neuronal oscillations in cortical networks. Science 304:1926–1929. CrossRef Medline

Landau AN, Fries P (2012) Attention samples stimuli rhythmically. Curr Biol 22:1000–1004. CrossRef Medline

Luck SJ (2005) An introduction to the event-related potential technique. Cambridge: MIT.

Romei V, Gross J, Thut G (2012) Sounds reset rhythms of visual cortex and corresponding human visual perception. Curr Biol 22:807–813. CrossRef Medline

Schroeder CE, Lakatos P (2009) Low-frequency neuronal oscillations as instruments of sensory selection. Trends Neurosci 32:9–18. CrossRef Medline

Song K, Meng M, Chen L, Zhou K, Luo H (2014) Behavioral oscillations in attention: rhythmic alpha pulses mediated through theta band. J Neurosci 34:4837–4844. CrossRef Medline

Thut G, Miniussi C, Gross J (2012) The functional importance of rhythmic activity in the brain. Curr Biol 22:R658–R663. CrossRef Medline

Tort AB, Komorowski R, Eichenbaum H, Kopell N (2010) Measuring phase-amplitude coupling between neuronal oscillations of different frequencies. J Neurophysiol 104:1195–1210. CrossRef Medline

REFERENCES

Abrams DA, Nicol T, Zecker S, Kraus N (2008) Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. J Neurosci 28:3958–3965.

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.

Allport DA (1968) Phenomenal simutaneity and the perceptual moment hypothesis. Br J Psychol Lond Engl 1953 59:395–406.

Andrews T, Purves D (2005) The wagon-wheel illusion in continuous light. Trends Cogn Sci 9:261–263.

Apteker RT, Fisher JA, Kisimov VS, Neishlos H (1995) Video acceptability and frame rate. IEEE Multimed 2:32–40.

Arai I, Yamada Y, Asaka T, Tachibana M (2004) Light-evoked oscillatory discharges in retinal ganglion cells are generated by rhythmic synaptic inputs. J Neurophysiol 92:715–725.

Arnal LH (2012) Predicting “When” Using the Motor System’s Beta-Band Oscillations. Front Hum Neurosci 6:225.

Arnal LH, Doelling KB, Poeppel D (2015) Delta-Beta Coupled Oscillations Underlie Temporal Prediction Accuracy. Cereb Cortex 25:3077-3085.

Arnal LH, Giraud A-L (2012) Cortical oscillations and sensory predictions. Trends Cogn Sci 16:390–398.

Arnal LH, Wyart V, Giraud A-L (2011) Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. Nat Neurosci 14:797–801.

Aru J, Aru J, Priesemann V, Wibral M, Lana L, Pipa G, Singer W, Vicente R (2015) Untangling cross-frequency coupling in neuroscience. Curr Opin Neurobiol 31:51–61.

Bahramisharif A, van Gerven MAJ, Aarnoutse EJ, Mercier MR, Schwartz TH, Foxe JJ, Ramsey NF, Jensen O (2013) Propagating neocortical gamma bursts are coordinated by traveling alpha waves. J Neurosci 33:18849–18854.

Ballou G (2005) Handbook for Sound Engineers. Burlington: Focal Press.

Barth DS, MacDonald KD (1996) Thalamic modulation of high-frequency oscillating potentials in auditory cortex. Nature 383:78–81.

Başar E, Başar-Eroglu C, Karakaş S, Schürmann M (2001) Gamma, alpha, delta, and theta oscillations govern cognitive processes. Int J Psychophysiol 39:241–248.

Bastos AM, Briggs F, Alitto HJ, Mangun GR, Usrey WM (2014) Simultaneous recordings from the primary visual cortex and lateral geniculate nucleus reveal rhythmic interactions and a cortical source for γ-band oscillations. J Neurosci 34:7639–7644.

Bastos AM, Litvak V, Moran R, Bosman CA, Fries P, Friston KJ (2015) A DCM study of spectral asymmetries in feedforward and feedback connections between visual areas V1 and V4 in the monkey. NeuroImage 108:460–475.

Bastos AM, Vezoli J, Bosman CA, Schoffelen J-M, Oostenveld R, Dowdall JR, De Weerd P, Kennedy H, Fries P (2015) Visual areas exert feedforward and feedback influences through distinct frequency channels. Neuron 85:390–401.

Bear MF, Connors BW, Paradiso MA (2006) Neuroscience: Exploring the Brain. Baltimore, MD: Lippincott Williams and Wilkins.

Bee MA, Klump GM (2004) Primitive auditory stream segregation: a neurophysiological study in the songbird forebrain. J Neurophysiol 92:1088–1104.

Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B Methodol 57:289–300.

Berens P (2009) CircStat: A Matlab Toolbox for Circular Statistics. J Stat Softw 31:1–31.

Berger H (1929) Über das Elektrenkephalogramm des Menschen. Arch Psychiatr Nervenkr 87:527–570.

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.

Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET (2000) Human temporal lobe activation by speech and nonspeech sounds. Cereb Cortex 10:512–528.

Bishop GH (1932) Cyclic changes in excitability of the optic pathway of the rabbit. Am J Physiol 103:213–224.

Blake R, Logothetis NK (2002) Visual competition. Nat Rev Neurosci 3:13–21.

Bollimunta A, Chen Y, Schroeder CE, Ding M (2008) Neuronal mechanisms of cortical alpha oscillations in awake-behaving macaques. J Neurosci 28:9976–9988.

Bonnefond M, Jensen O (2012) Alpha oscillations serve to protect working memory maintenance against anticipated distracters. Curr Biol 22:1969–1974.

Bonnefond M, Jensen O (2015) Gamma activity coupled to alpha phase as a mechanism for top-down controlled gating. PloS One 10:e0128667.

Brainard DH (1997) The Psychophysics Toolbox. Spat Vis 10:433–436.

Bregman AS (1978) Auditory streaming is cumulative. J Exp Psychol Hum Percept Perform 4:380–387.

Bregman AS (1994) Auditory Scene Analysis: The Perceptual Organization of Sound, Revised. Cambridge, MA: MIT Press.

Bregman AS, Campbell J (1971) Primary auditory stream segregation and perception of order in rapid sequences of tones. J Exp Psychol 89:244–249.

Bremer F (1958) Cerebral and cerebellar potentials. Physiol Rev 38:357–388.

Brett B, Krishnan G, Barth DS (1996) The effects of subcortical lesions on evoked potentials and spontaneous high frequency (gamma-band) oscillating potentials in rat auditory cortex. Brain Res 721:155–166.

Buffalo EA, Fries P, Landman R, Buschman TJ, Desimone R (2011) Laminar differences in gamma and alpha coherence in the ventral stream. Proc Natl Acad Sci U S A 108:11262–11267.

Buffalo EA, Fries P, Landman R, Liang H, Desimone R (2010) A backward progression of attentional effects in the ventral stream. Proc Natl Acad Sci U S A 107:361–365.

Busch NA, Dubois J, VanRullen R (2009) The phase of ongoing EEG oscillations predicts visual perception. J Neurosci 29:7869–7876.

Busch NA, VanRullen R (2010) Spontaneous EEG oscillations reveal periodic sampling of visual attention. Proc Natl Acad Sci U S A 107:16048–16053.

Buzsáki G, Anastassiou CA, Koch C (2012) The origin of extracellular fields and currents--EEG, ECoG, LFP and spikes. Nat Rev Neurosci 13:407–420.

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.

Calderone DJ, Lakatos P, Butler PD, Castellanos FX (2014) Entrainment of neural oscillations as a modifiable substrate of attention. Trends Cogn Sci 18:300–309.

Canavier CC (2015) Phase-resetting as a tool of information transmission. Curr Opin Neurobiol 31:206–213.

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:e14543.

Chakravarthi R, Vanrullen R (2012) Conscious updating is a rhythmic process. Proc Natl Acad Sci U S A 109:10599–10604.

