frontiers in
PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 31 May 2012
doi: 10.3389/fpsyg.2012.00170

Cortical oscillations in auditory perception and speech: evidence for two temporal windows in human auditory cortex

Huan Luo[1]* and David Poeppel[2]*

[1] State Key Laboratory of Brain and Cognitive Sciences, Institute of Biophysics, Chinese Academy of Sciences, Beijing, China
[2] Department of Psychology, New York University, New York, NY, USA

Edited by:
Marcela Pena, Catholic University of Chile, Chile

Reviewed by:
Andreas Keil, University of Florida, USA
Jackson Gandour, Purdue University, USA

***Correspondence:**
Huan Luo, State Key Laboratory of Brain and Cognitive Science, Institute of Biophysics, Chinese Academy of Sciences, Room 7302, 15 Datun Road, Chaoyang District, Beijing, 100101, China.
e-mail: luohuan@gmail.com;
David Poeppel, Department of Psychology, New York University, 6 Washington Place, New York, NY 10003,USA.
e-mail: david.poeppel@nyu.edu

Natural sounds, including vocal communication sounds, contain critical information at multiple time scales. Two essential temporal modulation rates in speech have been argued to be in the low gamma band (~20–80 ms duration information) and the theta band (~150–300 ms), corresponding to segmental and diphonic versus syllabic modulation rates, respectively. It has been hypothesized that auditory cortex implements temporal integration using time constants closely related to these values. The neural correlates of a proposed dual temporal window mechanism in human auditory cortex remain poorly understood. We recorded MEG responses from participants listening to non-speech auditory stimuli with different temporal structures, created by concatenating frequency-modulated segments of varied segment durations. We show that such non-speech stimuli with temporal structure matching speech-relevant scales (~25 and ~200 ms) elicit reliable phase tracking in the corresponding associated oscillatory frequencies (low gamma and theta bands). In contrast, stimuli with non-matching temporal structure do not. Furthermore, the topography of theta band phase tracking shows rightward lateralization while gamma band phase tracking occurs bilaterally. The results support the hypothesis that there exists multi-time resolution processing in cortex on discontinuous scales and provide evidence for an asymmetric organization of temporal analysis (asymmetrical sampling in time, AST). The data argue for a mesoscopic-level neural mechanism underlying multi-time resolution processing: the sliding and resetting of intrinsic temporal windows on privileged time scales.

Keywords: MEG, magnetoencephalography, timing, phase, phase coherence

INTRODUCTION

Mapping from input sounds (such as speech) to stored representations (such as words) involves the temporal analysis and integration of information on distinct – and perhaps even non-overlapping – timescales (Poeppel, 2003; Hickok and Poeppel, 2007; Poeppel et al., 2008). Multi-time resolution hypotheses of different types have been proposed to resolve the tension between information carried on different scales concurrently (Greenberg and Ainsworth, 2006; Giraud and Poeppel, 2012a,b). A "calculated simplification" is that the two main temporal scales in speech sounds are ~20–50 ms short scale and ~150–300 ms long scale signals, corresponding to segmental and syllabic rates respectively (Poeppel, 2003; Greenberg and Ainsworth, 2006).

Historically, the analysis of speech was dominated by research focusing on the rich *spectral properties of the acoustic signal* (see, e.g. Liberman, 1996, for many important experimental examples). That research forms the basis for much of our current understanding of how speech perception may function and has yielded many of the foundational insights into both the mental representation of speech and its neurobiological foundations. A second strand of research, somewhat more recent in its origin, has focused on the *temporal properties* of speech signals. Even a cursory glance at the acoustics of speech – whether as a waveform or as a spectrographic representation – reveals that different types of information appear to be carried on different timescales (for a review, see Rosen, 1992). For example, if one analyzes the broadband amplitude envelope of the signal (the type of information that the external ear actually receives, prior to the filterbank decomposition in the cochlea and subsequent auditory nuclei), relatively low modulation frequencies are visible in the signal (below 10 Hz, with peaks often lying between 4 and 6 Hz), with the timescale highly reminiscent of mean syllable duration across languages (Greenberg and Ainsworth, 2006; Pellegrino et al., 2011). By contrast, speech signals contain many rapid fluctuations that require decoding on a much shorter 10-of-ms-scale (e.g., voice onset time, certain formant transitions, onset bursts, etc.). The neural mechanisms for such multi-time resolution processing in human auditory cortex (and some possible hemispheric asymmetries) have been a focus of much recent work.

One hypothesis that has been investigated in a series of recent experiments suggests that the different integration time constants are consequences of intrinsic neuronal oscillations at different rates (Poeppel, 2003). In particular, it has been suggested that oscillatory activity in the theta band correlates closely with temporal "sampling" at the lower rates (relating, most probably, to envelope processing) and that the low gamma band correlates with more rapid information extraction (Poeppel, 2003; Ghitza, 2011; Giraud and Poeppel, 2012a,b). In short, the argument is that there is a close correspondence between neuronal oscillations and the temporal *parsing* of an input stream and *decoding* of sensory information. Recent neurophysiological experiments using magnetoencephalography (MEG), electroencephalography (EEG; e.g. Abrams et al., 2008), as well as concurrent EEG and fMRI (e.g. Giraud et al., 2007) have investigated some of these conjectures.

In a first MEG study aiming to link the modulation spectrum of speech to neural signals (Luo and Poeppel, 2007), participants were presented with naturally spoken sentences. An analysis based on inter-trial coherence of single trials revealed that the phase pattern of neural responses at a specific time scale tracked the stimulus dynamics. In particular, the *phase of the theta band* response showed both the requisite sensitivity and specificity to be interpreted as a neural marker for tracking details of the input signal; moreover, this phase pattern correlated closely with speech intelligibility. Given a standard interpretation of the theta band (∼3-8 Hz), it was suggested that an incoming natural speech stream is segmented and processed on the basis of a ∼200 ms sliding temporal window. In the context of speech, that would mean a parsing of the acoustic signal at roughly a syllabic rate. In a follow-up study using audiovisual movie clips, this data pattern was replicated and extended to the multi-sensory case (Luo et al., 2010). Both of these experimental results, building on coherence analyses of the neural data, support the important role that low modulation frequency brain information plays in perceptual analysis of speech and other auditory signals.

