Large-Scale Cortical Interactions During Auditory, Visual and Audiovisual 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

Based on the empirical findings on the human cortical locations responsive to auditory and/or visual stimuli, the objective of this work is to study the neural interactions within and across these cortical locations using electrocorticography (ECoG) signals. To this end, we obtain the brain functional connectivity networks under three basic speech conditions, namely auditory-only, visual-only and audiovisual. We show that the patterns of neural interactions largely depend on the condition of speech. In particular, under the auditory-only speech condition, the superior temporal gyrus directly communicates with the prefrontal cortex, and the prefrontal cortex interacts with the somatosensory cortex to process the auditory input. In the visual-only speech condition, the middle temporal gyrus and motor cortex directly exchange information, while the occipital lobe interacts with the prefrontal cortex. In the audio-visual speech condition, the prefrontal cortex communicates with the somatosensory cortex, and the superior temporal gyrus communicates with the occipital lobe to process the information. In addition, local communications within the motor cortex, somatosensory cortex, Wernick's area and superior temporal gyrus increase under all speech conditions. Several properties of the networks, including degree and strength of each node, are used to identify the areas that play key roles in speech processing for each speech condition. These analyses reveal that the superior temporal gyrus is highly involved in the auditory-only speech processing. In the visual-only speech condition, the occipital lobe, motor cortex, somatosensory cortex and prefrontal cortex play an active role in speech processing. The properties of the brain functional network under the audio-visual speech condition confirm the supra modality of speech by integrating the properties of the brain networks in the auditory-only and visual-only speech conditions.

1. Significance Statement

There exist empirical findings on the human cortical locations responsive to auditory and/or visual stimuli. It is also known that the speech perception is a multi-modal process that integrates both visual and acoustic information. However, the cortical interactions across the responsive brain areas are not well understood. Using high temporal and spatial resolution data collected from ECoG electrodes, we study the large-scale cortical interactions

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)

across responsive brain regions under three basic speech conditions. This study shows the importance of short- and long-range cortical interactions and identifies the areas that play an important role in speech processing under each speech condition. In this work, the brain functional networks are obtained based on partial correlation, which shows that the brain network is sparse and connected with large communicating distances.

2. Introduction

Human cognitive skills depend on various cortical and subcortical networks of neural interactions. 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 the speech perception is the well-known McGurk effect, where the auditory component of one sound paired with the visual component of another sound leads to the perception of a third sound. For example, the auditory /ba/ paired with the visual /ga/ yields 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 the 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; Joanisse et al., 2000; Mummery et al., 1999; Svirsky et al., 2004). 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 activations. These interactions can be represented by means of *functional connectivity*. In this work, we provide supporting evidence on the supra modality nature of speech processing in the human brain using a brain functional connectivity network analysis.

The functional connectivity is defined as a statistical dependency between the activities recorded from spatially remote brain regions. The concept of brain connectivity adds an important dimension to the functional mapping of cortical functions (Bressler and Tognoli, 2006). This concept suggests that the functional role of a cortical area is defined not only by its activation during a particular set of tasks but also by communicating with other remote cortical regions. Functional connectivity has been extensively studied using several functional neuro-imaging methods, particularly functional MRI (fMRI) (Koshino et al., 2005; Fair et al., 2007; Roy et al., 2009; Cordes et al., 2000; Cherkassky et al., 2006; Peltier et al., 2005). In this work, we study the brain neural interaction under three basic speech conditions: *auditory-only, visual-only* and *audio-visual* speech, using data collected from ECoG in human subjects implanted with subdural electrodes for epilepsy.

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 rela-

tionship 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. For the brain functional connectivity that is traditionally analyzed in terms of correlation and coherence, this issue can be addressed by means of partial correlation in the time domain and partial coherence in the frequency domain (Marrelec et al., 2006; Salvador et al., 2005; Sojoudi and Doyle, 2014). In this work, the patterns of the functional brain network under auditory, visual and audiovisual speech stimuli are studied in the time domain based on partial correlation.

After identifying the brain network structures, certain network properties, including degree and strength of each node, are used to study which brain areas are affected by the condition of speech and what roles they play in the speech processing. Moreover, we study the long- and short-range interactions between different brain regions to discover the communication patterns of these areas under each speech condition. In addition, different brain areas are clustered anatomically and their within and between community interactions are studied. The nodes of the brain connectivity networks are then clustered based on their patterns of connectivity and compared across the three speech conditions. Moreover, we show the importance of short- and long-range cortical interactions and identify the areas that play an important role in speech processing and show their dependence on the condition of speech.

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.

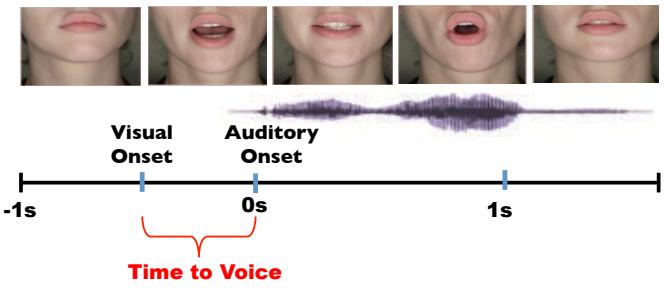


Figure 1: Visual and auditory onsets.

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 the 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 de-trended 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 time when no stimulus is presented) and the time interval $[0\text{s} \quad 0.5\text{s}]$ for the analysis of the three speech conditions. The number of trials chosen for the analysis is the same for all speech conditions. The study is performed on three subjects, resulting in similar conclusions. However, to simplify the presentation, we focus only on one subject in this paper, and the analysis for the other two subjects is provided in the supplementary materials. For the subjects under study, 50 trials are chosen on average for each analysis.

3.3. Brain Functional Connectivity Network

A subset of ECoG electrodes is selected for the analysis based on their high-gamma (70-190 Hz) responses. Since the magnitude and timing of high-gamma responses are highly correlated with increased firing rates in local cortical neurons, those electrodes whose corresponding signals show an increased high-gamma power relative to the baseline period are considered as responsive electrodes. These electrodes are then used to represent the nodes in the brain networks to be modeled. The brain functional networks for the three speech conditions are obtained based on the partial correlation coefficients. In what follows, we first review this concept together with a computational method for finding partial correlations.

Consider a variable set $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$, two random variables $x_i, x_j \in \mathbf{x}$, and a subset $\mathbf{y} \subseteq \mathbf{x} \backslash \{x_i, x_j\}$. We denote the partial correlation between x_i and x_j given $\mathbf{y}$ as $\rho_{x_i x_j . \mathbf{y}}$, which is defined as the correlation of the residuals R_{x_i} and R_{x_j} corresponding to the least-squares linear regression of x_i on $\mathbf{y}$ and x_j on $\mathbf{y}$, respectively. In the multivariate Gaussian setting, $\rho_{x_i x_j . \mathbf{y}} = 0$ if and only if x_i and x_j are conditionally independent given $\mathbf{y}$. In this case, the partial correlation can be computed efficiently without having to solve the regression problem, via inverting the correlation matrix for the union $\mathbf{y} \cup \{x_i, x_j\}$, provided that the correlation matrix is positive definite.

In practice, only a sample correlation matrix is available and the true (population) correlation matrix is unknown. In most fMRI studies, the number of measurements is low relative to the number of variables. Therefore, the sample partial correlation may not be directly computed through a matrix inversion. However, in neurophysiological recordings such as EEG and ECoG, the number of measurements is reasonably large compared to the number of recording sites and, hence, a good estimate of the partial correlation may be found by inverting and rescaling the sample correlation matrix. To model the functional connectivity between the recording sites of multichannel ECoG data, we first form the sample correlation matrices for all individual trials. These matrices are then used to compute the sample partial correlations. The sample partial correlation matrices are computed by normalizing the inverse sample correlation matrices with respect to the main diagonal elements.

