Causal Cortical Interactions in Speech Processing: Event Related Analysis of Human ECoG

Somayeh Sojoudi[a], Werner Doyle[b], Daniel Friedman[b], Patricia Dugan[b], Orrin Devinsky[b], Thomas Thesen[b]

[a]*Department of Electrical Engineering and Computer Sciences, University of California, Berkeley*
[b]*NYU Medical School, Comprehensive Epilepsy Center*

Abstract

In this work, we study the neural interactions within and across human cortical locations responsive to auditory and/or visual speech stimuli, using electrocorticography (ECoG) signals. The objective is to develop a framework for not only identifying the cortical areas responsive to the events but also understanding how they communicate in the presence of the stimuli. To this end, we first identify the cortical interactions based on the method of partial directed coherence (PDC). The strength of each node in the network is then evaluated and used as a tool for identifying the responsive cortical areas at different frequency bands. These results are then compared with the locations of the brain cortical areas responsive to the audio and/or visual stimuli, using high-gamma power and event related potential (ERP) analyses. We show that the strengths of the nodes associated with the responsive brain areas according to the high-gamma power analysis increase significantly in the high gamma band network. This suggests that the increased high-gamma power of the signal measured by the responsive electrodes would be a consequence of long-rang and short-range interactions rather than merely local interactions. Our study shows that the networks are sparsely connected at all frequencies. The strong long-range interactions occur in alpha and beta frequencies. We also study how different clusters of nodes communicate with one another under different speech conditions. For beta and lower frequencies, we show that the strengths of the connections between different clusters change significantly during the speech relative to the baseline. We also investigate the inward and outward patterns of the connections for each cluster to identify the ones that act as a source of information within the brain network.

1. Significance Statement

There is a large body of literature on the human cortical locations responsive to auditory and/or visual speech stimuli. However, the neural interactions between these areas are not well understood. Using ECoG data, we study cortical interactions across brain regions responsive to auditory and/or visual speech stimuli. Through network analysis, we identify the cortical regions that play important roles in speech processing, the frequencies at which disjoint brain regions communicate with one another, and the patterns of causal relationships between the activities of anatomically segregated brain regions under the three speech conditions. The framework developed in this work

Email addresses: sojoudi@berkeley.edu (Somayeh Sojoudi), wkd1@nyu.edu (Werner Doyle), daniel.friedman@nyumc.org (Daniel Friedman), patricia.dugan@nyumc.org (Patricia Dugan), od4@nyu.edu (Orrin Devinsky), thomas.thesen@med.nyu.edu (Thomas Thesen)

provides a tool for identifying the cortical areas that play critical roles in speech processing as well as evaluating how the patterns of causal interactions vary based on the condition of speech.

2. Introduction

Speech perception is a complex process that integrates both auditory and visual information and involves a wide network of neurons distributed across the frontal, parietal and temporal lobes as well as subcortical structures (Hall et al., 2005; Stein and Meredith, 1993; Wallace et al., 2004). An example of multi-sensory nature of speech perception is the well-known McGurk effect in which the auditory component of one sound paired with the visual component of another sound leads to the perception of a third sound. For instance, the auditory /ba/ paired with the visual /ga/ leads to the perception of /da/ (McGurk and MacDonald, 1976). There is a large body of literature on how cortical areas support the perception of speech and how observable mouth movements profoundly influence speech perception (Summerfield, 1992; Skipper et al., 2007; Jones and Callan, 2003; MacLeod and Summerfield, 1987; Schwartz et al., 2004; Jaekl et al., 2015; Tuomainen et al., 2005).

Understanding the neurophysiology of the audio-visual speech perception has implications for communicative disorders, where the multi-sensory speech integration is either impaired or may confer functional benefits such as dyslexia, autism, aphasia, cochlear implants and other hearing disorders (Redcay, 2008; Alcántara et al., 2004; Čeponienė et al., 2003; Godfrey et al., 1981). To better understand the neurophysiological processes underling the speech perception and other complex cognitive tasks, it is useful to measure how different cortical network components interact with one another in addition to measuring the location and timing of functional brain activation. These interactions can be modeled in the context of *effective connectivity*. The patterns of causal influences among network components under different functional conditions have been referred to as effective connectivity (Friston, 1994; Sporns, 2007).

There is a large body of literature on the role of neural oscillations in information processing. Recent studies suggest that the neurophysiological mechanisms responsible for task-related high gamma (60-200 Hz) activities are fundamental to neuronal functions (Bressler and Kelso, 2001; Engel et al., 2001; Jerbi et al., 2009; Palva et al., 2002; Rodriguez et al., 1999; Kaiser and Lutzenberger, 2005; Ball et al., 2008; Korzeniewska et al., 2011). These high gamma responses are best recorded with intracranial EEG (ECoG). Although the neural mechanisms underlying these responses are not fully known, recent studies have shown that the magnitude and timing of these responses are highly correlated with increased firing rates in local cortical neurons (Ray et al., 2008). To identify the cortical areas engaged during audio, visual and audiovisual speech conditions, we record ECoG in a human subject implanted with subdural electrodes for epilepsy surgery. The high-gamma oscillation powers of the simultaneously recorded local field potentials (LFPs) across multiple electrodes are then estimated. Afterwards, the responsive cortical areas are identified as the areas with higher power during the activity relative to the baseline (a period of time during which no stimulus is presented).

In brain connectivity studies, it is often desirable to understand how the brain communicates as a large network of many brain regions. In general, for three or more simultaneous signals measured from cortical areas, the relation-

ship between any two signals may be direct, mediated by a third signal, or a combination of both. To identify the direct interaction between two disjoint brain areas in the brain network, the effect of the remaining areas needs to be regressed out. In the context of brain effective connectivity, partial directed coherence is a method that addresses this issue and identifies the direct neural interactions in the brain network.

Methods based on multivariate autoregressive models are popular for measuring causal influences between distantly separated cortical sites and, therefore, for modeling the effective connectivity of cortical networks (Astolfi et al., 2007; Cadotte et al., 2008; Chen et al., 2009; Dauwels et al., 2010; Eichler, 2006; Gow Jr and Segawa, 2009; Pereda et al., 2005; Schlögl and Supp, 2006). Granger causality (Ding et al., 2006), directed transfer function (Kamiński et al., 2001), and partial directed coherence (Baccalá and Sameshima, 2001) are autoregressive-based approaches that have been widely used in fMRI and electrophysiological recordings to estimate the strength and directionality of the causal influences at different frequencies and under a variety of normal and pathological conditions (Roebroeck et al., 2005; Kaminski and Blinowska, 1991; Baccalá and Sameshima, 2001; Goebel et al., 2003; Kamiński et al., 2001; Brovelli et al., 2004; Miao et al., 2011).

In this work, we develop a statistical procedure for validating the network models and the analysis of the networks' properties. Some properties of the networks, including the degree and strength of each node, are obtained during the activities and compared with the ones during the baseline period. Moreover, the different brain areas whose electrical activities are recorded via ECoG electrodes are clustered anatomically and the interactions between those communities are studied for each condition.

3. Materials and Methods

3.1. Experiment

In this study, the subjects are presented with audio and video recordings of three basic speech conditions:

- Auditory only (**A**)
- Visual speech mouth movement (**V**)
- Audiovisual word-congruent (**AV**)

The auditory and visual speech stimuli designed for this study include multi-syllabic words with high lexical frequency. Each task involves detecting two target words of high lip-read probability in either modality. Figure 1 shows the onset of visual and auditory stimuli designed for this study.

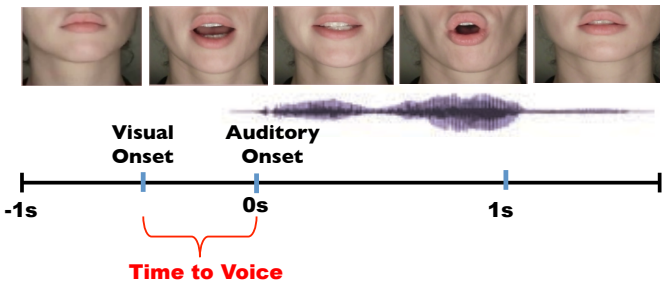


Figure 1: Visual and auditory onsets.

