Neuron
Article

Phase Patterns of Neuronal Responses Reliably Discriminate Speech in Human Auditory Cortex

Huan Luo[1,2] and David Poeppel[1,2,3,*]
[1] Neuroscience and Cognitive Science Program
[2] Department of Biology
[3] Department of Linguistics
University of Maryland College Park, College Park, MD 20742, USA
*Correspondence: dpoeppel@umd.edu
DOI 10.1016/j.neuron.2007.06.004

SUMMARY

How natural speech is represented in the auditory cortex constitutes a major challenge for cognitive neuroscience. Although many single-unit and neuroimaging studies have yielded valuable insights about the processing of speech and matched complex sounds, the mechanisms underlying the analysis of speech dynamics in human auditory cortex remain largely unknown. Here, we show that the phase pattern of theta band (4–8 Hz) responses recorded from human auditory cortex with magnetoencephalography (MEG) reliably tracks and discriminates spoken sentences and that this discrimination ability is correlated with speech intelligibility. The findings suggest that an ~200 ms temporal window (period of theta oscillation) segments the incoming speech signal, resetting and sliding to track speech dynamics. This hypothesized mechanism for cortical speech analysis is based on the stimulus-induced modulation of inherent cortical rhythms and provides further evidence implicating the syllable as a computational primitive for the representation of spoken language.

INTRODUCTION

Human speech signals contain rich dynamics in the amplitude and frequency domains, both of which contribute to speech comprehension (Shannon et al., 1995; Smith et al., 2002; Zeng et al., 2005), but the representation of such complex signals in human auditory cortex remains puzzling. This issue has been investigated extensively in animal neurophysiology using species-specific communication sounds (Machens et al., 2003, 2005; Narayan et al., 2006; Nelken, 2004; Nelken et al., 1999; Wang et al., 2003; Woolley et al., 2005). Many auditory cortical neurons produce stronger responses to conspecific vocalizations compared to other complex synthesized sounds (Hsu et al., 2004; Wang et al., 2003). Moreover, some recent studies demonstrate that single auditory neurons or ensembles fire in spiking patterns that reliably encode complex species-specific communication sounds, even in single trials (Machens et al., 2003; Narayan et al., 2006; Nelken, 2004; Wang et al., 2003).

Many neuroimaging studies with human subjects show that several cortical areas are significantly associated with speech processing. The cortical responses—mediated by large-scale assemblies of neurons—reflect detailed information about the spectral and temporal content of speech, words, or speech-like stimuli (Ahissar et al., 2001; Boemio et al., 2005; Elhilali et al., 2004; Giraud et al., 2000; Griffiths et al., 2004; Luo et al., 2006; Patel and Balaban, 2000; Scott et al., 2000, 2006; Suppes and Han, 2000; Suppes et al., 1997, 1998, 1999; Zatorre et al., 2002). However, the specific attributes of the macroscopic cortical responses collected in neuroimaging data that can track and discriminate natural speech signals are not well characterized, and how auditory information processing at such disparate scales is linked—what mechanisms can encode responses at the single-neuron level and couple these to responses in cortical cell assemblies—remains one of the most challenging questions in neuroscience (Logothetis et al., 2001; Shmuel et al., 2006).

In part, the present studies are motivated by previous work that identified correlations between neurophysiological responses as assessed by EEG and MEG and the acoustics of spoken language (Suppes et al., 1997, 1998, 1999; Suppes and Han, 2000; Ahissar et al., 2001). These studies were able to demonstrate that cortical responses in the time domain can discriminate single words and artificial simple sentences (Suppes et al., 1997); moreover, intelligibility (as tested with compression, i.e., the manipulation of acoustic envelope rate) correlated with auditory cortical responses (Ahissar et al., 2001). However, these experiments did not investigate intelligibility and discriminability in the same recording and using naturalistic materials. In addition, crucially, previous research did not speak to potential mechanisms underlying the analysis of spoken language. Here, we build on and extend the work by testing what kind of auditory cortical mechanism could form the basis for representing the acoustic structure of spoken sentences.

We hypothesized that the phase pattern of cortical rhythms might be one key representational mechanism, in particular rhythms commensurate with intelligible speech (Dau et al., 1997; Elhilali et al., 2003). This view is motivated by studies demonstrating that EEG and MEG signals are dominated by stimulus-induced changes in endogenous ongoing brain dynamics rather than by stimulus-evoked events (Makeig et al., 2002; Penny et al., 2002) and importantly, those inherent brain rhythms have been found to have functional significance in object perception (Engel et al., 2001; Hari and Salmelin, 1997). A final consideration derives from MEG experiments employing amplitude-modulated tone sequences. These show that the phase of the elicited response at the sound envelope modulation frequency reliably tracks the tone sequences (Luo et al., 2006; Patel and Balaban, 2000).

We recorded MEG signals from participants listening to spoken sentences and explored the phase-tracking hypothesis. To investigate whether this putative representational mechanism is correlated with speech intelligibility, we constructed for each sentence two additional types of degraded sentence signals (with different intelligibility levels) using the speech-noise chimera method (Smith et al., 2002). We found that the phase pattern of theta-band responses (4–8 Hz) from human auditory cortex (with right hemisphere lateralization) reliably discriminates the spoken sentence signals and that this tracking ability is correlated with sentence intelligibility in that theta phase tracking becomes less robust when the sentence stimulus is less intelligible. In addition, the theta-band power is not different prior to and during sentence presentation, confirming that it is the phase modulation of the intrinsic theta-band cortical rhythms that represents the incoming signals. Our results suggest that continuous speech is processed by an endogenous temporal window of ~200 ms (period of theta band), which resets and slides according to the speech dynamics. Because of the period duration, such a mechanism further implicates the syllable (mean duration crosslinguistically ~200 ms) as one computational primitive in cortical speech processing.

