ABSTRACT

Title of Document: TEMPORAL DYNAMICS OF MEG PHASE INFORMATION DURING SPEECH PERCEPTION: SEGMENTATION AND NEURAL COMMUNICATION USING MUTUAL INFORMATION AND PHASE LOCKING

Gregory B. Cogan, Ph.D. 2011

Directed by: Prof. William Idsardi, Department of Linguistics

The incoming speech stream contains a rich amount of temporal information. In particular, information on slow time scales, the delta and theta band (125 – 1000 ms, 1 – 8 Hz), corresponds to prosodic and syllabic information while information on faster time scales (20-40 ms, 25 – 50 Hz) corresponds to feature/phonemic information. In order for speech perception to occur, this signal must be segregated into meaningful units of analysis and then processed in a distributed network of brain regions. Recent evidence suggests that low frequency phase information in the delta and theta bands of the Magnetoencephalography (MEG) signal plays an important role for tracking and segmenting the incoming signal into units of analysis. This thesis utilized a novel method of analysis, Mutual Information (MI) to characterize the relative information contributions of these low frequency phases. Reliable information pertaining to the

stimulus was present in both delta and theta bands (3 – 5 Hz, 5 – 7 Hz) and information within each of these three sub-bands was independent of each other. A second experiment demonstrated that the information present in these bands differed significantly for speech and a non-speech control condition, suggesting that contrary to previous results, a purely acoustic hypothesis of this segmentation is not supported. A third experiment found that both low (delta and theta) and high (gamma) frequency information is utilized to facilitate communication between brain areas thought to underlie speech perception. Distinct auditory/speech networks that operated exclusively using these frequencies were revealed, suggesting a privileged role for these timescales for neural communication between brain regions. Taken together these results suggest that timescales that correspond linguistically to important aspects of the speech stream also facilitate segmentation of the incoming signal and communication between brain areas that perform neural computation.

TEMPORAL DYNAMICS OF MEG PHASE INFORMATION DURING SPEECH PERCEPTION: SEGMENTATION AND NEURAL COMMUNICATION USING MUTUAL INFORMATION AND PHASE LOCKING

By

Gregory B. Cogan

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in the partial fulfillment
of the requirements for the degree of
Ph.D.
2011

Advisory Committee:
Professor William Idsardi, Chair
Professor Jonathan Z. Simon
Professor Catherine E. Carr
Professor Norbert Hornstein
Professor David Poeppel

Acknowledgements

I would first like to thank my advisors for their wonderful guidance over the years as well as their patience. I'd like to especially thank David for his encouragement, as without him none of this would be possible. I have been extremely fortunate to have had the opportunity to work with all of you. I'd also like to thank my parents for having faith in me, and my entire family for their support. Lastly, I'd like to thank Tom for his advice and helping to keep things in perspective.

Table of Contents

List of Tables

List of Figures

General Introduction

Speech perception requires the transformation of a continuous acoustic waveform into segmented units of analysis that can be processed throughout the cortex. This is a particularly difficult problem as the input speech stream does not come pre-segmented (VanRullen & Koch 2003), nor is it obvious once segmented how the coordination of computations in different brain areas occurs.

Work from the animal literature can offer insight into plausible mechanistic explanations for both the parsing and the temporal dynamics of the network that underlie speech perception specifically and sensory perception/cognition more generally. Neuronal oscillations is a strong candidate for processing of this type of information because of its emphasis on the temporal aspect of operation and its ubiquity in the mammalian nervous system (Buzsáki & Draguhn 2004). Neuronal oscillatory phenomena offer an energy-efficient mode of processing that can both temporally segment an incoming continuous signal as well as dynamically coordinate neuronal operations throughout the brain.

Models of both sensory-input processing and network coordination come from computational work (Shamir et al. 2009, Wang 2010) as well as the animal literature (Laurent 2002, Lakatos et al. 2005). Incoming signals are processed in preferred phases of the underlying neuronal ensemble dynamics. This coordination offers a mechanistic explanation that links the incoming sensory input to both the underlying processing and the coordination of network dynamics that sub serve these computations. It is therefore reasonable to assume that there is a strong relationship between how the temporal

dynamics of the parsing of the incoming sensory signal and the subsequent coordination of network dynamics that underlie the processing and computation of these signals.

One candidate macroscopic mechanism for the parsing of the input signal is that low frequency portions of the neural signal track the slow amplitude fluctuations of the speech stream (Luo & Poeppel 2007, Abrams et al. 2008). These fluctuations (the envelope) correspond acoustically the peak of the modulation transfer function of the speech signal and linguistically to the average length of the syllable, making it an ideal candidate for linking purely acoustic features of the input signal with higher-order cognitive representations (Greenberg 2006, Greenberg et al. 1996).

The nature of the low frequency components of the neural signal that are purported to be responsible for the segmentation of the input acoustic signal is unclear. Previous work using magnetoencephalography (MEG) initially proposed an envelope tracking mechanism (Luo & Poeppel 2007). Mechanistically, the hypothesis was that endogenous oscillations within the theta band (4 – 8 Hz) reset to the acoustic transitions of the onsets of the envelope. This is in keeping with both EEG work (Abrams et al. 2008) and evasive single-unit recordings in macaque (Lakatos et al. 2005). A later hypothesis suggested that this response was simply the convolution of the evoked N1-P2 complex, an auditory onset response that is responsive to sounds onsets in general rather than slow amplitude fluctuations specifically (Howard & Poeppel 2010). The genesis of this response was therefore not the phase reset of endogenous oscillations but rather canonical onset responses responding to acoustic transients in the input signal. The response peaked within the theta band because of the duration of the onset response itself, rather than the temporal aspects of the salient components in the signal. It is important to

note however, that these two aspects could work in tandem: if the acoustic landmarks of the input are in tune with the duration of the onset response, this would presumably lead to an 'ideal' relationship between the input and the neural response.

More recent work using audio-visual stimuli (Luo et al. 2010) has demonstrated that this response extends into the delta band (1 – 3 Hz). Linguistically, information on this timescale corresponds to supra-segmental features and prosodic information. It is not clear however, how this finding would relate to the second hypothesis about the genesis of this response, as the N1-P2 complex is more transient that the timescale that low end of the delta response (1 Hz) would account for.

Work in the animal literature may shed some light on the low-level mechanistic features of this signal. Elhilali and colleagues (2004) have proposed that the preference for tracking low frequency envelope features (< 20 Hz) in the cortex is due to synaptic depression between monosynaptic connections between the Medial Geniculate Body (MGB) of the thalamus and primary auditory cortex that enables the cortex (exclusively) to track slower, more salient features of the input – the envelope. In this way, they propose that the envelope acts as a gating mechanism for the analysis of the fine structure, which is analyzed in short phasic bursts that last for approximately 100 ms (Elhalali et al. 2004). This allows for both a segmentation of the input signal as well as a mechanistic component that ensures that tracking of the fine structure does not exceed its adaptation duration.

What remains unclear is how this low-level mechanism relates to macroscopic signals (i.e. EEG, MEG) and how signals that contain the envelope (gating mechanism) but not the 'content' (fine structure) are processed. It could be for instance, that low

frequency amplitude modulated signals in the absence of fine structure (e.g. noise) engage sub-cortical responses to the input but are fundamentally altered in ways that affect this cortical gating process, or it could be that these are in fact distinct processes that interact, but do not rely on each other mechanistically to operate.

A unique methodological approach to these questions is Mutual Information analysis (MI). MI is a useful analysis technique that has been applied successfully to low-level recordings in non-humans (Kayser et al. 2009, Montemurro et al. 2008), but its usage in non-invasive human recordings has been limited (Magri et al. 2009). Kayser et al. (2009) demonstrated that in the macaque auditory cortex, the combination of local field potential (LFP) components of the signal and spike trains combined to offer more information that either constituent part did, displaying a synergistic relationship between these two components. This relationship has also been demonstrated in the visual domain (Montemurro et al. 2008), suggesting a more general cortical mechanism is responsible, rather than a specific auditory one.

While previous work in MEG has used inter-trial phase coherence and a phase dissimilarity function to quantity the consistency of the phase response across trials and the specific nature of this response, it is unable to shed light on the relationship between the signals themselves. For instance, it cannot assess whether or not the response within the theta band is tracking the same aspects of the signal as the delta band. Characterizing the nature of these interactions is an important component for elucidating both the nature of these response as well as the components of the signal that they are tracking. Furthermore, while previous work (Howard & Poeppel 2010) has demonstrated similar phase dissimilarity results for speech and reversed speech, it is not clear if responses

generated in each of these conditions in qualitatively the same. It could be for instance, that while quantitatively, the response is similar but it is in fact tracking an entirely different component of the input signal.

Once segmented, the resultant units of analysis then must be distributed among a network of different brain areas that are responsible for specific computations performed on the segments themselves. While the speech/language system was one of the first cognitive networks to be outlined (Lichtheim 1886), there has been surprisingly little attention paid to the network dynamics of this system in general, and the temporal components and their relation to the segmentation of the speech signal in particular.

Recent work using fMRI has shown that specific brain regions demonstrate high levels of correlation in the low frequency range of the BOLD response (0.01 to 0.1 Hz) during rest (Fox et al. 2005, Buckner et al. 2008). These networks are thought to represent coupling between disparate brain areas that underlie specific cognitive functions (Bressler & Menon 2010). Electrophysiological work examining these networks has implicated the involvement of a broad range of frequencies (Mantini et al. 2007).

A recent model of speech perception (Hickok & Poeppel 2000, 2004,2007) posits that the incoming speech stream is processed in disparate brain areas that are responsible for different computations performed on the signal. The organization of this model is split into a left-lateralized dorsal stream that maps the signal onto articulatory representations and a bilateral ventral stream that performs more meaning-centric computations. There is also a temporal component to this model, with the left hemisphere preferring to analyze incoming information on a fast time scale (gamma 25 –

50 Hz) and the right hemisphere preferring a slower window of analysis (theta 4 – 8 Hz). What remains unclear, is how the coordination between different brain areas, and consequently, different computations takes place.

This thesis therefore aimed to investigate these two components of speech perception: the initial parsing of the input signal and the temporal dynamics of the coordinate of computations within the speech perception network. Results of the first experiment demonstrate that the input signal is first parsed by low frequency phase information that does not map neatly onto canonical frequency bands (e.g. delta, theta), but rather tracks independent components of the input signal. Results of the second experiment show that these responses exhibit a degree of speech specificity as well as a qualitatively different tracking component in speech and non speech. Lastly, results of the third experiment suggest that timescales that are salient for speech perception itself (delta, theta, gamma) are also privileged timescales of network communication: both lateralized and bilateral networks were shown that operate exclusively using these frequencies. Together, this suggests that the incoming speech signal is first parsed into salient units of analysis via independent low frequency neural responses that display a degree of speech specificity and once parsed, the coordination of neural computations takes place using the same salient timescales that are prevalent in the speech stream itself.

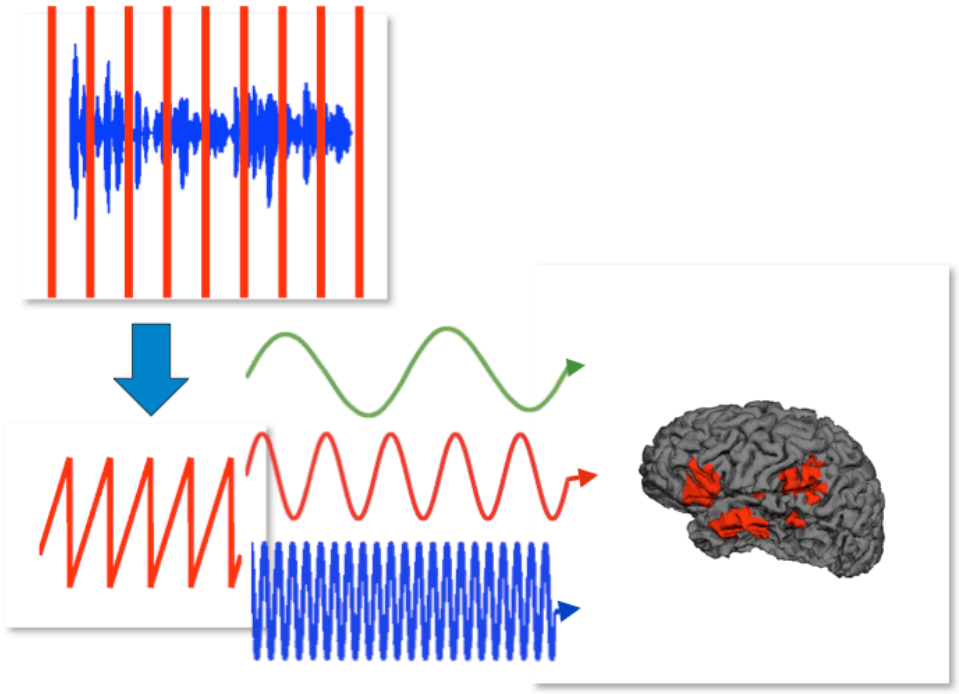

Fig 1: Parsing and network coordination model. The incoming speech stream is first parsed by the low frequency phase portion of the neural signal. Once parsed, the coordination of spatially distinct computations is carried out via phase locking in on three distinct time scales: 333-1000 ms (delta), 125-250 ms (theta), and 25-40 ms (gamma). Both the segmentation and the network dynamics of the system reflect the salient timescales of analysis of the speech stream itself.

Chapter 1: Mutual Information analysis of neural coding of speech by low frequency MEG phase information

1.1 INTRODUCTION

Recent evidence has suggested that low frequency phase information plays an important role in auditory perception (Kayser et al. 2009, Lakatos et al. 2005). Noninvasive studies using 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 (Luo & Poeppel 2007, Howard & Poeppel 2010, Greenberg 2006, Greenberg et al. 1996). 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 time scales corresponds to different aspects of the speech signal. The average length of the syllable is approximately 150-300 ms which corresponds to ~3-7 Hz, the heart of the theta band (Greenberg et al. 1996, Poeppel 2003). Longer time scales (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 time scales (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 if (and how) these low frequency electrophysiological responses are tuned to different aspects of the incoming speech signal and if so, if they are processed independently as linguistic information, 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 & 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 are 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.

In order 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 multi-unit recordings (MUA) and local field potentials (LFP) with non-human primates (Kayser et al. 2009, Strong et al. 1998, 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 (i) allow for a strong linkage between human and more low-level invasive analysis techniques on animals using a common assessment unit (the bit), and (ii) facilitate the study of the information capacity of the macroscopic electrophysiological signals that constitute MEG (and EEG and ECog) recording.

In the current 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: (i) The peak MI value should be within the theta band ($theta_{low}$/3-5 Hz, and $theta_{high}$ 5-7 Hz). If there is information within the delta band, then there should also be high MI values for 1-3 Hz. (ii) 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 ($theta_{low}$ and $theta_{high}$) should be heavily redundant, as they are processing the same aspect of the input signal. (iii) 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 sub-bands examined

should be independent, as they are simply tracking different temporal elements of the acoustic input signal independent of linguistic structure. (iv) Lastly, 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 sub-band is in fact part of the same multi-frequency process – the evoked response.

1.2 METHODS

Subjects

Eleven native English-speaking subjects (5 male, mean age 26.7) 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) and the University of Maryland institutional review board. All subjects were right-handed as assessed by the Edinburgh Inventory of Handedness (Oldfield 1971). Two subjects' data were not included in the analysis due to poor 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 seconds (sampling rate of 44.1 kHz) and each was spoken by a different speaker (American English pronunciation, 1 female). 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 sounds levels (~72 dB SPL). Four other tokens of each sentence were created in which a 1000 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 x 3 sentences + 4 tone sentences x 3 sentences = 108 trials) within 4 separate blocks. The order of sentences was randomized within each block, with a randomized inter-stimulus interval (ISI) between 800 and 1200 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 ms and 1000 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 1000 Hz, a notch filter at 60 Hz (to remove line noise), a 500 Hz on-line analog low pass filter, and no high-pass filter. Each subject's head position was assessed via five coils attached to anatomical landmarks both before and after the experiment to ensure that head movement was minimal. Headshape data were digitized using a three-dimensional digitizer (Polhemus). The data were noise reduced offline using the Continuously Adjusted Least-Squares Method (CALM -Adachi et al. 2002).

Data Analysis

Signal Processing

All data processing was done using MATLAB (MathWorks, Natick, MA). Figure 2 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, $theta_{low}$: 3-5 Hz, and $theta_{high}$: 5-7 Hz) using a 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, which in the current study is particularly important.

After filtering, the signal was then decimated by a factor of 4 (1000 Hz to 250 Hz). This had no effect on the overall results and was done strictly for computational speed purposes. The first 11 seconds of each sentence were analyzed so that after down-sampling there were 2750 data points for each trial within a given subject, channel, and sentence (Figure 2B).

The instantaneous phase information was then extracted from the Hilbert transform of the decimated signal:

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

$$\theta(t) = \arctan\frac{(H(t))}{(x(t))} \quad (2)$$

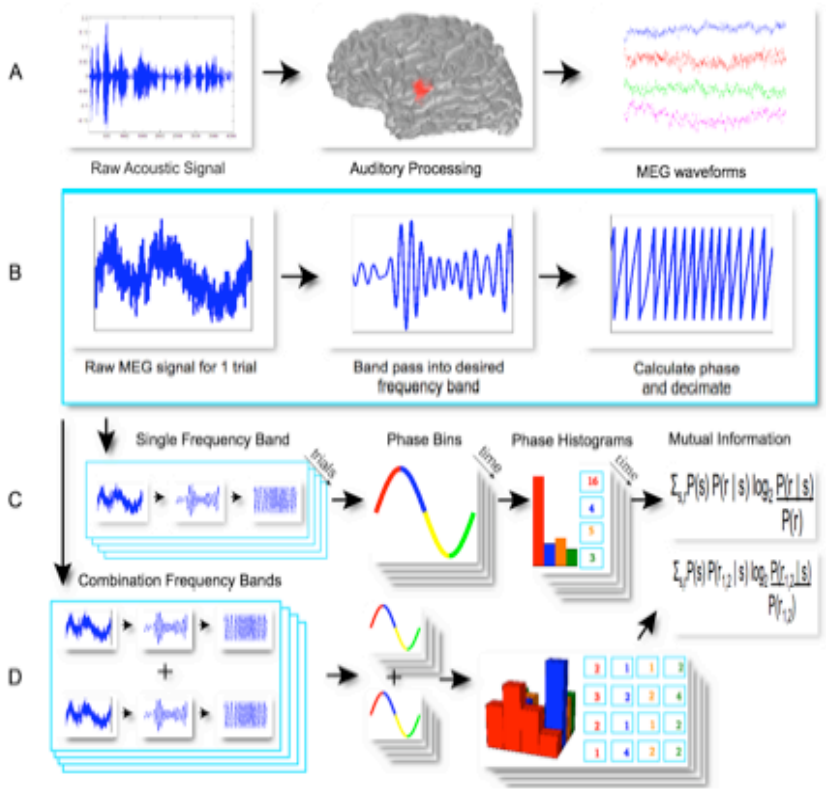

Figure 2. Outline of preprocessing and mutual information analysis (MI). A. Subjects listened to sentences, here shown as acoustic waveforms. 'Early' (acoustic, pre-semantic) processing of this information occurs in the auditory cortex, which can be measured effectively using MEG. B. Each signal from each trial, sentence and channel is band-passed into the frequency of interest, decimated and the phase is extracted from the Hilbert Transform. C. For each frequency sub-band, across trials, each phase response for each time bin is grouped into a 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 which is then used to assess the amount of information present in the combined frequency cases.

Mutual Information

All further analyses were done using the Information Breakdown Toolbox in Matlab (Magri et al. 2009). Mutual information (MI) between the response and stimulus was calculated using the following equation:

$$\mathrm{I(S;R)} = \sum_{r,s} P(s)P(r \mid s)\frac{P(r \mid s)}{P(r)} \tag{3}$$

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:

$$\mathrm{I(S;R)} = \mathrm{H(R)} - \mathrm{H(R|S)} \tag{4}$$

$$\mathrm{H(R)} = -\Sigma_r P(r) \log_2 P(r) \tag{5}$$

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

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 1/2750). 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 (Figure 2C), the response distribution was composed of phase responses that fit into 4 equally spaced bins: $-\pi$ to $-\pi/2$, $-\pi/2$ to 0, 0 to $\pi/2$, and $\pi/2$ to π. For the case of the frequency combinations (Figure 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 is that it is the minimum number of bins that adequately reflects the overall phase response and the second 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 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, $theta_{low}$, and $theta_{high}$) and also for each combination (delta + $theta_{low}$, delta + $theta_{high}$, and $theta_{low}$ + $theta_{high}$).