To understand how the elements of the correlation and partial correlation matrices are related, we denote the (sample) correlation matrix of the union $\mathbf{y} \cup \{x_i, x_j\}$ as R and the $(i,j)^{th}$ entry of R^{-1} as r_{ij}. The (sample) partial correlation between x_i and x_j is as follows:

$$\rho_{x_i x_j . \mathbf{x} \setminus \{x_j, x_j\}} = \frac{-r_{ij}}{\sqrt{r_{ii} r_{jj}}} \tag{1}$$

For each individual trial, the partial correlation matrix is first computed. Fisher z-transformation is then applied to every individual partial correlation matrix using the equation

$$z_{ij} := \frac{1}{2} \ln \left(\frac{1 + \rho_{ij}}{1 - \rho_{ij}} \right) \tag{2}$$

for the $(i, j)^{\text{th}}$ entry of the partial correlation matrix. The average is taken over all z-transformed matrices.

The statistical procedure for ruling out the possibility of existence/nonexistence of an edge in a functional network associated with one of the conditions is explained in the next section.

3.4. Statistical Tests

To form the brain functional network of a speech condition, consider two nodes i and j in the network. The corresponding z-transformed partial correlation coefficient between these two nodes is found for each trial, and then these coefficients are used to form a weight set $\mathbf{W}_{ij} = \{W_{ij}^1, W_{ij}^2, \ldots, W_{ij}^N\}$, where W_{ij}^k is the z-transformed partial correlation coefficient between electrodes i and j in trial k. The outliers are removed from $\mathbf{W_{ij}}$ and the remaining values are entered into a one-tailed one sample t-test with mean equal to 0 at the 5% significance level. Finally, a positive false discovery rate (pFDR) analysis was applied to the p-values obtained from the t-test using the procedure described by (Storey, 2003). More precisely, the positive false discovery rate is defined to be

$$pFDR = \mathcal{E}\left[\frac{V}{R} \middle| R > 0 \right] \tag{3}$$

where $\mathcal{E}[\cdot]$ denotes the expectation operator, and V and R are the number of false positives and the number of rejected null hypotheses, respectively.

The above-mentioned steps are illustrated in Figure 2. The results of the FDR analysis are used to determine whether the partial correlation between any two nodes in the brain functional network is significant in each speech condition. If the partial correlation between two nodes in the network is statistically significant, it will be represented by an edge in the brain functional connectivity graph. These steps are repeated for the baseline, auditory-only, visual-only and audio visual speech conditions, which will yield four brain functional connectivity graphs $\mathcal{G}_B$, $\mathcal{G}_A$, $\mathcal{G}_V$ and $\mathcal{G}_{AV}$, respectively.

Denote the set of all three speech conditions with $\mathcal{S}$, i.e., $\mathcal{S}$ = {auditory, visual and audio-visual}. Suppose $\mathbf{W}_{ij}^{c1}$ and $\mathbf{W}_{ij}^{c2}$ are two weight sets representing the partial correlations between nodes i and j under conditions c_1 and c_2, where $c_1 \in \mathcal{S}$ and c_2 = baseline. The values of $\mathbf{W}_{ij}^{c1}$ and $\mathbf{W}_{ij}^{c2}$ are entered into a two-tailed two sample t-test with mean equal to zero at the 5% significance level. Afterwards, an FDR analysis is applied to the p-values obtained from the t-test. Let $\bar{W}_{ij}^{c1}$ and $\bar{W}_{ij}^{c2}$ be the average weights obtained from the weight sets $\mathbf{W}_{ij}^{c1}$ and $\mathbf{W}_{ij}^{c2}$, respectively. If the difference between $\mathbf{W}_{ij}^{c1}$ and $\mathbf{W}_{ij}^{c2}$ is statistically significant and $\bar{W}_{ij}^{c1}$ is larger than $\bar{W}_{ij}^{c2}$, then the edge connecting nodes i and j in the brain functional network associated with condition c_1 will be shown in red; otherwise, the edge will be colored in blue. Likewise, if c_1 and c_2 both belong to the speech condition set $\mathcal{S}$, then the relationship between nodes i and j in the brain functional networks associated with those speech conditions can be compared statistically.

Given a graph, the *degree* of each node is defined as the number of links (edges) in the graph that connect the node to other nodes in the graph. For each of the four

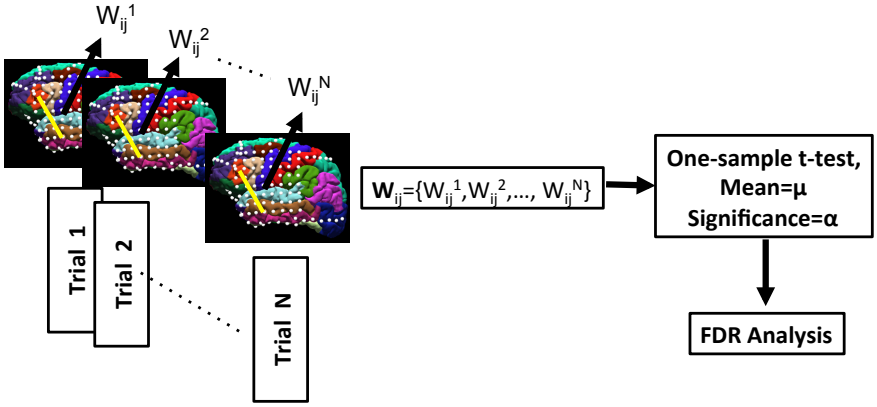


Figure 2: Statistical analysis of the functional connectivity.

functional brain networks $\mathcal{G}_B$, $\mathcal{G}_A$, $\mathcal{G}_V$ and $\mathcal{G}_{AV}$, the degree of each node is computed and reflected in the size of the circle showing the node in the graph. More precisely, the higher the degree of a node is, the larger the diameter of its corresponding circle is. Given a weighted graph, the *strength* of each node is defined as the sum of the weights of the links that connect the node to other nodes in the graph. We denote the strength of node i as S_i. In this work, we consider the weight of each edge in the graph as the value of the corresponding squared z-transformed partial correlation. In order to evaluate the dependency of the strength of a node to the condition of the speech, we form a union graph by taking the union of $\mathcal{G}_B$, $\mathcal{G}_A$, $\mathcal{G}_V$ and $\mathcal{G}_{AV}$, and denote it with $\bar{\mathcal{G}}$. The weights assigned to $\bar{\mathcal{G}}$ are variable and depend on the values of squared z-transformed partial correlation coefficients in each of the three speech conditions and the baseline period. Denote the "weighted union graphs" associated with the baseline, auditory, visual and audio-visual speech conditions as $\bar{\mathcal{G}}_B$, $\bar{\mathcal{G}}_A$, $\bar{\mathcal{G}}_V$ and $\bar{\mathcal{G}}_{AV}$, respectively.

To study the changes in the strength of each node from one speech condition to another, we form a *strength set* $\mathbf{S_{i_c}} = \{S^1_{i_c}, S^2_{i_c}, \ldots, S^N_{i_c}\}$, where $S^j_{i_c}$ denotes the strength of node i in trial j of the speech condition c. For every two conditions c_1 and c_2 in the speech condition set $\mathcal{S}$, $\mathbf{S_{i_{c_1}}}$ and $\mathbf{S_{i_{c_2}}}$ are computed based on the weighted union graphs. The outliers are then removed and the remaining values are entered into a two-tailed two-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. Let $\bar{S}_{i_{c_1}}$ and $\bar{S}_{i_{c_2}}$ be the average values obtained from the strengths sets $\mathbf{S_{i_{c_1}}}$ and $\mathbf{S_{i_{c_2}}}$, respectively. If $\bar{S}_{i_{c_1}}$ is larger than $\bar{S}_{i_{c_2}}$ and $\mathbf{S_{i_{c_1}}}$ is significantly different from $\mathbf{S}_{i_{c_2}}$ statistically, then the corresponding node in the graph of condition c_1 is shown in red and in blue otherwise. Likewise, the strength of every node is compared between each speech condition and the baseline.