RESULTS

Theta-Band Phase Pattern Discriminates Sentence-Level Acoustics

To investigate whether information in the MEG responses can be used to discriminate between different sentences, we developed an analysis that identifies the cortical activity patterns relevant to the representation of specific sentences *in single trials*. We call the response to trials for the same sentence conditions "within-group" signals. Correspondingly, we constructed "across-group" signals by randomly mixing trials from different stimulus conditions (Figure 1A). For each condition (both within-group and across-group conditions) and each recorded channel (157 channel whole-head acquisition), a spectrotemporal analysis of each trial's response profile (between 0 and 50 Hz) was performed to calculate the phase and power pattern. If the phase pattern at specific frequencies successfully discriminates between sentences, as we hypothesized, the phase patterns of within-group signals should be more similar across trials than those of across-group signals. The crosstrial phase coherence of each within-group signal was calculated and compared to that of the corresponding across-group signal (see Experimental Procedures). The dissimilarity in crosstrial phase coherence between the within-group and across-group signals, termed the "phase dissimilarity function" (Figure 1B), was determined as a function of frequency for each MEG channel.

We observed well-defined peaks in the 4–8 Hz frequency range in this phase dissimilarity function in many channels (Figure 1B, upper row), indicating that the phase pattern in the theta band discriminates between the different sentence stimuli. To assess whether the observed phase-based discrimination ability is accompanied by a corresponding discrimination ability in the power of the theta-band response, we calculated the "power dissimilarity function," characterizing the difference in the across-trial power coherence between "within-condition" and "across-condition" signals. There were no significant peaks in this analysis (Figure 1B, bottom row), confirming that stimulus discrimination is based on pure phase information. Furthermore, to examine whether the theta-band phase tracking is accompanied by stimulus-evoked theta-band power increase, we also compared the theta band power in baseline and during stimulus presentation, which showed no difference (paired t test, $n = 6$, $p = 0.21$). This analysis underscores that it is the phase modulation of the intrinsic ongoing brain rhythm in the theta band that discriminates the different sentence stimuli.

Auditory Cortex Origin of Theta-Band Phase Tracking

We divided the phase dissimilarity function into the five canonical electrophysiological frequency bands (theta, α, $\beta 1$, $\beta 2$, and γ) and examined the corresponding spatial distributions. The "theta phase dissimilarity distribution map" showed a clear auditory cortex origin (Figure 1C), matched with the dipolar pattern for typical auditory-evoked field distributions (Figure 2). However, the spatial distributions for other frequency ranges were noisy and not indicative of localized underlying activity (Figure 1C). This visual analysis strengthens the argument that it is the phase of theta-band activity in auditory cortex that tracks the sentence stimuli.

Crucially, a theta phase dissimilarity distribution map with auditory origin was observed in every subject (Figure 2, middle). For comparison, the contour maps for the M100/N1m, the largest and most robust auditory response originating in superior temporal cortex, are shown for each subject (Figure 2, left). This response is generated in superior temporal cortex roughly 100 ms after sound onset (Lütkenhöner and Steinstrater, 1998) and was elicited here in a pretest using 1 kHz pure-tone pips. Despite large differences in response amplitude, the two spatial

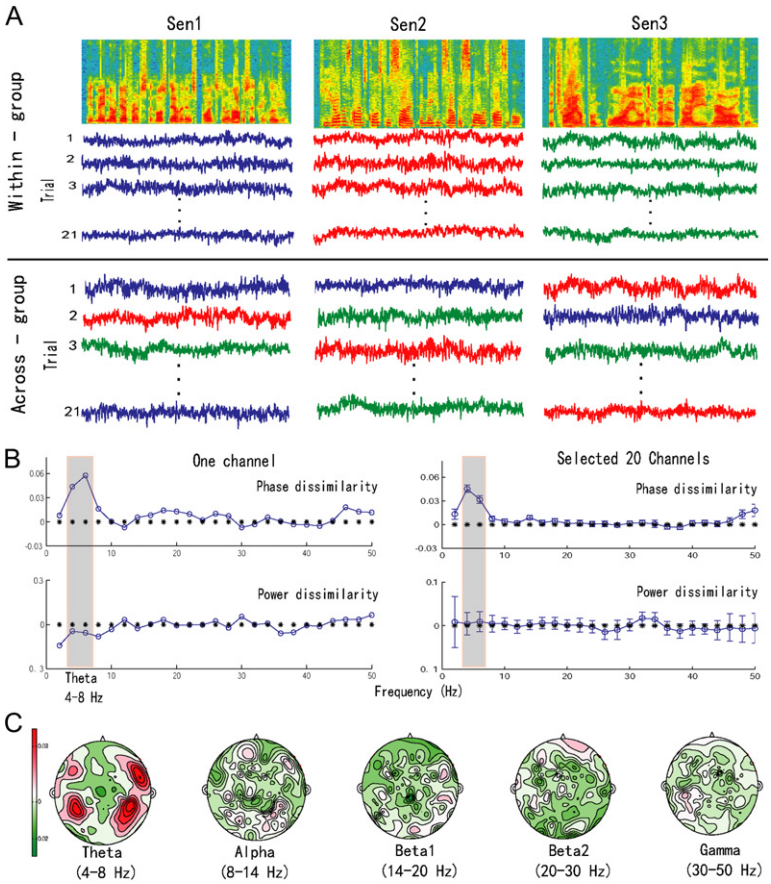

Figure 1. Spectrograms of Sentence Stimuli and Representative MEG Data for One Subject

(A) Example stimuli and single-trial responses (blue, red, green) from one channel. Within-group bins (same color) constitute responses to the same condition, across-group bins (mixed colors) to a random selection of trials across conditions.

(B) Left: phase dissimilarity function (upper) and power dissimilarity function (lower) as a function of frequency (0–50 Hz) for the same example channel. Gray box denotes the theta range (4–8 Hz) where the phase dissimilarity function shows peaks above 0. Right: averaged dissimilarity functions across 20 selected channels showing maximum phase dissimilarity values in theta band for same subject (mean and standard error).