However, how tightly these neurophysiological responses link to intelligibility *per se* and to the representation of *speech* units (versus features in the *acoustics* of speech) remains open and controversial (e.g. Howard and Poeppel, 2010). For example, there may exist attributes in the input signal that could be a prerequisite for recognition – although they are not in any obvious way related to traditional component features of speech. In order to obtain a more thorough perspective on the electrophysiological brain responses underlying speech recognition, in particular in the context of multi-time resolution hypotheses and the discontinuous sampling of information, it is necessary to pursue at least two further lines of investigation (among many other important perspectives). First, it will be helpful to investigate non-speech signals with respect to these kinds of neural responses. Insofar as acoustics play a critical role, the robust and well-replicated neural response profiles tested with speech will be able to be investigated with analytic signals in which the acoustic structure is fully controlled. Secondly, in the first set of experiments, the high modulation frequency/short timescale responses have not been consistently observed (Luo and Poeppel, 2007; Howard and Poeppel, 2010; Luo et al., 2010). Whereas low modulation frequency information is highly robust, easy to replicate, and attested in other techniques as well (e.g., Abrams et al., 2008; for the high sensitivity of human auditory cortex to low modulation frequencies, see Wang et al., 2011; Overath et al., 2012), the putative responses associated with rapid sampling, analysis, and decoding, in the gamma band spectral domain, have been more elusive. One possible reason for not finding short temporal window processing may lie in the behavioral tasks employed in those studies (namely none), such that coarse syllabic-level analysis was enough to achieve a general perception of the sentence (cf. Shannon et al., 1995). A second reason may lie in the acoustic structure of the materials themselves, in which the contribution of rapidly modulated information in the gamma bands was not highlighted in a way to elicit the response in a differential manner.

In the MEG experiment described here, listeners were presented with non-speech signals with varied temporal structures, and the recorded MEG responses were analyzed using a phase tracking coherence method, as employed in our previous studies, to examine the neuronal segmentation of auditory signals at different time scales. Three hypotheses were investigated. First, does the neuronal phase response lock to and follow stimulus dynamics in a way similar to speech signals? Although there exists tantalizing evidence for such time scales from fMRI and MEG during exposure to similar non-speech materials (Giraud et al., 2000; Boemio et al., 2005; Overath et al., 2008; Ding and Simon, 2009), it is not clearly established that the auditory system will lock to these rates in a similar manner when investigated with MEG. Second, if *any* auditory edge (i.e., occurring at any time scale) is sufficient to cause a phase resetting – that is to say acoustic discontinuities or transients occurring on any timescale are the triggering events for phase resetting – then the three stimulus types employed here should elicit a similar response profile – and the notion of different temporal windows loses its appeal. Alternatively, if acoustic discontinuities or edges reveal a grouping into different bands of neural response frequencies or oscillations, such a result would offer support for temporal windows that do not sample the space uniformly. Third, it has been suggested in previous work using various non-invasive approaches that there is an asymmetry with respect to the temporal sampling properties. These non-speech signals may further elucidate potential functional asymmetries of this type and provide potential explanations for why certain domains of perceptual experience appear to be lateralized in human auditory cortex. Anticipating what we describe here, the results show that stimuli with matching temporal structure to these two timescales (∼25 and ∼200 ms) successfully elicited reliable phase tracking at the corresponding cortical rhythms, whereas stimuli without the matching temporal property did not. Such observations are more consistent with the model that there exist non-overlapping sampling rates in auditory cortex.

MATERIALS AND METHODS

PARTICIPANTS

Twelve right-handed subjects (four female), all from the University of Maryland College Park undergraduate and graduate student population, with normal hearing, provided written informed consent before participating in the experiment.

STIMULI

Figure 1 illustrates the experimental materials. Three types of 5-s duration auditory stimuli were created by concatenating individual frequency-modulated segments with mean segment duration of 25, 80, and 200 ms, respectively (sampling frequency of 44.1 kHz). The segment duration values were selected to be well aligned with low gamma, high alpha, and theta band frequencies of the neuronal oscillations potentially subserving the cortical

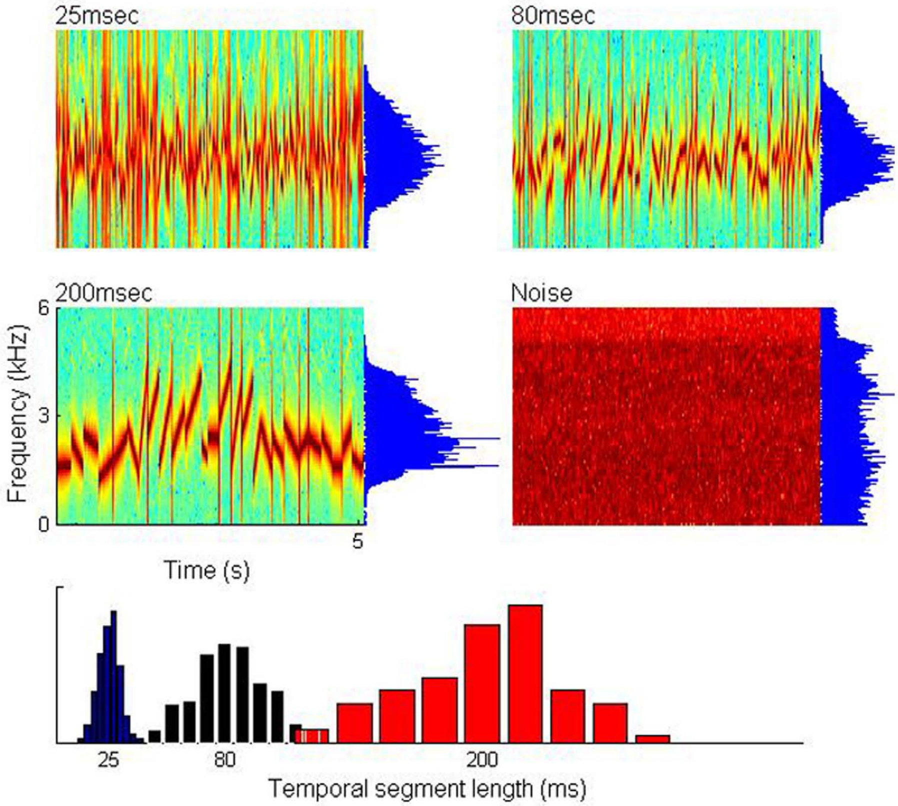

FIGURE 1 | Non-speech stimuli with varying temporal structure. Upper and middle panels: spectrograms and spectra of the three stimulus types with different temporal segment length (mean 25, 80, and 200 ms) and one noise control stimulus. Lower panel: segment length distribution for the three stimuli (blue: ∼25 ms; black: ∼80 ms; red: ∼200 ms).