In order to study the variability of the network properties under different speech conditions within and across groups of nodes (as opposed to individual nodes), the nodes of the network are clustered based on their anatomical locations. For a weighted union graph, the strength of a cluster is defined as the sum of the weights of the edges within the cluster. Assume that K is the number of clusters in the network. Let $\mathcal{S}^{cj}_k$ be the strength of cluster $k \in \{1, 2, \ldots, K\}$ in trial j under speech condition c. Moreover, define the strength of the connections between two clusters m and n as the sum of the weights of the edges that connect clusters m and n. Denote the strength of the connectivity between clusters m and n in trial j un-

der speech condition c with $\mathcal{S}_{mn}^{cj}$. To study how much the strength of a cluster and its connections to other clusters in the union graph change based on the condition of the speech, we first find the strength of each cluster and the strength of its connectivity to other clusters for each individual trial and speech condition. We use these numbers to form within and between clusters strength sets and denote them as $\mathbf{S}_k^c = \{\mathcal{S}_k^{c1}, \ldots, \mathcal{S}_k^{cN}\}$ and $\mathbf{S}_{mn}^c = \{\mathcal{S}_{mn}^{c1}, \ldots, \mathcal{S}_{mn}^{cN}\}$ for every $m, n \in \{1, \ldots, K\}$.

For every two conditions $c_1, c_2 \in \mathcal{S}$, we compute $\mathbf{S}^{c1}$, $\mathbf{S}^{c2}$, $\mathbf{S}_{mn}^{c1}$ and $\mathbf{S}_{mn}^{c2}$. Afterwards, $\mathbf{S}^{c1}$ with $\mathbf{S}^{c2}$ and $\mathbf{S}_{mn}^{c1}$ with $\mathbf{S}_{mn}^{c2}$ are entered into a two-tailed two-sample t-test with mean equal to zero at the 1% significance level, followed by an FDR correction. In addition to grouping the nodes of the networks based on their anatomical locations, it is useful to cluster them based on their pattern of connectivity as well. This type of clustering provides useful information on how different parts of the brain group together to process information. In this work, the nodes of each graph representing the brain connectivity network under a speech condition are clustered using the *Louvain* algorithm (Blondel et al., 2008; Reichardt and Bornholdt, 2006).

4. Results

In this work, we represent the functional connectivity networks associated with each speech condition with a network of nodes and edges, where each node represents a recording site and an edge between two nodes in the network represents a non-zero partial correlation between the variabilities of the signals recorded from those recording sites. Figure 3 shows the anatomical locations of the ECoG electrodes for the subject under study. A subset of these recording sites have been selected based on their high-gamma responses. The numbered yellow disks in Figure 3 show the anatomical locations of these recording sites. The cortical locations of the selected electrodes and their labels are provided in Table I.

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

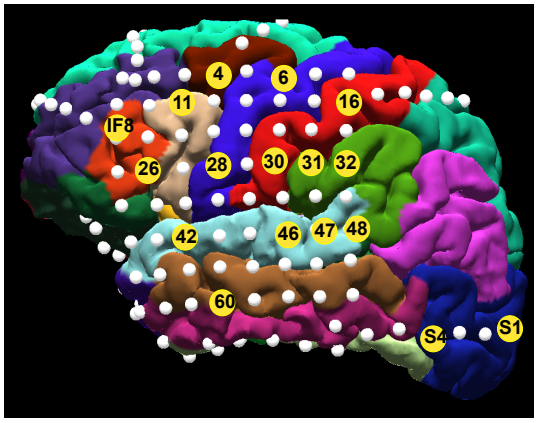


Figure 3: The locations of the selected ECoG electrodes for the subject under study.

In what follows, the approaches discussed earlier will be deployed to model the brain functional connectivity networks under three speech conditions as well as the baseline period. Certain properties of these networks will then be used to both identify the brain areas that respond to

speech stimuli and study how brain regions communicate with one another when the condition of speech changes from auditory-only to visual-only to audio-visual.

4.1. Brain Functional Connectivity Networks

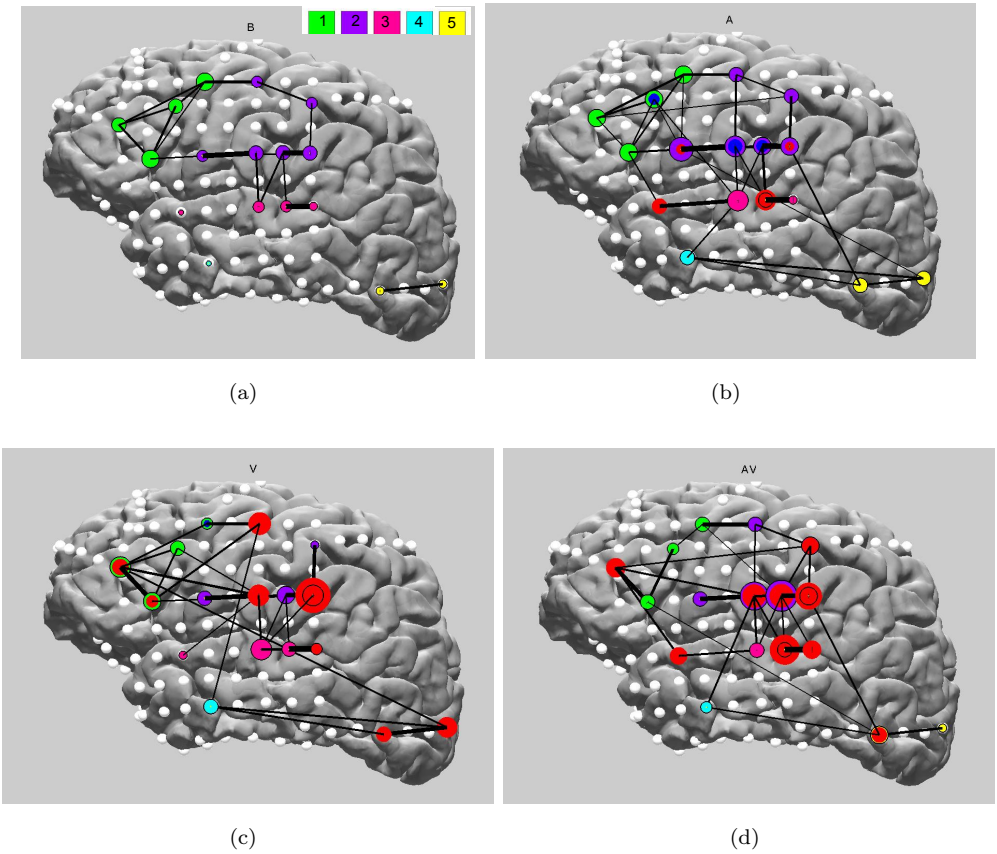
Using the methods described in Section 3.3, brain functional networks are obtained for all speech conditions and the baseline period. The graph representations of these networks, namely $\mathcal{G}_B$, $\mathcal{G}_A$, $\mathcal{G}_V$ and $\mathcal{G}_{AV}$, are given in Figure 4. Figure 4(a) shows the functional network during the baseline period. This brain network is highly sparse and there exist no significant long-range interactions between remote brain regions. However, the brain functional network changes under each speech condition and reveals an increased correlation between the activities or variabilities of anatomically remote brain areas, which indicates the role of those areas in the speech processing as well as the importance of their communications for processing speech.

In the functional connectivity graphs for all three speech conditions, the brain areas that form stronger and weaker connections to other regions are highlighted in red and blue, respectively. In particular, in presence of auditory-only speech stimuli, the brain regions whose activities are recorded via electrodes 28, 32, 42, 47 and 48 form stronger connections to other brain regions (Figure 4(b)). These regions include **pre-central, supra marginal, aSTG** and **pSTG**. Note that the strengths of these nodes are considered as the sum of the weights of the edges connected to these nodes in the union graph. More details on how these strengths are compared among different conditions can be found in Section 3.4.