3.2. Data

The data is collected using intracranial EEG recorded from three patients with pharmacologically intractable epilepsy with average or above IQ, normal verbal function and no evidence for functional reorganization. Electrode positions were determined through co-registration of pre implantation MRI with post implantation CT, and cortical gyral anatomy was obtained through surface rendering of the pre-implantation MRI. ECoG signals were recorded

from implanted grid and strips using a sampling rate of 512 Hz. The subdurally implanted 8×8 electrode grid and strips had an inter-electrode distance of 10 mm and an electrode diameter of 4 mm.

Prior to data analysis, the artifacts are detected and removed by visual inspection. The data is then detrended and demeaned. As shown in Figure 1, the visual onset time is set around -300ms, while time 0s indicates auditory onset. To avoid biasing the results of the analysis towards one condition over another, we choose the time interval $[-1\text{s} \quad -0.5\text{s}]$ as the baseline (the period when no stimuli is presented) and consider the time interval $[0 \quad 0.5\text{s}]$ for the analysis of all three conditions. The number of trials is chosen equally for all conditions. The study is performed on three subjects, resulting in similar observations and conclusions across different subjects. For each subject under study and each speech condition, 50 trials are used on average.

3.3. Responsive Electrodes, Power, and ERP Analysis

To identify the locations of the brain areas responsive to the events and their activation times, we study the changes in the time-frequency power spectrum. For the frequencies of 2-180 Hz, Hanning tapers are applied to fixed-time windows of 0.2, leading to a spectral resolution of 5 Hz. All spectral analyses are performed on individual trials before creating averages. To visualize the event-related changes, a normalization with respect to a baseline period is performed. To investigate the modulation of the measured signals with respect to the tasks, in addition to the time-frequency power analysis, the signals measured from individual channels during each individual trial are first low-pass filtered at 40 Hz. The ERP signals are then obtained by taking the average over the filtered signals followed by a baseline correction.

3.4. Brain Effective Connectivity Network

In this work, the causal interactions between different areas of the brain are obtained based on the method of partial directed coherence. This method will be briefly reviewed in this section.

The partial coherence function describes the interaction between two signals $x_i(t)$ and $x_j(t)$ after the influence of all other time series signals is discounted. To compute partial coherence for a system with n signals, a multivariate autoregressive model is first fitted to the recorded signals. The measured signals are treated as a vector output of a multivariate stochastic process, denoted $\mathbf{x}(t)$, and expressed as

$$\mathbf{x}(t) = \sum_{k=1}^{p} A(k)\mathbf{x}(t-k) + \epsilon(t) \tag{1}$$

where the $n \times n$ real-valued matrices $A(k)$ are the regression coefficients, and the n-dimensional stochastic process $\epsilon(t)$ is the error (or residuals) that is assumed to be independently and identically distributed (iid) and serially uncorrelated. The parameters of the autoregressive model are the coefficients $A(k)$ and the $n \times n$ residuals covariance matrix, denoted as $\Sigma_\epsilon \equiv \text{covariance}(\epsilon(t))$, which does not depend on time t by stationarity. The Akaike Information Criterion (Akaike, 1974) can be used to determine the optimal value of the model order p. The matrix coefficients of the model can be calculated by solving the Yule-Walker equations

$$\sum_{k=0}^{p} A(k)\Gamma(k-r) = 0 \tag{2}$$

where $A(0)$ is equal to the $n \times n$ identity matrix and $\Gamma(l)$ is an autocovariance matrix of $\mathbf{x}(t)$ at lag l. Given the autoregressive model (1), the power spectral density matrix at a frequency f, denoted as $S(f)$, can be found using the

equation

$$\mathbf{S}(f) = H(f)\Sigma_\epsilon H(f)^* = \begin{bmatrix} S_{11}(f) & S_{12}(f) & \cdots & S_{1n}(f) \\ S_{21}(f) & S_{22}(f) & \cdots & S_{2n}(f) \\ \vdots & \vdots & \ddots & \vdots \\ S_{n1}(f) & S_{n2}(f) & \cdots & S_{nn}(f) \end{bmatrix} \quad (3)$$

where $*$ is the conjugate operator, and the matrix $H(f)$ is a transfer function defined as

$$H(f) \equiv \left(I - \sum_{k=1}^{p} A(k) e^{-kf\sqrt{-1}} \right)^{-1} \quad (4)$$

The coherence between two arbitrary time series $x_i(t)$ and $x_j(t)$ at a frequency f can then be computed as

$$C_{ij}(f) = \frac{|S_{ij}(f)|^2}{S_i(f)S_j(f)} \quad (5)$$

The notion of partial coherence can be used to regress out the effects of all other signals on the statistical relationship between every two random processes. The partial cross-spectral density function $S_{ij|(\backslash ij)}$ associated with the two random processes $x_i(t)$ and $x_i(t)$ is defined as

$$S_{ij|(\backslash ij)}(f) = S_{ij}(f) - \mathbf{S}_{j(\backslash ij)}(f)\mathbf{S}^{-1}_{(\backslash ij)(\backslash ij)}(f)\mathbf{S}_{(\backslash ij)j}(f) \quad (6)$$

where $(\backslash ij)$ means "all but the i^{th} and j^{th}", and $\mathbf{S}^{-1}_{(\backslash ij)(\backslash ij)}(f)$ is the inverse of the spectral matrix remaining when the i^{th} and j^{th} rows and columns of $\mathbf{S}(f)$ have been removed.

The partial cross-spectral density function can be used to introduce the partial coherence. The partial coherence between two random processes $x_i(t)$ and $x_j(t)$ is defined as

$$\gamma_{ij|(\backslash ij)}(f) = \frac{|S_{ij|(\backslash ij)}(f)|^2}{S_{ii|(\backslash ij)}(f)S_{jj|(\backslash ij)}(f)} \quad (7)$$

A computationally efficient way of obtaining all pairwise partial coherences at all frequencies is via the inverse of the spectral density matrix (3). Define the matrix $G(f) = S^{-1}(f)$ and let $g_{ii}(f)$'s denote the diagonal elements of this matrix. Define also

$$H(f) = -h(f)G(f)h(f) \quad (8)$$

where $h(f)$ is a diagonal matrix whose diagonal elements are $g_{ii}^{-\frac{1}{2}}(f)$'s. Let $H_{ij}(f)$ be the $(i,j)^{th}$ element of $H(f)$. Then, the partial coherence between $x_i(t)$ and $x_j(t)$ satisfies the equation

$$\gamma_{ij}^2 = |H_{ij}(f)|^2 \quad (9)$$

Therefore, all pairwise partial coherence estimates can be computed simultaneously through the inverse of the spectral density matrix. $H_{ij}(f)$ can be factorized as

$$H_{ij}(f) = \frac{\bar{a}_i^H(f)\Sigma^{-1}\bar{a}_j(f)}{\sqrt{(\bar{a}_i^H(f)\Sigma^{-1}\bar{a}_i(f))(\bar{a}_j^H(f)\Sigma^{-1}\bar{a}_j(f))}} \quad (10)$$

where $\bar{a}_k(f)$ is the k^{th} column of the matrix

$$\bar{A}(f) := I - A(f) \quad (11)$$

where the matrix $A(f)$ is defined as

$$A(f) = \sum_{k=1}^{p} A(k) z^{-k}|_{z=e^{-i2\pi f}} \quad (12)$$

with

$$A(k) = \begin{bmatrix} a_{11}(k) & a_{12}(k) & \cdots & a_{1n}(k) \\ a_{21}(k) & S_{22}(k) & \cdots & a_{2n}(k) \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}(k) & a_{n2}(k) & \cdots & a_{nn}(k) \end{bmatrix} \quad (13)$$