analysis of such input signals. For each of the individual frequency-modulated segments, the starting frequency and ending frequency was randomly drawn from a uniform distribution of 1000–3,000 and 1,500–4,500 Hz, respectively, so that all of the individual frequency-modulated segments could be swept up or down, or kept flat. For each of the three stimuli with different mean segment durations (25, 80, 200 ms), the durations of each segment within the sound were drawn from a Gaussian distribution with a standard deviation equal to 0.2 of the corresponding mean segment duration (as shown graphically at the bottom of **Figure 1**). A single 5-s control white noise stimulus was constructed, with the power above 5,000 Hz filtered out. Each of these four stimuli were presented 21 times, pseudorandomly interleaved across conditions, at a comfortable loudness level (∼70 dB SPL), and subjects were instructed to passively listen to the stimuli and keep alert.

MEG DATA ACQUISITION

The MEG data were acquired in the Cognitive Neuroscience of Language Laboratory at the University of Maryland College Park. Neuromagnetic signals were recorded continuously with a 157 channel whole-head MEG system (5 cm baseline axial gradiometer SQUID-based sensors; KIT, Kanazawa, Japan) in a magnetically shielded room, using a sampling rate of 1000 Hz, a 60 Hz notch filter, and an online 100 Hz analog low-pass filter, with no high-pass filtering (recording bandwidth DC-100 Hz).

DATA ANALYSIS

In order to inspect the temporal waveforms, using a canonical event-related field analysis, raw MEG data, after noise reduction, were first smoothed using a 20-point moving average, epoched from −0.5 to 5s relative to sound stimulus onset, and then baseline corrected (0.5 s prestimulus interval). To extract auditory cortical responses, we selected 20 channels with a maximum M100 response elicited by a 1-kHz pure tone presented to each participant in a pretest. We then calculated the root mean square of the MEG responses across the 20 auditory channels per subject for the four types of auditory stimuli (∼25, ∼80, ∼200 ms, and noise) to visualize the aggregate auditory response across subjects.

The main data analysis builds on the *inter-trial phase coherence* methods developed in Luo and Poeppel (2007), Luo et al. (2010), and Howard and Poeppel (2010). The spectrogram of each single trial response (21 trials per stimulus) was calculated using a 500 ms time window in steps of 100 ms for each of the 157 MEG recording channels, and the calculated phase and power as a function of frequency and time were stored for further analysis. The "cross-trial phase coherence (*C*phase)" and "cross-trial power coherence (*C*power)" as a function of frequency, which quantifies the reliability of phase and power temporal patterns across trials for each specific stimulus condition in each frequency band, were calculated as:

$$C\text{phase}_i = \frac{\sum_{j=1}^{J}\left(\left(\frac{\sum_{n=1}^{N}\cos(\theta_{ij})}{N}\right)^2+\left(\frac{\sum_{n=1}^{N}\sin(\theta_{ij})}{N}\right)^2\right)}{J}$$

$$C\text{power}_i = \frac{\sum_{j=1}^{J}\left(\frac{\sqrt{\frac{\sum_{n=1}^{N}\left(A_{nij}{}^2-\overline{A_{ij}{}^2}\right)^2}{N}}}{\overline{A_{ij}{}^2}}\right)}{J}$$

where θ_{nij} and A_{nij} are the phase and amplitude at the frequency bin i and temporal bin *j* in trial *n*, respectively. *C*phase is in the range of [0 1]. Note that larger *C*phase value corresponds to strong cross-trial phase coherence, indicating that the sound stimuli with specific temporal structure elicit highly replicable phase pattern responses in each presentation trial, in other words, suggesting a reliable temporal segmentation of incoming sound. The corresponding frequency range (frequency bin *i*) in the *C*phase value represents the approximate window length of the temporal segmentation process. By comparing the *C*phase values at certain frequency bin (for example, theta band, ∼5 Hz) across different sound stimuli with varying temporal structures (∼25, ∼80, ∼200 ms, and noise), we can examine which stimuli elicit the most reliable temporal segmentation in terms of the ∼200 ms window length (the corresponding period of ∼5 Hz).

We subsequently focused only on the *C*phase and *C*power within three frequency ranges of interest (theta: 4∼8 Hz; alpha: 10∼14 Hz; low gamma: 38∼42 Hz), which were chosen based on the corresponding mean temporal segment durations of the three stimulus types (200, 80, 25 ms). For each subject, the average *C*phase and *C*power values within each of the three frequency ranges were calculated for all four stimulus conditions (200, 80, 25 ms, noise) and for all 157 channels, resulting in a 157 × 3 × 4 (channel × frequency × stimulus) dataset. The "phase coherence distribution maps" in each of the three frequency ranges and for each of the four stimulus conditions can then be constructed in terms of the corresponding *C*phase values of all 157 channels, and therefore there were 12 (3 × 4) phase coherence distribution maps for each subject.

To get a rough estimate of large-scale brain activity at each of the three frequencies of interest, and for all the four stimulus conditions, we first averaged the performance (*C*phase and *C*power separately) of all 157 MEG channels in each subject, and compared the mean values across the four stimulus conditions and three frequency ranges. Next, because of the apparent different distribution map for the different frequency ranges, we did a more detailed analysis for each frequency of interest, by selecting 50 channels for the three frequency ranges separately. For each subject, we averaged the "phase coherence distribution maps" for the same frequency range across all the four stimuli conditions, to eliminate any possible channel selection bias introduced by certain stimulus, and then selected the 50 channels with maximum values to stand for the represented channels for that frequency range. For each of the three frequency of interest in each subject, the performance of the selected 50 channels was then compared across stimulus conditions.

