

# A mutual information analysis of neural coding of speech by low-frequency MEG phase information

Gregory B. Cogan and David Poeppel

*J Neurophysiol* 106:554-563, 2011. First published 11 May 2011; doi:10.1152/jn.00075.2011

## You might find this additional info useful...

---

This article cites 30 articles, 6 of which can be accessed free at:

</content/106/2/554.full.html#ref-list-1>

This article has been cited by 2 other HighWire hosted articles

### **EEG Phase Patterns Reflect the Selectivity of Neural Firing**

Benedict Shien Wei Ng, Nikos K. Logothetis and Christoph Kayser

*Cereb. Cortex*, February , 2013; 23 (2): 389-398.

[\[Abstract\]](#) [\[Full Text\]](#) [\[PDF\]](#)

### **EEG Phase Patterns Reflect the Selectivity of Neural Firing**

Benedict Shien Wei Ng, Nikos K. Logothetis and Christoph Kayser

*Cereb. Cortex*, February 17, 2012; .

[\[Abstract\]](#) [\[Full Text\]](#) [\[PDF\]](#)

Updated information and services including high resolution figures, can be found at:

</content/106/2/554.full.html>

Additional material and information about *Journal of Neurophysiology* can be found at:

<http://www.the-aps.org/publications/jn>

---

This information is current as of September 18, 2014.

# A mutual information analysis of neural coding of speech by low-frequency MEG phase information

Gregory B. Cogan<sup>1,2,3</sup> and David Poeppel<sup>2</sup>

<sup>1</sup>Neuroscience and Cognitive Science, University of Maryland College Park, College Park, Maryland; and <sup>2</sup>Department of Psychology and <sup>3</sup>Center for Neural Science, New York University, New York, New York

Submitted 27 January 2011; accepted in final form 10 May 2011

**Cogan GB, Poeppel D.** A mutual information analysis of neural coding of speech by low-frequency MEG phase information. *J Neurophysiol* 106: 554–563, 2011. First published May 11, 2011; doi:10.1152/jn.00075.2011.—Recent work has implicated low-frequency (<20 Hz) neuronal phase information as important for both auditory (<10 Hz) and speech [theta (~4–8 Hz)] perception. Activity on the timescale of theta corresponds linguistically to the average length of a syllable, suggesting that information within this range has consequences for segmentation of meaningful units of speech. Longer timescales that correspond to lower frequencies [delta (1–3 Hz)] also reflect important linguistic features—prosodic/suprasegmental—but it is unknown whether the patterns of activity in this range are similar to theta. We investigate low-frequency activity with magnetoencephalography (MEG) and mutual information (MI), an analysis that has not yet been applied to noninvasive electrophysiological recordings. We find that during speech perception each frequency subband examined [delta (1–3 Hz), theta<sub>low</sub> (3–5 Hz), theta<sub>high</sub> (5–7 Hz)] processes independent information from the speech stream. This contrasts with hypotheses that either delta and theta reflect their corresponding linguistic levels of analysis or each band is part of a single holistic onset response that tracks global acoustic transitions in the speech stream. Single-trial template-based classifier results further validate this finding: information from each subband can be used to classify individual sentences, and classifier results that utilize the combination of frequency bands provide better results than single bands alone. Our results suggest that during speech perception low-frequency phase of the MEG signal corresponds to neither abstract linguistic units nor holistic evoked potentials but rather tracks different aspects of the input signal. This study also validates a new method of analysis for noninvasive electrophysiological recordings that can be used to formally characterize information content of neural responses and interactions between these responses. Furthermore, it bridges results from different levels of neurophysiological study: small-scale multiunit recordings and local field potentials and macroscopic magneto/electrophysiological noninvasive recordings.

oscillations; information theory; magnetoencephalography

RECENT EVIDENCE HAS SUGGESTED that low-frequency phase information plays an important role in auditory perception (Kaysers et al. 2009; Lakatos et al. 2005). Noninvasive studies using magnetoencephalography (MEG) have shown that for speech perception the peak of this response occurs within the high delta and theta bands (~3–8 Hz), which corresponds (acoustically) to the peak of the modulation spectrum and (linguistically) to the average length of a syllable (Greenberg 2006; Greenberg et al. 1996; Howard and Poeppel 2010; Luo and Poeppel 2007). A response component that has received less

attention in the electrophysiological speech perception literature is the delta band (1–3 Hz).

In terms of processing spoken language, information on these timescales corresponds to different aspects of the speech signal. The average length of the syllable is ~150–300 ms, which corresponds to ~3–7 Hz, the heart of the theta band (Greenberg et al. 1996; Poeppel 2003). Longer timescales (i.e., lower frequencies) correspond to other aspects of the linguistic structure of the signal, such as phrasal boundaries and suprasegmental prosodic information (Gandour et al. 2003; Rosen 1992). What remains unclear is whether or not, during speech perception, these aspects of the linguistic signal are processed separately, as reflected in the activity of the frequency bands that correspond to the relevant timescales (i.e., delta for phrasal boundaries/prosodic information and theta for syllabic information).

While speech information on these timescales is important for comprehension, it is unclear whether (and how) these low-frequency electrophysiological responses are tuned to different aspects of the incoming speech signal and, if so, whether they are processed independently as linguistic information or separately as acoustic information or if low-frequency information of the neural signal is simply tracking broadband sharp acoustic transitions in the speech stream (Howard and Poeppel 2010). Different interpretations are clearly possible.

It is also not well characterized how these elements interact early in the acoustic processing of the input. While much of modern linguistic theory would suggest that each of these components is processed separately, most of the models that posit distinct tiers for suprasegmental information and smaller phonological unit (e.g., syllabic) encoding are based on models of production (Levelt 1989; Dell 1986) and, consequently, the nature of the early perceptual encoding of these elements is not clear.

To answer these questions, a measure that can assess the amount of information in a particular signal and determine whether or not there is overlap between two different signals is needed. While recent work using a cross-trial phase coherence and phase dissimilarity analysis has been successful for the former, it cannot be applied to the latter (Luo and Poeppel 2007).

Here we apply an information-theoretic approach to this problem. Mutual information (MI) analysis is based on Shannon's pivotal work on information theory (Shannon 1948), and it allows for both the assessment of information quantity within a signal and the characterization of the relationship between different neural signals. It has been applied successfully predominantly to multiunit recordings [multiunit activity (MUA)]

Address for reprint requests and other correspondence: G. B. Cogan, Dept. of Psychology & Center for Neural Science, New York Univ., 6 Washington Pl., New York NY 10003 (e-mail: gcogan@nyu.edu).

and local field potentials (LFP) with nonhuman primates (Kays et al. 2009; Montemurro et al. 2008), but its use in noninvasive electrophysiological recordings has so far been limited (but see Magri et al. 2009). Adapting MI methods to noninvasive techniques on human subjects would therefore 1) allow for a strong linkage between human and more low-level invasive analysis techniques on animals using a common analysis technique and 2) facilitate the study of the information capacity of the macroscopic electrophysiological signals that constitute MEG [and EEG and electrocorticographic (ECoG)] recording.

In the present study, participants listened to auditory sentences while undergoing neuromagnetic recording. The phase attributes of the low-frequency MEG signal were analyzed. The hypotheses under consideration were as follows. 1) The peak MI value should be within the theta band ( $\theta_{\text{low}}$ : 3–5 Hz and  $\theta_{\text{high}}$ : 5–7 Hz). If there is information within the delta band, then there should also be high MI values for 1–3 Hz. 2) If this low-frequency information is parsed in a way that is reflective of the linguistic structure of the input, then information in the delta band (corresponding to phrasal boundaries/suprasegmental prosodic information) should be independent of information in the theta band, which by hypothesis aligns most closely with syllabic information. Conversely, information in each of the two theta bands ( $\theta_{\text{low}}$  and  $\theta_{\text{high}}$ ) should be heavily redundant, as they are processing the same aspect of the input signal. 3) If, however, activity in the low-frequency spectrum of the MEG signal corresponds to a purely acoustic processing stage of the input, then each of the three subbands examined should be independent, as they are simply tracking different temporal elements of the acoustic input signal independent of linguistic structure. 4) Finally, if the phase of the low-frequency portion of the MEG signal is simply the convolution of evoked responses to sharp acoustic transitions, then there should be high redundancy between all three bands, as each frequency subband is in fact part of the same multifrequency process—the evoked response.

