BRAIN CONNECTIVITY
Volume 5, Number 4, 2015

DOI: 10.1089/brain.2013.0218

Brain Networks Engaged in Audiovisual Integration During Speech Perception Revealed by Persistent Homology-Based Network Filtration

Heejung Kim,[1–3] Jarang Hahm,[1–3] Hyekyoung Lee,[1,2] Eunjoo Kang,[4] Hyejin Kang,[1,5] and Dong Soo Lee[1–3,6]

Abstract

The human brain naturally integrates audiovisual information to improve speech perception. However, in noisy environments, understanding speech is difficult and may require much effort. Although the brain network is supposed to be engaged in speech perception, it is unclear how speech-related brain regions are connected during natural bimodal audiovisual or unimodal speech perception with counterpart irrelevant noise. To investigate the topological changes of speech-related brain networks at all possible thresholds, we used a persistent homological framework through hierarchical clustering, such as single linkage distance, to analyze the connected component of the functional network during speech perception using functional magnetic resonance imaging. For speech perception, bimodal (audio-visual speech cue) or unimodal speech cues with counterpart irrelevant noise (auditory white-noise or visual gum-chewing) were delivered to 15 subjects. In terms of positive relationship, similar connected components were observed in bimodal and unimodal speech conditions during filtration. However, during speech perception by congruent audiovisual stimuli, the tighter couplings of left anterior temporal gyrus-anterior insula component and right premotor-visual components were observed than auditory or visual speech cue conditions, respectively. Interestingly, visual speech is perceived under white noise by tight negative coupling in the left inferior frontal region–right anterior cingulate, left anterior insula, and bilateral visual regions, including right middle temporal gyrus, right fusiform components. In conclusion, the speech brain network is tightly positively or negatively connected, and can reflect efficient or effortful processes during natural audiovisual integration or lip-reading, respectively, in speech perception.

Key words: audiovisual speech; fMRI; functional connectivity; graph filtration; persistent homology; single linkage dendrogram

Introduction

In humans, speech perception requires the integration of multisensory information. During natural conversation, we simultaneously perceive speech via sound (i.e., a speaker's voice) and visual cues (i.e., a speaker's articulatory movements). However, understanding speech in noisy environments is difficult and often requires a great deal of effort. In situations such as these, visual information can help us better understand a degraded auditory cue if it is congruent with auditory speech (Grant and Seitz, 2000). In contrast, if visual information is incongruent with or is competing with auditory speech cues, an individual might ignore visual information to recognize speech. When visual speech cues are presented without an auditory speech stimulus, subjects having no lip-reading experience will have great difficulty understanding speech. In other words, unimodal auditory or visual speech cues in a noisy environment could have different rates of success in terms of speech perception.

Speech perception has been investigated by identifying multisensory interactions based on stimuli congruency or intelligibility (Hickok and Poeppel, 2007; McGettigan

[1]Department of Nuclear Medicine, College of Medicine, Seoul National University, Seoul, Korea.
[2]Institute of Radiation Medicine, Medical Research Center, Seoul National University, Seoul, Korea.
[3]Interdisciplinary Program in Cognitive Science, Seoul National University, Seoul, Korea.
[4]Department of Psychology, Kangwon National University, Chuncheon, Korea.
[5]Data Science for Knowledge Creation Research Center, Seoul National University, Seoul, Korea.
[6]Department of Molecular Medicine and Biopharmaceutical Sciences, Graduate School of Convergence Science and Technology and College of Medicine or College of Pharmacy, Seoul National University, Seoul, Korea.

et al., 2012; Scott et al., 2000). Recently, functional integration across brain regions has been widely researched during speech perception or comprehension (Abrams et al., 2013; Schall and von Kriegstein, 2014; Yue et al., 2013). For example, under adverse listening conditions, increasing functional connectivity between frontal and parietal regions has been shown to facilitate speech comprehension (Obleser et al., 2007). However, these connectivity studies were examined in one speech condition such as during auditory-only speech perception (Schall and von Kriegstein, 2014), silent speech reading (Chu et al., 2013), or acoustic degradation condition (Obleser et al., 2007). These task choices made unclear how modality-specific brain regions and integration regions interact. We supposed that brain regions underlying speech perception could be activated not only during multimodal speech cue condition but also during unimodal speech cue conditions (e.g., auditory speech with visual noise and visual speech with auditory noise). Speech perception involves not only modality specific processing but also cross-modal integration. Therefore, we identified speech-related brain regions that were relatively activated regions (i.e., increased or decreased) in comparisons not only between each speech cue condition and control condition but also between unimodal speech cue and multimodal speech cue conditions.

We speculate that speech-related regions are differentially correlated in terms of bimodal congruent speech and unimodal speech cues with counterpart noise. Therefore, we hypothesize that brain network components engaged in bimodal congruent speech are different from those engaged in unimodal speech cues with counterpart noise.

To scrutinize the relationship between task-dependent, speech-related regions, we examined changes in topological structures of speech-related networks at every possible threshold using a multiscale framwork. Therefore, we attained a multiscale hierarchical framework, which has successfully modeled group characteristics of certain brain networks (Lee et al., 2012; Singh et al., 2008). The framework of persistent homology uses the all-threshold or threshold-free approach to disclose the topological structure of brain networks (Lee et al., 2012). Thresholding for edge weight has been widely used to determine the skeletal structure of a network for convenient visualization and interpretation. However, thresholding for network construction means to select relatively strong connected edges and to ignore others that possibly carry undisclosed network information (Bassett et al., 2012). Persistent homological methods circumvent this arbitrariness and retain hidden information inherent to the speech-related network. Several previous studies have already set measures for interpreting brain networks (Achard et al., 2006; Bassett and Bullmore, 2006; Reijneveld et al., 2007; Sporns, 2011; Sporns and Zwi, 2004; van den Heuvel et al., 2008); however, small-worldness and scale-free measures still employ *a priori*-determined thresholds of correlation. Moreover, we are interested in the connected components and changes in their relationship using multiscale hierarchical network modeling unlike the graph theoretical approach. Thus, persistent homology frameworks consider the sequence of networks obtained by varying all possible thresholds, a process called filtration.

In this study, we used functional magnetic resonance imaging (fMRI) to compare and visualize these hierarchical clustered structures of brain networks engaged in bimodal and unimodal speech perception. Using this approach, we were able to determine the difference in connected patterns of speech-related brain networks during bimodal audiovisual (AV) speech perception in contrast to auditory speech cues with visual noise (A-VN) or visual speech cues with auditory white noise (V-AN). We were also able to differentiate brain networks that were engaged in V-AN speech perception from those engaged in AV or A-VN speech perception.

Materials and Methods

Participants

In this study, 18 subjects without any history of neurological or psychiatric disorders participated in the fMRI experiments. Three of the subjects were excluded from the study after data collection due to large head-movement. Fifteen subjects (7 women and 8 men; mean age = 24.8 ± 5.1 years) were included. We had behavioral data for 12 subjects, due to technical problems, but not subjects' inattention or drowsiness. These subjects had normal or corrected-to-normal vision, and all were right handed (self-reported). They were native Korean speakers, and had no experience in lip-reading training. All subjects received monetary compensation for their participation and provided written informed consent before the experiment. The fMRI procedure used in this study complied with the ethical guidelines of the local ethics committees.