To characterize the *C*phase distribution map, we calculated the *C*phase values within each of the three frequency ranges of interest (theta: 4∼8 Hz; alpha: 10∼14 Hz; low gamma: 38∼42 Hz) for all 157 channels and for all four stimulus conditions (200, 80, 25 ms, noise), and examined the corresponding *C*phase distribution map for different frequency bands under different stimulus types. Furthermore, to investigate the lateralization of *C*phase distribution, for each of three frequency bands, we divided the 157 channels into LH channels and RH channels, and averaged *C*phase values within same hemisphere channels, for each of the four stimulus types, separately for each subject.

Finally, comparing different frequency ranges using spectrogram-based analysis in terms of fixed time windows may introduce different sensitivities to the different temporal properties of responses at different frequency ranges. For example, the employed 500 ms time window in steps of 100 ms sliding length, although appropriate for theta and alpha band, may not optimally capture the dynamics of phase and power response pattern in the gamma frequency range. Given this concern, we also did a control analysis in the gamma band (38∼42 Hz) using an induced wavelet transfer method (Complex Gaussian Wavelet) to determine the cross-trial phase and power coherence for each stimulus condition across all 157 channels.

RESULTS

We hypothesize that the two putative cortical temporal integration windows are neurally manifested in the phase pattern of the corresponding cortical rhythms. Moreover, a phase tracking mechanism might be closely related to the two intrinsic temporal windows and thus would be difficult to elicit at other oscillation frequencies. If two such intrinsic cortical temporal windows – manifested as oscillations – exist (Poeppel, 2003; Giraud et al., 2007; Giraud and Poeppel, 2012a,b), the stimuli with mean segment lengths of 25 and 200 ms should elicit reliable phase tracking at the corresponding cortical rhythms (∼40 and ∼5 Hz, respectively), because of the close match between the stimulus temporal structure and the intrinsic cortical temporal window. In contrast, the stimuli with ∼80 ms segment structure should, by extension, not elicit reliable phase tracking at the corresponding rhythms (∼12.5 Hz). Put differently, if all edges/acoustic discontinuities of a stimulus train are "equal," the response profile was predicted to be uniform for all three stimulus types; however, if there are preferences for certain temporal windows, then not all acoustic edges should be effective, only those edges aligning with the privileged windows (e.g., theta, gamma).

EVENT-RELATED MEG AUDITORY RESPONSE

As illustrated in upper panel of **Figure 2**, all four stimuli (∼25, ∼80, ∼200 ms, and noise) elicited typical auditory responses with a peak latency of around 150 ms, and then remained at a sustained level during the sound presentation. Note that these waveforms represent aggregate responses (RMS, root mean square) across 20 auditory channels per subject and thus are all positive values.

The contour map (**Figure 2**, lower right panel) corresponding to the MEG response pattern at ~150 ms window after sound onset, shows a relatively typical auditory topography. The evoked response around 150 ms after sound onset did not show any significant difference across the four types of stimuli [repeated measures one-way ANOVA, $F(3, 33) = 2.32$, $p = 0.093$].

Two temporally structured stimuli elicited stronger cross-trial phase coherence at corresponding frequency bands.

CROSS-TRIAL PHASE COHERENCE

The cross-trial phase coherence was calculated at the three stimulus-relevant frequency ranges (low gamma: 38~42 Hz; alpha: 10~14 Hz; theta: 4~8 Hz) corresponding to the three stimuli (~25, ~80, and ~200 ms). As a control, a noise stimulus containing no apparent temporal structure was used. As illustrated in second and third row panels of **Figure 3**, the results show that phase tracking at the time scales investigated was not accompanied by power tracking (**Figure 3**, middle panels) or stimulus-elicited power increase in the corresponding frequency ranges (**Figure 3**, lower panels), arguing against an "acoustics-only" interpretation. In contrast, compelling effects are observed in the phase patterns. As shown in the upper row of **Figure 3**, the results show phase tracking for both 50 selected auditory channels [two-way repeated measures ANOVA, stimulus × frequency interaction, $F(6, 66 = 5.93$, $p < 0.001$], and all 157 recorded channels $F(6, 66) = 3.42$, $p = 0.005$. Specifically, in the gamma frequency range, the ~25 ms stimulus (blue) elicited the most reliable phase pattern among the four stimulus conditions. In the theta frequency range, the ~200 ms stimulus (red) elicited the most reliable phase pattern. However, the ~80 ms stimulus (green) that has matching temporal structure to the alpha frequency range (~12.5 Hz) did not drive phase tracking efficiently. In addition, the noise stimulus that does not contain any explicit temporal structure did not drive phase tracking in any of the three frequency ranges tested here.

Comparing different frequency ranges using spectrogram-based analyses in terms of fixed time windows may introduce differential sensitivities to the different temporal properties of responses at different frequency ranges. For example, the employed 500 ms time window, in steps of 100 ms sliding length, although appropriate for the theta and alpha bands, may not optimally