In this latter case, the combination values were computed using the 10 channels for each individual frequency band that showed the highest MI values.

Bias Correction

Since the estimation of mutual information is dependent on the sampling of the probability distributions with results approaching their true value as more data are sampled, a multi-step 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). Mutual information is computed on the total data set, a randomization of half the trials, and a randomization of a quarter of the trials. As is illustrated in Figure 3, a quadratic function is then fit to the data points and the actual mutual information is taken to be the zero-crossing value. This new value reflects the estimated mutual information for an infinite number of trials and greatly reduces the finite sampling bias (Panzeri et al. 2007, Montemurro et al. 2007).

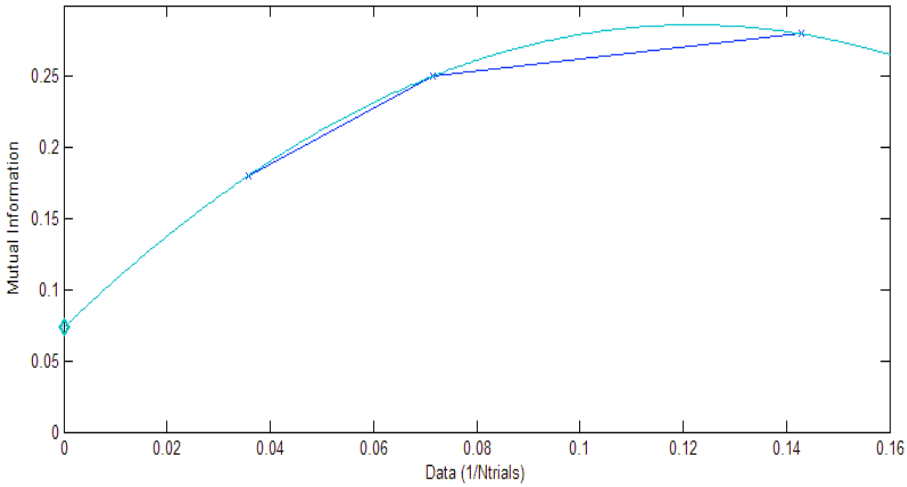

Figure 3. Bias correction: Quadratic extrapolation. Since the calculation of MI values depends on the sampling of the probability distribution, a larger number of trials will lead to a more accurate assessment of the information content. This method of bias reduction computes the MI values for the entire 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.

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:

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

Where I_{lin} is the linear sum of the MI values for each frequency band and I_{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):

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

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

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

Where $H_{ind}(R)$ replaces the summation of the marginal probabilities in equation (5) with the product of the marginal probabilities, and $H_{lin}(R)$ sums the probabilities from the single response cases (for the individual frequency responses in the present study). $I_{indcorr}$ represents the amount of noise correlation present that is independent of the stimuli. This

term can be thought of a measure of how similar two different neural responses are (in this case two different frequency bands) independent of which stimulus is presented:

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

Where $\chi(R)$ is the same as $H_{ind}(R)$ except that the normalization term (i.e. $P(r)$) is kept non-independent. Lastly, $I_{depcorr}$ 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 ΔI: the amount of information lost to a downstream decoder if noise correlations are ignored.

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

Where $H_{ind}(R|S)$ is calculated similarly to $H_{ind}(R)$, except that the product of the marginal probabilities is applied to equation (6) instead of (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, decimated, and the phase was extracted from the Hilbert transform. The phase value for each time point within each trial was binned using the same binning process as the MI analysis. Similarity between the template and the 3 comparison trials (one 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 non-match.

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 1000 times for each sentence, for each frequency and frequency combination (3 sentences x [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 using the same channels as in the MI analysis.

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

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

Since however, Chi square 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 square results are not a linear measure of the magnitude between variables and can therefore not 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 frequencies, the effect size of each of the 18 values for each subject was computed as a phi value:

$$\Phi = \sqrt{\frac{\chi^2}{N}} \tag{13}$$

Since this converts the chi square values into effect sizes, the performance of 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.

1.3 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 are shown in Figure 4 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 non-invasive recordings

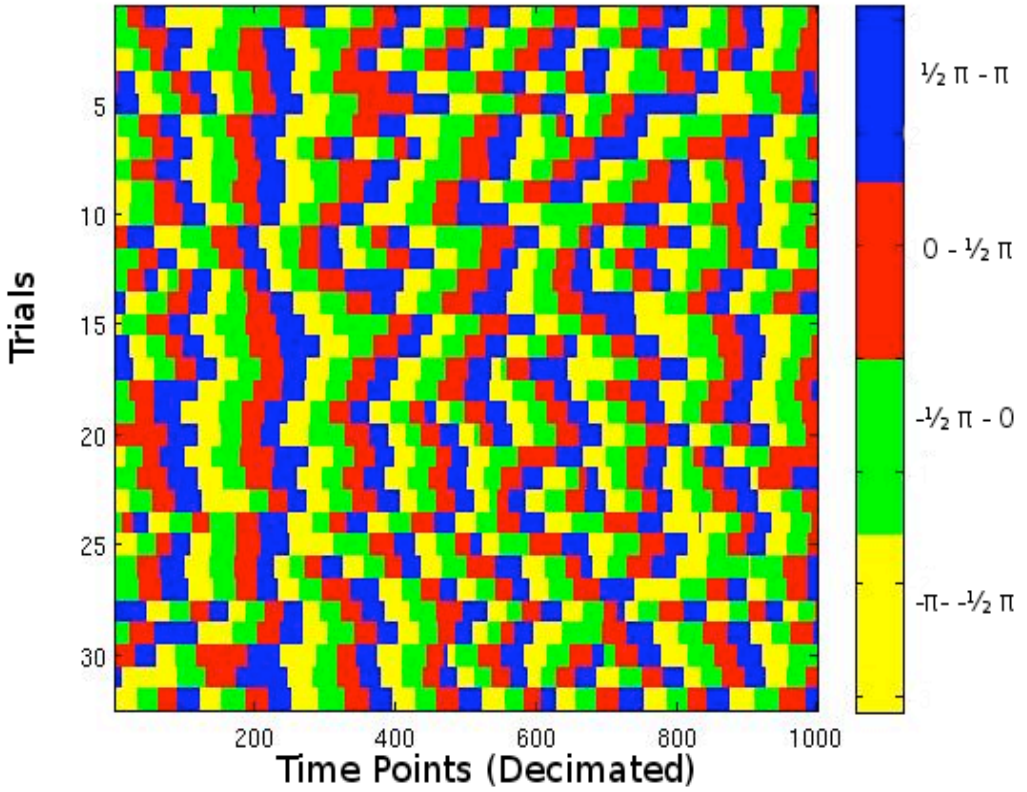

Figure 4. Binned phase response for a representative subject. Phase response taken from one subject, one channel, and a portion of one sentence. Note that the time points on the X-axis have been decimated by a factor of 4 and therefore reflect units of 4 ms each.

is manifest in the non-perfect 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 non-relational noise.

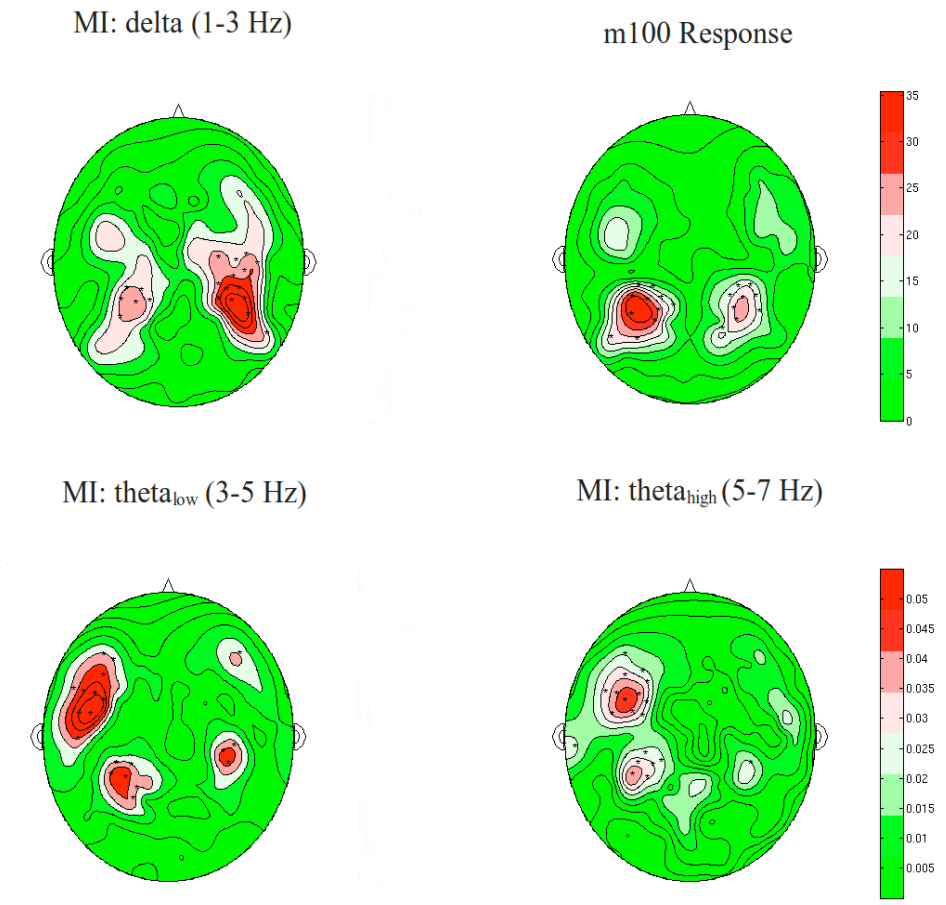

Figure 5. Topographic head-plots for a representative subject. MI is plotted for each frequency sub-band. Red denotes higher MI values and green lower values. As can been seen, the origin of the highest MI values compares favorably to the overall amplitude of the M100 response, consistent with an auditory origin for the channels yielding the MI values.

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 Figure 5. 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 (Pantev et al. 1990, Liégeois-Chauvel et al. 1994). 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).

MI values for each frequency band are plotted in Figure 6. Each value on the x-axis represents the center frequency of the filter utilized (see methods) and each bandwidth is 3 Hz. Results indicate that MI values are highest for delta, followed by $theta_{low}$ and then $theta_{high}$. 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 the 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 (approximately 22-28 Hz) for some subjects which is also present in the overall between subject plot. Unfortunately, no meaningful conclusions can be draw from these results as

the topography did not show an auditory response or for that matter, any coherent pattern at all (results not shown).

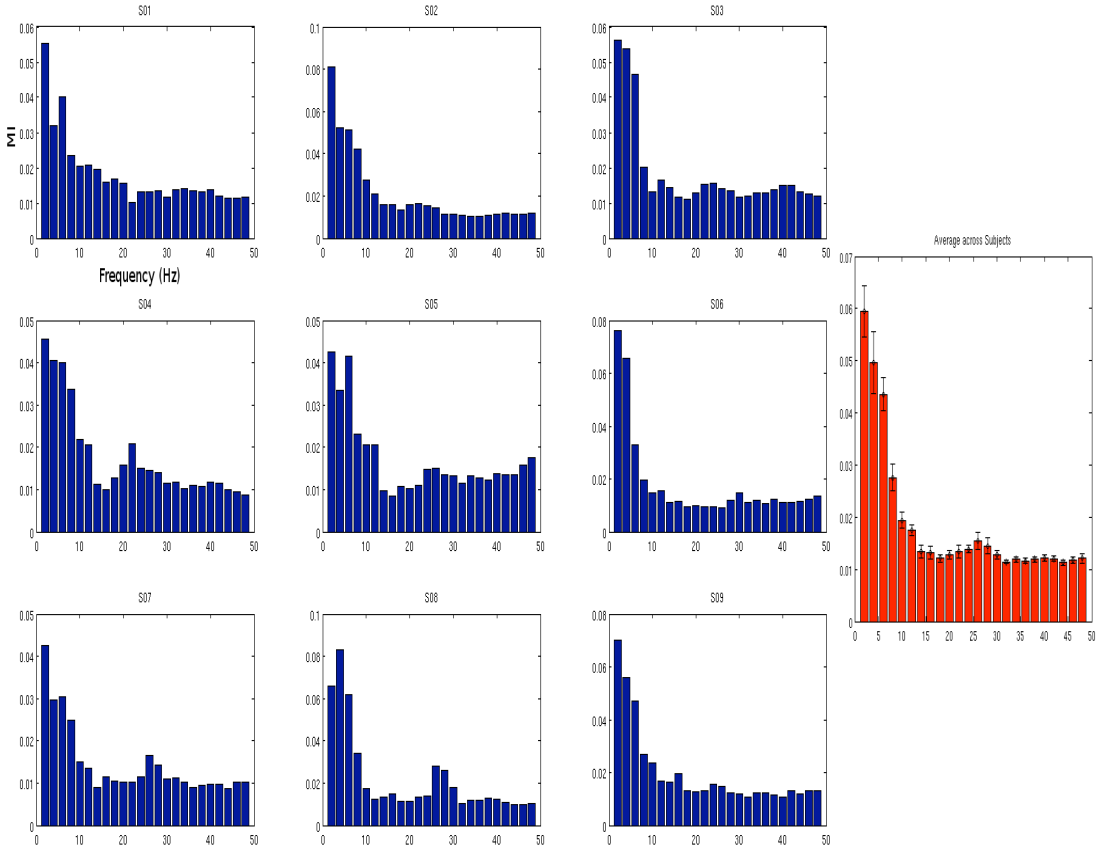

Figure 6. MI values for each Frequency band. The MI values are plotted here for each subject as a function of frequency. Each bar represents the average MI value across sentences for the top 10 channels. The plot on the right is the average across subjects (red). The MI values peaked in the low frequency range (< 8 Hz).

Figure 7 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 + $theta_{low}$, delta + $theta_{high}$, and $theta_{low}$ + $theta_{high}$ – 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, it surpassed it, albeit by a small amount.

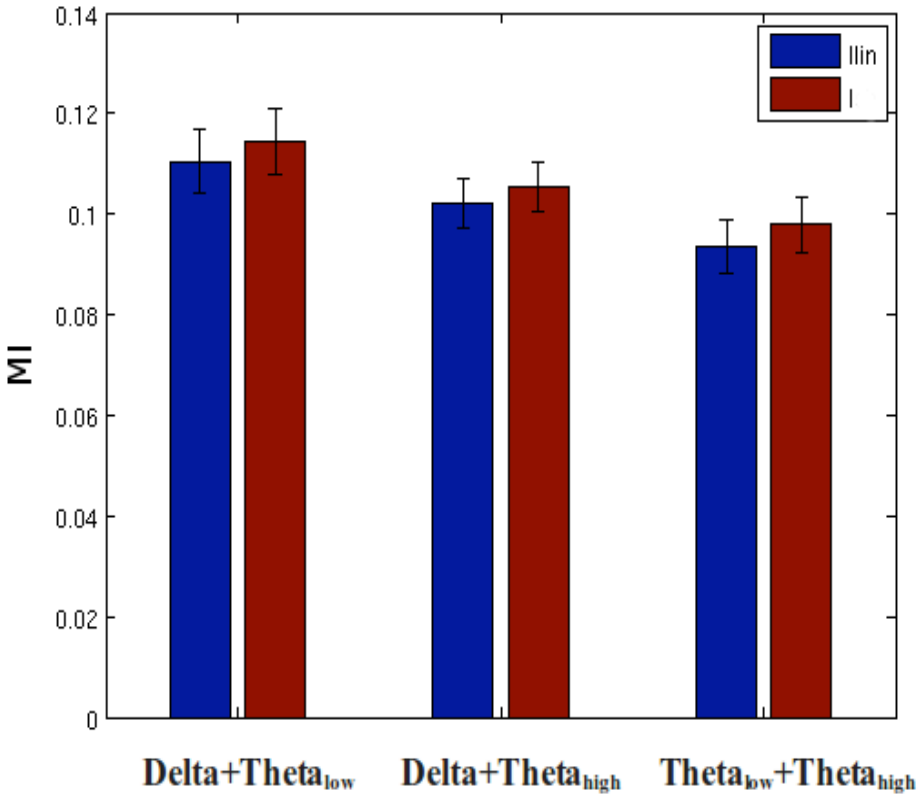

Figure 7. Average MI values for linear summations and combinations. The average MI values for the linear summation of each frequency sub-band (blue) and the combination of these sub-bands (red) is plotted. The combination values are quite

similar to the linear summation values, suggesting that each sub-band is in fact processing independent information.

Using the information breakdown method (Magri et al. 2009), 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:

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

This technique can be utilized to determine if 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). 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 due to 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 correlations. This would therefore undermine any interpretation of 'true' independence.

Results show that the amount of information lost due to stimulus independent noise correlations was extremely small, accounting for a loss of less than 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 + $theta_{low}$ and delta + $theta_{high}$, this value was effectively zero with only 1.7×10^{-6} and 4.2×10^{-6} bits lost for each combination respectively.

Conversely, stimulus dependent noise correlations led to an increase in information compared to 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 due to signal similarity was also extremely small (1.3×10^{-6} to 5×10^{-5} bits) suggesting that the role of 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 Figure 8, 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 inter-trial coherence for phase within the theta band was sufficient for this type of classification. The present study expands on these results by demonstrating both (i) that this result is obtainable with a different measure of information (MI vs. a phase dissimilarity function) and (ii) that information in the delta band also provides robust information that can lead to single trial based template classification (cf. Luo et al. 2010).

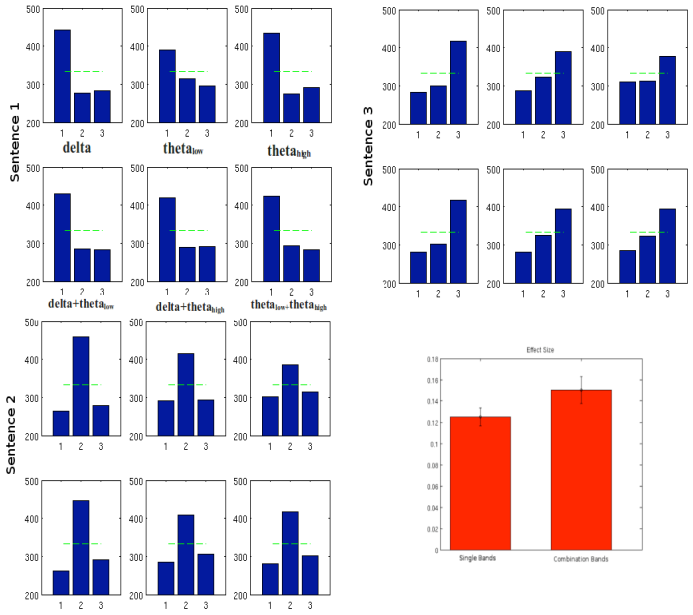

Figure 8. Classifier performance for a representative subject. Classifier data are shown for each frequency band and combination band for each sentence for a representative subject. The green dashed line represents chance performance. As is shown, each frequency band and combination successfully classified the template corresponding to a single trial of a given sentence with other tokens of that sentence. The effect size (red) 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.

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 Figure 6). This difference was found to be significant using 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.

1.4 DISCUSSION

The results of this study demonstrate novel insights into the relationship between different low frequency sub-bands of the phase of the macroscopic neural signal and offer a new approach to analyzing MEG data – mutual information analysis. The results also replicate and extend the findings of Luo and Poeppel (2007) and Luo et al. 2010 (see also Howard & 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 that the theta response is not only consistent across time but 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 mutual information between the response and the stimulus within this band is not simply due to longer periods, but rather that the information is meaningful. While earlier work (Luo & Poeppel 2007, Howard & Poeppel 2010) 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 approximately 4 seconds 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 approximately 30 seconds in length found 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 (~ 22-27 Hz) that appears in some subjects and slightly in the overall average (see figure 6) 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. MNE - Hämäläinen & Ilmoniemi 1994, 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 if and how information in the low frequency range is able to be dis-associated into separate frequency bands. The three relevant hypotheses were (i) 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 ($theta_{low}$ and $theta_{high}$); (ii) 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 & Poeppel 2010); or (iii) 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, $theta_{low}$, and $theta_{high}$) 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 less than 5% of the linear summated information). This suggests that any shared noise source common to all three responses is at best minimal, further supporting the notion of independence. 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.