## METHODS

### Subjects

Eleven native English-speaking subjects (5 men, 5 women; mean age 26.7 yr) with normal hearing and no history of neurological disorders provided informed consent according to the New York University University Committee on Activities Involving Human Subjects (NYU UCA/HS). All subjects were right-handed as assessed by the Edinburgh Inventory of Handedness (Oldfield 1971). Two subjects' data were not included in the analysis because of poor signal-to-noise ratio (SNR) as assessed by an independent auditory localizer in one case and a script malfunction in the other, leaving nine subjects for further analysis.

### Stimuli

Three different English sentences were obtained from a public domain internet audio book website (<http://librivox.org>). Each of the sentences was between 11 and 12 s (sampling rate of 44.1 kHz), and each was spoken by a different speaker (American English pronunciation, 1 woman). The sentences were delivered to the subjects' ears with a tubephone (E-A-RTONE 3A 50 ohm, Etymotic Research) attached to E-A-RLINK foam plugs inserted into the ear canal and presented at normal conversational sound levels (~72 dB SPL). Four

other tokens of each sentence were created in which a 1,000-Hz tone was inserted at a random time point in the second half of each sentence. The tone was 500 ms in length with 100-ms cosine on- and off-ramps and an amplitude equal to the average amplitude of the sentences. Each sentence was presented 32 times, and each "tone sentence" was presented once for a total of 108 trials (32 trials  $\times$  3 sentences + 4 tone sentences  $\times$  3 sentences = 108 trials) within 4 separate blocks. The order of sentences was randomized within each block, with a randomized interstimulus interval (ISI) between 800 and 1,200 ms.

### Task

Participants were instructed to listen to the sentences with their eyes closed. This was done to limit artifacts due to overt eye movements and blinks. The task was to press a response key as soon as they heard a tone (in the target tone sentences). This was a distracter task designed to keep subjects attentive and alert, and, as such, tone sentence trials were not analyzed.

After the sentence experiment, each participant's auditory response was characterized by a functional localizer: subjects listened to 100 repetitions each of a 1-kHz and a 250-Hz 400-ms sinusoidal tone, with a 10-ms cosine on- and off-ramp and an ISI that was randomized between 900 and 1,000 ms. This was done to assess the strength and characteristics of the auditory response for each subject, to facilitate identification of auditory-sensitive channels, and to confirm that subjects' heads were properly positioned.

### MEG Recordings

MEG data were collected on a 157-channel whole-head MEG system (5-cm baseline axial gradiometer SQUID-based sensors, KIT, Kanazawa, Japan) in an actively magnetically shielded room. Data were acquired with a sampling rate of 1,000 Hz, a notch filter at 60 Hz (to remove line noise), a 500-Hz online analog low-pass filter, and no high-pass filter. Each subject's head position was assessed via five coils attached to anatomic landmarks both before and after the experiment to ensure that head movement was minimal. Head shape data were digitized with a three-dimensional digitizer (Polhemus). The data were noise reduced off-line with the continuously adjusted least-squares method (CALM; Adachi et al. 2002).

### Data Analysis

**Signal processing.** All data processing was done with MATLAB (MathWorks, Natick, MA). Figure 1 provides a flowchart illustrating various steps in the analysis. For each subject, data were split into sentences, trials, and channels. The data were band-passed in frequency ranges of interest ( $\delta$ : 1–3 Hz,  $\theta_{\text{low}}$ : 3–5 Hz, and  $\theta_{\text{high}}$ : 5–7 Hz) with an 814-point two-way least-squares linear FIR filter, shifted backwards to compensate for phase delays due to the original filtering. The filters were designed to minimize spectral leakage and overlap in frequency (2-Hz bandwidth), which in the present study is particularly important.

After filtering, the signal was then decimated by a factor of 4 (1,000 Hz to 250 Hz). This had no effect on the overall results and was done strictly for computational speed purposes. The first 11 s of each sentence were analyzed so that after down-sampling there were 2,750 data points for each trial within a given subject, channel, and sentence (Fig. 1B).

The instantaneous phase information was then extracted from the Hilbert transform (computed in MATLAB) of the decimated signal:

$$H(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(\tau)}{t - \tau} d\tau \quad (1)$$

$$\theta(t) = \arctan \left[ \frac{H(t)}{x(t)} \right] \quad (2)$$

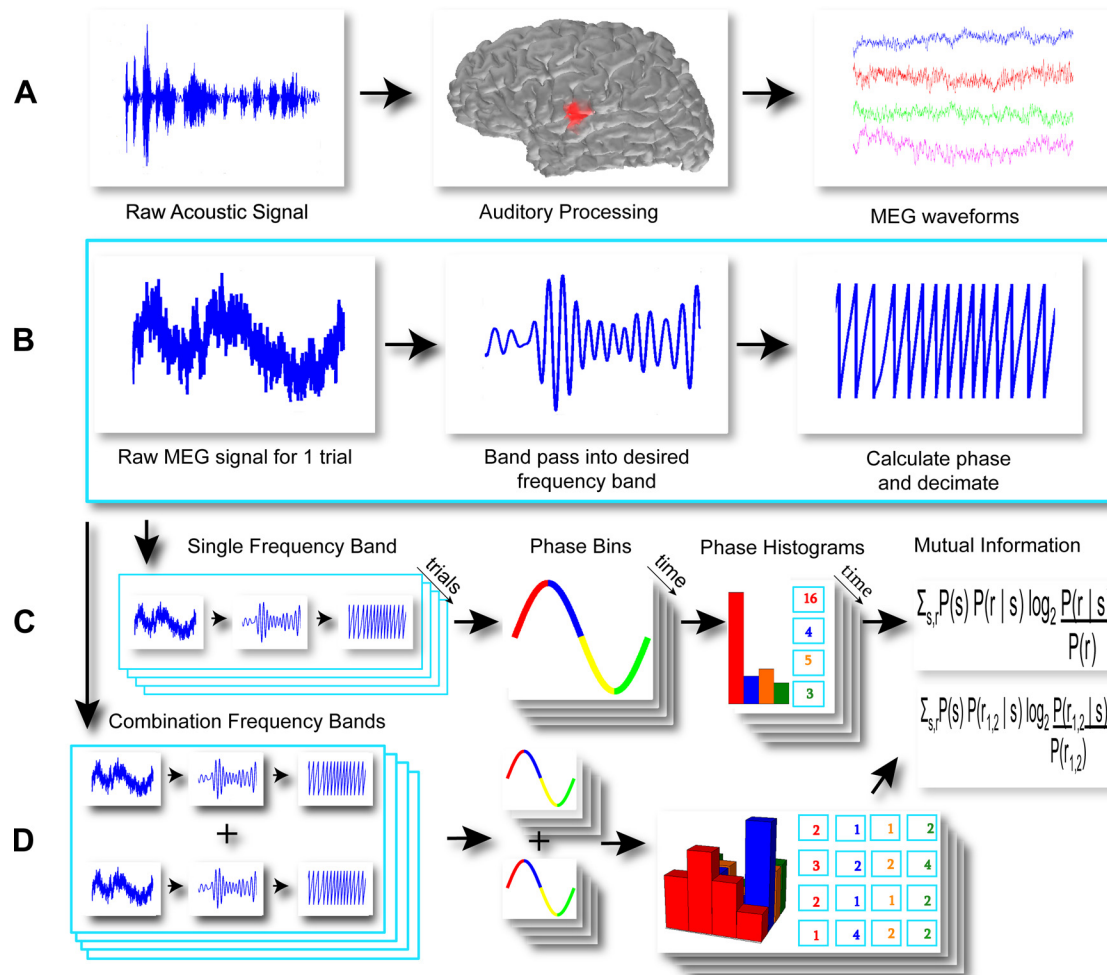


Fig. 1. Outline of preprocessing and mutual information (MI) analysis. **A**: subjects listened to sentences, here shown as acoustic waveforms. “Early” (acoustic, presemantic) processing of this information occurs in the auditory cortex, which can be measured effectively with magnetoencephalography (MEG). **B**: each signal from each trial, sentence, and channel is band-passed into the frequency of interest and decimated, and the phase is extracted from the Hilbert transform. **C**: for each frequency subband, across trials, each phase response for each time bin is grouped into 4 equally spaced bins. These values are then used to compute the MI values. **D**: for the combination frequency bands, the 4 bins from each single-frequency band response space are combined to form a 16-bin histogram, across trials for each time point, that is then used to assess the amount of information present in the combined frequency cases.