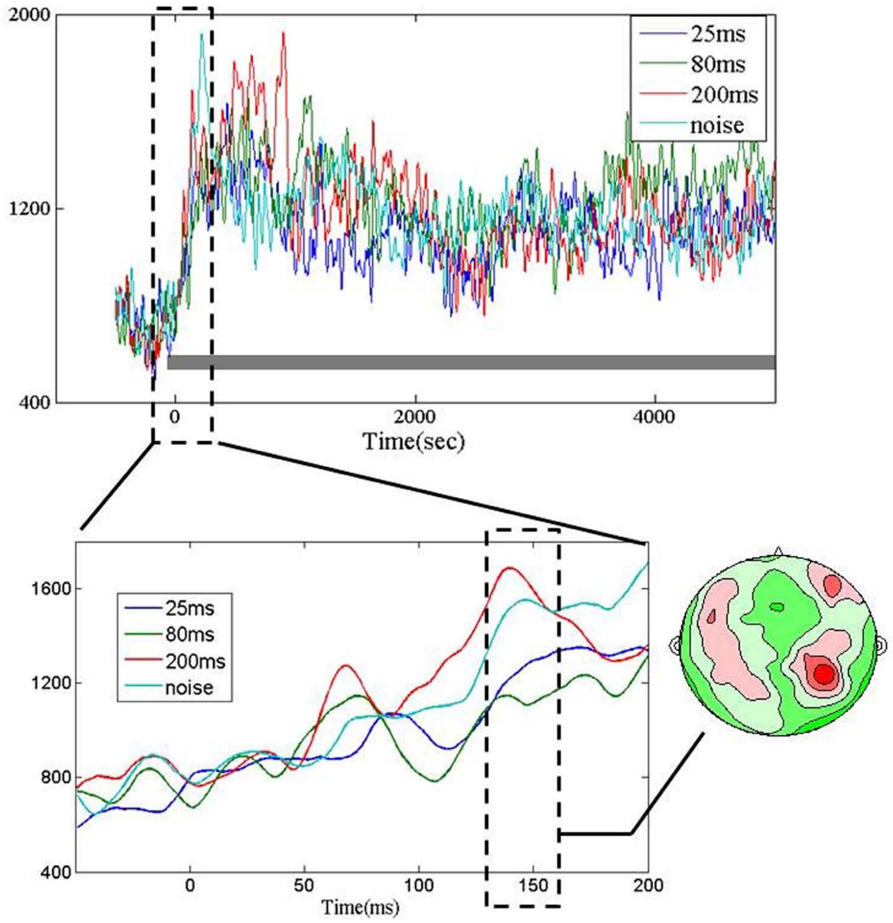

FIGURE 2 | Upper panel: grand average (root mean square, RMS) of MEG temporal waveforms in auditory channels (20 auditory channels per subject) elicited by the four types of sounds (black bar). Time 0 indicates the onset of sound stimuli. Lower panel, left: RMS of MEG auditory response waveform in the first 200 ms time window. Lower panel, right: Grand average of corresponding contour maps in the window of ~150 ms (dotted rectangle) after sound onset.

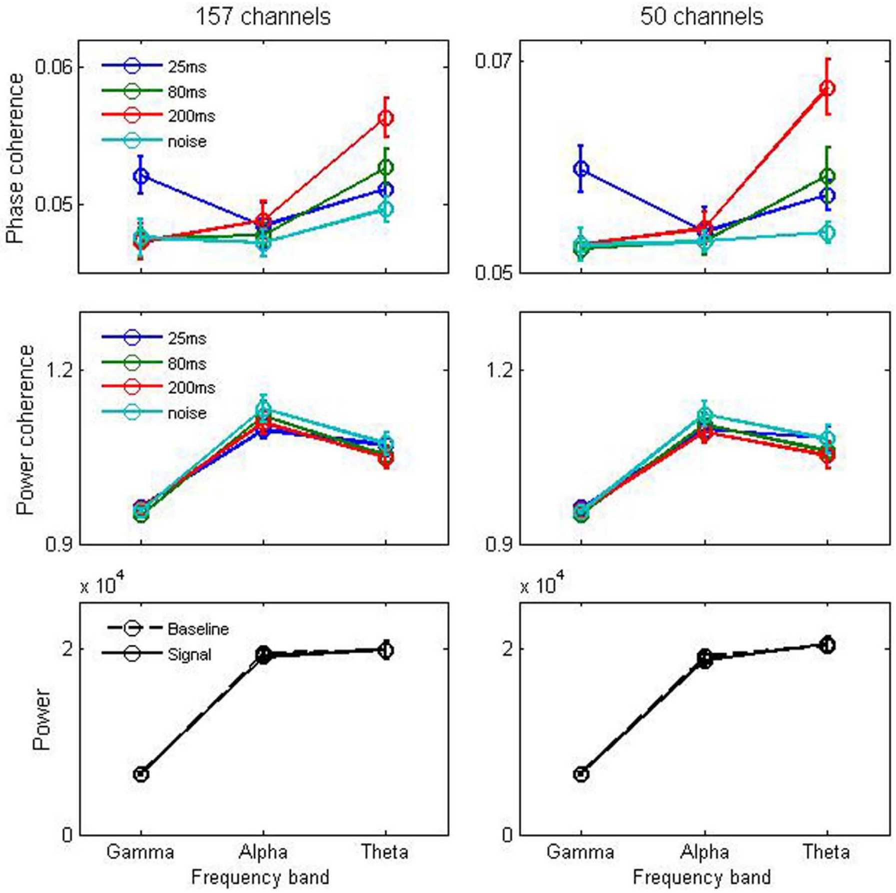

FIGURE 3 | Cross-trial phase and power coherence. Grand average of cross-trial phase and power coherence for all 157 MEG channels (left column) and for 50 frequency-dependent selected channels (right column). Upper panel: Cross-trial phase coherence at three different frequency bands (gamma: 38–42 Hz, alpha: 10–14 Hz, and theta: 4–6 Hz) for each of the four stimulus types (mean segment lengths of 25, 80, 200 ms, and noise). For the gamma and theta frequency bands, the matching stimulus (25 and 200 ms stimulus, respectively) elicits the strongest cross-trial phase coherence. Middle panel: cross-trial power coherence at three different frequency bands for each of the four stimulus types. Lower panel: power before (baseline) and after (signal) stimulus presentation at different frequency bands. Error bars represents standard error across 12 subjects.

capture the dynamics of phase and power response patterns in gamma frequency. Given this concern, we performed a control analysis in the gamma band (38∼42 Hz) using the induced wavelet transfer method (Complex Gaussian Wavelet), to determine the cross-trial phase and power coherence for each stimulus condition across all 157 channels. As was the case for the other analysis approach, stimuli with ∼25 ms mean segment duration elicited the strongest cross-trial phase coherence in the gamma band among the four stimulus conditions (one-way repeated ANOVA, $F(3, 33) = 3.22$, $p = 0.035$).

CROSS-TRIAL PHASE COHERENCE (*C*PHASE) DISTRIBUTION MAP

We characterized the cortical spatial distribution of the two temporal scale/phase patterns by studying the "*C*phase distribution map" in the frequency ranges of interest. As illustrated in **Figure 4**, the theta phase tracking, or the ∼200 ms time scale, mainly reflected an auditory cortical pattern (cf. Luo and Poeppel, 2007; Howard and Poeppel, 2010). A trend toward rightward lateralization for all of stimulus types with different temporal structure (**Figure 4**, lower panel) was observed, in which the matching stimulus (200 ms) elicited the strongest *C*phase values. Gamma phase tracking, on the other hand, shows a more distributed bilateral pattern (**Figure 4**, upper row) and the matching stimulus (25 ms) resulted in the strongest *C*phase topography among all stimulus types. Since the alpha rhythm did not show improved phase tracking with the corresponding stimulus (**Figure 3**), its phase tracking topography is much weaker than the other two distribution maps.