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 to 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 both 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 multi-unit 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 ΔI, proposed by Nirenberg et al. (2001), characterized the amount of information lost due to 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: one, that this value was only a very small fraction of the overall linear summation of information between responses (less than 5 %) and two, that 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 (Nirenberg & Latham 2003, Dan et al. 1998). While 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 to the single band cases, this increase was modest.

It is important to note that non-invasive techniques, while offering the advantages of have 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, Strong et al. 1998, Montemurro et al. 2007, Nirenberg et al. 2001) 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, multi-unit or LFP data, the current 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 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 towards 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 implication 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 ($theta_{low}$ and $theta_{high}$) demonstrated a linear summation of information, as opposed to redundancy. This suggests that each theta sub-band 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 sub-bands are not responding in this manner, merely that all three cannot be tracking this type of information (Howard & Poeppel 2010).

Secondly, the current study validates a unique methodological approach to non-invasive electrophysiological recordings. While previous work using MI has focused predominately 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 non-invasive electrophysiological human data and lead to meaningful results. The strength of this approach is that it characterizes the non-linear 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 non-stationary 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 time scale (Luo & Poeppel 2007), time constants associated with the neural response itself, or some combination of the two (Howard & 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 time scales that are thought to reflect time constants of the neural processing (Howard & 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 time scales (e.g. syllables, prosodic information), neural mechanisms track all aspects of the input

signal below approximately 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).

Chapter 2: Phase tracking of speech and non-speech: A mutual information analysis

2.1 INTRODUCTION

One of the most significant challenges of speech perception is determining how the brain turns a continuous stream of sounds into segments of meaningful units of analysis, roughly corresponding to first *parsing* the input stream into units used for *decoding*. Recent work using magnetoencephalography (MEG) and electroencephalography (EEG) has implicated low frequency phase information of the neural signal as a possible mechanism for this segmentation (Luo and Poeppel 2007, Luo et al. 2010, Howard & Poeppel 2010, Abrams et al. 2008). The initial hypothesis (Luo & Poeppel 2007) was that ongoing oscillations in the theta band (4 – 8 Hz) reset at the onset of syllables, tracking the acoustic correlates of the linguistic features of the input (but see Luo et al. 2010 for evidence that activity in delta , 1 – 3 Hz, also occurs).

Briefly, a neuronal phase pattern that corresponded to a particular sentence was compared to the phase pattern that corresponded to different sentences, and it was found that reliable phase information that occurred at set time points across trials of a single sentence within the theta band was not only consistent within a token sentence, but the patterns were unique to individual tokens and could be used to discriminate between different sentences – the phase dissimilarity function. A lowering of intelligibility led to a decrease in this response, leading to the hypothesis that it was specific both to speech content and the successful segmentation of the incoming signal.

This low frequency information corresponds to the peak of the modulation spectrum of the speech signal (Greenberg et al. 1996, Greenberg 2006). A speech signal can be divided into two corresponding components: the slow amplitude fluctuations that reflect the envelope of the signal and correspond mechanistically most closely to the opening and closing of the jaw (and, as such, vocalic information), and faster fine structure that contains more fine-grained frequency modulations that correspond to the movement of the articulators (Poeppel 2001, 2003). This first component, the slow, low frequency fluctuations of the envelope, is thought to be the acoustic component being tracked by the phase dissimilarity function (Luo & Poeppel 2007).

A recent model (Howard & Poeppel 2010), however, suggested that rather than reflect a reset of ongoing oscillations to linguistic features (i.e. syllables), the theta band phase response is simply the convolution of canonical evoked responses, the N1-P2 complex: a reliable onset response that responds to the onset of all sounds, including tones and brief clicks (Roberts et al. 2000, Howard & Poeppel 2009). This neural response would be sensitive to sharp acoustic transitions within the speech input. Evidence for this comes from contrasting speech with time-reversed speech (Howard & Poeppel 2010). Reversed speech is completely unintelligible but maintains many of the gross features of normal speech. If the phase of the neuronal response and consequently, the phase dissimilarity function, is dependent on intelligibility, then it should have been present in the normal speech condition and not the reversed condition as only in the former is the signal intelligible. The results demonstrated that phase dissimilarity was present and robust in both conditions, suggesting that this particular response was in fact largely driven by acoustics and not intelligibility (Howard & Poeppel 2010).

A model was presented that argued against the reset of ongoing oscillations hypothesis and in favor of a hypothesis that posited the underlying mechanism responsible for phase tracking as the evoked N1-P2 complex. Furthermore, this model suggested that the evoked response responsible for this phase dissimilarity was tracking acoustic transitions in the input rather than the ongoing changes in the envelope of the incoming speech signal, contrasting with previous work that found correlations in the right hemisphere between the electroencephalographic (EEG) signal and the envelope of the speech signal (Abrams et al. 2008).

A recent study by Deng and Srinivasan (2010) however, using both speech and reversed speech as stimuli, found correlations between the low passed envelope of the speech signal (< 30 Hz) and the frequency decomposed EEG signal for speech but not reversed speech, suggesting that the component measured (correlation of the power fluctuations of the neural signal with the speech envelope) is in fact sensitive to the intelligibility of speech, in contrast to the results of Howard and Poeppel (2010) using phase.

Taken together, there are two separate hypotheses related to the phase tracking of the incoming speech signal. The first is that ongoing oscillations reset themselves at the syllable boundaries and are consequently crucial for the meaningful segmentation of speech (Luo & Poeppel 2007). The response centers in the theta band because of the corresponding time scale of average length of the syllable (125 – 250 ms), which corresponds to the peak of the modulation spectrum of the speech envelope (Greenberg et al. 1996, Greenberg 2006). The second hypothesis is that the measured phase response is purely acoustic in nature and represents the evoked potential elicited by the onset of

sounds in general (Howard & Poeppel 2010). This hypothesis also states that it is these onsets that drive the low frequency phase response and not the envelope.

What remains unclear is whether or not the phase response in all studies examined is in fact the same response throughout. It is possible that while reversed speech elicits a robust evoked response and consequently high phase dissimilarity and inter-trial phase coherence, it is not in fact the same component that is causing the phase dissimilarity/ phase coherence in the speech condition. This would explain the results of Deng and Srinivasan (2010) who found envelope tracking in only the speech condition and not the reverse speech condition: top-down control over tracking of the input signal would be activated in the speech, but not the non-speech condition.

In order to further elucidate the acoustic contributions to the low frequency phase response as well as the nature of this response in speech and non-speech, a metric that can assess both the content of neural response as well as their relationship with one another, is required. Mutual information (MI) analysis was used to characterize not only the different phase responses, but also their interactions with each other (Cogan & Poeppel, in prep., Kayser et al. 2009, Strong et al. 1998, Magri et al. 2009, Montemurro et al. 2008).

2.2 METHODS

Subjects

Eight native English speaking subjects (5 male, mean age 28.9) 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) and the University of Maryland Institutional Review Board. Subjects were right-handed as assessed by the Edinburgh Inventory of Handedness (Oldfield 1971). One subjects' data were not included in the analysis due to saturation of channels by external noise.

Stimuli

Three different English sentences were obtained from a public domain internet audio book website (http://librivox.org). Each of the sentences was between 6 and 6.5 seconds (sampling rate of 44.1 kHz) and each was spoken by a different speaker (American English pronunciation, 1 female). 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 sounds levels (~72 dB SPL trials) within 4 separate blocks.

A second condition used stimuli that contained the envelope from the original sentences and a random Gaussian noise band carrier. These sentences were constructed by first band-passing the broadband speech signal into two separate frequency bands using a 500 point two-way least squares linear FIR filter, shifted backwards to compensate for phase delays due to the original filtering. The frequency bands were from 80 to 1240 Hz and 1240 Hz to 8820 Hz. The values were taken from a previous paper (Smith et al. 2002) that created speech chimeras, and are thought to reflect the spacing of the cochlear frequency map (Greenwood 1990). The envelope in each

frequency band was then extracted via the Hilbert Transform:

$$H(t) = \frac{1}{\pi} p.v \int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau} d\tau \quad (15)$$

A Gaussian white noise carrier was added to each of the two frequency-band envelopes. Each constructed speech envelope and noise carrier was then normalized against the maximum of the power of the corresponding frequency band of the original sentence and then summed together. The combined signal was then normalized again against the overall power of the original sentence.

These stimuli contained information corresponding to the original envelope of the signal but were entirely unintelligible. These stimuli therefore preserve a key feature of the original signal believed to be important for speech perception (the envelope – Smith et al. 2002, Luo & Poeppel 2007), but remove the intelligible content. A third condition was silence 'presented' for 6.5 seconds. This condition was not analyzed for the present study. Each condition contained 32 trials for a total of 192 trials analyzed. The order of conditions was randomized within each block, with a randomized inter-stimulus interval (ISI) between 800 and 1200 ms.

Task

Participants were instructed to listen to the sentences with their eyes closed. This was done to limit ocular artifacts. 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 ms and 1000 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 1000 Hz, a notch filter at 60 Hz (to remove line noise), a 500 Hz on-line analog low pass filter, and no high-pass filter. Each subject's head position was assessed via five coils attached to anatomical landmarks both before and after the experiment to ensure that head movement was minimal. Headshape data were digitized using a three-dimensional digitizer (Polhemus).

Signal Processing

The general flow of signal processing has been described in detail previously (Cogan & Poeppel in prep.). Briefly, firstly, for each trial, subject, condition, and sensor, the time series of the neuronal signal was decimated by a factor of four to reduce

computational overhead. Each signal was then band passed in frequencies of interest: delta: 1-3 Hz, $theta_{low}$: 3-5 Hz, and $theta_{high}$: 5-7 Hz. This was done using a 814 point two-way least squares linear FIR filter (rounded to 204 points after decimation) that was shifted backwards in time to compensate for delays caused by the filter. Only the first 6 seconds of data were analyzed for each trial, creating 1500 data points for each trial, condition, frequency, subject and sensor.

The phase information was then extracted using the Hilbert transform (equation 15):

$$\theta(t) = \arctan\frac{(H(t))}{(x(t))} \tag{16}$$

Mutual Information

All further mutual information analysis was computed using the Information Breakdown Toolbox in Matlab (Magri et al. 2009). Mutual information in this context assesses the amount of information between an input signal and the neural response (in this case low frequency phase information). Formally it is expressed as:

$$\mathrm{I(S;R)} = \sum_{r,s} P(s)P(r \mid s)\frac{P(r \mid s)}{P(r)} \tag{17}$$

Where P(s) is the probability of the stimulus, P(r) is the probability of the response and P(r | s) is conditional of the response given the stimulus. The average amount of mutual information (I(S;R)) between the set of stimuli (S) and responses (R) can be thought of as either the average amount of information that a response conveys about a stimulus, or the reduction of entropy of the response space that the stimulus provides (Kayser et al. 2009, Montemurro et al. 2008, Magri et al. 2009).

In the present study, each stimulus was a decimated time point (corresponding to 4 ms of the original speech signal) and each response was the phase of the neuromagnetic filtered signal in the three frequency bands of interest: for each subject, trial, condition, frequency band, and channel, the overall phase distribution was divided into four equally space bins: $-\pi$ to $-\pi/2$, $-\pi/2$ to 0, 0 to $\pi/2$, and $\pi/2$ to π. The MI was based on the probability distributions formed by the binned MEG responses for each stimulus (time point) across trials.

To assess the interactions between each frequency sub-bands, 16-bin histograms were utilized in which each bin reflected the co-occurrence of phase responses between the different frequency sub-bands, compared two at a time, for a total of three combination conditions: delta + $\text{theta}_{\text{low}}$, delta + $\text{theta}_{\text{high}}$ and $\text{theta}_{\text{low}}$ + $\text{theta}_{\text{high}}$. Both the individual frequency and the combination MI values were computed for all tokens of both the speech and envelope condition. The MI value was taken to be the average of the top ten channels for the individual bands and the combination frequency band analysis was performed on these ten channels for each condition and subject.

Synergy and Redundancy

The concept of synergy and redundancy has been outlined previously (Cogan & Poeppel in prep., Kayser et al. 2009, Montemurro et al. 2008). Briefly, for the frequency combinations, redundancy was defined as:

$$I_{redundancy} = I_{lin} - I_{tot} \quad (18)$$

Where I_{lin} is the linear sum of the MI values for each frequency band and I_{tot} is the MI value for the frequency combination. Negative redundancy is denoted as 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 can be broken down into its constituent parts as per the formalism of the information breakdown method (Magri et al. 2009):

$$I_{syn} = I_{sigsim} + I_{indcorr} + I_{depcorr} \quad (19)$$

I_{sigsim} is the amount of information lost due to correlations in the signal itself. $I_{indcorr}$ represents the amount of noise correlations present that is independent of the stimuli. This term can be thought of a measure of how similar two different neural responses are, independent of which stimulus is presented. $I_{depcorr}$ 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 ΔI: the amount of information lost to a downstream decoder if noise correlations are ignored.

Bias Correction

Since the core of the estimation of MI is based on a sampling of the probability distribution of the responses, a finite number of responses can bias the amount of MI calculated. A multi-step bias correction method was utilized (Kayser et al. 2009, Montemurro et al. 2007, 2008, Panzeri et al. 2007) in which for both the single frequency cases and the comparison cases, the MI values were calculated for a random shuffling of trials that equaled half the total number of trials and then for a random shuffling of trials that equaled a quarter of the total number of trials. A quadratic function was fit to the resulting three data points (all trials, half trials, and quarter trials) and the 'true' MI value was taken to be the zero crossing of this function (Strong et al. 1998).

A bootstrapping procedure was also utilized in which for each of the individual frequency and combination conditions, 20 time-shuffled versions of the stimulus were created and the mean of the resulting MI values produced was subtracted from the MI value obtained (Kayser et al. 2009, Montemurro et al. 2007, 2008, Panzeri et al. 2007). Lastly, for the frequency combination cases, to remove any within trial noise correlations, the MI value was calculated on a version of the data in which the values within each probability distribution within each stimulus (i. e. time point) were shuffled across trials but the marginal probabilities were held equivalent to the un-shuffled data values. Together, these three methods have been shown to significantly reduce bias associated with MI values (Kayser et al. 2009, Montemurro et al. 2007, 2008, Panzeri et al. 2007).

Classifier

Since MI only characterizes the specific relationship between a response and the stimulus presented, a classifier was used to determine if the information present as revealed by the MI analysis could be used to discriminate between tokens of the same stimulus category (i.e. speech and envelope). This was done by binning the individual frequency-specific phase responses of each token as in the MI analysis, and then computing the inner dot product between a trial of one sentence and a random trial from the same sentence as well as a random trial from each of the other two sentences within the stimulus category (i.e. speech and envelope). The highest sum of the inner dot products was taken to be the comparison that was most similar. This was repeated 1000 times for each individual frequency band, token, combination of bands, condition (envelope/ speech), and subject. As for the MI analysis, the top 10 channels for the individual frequency band for each subject, frequency band, and condition was utilized.

To assess statistical significance, a chi square was first performed for each comparison (2 conditions x 3 sentences x [3 individual frequency bands + 3 combination bands] = 36 classifier results per subject):

$$\chi^2 = \sum_{i=1}^{k} \frac{(x_i - \mu)^2}{\sigma^2} \tag{20}$$

and then converted to phi values to group across subjects:

$$\Phi = \sqrt{\frac{\chi^2}{N}} \tag{21}$$

Finally, paired 2 sample t-tests were performed using the phi values to test significance between the conditions, and individual frequency bands/ combinations within conditions.

2.3 RESULTS

Mutual Information

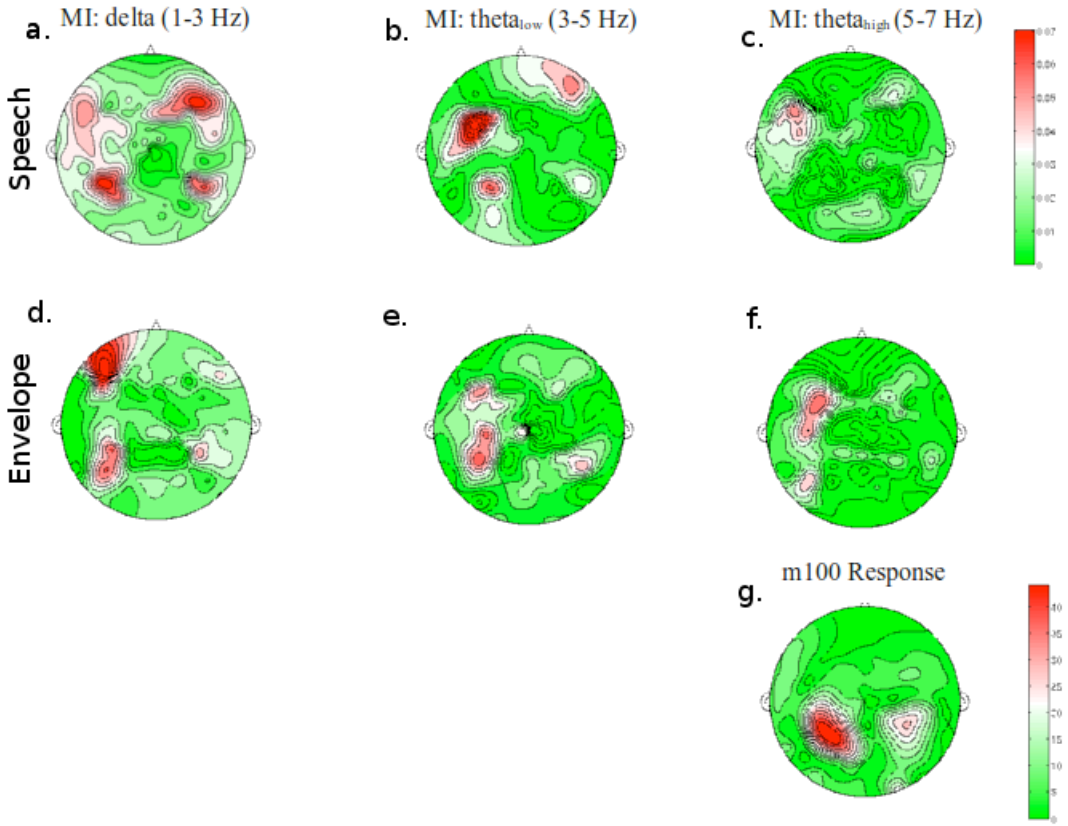

Figure 9. Topographic head-plots for a representative subject. a-c represents the topographic pattern for the speech condition for delta, theta$_{low}$ and theta$_{high}$

respectively, while d-f represents the same frequency bands for the envelope condition, with red indicative of higher MI values. The RMS of the m100 response is also plotted for the same subject for reference purposes (g). A clear auditory response is present in all frequencies and conditions.

The topographic maps for a typical subject compared to the m100 response (g) can be seen in Figure 9, with a-c representing delta, $theta_{low}$, and $theta_{high}$ for the speech condition and d-f representing the envelope condition the same frequency bands. The m100 response is thought to originate in auditory regions on the superior temporal gyrus, near the transverse temporal gyrus (Pantev et al. 1990, Liégeois-Chauvel et al. 1994). This demonstrates that a prototypical auditory response is produced in both conditions for all frequency bands.

MI was assessed for each individual frequency band (delta, $theta_{low}$, and $theta_{high}$) and each frequency combination (delta + $theta_{low}$, delta + $theta_{high}$, and $theta_{low}$ + $theta_{high}$), and for each subject and condition (speech and envelope). Results within the speech condition were similar to previous findings (Cogan & Poeppel in prep.), in that MI values were highest for delta. They differed slightly in that $theta_{high}$ provided slightly higher MI values than $theta_{low}$ (see Figure 10).

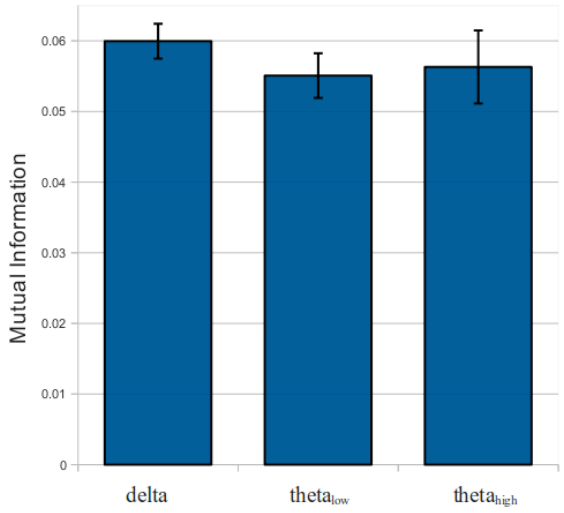

Figure 10. MI values for each individual band for the speech condition. Each bar represents the average MI value for the top 10 channels across subjects for each frequency sub-band of interest. Delta provides the highest MI values, followed by theta$_{high}$ and theta$_{low}$.