**Mutual information.** All further analyses were done with the Information Breakdown Toolbox in MATLAB (Magri et al. 2009; Pola et al. 2003). MI between the response and stimulus was calculated with the following equation:

$$I(S;R) = \sum_{r,s} P(s)P(r|s) \frac{P(r|s)}{P(r)} \quad (3)$$

where  $P(s)$  is the probability of observing a stimulus,  $P(r|s)$  is the probability of observing a response given a stimulus, and  $P(r)$  is the probability of observing a response across all stimuli and trials. The mutual information quantity  $I(S;R)$  between the stimulus and response can be thought of as the average amount of information that a single response provides about the stimulus. It can also be thought of as the reduction in entropy of the response space that the conditional probability of the response on the stimulus provides:

$$I(S;R) = H(R) - H(R|S) \quad (4)$$

$$H(R) = - \sum_r P(r) \log_2 P(r) \quad (5)$$

$$H(R|S) = - \sum_s P(s) \sum_r P(r|s) \log_2 P(r|s) \quad (6)$$

In the present study, the stimulus,  $s$ , is simply the value at each time point of the presented stimulus (i.e., the stimulus value at each down-sampled time point of each sentence so that the probability of

each stimulus is always 1/2,750). The MI analysis therefore makes no assumptions about the content of the signal itself, merely that it potentially changes as a function of time.

For the single-frequency case (Fig. 1C), the response distribution was composed of phase responses that fit into four equally spaced bins:  $-\pi$  to  $-\pi/2$ ,  $-\pi/2$  to 0, 0 to  $\pi/2$ , and  $\pi/2$  to  $\pi$ . For the case of the frequency combinations (Fig. 1D), the response distributions for each frequency were multiplied together to create a 16-bin distribution. Four bins were chosen for two reasons. The first reason is that this is the minimum number of bins that adequately reflects the overall phase response, and the second reason is due to pragmatic constraints: since the frequency combination case produces a number of bins that is equal to the square of the initial number of bins, an increase in the number of initial bins would lead to an exponential increase in the number of bins in the frequency combination case. Since there can only be a finite amount of trials, a greater number of initial bins would lead to a large number of instances in which there were zero values in a particular bin, which would skew the results.

The MI value was calculated for each subject, for each sentence, and for each channel across trials for each frequency band individually (delta, theta<sub>low</sub>, and theta<sub>high</sub>) and also for each combination (delta + theta<sub>low</sub>, delta + theta<sub>high</sub>, and theta<sub>low</sub> + theta<sub>high</sub>). In this latter case, the combination values were computed with the 10



channels for each individual frequency band that showed the highest MI values.

**Bias correction.** Since the estimation of MI is dependent on the sampling of the probability distributions, with results approaching their true value as more data are sampled, a multistep bias correction method was utilized (Kayser et al. 2009; Montemurro et al. 2007, 2008; Panzeri et al. 2007). The first step involved shuffling the values within each probability distribution within each stimulus (i.e., time point) across trials but holding the marginal probabilities equivalent to the unshuffled data values. This was done to **rule out incorrect conclusions about the amount of MI due to within-trial noise correlations.**

The second step utilized was introduced by Strong et al. (1998). MI is computed on the total data set, a randomization of half the trials, and a randomization of a quarter of the trials. As illustrated in Fig. 2, a quadratic function is then fit to the data points and the actual MI is taken to be the zero-crossing value. This new value reflects the estimated MI for an infinite number of trials and greatly **reduces the finite sampling bias** (Montemurro et al. 2007; Panzeri et al. 2007).

Finally, a **bootstrapping procedure** was utilized to remove any residual bias. Twenty time-shuffled trials were created, and the MI was assessed for each of the iterations. The mean of these iterations was then subtracted from the MI value obtained. **These three methods of bias correction—shuffling in time, quadratic extrapolation, and bootstrapping—have previously been found to reduce bias significantly** (Kayser et al. 2009; Montemurro et al. 2007, 2008; Panzeri et al. 2007).

**Synergy and redundancy.** For the frequency combinations, **redundancy was defined as:**

$$I_{\text{redundancy}} = I_{\text{lin}} - I_{\text{tot}} \quad (7)$$

where  $I_{\text{lin}}$  is the linear sum of the MI values for each frequency band and  $I_{\text{tot}}$  is the MI value for the frequency combination. Conversely, **negative redundancy can be termed synergy and is present if the combination of two signals provides more information than the sum of its parts** (Schneidman et al. 2003). The synergistic term was further broken down into its constituent parts as per the formalism of the information breakdown method (Magri et al. 2009; Pola et al. 2003):

$$I_{\text{syn}} = I_{\text{sigism}} + I_{\text{indcorr}} + I_{\text{depcorr}} \quad (8)$$

$I_{\text{sigism}}$  is the amount of information lost due to correlations in the signal and is calculated by subtracting the linear entropy from the independent entropy:

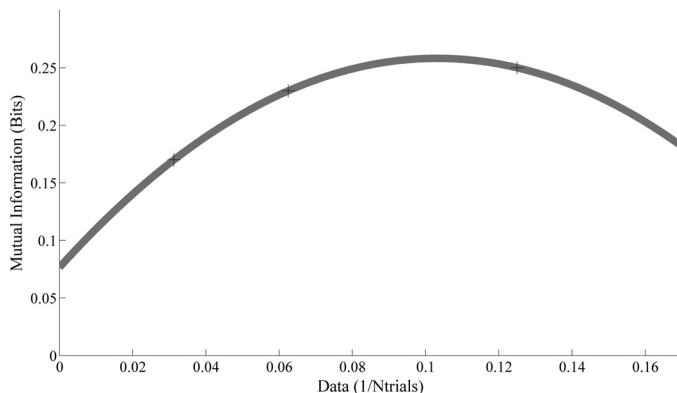


Fig. 2. Bias correction: quadratic extrapolation. Since the calculation of MI values depends on the sampling of the probability distribution, a larger number of trials will lead to a more accurate assessment of the information content. This method of bias reduction computes the MI values for the entire set of trials, a random set of half the trials, and a random set of a quarter of a trial. A quadratic function is fit to the data, and the zero-crossing is taken to be both the “true” MI value and what the MI value would be for an infinite number of trials.

$$I_{\text{sigism}} = H_{\text{ind}}(R) - H_{\text{lin}}(R) \quad (9)$$

where  $H_{\text{ind}}(R)$  replaces the summation of the marginal probabilities in Eq. 5 with the product of the marginal probabilities and  $H_{\text{lin}}(R)$  sums the probabilities from the single-response cases (for the individual frequency responses in the present study).  $I_{\text{indcorr}}$  represents the amount of noise correlation present that is independent of the stimuli. This term can be thought of as a measure of how similar two different neural responses are (in this case 2 different frequency bands) independent of which stimulus is presented:

$$I_{\text{indcorr}} = \chi(R) - H_{\text{ind}}(R) \quad (10)$$

where  $\chi(R)$  is the same as  $H_{\text{ind}}(R)$  except that the normalization term [i.e.,  $P(r)$ ] is kept nonindependent. Finally,  $I_{\text{depcorr}}$  represents the amount of information gained due to changes in noise correlations that are stimulus dependent. This term is therefore the most important for synergy, as both previous terms cannot contribute positively to synergy. This term was first introduced by Nirenberg et al. (2001) as  $\Delta I$ , **the amount of information lost to a downstream decoder if noise correlations are ignored.**

$$I_{\text{depcorr}} = [H(R) - H(R/S)] - [\chi(R) - H_{\text{ind}}(R/S)] \quad (11)$$

where  $H_{\text{ind}}(R/S)$  is calculated similarly to  $H_{\text{ind}}(R)$ , except that the product of the marginal probabilities is applied to Eq. 6 instead of Eq. 5.