(C) Phase dissimilarity distribution map for five frequency bands in same subject. Channels depicted with stronger red colors represent large phase dissimilarity values. The theta phase dissimilarity distribution map shows the "dipolar" distribution typical of auditory cortex responses.

maps show a good spatial match, consistent with an auditory cortex origin of the theta-band phase pattern. Note that the theta phase dissimilarity distribution map (Figure 2, middle) also shows right hemisphere lateralization. We tested the statistical significance of lateralization by comparing the averaged theta phase dissimilarity values (Figure 2) of all left hemisphere channels and all right hemisphere channels for each subject. A paired t test (two-tailed) shows significant asymmetry ($t = -3.35$, $df = 5$, $p = 0.02$).

Classification Performance

Having established the *sensitivity* of the theta-band phase pattern, it was of interest to evaluate its *specificity* with respect to sentence classification. For each subject, the 20 MEG channels with the largest theta phase dissimilarity were selected for further analysis. To verify that the theta-band phase pattern is sufficiently robust to discriminate among the sentence stimuli, a classification analysis was employed. For each sentence, the "theta phase pattern" as a function of time for one single trial response under one sentence condition was arbitrarily chosen as a template response for that sentence. The theta phase pattern of the remaining trials of all conditions was calculated, and their similarity to each of the three templates was defined as the distance to the templates. Responses were then classified to the closest sentence template. The classification was computed 1000 times for each of the 20 channels selected in each subject, by randomly choosing template combinations. The data from all subjects showed good classification performance (Figure 2, right). For each of the three sentences, trials were classified with higher proportion into the correct category than not, indicating that the theta phase pattern *could be relied on for sentence discrimination in single-trial responses*. Figure 3A shows the grand average of classification performance across the six subjects.

Discrimination Ability Correlates with Speech Intelligibility

Beyond successful sentence classification based on single-trial MEG data, it can be demonstrated that the phase of the theta-band response has compelling perceptual correlates. We show that (the discrimination ability of) the theta phase pattern correlates with intelligibility of the speech materials, by performing the same classification analysis on responses to degraded versions of the same sentences: speech-noise chimeras. Such a signal manipulation systematically changes the speech acoustics—and the associated perceptual intelligibility—by degrading the acoustic envelope and fine structure, both of which have been shown to be perceptually important

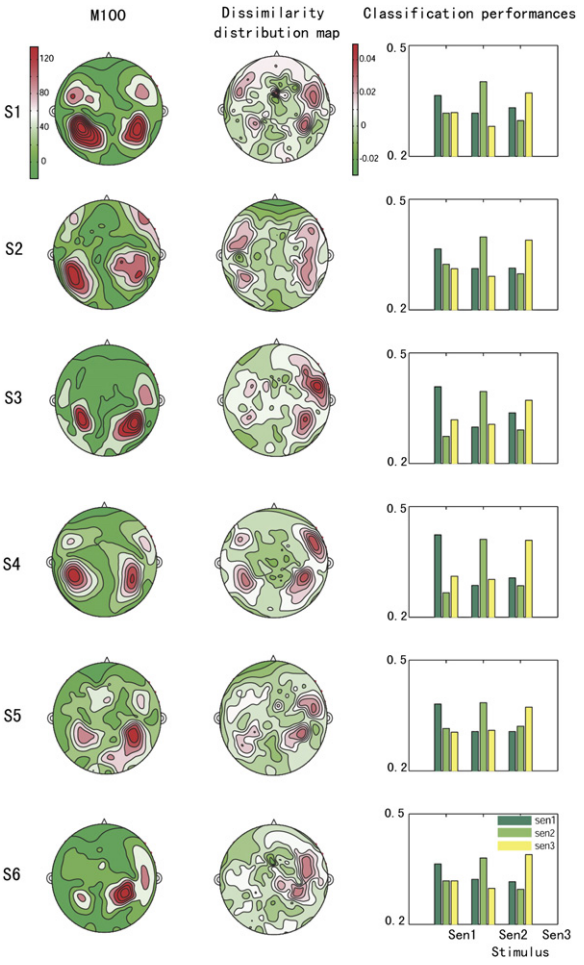

Figure 2. Auditory Cortex Identification, Theta Phase Dissimilarity Distribution Map, and Classification Performance for All Subjects

Left: M100 contour map for each subject. Red indicates large absolute response value at M100 peak latency. Middle: Theta phase dissimilarity distribution map. Right column: Classification performance. The horizontal axis represents the stimulus condition (Sen1, Sen2, Sen3) and the bar color represents the category (Sen1, Sen2, Sen3) this stimulus was classified to. The height of the bar represents the proportion that one single-trial to this stimulus condition (horizontal axis) was classified to this stimulus category (bar color). Note that the sum of the three clustered bars is 1.

(Drullman et al., 1994; Shannon et al., 1995; Zeng et al., 2005). We constructed two chimeras for each sentence, 4-band chimeras containing only acoustic envelope information (Env4) and 1-band chimaeras containing only fine structure information (Fin1) (see Experimental Procedures and see Figure S1 in the Supplemental Data available with this article online). The intelligibility level (proportion correct) of these degraded versions is 0.85 and 0.70, respectively, based on a previous study (Smith et al., 2002). The analysis of the MEG theta phase data reveals degraded classification performance (Figures 3B and 3C) compared to that of the original sentence stimuli

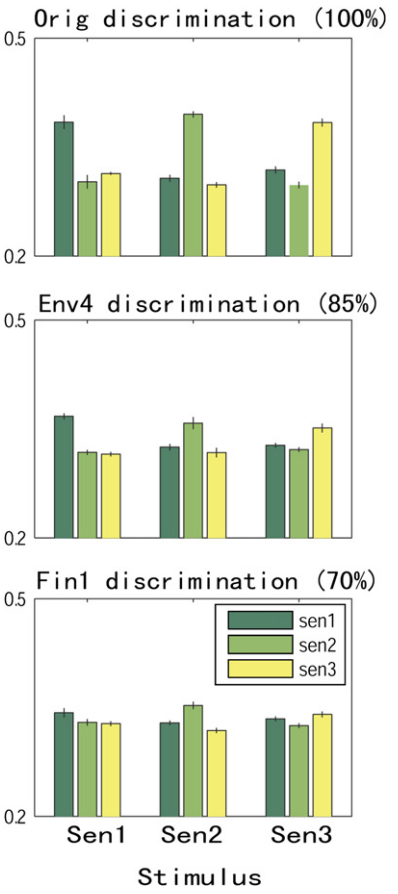

Figure 3. Classification Performance as a Function of Intelligibility (Mean and Standard Error)

Less-intelligible stimuli show parametrically degrading classification. Top: Discrimination of three original sentences. Middle: Discrimination of three Env4 sentences. Bottom: Discrimination of three Fin1 sentences. The percent value in each figure indicates the intelligibility score from a previous experiment (Smith et al., 2002).