The amount of information present in the combination bands was also similar to previous results, with MI values being slightly higher than the predicted linear values (see Figure 11). These results reproduce previous work using similar stimuli and methods and establish a baseline for comparison (Cogan & Poeppel in prep.).

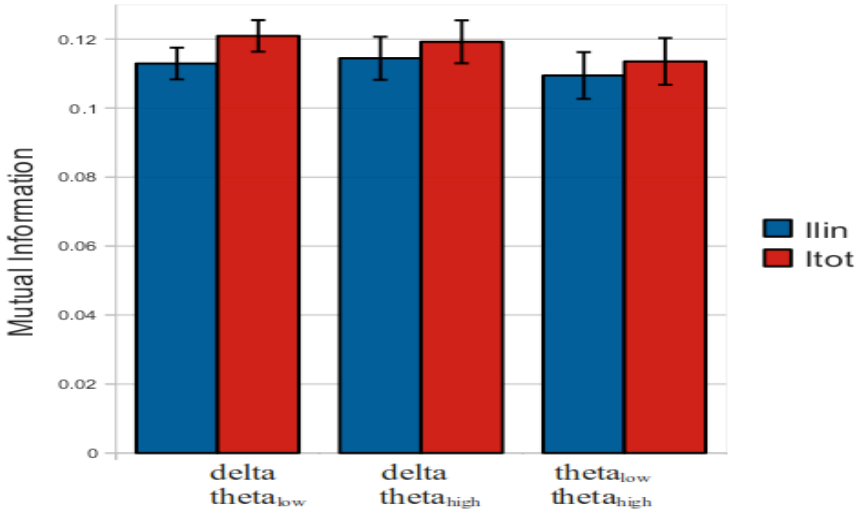

Figure 11. Combination MI values for the speech condition. The overall information present in the combination MI analysis (red) is plotted against the sum of the information present in the individual bands (blue). MI values were only slightly higher for the combination bands.

For the envelope condition, MI values showed a slightly different pattern with the highest value occurring in theta$_{low}$, followed by delta and then theta$_{high}$ (see Figure 12). Similar to the speech condition, the amount of information in the combination bands was slightly higher than the predicted linear combination value (Figure 13).

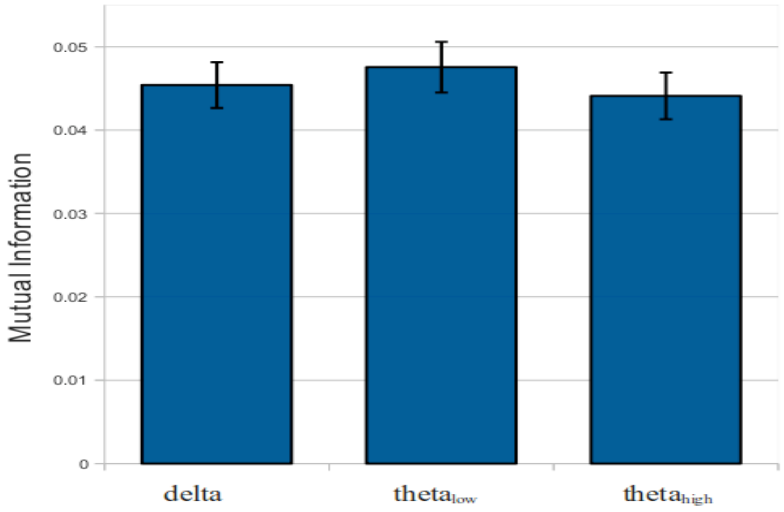

Figure 12. MI values for each individual band for the envelope condition. The overall convention for this plot is the same as for Figure 9, except it represents the individual MI values for the envelope condition instead of the speech condition. Unlike the speech condition, the peak response is in the $theta_{low}$ band (instead of the delta).

Overall, the amount of mutual information present between the signal and the response was much larger in all three bands examined for the speech condition as compared to the envelope condition (Figure 14). This contrasts with previous work that demonstrated that a phase dissimilarity function was similar for both speech and non-speech (reversed speech – Howard & Poeppel 2010). It is important to keep in mind however, that due to the differences between the metric utilized in the current study versus previous work, differences could also be due to non-linear aspects of the signal.

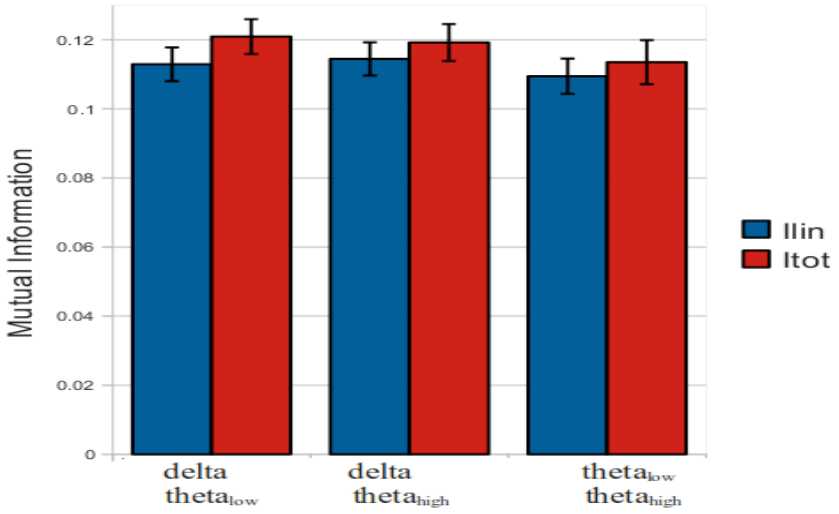

Figure 13. Combination MI values for envelope condition. Similar to the speech condition, the combination MI values (red) are only slightly higher than the sum of the individual bands (blue).

Furthermore, the amount of total information present in the combination bands was significantly higher for the speech condition as compared to the envelope condition. The amount of added information present (synergy) did not differ between speech and non-speech (data not shown), indicating that each condition displayed a near linear sum of information from the constituent frequency sub-bands.

In order to assess whether or not the information within each frequency band was in fact tracking the same aspects of the signal in the speech and envelope condition, MI values for each band were first compared to each other to establish a baseline of redundancy, and then compared between conditions to determine whether or not the

information contained within the envelope condition was simply a weaker version of the speech condition, or whether the response was fundamentally different.

For the within condition case, results show that the information is, as expected, quite redundant. It is important to remember that due to the trial shifting aspect of the bias correction, response are not compared directly with each other, but with a response at the same time point but in a different trial. Results can been seen in Figure 16a. Redundancy values for both the speech and the envelope conditions ranged from 54.12 – 61.07 %, which is close to the 50 % value expected for perfectly redundant, yet equally informative signals. Also as expected, the stimulus-independent noise correlations were also quite high, accounting for 109 to 112.98 % of the redundancy. These results establish that the information contained within each frequency band within each condition is heavily redundant, as is to be expected for responses that are consistent across trials.

Next, the MI values for each frequency band were compared across conditions. If the frequency phase information within each band is tracking the same elements of the signal, then these results should mimic the within condition results: MI combination value should be redundant. If however, in the envelope only condition, the low frequency phase portion of the MEG signal is tracking a different element of the signal, then the MI values should be independent. Results support the latter hypothesis. MI values across conditions but within each frequency bands were independent as can be seen in Figure 16b. While the I_{tot} values (amount of information obtained) were slightly less than the I_{lin} values, this difference was extremely small, accounting for less than 10

% of the predicted linear information values, in line with values that were previously found for independent retinal ganglion cells (Nirenberg et al. 2001).

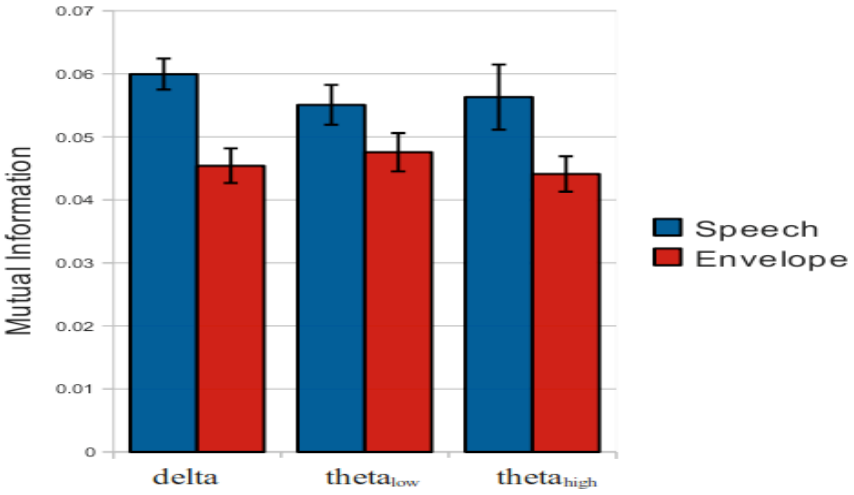

Figure 14. Comparison of individual frequency MI results between the speech and the envelope condition. The MI results for the speech condition (blue) were higher than the envelope condition (red).

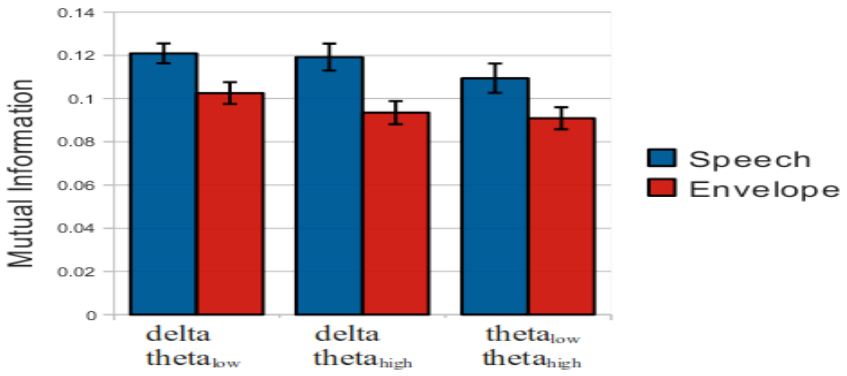

Figure 15. Comparison of combination frequency MI results between the speech and the envelope condition. The MI results for the speech condition (blue) outperformed the envelope condition (red)..

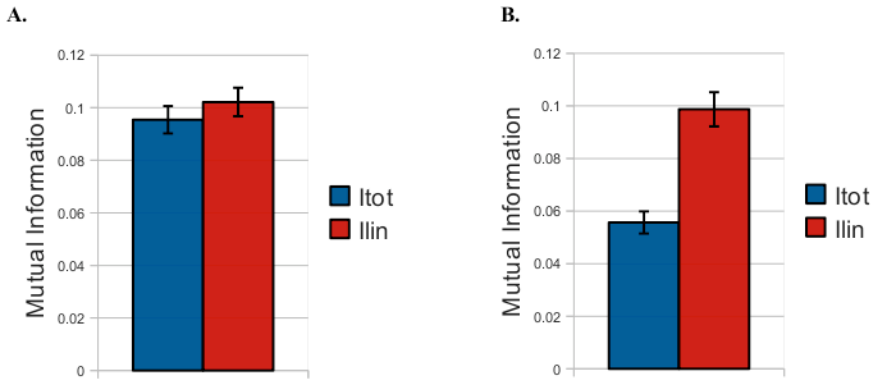

Fig 16. Average MI values for single frequency bands between and within conditions. The average MI values for all three sub-bands for the between conditions (a) and within conditions (b) demonstrate that information is redundant for within conditions and independent for between conditions. This suggests that

the phase tracking in the envelope condition is qualitatively different from the phase tracking in the speech condition.

Classifier Results

Classification results for a typical subject can be seen in Figure 17. 17a shows results for the speech condition. Each sentence token is presented in groups of six, with delta, $theta_{low}$ and $theta_{high}$, then delta + $theta_{low}$,delta + $theta_{high}$ and finally $theta_{low}$ + $theta_{high}$, moving clockwise. Results for the envelope condition are plotted in 17b. As can be seen, results for the frequency combinations performed better than the individual bands, suggesting that the independence of the information revealed by the MI analysis is also present in the classifier results and can be used to discriminate between tokens within conditions.

These Classifier results for the speech condition replicated the results of Cogan and Poeppel (in prep.). The mean phi score for the combination frequency classifier was 0.1153 compared to 0.1018 for the individual frequency based classifier (see figure 18a). This difference was found to be significant, $t(62) = 2.82$, $p = 0.007$. Results for the envelope condition also demonstrated higher phi values for the combination classifier as compared to the individual frequency classifier, with phi values of 0.0959 and 0.0769 respectively (see Figure 18b). This difference was also significant: $t(62) = 4.52$, $p = 0.0003$. While previous work using a phase dissimilarity function found robust and comparable phase tracking and phase classification using speech and non-speech (Howard & Poeppel 2010), the present study found a significant difference between speech and non-speech (envelope) conditions.

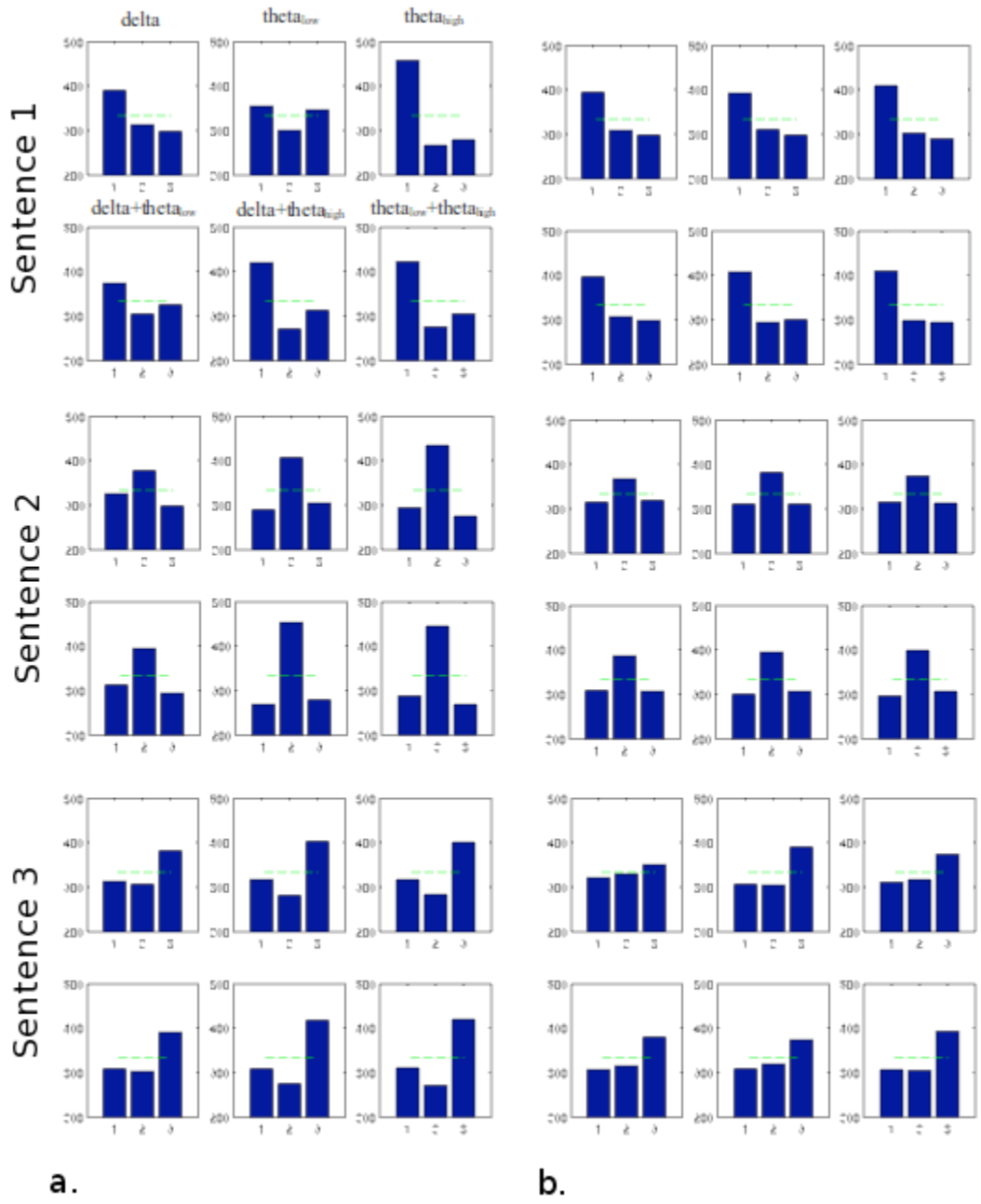

Figure 17. Classifier results for a representative subject. Classifier results are shown for a single representative subject. The top first three panels in the top row

represents the classifier results for sentence 1 for delta, $theta_{low}$ and $theta_{high}$ for the speech condition (a.), while the top 3 left panels represent the same frequency responses for the envelope condition (b.). The second row represents the combination classifier results for speech (left) and the envelope (right) condition. This convention is continues for sentences 2 (middle panels) and 3 (bottom panels). As can be seen, the classifier results for the speech condition outperformed the envelope condition in both the individual and the frequency combination results.

The higher MI values for speech versus the envelope condition could simply reflect the within token phase coherence across trials as opposed to the ability for the phase response to distinguish between different tokens both within the speech and envelope conditions. The comparison classifier results suggest that this is not the case. As can be seen in Figure 18b, both the single band and the combination band speech classifier outperformed the envelope classifiers. These results were significant, with $t(62) = 3.53$, $p = 0.004$ for the single band classifier comparison and $t(62) = 2.58$, $p = 0.012$ for the combination band classifier. While the average was higher (0.1018 versus 0.0959), the difference between the speech single band classifier and the envelope combination band classifier did not reach significance.

Taken together, both the MI and classifier results suggest that low frequency phase information occurs in at least three distinct independent bands: 1-3 Hz, 3-5 Hz, and 5 -7 Hz. While this separation is present for both speech and non-speech, the amount of information is far higher for speech versus non-speech (envelope). These results contrast with previous work (Howard & Poeppel 2010) that found no difference in a phase dissimilarity function for speech versus non-speech (reversed speech). Furthermore, phase tracking of the input signal for speech and non-speech differs qualitatively as the

MI values for within-frequency but across conditions were independent. This suggests that different elements of the signal are being tracked as opposed to the same elements being tracked less successfully.

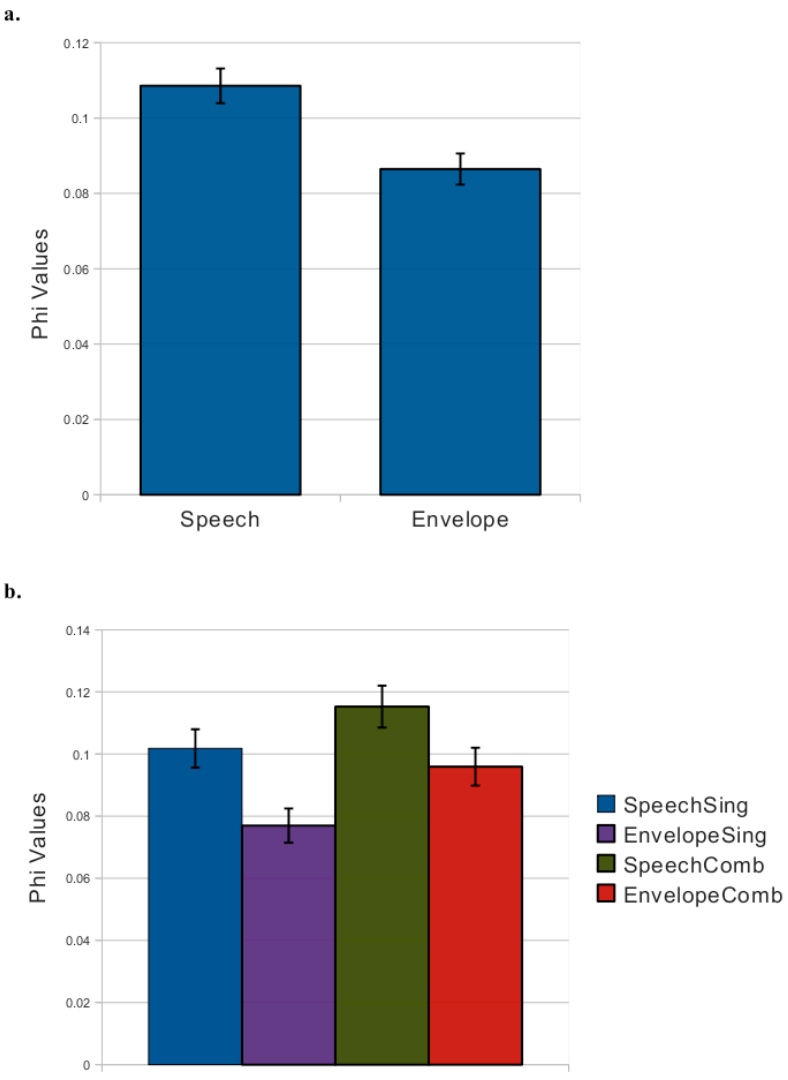

Figure 18. Phi values for classifier results. 17a. plots the overall mean of the classifier results for the speech versus the envelope condition and plot 17b plots the mean classifier results for the individual band speech condition (SpeechSing), the individual band envelope condition (EnvelopeSing), the speech combination results

(SpeechComb), and the envelope combination results (EnvelopeComb). Both the individual and the combination speech classifier outperformed the envelope classifiers.