**Classifier.** Classifier results were computed by comparing a random selected trial of each sentence (the template) with a random trial from the same sentence and a random trial from each of the other sentences. As in the case of the MI analysis, each trial was band-passed in the frequency region of interest and decimated, and the phase was extracted from the Hilbert transform. The phase value for each time point within each trial was binned with the same binning process as the MI analysis. Similarity between the template and the three comparison trials (1 from each sentence) was assessed by taking the inner dot product of the template and each comparison trial after the binning process. This created **values for each time bin that were either 1 for a match between the template and the comparison trial or a 0 for a nonmatch.**

**The average value across time points varied between 0 and 1 and was taken as the similarity between the template and the comparison.** The highest value of the three comparisons was taken to be the closest match. This was done **1,000 times for each sentence**, for each frequency and frequency combination [3 sentences  $\times$  (3 individual frequency bands + 3 combination bands) = 18 classifier results per subject]. In cases in which the template trial number matched the within-sentence comparison trial, another trial for the comparison sentence was chosen at random from the same sentence as the template. The process for the frequency combinations was the same as for the single-frequency version, except that instead of 4 phase bins 16 bins were utilized (as per the MI analysis). The classifier analysis was assessed with the same channels as in the MI analysis.

To assess the significance of the classifier results, a  $\chi^2$  analysis was done for each sentence and for each single frequency and each combination, for a total of 18  $\chi^2$  values for each subject:

$$\chi^2 = \sum_{i=1}^k \frac{(x_i - \mu)^2}{\sigma^2} \quad (12)$$

Since, however,  $\chi^2$  values are heavily dependent on the  $n$  value, significance via this measure of the effectiveness of the classifier is somewhat misleading. In other words, results could be made significant simply by choosing a large enough number of iterations of the classifier. Furthermore,  $\chi^2$  results are not a linear measure of the magnitude between variables and therefore cannot be summed or averaged.

With these issues in mind, **to assess whether or not the combination frequencies performed better in the classification than the single**

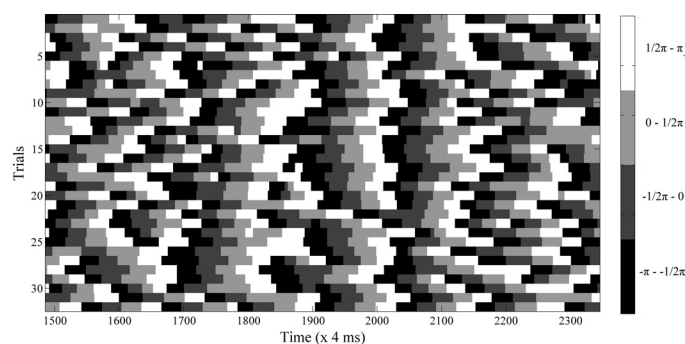


Fig. 3. Binned phase response for a representative subject. Phase response taken from 1 subject, 1 channel, and a portion of 1 sentence. Note that the time points on the  $x$ -axis have been decimated by a factor of 4 and therefore reflect units of 4 ms each.

frequencies, the effect size of each of the 18 values for each subject was computed as a  $\Phi$  value:

$$\Phi = \sqrt{\frac{\chi^2}{N}} \quad (13)$$

Since this converts the  $\chi^2$  values into effect sizes, the performance of a single-frequency band versus the combination bands can be assessed. The single-frequency band and combination frequency band values were then averaged for each sentence for each subject and then assessed for statistical significance via a paired two-sample  $t$ -test.

## RESULTS

### Mutual Information

The binning procedure produced phase values that occurred in one of four bins and were grouped across trials according to input sentence, channel, and frequency. An example of a portion of these values is shown in Fig. 3 for a single sentence, channel, and subject within the delta band. The structure of the data is quite apparent: time-locked phase responses are easily visible, although the noise inherent in noninvasive recordings is manifest in the nonperfect temporal alignment of these responses. This demonstrates that the ensuing MI calculations are in fact measuring a structured auditory response across trials as opposed to nonrelational noise.

The MI values within each frequency band when displayed by channel recapitulate the spatial distribution of a characteristic auditory response, as can be seen from a representative subject in Fig. 4. These responses were quite similar topographically to the responses known as the M100 or N1m, a response believed to originate from auditory regions on the superior temporal gyrus, near the transverse temporal gyrus (Liégeois-Chauvel et al. 1994; Pantev et al. 1990). This establishes that the response as assessed by MI for each of the three frequency bands analyzed originates from auditory regions in superior temporal cortex. Conversely, responses in higher-frequency bands did not elicit a reliable auditory response (data not shown).

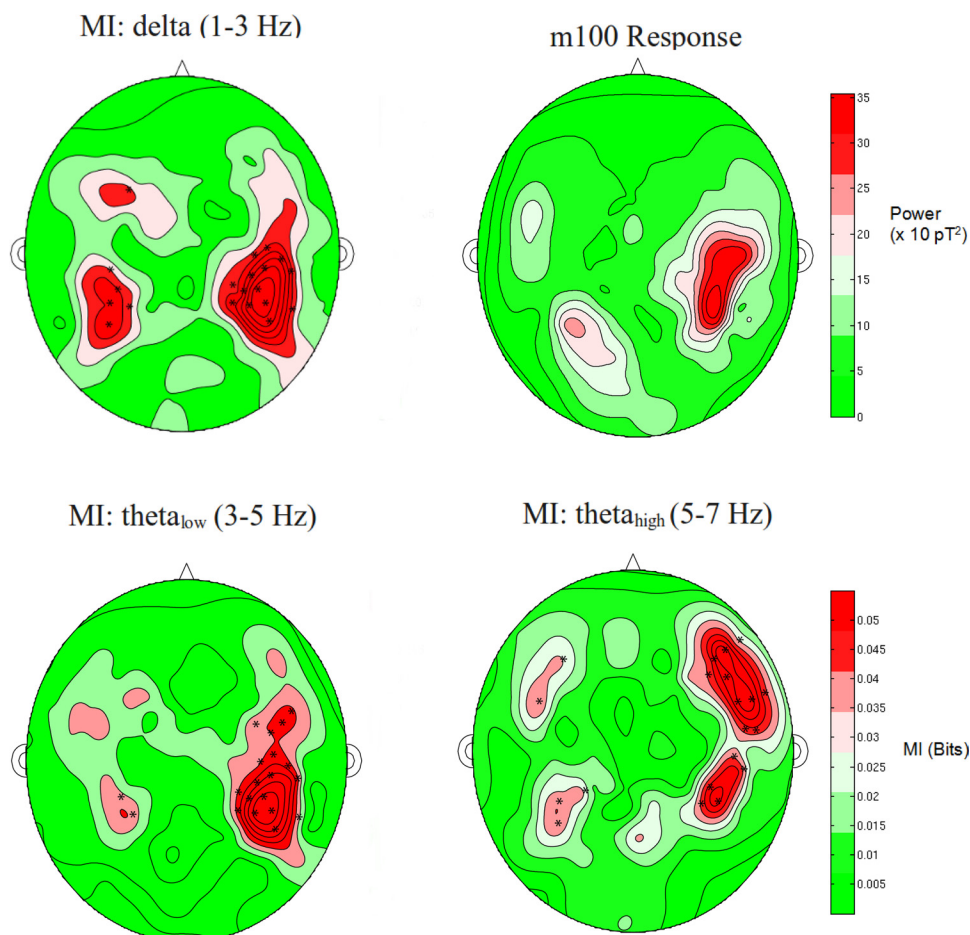


Fig. 4. Topographic head plots for a representative subject. MI is plotted for each frequency subband. Red denotes higher MI values and green lower values. As can be seen, the origin of the highest MI values compares favorably to the overall amplitude of the M100 response, consistent with an auditory origin for the channels yielding the MI values.