(Figure 3A). This difference in classification performance was statistically significant (one-way ANOVA, $F(2,15) = 31.4$, $p < 0.001$), even when all the 157 MEG channels were pooled together (one-way ANOVA, $F(2,15) = 13.3$, $p < 0.001$). In sum, the less intelligible a sentence is, the less reliable is the theta phase pattern. Remarkably, this suggests that the pattern typically observed using the speech transmission index (Elhilali et al., 2003) is well captured by the theta-band phase pattern.

Acoustic Category Membership

Finally, we tested whether the theta phase pattern could reflect "category membership" of Env4 and Fin1 responses to the corresponding original (undistorted) speech signal by doing the same classification across all nine stimulus conditions (3 sentences × 3 stimulus manipulations). The grand average of the nine-condition classification performance is summarized in a 9-by-9 classification matrix for illustration purposes (Figure 4A).

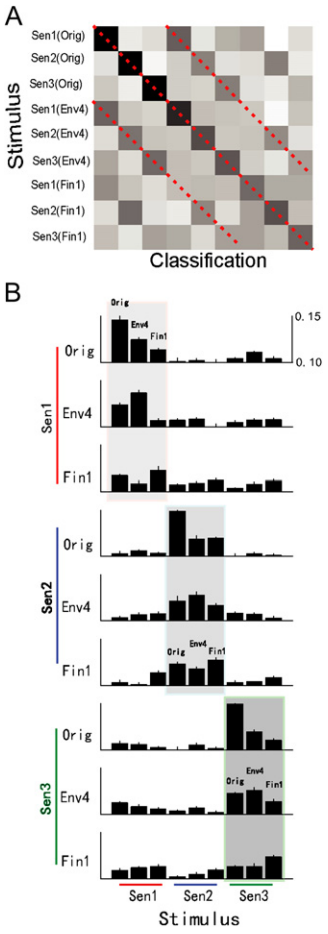

Figure 4. Theta Phase Pattern Reflects Category Membership

(A) Grand average of nine-condition classification matrix across six subjects. Each cell in the matrix represents the percent that a response trial for this stimulus condition (corresponding row) was classified to this stimulus category (corresponding column). The sum of each row is 1. Red lines indicate the main diagonal and subdiagonals, where the response was classified to stimulus itself or members in the same category (different versions of same sentence).

(B) Classification histograms for each of the nine stimulus conditions (3 sentences × 3 manipulated conditions). Rectangles indicate the range of corresponding correct category membership. For example, for all three versions of sentence 1 denoted by red vertical line (upper three rows), the rectangle covers the stimulus conditions all belonging to sentence 1 and should be classified into with higher percent than into other rectangles. Error bars indicate the standard error across six subjects.

The elements on main and subdiagonal axes denoted by red lines indicate the correct classification to the stimulus condition itself and the classification to other versions of the same sentence, respectively. These diagonal axes more or less showed peak values. Such clustering of different versions of the same sentence is shown more explicitly in Figure 4B. The three versions (Orig, Env4, and Fin1) of each sentence were predominantly classified into the corresponding sentence category (rectangular boxes) rather than into other groups. Moreover, among the three versions of each sentence, Fin1 stimuli showed the lowest classification performance, in accordance with the corresponding lower intelligibility scores.

Classification Performance Develops over Time

We examined the time course of the classification performance in terms of the theta-band phase pattern in each trial. We extracted the temporal segment (first 500 ms, first 1000 ms, first 2000 ms, first 3000 ms, and first 4000 ms) of recorded MEG responses and tested the same classification performance, as before. Interestingly, we observed a gradual development of the classification ability based on theta phase pattern. Specifically, the correct classification begins to emerge around 2000 ms from the beginning of speech sentence stimulus onset, as shown in Figure 5. We performed a one-way repeated-measures ANOVA on the classification performance at 500 ms, 1000 ms, and 2000 ms post-stimulus onset and confirmed a significant effect of time (one-way ANOVA, $F(2,10) = 4.12$, $p = 0.05$).

Modulation and Rate Controls

Originally, we amplitude modulated all of our sentence stimuli at 50 Hz, because initially we expected to find that some properties of 50 Hz responses could track speech dynamics. This hypothesis was based on previous experiments (Patel and Balaban, 2000; Luo et al., 2006) that showed that in amplitude-modulated tone sequences the response phase at the amplitude modulation frequency could track the tone sequence. We show that the observed theta-band phase discrimination ability does not depend on the 50 Hz amplitude modulation of the sentences, because all the stimuli were amplitude modulated at 50 Hz, and the observed discrimination ability was in the theta band, far away from the 50 Hz range. To verify this point, we ran control recordings using the same sentence stimuli without 50 Hz amplitude modulation and observed good classification performance (Figure 6, upper panel) based on the theta-band phase pattern and reasonable auditory cortex origin, supporting the argument that the observed theta-band phase tracking is not related to the 50 Hz amplitude modulation.