2.4 DISCUSSION

The present study investigated, using MEG, the nature of low frequency phase tracking of the acoustic input signal. Subjects listened to stimuli in two different conditions: speech and non-speech, specifically a manipulation in which the fine structure of the signal was replaced with Gaussian white noise and the envelope was left intact. Two bands of decomposition were chosen as this has been found to be below the level of intelligibility (Smith et al. 2002) so as to better separate the contributions from acoustics and intelligibility.

The neural signal was separated into three bands of interest: delta (1 – 3 Hz), $theta_{low}$ (3 – 5 Hz), and $theta_{high}$ (5 – 7 Hz) and analyzed using Mutual Information (MI), which has previously been shown to be robust at both identifying the low frequency phase contributions to speech perception, but also the relative contributions of sub-components within this low frequency range (Cogan & Poeppel in prep.).

Results for the speech condition were consistent with previous MI MEG results: Robust MI values were found in all three bands, with delta demonstrating the highest MI values. The amount of information present in the combination frequency cases (delta + $theta_{low}$, delta + $theta_{high}$, and $theta_{low}$ + $theta_{high}$) was only slightly higher than the amount of information present in the individual frequency bands, suggesting that the information

present in these sub-bands are largely independent. This was again consistent with previous results (Cogan & Poeppel in prep.).

Classifier results also showed that the signal measured was in fact able to discriminate between tokens, providing further evidence that the components being measured are specific to the token present and not simply noise-related. Furthermore, classifier results for the combination frequency bands were better than those of the individual frequency band classifiers, further supporting the claim of independence of these signals.

Results within the envelope condition were similar to the speech condition. Each sub-band examined contained reliable information, although the relative contributions of each band were different. The amount of MI present in the combination analysis was only slightly higher than the amount present in the individual bands, which was similar to the speech condition. Classifier results were able to distinguish between tokens, with combination band results outperforming the single band results.

The comparison between speech and the envelope only condition demonstrated higher MI results for each frequency band examined and higher combination results. These higher levels of information were also manifest in the classifier results, where both the speech single band classifiers and the speech combination band classifiers outperformed the envelope only ones.

Lastly, and crucially, comparisons of within frequency MI values both within and across conditions demonstrated that while combination MI values within a condition were redundant, values between conditions were independent. This suggests that the portion of the signal being tracked in each condition is qualitatively different with the

phase tracking in the envelope only condition tracking a different portion of the signal as compared to the speech condition.

Previous theories about the nature of this low frequency phase response posited that it was related to either intelligibility (Luo & Poeppel 2007) or purely acoustically driven (Howard & Poeppel 2010). The former study utilized two manipulations to reduce intelligibility based on a study by Smith et al. (2002). In one condition, the fine structure of the signal was maintained but the envelope removed, and in the other, the envelope maintained but the fine structure was removed, in both cases signals with reduced intelligibility. Their results demonstrated a reduced phase dissimilarity function response for the manipulated conditions as compared to the speech condition. They argued that the phase tracking response was therefore driven, at least in part, by intelligibility and the phase response reflected resetting of endogenous oscillations to syllabic boundaries. The second manipulation utilized was similar to the one used in the present study, however, in the present case, the signal was entirely unintelligible, which better separating the acoustic component from the intelligibility aspect of the input signal.

A later study (Howard & Poeppel 2010) using the same phase dissimilarity function, found no difference between speech and reversed speech. A model was proposed that suggested that that low frequency phase response was a concatenation of evoked responses. Furthermore, the response could be elicited by any sound that contained an onset, arguing for a purely acoustic interpretation of the phenomenon.

The present study supports the former interpretation over the latter, although with some caveats. Both MI and classifier results were higher for speech versus the envelope only condition, suggesting, contrary to previous results (Howard & Poeppel 2010), that

this response has a preference for speech compared to non-speech. The independence of the MI values within frequency but across conditions, also argue against a simple additive model in which both tracking of the envelope and sharp acoustic transitions in the fine structure occur and sum linearly to produce the measured response. If this were the case, the MI values across conditions would be heavily redundant instead of independent. Instead, the present study supports the hypothesis that either this phase response is driven at least in part by the intelligibility of the response itself, or that the fine structure component of the signal fundamentally alters the manner in which the low frequency-portion of the brain signal tracks the input.

The latter is unlikely for two reasons: firstly, this would be inconsistent with previous EEG results that found robust tracking of the speech envelope itself (Abrams et al. 2008) and with results that demonstrated a preferential tracking of the speech envelope over the envelope of reversed speech (Deng & Srinivasan 2010). Secondly, a purely acoustic interpretation of the cause of the phase tracking would posit that the fine structure imposes additional sharp acoustic transitions into the input as opposed to fundamentally altering the nature of the tracking itself.

Instead, it is argued that the intelligibility of the speech signal fundamentally alters the manner in which the signal is being tracked. It is important to note however, that phase tracking *does* occur with only bottom-up (i.e. acoustic) information as seen in previous results and the present study (Howard & Poeppel 2010), but that this phase tracking is fundamentally different in the non-speech case. What remains unclear is the relative contribution of the envelope tracking (Abrams et al. 2008, Deng & Srinivasan 2010) and the evoked response (Howard & Poeppel 2010). While it is tempting to

suggest that in the current results, envelope tracking is occurring in both conditions and the additional information reflects addition acoustic transition from the fine structure (i.e. evoked responses), the qualitative differences in tracking argue against this interpretation. Further work will have to be done to clarify the nature and the origin of the low frequency electrophysiological phase response, and it's relation both to envelope tracking/ evoked responses and which portion of the input signal is being tracked.

Chapter 3: The temporal dynamics of network communication during auditory and speech perception using MEG

3.1 INTRODUCTION

Cognition, broadly construed, is carried out via the interplay between localized computations that take place in discrete brain areas and the integration of these computations between areas (Varela et al. 01, Buzsáki & Draguhn 2004). While there has historically been a debate regarding the exact nature and scope of these localized computations (Fodor 1983), the existence of specific brain areas for specific functions is hardly controversial (Van Essen & Maunsell 1983, Hickok & Poeppel 2000, 2004, 2007). What is unclear however, is the manner in which these discrete computations are integrated both in time and space.

Studies of brain networks have been carried out on various spatial scales and in various models – both human and non-human, from collections of single neurons to large cortical areas (Pesaran 2008, Palva et al. 2010). Recent work using fMRI has provided evidence for networks that demonstrate correlation between areas even during inactivity (Biswal et al. 1997,Raichle et al. 2001, Raichle & Mintun 2006, Fox et al. 2005). The presence of these correlations suggests that specific brain regions form networks that carry out stereotypical tasks by acting in tandem, and that these correlations exist independently of whether or not an active task is being carried out.

These so-called 'resting networks' are characterized by correlations in the ultra-low frequency (0.001 to 0.1 Hz) of the BOLD response between brain areas; this

approach has revealed at least two characteristic networks (Fox et al. 2005) The first, a task positive network, is most active during specific cognitive tasks and contains the intraparietal sulcus, the inferior parietal lobule, the ventral orbital gyrus, the frontal eye fields (FEF), the interior precentral sulcus, the supplementary motor area (SMA)/ pre-SMA, the dorsal lateral prefrontal cortex (DLPFC), the medial temporal lobe, portions of the insula and the frontal operculum. The second, the default network, is most active during periods in which no task is being performed and may be responsible for self-referential activity (Fox et al. 2005). This network contains more medial areas: the posterior cingulate cortex, lateral parietal areas, the superior frontal cortex, the inferior temporal cortex, the parahippocampal gyurs, and the portions of the cerebellum (Fox et al., 2005, Buckner et al. 2008).

Hemodynamic (e.g. fMRI and PET) methods have excellent spatial localization ability but are lacking in temporal acuity. This poses a problem when attempting to assess temporal dynamics that underlie communication between nodes within a network. Electrophysiological measures such as electroencephalography (EEG) and magnetoencaphalography (MEG) provide much better temporal resolution and can therefore aid in the study of the temporal aspect of network dynamics. Evidence for these networks using electrophysiological techniques is somewhat scarce, but two recent studies (Mantini et al. 2007, de Pasquale et al. 2010, Morillon et al. 2010) found evidence for network connectivity using combinations of EEG, MEG and fMRI.

The first (Mantini et al. 2007) examined specific frequency-band EEG power and correlated the power waveforms within these bands with BOLD activity while subjects were at rest. They delineated six separate resting state networks. The first network

corresponded largely to the default-state network and the second corresponded to the dorsal attention network, a network believed to be important for top-down modulation of attention (Corbetta & Shulman 2002). The third was an occipital-based network that contained visual specific areas as well as areas at the occipital-temporal boundary. The fourth was an auditory-based network that included bilateral superior temporal regions, as well as the postcentral gyrus, and right inferior frontal gyrus. The fifth network was predominately motor-related (premotor, motor, SMA, and medial frontal regions), while the last network contained the medial-ventral prefrontal cortex, the anterior cingulate, the hypothalamus and the cerebellum and corresponded to a network responsible for self-referential mental activity.

It is also worth pointing out that in this study, each network had a characteristic frequency composition with each frequency band examined (delta, theta, alpha, beta, and gamma) being correlated with the BOLD signal in each network but at different levels of correlation. While no single network was associated with one band exclusively, there were differences among the different networks in relation to the relative strength of frequency contribution, with networks one and two receiving heavy contributions from alpha and beta, networks three and four (the auditory network) more associated with all frequency bands examined (except for gamma), and networks five and six more associated with gamma and beta (with a contribution to network five from alpha as well).

Taken together, there is evidence that cognition, broadly construed, at least in part, is carried out via the interplay between distinct brain regions that form specialized cognitive networks (Bressler & Menon 2010). These networks display a relatively high degree of encapsulation and occur at characteristic timescales that are different for each

network. This suggests that there are distributed neural networks that underlie specific cognitive tasks.

One system that has yet to receive full consideration in the study of network dynamics is the auditory/speech perception system. It was one of the first systems to be mapped out as a network of connected discrete brain areas. The Wernicke/ Lichtheim model (Lichtheim 1885) posited three separate modules (production, perception, and conceptual) linked together, that composed speech perception and production. Damage to anyone of the modules or the connections between them led to distinct characteristic deficits. For instance, damage to the connection between the motor output area (Broca's area) and the perception area (Wernicke's area) was hypothesized to lead to conduction aphasia in which, while both speech production and perception would be carried out without difficulty, repetition of words could not be performed.

More recent models (Hickok and Poeppel 2000, 2004, 2007, Scott and Johnsrude 2003) are laid out in a similar manner, in that the cognitive act in question, in this case language and speech perception, is carried out by the interplay between discrete brain areas that perform specific computations and the composite network formed by these discrete areas. Together, these two elements, discrete computations and network composition, contribute to the general act of speech comprehension/ language understanding.

The Hickok/Poeppel model (2000, 2004, 2007) posits a dual stream architecture of processing in which speech is first processed bilaterally by core auditory areas (dorsal superior temporal gyrus – dSTG) and a phonological network (mid and post superior temporal sulcus –STS) before splitting into a bilateral ventral network for lexical

interfacing in the middle temporal gyrus (MTG bilateral) and a left lateralized dorsal stream that maps speech input onto articulatory representations via the left posterior Sylvian region of the parietal/temporal boundary (area SPT) , premotor cortex, anterior insula and the prefrontal inferior frontal gyrus (pIFG). These two streams also interface with a conceptual network that is believed to be widely distributed across the brain.

Furthermore, this model of speech perception contains a temporal component as well as a spatial one. Incoming speech information is believed to be processed on two distinct 'privileged' time scales of analysis: a fast time scale of analysis that corresponds to the gamma frequency (~25-50 Hz, 20-40 ms) and a slow time scale in the range of the theta frequency (3-8 Hz, 125-333 ms - Poeppel 2001, 2003). The preference for each time scale of analysis differs hemispherically with the right hemisphere preferring to operate on the longer time scale while the left hemisphere prefers the shorter time scale. This hemispheric asymmetry helps explain results that demonstrate hemispheric differences for fast and slow signal modulations (Boemio et al. 2005).

Taken together, this suggests that speech is processed in a spatially specific network of brain regions that process the incoming auditory speech stream on specific time scales. What is unclear however, is twofold: Firstly, do these 'privileged' time scales correspond simply to the preferred scale of analysis of processing (the computations) or does the network itself communicate inter-areally at the same timescales (the network)? While electrophysiological work on network analysis has provided evidence for contributions from specific frequency bands (Mantini et al. 2007), the relationship between these bands and the speech/auditory system is unclear as these studies have been applied at rest.

Secondly, the majority of evidence for the specific computational roles and spatial locations of the areas involved in speech perception come from fine-grained localization (hemodynamic) methods, or aphasic/lesion data (e.g. Price 2009, Caramazza & Zurif 1976). While these methods can isolate specific computational roles for brain regions, they do not speak to communication between these areas (but see Giraud et al. 2007, Morillon et al. 2010). To date, the network aspect of the speech perception network has not been explored.

The present study seeks to characterize cognitive neuronal network activity during auditory and speech perception and determine whether or not the times scales that are salient for speech perception are also important for neuronal communication between brain areas. While previous work has implicated delta, theta, and gamma (Luo & Poeppel 2007, Luo et al. 2010, Howard & Poeppel 2010, Boemio et al. 2005) as being particularly important for speech perception, work on electrophysiological dynamics of brain networks at rest has implicated a much broader range of frequencies (Mantini et al. 2007). It is also unclear if areas associated with models of speech perception are isolated from other canonical networks such as the default network and dorsal attention network (Fox et al. 2005, Corbetta & Shulman 2002).

3.2 METHODS

Subjects

Eight native English speaking subjects (5 male, mean age 28.8) 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) and the University of Maryland institutional review board. All subjects were right-handed as assessed by the Edinburgh Inventory of Handedness (Oldfield 1971). One subjects' data were not included in the analysis due to saturation of channels by external noise.

Stimuli

Three different English sentences were obtained from a public domain internet audio book website (http://librivox.org). Each of the sentences was between 6 and 6.5 seconds (sampling rate of 44.1 kHz) and each was spoken by a different speaker (American English pronunciation, 1 female). 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 sounds levels (~72 dB SPL trials) within 4 separate blocks.

Two other conditions were included in the study. The first involved presenting stimuli that contained the envelope from the original sentences and a random Gaussian noise band carrier. These sentences were constructed by first band-passing the broadband speech signal into two separate frequency bands using a 500 point two-way least squares linear FIR filter, shifted backwards to compensate for phase delays due to the original filtering. The frequency bands were from 80 to 1240 Hz and 1240 Hz to 8820 Hz. The values were taken from a previous paper (Smith et al. 2002) and are thought to reflect the spacing of the cochlear frequency map (Greenwood 1990). The

envelope in each frequency band was extracted via the Hilbert Transform:

$$H(t) = \frac{1}{\pi} p.v \int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau} d\tau \quad (22)$$

and a Gaussian white noise carrier was added to each of the frequency-band envelopes. Each constructed speech envelope and noise carrier was then normalized against the maximum of the power of the corresponding frequency band of the original sentence and then summed together. The combined signal was then normalized again against the overall power of the original sentence. These stimuli contained information corresponding to the original envelope of the signal but were entirely unintelligible.

A third condition was simply silence presented for 6.5 seconds. Each condition contained 32 trials for a total of 224 trials (32 x 3 Sentences + 32 x 3 envelope/noise carriers + 32 x silent trials). The order of conditions was randomized within each block, with a randomized inter-stimulus interval (ISI) between 800 and 1200 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. There was also no overt task. This was done to restrict the interpretation of results to speech perception in its purest form and to remove any results due to either motor responses (button press) or anticipation of a motor response.

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 ms and 1000 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. After the competition of the subject recording, approximately 200 seconds of empty room data was recorded. This data was used to compute the noise covariance matrix for the Minimum Norm Estimation (MNE) Source localization (Hämäläinen et al. 1994).

MEG Recordings

Magnetoencephalography (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 1000 Hz, a notch filter at 60 Hz (to remove line noise), a 500 Hz on-line analog low pass filter, and no high-pass filter. Each subject's head position was assessed via five coils attached to anatomical landmarks both before and after the experiment to ensure that head movement was minimal. Headshape data were digitized using a three-dimensional digitizer (Polhemus).

Data Analysis

Signal Processing

Both the active data set and the empty room data were first de-noised using a time-shift Principled Component Analysis (tsPCA - de Cheveigné & Simon 2007). Ocular and cardiac artifacts were then removed using an Independent Component Analysis (ICA) algorithm (FastICA). Each of these data sets were high passed at 0.5 Hz and low passed at 100 Hz and converted to Native Neuromag format along with the fiducial marker measurement data via a conversion script.

Source reconstruction of neural sources was performed using the Minimum Norm Estimation (MNE – Hämäläinen et al. 1994). Briefly, the underlying neural current strengths and the measured MEG signals can be related via a linear transformation:

$$Y=AX+N \tag{23}$$

Where Y is a *m* by *t* matrix (sensors x time points) denoting the sensor space measurements, X is a 3*n* by *t* matrix (3 directions of underlying current x time points), A is a gain matrix (i. e. forward solution), and N is a noise term.

The solution is computed as follows:

$$X^{MNE} = RA^T (ARA^T + \lambda^2 C)^{-1} Y \tag{24}$$

Where R is the covariance matrix of the sensor data, C is the covariance matrix of the noise data, and λ^2 is a regularization parameter. The computational version of the equation is as follows:

$$X^{MNE} = R\tilde{A}^T (\tilde{A}R\tilde{A}^T + \lambda^2 I)^{-1} \tilde{Y} \quad (25)$$

Where R is equal to the covariance matrix of the sensor data, λ^2 is a regularization parameter and $\tilde{Y}$ and $\tilde{A}$ are spatially whitened versions of the data and gain matrices:

$$\tilde{Y} = C^{-1/2}Y \quad (26)$$

$$\tilde{A} = C^{-1/2}A \quad (27)$$

With C representing the noise covariance matrix. Whereas in the formal equation for calculation, the regularization parameter is applied to the noise covariance matrix, in the pre-whitened version, C is replaced by I, an identity matrix as in this case, $C^{-1/2}C = I$.

Each subject's structural MRI was reconstructed using the FreeSurfer suite to produce a 3D image of their MRI. This was used to localize neural activity onto the brain. The source space was setup such that each subjects' brain surface contained approximately 20480 'triangles' of localized dipoles. Due to computational constraints, this value was down-sampled using a triangulation procedure that recursively subdivides the inflated spherical surface into icosahedrons and then subdivides the number of triangles (sides) of these icosahedrons by a factor of four. This produces a source space

with 2562 sources per hemisphere, with an average source spacing of approximately 6.2 mm and a surface area of 39 mm^2.

The forward solution was then computed using this information as well as the boundary element model (BEM) information for the computed for a single compartment (homogenous) model for MEG data only. This reduced model has been shown to be effective for MEG (Huang et al. 1999).