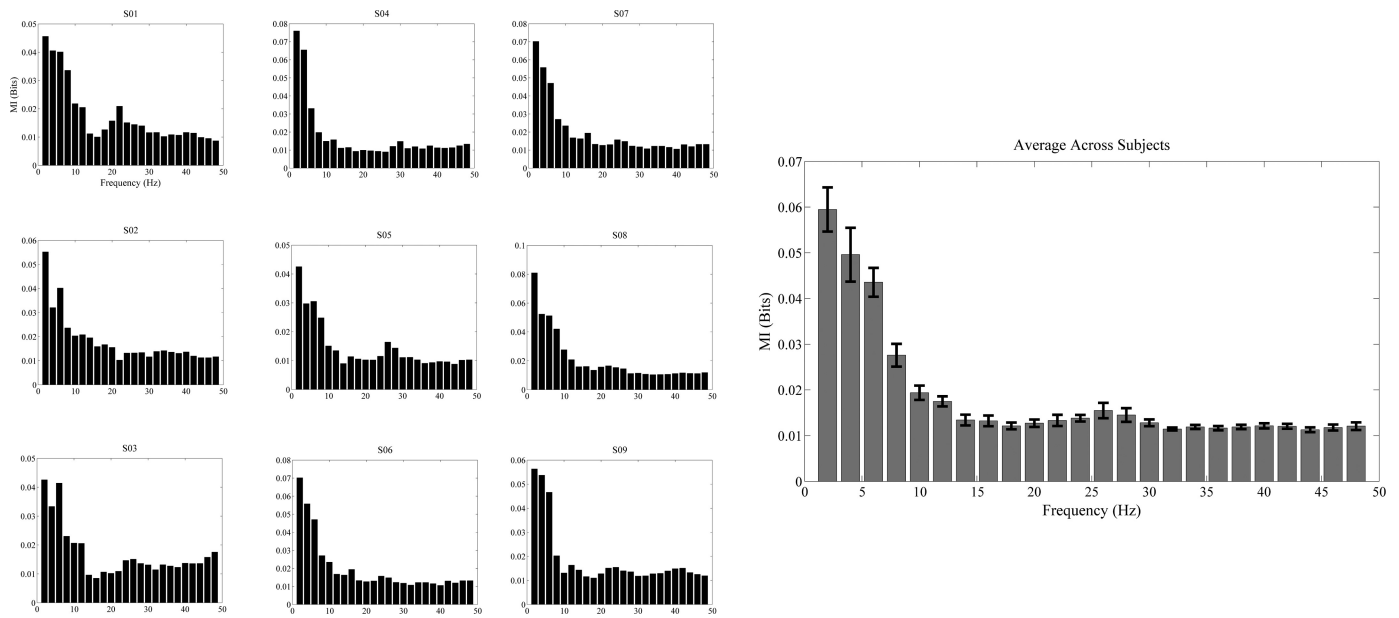


Fig. 5. MI values for each frequency band. The MI values are plotted here for each subject as a function of frequency. Each bar represents the average MI value across sentences for the top 10 channels. Plot on *right* is the average across subjects. MI values peaked in the low-frequency range (<8 Hz).

MI values for each frequency band are plotted in Fig. 5. Each value on the *x*-axis represents the center frequency of the filter utilized (see METHODS), and each bandwidth is 2 Hz. Results indicate that MI values are highest for delta, followed by  $\theta_{\text{low}}$  and then  $\theta_{\text{high}}$ . These results build upon and extend those of Luo and Poeppel (2007) (especially regarding the relevance of theta), Luo et al. (2010), which highlight the role (for audiovisual speech) of theta and delta, as well as Kayser et al. (2009). Luo and colleagues, using MEG, demonstrated a “privileged” role for the phase of theta band activity (4–6 Hz) during speech perception and delta and theta for the analysis of naturalistic movies, whereas Kayser et al. (2009) showed, in neurophysiological recordings, that the entire low-frequency range (<10 Hz) within auditory areas of macaque was particularly salient during the presentation of naturalistic movie scenes.

The results are robust for single subjects, with only minor variation present in the overall pattern. There is a small “bump” present in high beta/low gamma (~22–28 Hz) for some subjects, which is also present in the overall between-subject plot. Unfortunately, no meaningful conclusions can be drawn from these results as the topography did not show an auditory response or, for that matter, any coherent pattern at all (results not shown).

Figure 6 summarizes the comparison between the overall MI values predicted from the linear combination of the two composite individual frequency bands and the actual measured values obtained from the multiple-frequency band procedure (see METHODS). In all three combinations—delta +  $\theta_{\text{low}}$ , delta +  $\theta_{\text{high}}$ , and  $\theta_{\text{low}}$  +  $\theta_{\text{high}}$ —the combined MI values were higher than any of the individual frequency bands. Furthermore, each combination provided only slightly more than the linear sum of its individual subcomponents. This suggests that all three analyzed bands are processing independent aspects of the input signal, as the information present in the combined frequency band cases was not only equal to the

amount of information present in the individual band analyses but surpassed it, albeit by a small amount.

With the information breakdown method (Magri et al. 2009; Pola et al. 2003), the amount of total information provided by the frequency combinations was further examined. The total amount of information can be broken down into its constituent parts:

$$I_{\text{tot}} = I_{\text{lin}} + I_{\text{sigsim}} + I_{\text{indcorr}} + I_{\text{depcorr}} \quad (14)$$

This technique can be utilized to determine whether the linear combinations are in fact linear, as opposed to a combination of canceling opposing contributions from stimulus-independent and -dependent noise correlations (Magri et al. 2009; Pola et al. 2003). In other words, it could be the case that while the total amount of information present in the combination frequency bands is close to the value obtained from the linear summation of the two individual bands, this could be caused by a large increase in information due to stimulus-dependent noise correlations and an equally large reduction in information due to a response bias as reflective in the stimulus-independent noise

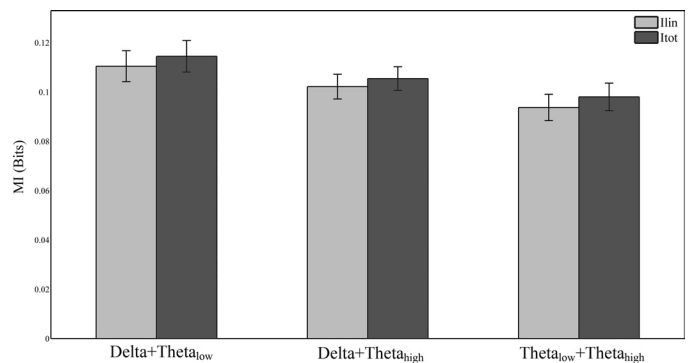


Fig. 6. Average MI values for linear summations and combinations. The average MI values for the linear summation of each frequency subband ( $I_{\text{lin}}$ ) and the combination of these subbands ( $I_{\text{tot}}$ ) are plotted. The combination values are quite similar to the linear summation values, suggesting that each subband is in fact processing independent information.

correlations. This would therefore undermine any interpretation of “true” independence.