Chapman RM, Shelburne SA, Bragdon HR (1970) EEG alpha activity influenced by visual input and not by eye position. Electroencephalogr Clin Neurophysiol 28:183–189.

Cherry EC (1953) Some Experiments on the Recognition of Speech, with One and with Two Ears. J Acoust Soc Am 25:975–979.

Chung K, McKibben N (2011) Microphone directionality, pre-emphasis filter, and wind noise in cochlear implants. J Am Acad Audiol 22:586–600.

Cigánek L (1969) Variability of the human visual evoked potential: normative data. Electroencephalogr Clin Neurophysiol 27:35–42.

Coutlee CG, Huettel SA (2012) The functional neuroanatomy of decision making: prefrontal control of thought and action. Brain Res 1428:3–12.

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.

Cummins F (2012) Oscillators and syllables: a cautionary note. Front Psychol 3:364.

Davis MH, Johnsrude IS (2003) Hierarchical processing in spoken language comprehension. J Neurosci 23:3423–3431.

Davis MH, Johnsrude IS (2007) Hearing speech sounds: top-down influences on the interface between audition and speech perception. Hear Res 229:132–147.

de Graaf TA, Gross J, Paterson G, Rusch T, Sack AT, Thut G (2013) Alpha-band rhythms in visual task performance: phase-locking by rhythmic sensory stimulation. PloS One 8:e60035.

Dehaene S (1993) Temporal Oscillations in Human Perception. Psychol Sci 4:264–270.

Deike S, Heil P, Böckmann-Barthel M, Brechmann A (2012) The Build-up of Auditory Stream Segregation: A Different Perspective. Front Psychol 3:461.

Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 134:9–21.

Denham SL, Winkler I (2014) Auditory perceptual organization. In: Oxford Handbook of Perceptual Organization, Johan Wagemans. Oxford University Press.

Deschênes M, Paradis M, Roy JP, Steriade M (1984) Electrophysiology of neurons of lateral thalamic nuclei in cat: resting properties and burst discharges. J Neurophysiol 51:1196–1219.

DeWitt I, Rauschecker JP (2012) Phoneme and word recognition in the auditory ventral stream. Proc Natl Acad Sci U S A 109:E505–E514.

Dietsch G (1932) Fourier-Analyse von Elektrencephalogrammen des Menschen. Pflüg Arch Gesamte Physiol 230:106–112.

Di Lollo V, Wilson AE (1978) Iconic persistence and perceptual moment as determinants of temporal integration in vision. Vision Res 18:1607–1610.

Ding N, Chatterjee M, Simon JZ (2013) Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. NeuroImage 88C:41–46.

Ding N, Simon JZ (2012a) Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. J Neurophysiol 107:78–89.

Ding N, Simon JZ (2012b) Emergence of neural encoding of auditory objects while listening to competing speakers. Proc Natl Acad Sci U S A 109:11854–11859.

Ding N, Simon JZ (2013) Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. J Neurosci 33:5728–5735.

Ding N, Simon JZ (2014) Cortical entrainment to continuous speech: functional roles and interpretations. Front Hum Neurosci 8:311.

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.

Drewes J, VanRullen R (2011) This is the rhythm of your eyes: the phase of ongoing electroencephalogram oscillations modulates saccadic reaction time. J Neurosci 31:4698–4708.

Drullman R, Festen JM, Plomp R (1994a) Effect of reducing slow temporal modulations on speech reception. J Acoust Soc Am 95:2670–2680.

Drullman R, Festen JM, Plomp R (1994b) Effect of temporal envelope smearing on speech reception. J Acoust Soc Am 95:1053–1064.

Dugué L, Marque P, VanRullen R (2011) The phase of ongoing oscillations mediates the causal relation between brain excitation and visual perception. J Neurosci 31:11889–11893.

Dugué L, Marque P, VanRullen R (2015) Theta oscillations modulate attentional search performance periodically. J Cogn Neurosci 27:945–958.

Durbin J (1960) The fitting of time series models. Rev Inst Int Stat 28:233–243.

Durgin FH, Tripathy SP, Levi DM (1995) On the filling in of the visual blind spot: some rules of thumb. Perception 24:827–840.

Dvorak D, Fenton AA (2014) Toward a proper estimation of phase-amplitude coupling in neural oscillations. J Neurosci Methods 225:42–56.

Eagleman DM, Sejnowski TJ (2000) Motion integration and postdiction in visual awareness. Science 287:2036–2038.

Edwards E, Chang EF (2013) Syllabic (~2-5 Hz) and fluctuation (~1-10 Hz) ranges in speech and auditory processing. Hear Res 305:113–134.

Eggermont JJ (1998) Representation of spectral and temporal sound features in three cortical fields of the cat. Similarities outweigh differences. J Neurophysiol 80:2743–2764.

Elhilali M, Ma L, Micheyl C, Oxenham AJ, Shamma SA (2009) Temporal coherence in the perceptual organization and cortical representation of auditory scenes. Neuron 61:317–329.

Engel AK, Fries P (2010) Beta-band oscillations--signalling the status quo? Curr Opin Neurobiol 20:156–165.

Eriksen CW, Spencer T (1969) Rate of information processing in visual perception: some results and methodological considerations. J Exp Psychol 79:1–16.

Essens PJ, Povel DJ (1985) Metrical and nonmetrical representations of temporal patterns. Percept Psychophys 37:1–7.

Fell J, Axmacher N (2011) The role of phase synchronization in memory processes. Nat Rev Neurosci 12:105–118.

Féron F-X, Frissen I, Boissinot J, Guastavino C (2010) Upper limits of auditory rotational motion perception. J Acoust Soc Am 128:3703–3714.

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.

Fishman YI, Reser DH, Arezzo JC, Steinschneider M (2001) Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. Hear Res 151:167–187.

Fitch WT (2013) Rhythmic cognition in humans and animals: distinguishing meter and pulse perception. Front Syst Neurosci 7:68.

Fontolan L, Morillon B, Liegeois-Chauvel C, Giraud A-L (2014) The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex. Nat Commun 5:4694.

Foxe JJ, Snyder AC (2011) The Role of Alpha-Band Brain Oscillations as a Sensory Suppression Mechanism during Selective Attention. Front Psychol 2:154.

Freeman JA, Nicholson C (1975) Experimental optimization of current source-density technique for anuran cerebellum. J Neurophysiol 38:369–382.

Frey JN, Mainy N, Lachaux J-P, Müller N, Bertrand O, Weisz N (2014) Selective modulation of auditory cortical alpha activity in an audiovisual spatial attention task. J Neurosci 34:6634–6639.

Frey JN, Ruhnau P, Weisz N (2015) Not so different after all: The same oscillatory processes support different types of attention. Brain Res. DOI: 10.1016/j.brainres.2015.02.017.

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.

Frisina RD (2001) Subcortical neural coding mechanisms for auditory temporal processing. Hear Res 158:1–27.

Friston K (2005) A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci 360:815–836.

Fujioka T, Trainor LJ, Large EW, Ross B (2012) Internalized timing of isochronous sounds is represented in neuromagnetic β oscillations. J Neurosci 32:1791–1802.

Funke K, Kisvarday F, Volgushev M, Worgotter F (2002) The visual cortex. In: Models of neural networks IV: early vision and attention, pp 131–160. New York: Springer.

Galambos R, Makeig S, Talmachoff PJ (1981) A 40-Hz auditory potential recorded from the human scalp. Proc Natl Acad Sci U S A 78:2643–2647.

Ghitza O (2001) On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception. J Acoust Soc Am 110:1628–1640.

Ghitza O (2011) Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. Front Psychol 2:130.

Ghitza O (2012) On the role of theta-driven syllabic parsing in decoding speech: intelligibility of speech with a manipulated modulation spectrum. Front Psychol 3:238.