The inverse operator was then computed using the forward solution as well as the noise covariance matrix computed for the empty room data. A depth weighting of 0.8 and a regularization parameter, λ^2, of 0.1 were used. The depth weighting function compensates for the superficial bias inherent in the MNE approach to source localization by adjusting the source covariance matrix to favor deeper sources. Values between 0.7 and 1 have been shown to minimize localization errors both in terms of depth and location (Lin et al. 2006). The regularization parameter weights the contribution of the covariance of the noise matrix and therefore is inversely related to the squared estimated SNR (Hämäläinen & Ilmoniemi 1994).

The values for each time point within trial in each condition was then transformed using the inverse solution into source space values for each of the 5124 vertices produced by the forward solution. The orientation of the sources were fixed to be normal to cortical surface as the primary source of the MEG signal is thought to originate from postsynaptic potentials of the apical dendrites of large pyramidal cells orientated perpendicular to the cortical surface (Hämäläinen et al. 1993).

Phase Locking Values

For each subject and condition, the raw source localized signal was first decimate by a factor of four to reduce computational overhead and then band passed into five frequencies of interest: Delta 1 – 3 Hz, Theta 4 -8 Hz, Alpha 9 – 13 Hz, Beta 14 – 24 Hz, and Gamma 25 – 50 Hz. For each band, subject, and condition, a two-ways least-squares linear FIR, shifted backwards in time to compensate for phase delays due to filtering was utilized. A 125 point filter was used for each band except for delta in which an 204 point filter was used to produce adequate frequency isolation due to narrow bandwidth.

Once band-passed, the phase was extracted from the Hilbert transform (equation 20):

$$\theta(x(t)) = \arctan\frac{(H(t))}{(x(t))} \tag{28}$$

Phase locking values were computed for each trial, vertex pair, condition and subject (Lachaux et al. 1999) using the following formula:

$$\mathrm{PLV} = \frac{1}{N}\left|\sum_{n=1}^{N} e^{i(\theta_1(t)-\theta_2(t))}\right| \tag{29}$$

Where $\theta_1(t)$ is the phase for signal 1 at time point *t* and $\theta_2(t)$ is the phase for signal 2 at time point *t*. Computationally, this was calculated as:

$$\text{PLV} = \frac{1}{N}\left|\sum_{n=1}^{N}(e^{i(\theta(f(t)))^T}e^{-i(\theta(f(t)))})\right| \tag{30}$$

Where θ(f((t))) represents angle of the filtered signal for each vertex. The PLV for each vertex combination for each frequency band, condition and subject was then averaged across trials.

Volume Conductance

A common problem when computing phase synchrony or any coherence measure using electrophysiological measures is the issue of volume conductance (Schoffelen & Gross 2009). Because both multiple sensors and localized source vertices can pick up portions of a single underlying neural source, a higher phase locking or coherence value can reflect either two separate underlying sources that are phase locked or coherent with one another or simply a single underlying source that is trivially coherent/phase locked with itself.

PLVs have the benefit that by computing the angle of the phase difference instead of the absolute value of the index, a preferred phase of synchrony between two separate sources can be revealed. Therefore, with this issue in mind, PLVs that had a preferred phase lag of either 0, +/- π were discarded. Since values are seldom at the exact point of these values, an error term was added to compensate for inaccuracies due to

rounding/noise such that values that were within 1 degree of the applicable band-passed signal above/below 0 or away from +/- π were also discarded. The rational behind this approach stems from the quasistatic approximation that underlies the generation of the MEG signal itself (Hämäläinen et al. 1993): Since a single source will take equal time to reach any given sensor/ reconstructed vertex, then a 'false-positive' produced by a high PLV that is in fact indicative of a single underlying neuronal source will have a preferred phase lag of 0 or +π/π.

This approach is similar to the Phase lag Index (PLI – Stam et al. 2007) in that it omits values centered around 0 and +/- π however the present approach maintains the actual index values themselves from the remaining 'true' values whereas PLI computes a sign function of the asymmetry of difference between signals. In this latter case, the index reflects not the strength of phase locking but rather the characterization of the peak of the distribution of phase lags. Put another way, a weakly coupled pair of signals will have the same PLI value as a strongly coupled pair if the distribution of the asymmetry is the same. The present study removes spurious values while maintaining a measure of the strength of phase locking.

Spurious effects due to the forward and inverse solution

While spurious effects due to volume conductance can be minimized using the above mentioned procedure, there can still be errors due to residual bias in the noise correlation matrix (which normalizes the inverse solution) or in the forward solution itself. These issues were accounted by construction a Gaussian white noise source

measurement file and performing the inverse solution using the same noise covariance matrix and forward solution as for the real data. For each subject, Gaussian white noise was simulated for each source vertex with a length equal to the total number of time points in a given condition (6 seconds x 1000 samples/ second x 32 = 192 000 time points). The amplitude of the noise was normalized to the standard deviation of the 'rest' condition data for the corresponding subject. This noise file was then preprocessed in the same manner as the real data (see above), creating five filtered versions of the noise corresponding to the filtered ranges of interest for each subject. Phase locking values and their corresponding matrices were computed as described above. This approach is similar to previous work (David et al. 2002, Palva et al. 2010) but differs in reference to the former in that the noise used was randomly generated as opposed to a shuffling in time and space of the original data as well as isolating the utilization of the noise specifically to rule out spurious connections as opposed to generating a generative distribution as an approximation of the statistical null hypothesis.

Statistical Significance

To test for statistical significance, a surrogate distribution of time-shifted PLVs was created. This was done by splitting a time series of data corresponding to one trial into two halves that varied between +/- of a third of the duration of the trial and then reversing the halves and computing the PLV between the time shifted signal and a corresponding 'normal' signal. This was performed for each condition and frequency in

each subject for a random sampling of 100 000 vertex pairs. This number was found to be the lowest number of iterations that did not under or over-estimate the variability of the surrogate distribution. The actual PLVs for each subject, frequency band, and condition were then statistically compared to this surrogate distribution of values. This had the added effect of controlling for spurious changes in PLVs due to overall power increases (Schoffelen & Gross 2009) as each surrogate data's mean and standard deviation would reflect the specific intrinsic noisiness of the data. A Bonferonni correction was utilized with a corrected alpha of 0.05 so that the minimum normalized value (z-score) of the surrogate data to reach significance was 6.64. Values below this threshold for each frequency band, condition, and subjects were removed.

Brain Space Mapping

Since each source reconstruction is unique to each individual brain, each subject's brain reconstruction was converted to labeled brain regions using an automatic cortical parcellation algorithm based on sulci and gyri (Destrieux et al. 2010). Labels that corresponded to non-cortical areas were removed and to correct for biases due to inhomogeneous label size, large labels were split and small labels were joined to produce 164 labels per subject brain reconstruction (82 per hemisphere). The table of labels can be seen in Table 1.The PLV for each label was computed from the average each of the vertices within that label that met the statistical criterion for significance (see above).

Brain Area	Abbreviation		
Fronto-Marginal Gyrus	FM	Heschl's Gyrus	HG
Posterior Occipital Gyrus and Sulcus	OGSp	Posterior Superior Temporal Gyrus	STGp
Anterior Occipital Gyrus and Sulcus	OGSa	Anterior Superior Temporal Gyrus	STGa
Paracentral Lobule and Sulcus	PLS	Planum Polare	PP
Subcentral Gyrus and Sulcus	gsSUB	Posterior Planum Temporale	PTp
Anterior Cingulate Gyrus	CINGa	Anterior Planum Temporale	PTa
Mid Anterior Cingulate Gyrus	CINGma	Posterior Inferior Temporal Gyrus	IFGp
Posterior Mid-Post Cingulate Gyrus	CINGpmp	Anterior Inferior Temporal Gyrus	IFGa
Anterior Mid-Post Cingulate Gyrus	CINGamp	Posterior Middle Temporal Gyrus	MTGp
Posterior dorsal Cingulate Gyrus	CINGpd	Anterior Middle Temporal Gyrus	MTGa
Cuneus	CUN	Occipital Pole	OP
Inferior Frontal Gyrus – Pars Operculum	IFGo	Temporal Pole	TP
Inferior Frontal Gyrus – Pars Triangularis	IFGt	Posterior Calcarine Fissure	CALCp
Posterior Middle Frontal Gyrus	gFRONTmidp	Anterior Calcarine Fissue	CALCa
Anterior Middle Frontal Gyrus	gFRONTmida	Posterior Central Sulcus	sCENTp
Posterior Superior Frontal Gyrus	gFRONTsupa	Anterior Central Sulcus	sCENTa
Mid-Posterior Superior Frontal Gyrus	gFRONTsupmp	Cingulate Marginalis	sCINGM
Mid-Anterior Superior Frontal Gyrus	gFRONTsupma	Inferior Circular Insula	CINSi
Anterior Superior Frontal Gyrus	gFRONTsupa	Middle Frontal Sulcus	sFRONTmid
Short Insular Gyri	CINSinf	Posterior Superior Circular Insula	CINSsupp
Posterior Middle Occipital Gyrus	gMOp	Anterior Superior Circular Insula	CINSsupa
Anterior Middle Occipital Gyrus	gMOa	Posterior Inferior Frontal Sulcus	sFRONTinfp
Superior Occipital Gyrus	gSO	Anterior Inferior Frontal Sulcus	sFRONTinfp
Lateral-Occipital-Temporal Gyrus	gLOT	Posterior Frontal Sulcus	FSp
Posterior Lingual Gyrus	gLINGp	Anterior Frontal Sulcus	FSa
Anterior Lingual Gyrus	gLINGa	Posterior Intraparietal Sulcus	ISp
Parahippocampal Gyrus	gPARAH	Anterior Intraparietal Sulcus	ISa
Posterior Orbital Gyrus	gORBp	Posterior Lingual Sulcus	sLINGp
Anterior Orbital Gyrus	gORBa	Anterior Lingual Sulcus	sLINGa
Posterior Angular Gyurs	gANGp	Orbital Sulcus	sORBs
Anterior Angular Gyrus	gANGa	Occipital-Parietal Sulcus	OPSULC
Posterior Supramarginal Gyrus	gSUPRAp	Posterior Postcentral Sulcus	sPOSTCENTp
Anterior Supramarginal Gyrus	gSUPRAa	Anterior Postcentral Sulcus	sPOSTCENTa
Posterior Parietal Lobule	pPL	Inferior Precentral Sulcus	sPRECENTinf
Anterior Parietal Lobule	aPL	Superior Precentral Sulcus	sPRCENTsup
Posterior Postcentral Gyrus	gPOSTCENTp	Suborbital Sulcus	sSUBORBs
Anterior Postcentral Gyrus	gPOSTCENTa	Subparietal Sulcus	sSUBP
Posterior Precentral Gyrus	gPRECENTp	Posterior Superior Temporal	STSp

		Sulcus	
Anterior Precentral Gyrus	gPRECENT	Posterior Mid Superior Temporal Sulcus	STSmp
Posterior Precuneus	PRECUNp	Anterior Mid Superior Temporal Sulcus	STSma
Anterior Precuneus	PRECUNa	Anterior Temporal Sulcus	STSa

Table 1: Brain Labels. After the parcellation algorithm, small brain labels were joined and large ones split to create 82 labels per hemisphere. Note that the labels listed here occur in both hemispheres. The abbreviations are the same as used in the figures.

Remaining values that centered on 0 and +/- π were removed (see above). Residual bias due to the forward /inverse solution were removed by subtracting the random Gaussian white noise data set from the actual values. This was done for each condition, subject, and frequency band.

Graph Theoretic Analysis

In graph network analysis, a network is composed of a series of vertices linked by edges (Bullmore and Sporns 2009). Each vertex can be considered a node and each edge a connection between two nodes. In this study, each vertex is a brain region (as broken down via the parcellation algorithm listed above) and each edge is a PLV between two edges. The PLV acts as the weight between each vertex. The graphs produced here are therefore said to be weighted as opposed to binary (in which an edge is considered present or not present rather than associated with a value). It is important to note that these edges are unidirectional as causality cannot be inferred from phase (but see Nolte et al. 2008 for an interesting approach). Unlike typical graphs, each graph produced here could have

connections with itself. This is because each label is composed of a number of underlying vertices and it is possible that each of the vertices that compose a label is phase locked to each other, leading to significant PLVS within a label. Within the speech and envelope condition, the PLV values within subjects were collapsed across tokens to facilitate appropriate statistical comparisons with the rest condition.

In order to isolate subsystems of interest, each of the auditory (speech and envelope) connectivity matrices were averaged across subjects within each frequency band and a modularity algorithm was applied to the weighted networks (Blondel et al. 2008) to isolate modules of interest. Briefly, in graph theoretical terms, a module is a group of vertices that contains a high amount of edges between vertices within the module but only a few between other modules (Newman 2006). The modularity algorithm therefore produces groups of vertices that are highly interconnected with each other but have minimal connections with other vertices not within a given module. The modularity index defines the strength of this modularity and it ranges from 0 to 1 with 1 being completely modular and 0 being essentially random (everything connected to everything).

Density was calculated for each graph for each condition (collapsed across tokens in the case of the speech and envelope conditions), subject, and frequency band. Density reflects the number of connections within a graph that exist as a fraction of total possible connections (Bullmore & Sporns 2010). This was also calculated for each module of interest as produced by the modularity algorithm.

3.3 RESULTS

The modularity analysis revealed characteristic auditory/speech modules in three distinct frequency bands: delta 1-3 Hz, theta 4-8 Hz, and gamma 25-50 Hz. The two remaining bands (alpha 9-14 Hz and beta 15-24 Hz) did not show clear modular organization. As can be seen in Figures 19-22, delta and theta composed networks of similar topology with large bilateral network and two smaller lateralized networks. The overall modularity indices for delta and theta were 0.6309 and 0.4627 respectively.

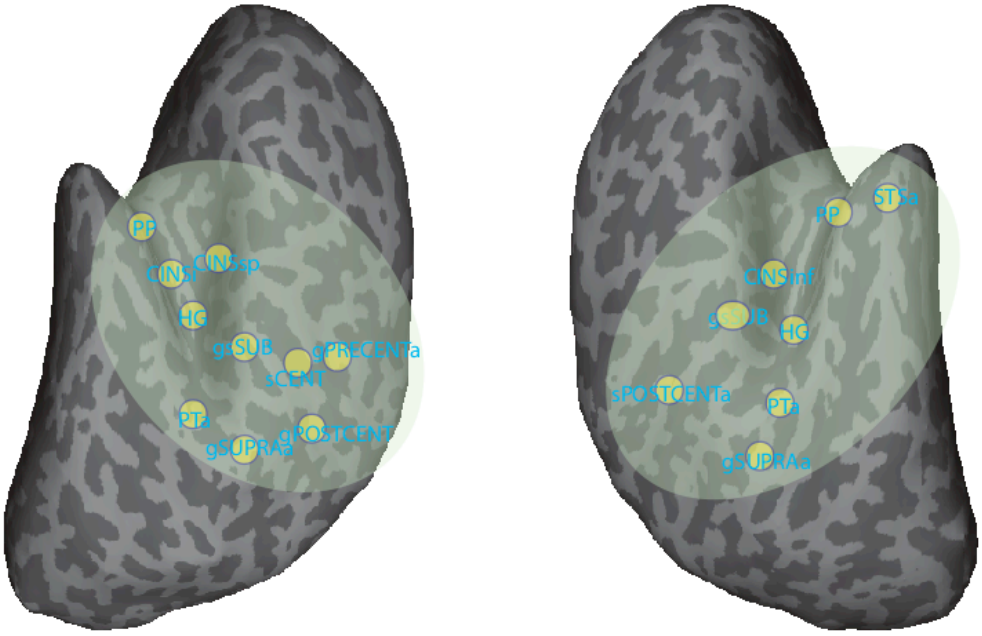

Figure 19: Delta Lateralized Networks. The Delta network lateralized networks are quite sparse, and are composed of only lateral areas. They include temporal,

parietal, motor and insular regions. Each of the left and right networks is quite similar topologically.

3.385, p = 0.009 for envelope versus rest in the bilateral delta network, and $t(6)$ =8.52, p = 0.0001 for speech versus rest and $t(6)$ = 4.18, p = 0.006 for envelope versus rest in the bilateral theta network.

Within the two lateralized networks for both frequencies, there were no significant differences between PLVs in any of the three conditions (speech, envelope and rest). This contrasted with the more extensive bilateral network in which both auditory conditions differed from rest, $t(6)$ =2.78, p = 0.0321 for speech versus rest and $t(6)$ =

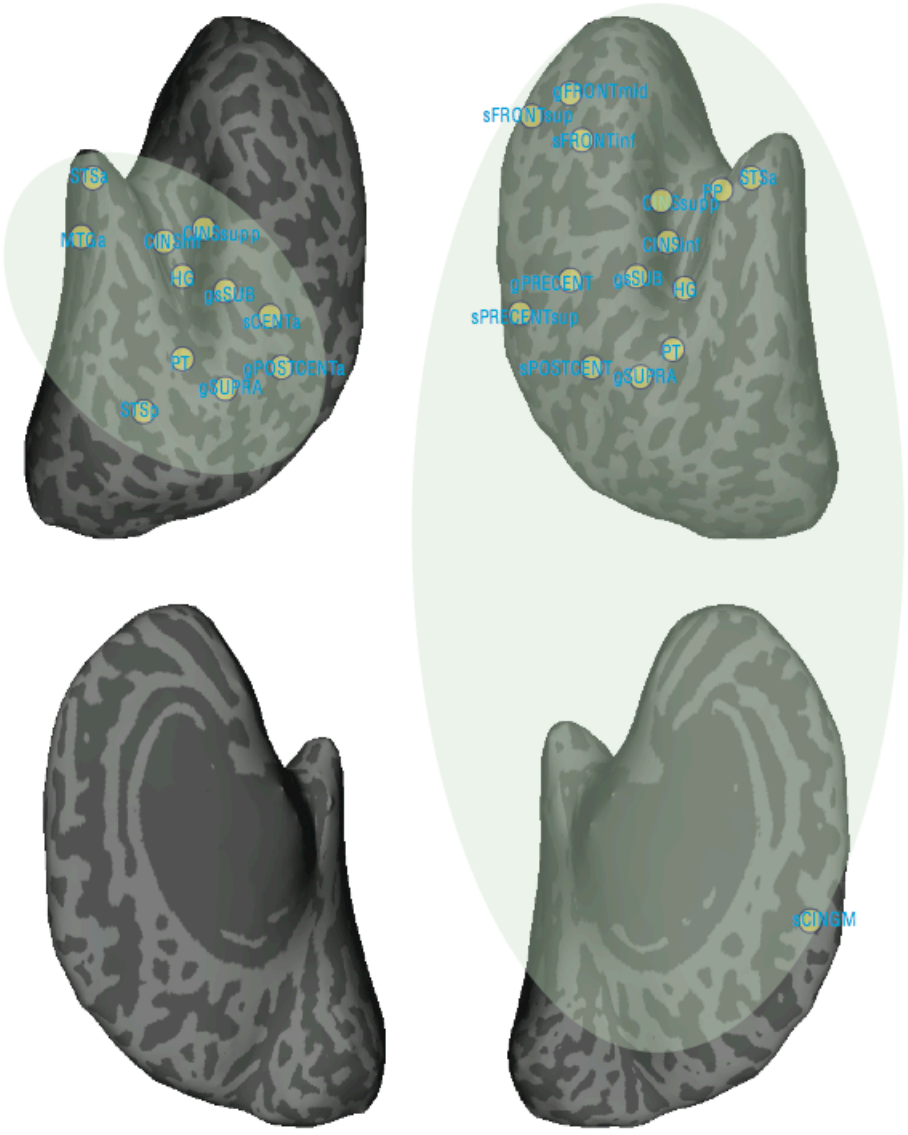

Figure 20: Theta Lateralized Networks. The Theta lateralized networks are more extensive than the Delta Lateralized networks, but share similar topological features. More Frontal areas are present in the right network than in the left, as

well as one medial region: A portion of the Cingulate. The larger scale of the right network suggests a bias towards communication on the timescale of theta within the right hemisphere, although PLVs were not significantly different from rest.

