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

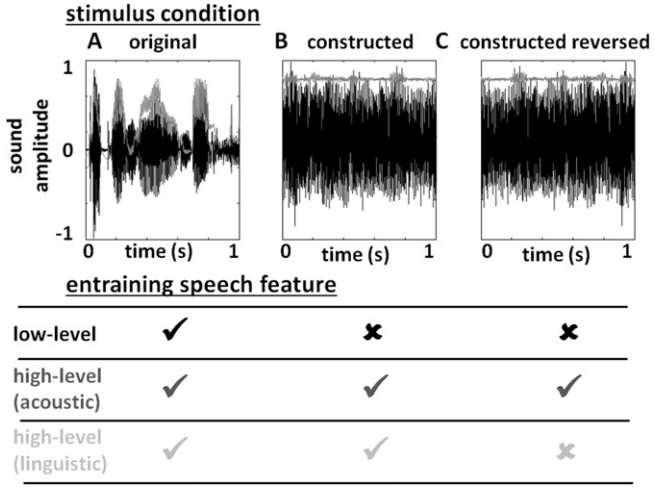

Fig. 1. The contribution of low- and high-level components of speech sound to phase entrainment was studied in three conditions (in the upper panel, for each condition, 1 s of an exemplary stimulus is shown in black, with its signal envelope in gray). Original speech snippets (A) were presented, entailing 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

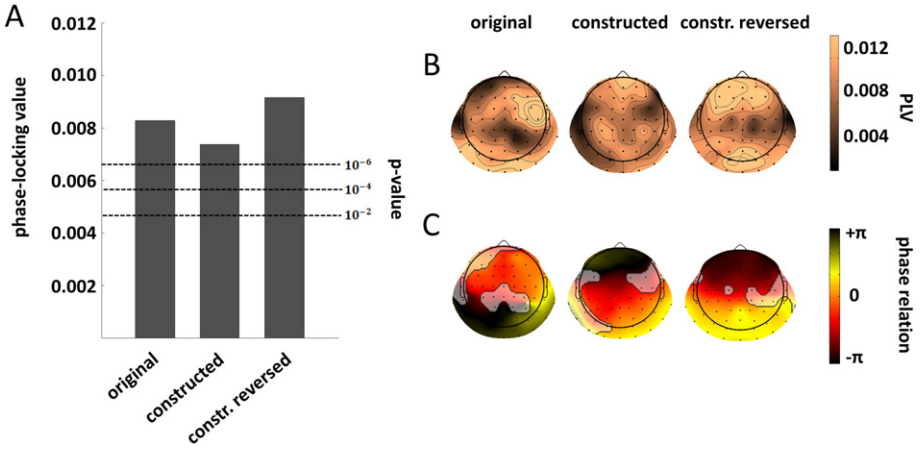

Fig. 2. Phase-locking between EEG signal and original signal envelope (i.e., phase entrainment) in the three conditions. A. The average phase-locking across channels is shown as bars. P-values of phase-locking (obtained by a permutation procedure; see Materials and Methods) are shown as dashed lines, 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

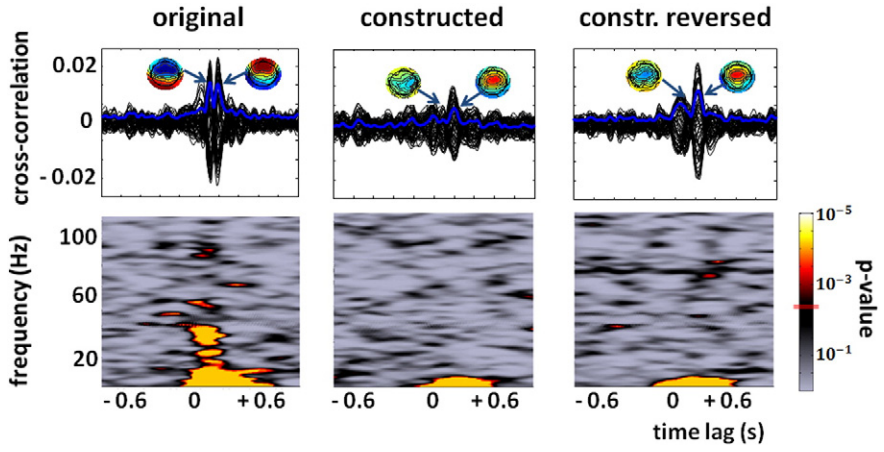

Fig. 3. Top panels: Cross-correlation between original signal envelope and EEG signal (both unfiltered) for all channels (black lines) and the standard deviation across channels (blue line). Only the original condition shows a peak in standard deviation at ~110 ms (time lag between speech and EEG), indicating 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.

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.

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.

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.