RIGHT HEMISPHERE LATERALIZATION OF THETA BAND *C*PHASE DISTRIBUTION MAP

To characterize the potential hemispheric lateralization of the *C*phase distribution maps, we compared the *C*phase values between left hemisphere (LH) and right hemisphere (RH) channels, for each of the four stimuli. As illustrated in **Figure 5**, gamma phase tracking shows no significant difference in *C*phase between LH and RH (**Figure 5**, upper panel, left) for the matching ∼25 ms stimulus (paired t-test, df = 11, $p = 0.79$), consistent with the corresponding bilateral *C*phase distribution map. Interestingly, theta

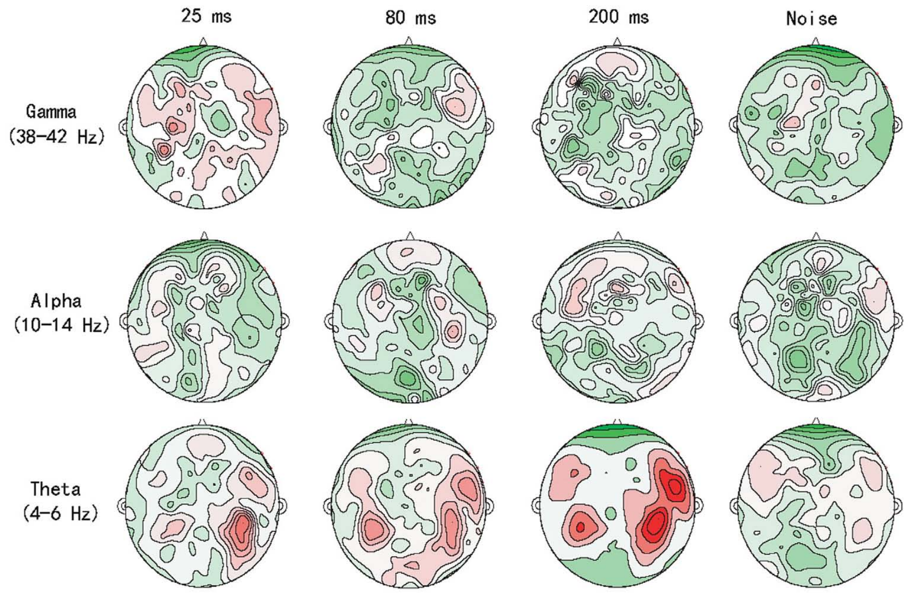

FIGURE 4 | Cross-trial phase coherence (*C*phase) distribution maps. Grand average of *C*phase distribution maps for gamma (upper panels), alpha (middle panels), and theta (lower panels) for the four stimulus types (25, 80, 200 ms, noise) from left to right. All maps are plotted on the same scale from 0.03 to 0.07 (green to red). Note that for gamma and theta bands (upper and lower rows), the ∼25 and the ∼200 ms stimuli respectively elicited strongest phase coherence among all four stimulus conditions, whereas for the alpha band (middle row), the ∼80 ms stimulus shows no notable difference.

phase tracking shows significantly larger *C*phase values in RH than in LH channels (**Figure 5**, lower panel) for the ∼200 ms stimulus (paired one-tailed t-test, df = 11, $p = 0.04$). This finding is reflected in the corresponding theta *C*phase distribution map, indicating a clear auditory cortex origin with RH lateralization.

The results are consistent with previous data using fMRI and MEG (Boemio et al., 2005; Luo and Poeppel, 2007) and reminiscent of patterns with similar lateralization (Giraud et al., 2007; Abrams et al., 2008). The weak alpha phase tracking did not reveal hemispheric lateralization effects (**Figure 5**, middle panel) for the matching ∼80 ms stimuli (paired t-test, df = 11, $p = 0.94$).

DISCUSSION

In this MEG experiment, we deployed non-speech stimuli with specified temporal structures to explore the neural correlates of processing over the different time scales. Three related hypotheses were investigated. First, we found that our analytic non-speech stimuli elicited neuronal phase tracking in the same manner as has been demonstrated repeatedly for speech signals. Second, we determined that temporal structure is neurally reflected in a non-uniform manner, in that neuronal oscillations phase-lock (and "sample") auditory stimuli over distinct time scales (∼25 and ∼200 ms). Third, the two cortical temporal scales – a longer one (associated preferentially with RH mechanisms) and a shorter one (represented more bilaterally) – undergo pure phase regularization and resetting to process and track incoming stimulus temporal transients, in an asymmetric manner.

The present results are consistent with our previous findings (Boemio et al., 2005; Luo and Poeppel, 2007) and support some current conjectures about functional anatomy and lateralization (Poeppel, 2003; Giraud et al., 2007; Hickok and Poeppel, 2007; Giraud and Poeppel, 2012a,b). Importantly, the data presented here reveal a potential mechanism underlying multi-time resolution processing: the sliding and resetting of intrinsic temporal windows. We provide a way to extract appropriate temporal processing information from the recorded brain signals that is also naturally linked to recent findings in neuronal interactions through phase synchronization (Womelsdorf et al., 2007).

Influential psychophysical research has shown that modulation frequencies below approximately 16 Hz suffice to yield intelligible speech, even when relatively few spectral bands are used (that is to say, the spectral composition of the stimulus is impoverished), and even when the carrier is noise rather than the fine structure associated with the original speech stimulus (Drullman et al., 1994a,b; Shannon et al., 1995; Kanedera et al., 1999; Elliott and Theunissen, 2009). Both behavioral and neurobiological imaging data (Ahissar et al., 2001; Luo and Poeppel, 2007) demonstrate compellingly that the integrity of the *low modulation frequency speech envelope* is required for successful intelligibility. More colloquially, the rate of syllables is a critical determinant of spoken language recognition. Zion-Golumbic et al. (in press) provide a recent perspective on the role of the speech envelope for parsing the signal and outline the role in attentional processes as well as for predictive processing, facilitated at this longer timescale. Ghitza (2011) describes a computational model that outlines the steps by which these rates lead to the parsing and decoding of speech input.