Density for each network was also larger than in the rest condition for both delta and theta (see figure 24a,b.). Each lateralized network showed a larger overall density for each of the two auditory conditions versus rest: for the right delta network $t(6) = 3.98$ $p = 0.0073$ and envelope $t(6) = 3.34$, $p = 0.016$, and for the left delta network $t(6) = 4.26$ $p = 0.0053$ and envelope $t(6) = 4.44$, $p = 0.0044$ For theta, the left lateralized network did not show a difference in density for either comparison while the right lateralized network for speech versus rest $t(6) = 3.16$ $p = 0.02$, and envelope versus rest, $t(6) = 26.13$ $p = 0.0008$, showed an increase in density. The bilateral networks for delta and theta also showed a significant increase in density for the speech versus rest contrast $t(6) = 2.52$ $p = 0.045$ and ($t(6) = 2.78$, $p = 0.028$ for delta and theta respectively, as well as the envelope versus rest contrast, $t(6) = 2.52$ $p = 0.045$ and $t(6) = 3.84$, $p = 0.009$.

The modularity of the gamma band was different from the lower two frequency bands. It had a modularity index of 0.3978 and as can be seen in figure 23, it was composed of two large lateralized networks that spanned from frontal regions to parietal cortex. The left lateralized network demonstrated higher PLVs for auditory versus rest but not between auditory conditions with $t(6) = 5.49$ $p = 0.002$ for speech and $t(6) = 4.4$, $p = 0.005$ for envelope versus rest. The right lateralized network demonstrated similar results with $t(6) = 2.67$ $p = 0.037$ for speech and $t(6) = 2.8$, $p = 0.031$ for envelope versus rest. As can be seen in figure 23c, density for gamma was only significant in the left

lateralized network with speech versus rest at $t(6) = 2.51$ $p = 0.046$ and envelope versus rest at $t(6) = 3.26$, $p = 0.017$.

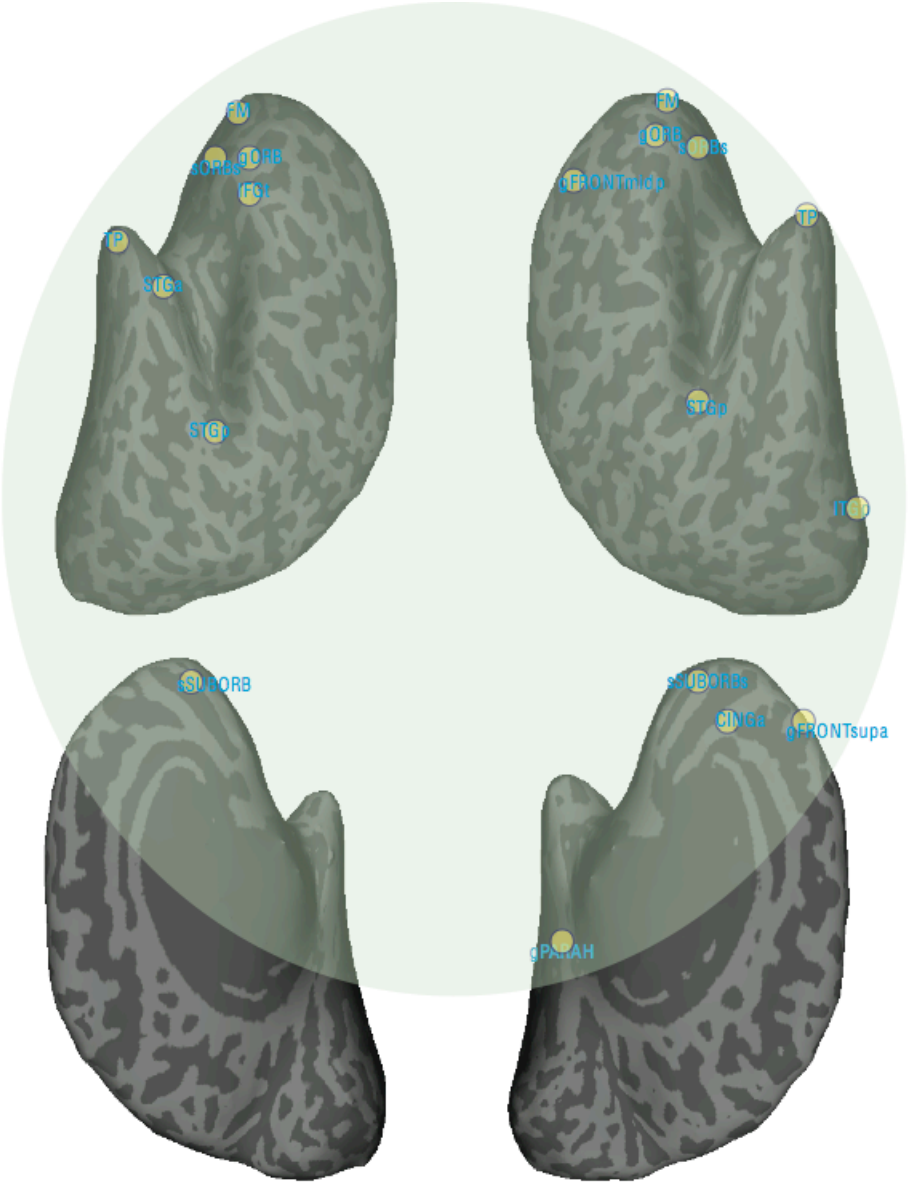

Figure 21: Delta Bilateral Network. The bilateral Delta network is quite sparse, but includes a wider range of areas than either of the lateralized networks. Far more

frontal regions are present, as well as medial areas. The sparse and widespread distribution, combined with the slow scale of integration (delta), suggests that this network integrates large 'chunks' of information over long distances.

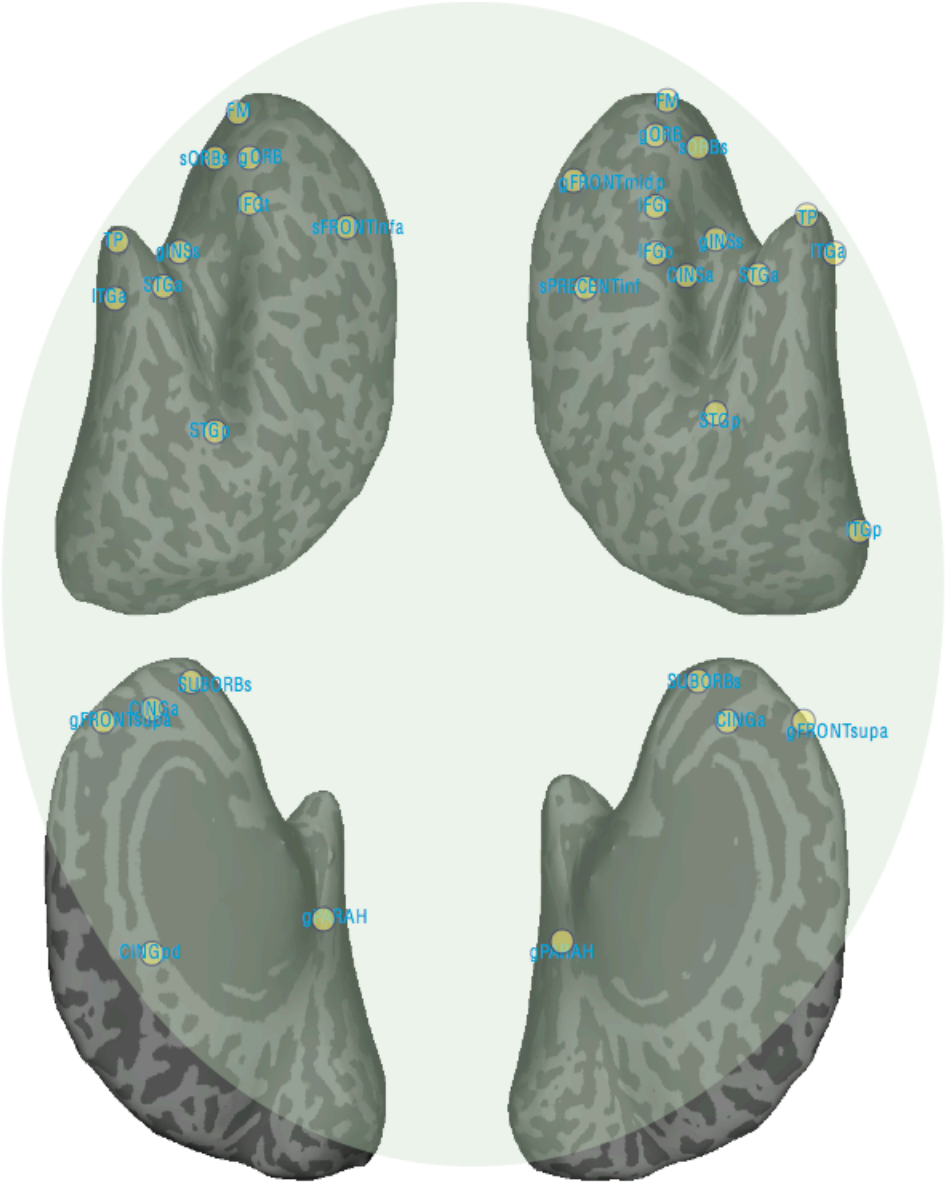

Figure 22: Theta Bilateral Network: Like the Delta bilateral network, the Theta bilateral network is biased towards frontal sources. As in the lateralized networks, there is a bias towards right frontal areas, again suggesting a preference for operation on this timescale within this hemisphere. Pars Triangularis (IFGt) is present bilaterally, although the significance of this is not clear (see text).

The remaining two bands, alpha and beta, failed to demonstrate a clear auditory network both in terms of localization and modularity index. The modularity index for these bands was 0.1149 and 0.2055 respectively. Statistical tests comparing all three conditions in the modules that contained auditory areas that were produced did not reach statistical significance. Mean density values for these networks can be seen in figure 23.

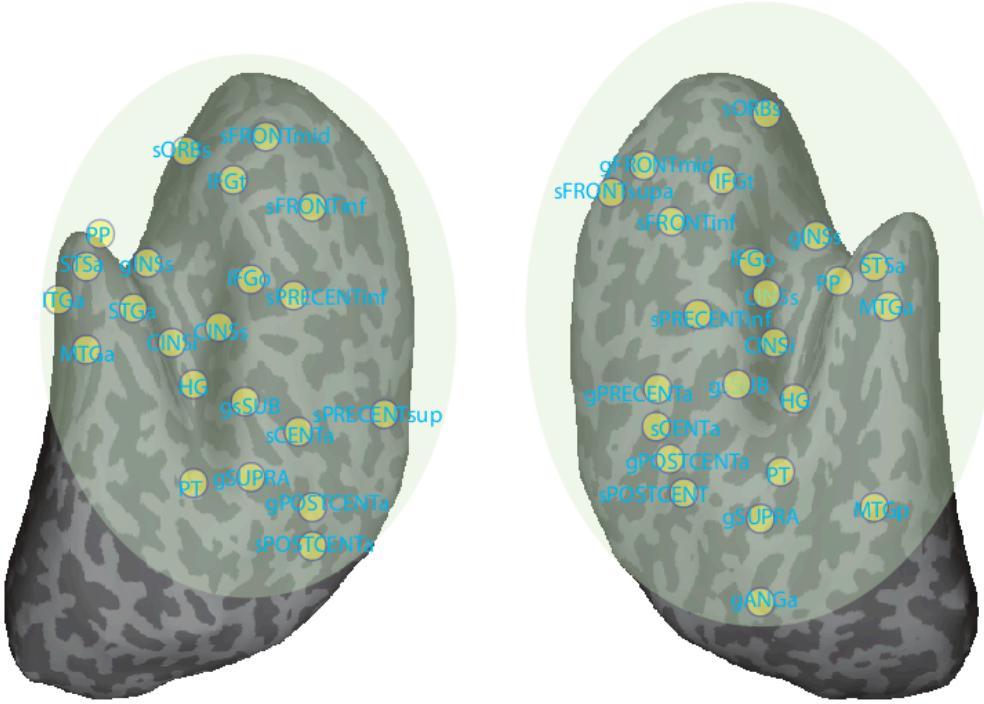

Figure 23. Gamma Lateralized Networks. Each of the Gamma lateralized networks is far more extensive than their low frequency counterparts. Each network includes parietal, temporal, frontal, and portion of the Insula. The extension and density of these networks suggest that operations in this timescale occur largely on smaller distances, but more extensively.

In terms of structure, both the low frequency lateralized networks are quite similar. They are composed of core auditory cortex, some motor regions, portions of the temporal cortex, portions of the parietal lobe, and the portions of the insular cortex. The bilateral networks contain a larger number of areas and are composed mostly of anterior (frontal and prefrontal areas). What is quite interesting, is that there are surprisingly few connections between each of the lateral and bilateral modules suggesting that they are in fact distinct networks (data not shown). This is further supported by their extremely similar structure across frequencies and in each hemisphere (see figures 19-22).

It is also worth mentioning that these modules did not adhere to the overall structure of the Hickok and Poeppel model (2000,2004,2007) which posits a dual stream, ventral/ dorsal split with the dorsal stream connecting core auditory areas parietal areas (area STP) with motor areas, prefrontal areas and the anterior insula while the ventral stream connects core auditory areas with posterior STS, posterior inferior temporal regions, and anterior and posterior middle temporal cortex. The current results suggest that while the core of the dorsal network is maintained in the low frequency lateral netwoks, the prefrontal cortex portion was better represented in the bilateral module.

Bilateral prefrontal components were also seen in the theta bilateral network with left and right pars opercularis and pars triangularis being present in the bilateral theta

network (but not the delta – see figures 21 and 22). The significance of this is unclear as the majority of studies have found that speech perception is lateralized in the prefrontal

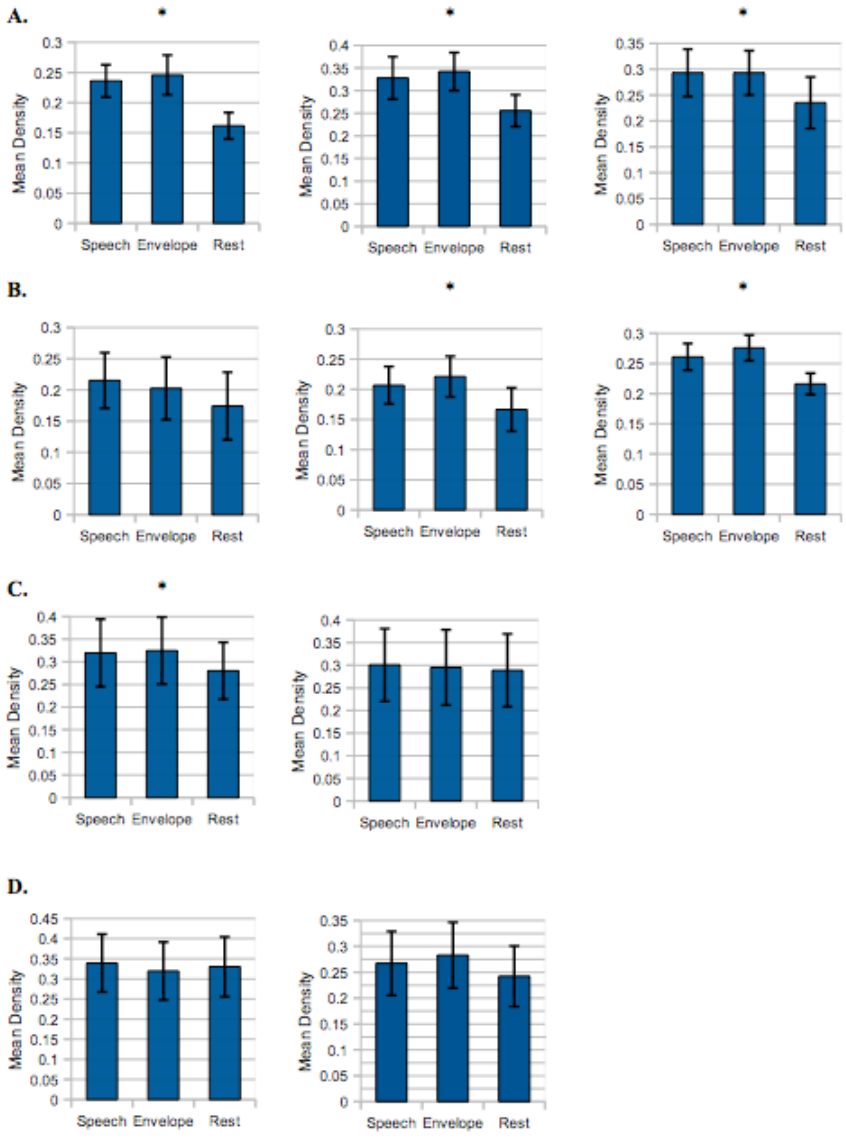

Figure 24. Mean Density Values. A) Mean Density values for the left, right and bilateral Delta Networks. '*' indicates that the two auditory conditions (speech and

envelope) differed significantly from rest. B) Mean Density values for left, right and bilateral Theta Networks. Note that only the right and bilateral Theta networks were significantly different from rest. C) Mean Density Values for the left and right Gamma Network. Only the left network was significantly different from rest. D) Mean density values for an extensive network for Alpha and Beta. Note that the density did not change from rest for either frequency band network.

regions, although a study by Binder et al. (2000) found bilateral activation in a number of conditions including tones, pseudowords, reversed words and speech.

PLVs within each of the lateralized networks did not differ significantly from rest suggesting that these regions are intrinsically coupled. Giraud et al. (2007) found correlations between power in the theta band (define there as 3-6 Hz) within bilateral Heschl's gyrus, the left and right anterior temporal lobe as well as right motor and premotor areas. This suggests that there are intrinsic oscillations within these frequencies during rest. While the lateralized networks posited here are more extensive, the difference in methodologies (phase locking with MEG vs. EEG and BOLD correlation) could at least in part explain the discrepancy.

The PLVs did differ however, between auditory and rest conditions in the bilateral network in both low frequencies. This suggests that coupling within this network is stimulus dependent and may represent a task-specific network as opposed to a more 'intrinsically wired' network as in the lateralized modules.

The overall density changed for the bilateral network (see figure 24), suggesting that more areas within this module were being recruited during auditory/speech perception. Both the right delta and theta networks demonstrated an increase in density

as well, while only the right theta module did not. This might be related to the preference for the right hemisphere to process incoming signals on the time scale of theta (150-333 ms Poeppel 2001,2003).

The gamma networks were composed of mostly the same areas as the low frequency networks, but were more extensive and completely bilateral (figure 23). Once again, connections were found to the right prefrontal cortex as well as the left. PLVs changed between auditory conditions and rest in both networks. As can be seen in figure 23c, density was larger only for the left network, which once again may relate to the preference for the left hemisphere to process information on that timescale (Poeppel 2001, 2003).

The overall size of the network was much larger than either of the lateralized low frequency network and each averaged lateralized gamma module was approximately the size (in terms of nodes) of the averaged bilateral theta module (see figures 22 and 23). It is also worth pointing out that each of the gamma networks was far more dense than either of the low frequency networks with a density of 0.3 (meaning that out of all possible connections within that module, 30 % existed) for the gamma networks and a mean of less than 0.01 for the low frequency networks indicating both that many more regions within each module communicate at the gamma frequency than at the theta or delta. It is also worth pointing out that while density of gamma within each module is higher than the average density of the entire gamma network (i. e. the entire cortex – 0.1428), the theta and delta networks are much more sparse than their overall density, with the overall density for delta and theta being 0.0169 and 0.0450 respectively.

The results combined suggest that distinct networks within the temporal, frontal, prefrontal, and parietal cortex process auditory information. These networks can be characterized both by their preferred timescale of communication and by their topographic extension. Low frequency networks (delta and theta) each compose 3 distinct networks, 2 smaller lateralized networks and 1 more extensive bilateral network that are quite sparse. Gamma operates in two large lateralized networks that are extremely dense. Together, these results suggest that incoming auditory information communicates between specific brain areas on two distinct time scale: a slow rate between 125 and 1000 ms and a much faster rate of between 25 and 40 ms.

3.4 DISCUSSION

The present study sought to characterize the network dynamics that underlie speech and auditory perception. While the speech perception/production system was one of the first brain systems to be mapped out (Lichtheim 1886), the study of the actual network dynamics of the system has as of yet been largely neglected.

These early models of speech perception and production posited a 'house' model in which three areas were linked: an anterior prefrontal 'output' area (prefrontal inferior frontal gyrus - pIFG) linked to both an articulatory motor area (premotor/motor cortex) and a posterior receptive area (area spt). A later model from Hickok and Poeppel (2000,2004,2007) split this system into two distinct subsystems: a lateralized dorsal stream that mapped input onto articulatory representations via the premotor, anterior

insular and prefrontal inferior frontal gyrus and a bilateral ventral stream that corresponds to a lexical interface composed of middle and inferior temporal regions.