Figure 4(c) shows the brain functional network under the visual-only speech condition. The regions whose strengths are increased significantly by the stimuli are the ones corresponding to electrodes, 6, IF8, 26, 30, 32, S1, and S4. These regions are **precentral, pars triangularis, postcentral, supra marginal** and **lateral occipital cortex**. Moreover, under the audio-visual speech condition, the regions associated with electrodes IF8, 16, 30, 31, 32, 42, 47, 48, and S4 form significantly stronger connections with other nodes in the functional network. These regions include **pars triangularis, postcentral, supra marginal, aSTG, pSTG** and **lateral occipital cortex**, and are shown in Figure 4(d).

The methods explained in Subsection 3.4 are used to study how the strength of each node in the brain functional network is effected by the speech condition. The results of this study are provided in Figure 5. Each node highlighted in red in any of these graphs has the property that its strength increases significantly under a speech condition. Similarly, if a node becomes significantly week under a speech condition, it is highlighted in blue. Figure 5(a) shows the map of all nodes that are significantly stronger or weaker during the auditory-only condition relative to the visual-only speech condition. The nodes with larger differences are shown by larger diameters. In particular, the nodes associated with **caudal middle frontal, precentral, aSTG, mSTG** and **pSTG** are significantly stronger under the auditory-only speech condition relative to the visual-only speech condition. On the other hand, the nodes associated with **pars triangularis** and **postcentral** become weaker under the auditory-only condition in comparison with the visual-only condition, while the remaining nodes remain relatively unchanged.

Note that the strengths of the nodes can be used to identify the areas that are highly engaged in the speech processing. These areas largely depend on the speech condition. Figure 5(b) shows the map of the nodes that are significantly stronger or weaker during the auditory-only



(a) (b)

(c) (d)

Figure 4: The brain functional connectivity networks representing the cortical interactions during: (a) baseline, (b) auditory-only speech condition, (c) visual-only speech condition, and (d) audio-visual speech condition. In these graphs, the nodes that are significantly stronger during each speech condition relative to the baseline are highlighted in red. The nodes that show a decrease in their strengths are shown in blue.

condition relative to the audio-visual speech condition. As illustrated in the map, the nodes located in **aSTG, mSTG** and **pSTG** show increasing strengths during the auditory-only condition, while the strengths of the nodes in **pars triangularis** and **precentral** decrease significantly in the auditory-only condition (or equivalently increase significantly during the auditory-visual speech condition). Figure 5(c) shows a nodal map that can be used to compare the strength of each node in the visual-only condition relative to the audio-visual condition. In the presence of visual-only speech stimuli, the nodal strengths in **supra marginal, mSTG, pSTG** and **lateral occipital cortex** increase significantly, while the nodes located in **pars opercularis** and **aSTG** form stronger connections to other nodes under the audio-visual stimuli.

So far, we have used nodal interactions to identify brain regions that are highly active in speech processing, to understand how they interact with other parts of the brain in order to process speech, and to study the dependency of certain properties of the network on the condition of speech. In order to study the brain neural interactions within and between larger brain segments, we cluster the nodes of each network according to their anatomical locations into 5 different clusters as follows:

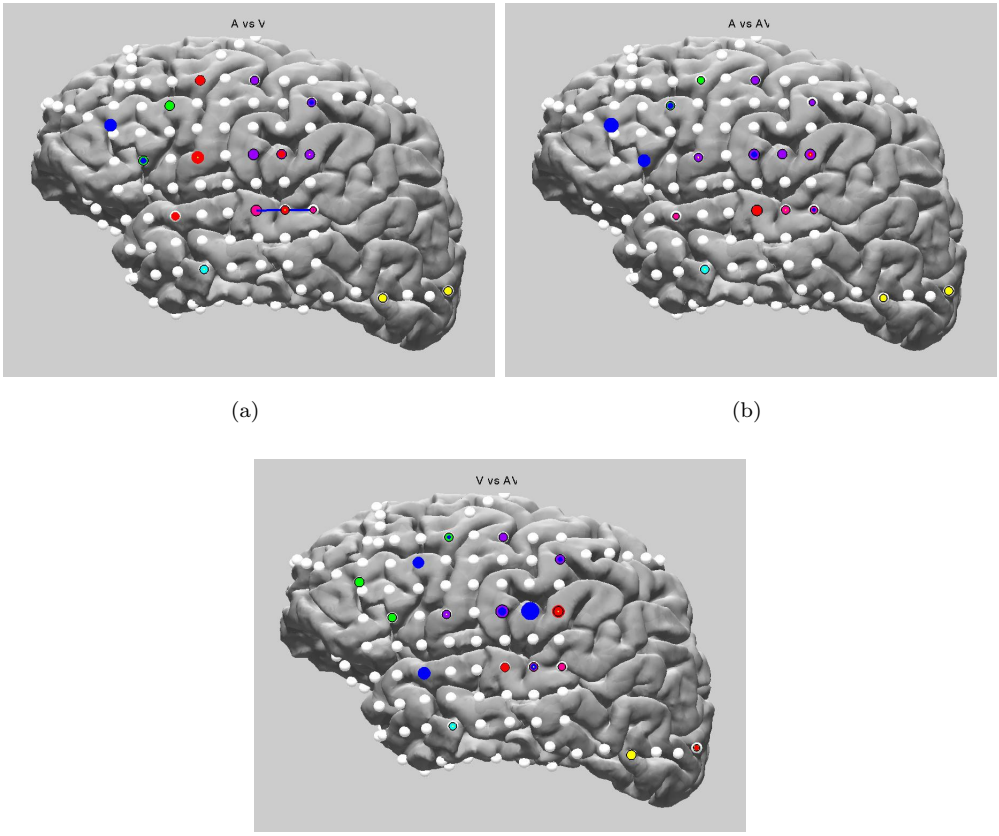
Cluster 1: $C_1 = \{4, 11, 26, IF8\}$. These nodes/electrodes are located in the prefrontal cortex.

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.



(a) (b)

(c)

Figure 5: The nodal maps comparing the strength of each node in one speech condition relative to another speech condition: (a) map of nodal strength changes in the auditory-only condition relative to the visual-only condition, (b) map of nodal strength changes in the auditory-only condition relative to the audio-visual condition, (c) map of nodal strength changes in the visual-only condition relative to the audio-visual condition. The stronger nodes in one condition in comparison with another condition are shown in red. The weaker nodes are shown in blue.

The five clusters defined above are shown in Figure 4(a). The nodes belonging to the same cluster of the network are shown in the same color. In what follows, we investigate: i) how the strength of each cluster changes from one speech condition to another, ii) how the links that connect these clusters to one another are effected by the condition of speech.

4.2. Cluster Connectivity

The nodes of the brain functional networks depicted in Figure 4 are first clustered into the five clusters mentioned before. The strength of the connections within each cluster and the strength of the connections between every two clusters in the network are then evaluated for each speech condition. Figure 6(a) shows how the strength of each cluster changes from one speech condition to another. In particular, the strengths of clusters 2 and 3 increase significantly during all three speech conditions. These clusters span the **primary motor cortex, primary somantosensory cortex, wernicke's area** and **superior temporal gyrus**. The brain areas associated with clusters 2 and 3 exhibit not only a high level of correlated activities under speech stimuli relative to the baseline, but also significant differences from one speech condition to another. These differences are highlighted in Figure 6(b). For instance, the strength of the connectivity within the primary motor cortex, primary somantosensory cortex and wernicke's area is significantly higher during the auditory-only speech relative to the visual-only speech condition. Moreover, the strength of the connectivity within the superior temporal gyrus is significantly different in the audio-visual condition relative to the two other speech conditions. Furthermore, the strength of cluster 5 associated with the occipital lobe decreases during the auditory-only speech and increases during the visual-only and audio-visual speech conditions.