Ghitza O (2013) The theta-syllable: a unit of speech information defined by cortical function. Front Psychol 4:138.

Ghitza O (2014) Behavioral evidence for the role of cortical θ oscillations in determining auditory channel capacity for speech. Front Psychol 5:652.

Ghitza O, Giraud A-L, Poeppel D (2012) Neuronal oscillations and speech perception: critical-band temporal envelopes are the essence. Front Hum Neurosci 6:340.

Ghitza O, Greenberg S (2009) On the possible role of brain rhythms in speech perception: intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. Phonetica 66:113–126.

Gho M, Varela FJ (1988) A quantitative assessment of the dependency of the visual temporal frame upon the cortical rhythm. J Physiol 83:95–101.

Gilbert CD, Li W (2013) Top-down influences on visual processing. Nat Rev Neurosci 14:350–363.

Giraud A-L, Poeppel D (2012) Cortical oscillations and speech processing: emerging computational principles and operations. Nat Neurosci 15:511–517.

Goldwyn JH, Mc Laughlin M, Verschooten E, Joris PX, Rinzel J (2014) A model of the medial superior olive explains spatiotemporal features of local field potentials. J Neurosci 34:11705–11722.

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 electrocorticographic investigation. J Neurosci 31:18556–18567.

Gould IC, Rushworth MF, Nobre AC (2011) Indexing the graded allocation of visuospatial attention using anticipatory alpha oscillations. J Neurophysiol 105:1318–1326.

Grahn JA (2012) See what I hear? Beat perception in auditory and visual rhythms. Exp Brain Res 220:51–61.

Granger CWJ (1969) Investigating Causal Relations by Econometric Models and Cross-spectral Methods. Econometrica 37:424–438.

Gray MJ, Frey H-P, Wilson TJ, Foxe JJ (2015) Oscillatory recruitment of bilateral visual cortex during spatial attention to competing rhythmic inputs. J Neurosci 35:5489–5503.

Green T, Faulkner A, Rosen S (2002) Spectral and temporal cues to pitch in noise-excited vocoder simulations of continuous-interleaved-sampling cochlear implants. J Acoust Soc Am 112:2155–2164.

Greenberg S (1998) A syllable-centric framework for the evolution of spoken language. Behav Brain Sci 21:518.

Griffiths TD, Warren JD (2004) What is an auditory object? Nat Rev Neurosci 5:887–892.

Grimsley JMS, Shanbhag SJ, Palmer AR, Wallace MN (2012) Processing of communication calls in Guinea pig auditory cortex. PloS One 7:e51646.

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.

Grothe B (2000) The evolution of temporal processing in the medial superior olive, an auditory brainstem structure. Prog Neurobiol 61:581–610.

Gutschalk A, Oxenham AJ, Micheyl C, Wilson EC, Melcher JR (2007) Human cortical activity during streaming without spectral cues suggests a general neural substrate for auditory stream segregation. J Neurosci 27:13074–13081.

Guttman SE, Gilroy LA, Blake R (2005) Hearing what the eyes see: auditory encoding of visual temporal sequences. Psychol Sci 16:228–235.

Haegens S, Händel BF, Jensen O (2011a) Top-down controlled alpha band activity in somatosensory areas determines behavioral performance in a discrimination task. J Neurosci 31:5197–5204.

Haegens S, Nácher V, Luna R, Romo R, Jensen O (2011b) 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.

Hanslmayr S, Aslan A, Staudigl T, Klimesch W, Herrmann CS, Bäuml K-H (2007) Prestimulus oscillations predict visual perception performance between and within subjects. NeuroImage 37:1465–1473.

Harter MR (1967) Excitability cycles and cortical scanning: a review of two hypotheses of central intermittency in perception. Psychol Bull 68:47–58.

Heil P, Peterson AJ (2015) Basic response properties of auditory nerve fibers: a review. Cell Tissue Res.

Henry MJ, Herrmann B (2012) A precluding role of low-frequency oscillations for auditory perception in a continuous processing mode. J Neurosci 32:17525–17527.

Henry MJ, Herrmann B, Obleser J (2014) Entrained neural oscillations in multiple frequency bands comodulate behavior. Proc Natl Acad Sci U S A 111:14935–14940.

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.

Herrmann CS (2001) Human EEG responses to 1-100 Hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. Exp Brain Res 137:346–353.

Herrmann CS, Rach S, Neuling T, Strüber D (2013) Transcranial alternating current stimulation: a review of the underlying mechanisms and modulation of cognitive processes. Front Hum Neurosci 7:279.

Herrmann CS, Rach S, Vosskuhl J, Strüber D (2014) Time-frequency analysis of event-related potentials: a brief tutorial. Brain Topogr 27:438–450.

Herrmann CS, Strüber D, Helfrich RF, Engel AK (2015) EEG oscillations: From correlation to causality. Int J Psychophysiol. DOI: 10.1016/j.ijpsycho.2015.02.003.

Hickok G, Farahbod H, Saberi K (2015) The Rhythm of Perception: Entrainment to Acoustic Rhythms Induces Subsequent Perceptual Oscillation. Psychol Sci 26:1006-1013.

Hickok G, Poeppel D (2007) The cortical organization of speech processing. Nat Rev Neurosci 8:393–402.

Hillyard SA, Teder-Sälejärvi WA, Münte TF (1998) Temporal dynamics of early perceptual processing. Curr Opin Neurobiol 8:202–210.

Hirsh IJ, Sherrick CE (1961) Perceived order in different sense modalities. J Exp Psychol 62:423–432.

Holcombe AO, Clifford CWG, Eagleman DM, Pakarian P (2005) Illusory motion reversal in tune with motion detectors. Trends Cogn Sci 9:559–560.

Holcombe AO, Seizova-Cajic T (2008) Illusory motion reversals from unambiguous motion with visual, proprioceptive, and tactile stimuli. Vision Res 48:1743–1757.

Honing H, Merchant H, Háden GP, Prado L, Bartolo R (2012) Rhesus monkeys (Macaca mulatta) detect rhythmic groups in music, but not the beat. PloS One 7:e51369.

Horschig JM, Jensen O, van Schouwenburg MR, Cools R, Bonnefond M (2014) Alpha activity reflects individual abilities to adapt to the environment. NeuroImage 89:235–243.

Horton C, D'Zmura M, Srinivasan R (2013) Suppression of competing speech through entrainment of cortical oscillations. J Neurophysiol 109:3082–3093.

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.

Hutcheon B, Yarom Y (2000) Resonance, oscillation and the intrinsic frequency preferences of neurons. Trends Neurosci 23:216–222.

Hutchinson CV, Ledgeway T (2006) Sensitivity to spatial and temporal modulations of first-order and second-order motion. Vision Res 46:324–335.

Hyafil A, Fontolan L, Kabdebon C, Gutkin B, Giraud A-L (2015) Speech encoding by coupled cortical theta and gamma oscillations. eLife 4.

İlhan B, VanRullen R (2012) No counterpart of visual perceptual echoes in the auditory system. PloS One 7:e49287.

Immerseel LV, Peeters S (2003) Digital implementation of linear gammatone filters: Comparison of design methods. Acoust Res Lett Online 4:59–64.

Iversen JR, Patel AD, Ohgushi K (2008) Perception of rhythmic grouping depends on auditory experience. J Acoust Soc Am 124:2263–2271.

Jacobs J, Kahana MJ, Ekstrom AD, Fried I (2007) Brain oscillations control timing of single-neuron activity in humans. J Neurosci 27:3839–3844.

Jans B, Peters JC, De Weerd P (2010) Visual spatial attention to multiple locations at once: the jury is still out. Psychol Rev 117:637–684.

Jaspers HH, Andrews HL (1938) Electroencephalography III. Normal differentiations of occipital and precentral regions in man. Arch Neurol Psychiatry 39:96–115.