The partial directed coherence factor from variable j to

variable i is defined as

$$\pi_{ij}(f) = \frac{\bar{A}_{ij}(f)}{\sqrt{\bar{a}_j^H \Sigma^{-1} \bar{a}_j}(f)} \tag{14}$$

where $\bar{A}_{ij}(f)$ is the $(i,j)^{\text{th}}$ element of the matrix $\bar{A}(f)$, given as

$$\bar{A}_{ij}(f) = \begin{cases} 1 - \sum_{k=1}^{p} a_{ij}(k) e^{-i2\pi fk}, & \text{if } i = j \\ -\sum_{k=1}^{p} a_{ij}(k) e^{-i2\pi fk}, & \text{otherwise} \end{cases} \tag{15}$$

3.5. Statistical Analysis of Effective Connectivity Networks

The objective is to obtain brain effective networks under various speech conditions for different frequency bands. In particular, we study the brain effective networks over theta (5-8 Hz), alpha (9-13 Hz), beta (14-32 Hz), gamma (33-55 Hz), and high gamma (70-170 Hz) frequency bands. In order to assess the significance of any causal interaction in the effective network associated with each frequency band of interest, the following steps are taken. First, the average partial directed coherence value is computed over the desirable frequency band for each individual trial. These values are then entered into a one-tailed student t-test with the mean 0.1 and the 1% significance level. In this work, we consider the connection from i to j significant if the value of PDC from i to j is greater than 0.1, according to Schnider et al. (1989). Finally, a positive false discovery rate (pFDR) analysis is applied to the p-values obtained from the t-test using the procedure described in Storey (2003). The results of the FDR analysis are then used to determine whether the value of the partial directed coherence between two parts of the brain is significant under each of the speech conditions. To identify the direction of the information flow between two brain areas x and y, we compare the values of the partial directed coherence from x to y and vise versa. In this study, we assume that the network is uni-directional. Therefore, the larger component determines the direction of the edge in the brain effective network.

Denote the set of all speech conditions as well as the baseline with $\mathcal{S}$, i.e., $\mathcal{S} = \{\text{baseline, auditory, visual and audio-visual}\}$. Suppose $\mathbf{W}_{ij}^{c1}$ and $\mathbf{W}_{ij}^{c2}$, are the weight sets obtained for two conditions c_1 and c_2, where $c_1, c_2 \in \mathcal{S}$. The values of $\mathbf{W}_{ij}^{c1-c2} = (\mathbf{W}_{ij}^{c1})^2 - (\mathbf{W}_{ij}^{c2})^2$ are entered into a two-tailed one sample t-test with mean equal to zero at the 5% significance level. Finally, an FDR analysis is applied to the p-values obtained from the t-test. If the value of $\mathbf{W}_{ij}^{c1-c2}$ is statistically significant and the weight of the edge connecting nodes i and j in condition c_1 is on average larger than the one in condition c_2, the edge is shown in red; otherwise, it is shown in blue.

For illustration purposes, each node of the graphs obtained here is shown by a circle whose size depends on the *degree* of that node (defined as the number of edges connected to the node). The *strength* of a node i, denoted by S_i, is defined as the sum of the weights of the edges connected to that node. To compare the strengths of the nodes in different graphs, we first form a graph by assuming that all nodes of the network are connected to one another. In other words, we do not threshold the values of the PDC matrix. The main reason for not thresholding the matrices is that the values of PDC are relatively lower in high-gamma frequency band than in lower frequencies. To be consistent across different frequencies, we keep the value of thresholding constant and this may result in networks with no edges in high frequencies. Although in high frequencies, there might be no significant edges (relative

to lower frequencies) between the nodes and therefore, the significance of the nodes cannot be computed. However, as shown later in the paper, although a single edge might not be statistically significant, the aggregated values of PDC's across multiple edges may result in a significant difference in a speech condition.

To find a change in the strength of each node during each speech condition, we form a *strength set* $\mathbf{S_{i_c}} = \{S^1_{i_c}, S^2_{i_c}, \cdots, S^N_{i_c}\}$ for each node i and for all trials under condition $c \in \mathcal{S}$. For every two conditions c_1 and c_2 in $\mathcal{S}$, $\mathbf{S_{i_{c_1}}} - \mathbf{S_{i_{c_2}}}$ is computed based on the weighted graphs. These values are entered into a two-tailed one sample t-test with mean equal to zero at the 10% significance level. An FDR analysis is then applied to the results of the t-test. If the average $\mathbf{S_{i_{c_1}}}$ is larger than the average $\mathbf{S_{i_{c_2}}}$ and statistically significant, the corresponding node in the graph of condition $c1$ is shown in red, and is shown in blue otherwise.

In order to study the differences across larger brain areas rather than individual recording sites, the electrodes are clustered based on their anatomical locations. For a weighted graph, we define the *inter-cluster strength*, denoted by $\mathcal{S}_{I_{km}}$ for every $k, m \in \{1, 2, \cdots, K\}$, as the sum of the weights of the edges that connect clusters k and m. To compare the strengths inside each cluster and between every two clusters $k, m \in \{1, 2, \cdots, K\}$ under different speech conditions, $S^{c^j}_{c_k}$ and $\mathcal{S}^{c^j}_{I_{km}}$ are found for each condition $c \in \mathcal{S}$ and trial $i \in \{1, 2, \cdots, N\}$. For every two conditions $c_1, c_2 \in \mathcal{S}$, the differences $S^{{c_1}^j}_{c_k} - S^{{c_2}^j}_{c_k}$ and $\mathcal{S}^{{c_1}^j}_{I_{km}} - \mathcal{S}^{{c_2}^j}_{I_{km}}$ are also computed. These differences are all entered into a two-tailed t-test with mean equal to zero at the 1% significance level, followed by an FDR correction.

4. Results

The brain effective connectivity networks, corresponding to each of the speech conditions and the baseline period, are represented by networks of nodes and directed edges, where each node represents one recording site and each edge represents the partial directed coherence from one recording site to another. Figure 2 shows the anatomical locations of the subdural electrodes used for ECoG recording in one of the subjects. The numbered yellow disks in Figure 2 show the anatomical locations of the recording sites selected for analysis. The cortical locations of the selected electrodes and their labels are given in Table I.

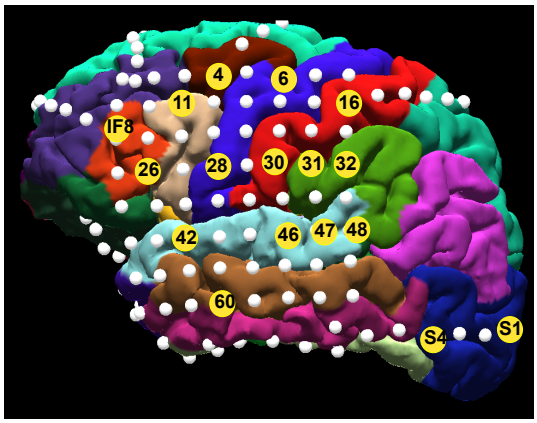


Figure 2: The locations of the selected ECoG electrodes for the subject under study.

In what follows, the approaches discussed earlier will be deployed to model the brain effective connectivity networks during the three speech conditions. These networks will then be used to study and compare how different brain regions communicate with one another under each speech condition.

4.1. Effective Connectivity Networks: Single Subject analysis

Using the method described in Section 3.4, the brain effective connectivity networks are obtained for each of the

Table 1: The cortical locations of the selected ECoG electrodes.