Speech itself contains information spanning multiple timescales but a subset of these are thought to be 'privileged': activity in the delta/theta band corresponds to prosodic and syllabic information (respectively) and activity in the gamma band corresponds to the size of a phoneme (Greenberg 2006). Recent work has demonstrated the importance of the former (Luo & Poeppel 2007, Luo et al. 2010, Howard & Poeppel 2010) as well as the latter (Boemio et al. 2005) for auditory and speech perception. What is unclear however, is if communication between brain areas thought to underlie speech perception communicate on time scales that are believed to be particularly salient.

With these issues in mind, the present study used MEG to measure neuronal phase coupling between brain areas and had subjects listen to speech, sentences that were manipulated versions of the same sentences and also listen to nothing (no stimulus). The manipulation utilized for the speech stimuli maintained the overall gross amplitude fluctuations (envelope) while removing the frequency transitions that occur at shorter time scales (fine structure). This creates an auditory signal that is quite similar to speech in a way that is thought to be particularly salient for speech perception and yet makes them entirely unintelligible (Smith et al. 2002).

To examine communication between brain areas, the entire set of data collected was reconstructed in source space using MNE (Hämäläinen & Ilmoniemi 1994) and phase locking values (Lachaux et al. 1999) were computed between each reconstructed location. These values were computed in five distinct frequency ranges: - delta 1-3 Hz,

theta 4-8 Hz, alpha 8-14 Hz, beta 15-24 Hz and gamma 25-50 Hz for each condition: speech, envelope and rest.

To compare across subjects, each of the reconstructed vertices for each condition, frequency band and subject were converted to brain area labels using an automatic parcellation algorithm (Destrieux et al. 2010) and then small labels were combined and large labels separated leaving 164 areas per brain (82 per hemisphere, see table 1).

A modularity algorithm was applied (Blondel et al. 2008) to isolate different subsystems and extract those believed to be auditory/speech related. This was done for each frequency band on the average between the speech and envelope conditions. PLVs and overall density values were calculated for all subjects in each frequency band and conditions in each of the modules identified.

Results suggest that there are distinct subsystems that underline auditory/speech perception and that these modules can be characterized by topology and preferred time scale of inter-areal communication. Low frequency (delta and theta) networks occurred in three distinct subsystems: two lateralized systems that were largely composed of temporal, parietal, insula, and pre/motor areas and a more extensive bilateral network that was mainly composed of pre/frontal, temporal and portions of the cingulate. These networks were highly modularized, with a modularity score of 0.6309 for delta and 0.4627 for theta suggesting that activity within each of these networks is heavily encapsulated.

PLVs differed between rest and auditory conditions only in the bilateral network, suggesting both that the lateralized networks reflect more intrinsic coupling (Giraud et al.

2007, Morillon et al. 2010) and that the bilateral network is functionally activation specifically for auditory/speech perception.

Density differed in all three networks for both frequency bands were significantly different in the speech and envelope condition as compared to rest, except for the left-lateralized theta network. This might be due to a preference for processing on the time scale of theta by the right hemisphere (Poeppel 2001, 2003). All three of the low frequency networks were quite sparse both in absolute terms and in reference to the overall density of the entire brain network at these frequencies.

Coupling on the timescale of gamma demonstrated a different network structure. Two large lateralized networks were present that were composed of a wide range of speech/auditory areas including areas in the pre/frontal cortex, temporal lobes, pre/motor areas, insula, and portions of the parietal lobe. Each of the lateralized networks was largely the same in terms of areas encapsulated. PLVs for both lateralized gamma networks displayed higher PLVs than the rest condition while overall density within each module was only significantly different from rest for the left lateralized network. This could be due to the hypothesized preference of the left hemisphere for processing data on the timescale that corresponds to gamma (Poeppel 2001, 2003). The average density of each of the lateralized modules was quite high, both in absolute terms and relative to the overall gamma network density: 0.3 for the left lateralized network and 0.28 for the right lateralized network versus 0.1428 for the overall network.

Activity in both of the remaining bands – alpha and beta produced very low modularity indices and failed to reveal any clear auditory/speech networks. Furthermore, both PLV and density values for the extremely broad networks did not differ between rest

and the auditory conditions. This suggests that communication on these timescales between brain regions is not particularly salient for auditory/speech perception.

Together, these results suggest that auditory and speech perception is carried out in part through the communication and phase locking between different brain areas. These areas form distinct networks that can be characterized by both their spatial topology and their temporal dynamics. This temporal component corresponds to the salient timescales of speech perception – delta/theta and gamma- suggesting that not only is incoming information preferentially processed on these timescales, but that this selectivity extends to communication between brain areas.

While none of the networks demonstrated a clear one to one correspondence between the dual stream model (Hickok and Poeppel 2000, 2004, 2007), each of the areas contained in the model were present in the modules revealed in this study. There is also no a priori reason to suspect that modules grouped based on their preferred timescales of communication reflect similarities in computational role. It is likely that while processing and communication occurs on distinct and privileged time scales, the fundamental components of these computations are quite different.

The differences in topologies and preferred frequencies of each of the networks combined with the density/PLV results suggest that the lateralized low frequency networks are intrinsic and 'hard-wired' (Giraud et al. 2007). PLVs for the auditory conditions in each of these lateralized networks did not differ from the rest condition suggesting that phase locking occurs between brain regions that compose these modules regardless of external input. The more fronto-centered, bilateral network at the same frequencies displayed higher PLVs for auditory stimuli than for rest suggesting this

network integrates information between hemispheres during auditory/speech perception. The high density and more extensive nature of the bilateral gamma networks suggest that it is responsible for more widespread yet localized communication between different brain areas.

Taken together, these results demonstrate for the first time, active network dynamics of the auditory/speech perception system. Communication between brain areas is carried out on the same privileged time scales as are thought to be important for speech perception itself (Poeppel 2001, 2003). This suggests that the important of both low (delta/theta) and high (gamma) frequency information extends to the role of inter-areal communication.

Bibliography

Abrams, D. A., Nicol, T., Zecker, S., Kraus, N. (2008). Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. *Journal of Neuroscience,* 28(15), pp. 3958-3965.

Adachi, Y., Shimogawara, M., Higuchi, M., Haruta, Y., Ochiai, M. (2002). Reduction of non-periodic environmental magnetic noise in MEG measurement by continuously adjusted least squared method. *IEEE Transactions on Applied Superconductivity,* 11(1), pp. 668-672.

Binder, J. R., Frost, J. A, Hammeke, T. A., Bellgowan, P. S. F., Springer, J. A., Kauffman, J. N., & Possing, E. T. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex,* 20(12), pp. 512-528.

Biswall, B. B., Van Kylen, J., & Hyde, J. S. (1997). Simultaneous assessment of flow and BOLD signals in resting-state functional connectivity maps. *NMR in Medicine,* 10, pp. 165-170.

Blondel, V. D., Guillaume, J-L., Lambiotte, R., & Lefebrve, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment,* 10, pp. 1-11.

Boemio, A., Fromm, S., Braun, A., & Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience* 8, pp. 389-395.

Bressler, S. L., & Menon, V. (2010). Large-scale brain networks in cognition: Emerging methods and principles. *Trends in Cognitive Sciences,* 14(6), pp. 277-290.

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences,* 1124, pp. 1-38.

Bullmore, E., & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience,* 10, pp. 186-198.

Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science,* 304(5679), pp. 1926-1929.

Caramazza, A., Zurif, E. A. (1976). Disassociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain and Language,* 3(4), pp. 572-582.

De Cheveigné, A., & Simon, J. Z. (2007). Denoising based on time-shift PCA. *Journal of Neuroscience Methods,* 165, pp. 297-305.

Cogan, G. B. & Poeppel. D. (in prep.). Mutual Information analysis of neural coding of speech by low frequency MEG phase information**.**

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience,*3(3), pp. 201-215.

Dan, Y., Alonso, J-M., Usrey, W. M., & Reid, R. C. (1998). Coding of visual information by precisely correlated spikes in the lateral geniculate nucleus. *Nature Neuroscience,* 1(6), pp. 501-506.

David, O., Garnero, L., Cosmelli, D., & Varela, F. J. (2002). Estimation of neural dynamics from MEG/EEG cortical density maps: Application to the reconstruction of large-scale cortical synchrony. *IEEE Transactions on Biomedical Engineering,* 49(9), pp. 975-987.

Dell, G. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review,* 93(3), pp. 283-321.

Destrieux, C., Fischl, B., Dale, A., & Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage,* 53, pp. 1-15.

Ding, S., Srinivasan, R., (2010). Semantic and acoustic analysis of speech by functional networks with distinct time scales. *Brain Research*, 1446, pp. 132-144.

Elhilali, M., Fritz, J. B., Klein, D. J., Simon, J. Z., & Shamma, S. A. (2004). Dynamics of precise spike timing in primary auditory cortex. *Journal of Neuroscience,* 24(5), pp. 1159-1172.

Fodor, J. The Modularity of Mind. Cambridge: MIT Press, 1983.

Fox, M. D., Synder, A., Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M., E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences of the United States of America,* 102(27), pp. 9673-9678.

Gandour, J., Dzemidzic, M., Wong, D., Lowe, M., Tong, Y., Hsieh, L., Satthamnuwong, N., & Lurito, J. (2003). Temporal integration of speech prosody is shaped by language experience: An fMRI study. *Brain and Language,* 84, pp. 318-336.

Giraud, A. L., Kleinschmeidt, A., Poeppel, D., Lund, T. E., Frackowiak, R. S. J, & Laufs, H. (2007). Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron,* 56(6), pp. 1127-1134.

Greenberg, S. (2006). “A Multi-Tier Framework for Understanding Spoken Language”. In “Listening to Speech: An Auditory Perspective:, Eds: Greenberg, S. & Ainsworth, W. A. Lawrence Erlbaum Associates, Mahwah, NJ.

Greenberg, S., Hollenback, J. and Ellis, D. (1996). Insights into spoken language gleaned from phonetic transcriptions of the Switchboard Corpus. *In Proceedings of the Fourth International Conference on Spoken Language (ICSLP), Philadelphia, PA*. pp. S24-S27.

Greenwood, D. (1990). A cochlear frequency-position function for several species – 29 years later. *Journal of the Acoustical Society of America,* 87(6), pp. 2592-2605.

Hämäläinen, M. S., & Ilmoniemi, R. J. (1994). Interpreting magnetic fields of the brain: Minimum norm estimates. *Medical & Biological Engineering & Computing,* 32(1), pp. 35-42.

Hämäläinen, M. S., Hari, R., Ilmoneimi, R. J., Knuutila, J., Lounasmaa, O. V. (1993). Magnetoencephalography – theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, 65(2), pp. 413-497.

Huang., M. X., Mosher, J. C., & Leahy, R. M. (1999). A sensor-weighted overlapping-sphere head model and exhaustive head model comparison for MEG. *Physics in Medicine and Biology,* 44, pp. 423-440.

Hickok, G. & Poeppel, D. (2000). Towards a Functional Anatomy of Speech Perception. *Trends in Cognitive Sciences,* 4, pp. 131-138.

Hickok, G. & Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition,* 92, pp. 67-99.

Hickok, G & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Neuroscience Reviews,* 8, pp. 393-402.

Howard, M. F., & Poeppel, D. (2010). Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. *Journal of Neurophysiology.*

Jeong, J., Gore, J. C., & Peterson, B. S. (2001). Mutual information analysis of the EEG in patients with Alzheimer's disease. *Clinical Neurophysiology,* 112, pp. 827-835.

Kayser, C., Montemurro, M. A., Logothetis, N. K., & Panzeri, S. (2009). Spike-phase coding boosts and stabilizes information carried by spatial and temporal spike patterns. *Neuron,* 61(4), pp. 597-608.

Lachaux. J-P., Rodriguez, E., Martinerie, J., & Valera, F. J. (1999). Measuring phase synchrony in brain signals. *Human Brain Mapping*, 8, pp. 194-208.

Lakatos, P., Ankoor, S. S., Knuth, K. H., Ulbert, I., Karmos, G., & Schroeder, C. E. (2005). An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *Journal of Neurophysiology,* 94, pp. 1904-1911.

Laurent, G. (2002). Olfactory network dynamics and the coding of multidimensional signals. *Nature Reviews Neuroscience,* 3, pp. 884-895.

Levelt, W. J. M. (1989). Speaking: From intention to articulation. Cambridge, MA: MIT Press.

Lichtheim, L. (1885). On aphasia. *Brain*, 8, pp. 433-484.

Liégeois-Chauvel, C., Musolino, A., Badier, J. M., Marquis, P., & Chauvel, P. (1994). Evoked potentials from the auditory cortex in man: Evaluation and topography of the middle latency components. *Electroencephalography and Clinical Neurophysiology,* 92(3), pp. 204-214.

Lin, F-H., Witzel, T., Ahlfors, S. P, Shufflebeam, S. M., Belliveau, J. W., & Hämäläinen, M. (2006). Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. *NeuroImage,* 31, pp. 160-171.

Luo, H., Liu, Z., & Poeppel, D. (2010). Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS Biology,* 8(8), pp. 1-13.

Luo, H. & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron,* 54(6), pp. 1001-1010.

Magri, C., Whittingstall, K., Singh, V., Logothetis, N. K., & Panzeri, S. (2009). A toolbox for the fast information analysis of multiple-site LFP, EEG, and spike train recordings. *BioMed Central Neuroscience, 10(81),* pp. 1-24.

Mantini, D., Perrucci, M. G., Del Gratta, C., Romani, G. L., & Corbetta, M. (2007). *Proceedings of the National Academy of Sciences of the United States of America,* 104(32), pp. 13170-13175.

Montemurro, M. A., Rasch, M. J., Murayama, Y., Logothetis, N. K., and Panzeri, S. (2008). Phase-of-firing coding of natural visual stimuli in primary visual cortex. *Current Biology, 18*, pp. 375-380.

Montemurro, M. A., Senatore, R., and Panzeri, S. (2007). Tight data-robust bounds to mutual information combining shuffling and model selection techniques. *Neural Computation 19*, pp. 2913-2957.

Morillon, B., Lehongre, K., Frackowiak, R. S. J, Ducorps, A., Kleinschmeidt, A., Poeppel, D. & Giraud, A. L. (2010). Neurophysiological origin of human brain asymmetry for speech and language. *Proceedings of the National Academy of Sciences of the United States of America,* 107(43), pp. 18688-18693.

Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America, 103(23),* pp. 8577-8582.

Nirenberg, S., Carcieri, S. M., Jacobs, A. L., & Latham, P. E. (2001). Retinal ganglion cells act largely as independent encoders. *Nature, 411,* pp. 698-701.

Nirenberg, S., & Latham, P. E. (2003). Decoding neuronal spike trains: How important are correlations? *Proceedings of the National Academy of Sciences of the United States of America,* 100(12), pp. 7348-7353.

Nolte, G., Ziehe, A., Krämer, N., Popescu, F., & Müller, K-R., (2008). Comparison of Granger causality and phase slope index. *Journal of Machine Learning Research Workshop and Conference Proceedings*

Nolte, G., Ziehe, A., Nikulin, V. V., Schlögl, A., Krämer, N., Brismar, T. & Müller, K-R.(2008). Robustly estimating the flow direction of information in complex physical systems. *Physical Review Letters*, 100, pp 1-4.

Oldfield, R.C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*, pp. 97-113.

Pannenkamp, A., Toepel, U., Alter, K., Hahne, A., & Friederici, A. D. (2005). Prosody-driven sentence processing: An event-related brain potential study. *Journal of Cognitive Neuroscience, 17(3),* pp. 407-421.

Pantev C, Hoke M, Lehnertz K, Lütkenhöner B, Fahrendorf G, & Stöber U.. (1990). Identification of sources of brain neuronal activity with high spatiotemporal resolution through combination of neuromagnetic source localization (NMSL) and magnetic resonance imaging (MRI). *Electroencephalography and Clinical Neurophysiology,* 75(3), pp. 173-184.

Panzeri, S., Senatore, R., Montemurro, M. A., and Petersen, R. S. (2007). Correcting for the sampling bias problem in spike train information measures. *Journal of Neurophysiology 98*, pp. 1064-1072.

Picton, T. W., John, M. S., Dimitrijevic, A., & Purcell, D. (2003). Human auditory steady-state responses. *International Journal of Audiology,* 42(4), 177-219.

Riete, M., Adams, M., Simon, J., Teale, P., Sheeder, J., Richardson, D., & Grabbe, R. (1994). Auditory M100 component 1: Relationship to Heschl's gyri. *Cognitive Brain Research, 2(1),* pp. 13-20.

Rosen, S. (1992). Temporal information in speech: Acoustic, auditory, and linguistic aspects. *Philosophical Transactions: Biological Sciences,* 336(1278), pp. 367-373.

Palva, J. M., Monto, S., Kulashekhar, S., & Palva, S. (2010). Neuronal synchrony reveals working memory networks and predicts individual memory capacity. *Proceedings of the National Academy of Sciences of the United States of America,* 107(16), pp. 7580-7585.

de Pasquale, F., Della Penna, S., Snyder, A. Z, Lewis, C., Mantini, D., Marzetti, L., Belardinelli, P., Ciancetta, L., Pizzella, V., Romani, G. L., & Corbetta, M. (2010). Temporal dynamics of spontaneous MEG activity in brain networks. *Proceedings of the National Academy of Sciences of the United States of America,* 107(13), pp. 6040-6045.

Pesaran, B., Nelson, M. J., & Anderson, R. A. (2008). Free choice activates a decision circuit between frontal and parietal cortex. *Nature,* 453, pp. 406-409.

Poeppel, D. (2001). Pure word deafness and the bilateral processing of the speech code. *Cognitive Science*, 21(5), pp. 679-693.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows:

cerebral lateralization as ‘asymmetric sampling in time’. *Speech Communication*, 41, pp. 245-255.

Price, C. J. (2010). The anatomy of language: A review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences,* 1191, pp. 62-88.

Raichle, M. E., MacLeod, A. M., Snyder, A., Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). *Proceedings of the National Academy of Sciences of the United States of America,* 98(2), pp. 676-682.

Roberts, T. P., Ferrarri, P., Shufflebeam, S. M., & Poeppel, D. (2000). Latency of the auditory evoked neuromagnetic field components: Stimulus dependence and insights towards perception. *Journal of Clinical Neurophysiology,* 17(2), pp. 114-129.

Raichle, M. E., & Mintun, M A., (2006). Brain work and brain imaging. *Annual Review of Neuroscience,* 29, pp. 449-476.

Schneidman, E., Bialek, W., & Berry II, M. J. (2003). Synergy, redundancy, and independence in population codes. *Journal of Neuroscience, 23(37),* pp. 11539-11553.

Schoffelen, J-M., & Gross, J. (2009). Source connectivity analysis with MEG and EEG. *Human Brain Mapping*, 30(6), pp. 1857-1865.

Scott,S.K., Johnsrude,I. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences* 26(2), 100-107.

Shamir, M., Ghitza, O., Epstein, S., & Kopell, N. (2009). Representation of time-varying stimuli by a network exhibiting oscillations on a faster time scale. *PLoS Computational Biology,* 5(5), pp. 1-12.

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

Stam. C. J., Nolte, G., & Daffertshofer, A. (2007). Phase lag index: Assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources. *Human Brain Mapping*, 28(11), pp. 1178-1193.

Strong, S. P., Koberle, R., de Ruyter van Steveninck, R. R., & Bialek, W. (1998). Entropy and information in neural spike trains. *Physical Review Letters, 80(1),* pp. 197-200.

Van Essen, D. C., & Maunsell, J. H. R. (1983). Hierarchical organization and functional streams in visual cortex. *Trends in Neurosciences,* 6, pp. 370-375.

VanRullen, R., & Koch, C. (2003). Is perception discrete or continuous?. *Trends in Cognitive Sciences,* 7(5), pp. 207-213.

Van Veen, B. D., Van Drongelen, W., Yuchtman, M., & Suzuki, A. (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transactions on Biomedical Engineering,* 44, pp. 867-880.

Varela, F., Lachaux, J-P., Rodriguez, E., & Martinerie, J. (2001). *Nature Reviews Neuroscience,* 2, pp. 229-239.

Wang, X. J. (2010). Neurophysiological and computational principles of cortical rhythms in cognition. *Physiological Reviews,* 90(3), pp. 1195-1268.