Jensen O, Bonnefond M, Marshall TR, Tiesinga P (2015) Oscillatory mechanisms of feedforward and feedback visual processing. Trends Neurosci 38:192–194.

Jensen O, Bonnefond M, VanRullen R (2012) An oscillatory mechanism for prioritizing salient unattended stimuli. Trends Cogn Sci 16:200–206.

Jensen O, Gips B, Bergmann TO, Bonnefond M (2014) Temporal coding organized by coupled alpha and gamma oscillations prioritize visual processing. Trends Neurosci 37:357–369.

Jensen O, Mazaheri A (2010) Shaping functional architecture by oscillatory alpha activity: gating by inhibition. Front Hum Neurosci 4:186.

Joliot M, Ribary U, Llinás R (1994) Human oscillatory brain activity near 40 Hz coexists with cognitive temporal binding. Proc Natl Acad Sci U S A 91:11748–11751.

Jones EG (2002) Thalamic circuitry and thalamocortical synchrony. Philos Trans R Soc Lond B Biol Sci 357:1659–1673.

Kahneman D (1973) Attention and effort. Englewood Cliffs, NJ: Prentice-Hall.

Kavanagh KT, Domico WD (1986) High-pass digital filtration of the 40 Hz response and its relationship to the spectral content of the middle latency and 40 Hz responses. Ear Hear 7:93–99.

Kay LM (2014) Circuit oscillations in odor perception and memory. Prog Brain Res 208:223–251.

Keil J, Müller N, Hartmann T, Weisz N (2014) Prestimulus beta power and phase synchrony influence the sound-induced flash illusion. Cereb Cortex 24:1278–1288.

Keitel C, Quigley C, Ruhnau P (2014) Stimulus-driven brain oscillations in the alpha range: entrainment of intrinsic rhythms or frequency-following response? J Neurosci 34:10137–10140.

Kerlin JR, Shahin AJ, Miller LM (2010) Attentional gain control of ongoing cortical speech representations in a “cocktail party.” J Neurosci 30:620–628.

Ketz NA, Jensen O, O’Reilly RC (2015) Thalamic pathways underlying prefrontal cortex-medial temporal lobe oscillatory interactions. Trends Neurosci 38:3–12.

Kinnunen T, Li H (2010) An overview of text-independent speaker recognition: from features to supervectors. Speech Commun 52:12–40.

Klimesch W, Sauseng P, Hanslmayr S (2007) EEG alpha oscillations: the inhibition-timing hypothesis. Brain Res Rev 53:63–88.

Kline KA, Eagleman DM (2008) Evidence against the temporal subsampling account of illusory motion reversal. J Vis 8:13.1–5.

Kline K, Holcombe AO, Eagleman DM (2004) Illusory motion reversal is caused by rivalry, not by perceptual snapshots of the visual field. Vision Res 44:2653–2658.

Koepsell K, Wang X, Vaingankar V, Wei Y, Wang Q, Rathbun DL, Usrey WM, Hirsch JA, Sommer FT (2009) Retinal oscillations carry visual information to cortex. Front Syst Neurosci 3:4.

Kohler MA (1997) A comparison of the new 2400 bps MELP Federal Standard with other standard coders. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, 1997. ICASSP-97, pp 1587–1590 vol.2.

Köppl C (1997) Phase locking to high frequencies in the auditory nerve and cochlear nucleus magnocellularis of the barn owl, Tyto alba. J Neurosci 17:3312–3321.

Kösem A, Gramfort A, van Wassenhove V (2014) Encoding of event timing in the phase of neural oscillations. NeuroImage 92:274–284.

Kotz SA, Schmidt-Kassow M (2015) Basal ganglia contribution to rule expectancy and temporal predictability in speech. Cortex 68:48-60.

Krawczyk DC (2002) Contributions of the prefrontal cortex to the neural basis of human decision making. Neurosci Biobehav Rev 26:631–664.

Krumbholz K, Patterson RD, Seither-Preisler A, Lammertmann C, Lütkenhöner B (2003) Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. Cereb Cortex 13:765–772.

Kubovy M (1988) Should we resist the seductiveness of the Space:Time::Vision:Audition analogy? J Exp Psychol Hum Percept Perform 14:318–320.

Lachaux J-P, Axmacher N, Mormann F, Halgren E, Crone NE (2012) High-frequency neural activity and human cognition: past, present and possible future of intracranial EEG research. Prog Neurobiol 98:279–301.

Lachaux JP, Rodriguez E, Martinerie J, Varela FJ (1999) Measuring phase synchrony in brain signals. Hum Brain Mapp 8:194–208.

Lakatos P, Barczak A, Ross D, McGinnis T, Javitt DC, O'Connell MN (submitted) Global temporal dynamics of attention and its distinct neuronal signatures.

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, Musacchia G, O'Connel MN, Falchier AY, Javitt DC, Schroeder CE (2013a) The spectrotemporal filter mechanism of auditory selective attention. Neuron 77:750–761.

Lakatos P, O'Connell MN, Barczak A, Mills A, Javitt DC, Schroeder CE (2009) The leading sense: supramodal control of neurophysiological context by attention. Neuron 64:419–430.

Lakatos P, Pincze Z, Fu K-MG, Javitt DC, Karmos G, Schroeder CE (2005a) Timing of pure tone and noise-evoked responses in macaque auditory cortex. Neuroreport 16:933–937.

Lakatos P, Schroeder CE, Leitman DI, Javitt DC (2013b) Predictive suppression of cortical excitability and its deficit in schizophrenia. J Neurosci 33:11692–11702.

Lakatos P, Shah AS, Knuth KH, Ulbert I, Karmos G, Schroeder CE (2005b) An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. J Neurophysiol 94:1904–1911.

Lakatos S, Shepard RN (1997) Constraints common to apparent motion in visual, tactile, and auditory space. J Exp Psychol Hum Percept Perform 23:1050–1060.

Lalor EC, Foxe JJ (2010) Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. Eur J Neurosci 31:189–193.

Lalor EC, Pearlmutter BA, Reilly RB, McDarby G, Foxe JJ (2006) The VESPA: a method for the rapid estimation of a visual evoked potential. NeuroImage 32:1549–1561.

Lalor EC, Power AJ, Reilly RB, Foxe JJ (2009) Resolving precise temporal processing properties of the auditory system using continuous stimuli. J Neurophysiol 102:349–359.

Lamme VA, Spekreijse H (2000) Modulations of primary visual cortex activity representing attentive and conscious scene perception. Front Biosci 5:D232–D243.

Landau AN, Fries P (2012) Attention samples stimuli rhythmically. Curr Biol 22:1000–1004.

Langner G, Schreiner CE (1988) Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. J Neurophysiol 60:1799–1822.

Large EW, Jones MR (1999) The dynamics of attending: How people track time-varying events. Psychol Rev 106:119–159.

Lee HS, Ghetti A, Pinto-Duarte A, Wang X, Dziewczapolski G, Galimi F, Huitron-Resendiz S, Piña-Crespo JC, Roberts AJ, Verma IM, Sejnowski TJ, Heinemann SF (2014) Astrocytes contribute to gamma oscillations and recognition memory. Proc Natl Acad Sci U S A 111:E3343–E3352.

Legatt AD, Arezzo J, Vaughan HG (1980) Averaged multiple unit activity as an estimate of phasic changes in local neuronal activity: effects of volume-conducted potentials. J Neurosci Methods 2:203–217.

Levinson N (1947) The Wiener RMS error criterion in filter design and prediction. J Math Phys 25:261–278.

Liégeois-Chauvel C, Lorenzi C, Trébuchon A, Régis J, Chauvel P (2004) Temporal envelope processing in the human left and right auditory cortices. Cereb Cortex 14:731–740.

Lisman JE, Jensen O (2013) The θ-γ neural code. Neuron 77:1002–1016.