Task design for speech perception

We used the same experimental design as the one used in a previous positron emission tomography study (Kang et al., 2006). More specifically, we employed the block design paradigm. In all four conditions, moving audiovisual stimuli were used as bimodal inputs; visual stimuli were faces, and auditory stimuli were sounds (Table 1). The auditory speech cue was delivered by a male voice, and the visual speech cue was delivered by mouth movement. The stimuli consisted of simple sentences, which were of 3 to 4 unit lengths in Korean (e.g., ''A sister is a man''). For visual nonspeech cues, a face with unopened mouth movements (like chewing gum) was used, and white noise was used for auditory nonspeech cues. The four conditions were as follows: In the AV condition, both auditory speech sounds and congruent visual mouth movements were presented as bimodal speech cues. In the A-VN condition, speech sounds were delivered along with visual nonspeech stimuli. In the V-AN condition, visual speech cues with auditory nonspeech cues were delivered. In the control (C) condition, white noise and unopened mouth movements were presented. During the AV, A-VN,

Table 1. Stimuli Design

		Auditory stimuli	
		Speech sound	*White noise*
Visual stimuli	Speech mouth	AV	V-AN
	Gum-chewing mouth	A-VN	C

A-VN, auditory speech cues with visual noise; AV, audiovisual; C, control; V-AN, visual speech cues with auditory white noise.

and V-AN conditions, subjects were required to press a button for semantically plausible sentences. The probability of occurrence of semantically true sentences was 50%. In the C condition, subjects were required to alternatively press the button on a trial to maintain attention.

Subjects performed the condition through two sessions; each condition consisted of 8 blocks (4 trials/block), and 32 trials were presented per condition. The order of each block was randomized. All visual stimuli were delivered via an LCD-monitor through a back-projected mirror, and auditory stimuli were delivered binaurally via a headset. Visual face stimuli in the screen subtended an angle of about 6° at the subjects' eye level. During presentation of auditory stimuli, auditory intensity was subjectively matched by each subject. Each block, which consisted of four sentences, was delivered every 28.5 sec. Total time per session was 9 min 45 sec, including eight fixation blocks.

Data acquisition and preprocessing

FMRI data were acquired on a GE 1.5T MRI scanner using a standard whole head coil. We obtained 195 T2*-weighted echo-planar image (EPI) volumes in each of the 20 axial

Table 2. Significant Brain Regions During Speech Perception and 12 Selected Nodes for Speech-Related Brain Networks

Contrasts	*L/R*	*Region*	*BA*	*Talairach* x	y	z	p-*Value* *(FWE-corr)*	k	t-*Value*	*Network node*
AV > C										
	L	*Posterior insula*	*BA13*	*−45*	*−16*	*2*	*0.012*	*380*	*6.83*	*Node 04*
	L	*Superior temporal gyrus*	*BA22*	*−45*	*−22*	*−4*			*5.46*	*Node 05*
A-VN > C										
	L	Superior temporal gyrus	BA22	−53	−13	0	0.002	613	6.15	
	L	*Superior temporal gyrus*	*BA21*	*−53*	*−26*	*−1*			*4.96*	*Node 06*
	L	Superior temporal gyrus	BA21	−49	−6	−10			4.91	
	L	*Lingual gyrus*	*BA18*	*−11*	*−86*	*3*	*0.001*	*631*	*5.65*	*Node 10*
	L	Lingual gyrus	BA19	−22	−67	3			4.90	
	R	*Cuneus*	*BA17*	*8*	*−70*	*9*			*4.63*	*Node 11*
V-AN > C										
	L	*Inferior frontal gyrus*	*BA44/45*	*−45*	*18*	*12*	0.023	1354	*4.97*	*Node 01*
	L	Inferior frontal gyrus	BA44	−45	13	7			4.41	
	L	*Anterior insula*	*BA13*	*−38*	*13*	*5*			*3.86*	*Node 03*
	L	*Anteiror superior temporal gyrus*	*BA38*	*−45*	*3*	*−9*			*3.84*	*Node 07*
AV > A-VN										
				N.S.						
AV > V-AN										
				N.S.						
A-VN > AV	R	Cuneus	BA17	2	−85	7	0.000	948	6.22	
	L	Cuneus	BA18	−9	−76	19			5.48	
	L	Cuneus	BA18	−16	−86	25			5.09	
A-VN > V-AN										
	L	Cuneus	BA17	−9	−85	10	0.000	1685	6.92	
	R	Cuneus	BA17	8	−66	9			6.92	
	L	Cuneus	BA18	−2	−84	21			6.18	
V-AN > AV										
	R	*Premotor*	*BA6*	*49*	*−7*	*37*	*0.001*	*611*	*6.00*	*Node 02*
	R	Premotor	BA6	37	−10	47			5.77	
	R	Premotor	BA6	28	−14	52			4.86	
V-AN > A-VN										
	R	*Fusiform gyrus*	*BA37*	*49*	*−55*	*−13*	*0.028*	*1489*	*6.27*	*Node 09*
	R	Middle occipital gyrus	BA19	30	−71	14			4.96	
	R	*Middle temporal gyrus*	*BA37*	*45*	*−63*	*−3*			*4.94*	*Node 08*
	R	*Anterior cingulate gyrus*	*BA32*	*23*	*16*	*30*	*0.040*	*1366*	*5.01*	*Node 12*
	R	Middle frontal gyrus	BA6	43	4	40			4.96	

The significance was set to a corrected level of $p < 0.05$ for multiple comparisons at the cluster level. To identify regions, we converted coordinates from the MNI to Talairach space using icbm2tal transform (Lancaster et al., 2007), and used the Talairach Daemon program to identify Brodmann's area. These local maxima are located more than 8 mm apart. The criteria for node selection were one node of local maxima of cluster or a different Brodmann's region. We extracted the time course of each node within a 4 mm radius. These details are described in the Materials and Methods section of our main text.

Selected network nodes are presented in italics.

BA, Brodmann area; FWE, familywise error rate; MNI, Montreal Neurological Institute.

planes parallel to the anterior and posterior commissure line (5 mm thick without gap, matrix size = 64 × 64, voxel size = 3.75 × 3.75 × 5 mm, TR = 3000 msec, TE = 60 msec, and flip angle = 90°). A high-resolution T1-weighted spoiled gradient-recalled 3D MRI sequence was obtained for anatomical reference. Functional images (EPIs) were preprocessed using Statistical Parametric Mapping (SPM 8; University College of London, London, United Kingdom). The first five volumes of the functional image were discarded to allow for signal stabilization. Before the preprocessing step, we detected and fixed bad slices using ARTrepair (http://cibsr.stanford.edu/tools/human-brain-project/artrepair-software.html). It was used to remove noise spike and to repair bad slices with a particular scan, and bad slices were repaired by interpolation between adjacent slices. EPI functional images were realigned for movement correction using rigid-body transformation. T1-weighted anatomical images were co-registered to the mean EPI using mutual information. To register to Montreal Neurological Institute (MNI) standard space, the co-registered T1-weighted image was used to calculate the normalization parameter by an affine and nonlinear algorithm, which was applied to all EPIs. To increase the signal-to-noise ratio, spatially normalized EPIs were smoothed with an 8 mm full-width at half-maximum Gaussian kernel. In the first-level individual analysis, the fMRI signal was regressed out of the confounding effect of movement using six rigid-body motion parameters. The high-pass filter cutoff was 128 sec to remove the effects of slow signal drifts. These basis functions can generate a residual forming matrix as confounds in the design matrix, thus removing these factors in the generalized linear model (GLM). Statistical parametric maps of the *t*-statistics were generated in individual analysis using GLM. In the second-level random-effect analysis, each contrast image was used to test group significance levels. Statistical maps were thresholded at a significant level of $p < 0.001$ with a minimum extent threshold of 20 contiguous voxels. Significance was set to $p < 0.05$ at the cluster level after correction for multiple comparisons (Friston et al., 1994) (Table 2).