Finally, to explore the possibility that it is the theta-band power in the modulation spectra of the sentences themselves that drives the observed theta-band phase tracking, we ran a control recording using the same sentence stimuli at a compression ratio of 0.5, which has a dramatically different acoustic structure compared to the original speech while still remaining reasonably intelligible. We still observed adequate theta phase classification performance with auditory cortex origin (Figure 6, lower panel). This suggests that the theta-band phase pattern is not simply stimulus-acoustics driven, but closely related to the intrinsic *cortical* processing of speech. Furthermore, we analyzed one subject's MEG responses to an unintelligible version of the same sentences, Env1 chimeras that contain only acoustic envelope information (Figure S1), and found that the theta phase tracking disappeared

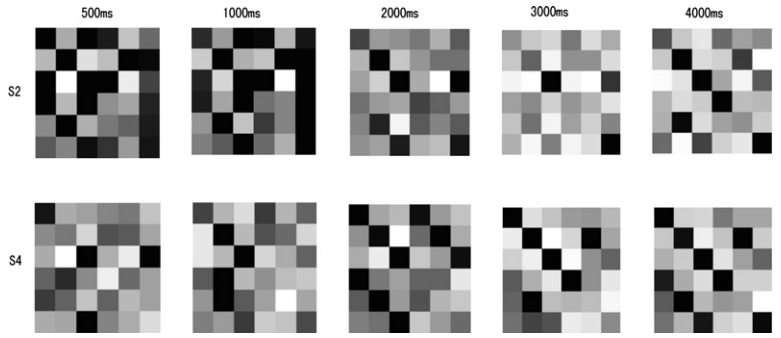

Figure 5. Classification Performance Develops over Time in Each Trial

Sample classification matrices as a function of integration time for two subjects. A six-condition (Original and Env4 versions of three sentences) classification analysis is shown. For example, 500 ms classification performance was calculated on only the first 500 ms of response, 1000 ms classification performance was calculated on the first 1000 ms of response, and so on. Unsurprisingly, because of the long period of theta (~200 ms), the MEG-recorded response must be collected over several periods before it becomes a robust discriminator. For subject 2, robust discrimination ability emerged around 2000 ms, and for subject 4, the discrimination ability emerged around 3000 ms.

(Figure S2), confirming the tight relationship between theta phase tracking and speech intelligibility.

DISCUSSION

We demonstrate that specific response attributes in *single trials* of MEG-derived auditory cortical responses suffice to discriminate among sentence-level acoustic stimuli. In particular, the ongoing phase pattern of endogenous theta-band responses in human auditory cortex robustly tracks sentence-level acoustics associated with intelligible speech. The discrimination performance evolves over the time of a trial and is strongly present by 1000–2000 ms post-stimulus onset. The ability to distinguish among stimuli is correlated with sentence intelligibility: the less intelligible the speech signal, the worse is the theta phase tracking performance. The observed pattern is consistent with a single or a complex generator in auditory cortex (Figure 2). We believe that the functional connectivity across areas is likely to form the relevant substrate (see, e.g., Price et al. [2005]). This view is also more consistent with our own functional anatomic perspective (Hickok and Poeppel, 2007). Accordingly, it is our hypothesis that the measured theta response reflects the interaction between core and belt (and perhaps parabelt) auditory areas. Cumulatively, the data demonstrate a tight link between the ability of auditory cortical neuronal populations to employ theta-band phase-tracking and the acoustic prerequisites of speech intelligibility.

Modulation of Ongoing Cortical Dynamics

A key aspect of our results concerns the nature of the phase tracking mechanism. Phase tracking was not accompanied by corresponding tracking of theta-band *power* (and theta-band power increases from baseline), suggesting that our data are a consequence of a pure phase modulation mechanism. It has been argued that event-related potentials in MEG/EEG responses are generated by a superposition of evoked oscillations at various frequencies and that in response to stimulus presentation these intrinsic rhythms undergo significant phase resetting or amplitude changes (Basar, 1998; Engel et al., 2001; Hari and Salmelin, 1997; Llinas, 2001; Makeig et al., 2002; Penny et al., 2002). An important recent study recording from A1 of awake macaques (Lakatos et al., 2007) revealed phase modulation in the context of a multisensory interaction mechanism: somatosensory inputs enhanced auditory processing by resetting the phase of ongoing neuronal oscillations in A1 so that the accompanying auditory input arrived during a high-excitability phase. The "theta phase tracking" observed here concurs with these findings and in turn supports an interpretation of MEG/EEG activity as representing endogenous brain states and stimulus-induced modulation (e.g., phase modulation) of these rhythms that are core attributes of the system.

Cortical Processing of Speech Signals

The acoustic structure of human speech contains rich dynamics on multiple temporal scales (Rosen, 1992), from ~20 ms to ~200 ms (longer time scales, at the phrasal level, are not relevant in the present context). Accumulating evidence from speech recognition studies demonstrates that comprehension does not require a detailed spectral analysis of the signal and only a coarse representation suffices (Drullman et al., 1994; Greenberg and Ainsworth, 2006; Shannon et al., 1995). Furthermore, the theta band (4–8 Hz) corresponds to a temporal window of 125–250 ms, matched to mean syllable length across languages (Greenberg et al., 2003; Greenberg and Ainsworth, 2006). Critically, the syllable has been suggested as a fundamental unit for speech perception and production, and robust information regarding the sequence of syllables in continuous speech is essential for spoken-language understanding (Greenberg et al., 2003; Greenberg and Ainsworth, 2006). The observed tracking ability of the theta-band response, a relatively long temporal processing scale of ~200 ms, correlates well with these data and suggests that speech signals are processed (among others) by a relatively slow syllabic-level analysis