Llinás RR (1988) The intrinsic electrophysiological properties of mammalian neurons: insights into central nervous system function. Science 242:1654–1664.

Llinás RR, Paré D (1991) Of dreaming and wakefulness. Neuroscience 44:521–535.

Lopes da Silva F (1991) Neural mechanisms underlying brain waves: from neural membranes to networks. Electroencephalogr Clin Neurophysiol 79:81–93.

Lopes da Silva F (2013) EEG and MEG: relevance to neuroscience. Neuron 80:1112–1128.

Lopes da Silva FH, Pijn JP, Velis D, Nijssen PC (1997) Alpha rhythms: noise, dynamics and models. Int J Psychophysiol 26:237–249.

Lorincz ML, Kékesi KA, Juhász G, Crunelli V, Hughes SW (2009) Temporal framing of thalamic relay-mode firing by phasic inhibition during the alpha rhythm. Neuron 63:683–696.

Luck SJ (2005) An Introduction to the Event-Related Potential Technique. Cambridge, Massachusetts: The MIT Press.

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 (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.

McCarthy J, Sasse M, Miras D (2004) Sharp or smooth? Comparing the effects of quantization vs. frame rate for streamed video. In: Proc. SIGCHI Conference on Human Factors in Computing Systems, Vienna, Austria, 24–29 Apr 2004, pp 535–542. Aarhus, Denmark: Interaction Design Foundation.

Macdonald JSP, Cavanagh P, VanRullen R (2014) Attentional sampling of multiple wagon wheels. Atten Percept Psychophys 76:64–72.

Mack A, Rock I (1998) Inattentional blindness. Cambridge, MA: MIT Press.

Macmillan NA, Creelman CD (2004) Detection Theory: A User's Guide. Taylor & Francis.

Maier A, Adams GK, Aura C, Leopold DA (2010) Distinct superficial and deep laminar domains of activity in the visual cortex during rest and stimulation. Front Syst Neurosci 4.

Makeig S, Westerfield M, Jung TP, Enghoff S, Townsend J, Courchesne E, Sejnowski TJ (2002) Dynamic brain sources of visual evoked responses. Science 295:690–694.

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.

Mazzoni A, Whittingstall K, Brunel N, Logothetis NK, Panzeri S (2010) Understanding the relationships between spike rate and delta/gamma frequency bands of LFPs and EEGs using a local cortical network model. NeuroImage 52:956–972.

McMahon CM, Patuzzi RB (2002) The origin of the 900 Hz spectral peak in spontaneous and sound-evoked round-window electrical activity. Hear Res 173:134–152.

McMullan AR, Hambrook DA, Tata MS (2013) Brain dynamics encode the spectrotemporal boundaries of auditory objects. Hear Res 304:77–90.

Mendelson JR, Cynader MS (1985) Sensitivity of cat primary auditory cortex (AI) neurons to the direction and rate of frequency modulation. Brain Res 327:331–335.

Merchant H, Grahn J, Trainor L, Rohrmeier M, Fitch WT (2015) Finding the beat: a neural perspective across humans and non-human primates. Philos Trans R Soc Lond B Biol Sci 370:20140093.

Merzenich MM, Brugge JF (1973) Representation of the cochlear partition of the superior temporal plane of the macaque monkey. Brain Res 50:275–296.

Mesgarani N, Chang EF (2012) Selective cortical representation of attended speaker in multi-talker speech perception. Nature 485:233–236.

Mesgarani N, Cheung C, Johnson K, Chang EF (2014) Phonetic feature encoding in human superior temporal gyrus. Science 343:1006–1010.

Michel CM, Murray MM, Lantz G, Gonzalez S, Spinelli L, Grave de Peralta R (2004) EEG source imaging. Clin Neurophysiol 115:2195–2222.

Micheyl C, Tian B, Carlyon RP, Rauschecker JP (2005) Perceptual organization of tone sequences in the auditory cortex of awake macaques. Neuron 48:139–148.

Miller GA, Licklider JCR (1950) The intelligibility of interrupted speech. J Acoust Soc Am 22:167–173.

Millman RE, Johnson SR, Prendergast G (2015) The Role of Phase-locking to the Temporal Envelope of Speech in Auditory Perception and Speech Intelligibility. J Cogn Neurosci 27:533–545.

Millman RE, Prendergast G, Hymers M, Green GGR (2013) Representations of the temporal envelope of sounds in human auditory cortex: can the results from invasive intracortical “depth” electrode recordings be replicated using non-invasive MEG “virtual electrodes”? NeuroImage 64:185–196.

Minlebaev M, Colonnese M, Tsintsadze T, Sirota A, Khazipov R (2011) Early γ oscillations synchronize developing thalamus and cortex. Science 334:226–229.

Mitzdorf U (1985) Current source-density method and application in cat cerebral cortex: investigation of evoked potentials and EEG phenomena. Physiol Rev 65:37–100.

Moerel M, De Martino F, Formisano E (2012) Processing of natural sounds in human auditory cortex: tonotopy, spectral tuning, and relation to voice sensitivity. J Neurosci 32:14205–14216.

Moore BCJ (2003) An Introduction to the Psychology of Hearing. Academic Press.

Moore BCJ, Gockel HE (2012) Properties of auditory stream formation. Philos Trans R Soc Lond B Biol Sci 367:919–931.

Morillon B, Hackett TA, Kajikawa Y, Schroeder CE (2015) Predictive motor control of sensory dynamics in auditory active sensing. Curr Opin Neurobiol 31C:230–238.

Morillon B, Schroeder CE (2015) Neuronal oscillations as a mechanistic substrate of auditory temporal prediction. Ann N Y Acad Sci 1337:26–31.

Morillon B, Schroeder CE, Wyart V (2014) Motor contributions to the temporal precision of auditory attention. Nat Commun 5:5255.

Müller N, Weisz N (2012) Lateralized auditory cortical alpha band activity and interregional connectivity pattern reflect anticipation of target sounds. Cereb Cortex 22:1604–1613.

Nelken I (2008) Processing of complex sounds in the auditory system. Curr Opin Neurobiol 18:413–417.

Nelken I, Bizley J, Shamma SA, Wang X (2014) Auditory cortical processing in real-world listening: the auditory system going real. J Neurosci 34:15135–15138.

Nelken I, Las L, Ulanovsky N, Farkas D (2005) Are speech, pitch, and space early or late in the auditory system? In: The auditory cortex: a synthesis of human and animal research. Mahwah, NJ: Lawrence Erlbaum Associates.

Neuenschwander S, Castelo-Branco M, Singer W (1999) Synchronous oscillations in the cat retina. Vision Res 39:2485–2497.

Neuling T, Rach S, Wagner S, Wolters CH, Herrmann CS (2012) Good vibrations: oscillatory phase shapes perception. NeuroImage 63:771–778.

Ng BSW, Logothetis NK, 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 (1994) Motion extrapolation in catching. Nature 370:256–257.

Nourski KV, Reale RA, Oya H, Kawasaki H, Kovach CK, Chen H, Howard MA 3rd, Brugge JF (2009) Temporal envelope of time-compressed speech represented in the human auditory cortex. J Neurosci 29:15564–15574.

Nozaradan S (2014) Exploring how musical rhythm entrains brain activity with electroencephalogram frequency-tagging. Philos Trans R Soc Lond B Biol Sci 369:20130393.

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.

Obleser J, Eisner F, Kotz SA (2008) Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. J Neurosci 28:8116–8123.

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.

O'Connell MN, Barczak A, Ross D, McGinnis T, Schroeder CE, Lakatos P (submitted) Multi-scale entrainment of coupled neuronal oscillations in primary auditory cortex.

O'Connell MN, Barczak A, Schroeder CE, Lakatos P (2014) Layer specific sharpening of frequency tuning by selective attention in primary auditory cortex. J Neurosci 34:16496–16508.