Figures 6(c)–(e) illustrate how the strength of the connections between every two clusters in the network changes from one speech condition to another. As shown in Figure 6(c), the correlation between the activities of C_1 (prefrontal cortex) and C_2 (motor, somatosensory and wernicke's area) increases under the auditory-only speech condition. The connection between C_1 and C_3 (superior temporal gyrus) also becomes stronger, though insignificantly. On the other hand, the strength of the connections between C_2 and C_4 (middle temporal gyrus) decreases significantly under the auditory-only stimuli. Figure 6(d) shows the changes in the values of partial correlations between different clusters under the visual-only speech condition. In particular, the correlated activities increase between C_1 and C_2, and decrease between C_3 and C_4 under the visual-only speech. In the audio-visual speech experiment, the correlated activities between C_1 and C_2, C_2 and C_3, C_2 and C_5 (occipital lobe) increase significantly in comparison with the baseline period. These results are shown in Figure 6(e).

In addition to investigating how the strengths of the interactions within each cluster and between every two clusters in the network change from one speech condition to another, it is useful to compare the number of edges within and between the anatomically clustered nodes under each speech condition. These numbers are given in Figure 7. A quick comparison reveals that nodes of the 3^{rd} cluster, i.e., superior temporal gyrus, are more connected during all three speech conditions compared to the baseline. However, the nodes of the network in the cluster associated with the prefrontal cortex have less connectivity under the three speech conditions. Moreover, the number of interactions within the second cluster, which spans the

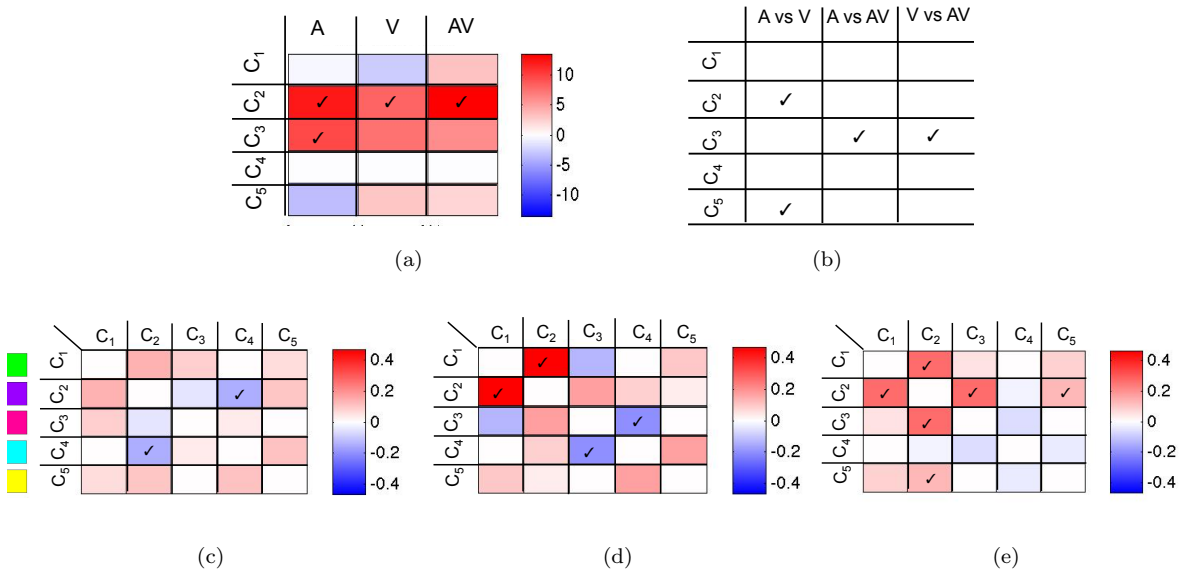


Figure 6: Strength of the connections within each cluster and between every two clusters in the network: (a) strength of the connections within each of the five clusters under the auditory-only, visual-only and audio-visual speech conditions compared to the baseline, (b) comparison of the strength of each cluster for each speech condition, (c) the inter-cluster strengths during the auditory-only speech stimuli in comparison to the baseline, (d) the inter-cluster strengths during the visual-only speech stimuli relative to the baseline, (e) the inter-cluster strengths during the audio-visual condition relative to the baseline. The ticked boxes show the conditions under which the strengths of the interactions within a cluster decrease/increase significantly. In all these graphs, the increased strengths are shown in red and the decreased strengths are shown in blue.

primary motor cortex, primary somatosensory cortex and wernicke's area, increases significantly under the auditory-only and audio-visual speech conditions. Figure 7 also shows the connectivity between every two clusters in the brain network. Under all three speech conditions, the connectivity between prefrontal cortex and primary motor cortex, primary somatosensory cortex and wernicke's area increases significantly. Similarly, the number of links connecting the primary motor cortex, primary somatosensory cortex, and wernicke's area to the superior temporal gyrus increases significantly under the three speech conditions.

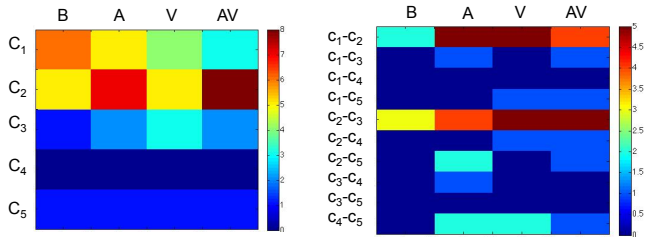
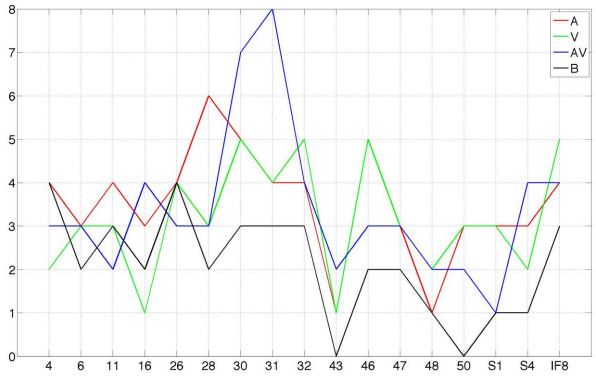


Figure 7: Number of neural interactions within and between anatomically clustered nodes in the brain functional networks.

The degree of a node in the brain functional network could potentially provide useful information on the importance of that node/area in the network for processing information under certain conditions. Figure 8 shows how the degree of a node in the brain connectivity map changes from one speech condition to another. For instance, nodes 31 and 32 located in the wernicke's area have the highest degree of connectivity in the brain functional network associated with the audio-visual speech condition. These nodes can be considered as central nodes in the brain functional network during the audio-visual speech processing. On the other hand, node 28 located in the precentral region is a central node in the auditory speech processing. For processing the visual-only speech, nodes 30, 31, 46 and IF8 located in the postcentral, supra marginal, mSTG and pars triangularis become the central nodes, as shown in the graph.