Brain network construction involved in speech perception

Weighted matrix based on interregional correlation. The first step in forming a brain network is to define a node. For the construction of the brain network involved in speech perception, we identified 12 regions of interest (ROIs) as nodes based on the results of a second-level analysis through a random-effect model (Fig. 1 and Table 2). The criterion for node selection was as follows: The node had to be one of the local maxima of each significant cluster. However, if the significant cluster comprised more than one Brodmann area (BA), we considered them as separate nodes. In addition, if there were any anatomically identical nodes across conditions, we included only the node that had greater statistical significance (*t*-value) among them.

Second, we defined edges linking nodes as follows: We extracted the time series of each node for the individual fMRI that calculated weighted mean signal intensity as the first principal component of all voxel time series within a sphere (radius 4 mm) at coordinates of each node. We analyzed the time courses of only the last seven scans for each speech condition block considering a hemodynamic delay of 6 sec. The seven time courses were then averaged within each block, yielding time courses of four blocks for each session, speech condition, and subject. Since the number of blocks in each speech condition was eight per individual, it was not enough to estimate correlation coefficients between time courses of ROIs. Thus, we concatenated an individual's eight data points as a fixed-effect model, so that the data matrix contained 120 (8 blocks × 15 subjects) rows and 12 (ROIs) columns for each speech condition in each group level.

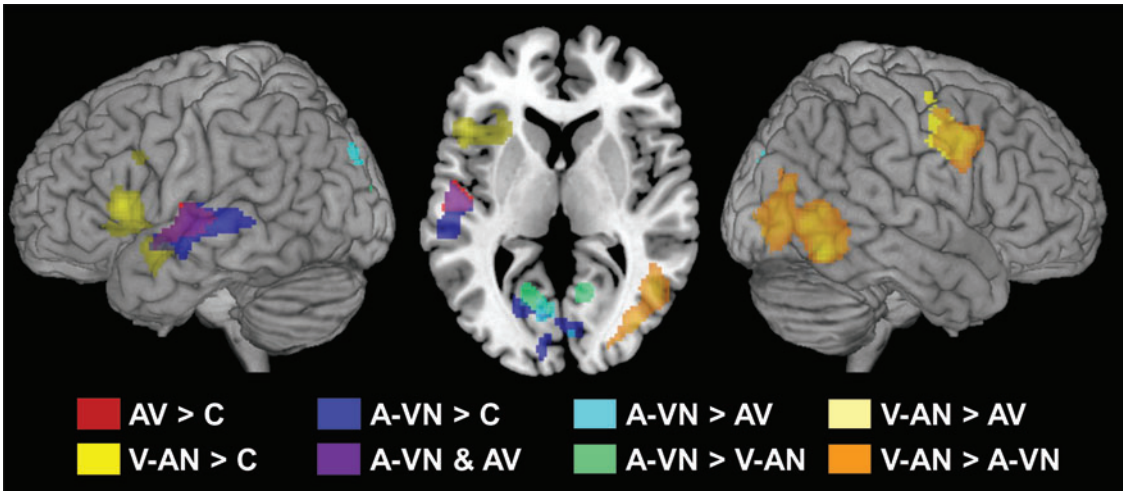

FIG. 1. Significant activation map of each speech perception condition. Brain regions showing significant activation in each contrast between speech conditions are overlaid (corrected $p < 0.05$ at cluster level). Red, blue, and yellow represent the activation map during audiovisual (AV), auditory speech cues with visual noise (A-VN), and visual speech cues with auditory white noise (V-AN) conditions compared with the control condition, respectively. Purple indicates common areas of activation during AV and A-VN conditions compared with the control condition. Cyan and light green represent greater activated regions during the A-VN condition compared with AV or V-AN conditions, and light yellow and orange during the V-AN compared with AV or A-VN conditions, respectively. Color images available online at www.liebertpub.com/brain

Pearson correlation coefficients were computed across the 120 data points between ROIs for each speech condition, yielding a {12×12} matrix of speech-related brain network for each condition. In this study, we separately analyzed positive and negative correlation values. The {12×12} distance matrix of each condition was computed by one minus the absolute correlation coefficient (Fig. 2C).

Speech perception networks in terms of persistent homology. For construction of the connectivity matrix, we applied a new multi-scale weighted network modeling called graph filtration to the distance matrix (Lee et al., 2012). To overcome arbitrariness of thresholding, using filtration, we observed topological changes in the network over the entire range of thresholds. The binary network was obtained by connecting two nodes if their distance was smaller than the threshold, ε. The binary network with the smaller ε was a sub-network of the network with the larger ε. So, if we estimated binary networks by increasing thresholds, we obtained the nested sequence of binary networks (Fig. 2D). This procedure is known as graph filtration. During filtration, we can extract the topological invariant feature, the Betti number, which defines the shape of the topological space in algebraic topology (Carlsson and Memoli, 2010; Edelsbrunner et al., 2008). We were especially interested in the 0th Betti number that counts the number of connected components in a network (Fig. 2E). In graph filtration, when the initial threshold was equal to 0, all nodes were disconnected and the number