O'Connell MN, Falchier A, McGinnis T, Schroeder CE, Lakatos P (2011) Dual mechanism of neuronal ensemble inhibition in primary auditory cortex. Neuron 69:805–817.

Oever S ten, Schroeder CE, Poeppel D, van Atteveldt N, Zion-Golumbic E (2014) Rhythmicity and cross-modal temporal cues facilitate detection. Neuropsychologia 63:43–50.

Oshurkova E, Scheich H, Brosch M (2008) Click train encoding in primary and non-primary auditory cortex of anesthetized macaque monkeys. Neuroscience 153:1289–1299.

O'Sullivan JA, Power AJ, Mesgarani N, Rajaram S, Foxe JJ, Shinn-Cunningham BG, Slaney M, Shamma SA, Lalor EC (2015) Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. Cereb Cortex 25:1697–1706.

O'Sullivan JA, Shamma SA, Lalor EC (2015) Evidence for Neural Computations of Temporal Coherence in an Auditory Scene and Their Enhancement during Active Listening. J Neurosci 35:7256–7263.

Otero-Millan J, Troncoso XG, Macknik SL, Serrano-Pedraza I, Martinez-Conde S (2008) Saccades and microsaccades during visual fixation, exploration, and search: foundations for a common saccadic generator. J Vis 8:21.1–18.

Pachitariu M, Lyamzin DR, Sahani M, Lesica NA (2015) State-dependent population coding in primary auditory cortex. J Neurosci 35:2058–2073.

Palmer AR, Russell IJ (1986) Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells. Hear Res 24:1–15.

Park H, Ince RAA, Schyns PG, Thut G, Gross J (2015) Frontal Top-Down Signals Increase Coupling of Auditory Low-Frequency Oscillations to Continuous Speech in Human Listeners. Curr Biol 25:1649–1653.

Patel AD, Iversen JR, Chen Y, Repp BH (2005) The influence of metricality and modality on synchronization with a beat. Exp Brain Res 163:226–238.

Pavlides C, Greenstein YJ, Grudman M, Winson J (1988) Long-term potentiation in the dentate gyrus is induced preferentially on the positive phase of theta-rhythm. Brain Res 439:383–387.

Peelen MV, Kastner S (2014) Attention in the real world: toward understanding its neural basis. Trends Cogn Sci 18:242–250.

Peelle JE, Davis MH (2012) Neural Oscillations Carry Speech Rhythm through to Comprehension. Front Psychol 3:320.

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.

Petkov CI, Kang X, Alho K, Bertrand O, Yund EW, Woods DL (2004) Attentional modulation of human auditory cortex. Nat Neurosci 7:658–663.

Petkov CI, Kayser C, Steudel T, Whittingstall K, Augath M, Logothetis NK (2008) A voice region in the monkey brain. Nat Neurosci 11:367–374.

Pfurtscheller G, Stancák A, Neuper C (1996) Event-related synchronization (ERS) in the alpha band--an electrophysiological correlate of cortical idling: a review. Int J Psychophysiol 24:39–46.

Piantoni G, Kline KA, Eagleman DM (2010) Beta oscillations correlate with the probability of perceiving rivalrous visual stimuli. J Vis 10:18.

Picazio S, Veniero D, Ponzo V, Caltagirone C, Gross J, Thut G, Koch G (2014) Prefrontal control over motor cortex cycles at beta frequency during movement inhibition. Curr Biol 24:2940–2945.

Pitts W, McCulloch WS (1947) How we know universals; the perception of auditory and visual forms. Bull Math Biophys 9:127–147.

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, Emmorey K, Hickok G, Pylkkänen L (2012) Towards a new neurobiology of language. J Neurosci 32:14125–14131.

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.

Pöppel E (1997) A hierarchical model of temporal perception. Trends Cogn Sci 1:56–61.

Posner MI (1994) Attention: the mechanisms of consciousness. Proc Natl Acad Sci U S A 91:7398–7403.

Povel DJ, Okkerman H (1981) Accents in equitone sequences. Percept Psychophys 30:565–572.

Power AJ, Mead N, Barnes L, Goswami U (2013) Neural entrainment to rhythmic speech in children with developmental dyslexia. Front Hum Neurosci 7:777.

Práwdicz-Neminski WW (1925) Zur Kenntnis der elektrischen und der Innervationsvorgänge in den funktionellen Elementen und Geweben des tierischen Organismus. Elektrocerebrogramm der Säugetiere. Pflüg Arch Gesamte Physiol 209:362–382.

Pressnitzer D, Sayles M, Micheyl C, Winter IM (2008) Perceptual organization of sound begins in the auditory periphery. Curr Biol 18:1124–1128.

Purushothaman G, Marion R, Li K, Casagrande VA (2012) Gating and control of primary visual cortex by pulvinar. Nat Neurosci 15:905–912.

Purves D, Paydarfar JA, Andrews TJ (1996) The wagon wheel illusion in movies and reality. Proc Natl Acad Sci U S A 93:3693–3697.

Rager G, Singer W (1998) The response of cat visual cortex to flicker stimuli of variable frequency. Eur J Neurosci 10:1856–1877.

Rajkai C, Lakatos P, Chen C-M, Pincze Z, Karmos G, Schroeder CE (2008) Transient cortical excitation at the onset of visual fixation. Cereb Cortex 18:200–209.

Rauschecker JP (2015) Auditory and visual cortex of primates: a comparison of two sensory systems. Eur J Neurosci 41:579–585.

Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat Neurosci 12:718–724.

Reddy L, Rémy F, Vayssière N, VanRullen R (2011) Neural correlates of the continuous Wagon Wheel Illusion: a functional MRI study. Hum Brain Mapp 32:163–170.

Regan D (1977) Steady-state evoked potentials. J Opt Soc Am 67:1475–1489.

Repp BH, Penel A (2002) Auditory dominance in temporal processing: New evidence from synchronization with simultaneous visual and auditory sequences. J Exp Psychol Hum Percept Perform 28:1085–1099.

Rezec A, Krekelberg B, Dobkins KR (2004) Attention enhances adaptability: evidence from motion adaptation experiments. Vision Res 44:3035–3044.

Rimmele JM, Zion Golumbic E, Schröger E, Poeppel D (2015) The effects of selective attention and speech acoustics on neural speech-tracking in a multi-talker scene. Cortex 68:144-154.

Rinne T, Stecker GC, Kang X, Yund EW, Herron TJ, Woods DL (2007) Attention modulates sound processing in human auditory cortex but not the inferior colliculus. Neuroreport 18:1311–1314.

Rohracher H (1935) Die gehirnelektrischen Erscheinungen bei geistiger Arbeit. Z Psychol 136:308–324.

Romei V, Brodbeck V, Michel C, Amedi A, Pascual-Leone A, Thut G (2008) Spontaneous fluctuations in posterior alpha-band EEG activity reflect variability in excitability of human visual areas. Cereb Cortex 18:2010–2018.

Romei V, Gross J, Thut G (2012) Sounds reset rhythms of visual cortex and corresponding human visual perception. Curr Biol 22:807–813.

Romo R, de Lafuente V (2013) Conversion of sensory signals into perceptual decisions. Prog Neurobiol 103:41–75.

Rose MM, Moore BCJ (2005) The relationship between stream segregation and frequency discrimination in normally hearing and hearing-impaired subjects. Hear Res 204:16–28.

Roux F, Wibral M, Singer W, Aru J, Uhlhaas PJ (2013) The phase of thalamic alpha activity modulates cortical gamma-band activity: evidence from resting-state MEG recordings. J Neurosci 33:17827–17835.

Ruhnau P, Hauswald A, Weisz N (2014) Investigating ongoing brain oscillations and their influence on conscious perception - network states and the window to consciousness. Front Psychol 5:1230.