An alternative way of visualizing how different parts of the brain interact with one another under different condi-



(a)

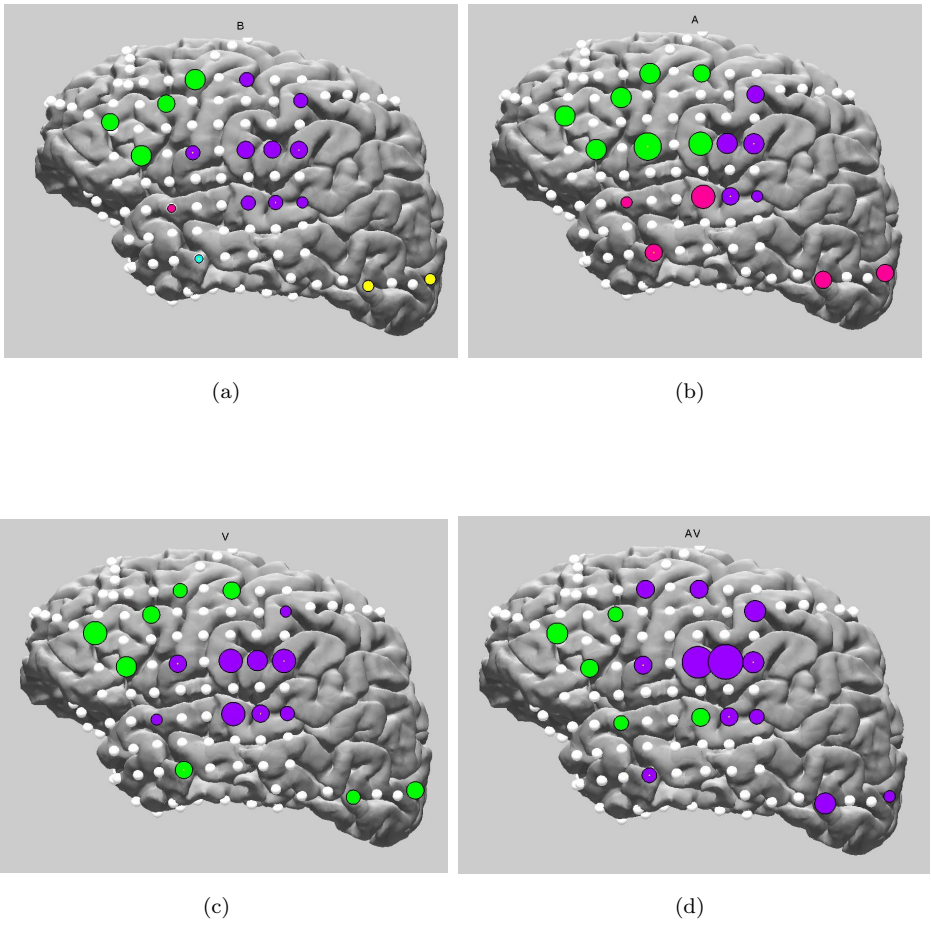
Figure 8: Degrees of the nodes in the brain networks for the baseline period (black), auditory-only (red), visual-only (green), and audio-visual (blue) speech conditions.

tions is to cluster them based on their connectivity, rather than their anatomical locations as considered before. In order to cluster the nodes of the brain networks based on their connectivity, the Louvain clustering algorithm is used and the results of this algorithm are provided in Figure 9. All nodes of the network in the same cluster are shown with the same color. Figure 9(a) shows the map of the clustered nodes of the network associated with the baseline, i.e., $\mathcal{G}_B$. We can observe that the clusters are mostly local, and that the outcome of this clustering strategy is somewhat similar to the anatomical clustering performed before. The exception is that the nodes located in the superior temporal gyrus are clustered with the ones located in the motor cortex, somatosensory cortex and wernicke's area (i.e., cluster C_2 in the anatomical clustering).

Figure 9(b) shows the anatomical locations of the nodes in the brain network $\mathcal{G}_A$ and their clustering in the auditory-only speech condition. This map unveils the clustering of some areas in the occipital lobe with middle temporal gyrus and a part of the superior temporal gyrus. On the other hand, a part of the superior temporal gyrus is clustered with the somatosensory cortex and wernicke's area. In addition, the nodes located in the prefrontal lobe are clustered with the nodes located in the motor cortex and somatosensory cortex. In Figure 9(c), it is shown which parts of the brain interact with one another for processing visual-only speech information. According to this map, some areas in the prefrontal lobe, primary motor cortex and middle temporal gyrus directly communicate with the areas located in the occipital lobe. On the other hand, the superior temporal gyrus is clustered with the primary somatosensory cortex, wernicke's area and a part of the motor cortex. The clustering of the brain network under the audio-visual-speech condition, namely $\mathcal{G}_{AV}$, is given in Figure 9(d). This map shows the clustering of the nodes located in the occipital lobe with the ones in the middle temporal gyrus, motor cortex, somatosensory cortex and wernicke's area. Moreover, a part of the superior tempera gyrus is clustered with the prefrontal cortex in the graph associated with the audio-visual speech condition.

5. Discussion

Speech perception is a multi-modal process that involves both auditory and visual information. There exist empirical findings on the human cortical locations responsive to auditory and/or visual stimuli. The purpose of this work has been the development of a framework under which the communication networks underlying the auditory-only, visual-only and audio-visual speech processing in the brain could be identified. In particular, we have studied how disjoint brain regions associated with audio/visual speech interact with one another under different speech conditions and how the brain functional network changes from one condition to another. The patterns of "direct" neural interactions during these activities have been obtained based on the partial correlation method. Each experiment has been repeated multiple times and



(a) (b)

(c) (d)

Figure 9: Maps of the nodes in the brain networks that are clustered together using the Louvain clustering algorithm for: (a) baseline period, (b) auditory-only speech condition, (c) visual-only speech condition, (d) audio-visual speech condition. All nodes belonging to the same cluster are shown with the same color.

these trials have been used to statistically validate the results of the analysis. The data used in this study was collected from three epilepsy patients, implanted with subdural electrodes with normal verbal function and no evidence for functional reorganization.

Using brain signals recorded from ECoG, the cortical locations responding to auditory and/or visual stimuli are first identified. This work focuses on the temporal time correlation between the activities of disjoint brain areas under the three basic speech conditions. In order to capture the direct interactions between different areas, we calculate partial correlations rather than correlation coefficients. We employ the values of the partial correlation coefficients to form weighted graphs that represent the brain functional connectivities under different speech conditions.

Our models of the brain functional networks suggest that the functional brain networks within and across responsive cortical areas under auditory and/or visual speech conditions are sparse and that the connections are not randomly dispersed among cortical areas. Several properties of the networks, including the degree and strength of each node, are compared among different conditions. In addition, disjoint brain areas whose activities are recorded via ECoG electrodes are clustered anatomically and their within and between community interactions are studied for each condition.

Comparing the graphs of the brain connectivity networks reveals that most of the correlated activities are local during the baseline period, while anatomically remote brain areas interact with one another to process speech under the speech conditions. In particular, long-range connections connect the superior temporal gyrus to the prefrontal cortex and the prefrontal cortex to the somatosensory cortex under auditory-only stimuli. Under the visual-only condition, the middle temporal gyrus is connected to the motor cortex, and the occipital lobe is connected

to the prefrontal cortex. Furthermore, long-range connections link the prefrontal cortex to the somatosensory cortex, superior temporal gyrus and occipital lobe under audio-visual speech stimuli. In addition, the correlated local activities within the motor cortex, somatosensory cortex and Wernick's area, as well as the superior temporal gyrus, increase significantly under the three speech conditions.

In addition to studying the neural interactions represented by the edges in the brain functional network, the strengths of the nodes are measured as the sum of the weights of the edges connected to each node. The nodal strength analysis unveils that the areas involved in the auditory speech processing include the superior temporal gyrus. In the visual-only speech processing, the nodes located in the occipital lobe, motor cortex, somatosensory cortex and prefrontal cortex become significantly stronger. Under audio-visual speech stimuli, those nodes of the network that are activated include the ones activated during the auditory-only or visual-only speech condition. Comparing the degrees of the nodes in the brain functional networks shows that the somatosensory cortex and Wernick's area have the highest degree of connectivity during the audio-visual speech processing.

In this work, the nodes of the brain functional networks are clustered both in terms of their anatomical locations and their pattern of connectivity. In the anatomical clustering, the strength of each cluster and its connection to other clusters are obtained and compared among different speech conditions. A comparison of the anatomically clustered nodes in the brain networks shows that there is a strong correlation between the activities of the prefrontal cortex, motor cortex and somatosensory cortex under the three speech stimuli. It also shows a strong connectivity between the motor cortex, somatosensory cortex, Wernick's area and superior temporal gyrus, especially under the audio-visual speech condition. Furthermore, the strength of the connections improves significantly within the superior temporal gyrus, motor area and somatosensory cortex under all three speech conditions. In addition to the anatomical clustering, the nodes of the brain functional networks are clustered together based on their pattern of connectivity using the Louvain clustering algorithm.