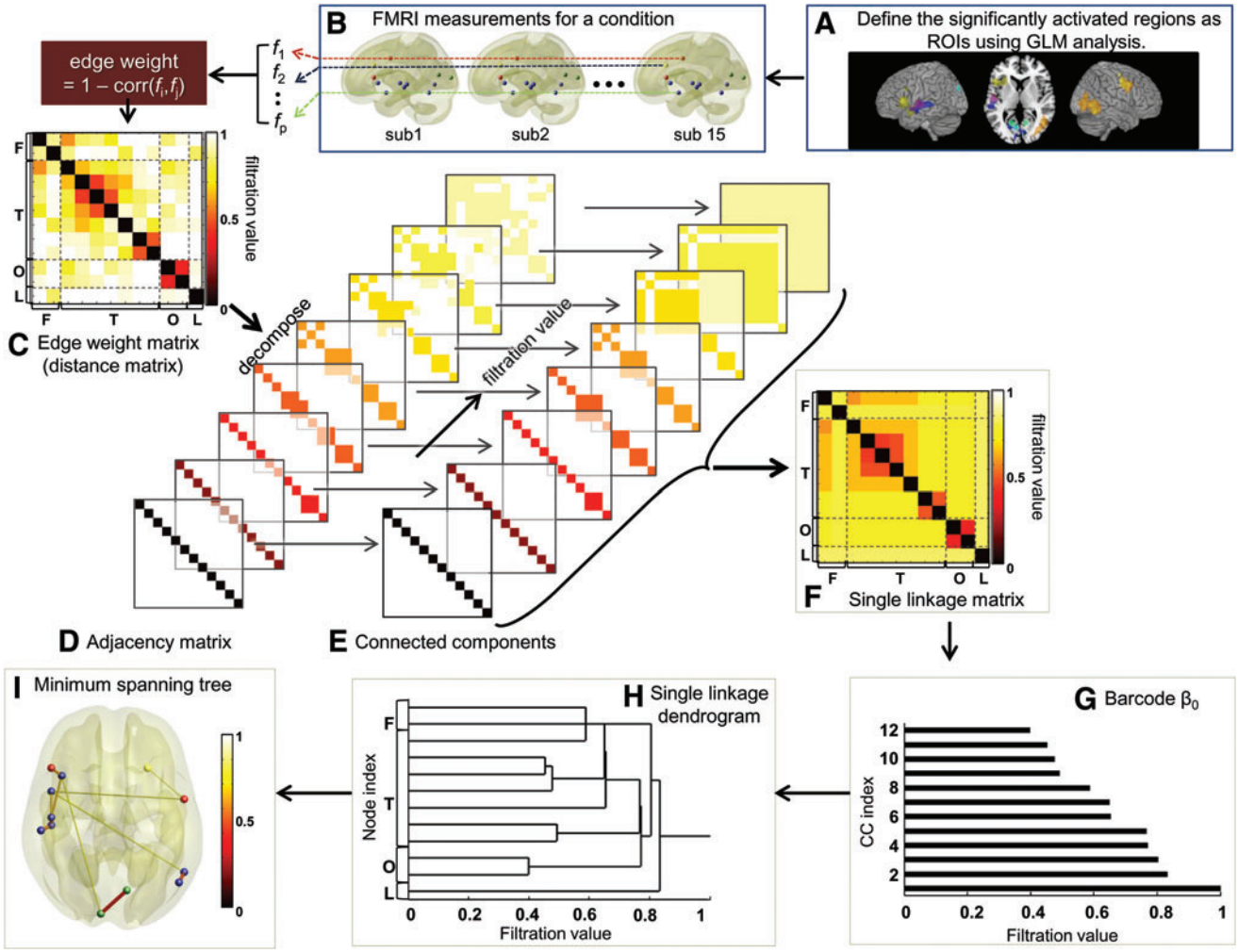

FIG. 2. Schematic overview using graph filtration for the functional connectivity analysis. To select a network node, **(A)** we defined the significantly activated areas (corrected $p < 0.05$ at cluster level) at the second-level analysis. **(B)** For these 12 regions of interest (ROIs), the value of speech perception effect in each condition was extracted from the functional magnetic resonance imaging (fMRI) time courses of each individual's raw images. **(C)** The distance (one minus positive or |negative| correlation coefficient) matrices of the 12 ROIs was computed for each condition. **(D)** Given the threshold ε, the corresponding adjacency matrices were composed by filtration. **(E)** During filtration, the topological invariant feature was extracted, called 0th Betti number, which counts the number of connected components in the network. Connected components were represented using varying colors and the remaining regions as white at each threshold. **(F)** Single linkage matrices were formed for each condition to express the network characteristics as a single matrix. **(G)** Changing patterns of 0th Betti numbers during filtration were visualized by a barcode. **(H)** We rearranged the bars in the barcode according to the node indices of location in the brain. Then, this yielded a single linkage dendrogram for each condition. **(I)** Edge connections are visualized as a minimum spanning tree, which is the nonredundant skeleton of each network. F, frontal; L, limbic; O, occipital regions; T, temporal. Color images available online at www.liebertpub.com/brain

of connected components was equal to the number of nodes. By increasing the threshold, nodes started becoming connected with each other and the number of connected components decreased. Once all nodes were connected, a single connected component remained and graph filtration continued to not have any change in the 0th Betti number. A change of the 0th Betti number during filtration could be visualized by the barcode, where vertical and horizontal axes represented the indices of connected components and the threshold, respectively (Fig. 2G). One bar represented each connected component, and the number of bars at the specific threshold was the same as the number of connected components. Thus, algebraic topology seeks the global shape of translation-, rotation-, and scale-invariant networks without considering node location; however, in our investigation, nodes were regions in the brain network. To determine which node was connected to which node, we rearranged the bars in the barcode according to the node indices of location. This yielded a single linkage dendrogram, which is a widely used agglomerative hierarchical cluster (Fig. 2F). We represented the change of connected network structures in two different forms: a single linkage matrix and minimum spanning tree (Fig. 2F, I). The single linkage matrix was used to describe the characteristics of networks and to yield the difference between networks. The minimum spanning tree served as the nonredundant skeleton of the network and was appropriate for visualization to yield tree-like outputs.

Comparisons of brain networks for speech perception conditions. To compare networks between two speech perception conditions, differences between d_{ij}, which is a value of each entry of single linkage matrices, were compared with the null distribution (Fig. 3). Nonparametric paired permutation testing was used to test the probability that the observed difference between two different conditions occurred by chance (the null hypothesis). To make a null distribution of single linkage matrices (e.g., AV and A-VN conditions) of d_{ij}, we went back to the data matrix consisting of 120 data points by 12 ROIs where the 120 data points comprised eight data points for each of the 15 subjects. For two conditions to be compared with each other, 16 data points for each subject were randomly assigned into two pseudo-condition data matrices (e.g., pseudoAV and pseudoA-VN). On each

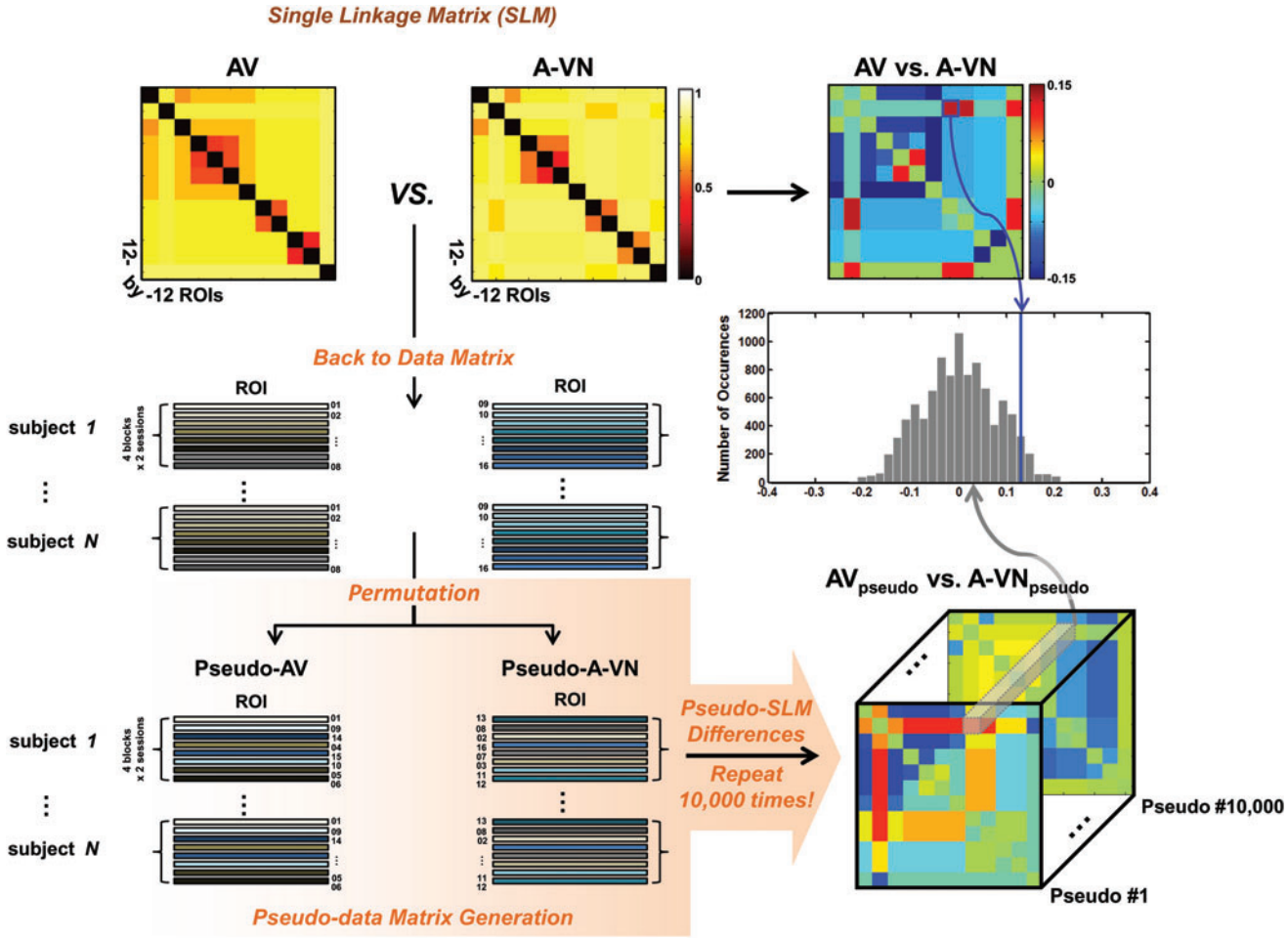

FIG. 3. Nonparametric paired permutation testing based on pseudo-conditions made 10,000 times. To make null distributions of single linkage matrices, d_{ij} (e.g., AV and A-VN conditions) of distance, fMRI values of each ROI (i: 1–12) of randomly chosen individuals were exchanged to yield two pseudo-condition (pseudoAV and pseudoA-VN) data matrices. On each pseudo-data matrix, single linkage matrices were generated. The difference of each d_{ij} was compared with the null distribution. Significance was set as $p < 0.01$ (two tailed). Color images available online at www.liebertpub.com/brain

pseudo-data matrix, single linkage matrices were generated and differences between two pseudo-conditions were computed. This procedure was repeated 10,000 times to generate the null distribution of each d_{ij}. Significance was set as $p < 0.01$ (two-tailed) by permutation test.