Results show that the amount of information lost because of stimulus-independent noise correlations was extremely small, accounting for a loss of  $<1\%$  of the total amount of information predicted by the linear summation of the individual frequency bands. In fact, for two of the three combinations,  $\Delta + \theta_{\text{low}}$  and  $\Delta + \theta_{\text{high}}$ , this value was effectively zero, with only  $1.7 \times 10^{-6}$  and  $4.2 \times 10^{-6}$  bits lost for each combination, respectively.

Conversely, stimulus-dependent noise correlations led to an increase in information compared with the amount predicted by a linear summation, but again this value was quite low with all three combinations, presenting values that were less than a 5% gain referenced to the predicted linear values. These values are within the range of a previous study that suggested that retinal ganglion cells provide independent information (Nirenberg et al. 2001).

In all three combinations, the amount of information lost because of signal similarity was also extremely small ( $1.3 \times 10^{-6}$  to  $5 \times 10^{-5}$  bits), suggesting that the similarity of the input signal played a negligible role for the combined MI

values. Together, this suggests that information in the low-frequency phase response of MEG contains independent information about the input speech signal. Not only did the information present in the combined frequency band analyses contain slightly more than the information values predicted by the linear summation of the individual frequency bands, a near-perfect independent information content from each constituent frequency band, the amount of information present due to both stimulus-independent noise correlations and stimulus-dependent noise correlations was quite small relative to the total amount of information present. These latter results suggest that the measured information values do not reflect opposite canceling sources of noise correlation but rather “true” information independence.

### Classifier Results

The classifier results, shown for a representative subject in Fig. 7, demonstrate for the individual frequency bands that all three bands produced robust single trial-based template classifications. This is consistent with Luo and Poeppel (2007), who found that intertrial coherence for phase within the theta band

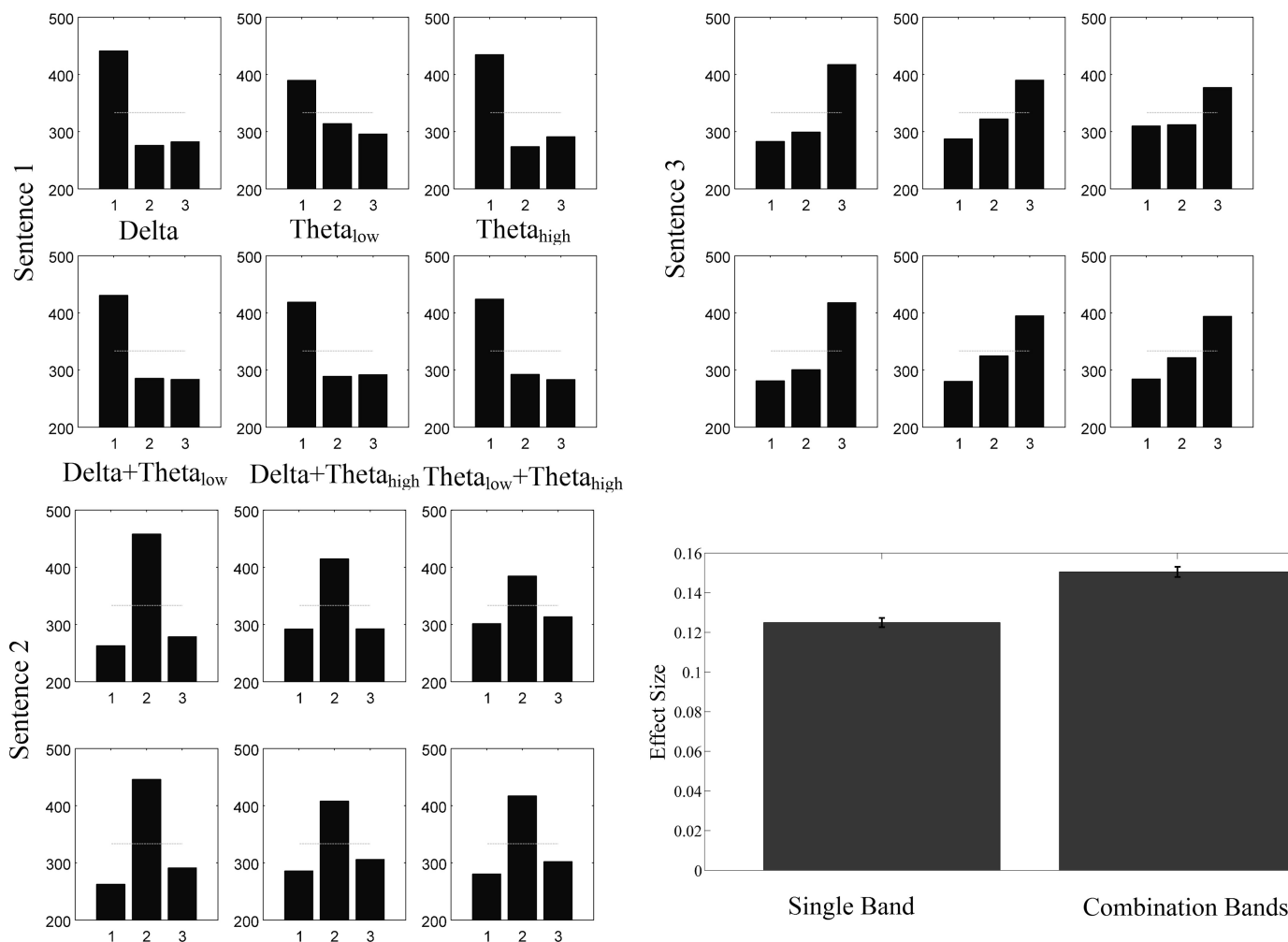


Fig. 7. Classifier performance for a representative subject. Classifier data are shown for each frequency band and combination band for each sentence for a representative subject. y-Axis represents absolute numbers of correct identifications out of 1,000 iterations (see text). Dashed line represents chance performance. As shown, each frequency band and combination successfully classified the template corresponding to a single trial of a given sentence with other tokens of that sentence. The effect size for the combination bands was significantly higher than that for the individual bands, suggesting that more information is available to the classifier in the combination cases.



was sufficient for this type of classification. The present study expands on these results by demonstrating both 1) that this result is obtainable with a different measure of information (MI vs. a phase dissimilarity function) and 2) that information in the delta band also provides robust information that can lead to single trial-based template classification (cf. Luo et al. 2010).

The combination band classifier results were also robust, indicating that the information from the combination of two different frequency bands can also be used to classify categorical membership of individual sentence tokens based on a single trial template. Furthermore, the average  $\Phi$  value (effect size) was larger for the combination frequency bands than it was for the individual frequency bands: 0.15 and 0.12, respectively (see Fig. 6). This difference was found to be significant by a paired  $t$ -test:  $t(26) = -4.64$ ,  $P < 5 \times 10^{-5}$ .

It is worth noting that the information utilized for the classification analysis is not entirely homologous to the information assessed in the MI analysis. Since MI is based upon the distribution of responses for a given stimulus across trials, it is impossible to utilize this distribution for single-trial classification. Nonetheless, this result is significant given that it validates the discrete binning process as an appropriate division of phase information and it also supports the notion that each frequency band investigated does in fact contain some independent information, as the classifier performance for the combination frequency bands was significantly higher than the single-frequency band cases.