In complementary fashion to the delta–theta, longer timescale phenomena, it is clear that the rapidly modulated information contained in speech signals is important for decoding the input. Acoustic features such as burst duration, voice onset time, frequency excursion of formants, and other short duration signal

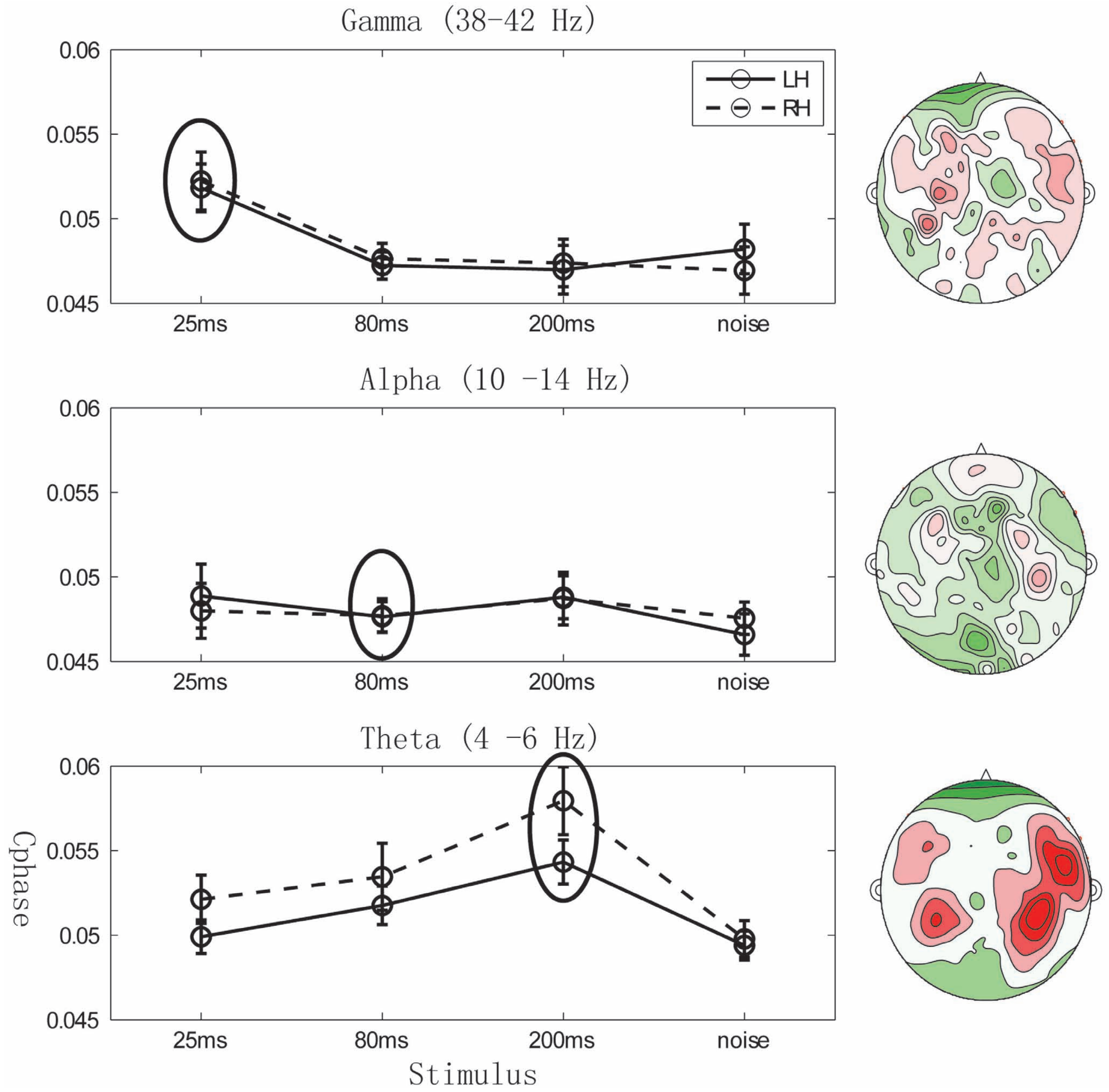

FIGURE 5 | *C*phase hemispheric lateralization. Grand average of *C*phase across left hemisphere (LH) channels and right hemisphere (RH) channels at different frequency bands for four different stimulus types. Error bars represents standard error across 12 subjects. Right contour maps: Grand average of "*C*phase distribution maps" in the gamma (upper), alpha (middle), and theta (lower) frequency bands for the corresponding matching stimulus (25, 80, and 200 ms respectively). All maps are plotted on the same scale (0.03–0.07 green to red).

attributes – often infelicitously summarized as the fine structure – play a critical role in the correct analysis of naturalistic spoken language. The seminal work of Fletcher (1953), Liberman (1996), Stevens (2000), and many others underscores the profound relevance of short duration, high modulation frequency acoustic cues.

Although a clear oversimplification, one useful subdivision is, therefore, between information carried at a time scale of roughly 150–300 ms (corresponding, roughly, to syllable duration) and information at a time scale in the 10 s of milliseconds (corresponding, roughly, to local short duration acoustic features). Both sources of information are likely crucial for recognition in ecological contexts. Indeed, recent models of speech perception argue that the syllabic scale, low temporal modulation frequency information may serve to *parse* the signal into manageable chunks whereas the shorter duration and higher modulation frequency information is likely used to *decode* the signal (Ghitza, 2011; Giraud and Poeppel, 2012a,b). Interestingly, such dual discrete temporal processing has also been suggested in visual perception (VanRullen and Koch, 2003; Holcombe, 2009). But both types of data are necessary for the brain to link the incoming acoustic information to stored mental representations, or, in short, words.

If such models are on the right track, evidence for phase tracking at both rates is necessary. In two of our recent MEG studies linking the modulation spectrum of speech to neural oscillations (Luo and Poeppel, 2007; Luo et al., 2010), the results, building on coherence analyses of the neural data, support the important role that low modulation frequency brain information plays in perceptual analysis of speech signals. A further experiment with naturalistic speech, now using a rather difference approach, namely mutual information analyses, provided more data for the generalization that the delta and theta bands in the neurophysiological response to speech play a privileged role (Cogan and Poeppel, 2011). Critically, there is consistent evidence for the position that intact information at these time scales is essential for successful intelligibility (Ahissar et al., 2001; Luo and Poeppel, 2007).