Results

Behavioral performance

Behavioral responses of the speech conditions were analyzed for 12 participants. The behavioral performances of three conditions (AV, A-VN, and V-AN) were significantly different ($F[2, 32] = 18.1$, $p < 0.0001$). Percentages of correct response were 75%, 57%, and 31% for the AV, A-VN, and V-AN conditions, respectively. Subjects had significantly lower rates of correct answers in the V-AN condition compared with the other two speech perception conditions [*post hoc* analysis, Scheffe's *t*-test ($p < 0.005$)].

Regions of interest selected by their engagement during three speech perception conditions

We identified speech-related brain regions with areas showing relatively increased or decreased activation not only between each speech condition and control condition but also between bimodal and unimodal speech conditions (Fig. 1 and Table 2). The left superior temporal region was involved in all types of speech perception. The anterior part of superior temporal regions was significantly activated during the V-AN condition. Additional regions showing significant activation were observed in the posterior insula during the AV condition, visual areas during the A-VN condition, and inferior frontal gyrus and anterior insula during the V-AN condition compared with the control condition. In comparisons between speech conditions, there was no significant activation during the AV condition compared with the other two speech conditions, respectively. Bilateral cuneus were significantly involved in the A-VN condition compared with AV or V-AN conditions. During the V-AN condition, significant activation was observed in the right premotor (BA6) region compared with during the AV condition and in the fusiform, middle occipital, middle temporal, anterior cingulate, and middle frontal (BA6) regions compared with during the A-VN condition.

To construct a task-dependent functional network, we selected 12 ROIs consisting of both conjoint and disjoint activated regions (Table 2 and Fig. 2A). Selected regions were two frontal, seven temporal, one limbic, and two occipital regions as follows: left inferior frontal gyrus (BA44/45) and right premotor (BA6), left anterior and posterior insular

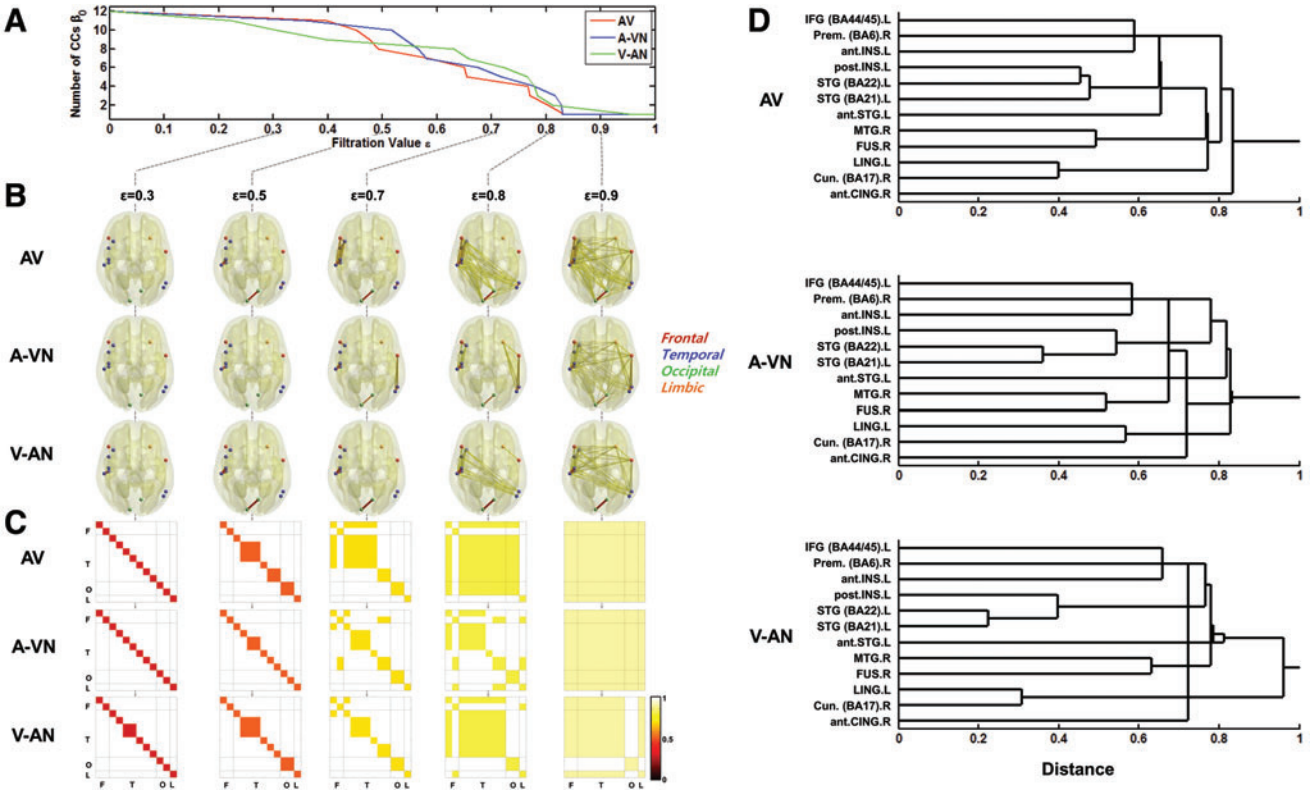

FIG. 4. Changing pattern of the connected component during graph filtration in positive coupling. **(A)** The numbers of connected components depending on the filtration value, ε, are presented as barcodes for speech perception conditions. The red line represents the AV condition, blue for the A-VN condition, and green for the V-AN condition. The vertical axis shows the number of connected components, and the horizontal axis shows filtration values. **(B)** Connected edges of the brain are shown according to the filtration values, 0.3, 0.5, 0.7, 0.8, and 0.9. **(C)** Connected components of speech perception conditions according to the same filtration vales as **(B)**. The darker the color is, the stronger the connection and the earlier merging occurs during filtration. **(D)** Single linkage dendrograms of speech perception conditions are displayed. The merging pattern of the brain network is displayed for each speech perception condition during filtration. F, frontal; L, limbic; O, occipital regions; T, temporal. Color images available online at www.liebertpub.com/brain

cortices, left superior temporal gyri (BA22, BA21, and BA38), right middle temporal gyrus (BA37), right fusiform gyrus, left lingual gyrus, right cuneus, and anterior cingulate cortex.

Network connectivity pattern: a positive coupling

We examined changes in the numbers of connected components during filtration in each speech condition (Fig. 4A). We also visualized the connected regions at five selected distance thresholds, 0.3, 0.5, 0.7, 0.8, and 0.9 from the left to the right panels in Figure 4 and Supplementary Movies S1 and S2 (Supplementary Data are available online at www.liebertpub.com/brain). During filtration, the clustering patterns of earlier or later merging of regions were different in each speech perception condition. The more tightly coupled areas were found to show earlier merging during filtration of the correlation matrix of brain regions. The earliness or lateness of merging of the connected regions during filtration represents the hierarchical clustered structures associated with single linkages of brain networks.