Region of interest	Labels
Caudal middle frontal	4
Precentral	6, 28
Pars opercularis	11
Postcentral	16, 30
Pars triangularis	26, IF8
Supra marginal	31, 32
aSTG	42
mSTG	46
pSTG	47, 48
aMTG	60
Lateral occipital	S1, S4

speech conditions for the five frequency bands of interest. These graphs are shown in Table 2. In the connectivity graphs for the three speech conditions, those nodes in the brain networks that form stronger connections to other regions relative to the baseline are highlighted in red. Similarly, the nodes with reduced connections are shown in blue. It can be seen that, in presence of auditory-only speech stimuli, the strengths of the nodes located in the superior temporal gyrus (electrodes 47 and 48) increase significantly over the theta, alpha and beta bands. In the gamma frequency band, the strengths of the nodes corresponding to the electrodes 46 (mSTG), 47 and 48 (pSTG), 31 (supra marginal), 26 (pars triangularis) increase significantly. In the high gamma frequency bands, the nodes of the brain network located in mSTG, pSTG, aMTG, caudal middle frontal and pars triangularis become significantly stronger under the auditory-only speech condition.

In the auditory-only speech condition, the increased nodal strengths are mostly local and within the auditory cortex for frequencies below the beta band. In gamma and specially high-gamma frequencies, the areas involved in the speech processing spread out to a wider cortical area including the auditory cortex and frontal lobe. In the visual-only speech condition, the strengths of the nodes associated with electrodes 31 and 32 (supra marginal), 30 (postcentral), 6 (precentral), S4 (lateral occipital) and 46 (mSTG) increase significantly. In particular, in gamma and high-gamma frequencies, the nodes located in the visual cortex, motor cortex and wernicke's areas form significantly stronger connections with the rest of the brain. Under the audio-visual speech condition, the nodes that are significantly strong during the auditory-only or visual-only speech condition become also significantly stronger relative to baseline. Under the auditory-visual condition, those nodes that are significantly stronger in the theta frequency network are mostly located in the auditory cortex, motor cortex and the wernicke's area. However, in the alpha, gamma and high-gamma frequencies, the cortical locations of the strong nodes include visual cortex as well.

In all of the three speech conditions, the maximum connectivity occurs in the alpha and beta frequency bands, meaning that the short- and long-range cortical interactions are more significant in these frequencies. According to the graphs of the brain effective networks, indicating the strong nodes in different speech conditions and distinct frequency bands, although in the gamma and high gamma frequency bands, each individual link connecting one node to another may not be strong on its own, the summation of such links connecting an individual node to multiple brain areas can be highly strong relative to the baseline resulting in a node with a significantly increased strength in a speech condition relative to baseline. Since these links include both inputs from other nodes in the

network into a node and output from that nodes to the rest of the network, this could be indicated as how active that node or brain region is in a speech condition.

In addition to the identification of the pattern of connectivity in each speech modality, it is desirable to study the relationship between the significantly strong nodes in the brain effective networks and the responsive electrodes. The responsive electrodes are chosen according to their high-gamma response since the magnitude and timing of high gamma (70-200 Hz) responses are highly correlated with the increased firing rates in local cortical neurons. On the other hand, the strong nodes are identified by comparing the strengths of different nodes, where the strength of a node in a speech condition or during the baseline period is defined as the sum of the weights (PDC value) of the links that connect the node to the rest of the network. To streamline the presentation, we focus our analysis on only the high-gamma frequency band. Comparing the power spectrum of the individual electrodes (nodes) with the strengths of the corresponding nodes reveals that a node whose high-gamma power increases significantly relative to the baseline in a condition would likely show a significant increase in its strength in the high-gamma effective network as well. However, note that the strength of the node is defined by the weights of the causal interactions between different nodes of the network. This indicates that through our network analysis we are able to not only identify the responsive areas to the stimulus but also study how those areas interact with one another in presence of the stimulus. Figures 3-5 show the power spectrum of the responsive electrodes, their ERP signals, the locations of the responsive electrodes on the brain map, and the highlighted strong nodes in high-gamma networks, for the auditory-only, visual-only and audio-visual only speech conditions, respectively.

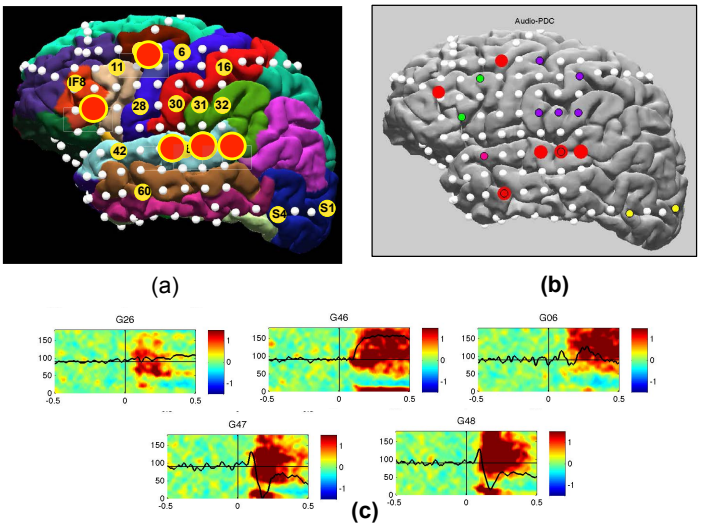


Figure 3: For the auditory-only speech condition, (a) shows the locations of the responsive electrodes which are highlighted in red on the brain map, (b) shows the significantly stronger nodes (highlighted in red) in the high-gamma frequency network, and (c) shows the power spectrum and ERP signals of the responsive electrodes.