## DISCUSSION

The results of this study demonstrate novel insights into the relationship between different low-frequency subbands of the phase of the macroscopic neural signal and offer a new approach to analyzing MEG data—MI analysis. The results also replicate and extend the findings of Luo and Poeppel (2007) and Luo et al. 2010 (see also Howard and Poeppel 2010). The MI results demonstrate that phase information in the theta band (here 3–7 Hz) contains strong relational information between the neural response and the input signal. The classifier results demonstrate not only that the theta response is consistent across time but also that it is discriminant in that it can be used on a single-trial basis to distinguish between different sentences. The MI results also demonstrate a particularly strong role for phase information in the delta band (here 1–3 Hz). This delta band phase information was also able to discriminate between different sentences based on a single trial template classifier analysis, suggesting that the elevated MI between the response and the stimulus within this band is not simply due to higher power (and consequently potentially higher SNR) but rather the information is meaningful. While earlier work (Howard and Poeppel 2010; Luo and Poeppel 2007) failed to show this effect within the delta band, this could be due to limitations based on their choice of frequency decomposition and the length of the stimuli utilized. In both previous studies, sliding windows of 500 ms were used while performing a moving window Fourier transform of the neural signal. This choice of window length would make it difficult to assess information within the low end of the delta range ( $<2$  Hz), as the frequency resolution would not be sufficiently accurate. Both previous studies also utilized stimuli that were  $\sim 4$  s in length, which would produce at most only 12 cycles (contrast

with 33 cycles for the present study) of the delta response. Results from Luo et al. (2010), using a cross-trial phase coherence analysis and stimuli that were  $\sim 30$  s in length, showed robust values within the delta band (2 Hz). While this result validates the latter concern, it unfortunately cannot address the first concern, as information below 2 Hz is still not present.

The small peak in the high beta/low gamma region ( $\sim 22$ – $27$  Hz) that appears in some subjects and slightly in the overall average (see Fig. 5) was unfortunately not sufficiently above noise to produce a characteristic topographic pattern. It is therefore unclear what this peak reflects. Future work using source-reconstruction methods [e.g., minimum norm estimate (MNE; Hämäläinen and Ilmoniemi 1994), linearly constrained minimum variance (LCMV) beamformers (Van Veen et al. 1997)] will perhaps be able to elucidate both the spatial distribution and the nature of this minor peak.

The overall goal of this study was to determine whether and how information in the low-frequency range is able to be dis-associated into separate frequency bands. The three relevant hypotheses were 1) that the information could correspond to different linguistic aspects of the speech signal (delta: prosody/suprasegmental information and theta: syllabic information) and show independent information for delta and theta, but not between the two theta bands ( $\theta_{\text{low}}$  and  $\theta_{\text{high}}$ ); 2) that the low-frequency range was simply tracking sharp broadband acoustic transients in the signal holistically, and therefore all three bands would be redundant (Howard and Poeppel 2010); or 3) that each band was in fact tracking different elements of the acoustic signal, and therefore each band would demonstrate independent tracking of information.

The present results are most consistent with the third hypothesis: all three bands (delta,  $\theta_{\text{low}}$ , and  $\theta_{\text{high}}$ ) showed a linear summation of information. This information independence was not due to a cancellation of opposing sources of noise correlation (as both signal-independent and signal-dependent noise correlations contributed  $<5\%$  of the linear summated information). This suggests that any shared “meaningful” noise source common to all three responses is at best minimal, further supporting the notion of independence. It is worth noting that this result does not necessarily mean that there are no shared noise components, but merely that if there are larger shared noise components they are minimally informative. Furthermore, the information present in the frequency band combinations outperformed the single-frequency bands in a single trial template-based classifier. This adds further support to the notion that the information within each of these bands is independent.

The presence of information independence does not, however, directly demonstrate that each of these bands corresponds to separate linguistic features. While it could be the case that each of the three bands examined corresponds to different linguistic features, it could also be the case that each band is simply tracking different frequency-specific acoustic features of the stimulus. Future work will have to be done to elucidate the nature of the specific stimulus features that drive these responses.

We reemphasize that the classifier utilized in the present study used aspects of the data that, while similar, were not in fact homologous to the information in the MI analysis. MI involves computations based on the entire response space,

whereas the classifier compares single trials to other single trial templates and therefore relies on single data points to produce a classification result. An MI classifier could be computed using the entire sentence as a probability distribution (similar to how MI is generally computed, relating one signal to another as opposed to a response to a signal, e.g., Jeong et al. 2001); however, this would remove the key component of the entire analysis, the specific relationship between the response and the stimulus. In this case, the information gained by examining the data at each time point would be lost. Nonetheless, the classifier results do lend credence to the binning process as an appropriate division of the phase responses as well as demonstrating that each frequency band being investigated does in fact contain complementary information.

### *Relationship to Decoders*

A common theme to early work using MI in single and multiunit recording has been the notion of an ideal decoder (Schneidman et al. 2003; Strong et al. 1998). In this framework, correlations between single units are seen as a source of ambiguity on a downstream decoder, as it would be unclear what portion of the signal was due to independent information computed from different single units in a population and what portion was due to signal correlations between different neurons. The quantity  $\Delta I$ , proposed by Nirenberg et al. (2001), characterized the amount of information lost because of these noise correlations from the perspective of a downstream decoder.

In the present study, this quantity is computed as stimulus-dependent noise correlations. It is worth pointing to two aspects of the results of this quantity: 1) this value was only a very small fraction of the overall linear summation of information between responses (<5%) and 2) this quantity was positive, denoting the fact that noise correlations that were dependent on the signal actually added information to the response. This type of result has led some to suggest that these correlations could act as a third channel of information (Dan et al. 1998; Nirenberg and Latham 2003). While it is an intriguing possibility, the results of this study do not support this hypothesis at the macroscopic level: while stimulus-dependent noise correlations did add to the information present, the amount relative to the linear summation of information was negligible. Furthermore, while the classifier results demonstrated an increase in effect size for the combination frequency results compared with the single-band cases, this increase was modest.

It is important to note that noninvasive techniques, while offering the advantages of having resolution that extends to the whole head and being applicable to general human populations, are also inherently noisier than more invasive methodologies such as MUA and LFP. This would explain the overall low values of information relative to previous work using this technique (Kayser et al. 2009; Montemurro et al. 2007; Nirenberg et al. 2001; Strong et al. 1998) that obtained values that were at least double those obtained in this study. While previous work using MI has predominantly been applied to single-unit, multiunit, or LFP data, the present investigation examined MEG data.

Another possible concern regards the filtering processes: in all three frequency bands examined, there was overlap between the frequency ranges being analyzed. Unfortunately, obtaining

reliable results by using narrower frequency ranges or with sharper edges would have involved filter orders larger than the data sets being analyzed. Also, using different FIR filters with different (slightly larger) filter orders did not result in qualitatively different results (data not shown). Furthermore, any shared spectral information due to the overlap in the frequencies being filtered would result in a shift toward redundancy, as shared portions of the MEG signal would now be present in two separate frequency bands. Given that the results obtained here demonstrate that each band is in fact independent, it is unlikely that this overlap contributed to the results.

The larger implications of these results are twofold. The first is that **the low-frequency content of the MEG signal contains independent information about the acoustic signal**. This division cannot be attributed directly to linguistic units of representation per se, as information within the two theta bands ( $\theta_{\text{low}}$  and  $\theta_{\text{high}}$ ) demonstrated a linear summation of information, as opposed to redundancy. This suggests that **each theta subband analyzed here is in fact tracking different elements of the input speech signal**. The results do not support a model in which each band is in fact responding to sharp broadband acoustic transitions, as this would lead to redundancy between all three bands. This does not mean, however, that any of the three subbands are not responding in this manner, merely that all three cannot be tracking this type of information (Howard and Poeppel 2010).

Second, the present study validates a unique methodological approach to noninvasive electrophysiological recordings. While previous work using MI has focused predominantly on lower-level invasive animal recordings (Kayser et al. 2009; Montemurro et al. 2008), the results of this study suggest that it can also be applied to noninvasive electrophysiological human data and lead to meaningful results. The strength of this approach is that it characterizes the nonlinear relationship between a response and a stimulus, and it can also be used to qualify and quantify the relationship between different neural responses. Further work will be needed to produce an appropriate model of the MEG signal(s) being analyzed that leads to these types of results. It is a particularly challenging endeavor as it requires accurately modeling both the specific characterizations of the various noise sources (e.g., external, internal noise sources) and the nonstationary elements of the overall MEG signal.