The connected components in terms of positive relationships were observed between speech-related regions. Interestingly, on the single linkage matrix representation by each filtration value, in terms of connectivity strength, several of the same connected components were formed at different filtration values among speech conditions (Fig. 4B, C). We can visualize details of the connected pattern by barcodes representing the number of connected components (Fig. 4A) and dendrogram indexing of connected components during filtration (Fig. 4D).

During the AV condition, the left temporal component and left fronto-temporal component were tightly coupled (Fig. 4C). The changes of connected components can be visualized during filtration by dendrogram (Fig. 4D). More specifically, a visual component was observed in the earlier filtration value after the connected components were observed in the order of the right middle temporal fusiform, the left temporal component [superior temporal (BA22)-left posterior insula-left superior temporal component, and the left inferior frontal (BA44/45)-the left insula component]. Merging between the components spread from fronto-temporal and occipital components.

During the A-VN condition, the connected component of the left superior temporal region (BA21 and BA22) was formed at an earlier filtration value. Connected components were observed in the visual, right middle temporal fusiform, left superior temporal-left posterior insula, and left inferior frontal-left anterior insula regions at filtration values from $\varepsilon = 0.5$ to $\varepsilon = 0.6$. Next, connected components of the right premotor (BA6)–right temporal yielded and then merged to the right anterior cingulate. After $\varepsilon = 0.6$, sporadic clusters showed up from fronto-temporal-occipital regions.

During V-AN, the first coupled cluster of brain regions was the superior temporal component (BA21 and BA22)

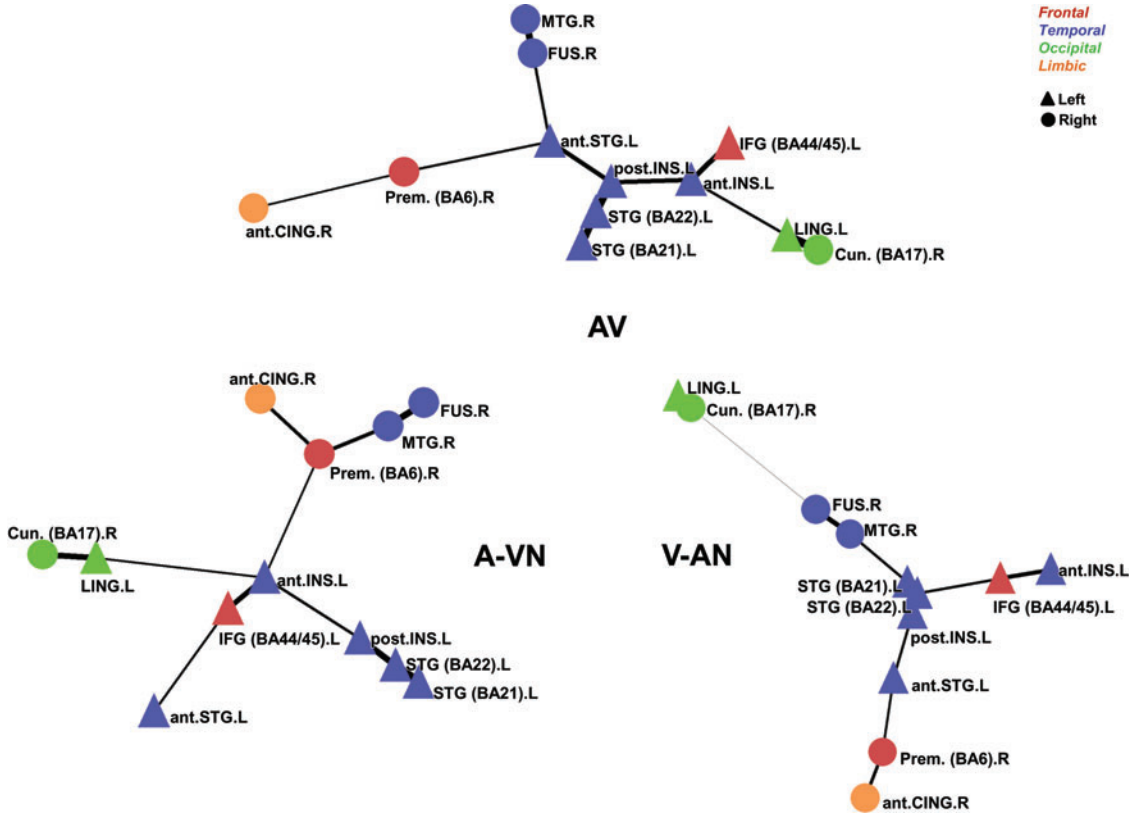

FIG. 5. Minimum spanning tree of the brain network for each speech perception condition in positive coupling. In positive coupling, similar connected components were commonly observed as follows: visual, superior temporal, right middle temporal fusiform, and left inferior frontal-anterior insula components. Red color denotes frontal; blue, temporal; green, occipital; and orange, limbic regions. The circle denotes the right hemisphere, and the triangle indicates the left. Color images available online at www.liebertpub.com/brain

and visual component at $\varepsilon = 0.22$ and $\varepsilon = 0.31$, respectively. As filtration proceeded, superior temporal regions were coupled to the left posterior insula and the connected component of the right middle temporal fusiform at $\varepsilon = 0.40$, and the left inferior frontal-left anterior insula components were yielded at $\varepsilon = 0.66$. The right premotor–the right anterior cingulate component was then formed. The merging of connected components spread from frontal-temporal and occipital regions. At last, the visual component was merged to the fronto-temporal components after $\varepsilon = 0.96$, and the connection slowly grew to yield a final giant component.

The minimum spanning tree also represents a network that has the minimum number of edges, which does not allow redundant connections within connected clusters (Alexander-Bloch et al., 2010; Ciftci, 2011; Lee et al., 2006). The minimum spanning trees of the AV, A-VN, and V-AN conditions are shown in Figure 5. Notably, in positive relationships, similar connected components were observed among AV, A-VN, and V-AN speech conditions. For example, the visual component, superior temporal component, including the posterior insula, inferior frontal (BA44/45)-anterior insula component, and right middle temporal-fusiform component, were commonly yielded.

Network connectivity pattern: a negative coupling

Connected components in negative coupling were observed between speech-related regions. By barcode, the number of connected regions decreased earlier during the V-AN condition (Fig. 6A), which meant that the correlation between regions was stronger and coupling between speech-related brain regions was tighter (Fig. 7). During the AV condition, the connected component of visual and right middle temporal/fusiform was extended to temporal and frontal regions (Fig. 6B–D). During the A-VN condition, a temporo-visual component was observed. Later, the left inferior frontal region was merged to the temporo-visual component. During the V-AN, unlike the other two speech conditions, visual and right temporal components were formed earlier; later, the left inferior frontal region was included at $\varepsilon = 0.79$. The clusters that included temporal and occipital regions then merged to the right premotor region, with the left anterior superior temporal region merging shortly thereafter.