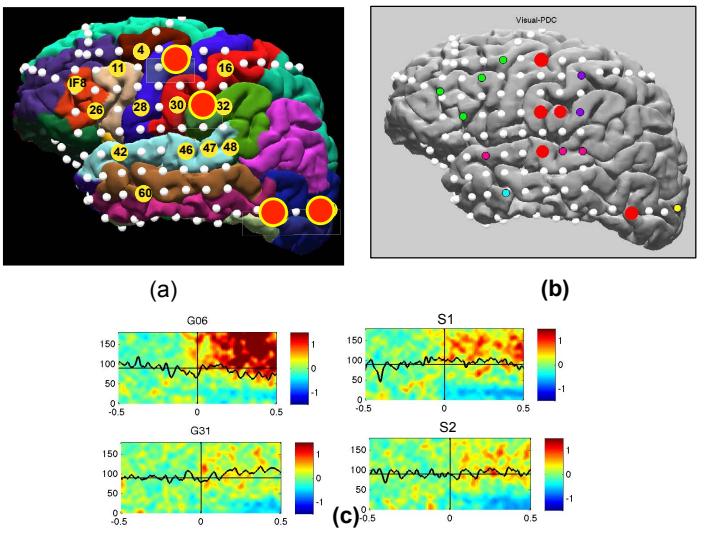
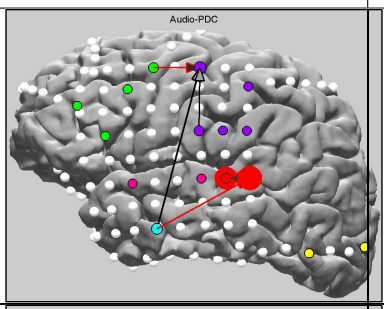
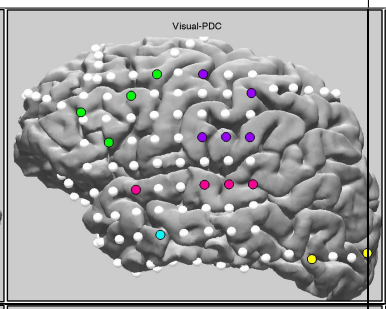
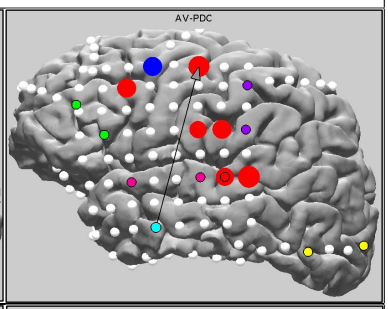
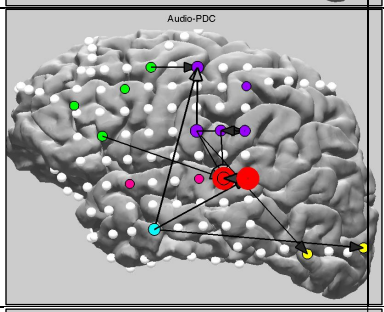
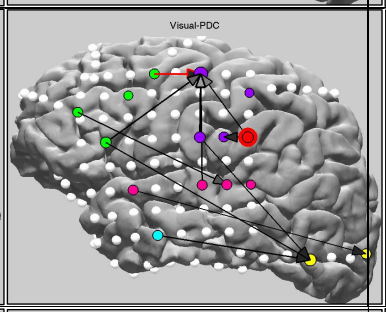
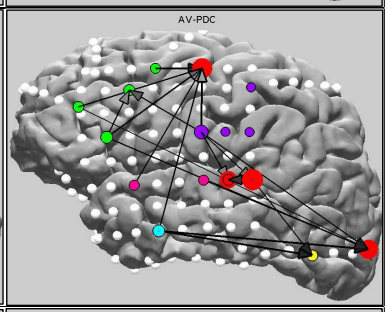
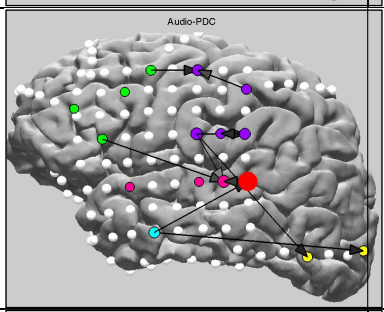
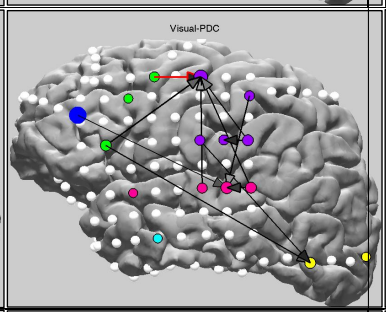
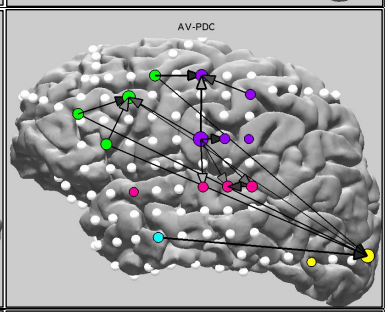
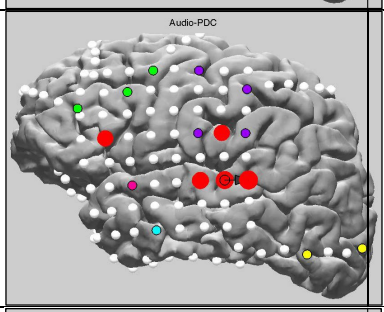
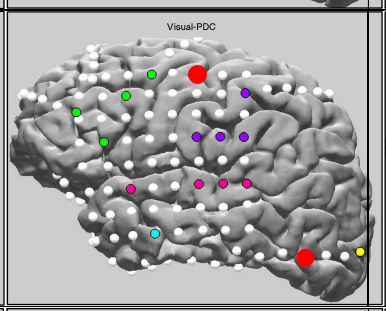
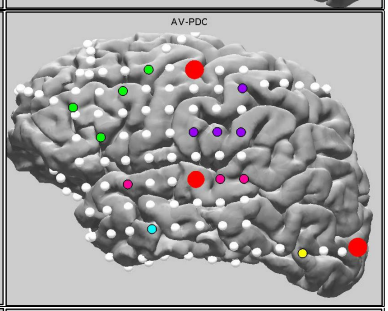
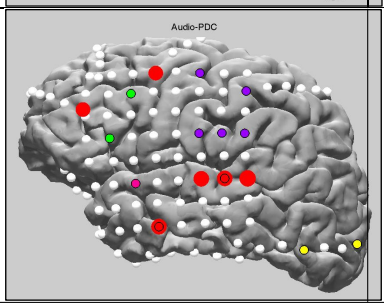
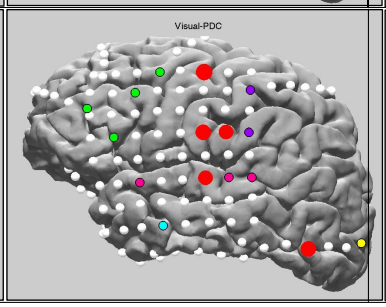
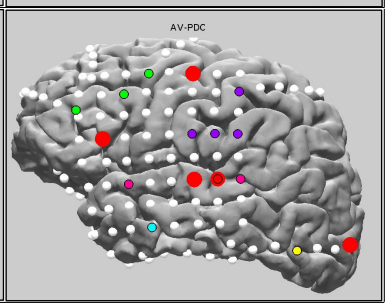


Figure 4: For the visual-only speech condition, (a) shows the locations of the responsive electrodes which are highlighted in red on the brain map, (b) shows the significantly stronger nodes (highlighted in red) in the high-gamma frequency network, and (c) shows the power spectrum and ERP signals of the responsive electrodes.

Since we have kept the threshold constant across different frequency bands to identify the important frequency bands in terms of the significance of causal interactions between different brain areas, this yielded no significant edge in the high gamma frequency networks, as shown in Table 2. On the other hand, our approach for the iden-

Table 2: Brain effective connectivity networks under the three basic speech conditions for different frequency bands.

Freq.	Auditory speech	Visual speech	Audio-visual speech
Theta			
Alpha			
Beta			
Gamma			
High gamma			

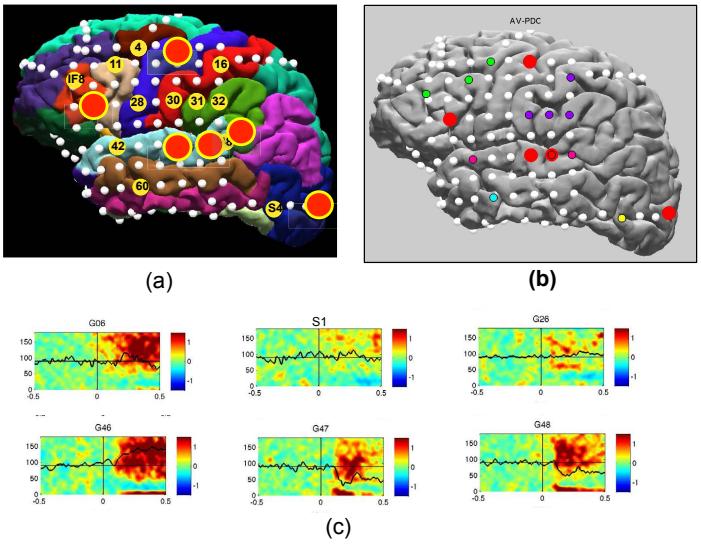


Figure 5: For the audio-visual speech condition, (a) shows the locations of the responsive electrodes which are highlighted in red on the brain map, (b) shows the significantly stronger nodes (highlighted in red) in the high-gamma frequency network, and (c) shows the power spectrum and ERP signals of the responsive electrodes.

tification of those nodes that show a significant increase in their strength in a condition relative to the baseline is based on un-thresholded PDC. Therefore, to identify the edges in the high frequency networks, we have lowered the threshold and the results of this analysis are given in Figure 6. Although these connections may not be sufficiently significant individually, they add up and result in activating a specialized part of the brain to a stimulus. Hence, one can conclude that the increased power in high-gamma frequencies may also be the result of both long- and short-range interactions rather than merely local interactions.

So far, we have studied the nodal interactions within the brain effective networks under the three speech conditions. In order to study the brain neural interactions between larger brain segments, we cluster the nodes of the networks according to their anatomical locations into 5 different clusters, as listed below:

Cluster 1: $C_1 = \{4, 11, 26, IF8\}$. These nodes (electrodes) are located in the prefrontal cortex.