Rushworth MFS, Kolling N, Sallet J, Mars RB (2012) Valuation and decision-making in frontal cortex: one or many serial or parallel systems? Curr Opin Neurobiol 22:946–955.

Saalmann YB, Pinsk MA, Wang L, Li X, Kastner S (2012) The pulvinar regulates information transmission between cortical areas based on attention demands. Science 337:753–756.

Saberi K, Perrott DR (1999) Cognitive restoration of reversed speech. Nature 398:760.

Saenz M, Langers DRM (2014) Tonotopic mapping of human auditory cortex. Hear Res 307:42–52.

Salazar RF, Dotson NM, Bressler SL, Gray CM (2012) Content-specific fronto-parietal synchronization during visual working memory. Science 338:1097–1100.

Samaha J, Bauer P, Cimaroli S, Postle BR (2015) Top-down control of the phase of alpha-band oscillations as a mechanism for temporal prediction. Proc Natl Acad Sci U S A 112:8439-8444.

Sauseng P, Klimesch W, Gruber WR, Hanslmayr S, Freunberger R, Doppelmayr M (2007) Are event-related potential components generated by phase resetting of brain oscillations? A critical discussion. Neuroscience 146:1435–1444.

Sauseng P, Klimesch W, Heise KF, Gruber WR, Holz E, Karim AA, Glennon M, Gerloff C, Birbaumer N, Hummel FC (2009) Brain oscillatory substrates of visual short-term memory capacity. Curr Biol 19:1846–1852.

Schepers IM, Hipp JF, Schneider TR, Röder B, Engel AK (2012) Functionally specific oscillatory activity correlates between visual and auditory cortex in the blind. Brain J Neurol 135:922–934.

Schnupp JWH, Hall TM, Kokelaar RF, Ahmed B (2006) Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex. J Neurosci 26:4785–4795.

Schouten J (1967) Subjective stroboscopy and a model of visual movement detectors. In: Models for the perception of speech and visual form, pp 44–45. Cambridge, MA: MIT Press.

Schreiber T (2000) Measuring information transfer. Phys Rev Lett 85:461–464.

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.

Schroeder CE, Mehta AD, Givre SJ (1998) A spatiotemporal profile of visual system activation revealed by current source density analysis in the awake macaque. Cereb Cortex 8:575–592.

Schroeder CE, Wilson DA, Radman T, Scharfman H, Lakatos P (2010) Dynamics of Active Sensing and perceptual selection. Curr Opin Neurobiol 20:172–176.

Scott SK, Blank CC, Rosen S, Wise RJ (2000) Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123 Pt 12:2400–2406.

Shamma SA, Elhilali M, Micheyl C (2011) Temporal coherence and attention in auditory scene analysis. Trends Neurosci 34:114–123.

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. Science 270:303–304.

Sheffert SM, Pisoni DB, Fellowes JM, Remez RE (2002) Learning to recognize talkers from natural, sinewave, and reversed speech samples. J Exp Psychol Hum Percept Perform 28:1447–1469.

Shepard RN (1964) Circularity in Judgments of Relative Pitch. J Acoust Soc Am 36:2346–2353.

Shipley T (1964) Auditory flutter-driving of visual flicker. Science 145:1328–1330.

Simon JZ (2015) The encoding of auditory objects in auditory cortex: insights from magnetoencephalography. Int J Psychophysiol 95:184–190.

Simpson WA, Shahani U, Manahilov V (2005) Illusory percepts of moving patterns due to discrete temporal sampling. Neurosci Lett 375:23–27.

Smith PH, Bartlett EL, Kowalkowski A (2007) Cortical and collicular inputs to cells in the rat paralaminar thalamic nuclei adjacent to the medial geniculate body. J Neurophysiol 98:681–695.

Snyder JS, Alain C, Picton TW (2006) Effects of attention on neuroelectric correlates of auditory stream segregation. J Cogn Neurosci 18:1–13.

Song K, Meng M, Chen L, Zhou K, Luo H (2014) Behavioral oscillations in attention: rhythmic α pulses mediated through θ band. J Neurosci 34:4837–4844.

Soon IY, Koh SN, Yeo CK (1998) Noisy speech enhancement using discrete cosine transform. Speech Commun 24:249–257.

Spaak E, Bonnefond M, Maier A, Leopold DA, Jensen O (2012) Layer-specific entrainment of γ-band neural activity by the α rhythm in monkey visual cortex. Curr Biol 22:2313–2318.

Spaak E, de Lange FP, Jensen O (2014) Local entrainment of α oscillations by visual stimuli causes cyclic modulation of perception. J Neurosci 34:3536–3544.

Stapells DR, Linden D, Suffield JB, Hamel G, Picton TW (1984) Human auditory steady state potentials. Ear Hear 5:105–113.

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.

Steinschneider M, Nourski KV, Fishman YI (2013) Representation of speech in human auditory cortex: is it special? Hear Res 305:57–73.

Steinschneider M, Nourski KV, Rhone AE, Kawasaki H, Oya H, Howard MA (2014) Differential activation of human core, non-core and auditory-related cortex during speech categorization tasks as revealed by intracranial recordings. Front Neurosci 8:240.

Steinschneider M, Reser D, Schroeder CE, Arezzo JC (1995) Tonotopic organization of responses reflecting stop consonant place of articulation in primary auditory cortex (A1) of the monkey. Brain Res 674:147–152.

Steriade M, Gloor P, Llinás RR, Lopes de Silva FH, Mesulam MM (1990) Report of IFCN Committee on Basic Mechanisms. Basic mechanisms of cerebral rhythmic activities. Electroencephalogr Clin Neurophysiol 76:481–508.

Steriade M, McCormick DA, Sejnowski TJ (1993) Thalamocortical oscillations in the sleeping and aroused brain. Science 262:679–685.

Stilp CE, Kiefte M, Alexander JM, Kluender KR (2010) Cochlea-scaled spectral entropy predicts rate-invariant intelligibility of temporally distorted sentences. J Acoust Soc Am 128:2112–2126.

Strauß A, Kotz SA, Scharinger M, Obleser J (2014a) Alpha and theta brain oscillations index dissociable processes in spoken word recognition. NeuroImage 97:387–395.

Strauß A, Wöstmann M, Obleser J (2014b) Cortical alpha oscillations as a tool for auditory selective inhibition. Front Hum Neurosci 8:350.

Stroud JM (1956) The fine structure of psychological time. In: Information theory in psychology: problems and methods, pp 174–207. New York, NY, US: Free Press.

Suied C, Drémeau A, Pressnitzer D, Daudet L (2013) Auditory Sketches: Sparse Representations of Sounds Based on Perceptual Models. In: From Sounds to Music and Emotions (Aramaki M, Barthet M, Kronland-Martinet R, Ystad S, eds), pp 154–170 Lecture Notes in Computer Science. Springer Berlin Heidelberg. Available at: http://link.springer.com/chapter/10.1007/978-3-642-41248-6_9 [Accessed October 29, 2014].

Teki S, Chait M, Kumar S, Kriegstein K von, Griffiths TD (2011) Brain bases for auditory stimulus-driven figure-ground segregation. J Neurosci 31:164–171.

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.

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.

Thut G, Miniussi C (2009) New insights into rhythmic brain activity from TMS-EEG studies. Trends Cogn Sci 13:182–189.

Thut G, Miniussi C, Gross J (2012) The functional importance of rhythmic activity in the brain. Curr Biol 22:R658–R663.

Thut G, Nietzel A, Brandt SA, Pascual-Leone A (2006) Alpha-band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. J Neurosci 26:9494–9502.

Tian B, Reser D, Durham A, Kustov A, Rauschecker JP (2001) Functional specialization in rhesus monkey auditory cortex. Science 292:290–293.

Tort ABL, Komorowski R, Eichenbaum H, Kopell N (2010) Measuring phase-amplitude coupling between neuronal oscillations of different frequencies. J Neurophysiol 104:1195–1210.