Determining the specific correspondence between the acoustic signal and the neural signal as reflected in the phase responses of the individual frequency bands is important for future research. It is not clear, for instance, whether temporal periods of high MI values reflect portions of the input signal that preferentially drive the response or simply portions of the MEG signal that are less contaminated by noise. Put differently, it is not clear whether or not the neural phase response measured in this study is stationary. This clarification would shed light on whether or not there are specific portions of the input auditory signal that are particularly salient for this particular response or whether the neural phase response as measured by MEG is only tracking a portion of the "true" signal.

It is also unclear whether the relevant frequency bands being investigated reflect tracking of components of the input signal that occur at that timescale (Luo and Poeppel 2007), time constants associated with the neural response itself, or some

combination of the two (Howard and Poeppel 2010). While it is intuitive to think that a neural response at a particular frequency band reflects tracking of an input component at the same corresponding frequency [e.g., auditory steady-state response (aSSR) (Picton et al. 2003)], evoked responses, for instance, occur on characteristic timescales that are thought to reflect time constants of the neural processing (Howard and Poeppel 2010) associated with the input rather than the specific temporal qualities of the input itself.

The present study employed three bands of frequency decomposition, but this does not necessarily mean that these reflect specific “privileged” divisions of the neural signal. Rather they were chosen as a compromise between the hypothesis-driven investigation and frequency decomposition limitations. It could be that there are in fact no privileged frequency bands within the low-frequency range of the neural signal. This would suggest that rather than tracking specific events that occur on particular timescales (e.g., syllables, prosodic information), neural mechanisms track all aspects of the input signal below ~10 Hz. This hypothesis would be more in line with low-level studies that find a broad peak of activation within the low-frequency range rather than specific peaks corresponding to components of the input signal (Kayser et al. 2009).

#### ACKNOWLEDGMENTS

We thank Jonathan Simon and William J. Idsardi for their analytical support and Jeff Walker for his excellent technical assistance.

#### GRANTS

This work was supported by National Institute of Deafness and Other Communications Disorders Grant 2R01DC-05660 to D. Poeppel.

#### DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

#### REFERENCES

- Adachi Y, Shimogawara M, Higuchi M, Haruta Y, Ochiai M. Reduction of non-periodic environmental magnetic noise in MEG measurement by continuously adjusted least squared method. *IEEE Trans Appl Superconduct* 11: 668–672, 2002.
- Dan Y, Alonso JM, Usrey WM, Reid RC. Coding of visual information by precisely correlated spikes in the lateral geniculate nucleus. *Nat Neurosci* 1: 501–506, 1998.
- Dell G. A spreading-activation theory of retrieval in sentence production. *Psychol Rev* 93: 283–321, 1986.
- Gandour J, Dziedzic M, Wong D, Lowe M, Tong Y, Hsieh L, Saththamnuwong N, Lurito J. Temporal integration of speech prosody is shaped by language experience: an fMRI study. *Brain Lang* 84: 318–336, 2003.
- Greenberg S. A multi-tier framework for understanding spoken language. In: *Listening to Speech: an Auditory Perspective*, edited by Greenberg S, Ainsworth WA. Mahwah, NJ: Erlbaum, 2006.
- Greenberg S, Hollenback J, Ellis D. Insights into spoken language gleaned from phonetic transcriptions of the Switchboard Corpus. *Proc Fourth Intl Conf on Spoken Language (ICSLP)*, Philadelphia, PA, 1996, S24–S27.
- Hämäläinen MS, Ilmoniemi RJ. Interpreting magnetic fields of the brain: minimum norm estimates. *Med Biol Eng Comput* 32: 35–42, 1994.
- Howard MF, Poeppel D. Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. *J Neurophysiol* 104: 2500–2511, 2010.
- Jeong J, Gore JC, Peterson BS. Mutual information analysis of the EEG in patients with Alzheimer's disease. *Clin Neurophysiol* 112: 827–835, 2001.
- Kayser C, Montemurro MA, Logothetis NK, Panzeri S. Spike-phase coding boosts and stabilizes information carried by spatial and temporal spike patterns. *Neuron* 61: 597–608, 2009.
- Lakatos P, Ankoor SS, Knuth KH, Ulbert I, Karmos G, Schroeder CE. An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *J Neurophysiol* 94: 1904–1911, 2005.
- Levelt WJM. *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press, 1989.
- Liégeois-Chauvel C, Musolino A, Badier JM, Marquis P, Chauvel P. Evoked potentials from the auditory cortex in man: evaluation and topography of the middle latency components. *Electroencephalogr Clin Neurophysiol* 92: 204–214, 1994.
- Luo H, Liu Z, Poeppel D. Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS Biol* 8: 1–13, 2010.
- Luo H, Poeppel D. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54: 1001–1010, 2007.
- Magri C, Whittingstall K, Singh V, Logothetis NK, Panzeri S. A toolbox for the fast information analysis of multiple-site LFP, EEG and spike train recordings. *BMC Neurosci* 10: 81, 2009.
- Montemurro MA, Rasch MJ, Murayama Y, Logothetis NK, Panzeri S. Phase-of-firing coding of natural visual stimuli in primary visual cortex. *Curr Biol* 18: 375–380, 2008.
- Montemurro MA, Senatore R, Panzeri S. Tight data-robust bounds to mutual information combining shuffling and model selection techniques. *Neural Comput* 19: 2913–2957, 2007.
- Nirenberg S, Carcieri SM, Jacobs AL, Latham PE. Retinal ganglion cells act largely as independent encoders. *Nature* 411: 698–701, 2001.
- Nirenberg S, Latham PE. Decoding neuronal spike trains: how important are correlations? *Proc Natl Acad Sci USA* 100: 7348–7353, 2003.
- Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9: 97–113, 1971.
- Pannenkamp A, Toepel U, Alter K, Hahne A, Friederici AD. Prosody-driven sentence processing: an event-related brain potential study. *J Cogn Neurosci* 17: 407–421, 2005.
- Pantev C, Hoke M, Lehnertz K, Lütkenhöner B, Fahrendorf G, Stöber U. Identification of sources of brain neuronal activity with high spatiotemporal resolution through combination of neuromagnetic source localization (NMSL) and magnetic resonance imaging (MRI). *Electroencephalogr Clin Neurophysiol* 75: 173–184, 1990.
- Panzeri S, Senatore R, Montemurro MA, Petersen RS. Correcting for the sampling bias problem in spike train information measures. *J Neurophysiol* 98: 1064–1072, 2007.
- Picton TW, John MS, Dimitrijevic A, Purcell D. Human auditory steady-state responses. *Int J Audiol* 42: 177–219, 2003.
- Poeppel D. The analysis of speech in different temporal integration windows: cerebral lateralization as “asymmetric sampling in time”. *Speech Commun* 41: 245–255, 2003.
- Pola G, Thiele A, Hoffman KP, Panzeri S. An exact method to quantify coding the information transmitted by the different mechanisms of correlational coding. *Network* 14: 35–60, 2003.
- Reite M, Adams M, Simon J, Teale P, Sheeder J, Richardson D, Grabbe R. Auditory M100 component 1: relationship to Heschl's gyri. *Brain Res Cogn Brain Res* 2: 13–20, 1994.
- Rosen S. Temporal information in speech: acoustic, auditory, and linguistic aspects. *Philos Trans R Soc Lond B Biol Sci* 336: 367–373, 1992.
- Schneidman E, Bialek W, Berry MJ 2nd. Synergy, redundancy, and independence in population codes. *J Neurosci* 23: 11539–11553, 2003.
- Shannon CE. A mathematical theory of communication. *Bell Syst Tech J* 27: 379–423, 623–656, 1948.
- Strong SP, Koberle R, de Ruyter van Steveninck RR, Bialek W. Entropy and information in neural spike trains. *Phys Rev Lett* 80: 197–200, 1998.
- Van Veen BD, Van Drongelen W, Yuchtman M, Suzuki A. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Trans Biomed Eng* 44: 867–880, 1997.