Cluster 2: $C_2 = \{6, 28, 16, 30, 31, 32\}$. These nodes are located in the primary motor cortex, primary somatosensory cortex and wernicke's area.

Cluster 3: $C_3 = \{42, 46, 47, 48\}$. These nodes are located in the superior temporal gyrus.

Cluster 4: $C_4 = 60$. This node is located in the middle temporal gyrus.

Cluster 5: $C_5 = \{S1, S4\}$. These electrodes are located in the occipital lobe.

The nodes within one community or cluster are shown by the same color. The nodes in clusters 1-5 are colored in green, purple, magenta, blue and yellow, respectively. For the subjects under study, the strengths of the connections between clusters are obtained in brain effective networks under the three speech conditions and for each individual trial and the compared statistically, as explained in Subsection 3.5. The anatomical clustering for the three subjects is shown in figure 7.

In this part, we study the inter-cluster strengths in the brain effective networks for the subjects under study during the three speech conditions and also compare them to the strengths of the connections during the baseline period. In these networks, if the connection from one cluster to another is significantly stronger during a speech condition relative to the baseline for at least two subjects, we draw an arrow between their corresponding nodes in the 5-node network. Figure 8 shows the results of this analysis. We see that, in all of the three speech conditions, the frequencies in which the strengths of the connections from one cluster to another increase significantly are theta, alpha and beta frequencies.

By lowering the threshold, we can reveal more edges that may not be highly strong in high frequencies but are still stronger during the activity relative to the baseline.

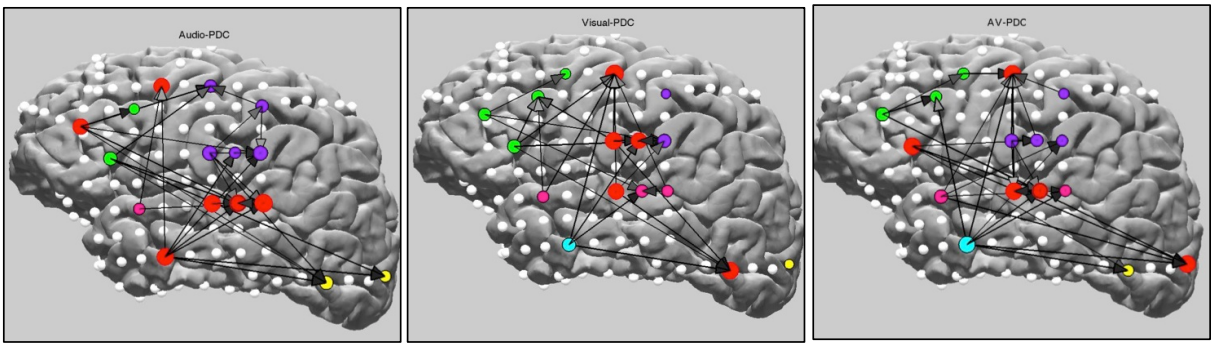


Figure 6: Brain effective connectivity networks under the three basic speech conditions associated with high-gamma frequencies.

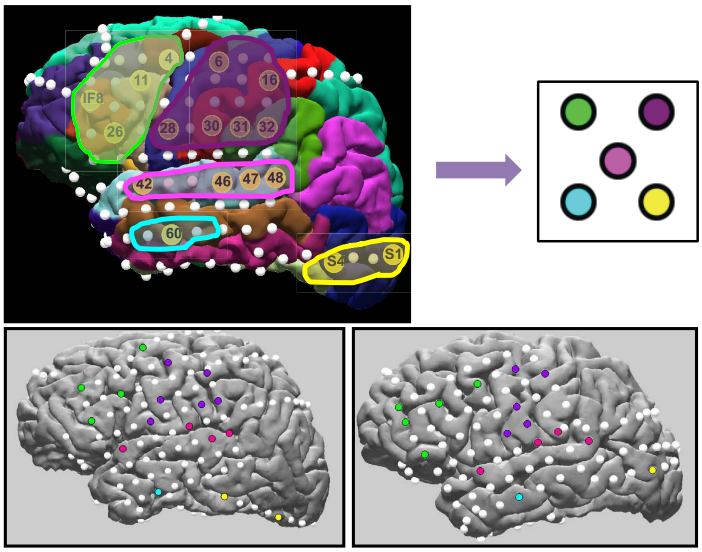


Figure 7: Anatomical clustering for one subject: each cluster is represented by a node in the 5-node network on the right. Similarly, the nodes are clustered based on their anatomical locations into 5 clusters for two other subjects in the bottom panels.

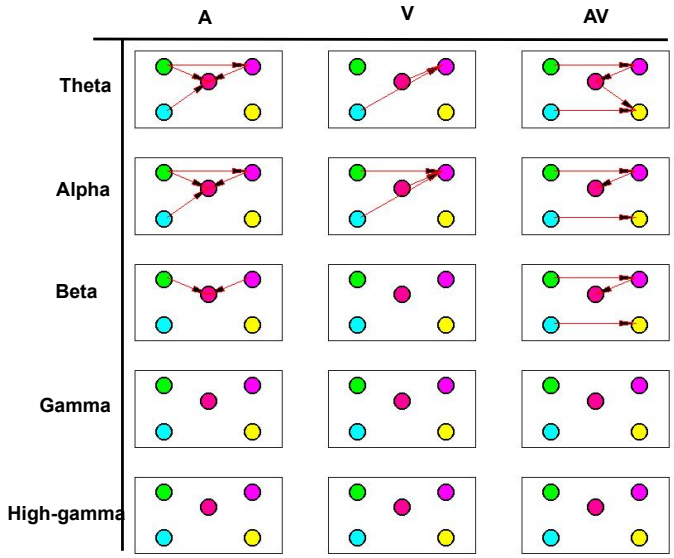


Figure 8: The inter-cluster strengths during the auditory-only, visual-only, and audio-visual speech stimuli are compared to the baseline at different frequency bands.

The results associated with the case of a lower threshold are given in Figure 9. This type of analysis is in particular interesting because it can be used to identify the areas that act as sources of the activity as well as the areas that are mostly influenced in the activity within the brain network (namely, sinks). The results of this study are summarized in Figure 10.

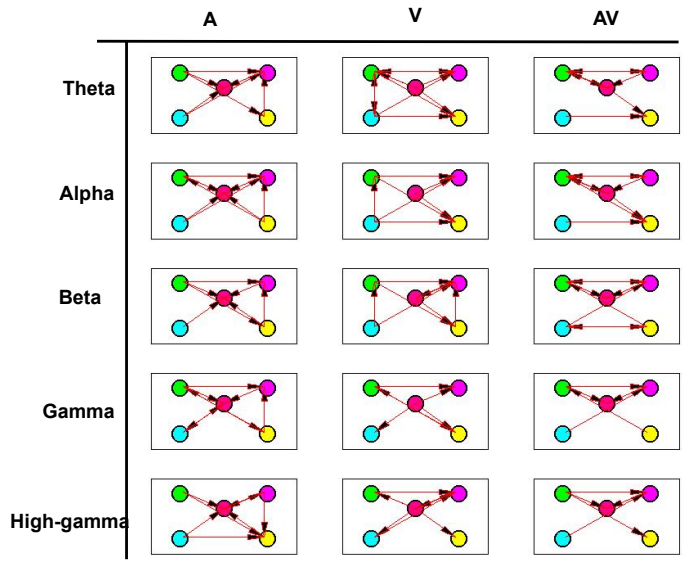


Figure 9: The inter-cluster strengths during the auditory-only, visual-only, and audio-visual speech stimuli are compared to the baseline at different frequency bands, associated with a lower threshold.