The high temporal and spatial resolutions of ECoG data enable us to accurately identify both the cortical locations responsible for information processing and their interactions. However, a main limitation of the ECoG technology is that it does not cover all possible brain areas. Therefore, the role of unmeasured cortical activities in processing the external stimuli remains unclear. Although our studies of the brain speech processing is no exception, we have selected our subjects based on their ECoG coverage and ensured that most of the brain areas known to play important roles in language and speech processing have been covered and their activities have been recorded by the ECoG electrodes. However, it is difficult, if not impossible, to avoid the differences across individual subjects. More specifically, the locations of the ECoG electrodes would be slightly different across the subjects even though we have attempted to maximize the similarity of the ECoG coverage between the subjects. Due to this variability across the subjects, the functional connectivities are studied separately for each subject.

References

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.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment 2008 (10), P10008.

Bressler, S. L., Tognoli, E., 2006. Operational principles of neurocognitive networks. International Journal of Psychophysiology 60 (2), 139–148.

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.

Cherkassky, V. L., Kana, R. K., Keller, T. A., Just, M. A., 2006. Functional connectivity in a baseline resting-state network in autism. Neuroreport 17 (16), 1687–1690.

Cordes, D., Haughton, V. M., Arfanakis, K., Wendt, G. J., Turski, P. A., Moritz, C. H., Quigley, M. A., Meyerand, M. E., 2000. Mapping functionally related regions of brain with functional connectivity mr imaging. American Journal of Neuroradiology 21 (9), 1636–1644.

Fair, D. A., Schlaggar, B. L., Cohen, A. L., Miezin, F. M., Dosenbach, N. U., Wenger, K. K., Fox, M. D., Snyder, A. Z., Raichle, M. E., Petersen, S. E., 2007. A method for using blocked and event-related fmri data to study ?resting state? functional connectivity. Neuroimage 35 (1), 396–405.

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.

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.

Joanisse, M. F., Manis, F. R., Keating, P., Seidenberg, M. S., 2000. Language deficits in dyslexic children: Speech perception, phonology, and morphology. Journal of experimental child psychology 77 (1), 30–60.

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.

Koshino, H., Carpenter, P. A., Minshew, N. J., Cherkassky, V. L., Keller, T. A., Just, M. A., 2005. Functional connectivity in an fmri working memory task in high-functioning autism. Neuroimage 24 (3), 810–821.

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

Marrelec, G., Krainik, A., Duffau, H., Pélégrini-Issac, M., Lehéricy, S., Doyon, J., Benali, H., 2006. Partial correlation for functional brain interactivity investigation in functional mri. Neuroimage 32 (1), 228–237.

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

Mummery, C. J., Ashburner, J., Scott, S. K., Wise, R. J., 1999. Functional neuroimaging of speech perception in six normal and two aphasic subjects. The Journal of the Acoustical Society of America 106 (1), 449–457.

Peltier, S. J., Kerssens, C., Hamann, S. B., Sebel, P. S., Byas-Smith, M., Hu, X., 2005. Functional connectivity changes with concentration of sevoflurane anesthesia. Neuroreport 16 (3), 285–288.

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.

Reichardt, J., Bornholdt, S., 2006. Statistical mechanics of community detection. Physical Review E 74 (1), 016110.

Roy, A. K., Shehzad, Z., Margulies, D. S., Kelly, A. C., Uddin, L. Q., Gotimer, K., Biswal, B. B., Castellanos, F. X., Milham, M. P., 2009. Functional connectivity of the human amygdala using resting state fmri. Neuroimage 45 (2), 614–626.

Salvador, R., Suckling, J., Coleman, M. R., Pickard, J. D., Menon, D., Bullmore, E., 2005. Neurophysiological architecture of functional magnetic resonance images of human brain. Cerebral cortex 15 (9), 1332–1342.

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.

Sojoudi, S., Doyle, J., 2014. Study of the brain functional network using synthetic data. Allerton.

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.

Svirsky, M. A., Teoh, S.-W., Neuburger, H., 2004. Development of language and speech perception in congenitally, profoundly deaf children as a function of age at cochlear implantation. Audiology and Neurotology 9 (4), 224–233.

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

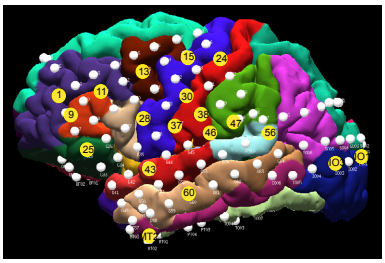
Wallace, M. T., Ramachandran, R., Stein, B. E., 2004. A revised view of sensory cortical parcellation. Proceedings of the National Academy of Sciences of the United States of America 101 (7), 2167–2172.

6. Appendix

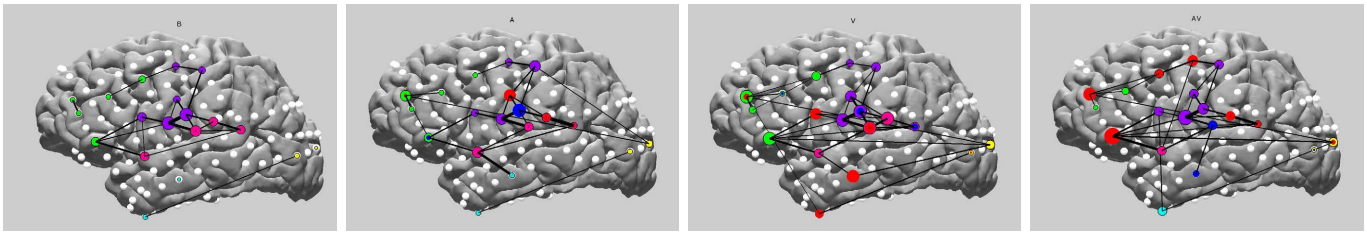
The results of the brain functional network modeling and analysis have been presented in the main body of the paper for only one subject. In order to provide more supporting evidence, we have studied the neural interactions for two other subjects under three speech conditions. The results of these studies are summarized in Figures 10 and 11.

Figure 10(a) shows the locations of the ECoG electrodes implanted on the brain of Subject 2. The brain functional connectivity networks representing the cortical interactions for this subject are shown in Figures 10(b)-(e) during the baseline period, auditory-only, visual-only, and audio-visual speech conditions, respectively. These graphs show that the interactions are short range and mostly local during the baseline period, while the processing speech would require both short- and long-range cortical interactions. Furthermore, the increased nodal strengths in the middle and superior temporal gyrus, prefrontal cortex, motor area, somatosensory cortex and occipital lobe reveal the importance of these areas in processing speech.

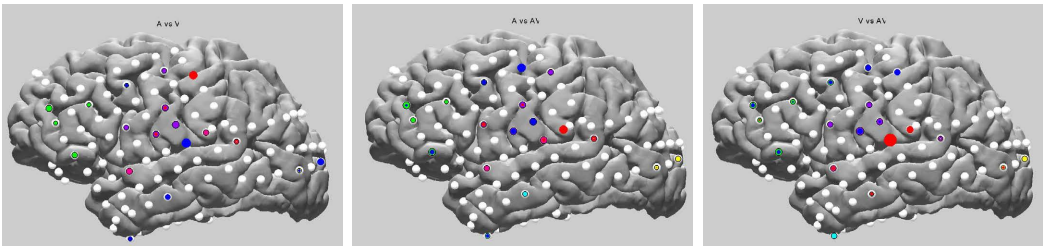
Figures 10(f)-(h) show the maps of the nodal strengths, which are compared between each two speech conditions. The nodes of the brain networks are clustered together based on their connectivity, and the results of these clusterings are provided in Figures 10(i)-(l). Finally, the degree of each node is calculated under the three speech conditions and the baseline in Figure 10(m). This graph shows that the node located in the somatosensory cortex has the highest degree of connectivity and its degree is highest during the audio-visual speech experiment. Similar results for Subject 3 are presented in Figure 11.