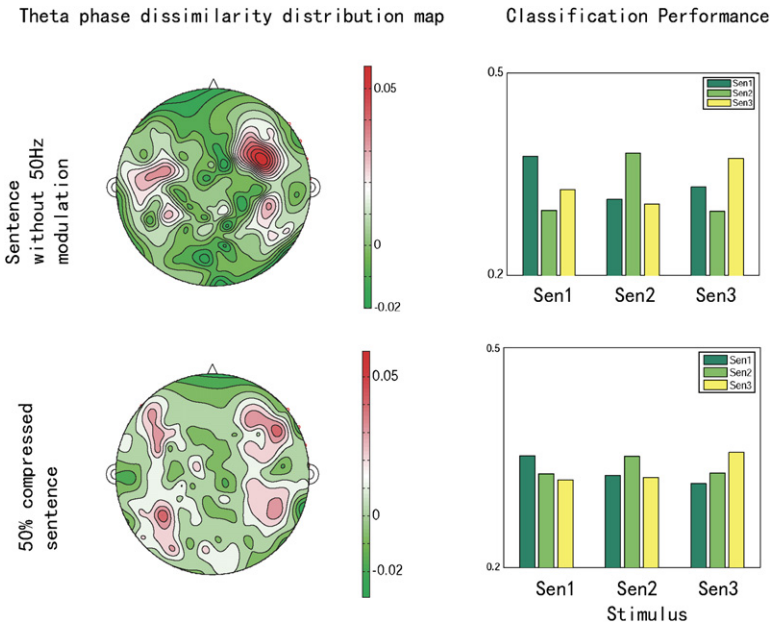

Figure 6. Performance of Two Control Subjects

Upper panels: contour map and classification performance of one control subject using three sentences without amplitude modulation. Lower panels: contour map and classification performance of one subject using the same three sentences at a compression ratio of 0.5.

rhythm in human auditory cortex. The theta phase patterns for distinct sentence stimuli presumably differ due to the variation in syllable structure and timing across sentences (given English syllable structure—a different result might be obtained in a syllable-timed language such as French). The observed correlation between theta phase tracking and intelligibility could be due to the blurring of syllable structure introduced by acoustically degrading sentences. The *sustained* theta-band phase tracking for 0.5-compressed speech and its *disappearance* for fully unintelligible sentential stimuli (Figure S2) imply that tracking is not simply stimulus-acoustics driven, but rather reflects an internal stable processing rhythm that is ideally suited to match the gross statistical temporal structure of speech. The information at other temporal scales, for example at the segmental scale (20–80 ms duration) is also crucial for speech perception (Poeppel, 2003; Hickok and Poeppel, 2007) but may not be easily observed and efficiently elicited in the current experimental and analysis paradigms. Why do our data show robust phase resetting at theta, but not at other frequencies? Because the materials and task we used demanded an assessment of intelligibility, and since intelligibility is predominantly mediated by low modulation frequency syllabic information, we hypothesized that the cortical response commensurate with that time scale, ~150–250 ms, would be preferentially modulated. We surmise that if we change the task demands, for example by requiring attention to specific, perhaps phonemic, representations, we will upregulate other response frequencies, including the γ response.

Similarities and Differences to Related Studies

We observe that the ongoing theta phase pattern reliably represents and discriminates spoken sentences, in agreement with previous work demonstrating low-frequency (<10Hz) brain wave representation of words and simple sentences (Suppes and Han, 2000; Suppes et al., 1997, 1998). The correlation between the phase tracking and speech intelligibility also matches relevant previous research, in particular an MEG experiment revealing that cortical responses show decreased tracking performance for compressed speech (Ahissar et al., 2001). However, there are several distinct and novel aspects of the findings presented here. First, we employed natural continuous spoken sentences and therefore the observed discrimination ability of theta phase pattern was at the ecologically natural sentence-level, whereas in previous work (Suppes et al., 1998) artificial short sentences designed to have clearly delineated word boundaries were used (equivalent to having spaces between printed words), and their results thus mainly indicated word-level representation in brain waves. Second, we systematically changed speech intelligibility by degrading both acoustic envelope and fine-structure information, a method often used in speech recognition studies. In contrast, the previous work (notably Ahissar et al. [2001]) used a very different intelligibility manipulation, compressed speech, in which only the acoustic envelope rate was modified. These authors also employed a different analysis method (PCA) and found that the cortical response failed to track the speeded acoustic envelope of speech stimuli with the accompanying decrease in intelligibility. In addition, they did not report any sentence-level discrimination ability in the cortical response. Third, we discovered a natural speech representation mechanism—*phase modulation of the internal theta rhythm*—that was neither observed nor implicated in previous studies. In sum, our experiments to our knowledge are the first to directly show that a special cortical response mechanism, the theta

phase pattern, plays a central role in encoding natural spoken sentences and has compelling perceptual correlates.

Two Hundred Millisecond Temporal Window

It has been hypothesized that perception relies on discrete processing epochs, and that the external stimulus is translated into internal information "chunks" on certain temporal scales, a view that accounts for many psychophysical results (Poeppel, 2003; Pöppel, 1997; VanRullen and Koch, 2003). Such a discrete sampling window concept is partially supported by the observation of cortical oscillations at certain frequencies. Our results suggest that sentence stimuli are continuously segmented and processed by an endogenous temporal window of ~200 ms duration, a value commensurate with one crucial aspect of the statistical temporal structure of speech, roughly the syllable flow, and therefore are also matched to the discrete sampling processing concept. The putative sampling window of ~200 ms—biased toward the right hemisphere in our data as well as in other recent studies (Boemio et al., 2005; Zatorre et al., 2002), even though we are presenting speech—undergoes a timing regularization and resets in a pattern closely tied to the dynamic structure of speech. Such an explanation also supports the rightward lateralization of a hypothesized long temporal window in speech and hearing (Boemio et al., 2005; Poeppel, 2003). Further studies using complex sounds with similar temporal structure need to be done to investigate whether the observed theta phase tracking is specific to speech processing or reflects a generic computation in human auditory cortex.

EXPERIMENTAL PROCEDURES

Subjects and MEG Data Acquisition

Six right-handed native English speakers provided informed consent before participating in the experiment. 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 and an online 100 Hz analog low-pass filter, with no high-pass filtering.

Stimuli

Three spoken sentences ("*It made no difference that most evidence points to an opposite conclusion.*"; "*He held his arms close to his sides and made himself as small as possible.*"; "*The triumphant warrior exhibited naive heroism.*") with sampling frequency 16 kHz were selected from the DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus (TIMIT). Two of the sentences are spoken by a female and one is spoken by a male, and they range in duration from 4000–4700 ms. For each sentence, we constructed four types of speech-noise chimeras (Env4, Fin1, Env1, Fin8), the spectrograms for which are shown in the Supplemental Data. These speech-noise chimaeras contain speech information in either their envelope (ENV) or their fine structure (FIN); another important manipulated variable is the number of frequency bands into which the signal is split (Smith et al., 2002). The intelligibility scores for Env4, Fin1, Env1, and Fin8 were shown to be 0.85, 0.7, 0.05, and 0.2, respectively (Smith et al., 2002). Correspondingly, they can be separated into intelligible (original, Env4, and Fin1) and unintelligible (Env1 and Fin8) speech signals. The original and chimeric signals were then amplitude modulated at 50 Hz.