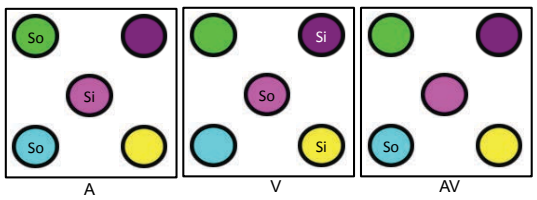


Figure 10: Sources (marked as "So") and sinks (marked as "Si") in the speech conditions.

Figure 10 shows that, in the auditory-only speech condition, cluster 1 (located in the prefrontal cortex) and cluster 4 (located in the middle temporal gyrus) act as sources, whereas the third cluster located in the auditory cortex acts as a sink in the 5-node network. On the other hand, in the visual-only speech, the auditory cortex acts as a source, whereas the visual cortex and the second cluster located in the primary motor cortex, primary somatosensory cortex and wernicke's area are sinks and influenced by the activities in other regions of the brain. In the auditory-visual condition, all areas communicate with one another in both directions with the exception of cluster 3.

In addition to the comparison between each condition and the baseline connectivity, it is desirable to make a comparison between the three speech conditions and the strengths of the edges in the 5-nodes networks. The results of this comparison are shown in Figure 11. In particular, at beta and lower frequencies, some connections become stronger and some other connections become weaker, during the auditory-visual condition compared to both auditory-only and visual-only conditions. Moreover, certain connections under the audio-visual speech stimuli become weaker in one of the two uni-modal conditions and stronger in the other condition. In particular, the connections from the middle temporal gyrus to the primary motor cortex, primary somatosensory cortex and wernicke's area as well as the connections from the prefrontal cortex to the occipital lobe are weaker in audio-visual condition relative to both auditory-only and visual-only conditions in beta and lower frequencies. However, in these frequencies, the connections from the prefrontal cortex to the primary motor cortex, primary somatosensory cortex and wernicke's area, and from the superior temporal gyrus to the occipital lobe are stronger in the audio-visual condition compared to the other two conditions. In the high gamma frequencies, the strengths of the connections between different parts of the brain in the audio-visual condition are weaker than one of the two other speech conditions and stronger than the other condition.

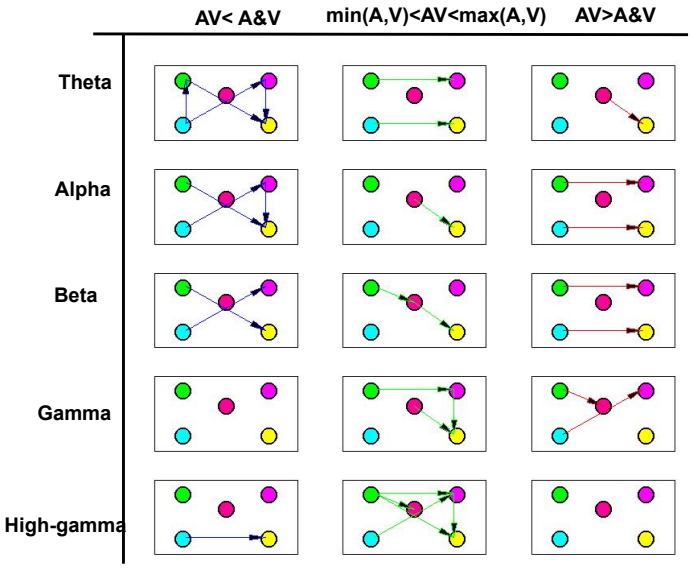


Figure 11: Sources and sinks in the speech conditions.

5. Discussion

It is known that the speech perception is a multi-modal process that involves both auditory and visual information. In this work, we have studied how disjoint brain regions communicate with one another in order to process speech under three basic conditions: auditory-only, visual-only and audio-visual speech. The high-level goal of this study is to better understand how different parts of the brain would communicate with one another under these speech conditions at different frequencies and how these networks would change from one condition to another. These neural interactions have been molded based on the partial directed coherence technique in order to capture the direct interactions between them.

Our study shows that the significant long and short neural interactions occur in beta and lower frequencies

in all of the three speech conditions. By measuring the strength of each node in the brain connectivity network—reflecting how strongly the node affects other brain areas and/or is influence by the activities of other brain areas—we have identified the brain areas involved in the speech processing in different frequencies. In particular, through the analysis of the strengths of the nodes in each speech condition, we have shown that the superior temporal gyrus and prefrontal cortex play an important role in the auditory-only speech condition. In the visual-only speech condition, the motor cortex and occipital lobe are highly active in communication with the rest of the brain in processing the visual information. The areas that were already involved in the auditory-only and visual-only speech, i.e., superior temporal gyrus and prefrontal, motor cortex, occipital lobe, are again involved in the speech processing in the audio-visual speech condition.

By identifying the strong nodes in the brain networks of different speech conditions and comparing with the power spectrum of the ECoG electrodes, we can observe a consistency in the locations of the brain areas that are engaged in processing the speech information. For example, consider the high-gamma power of the signals recorded from the ECoG electrode. Although we can identify the locations of the brain regions that respond to the stimulus through the high-gamma responses, we are able to both identify the locations of the responsive brain areas and study their interactions with one another under different speech conditions through the network analysis approach developed here.

Our models of the brain effective networks suggest that the effective brain networks within and across responsive cortical areas during auditory and/or visual tasks are sparse and the connections are not randomly dispersed among cortical areas. It is also observed that the strong long-range interactions occur in alpha and beta frequencies. Furthermore, in the auditory-only speech condition, the regions that are involved in the speech processing include the superior temporal gyrus and prefrontal cortex. In the visual-only speech condition, the motor cortex and lateral occipital lobe are the regions involved in the process. Furthermore, in the presence of the audio-visual speech, all of the four brain regions mentioned above are engaged in the activity. This is consistent with the supramodality of speech processing.

In addition to studying how individual nodes in the network communicate with one another, we have studied how clusters of nodes communicate with one another under different speech conditions. The nodes of the network are clustered based on their anatomical locations. This study is performed on the data collected from three different subjects. The results of this study show that the strengths of the connections between different clusters change significantly relative to the baseline in beta band and lower frequencies. We have also studied the inward and outward patterns of the connections for each cluster to identify the ones that play the role of a source within the brain network. In the auditory-only speech condition, the auditory cortex acts as a sink receiving information from the frontal lobe as well as from the superior temporal gyrus that acts as the source in the network. Interestingly, in the visual-only speech condition, the auditory cortex acts as a source forwarding the information to the visual cortex. This result supports the speculation that the visual process may involve the auditory pathways.

The main focus of our analysis has been on modeling the "direct" interactions between different parts of the brain. Therefore, we have used the partial directed co-

herence method to identify those connections in the network. The high temporal and spatial resolutions of ECoG data enable us to more accurately identify the cortical locations responsible for information processing and their interactions. However, unlike fMRI, ECoG electrodes cannot monitor the activities of all regions of the brain and, therefore, the role of some regions in processing the external stimuli remains unclear. Nevertheless, our study considers most of the regions that are known to play an important role in language and speech processing, and their activities are recorded by the ECoG electrodes.

References

Akaike, H., 1974. A new look at the statistical model identification. Automatic Control, IEEE Transactions on 19 (6), 716–723.

Alcántara, J. I., Weisblatt, E. J., Moore, B. C., Bolton, P. F., 2004. Speech-in-noise perception in high-functioning individuals with autism or asperger's syndrome. Journal of Child Psychology and Psychiatry 45 (6), 1107–1114.

Astolfi, L., Cincotti, F., Mattia, D., Marciani, M. G., Baccala, L. A., de Vico Fallani, F., Salinari, S., Ursino, M., Zavaglia, M., Ding, L., et al., 2007. Comparison of different cortical connectivity estimators for high-resolution eeg recordings. Human brain mapping 28 (2), 143–157.

Baccalá, L. A., Sameshima, K., 2001. Partial directed coherence: a new concept in neural structure determination. Biological cybernetics 84 (6), 463–474.