Townsend JT (1990) Serial vs. Parallel Processing: Sometimes They Look Like Tweedledum and Tweedledee but They Can (and Should) Be Distinguished. Psychol Sci 1:46–54.

Treisman AM (1969) Strategies and models of selective attention. Psychol Rev 76:282–299.

Uhlhaas PJ, Pipa G, Lima B, Melloni L, Neuenschwander S, Nikolić D, Singer W (2009) Neural synchrony in cortical networks: history, concept and current status. Front Integr Neurosci 3:17.

Uppenkamp S, Johnsrude IS, Norris D, Marslen-Wilson W, Patterson RD (2006) Locating the initial stages of speech-sound processing in human temporal cortex. NeuroImage 31:1284-1296.

Usrey WM, Reid RC (1999) Synchronous activity in the visual system. Annu Rev Physiol 61:435–456.

van Kerkoerle T, Self MW, Dagnino B, Gariel-Mathis M-A, Poort J, van der Togt C, Roelfsema PR (2014) Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. Proc Natl Acad Sci U S A 111:14332–14341.

Van Noorden LPAS (1975) Temporal coherence in the perception of tone sequences. PhD thesis, Eindhoven University of Technology.

VanRullen R (2006) The continuous Wagon Wheel Illusion is object-based. Vision Res 46:4091–4095.

VanRullen R (2007) The continuous Wagon Wheel Illusion depends on, but is not identical to neuronal adaptation. Vision Res 47:2143–2149.

Vanrullen R, Busch NA, Drewes J, Dubois J (2011) Ongoing EEG Phase as a Trial-by-Trial Predictor of Perceptual and Attentional Variability. Front Psychol 2:60.

VanRullen R, Carlson T, Cavanagh P (2007) The blinking spotlight of attention. Proc Natl Acad Sci U S A 104:19204–19209.

Vanrullen R, Dubois J (2011) The psychophysics of brain rhythms. Front Psychol 2:203.

VanRullen R, Koch C (2003) Is perception discrete or continuous? Trends Cogn Sci 7:207–213.

VanRullen R, Macdonald JSP (2012) Perceptual echoes at 10 Hz in the human brain. Curr Biol 22:995–999.

Vanrullen R, McLelland D (2013) What goes up must come down: EEG phase modulates auditory perception in both directions. Front Psychol 4:16.

VanRullen R, Pascual-Leone A, Battelli L (2008) The continuous Wagon wheel illusion and the "when" pathway of the right parietal lobe: a repetitive transcranial magnetic stimulation study. PloS One 3:e2911.

VanRullen R, Reddy L, Koch C (2005) Attention-driven discrete sampling of motion perception. Proc Natl Acad Sci U S A 102:5291–5296.

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.

VanRullen R, Zoefel B, Ilhan B (2014) On the cyclic nature of perception in vision versus audition. Philos Trans R Soc Lond B Biol Sci 369:20130214.

van Santen JP, Sperling G (1985) Elaborated Reichardt detectors. J Opt Soc Am A 2:300–321.

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, Grzeczkowski L (2015) Visual-induced expectations modulate auditory cortical responses. Front Neurosci 9:11.

Varela FJ, Toro A, John ER, Schwartz EL (1981) Perceptual framing and cortical alpha rhythm. Neuropsychologia 19:675–686.

Wall C, Kozak WM, Sanderson AC (1979) Entrainment of oscillatory neural activity in the cat's lateral geniculate nucleus. Biol Cybern 33:63–75.

Walter WG (1936) The location of cerebral tumours by electro-encephalography. Lancet 2:305–308.

Wang X, Lu T, Snider RK, Liang L (2005) Sustained firing in auditory cortex evoked by preferred stimuli. Nature 435:341–346.

Wang X, Merzenich MM, Beitel R, Schreiner CE (1995) Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. J Neurophysiol 74:2685–2706.

Weisz N, Moratti S, Meinzer M, Dohrmann K, Elbert T (2005) Tinnitus perception and distress is related to abnormal spontaneous brain activity as measured by magnetoencephalography. PLoS Med 2:e153.

Weisz N, Obleser J (2014) Synchronisation signatures in the listening brain: a perspective from non-invasive neuroelectrophysiology. Hear Res 307:16–28.

Weisz N, Wühle A, Monittola G, Demarchi G, Frey J, Popov T, Braun C (2014) Prestimulus oscillatory power and connectivity patterns predispose conscious somatosensory perception. Proc Natl Acad Sci U S A 111:E417–E425.

Wertheimer M (1912) Experimentelle Studien uber das Sehen von Bewegung. Zeit Psychol 61:161-265.

Wessinger CM, VanMeter J, Tian B, Van Lare J, Pekar J, Rauschecker JP (2001) Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. J Cogn Neurosci 13:1–7.

Whittingstall K, Logothetis NK (2009) Frequency-band coupling in surface EEG reflects spiking activity in monkey visual cortex. Neuron 64:281–289.

Wiegrebe L, Winter IM (2001) Temporal representation of iterated rippled noise as a function of delay and sound level in the ventral cochlear nucleus. J Neurophysiol 85:1206–1219.

Wild CJ, Yusuf A, Wilson DE, Peelle JE, Davis MH, Johnsrude IS (2012) Effortful listening: the processing of degraded speech depends critically on attention. J Neurosci 32:14010–14021.

Wilsch A, Henry MJ, Herrmann B, Maess B, Obleser J (2015) Alpha Oscillatory Dynamics Index Temporal Expectation Benefits in Working Memory. Cereb Cortex 25:1938-1946.

Wöstmann M, Herrmann B, Wilsch A, Obleser J (2015) Neural alpha dynamics in younger and older listeners reflect acoustic challenges and predictive benefits. J Neurosci 35:1458–1467.

Wutz A, Melcher D (2014) The temporal window of individuation limits visual capacity. Front Psychol 5:952.

Wutz A, Weisz N, Braun C, Melcher D (2014) Temporal windows in visual processing: “prestimulus brain state” and “poststimulus phase reset” segregate visual transients on different temporal scales. J Neurosci 34:1554–1565.

Zaehle T, Rach S, Herrmann CS (2010) Transcranial alternating current stimulation enhances individual alpha activity in human EEG. PloS One 5:e13766.

Zion Golumbic E, Cogan GB, Schroeder CE, Poeppel D (2013a) Visual input enhances selective speech envelope tracking in auditory cortex at a “cocktail party.” J Neurosci 33:1417–1426.

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 (2013b) Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party.” Neuron 77:980–991.

Zion Golumbic EM, Poeppel D, Schroeder CE (2012) Temporal context in speech processing and attentional stream selection: a behavioral and neural perspective. Brain Lang 122:151–161.

Zoefel B, Heil P (2013) Detection of Near-Threshold Sounds is Independent of EEG Phase in Common Frequency Bands. Front Psychol 4:262.

Zoefel B, Reddy Pasham N, Brüers S, VanRullen R (2015) The ability of the auditory system to cope with temporal subsampling depends on the hierarchical level of processing. Neuroreport 26:773-778.

Zoefel B, Sokoliuk R (2014) Investigating the rhythm of attention on a fine-grained scale: evidence from reaction times. J Neurosci 34:12619–12621.

Zoefel B, VanRullen R (2015) Selective perceptual phase entrainment to speech rhythm in the absence of spectral energy fluctuations. J Neurosci 35:1954–1964.

Zoefel B, VanRullen R (in press) EEG oscillations entrain their phase to high-level features of speech sound. NeuroImage.

Zoefel B, VanRullen R (submitted) The role of high-level processes for oscillatory phase entrainment to speech sound.

Zumer JM, Scheeringa R, Schoffelen J-M, Norris DG, Jensen O (2014) Occipital alpha activity during stimulus processing gates the information flow to object-selective cortex. PLoS Biol 12:e1001965.