Experimental Procedures

In an initial scan, the participants were presented with 1 kHz tone pips (duration 50 ms) to determine their M100 evoked responses. Subjects were then told to listen to the (original and degraded) versions of spoken sentences. On each speech trial, two sentences were presented sequentially with a 1 s interval between them; subjects were instructed to indicate by button-press whether they were same or different sentences. The first one was always drawn from the intelligible set (original, Env4, Fin1), the second one was always unintelligible (Env1, Fin8). Each of the nine intelligible conditions (three sentences, three intelligible conditions) was presented 21 times at a comfortable loudness level (~70 dB). Eleven other duration-matched sentences from the TIMIT database were selected and their unintelligible versions (Env1, Fin8) were constructed. These unintelligible speech stimuli were randomly selected as the second stimulus in each speech trial. Only cortical responses to intelligible stimuli were extracted for further analysis.

Data Analysis

All response trials to the same speech stimulus (21 trials of each) are termed within-group signals (three within-group signals corresponding to three original sentences). Seven response trials (one-third of the 21 trials for each stimulus condition) are randomly chosen from each of the three within-group signals and combined to construct a 21-trial across-group signal. Three across-group signals are constructed by repeating the random combination procedure three times. For each of the six 21-trial signals (three within-group and three across-group signals), the spectrogram of the first 4000 ms of each single trial response was calculated using a 500 ms time window in steps of 100 ms for each of the 157 MEG recording channels. The phase and power were calculated as a function of frequency and time and were stored for further analysis. The "crosstrial phase coherence" (*Cphase*) and "crosstrial power coherence" (*Cpower*) were calculated as

$$Cphase_{ij} = \left(\frac{\sum_{n=1}^{N}\cos(\theta_{nij})}{N}\right)^2 + \left(\frac{\sum_{n=1}^{N}\sin(\theta_{nij})}{N}\right)^2$$

$$Cpower_{ij} = \frac{\sqrt{\frac{\sum_{n=1}^{N}\left(A_{nij}^2 - \overline{A_{ij}^2}\right)^2}{N}}}{\overline{A_{ij}^2}}$$

where θ_{nij} and A_{nij} are the phase and amplitude at the frequency bin i and temporal bin j in trial n, respectively. *Cphase* is in the range of [0 1]. Note that a larger *Cphase* value corresponds to strong crosstrial phase coherence, whereas a smaller *Cpower* value corresponds to strong crosstrial power coherence. These calculated crosstrial coherence parameters (*Cphase*,*Cpower*) were compared between each of the three within-group signals and each of three across-group signals separately. The dissimilarity function for each frequency bin i was defined as

$$Dissimilarity\_phase_i = \frac{\sum_{j=1}^{J} Cphase_{ij,within}}{J} - \frac{\sum_{j=1}^{J} Cphase_{ij,across}}{J}$$

$$Dissimilarity\_power_i = \frac{\sum_{j=1}^{J} Cpower_{ij,across}}{J} - \frac{\sum_{j=1}^{J} Cpower_{ij,within}}{J}.$$

The resulting three dissimilarity functions (three within-group-across-group pairs) were averaged. Each of the 157 MEG channels has two dissimilarity functions as a function of frequency (*Dissimilarity_phase*, *Dissimilarity_power*), in which a value significantly above 0 indicates larger crosstrial coherence of within-group signals than across-group signals.

The *Dissimilarity_phase* function was then divided into the five canonical electrophysiological frequency bands (theta, 4∼8 Hz; alpha, 8∼14 Hz; beta1, 14∼20 Hz; beta2, 20∼30 Hz; gamma, 30∼50 Hz), and the average values within each frequency band were calculated, resulting in five *Dissimilarity_phase* values for the five frequency bands, respectively. Phase dissimilarity distribution maps for the five frequency bands were then constructed separately in terms of the corresponding *Dissimilarity_phase* value of all 157 channels in this frequency band. For each subject, the 20 channels with maximum *Dissimilarity_phase* value in the theta band (4∼8 Hz) were selected for further classification and grand average analysis. Note that the selected 20 channels correspond to channels with stronger red color in the theta phase dissimilarity distribution map.

In the classification analysis, the classification was computed 1000 times, for all 21 trials for each stimulus condition and for all the selected 20 channels in each subject, by randomly choosing template combinations. The classification results were then averaged to be in the range from 0 to 1, indicating the percent that an empirical single-trial response to a specific stimulus condition is classified to one stimulus condition.

Supplemental Data

The Supplemental Data can be found with this article online at http://www.neuron.org/cgi/content/full/54/6/1001/DC1/.

ACKNOWLEDGMENTS

This work is supported by NIH R01 DC05660 to D.P. We thank Jeff Walker for his technical assistance, Susannah Hoffman for help with the manuscript, and Allen R. Braun, Catherine Carr, Mary Howard, Shihab Shamma, and Jonathan Z. Simon for their thoughtful comments on the work.

Received: March 26, 2007
Revised: May 11, 2007
Accepted: June 4, 2007
Published: June 20, 2007

REFERENCES

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

Basar, E. (1998). Brain function and oscillations (Berlin: Springer).

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

Dau, T., Kollmeier, B., and Kohlrausch, A. (1997). Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration. J. Acoust. Soc. Am. *102*, 2906–2919.

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

Elhilali, M., Chi, T., and Shamma, S.A. (2003). A spectro-temporal modulation index (STMI) for assessment of speech intelligibility. Speech Comm. *41*, 331–348.