Ball, T., Demandt, E., Mutschler, I., Neitzel, E., Mehring, C., Vogt, K., Aertsen, A., Schulze-Bonhage, A., 2008. Movement related activity in the high gamma range of the human eeg. Neuroimage 41 (2), 302–310.

Bressler, S. L., Kelso, J., 2001. Cortical coordination dynamics and cognition. Trends in cognitive sciences 5 (1), 26–36.

Brovelli, A., Ding, M., Ledberg, A., Chen, Y., Nakamura, R., Bressler, S. L., 2004. Beta oscillations in a large-scale sensorimotor cortical network: directional influences revealed by granger causality. Proceedings of the National Academy of Sciences of the United States of America 101 (26), 9849–9854.

Cadotte, A. J., DeMarse, T. B., He, P., Ding, M., 2008. Causal measures of structure and plasticity in simulated and living neural networks. PloS one 3 (10), e3355.

Čeponienė, R., Lepistö, T., Shestakova, A., Vanhala, R., Alku, P., Näätänen, R., Yaguchi, K., 2003. Speech–sound-selective auditory impairment in children with autism: they can perceive but do not attend. Proceedings of the National Academy of Sciences 100 (9), 5567–5572.

Chen, H., Yang, Q., Liao, W., Gong, Q., Shen, S., 2009. Evaluation of the effective connectivity of supplementary motor areas during motor imagery using granger causality mapping. Neuroimage 47 (4), 1844–1853.

Dauwels, J., Vialatte, F., Musha, T., Cichocki, A., 2010. A comparative study of synchrony measures for the early diagnosis of alzheimer's disease based on eeg. NeuroImage 49 (1), 668–693.

Ding, M., Chen, Y., Bressler, S. L., 2006. Granger causality: Basic theory and application to neuroscience. Handbook of time series analysis: recent theoretical developments and applications, 437.

Eichler, M., 2006. On the evaluation of information flow in multivariate systems by the directed transfer function. Biological cybernetics 94 (6), 469–482.

Engel, A. K., Fries, P., Singer, W., 2001. Dynamic predictions: oscillations and synchrony in top–down processing. Nature Reviews Neuroscience 2 (10), 704–716.

Friston, K. J., 1994. Functional and effective connectivity in neuroimaging: a synthesis. Human brain mapping 2 (1-2), 56–78.

Godfrey, J. J., Syrdal-Lasky, K., Millay, K. K., Knox, C. M., 1981. Performance of dyslexic children on speech perception tests. Journal of experimental child psychology 32 (3), 401–424.

Goebel, R., Roebroeck, A., Kim, D.-S., Formisano, E., 2003. Investigating directed cortical interactions in time-resolved fmri data using vector autoregressive modeling and granger causality mapping. Magnetic resonance imaging 21 (10), 1251–1261.

Gow Jr, D. W., Segawa, J. A., 2009. Articulatory mediation of speech perception: a causal analysis of multi-modal imaging data. Cognition 110 (2), 222–236.

Hall, D. A., Barrett, D. J., Akeroyd, M. A., Summerfield, A. Q., 2005. Cortical representations of temporal structure in sound. Journal of neurophysiology 94 (5), 3181–3191.

Jaekl, P., Pesquita, A., Alsius, A., Munhall, K., Soto-Faraco, S., 2015. The contribution of dynamic visual cues to audiovisual speech perception. Neuropsychologia 75, 402–410.

Jerbi, K., Ossandón, T., Hamame, C. M., Senova, S., Dalal, S. S., Jung, J., Minotti, L., Bertrand, O., Berthoz, A., Kahane, P., et al.,

2009. Task-related gamma-band dynamics from an intracerebral perspective: Review and implications for surface eeg and meg. Human brain mapping 30 (6), 1758–1771.

Jones, J. A., Callan, D. E., 2003. Brain activity during audiovisual speech perception: an fmri study of the mcgurk effect. Neuroreport 14 (8), 1129–1133.

Kaiser, J., Lutzenberger, W., 2005. Human gamma-band activity: a window to cognitive processing. Neuroreport 16 (3), 207–211.

Kaminski, M., Blinowska, K., 1991. A new method of the description of the information flow in the brain structures. Biological cybernetics 65 (3), 203–210.

Kamiński, M., Ding, M., Truccolo, W. A., Bressler, S. L., 2001. Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. Biological cybernetics 85 (2), 145–157.

Korzeniewska, A., Franaszczuk, P. J., Crainiceanu, C. M., Kuś, R., Crone, N. E., 2011. Dynamics of large-scale cortical interactions at high gamma frequencies during word production: event related causality (erc) analysis of human electrocorticography (ecog). Neuroimage 56 (4), 2218–2237.

MacLeod, A., Summerfield, Q., 1987. Quantifying the contribution of vision to speech perception in noise. British journal of audiology 21 (2), 131–141.

McGurk, H., MacDonald, J., 1976. Hearing lips and seeing voices.

Miao, X., Wu, X., Li, R., Chen, K., Yao, L., 2011. Altered connectivity pattern of hubs in default-mode network with alzheimer's disease: an granger causality modeling approach. PloS one 6 (10), e25546.

Palva, S., Palva, J. M., Shtyrov, Y., Kujala, T., Ilmoniemi, R. J., Kaila, K., Naatanen, R., 2002. Distinct gamma-band evoked responses to speech and non-speech sounds in humans. J. Neurosci 22 (4), 211RC.

Pereda, E., Quiroga, R. Q., Bhattacharya, J., 2005. Nonlinear multivariate analysis of neurophysiological signals. Progress in neurobiology 77 (1), 1–37.

Ray, S., Crone, N. E., Niebur, E., Franaszczuk, P. J., Hsiao, S. S., 2008. Neural correlates of high-gamma oscillations (60–200 hz) in macaque local field potentials and their potential implications in electrocorticography. The Journal of Neuroscience 28 (45), 11526–11536.

Redcay, E., 2008. The superior temporal sulcus performs a common function for social and speech perception: implications for the emergence of autism. Neuroscience & Biobehavioral Reviews 32 (1), 123–142.

Rodriguez, E., George, N., Lachaux, J.-P., Martinerie, J., Renault, B., Varela, F. J., 1999. Perception's shadow: long-distance synchronization of human brain activity. Nature 397 (6718), 430–433.

Roebroeck, A., Formisano, E., Goebel, R., 2005. Mapping directed influence over the brain using granger causality and fmri. Neuroimage 25 (1), 230–242.

Schlögl, A., Supp, G., 2006. Analyzing event-related eeg data with multivariate autoregressive parameters. Progress in brain research 159, 135–147.

Schnider, S., Kwong, R., Lenz, F., Kwan, H., 1989. Detection of feedback in the central nervous system using system identification techniques. Biological cybernetics 60 (3), 203–212.

Schwartz, J.-L., Berthommier, F., Savariaux, C., 2004. Seeing to hear better: evidence for early audio-visual interactions in speech identification. Cognition 93 (2), B69–B78.

Skipper, J. I., van Wassenhove, V., Nusbaum, H. C., Small, S. L., 2007. Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. Cerebral Cortex 17 (10), 2387–2399.

Sporns, O., 2007. Brain connectivity. Scholarpedia 2 (10), 4695.

Stein, B. E., Meredith, M. A., 1993. The merging of the senses. The MIT Press.

Storey, J. D., 2003. The positive false discovery rate: A bayesian interpretation and the q-value. Annals of statistics, 2013–2035.

Summerfield, Q., 1992. Lipreading and audio-visual speech perception. Philosophical Transactions of the Royal Society B: Biological Sciences 335 (1273), 71–78.

Tuomainen, J., Andersen, T. S., Tiippana, K., Sams, M., 2005. Audio–visual speech perception is special. Cognition 96 (1), B13–B22.

Wallace, M. T., Ramachandran, R., Stein, B. E., 2004. A revised view of sensory cortical parcellation. Proceedings of the National Academy of Sciences of the United States of America 101 (7), 2167–2172.