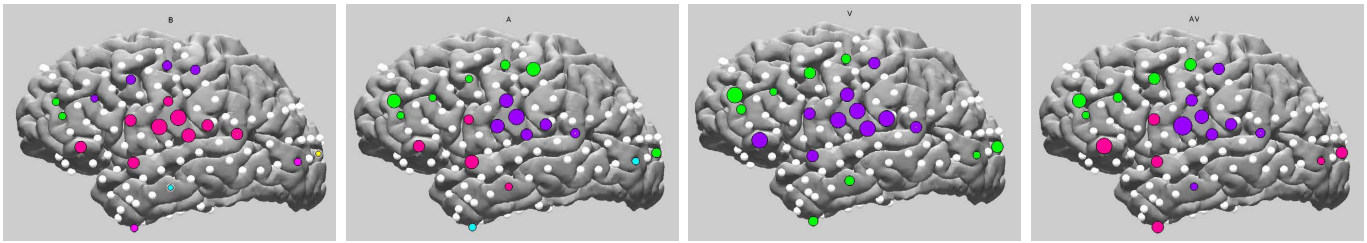
(a)



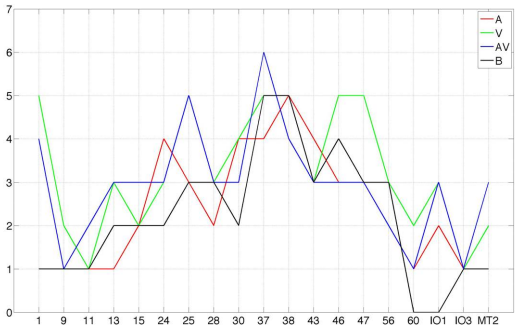
(b) (c) (d) (e)



(f) (g) (h)

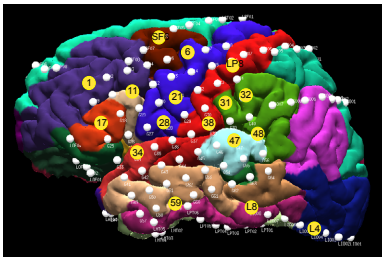


(i) (j) (k) (l)

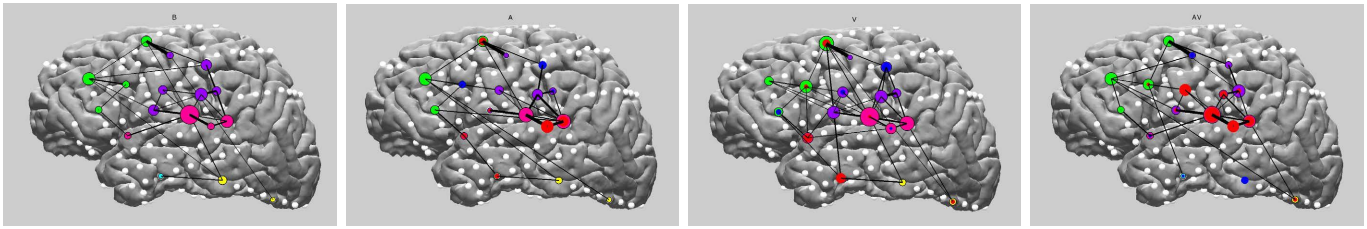


(m)

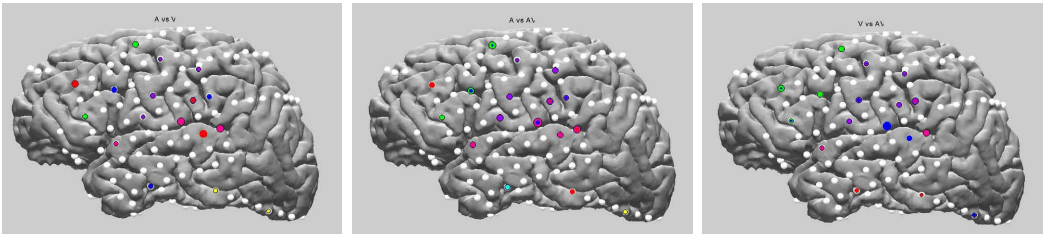
Figure 10: a) The locations of the selected ECoG electrodes for Subject 2. The brain functional connectivity networks representing the cortical interactions for this subject are shown during: (b) the baseline period, (c) the auditory-only speech condition, (d) the visual-only speech condition, and (e) the audio-visual speech condition. The nodal maps comparing the strength of each node are shown in: (f) the auditory-only condition relative to the visual-only condition, (g) the auditory-only condition relative to the audio-visual condition, and (h) the visual-only condition relative to the audio-visual condition. The maps of the nodes in the brain networks that are clustered together based on their connectivity using the Louvain clustering algorithm are given for: (i) the baseline period, (j) the auditory-only speech condition, (k) the visual-only speech condition, and (l) the audio-visual speech condition. The degrees of the nodes in the brain networks of the subject under study during the baseline period (black), auditory-only (red), visual-only (green), and audio-visual (blue) speech conditions are given in Figure (m).



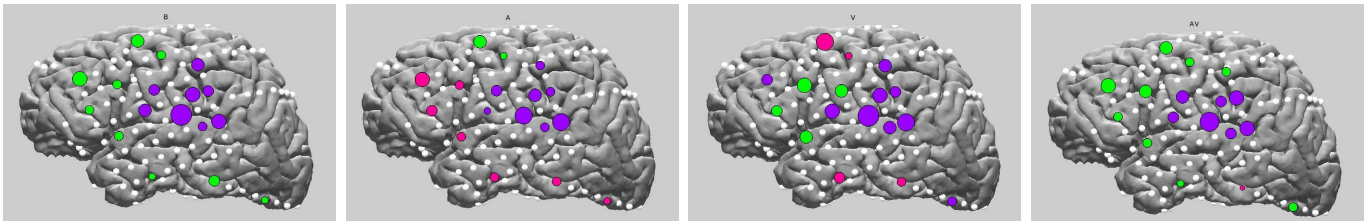
(a)



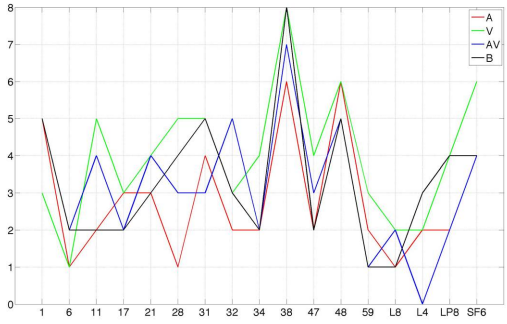
(b) (c) (d) (e)



(f) (g) (h)



(i) (j) (k) (l)



(m)

Figure 11: a) The locations of the selected ECoG electrodes for Subject 2. The brain functional connectivity networks representing the cortical interactions for this subject are shown during: (b) the baseline period, (c) the auditory-only speech condition, (d) the visual-only speech condition, and (e) the audio-visual speech condition. The nodal maps comparing the strength of each node are shown in: (f) the auditory-only condition relative to the visual-only condition, (g) the auditory-only condition relative to the audio-visual condition, and (h) the visual-only condition relative to the audio-visual condition. The maps of the nodes in the brain networks that are clustered together based on their connectivity using the Louvain clustering algorithm are given for: (i) the baseline period, (j) the auditory-only speech condition, (k) the visual-only speech condition, and (l) the audio-visual speech condition. The degrees of the nodes in the brain networks of the subject under study during the baseline period (black), auditory-only (red), visual-only (green), and audio-visual (blue) speech conditions are given in Figure (m).