Elhilali, M., Fritz, J.B., Klein, D.J., Simon, J.Z., and Shamma, S.A. (2004). Dynamics of precise spike timing in primary auditory cortex. J. Neurosci. *24*, 1159–1172.

Engel, A.K., Fries, P., and Singer, W. (2001). Dynamic predictions: oscillations and synchrony in top-down processing. Nat. Rev. Neurosci. *2*, 704–716.

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.

Greenberg, S., and Ainsworth, W. (2006). Listening to Speech: an Auditory Perspective (Mahwah, NJ: Erlbaum).

Greenberg, S., Carvey, H., Hitchcock, L., and Chang, S. (2003). Temporal properties of spontaneous speech—a syllable-centric perspective. J. Phonetics *31*, 465–485.

Griffiths, T.D., Warren, J.D., Scott, S.K., Nelken, I., and King, A.J. (2004). Cortical processing of complex sound: a way forward? Trends Neurosci. *27*, 181–185.

Hari, R., and Salmelin, R. (1997). Human cortical oscillations: a neuromagnetic view through the skull. Trends Neurosci. *20*, 44–49.

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

Hsu, A., Woolley, S.M., Fremouw, T.E., and Theunissen, F.E. (2004). Modulation power and phase spectrum of natural sounds enhance neural encoding performed by single auditory neurons. J. Neurosci. *24*, 9201–9211.

Lakatos, P., Chen, C.M., O'Connell, M.N., Mills, A., and Schroeder, C.E. (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. Neuron *53*, 279–292.

Llinas, R.R. (2001). I of the Vortex: from Neurons to Self (Cambridge, MA: The MIT Press).

Logothetis, N.K., Pauls, J., Augath, M., Trinath, T., and Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. Nature *412*, 150–157.

Luo, H., Wang, Y., Poeppel, D., and Simon, J.Z. (2006). Concurrent encoding of frequency and amplitude modulation in human auditory cortex: MEG evidence. J. Neurophysiol. *96*, 2712–2723.

Lütkenhöner, B., and Steinstrater, O. (1998). High-precision neuromagnetic study of the functional organization of the human auditory cortex. Audiol. Neurootol. *3*, 191–213.

Machens, C.K., Schutze, H., Franz, A., Kolesnikova, O., Stemmler, M.B., Ronacher, B., and Herz, A.V.M. (2003). Single auditory neurons rapidly discriminate conspecific communication signals. Nat. Neurosci. *6*, 341–342.

Machens, C.K., Gollisch, T., Kolesnikova, O., and Herz, A.V. (2005). Testing the efficiency of sensory coding with optimal stimulus ensembles. Neuron *47*, 447–456.

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

Narayan, R., Grana, G., and Sen, K. (2006). Distinct time scales in cortical discrimination of natural sounds in songbirds. J. Neurophysiol. *96*, 252–258.

Nelken, I. (2004). Processing of complex stimuli and natural scenes in the auditory cortex. Curr. Opin. Neurobiol. *14*, 474–480.

Nelken, I., Rotman, Y., and Bar Yosef, O. (1999). Responses of auditory-cortex neurons to structural features of natural sounds. Nature *397*, 154–157.

Patel, A.D., and Balaban, E. (2000). Temporal patterns of human cortical activity reflect tone sequence structure. Nature *404*, 80–84.

Penny, W.D., Kiebel, S.J., Kilner, J.M., and Rugg, M.D. (2002). Event-related brain dynamics. Trends Neurosci. *25*, 387–389.

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

Pöppel, E. (1997). A hierarchical model of temporal perception. Trends Cogn. Sci. *1*, 56–61.

Price, C., Thierry, G., and Griffiths, T. (2005). Speech-specific auditory processing: where is it? Trends Cogn. Sci. *9*, 271–276.

Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. Philos. Trans. R. Soc. Lond. B Biol. Sci. *336*, 367–373.

Scott, S.K., Blank, C.C., Rosen, S., and Wise, R.J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. Brain *123*, 2400–2406.

Scott, S.K., Rosen, S., Lang, H., and Wise, R.J. (2006). Neural correlates of intelligibility in speech investigated with noise vocoded speech—a positron emission tomography study. J. Acoust. Soc. Am. *120*, 1075–1083.

Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). Speech recognition with primarily temporal cues. Science *270*, 303–304.

Shmuel, A., Augath, M., Oeltermann, A., and Logothetis, N.K. (2006). Negative functional MRI response correlates with decreases in neuronal activity in monkey visual area V1. Nat. Neurosci. *9*, 569–577.

Smith, Z.M., Delgutte, B., and Oxenham, A.J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. Nature *416*, 87–90.

Suppes, P., and Han, B. (2000). Brain-wave representation of words by superposition of a few sine waves. Proc. Natl. Acad. Sci. USA *97*, 8738–8743.

Suppes, P., Lu, Z.L., and Han, B. (1997). Brain wave recognition of words. Proc. Natl. Acad. Sci. USA *94*, 14965–14969.

Suppes, P., Han, B., and Lu, Z.L. (1998). Brain-wave recognition of sentences. Proc. Natl. Acad. Sci. USA *95*, 15861–15866.

Suppes, P., Han, B., Epelboim, J., and Lu, Z.L. (1999). Invariance between subjects of brain wave representations of language. Proc. Natl. Acad. Sci. USA *96*, 12953–12958.

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

Wang, X.Q., Lu, T., and Liang, L. (2003). Cortical processing of temporal modulations. Speech Comm. *41*, 107–121.

Woolley, S.M., Fremouw, T.E., Hsu, A., and Theunissen, F.E. (2005). Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. Nat. Neurosci. *8*, 1371–1379.

Zatorre, R.J., Belin, P., and Penhune, V.B. (2002). Structure and function of auditory cortex: music and speech. Trends Cogn. Sci. *6*, 37–46.

Zeng, F.G., Nie, K., Stickney, G.S., Kong, Y.Y., Vongphoe, M., Bhargave, A., Weit, C.G., and Cao, K. (2005). Speech recognition with amplitude and frequency modulations. Proc. Natl. Acad. Sci. USA *102*, 2293–2298.