However, based on these data, some critical questions remained unanswered. Because in these studies few effects were visible in the gamma range, it has not been clear to what extent phase coherence analyses for speech would reflect higher-frequency, gamma band effects. At least for non-speech signals with the requisite structure, we can answer that question in the affirmative. Second, it had not been established to what extent the observed effects reflected speech-driven or acoustics-driven effects. Some data suggest the latter interpretation. One experiment, using speech as stimuli, highlights the issue: if listeners are presented with backward speech (with only a medium amount of exposure and no demonstrable intelligibility), the phase pattern of the theta band response *still* shows the characteristic response profile driven by theta phase (Howard and Poeppel, 2010). This finding argues that what the neuronal phase pattern is tracking does depend on acoustics of speech but does not depend on comprehension *per se.* The "onsets of reversed syllables" may be the causal factor in the acoustics. However, on a purely acoustic view, many (or any) auditory edges should lead to the response profile typically observed in such phase tracking experiments. What we find is that auditory edges, or acoustic discontinuities yielding phase resetting, are critical precursors. But, crucially, we show that not all edges are created equal. Information, and in this case edges distributed within two distinct time windows, are privileged, suggesting that the auditory worlds is "sampled" using two discontinuous temporal integration windows.

ACKNOWLEDGMENTS

This work was supported by NIH 2R01DC05660 to David Poeppel as well as the Ministry of Science and Technology of China Grants 2012CB825500, National Nature Science Foundation of China Grant 31171075, National Nature Science Foundation of China Grant 91120019. Jeff Walker provided expert technical support.

REFERENCES

Abrams, D. A., Nicol, T., Zecker, S., and 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., 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.

Boemio, A., Fromm, S., Braun, A., and Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nat. Neurosci.* 8, 389–395.

Cogan, G. B., and Poeppel, D. (2011). A mutual information analysis of neural coding of speech by low-frequency MEG phase information. *J. Neurophysiol.* 106, 554–563.

Ding, N., and Simon, J. Z. (2009). Neural representations of complex temporal modulations in the human auditory cortex. *J. Neurophysiol.* 102, 2731–2743.

Drullman, R., Festen, J. M., and Plomp, R. (1994a). Effect of reducing slow temporal modulations on speech reception. *J. Acoust. Soc. Am.* 95, 2670–2680.

Drullman, R., Festen, J. M., and Plomp, R. (1994b). Effect of temporal envelope smearing on speech reception. *J. Acoust. Soc. Am.* 95, 1053–1064.

Elliott, T. M., and Theunissen, F. E. (2009). The modulation transfer function for speech intelligibility. *PLoS Comput. Biol.* 5, e1000302. doi:10.1371/journal.pcbi.1000302

Fletcher, H. (1953). *Speech and Hearing in Communication.* Princeton: Van Nostrand Co.

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

Giraud, A. L., Kleinschmidt, A., Poeppel, D., Lund, T. E., Frackowiak, R. S. J., and Laufs, H. (2007). Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron* 56, 1127–1134.

Giraud, A. L., Lorenzi, C., Ashburner, J., Wable, J., Johnsrude, I., Frackowiak, R., and Kleinschmidt, A. (2000). Representation of the temporal envelope of sounds in the human brain. *J. Neurophysiol.* 84, 1588–1598.

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

Giraud, A. L., and Poeppel, D. (2012b). "Speech perception from a neurophysiological perspective," in *The Human Auditory Cortex. Springer Handbook of Auditory Research*, eds D. Poeppel, T. Overath, A. Popper, and R. Fay (New York: Springer), 225–260.

Greenberg, S., and Ainsworth, W. (2006). *Listening to Speech: An Auditory Perspective.* Mahwah: Lawrence Erlbaum Association, Inc.

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

Holcombe, A. O. (2009). Seeing slow and seeing fast: two limits on perception. *Trends Cogn. Sci. (Regul. Ed.)* 13, 216–221.

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.

Kanedera, N., Arai, T., Hermansky, H., and Pavel, M. (1999). On the Relative importance of various components of the modulation spectrum for automatic speech recognition. *Speech Commun.* 28, 43–55.

Liberman, A. M. (1996). *Speech: A Special Code.* Cambridge: The MIT Press.

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. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54, 1001–1010.

Overath, T., Kumar, S., von Kriegstein, K., and Griffiths, T. D. (2008). Encoding of spectral correlation over time in auditory cortex. *J. Neurosci.* 28, 13268–13273.

Overath, T., Zhang, Y., Sanes, D. H., and Poeppel, D. (2012). Sensitivity to temporal modulation rate and spectral bandwidth in the human auditory system: fMRI evidence. *J. Neurophysiol.* 107, 2042–2056.

Pellegrino, F., Coupé, C., and Marsico, E. (2011). Across-language perspective on speech information rate. *Language* 87, 539–558.

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, W. J., and van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philos. Trans. R. Soc. Lond.* 363, 1071–1086.

Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 336, 367.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science* 270, 303.

Stevens, K. N. (2000). *Acoustic Phonetics.* Cambridge: The MIT Press.

VanRullen, R., and Koch, C. (2003). Is perception discrete or continuous? *Trends Cogn. Sci. (Regul. Ed.)* 7, 207–213.

Wang, Y., Ding, N., Ahmar, N., Xiang, J., Poeppel, D., and Simon, J. Z. (2011). Sensitivity to temporal modulation rate and spectral bandwidth in the human auditory system: MEG evidence. *J. Neurophysiol.* 107, 2033–2041.

Womelsdorf, T., Schoffelen, J. M., Oostenveld, R., Singer, W., Desimone, R., Engel, A. K., and Fries, P. (2007). Modulation of neuronal interactions through neuronal synchronization. *Science* 316, 1609–1612.

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

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 22 February 2012; accepted: 10 May 2012; published online: 31 May 2012.

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

This article was submitted to Frontiers in Language Sciences, a specialty of Frontiers in Psychology.