Differential connectivity engaged in bimodal speech perception from unimodal speech condition

First, in positive coupling, significant differences were examined between single linkage matrices at $p < 0.01$ using the

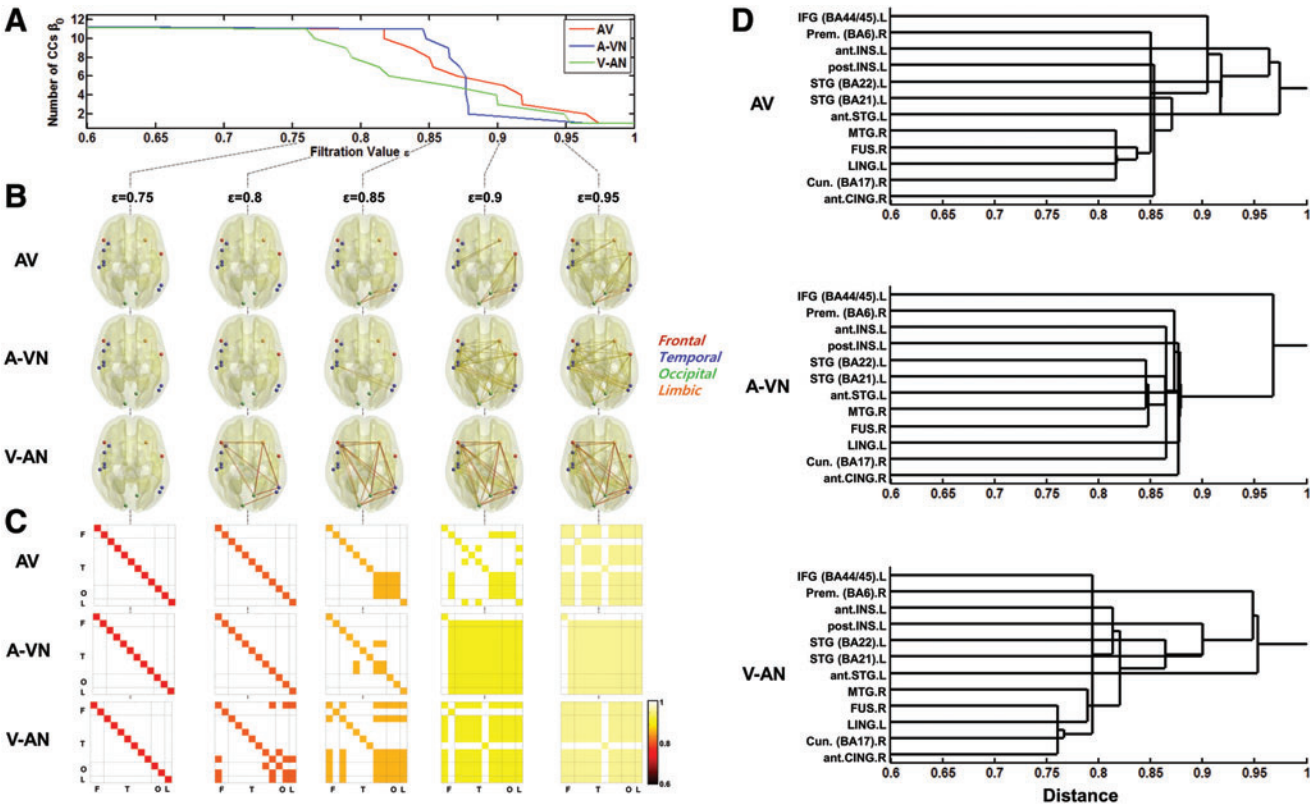

FIG. 6. Changing pattern of the connected component during graph filtration in terms of negative relationship. **(A)** The numbers of connected components depending on the filtration value, ε, are presented as barcodes for speech perception conditions. Red line represents the AV condition, blue for the A-VN condition, and green for the V-AN condition. The vertical axis shows the number of connected components, and the horizontal axis shows filtration values. **(B)** Connected edges of the brain are shown according to the filtration values, 0.75, 0.8, 0.85, 0.9, and 0.95. **(C)** Connected components of speech perception conditions according to the same filtration vales as **(B)**. The darker the color is, the stronger the connection and the earlier it merges during filtration. **(D)** Single linkage dendrograms of speech perception conditions are displayed. The merging pattern of the brain network is displayed for each speech perception condition during filtration. F, frontal; L, limbic; O, occipital regions; T, temporal. Color images available online at www.liebertpub.com/brain

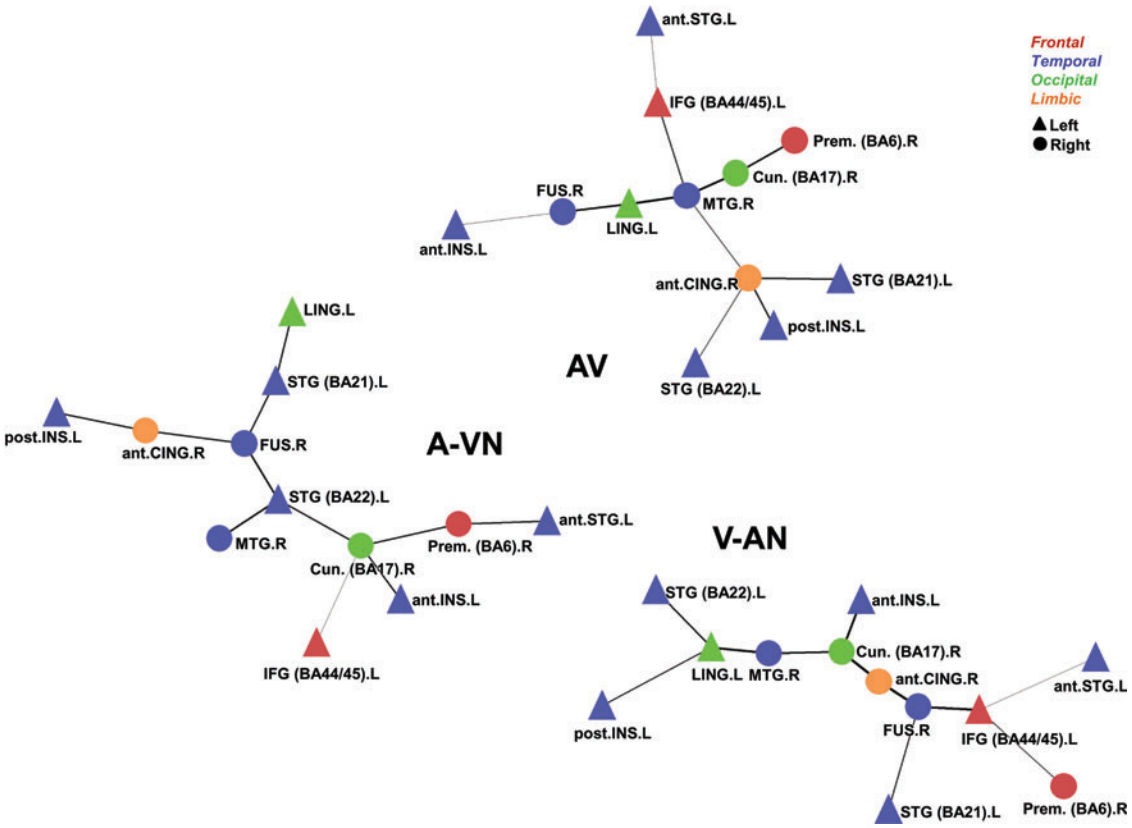

FIG. 7. Minimum spanning tree of the brain network for each speech perception condition in terms of negative relationship. Left inferior frontal region was tightly connected to the visual region, including the right middle temporal and fusiform during the V-AN condition rather than the A-VN condition. Red color denotes frontal; blue, temporal; green, occipital; and orange, limbic regions. The circle denotes the right hemisphere, and the triangle indicates the left. Color images available online at www.liebertpub.com/brain

nonparametric permutation test (Fig. 3). Since a single linkage matrix assumes that connected brain regions act as a single cluster (regardless of distant in space), during the AV condition the single linkage between the left anterior superior temporal and left anterior insula was stronger than during the A-VN condition (Fig. 8). The single linkage between right premotor (BA6) and visual areas (left lingual and right cuneus) was also stronger than during the V-AN condition. No significant difference of positive coupling was observed between A-VN and V-AN conditions.

Second, in negative coupling, no significant difference of single linkage distance was observed between the AV and unimodal speech conditions at $p < 0.01$. During the V-AN condition, the left inferior frontal (BA44/45) was significantly earlier connected to the right anterior cingulate, left anterior insula, and visual regions, including left middle temporal and right fusiform gyri (Fig. 9).

Discussion

We investigated the changes of network topology among speech-related brain regions using multiscale hierarchical network modeling during speech perception. We found that, regardless of bimodal or unimodal speech cues, similar connected components were observed among speech-related brain regions; however, the merging pattern of each connected component was different. During a congruent audio-visual-speech condition, tighter positive couplings were observed in the left anterior temporal component and the right premotor-visual component compared with auditory or visual speech condition, respectively. Noisy auditory or visual stimuli hampered positively tighter integration within this network. Interestingly, during lip-reading, the left inferior frontal (BA 44) region was negatively connected to right anterior cingulate and visual regions compared with its connectivity during the auditory speech condition.

We suspect that congruent audiovisual speech perception needs smaller areas of neural substrates than auditory speech perception with irrelevant visual cues (Kang et al., 2006). While no area outside of the temporal cortex was activated during AV speech perception, during both A-VN and V-AN conditions, additional areas were activated in the visual and left inferior frontal regions. As for audiovisual integration, we could only conclude that superior temporal recruitment showed under-additivity during audiovisual speech perception (Kang et al., 2006). However, when we scrutinized networks of speech-related regions, the left temporal component was found to be tightly coupled, especially the left anterior superior temporal and anterior insula, compared with the auditory speech condition. The left anterior superior temporal region is involved in intelligible speech, which acts as an interface between speech-sound representations and word meaning (Scott et al., 2000; Spitsyna et al., 2006). The left anterior insula has been known to be responsible

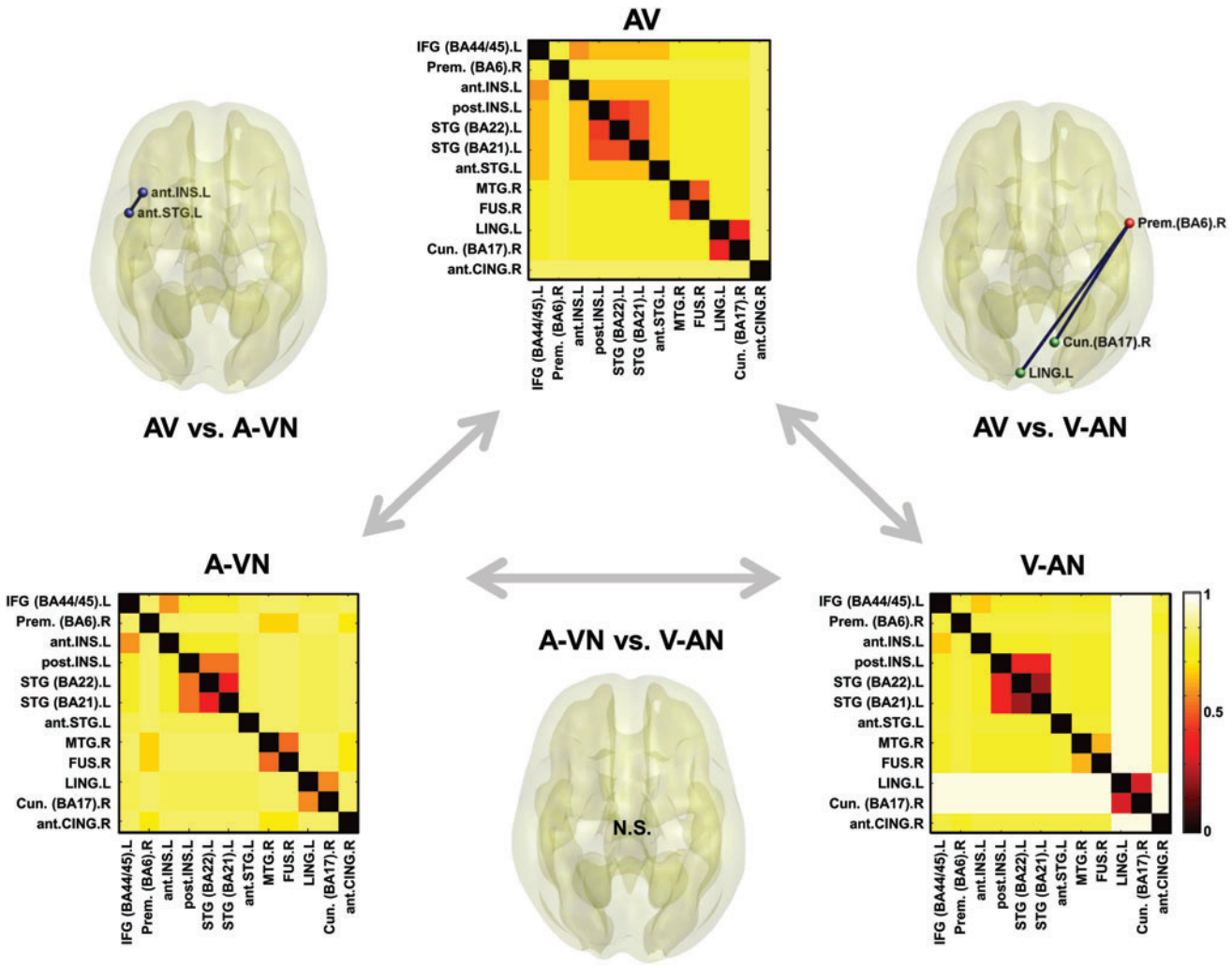

FIG. 8. Differences between single linkage matrices of speech perception conditions in positive coupling. We performed nonparametric permutation test at $p < 0.01$. Single linkage matrices are shown at the top for the AV condition, at the left for the A-VN condition, and at the right for the V-AN condition. The darker the color is, the stronger (tighter and earlier merge during filtration) the connection is. In 3D-rendered brain images, significantly different connections between these speech perception conditions are displayed: AV versus A-VN in the top-left, AV versus V-AN in the top-right, and A-VN versus V-AN in the bottom-center. Color images available online at www.liebertpub.com/brain

for speech motor control such as coordinating speech articulation (Ackermann and Riecker, 2004, 2010). Although two anterior temporal regions were activated during V-AN condition compared with control condition, their relationship was different between AV and A-VN conditions, but not between AV and V-AN conditions. Therefore, positively tight coupling of the left anterior temporal component can reflect speech-sound meanings and speech motor control based on visual speech cues for efficient audiovisual speech integration.

The motor cortex, especially the left motor cortex, plays an active role in speech perception, as it is involved in the discrimination of speech sounds (Bartoli et al., 2013; Osnes et al., 2011). This left motor system seems to contribute to speech perception under adverse listening conditions, such as under conditions of degraded or incomplete speech but with recognizable phonetic signals (Osnes et al., 2011). On the other hand, the right premotor cortex, as detected by the subtractive design of our study, was suppressed in the AV condition compared with V-AN (Fig. 1 and Table 2) or control (data not shown) conditions. This might be associated with a suppressive role for recognizable phonetic signals during AV conditions. Although there was no statistical difference, coupling of the right premotor and visual regions during the AV condition was similar to that during the A-VN condition, but not during the V-AN condition (Fig. 4). It should be noted that strong or weak connectivity is different from increased or decreased activation. The tight coupling of the right premotor and visual regions during the AV could be associated with a strong efficient relationship between visual input and the speech motor system without mapping speech onto meaning (Krieger-Redwood et al., 2013).

The visual stimuli we used in this investigation was silent speech, and evoked activation in the left superior temporal sulcus/gyrus (Calvert et al., 1997). Activation in the left

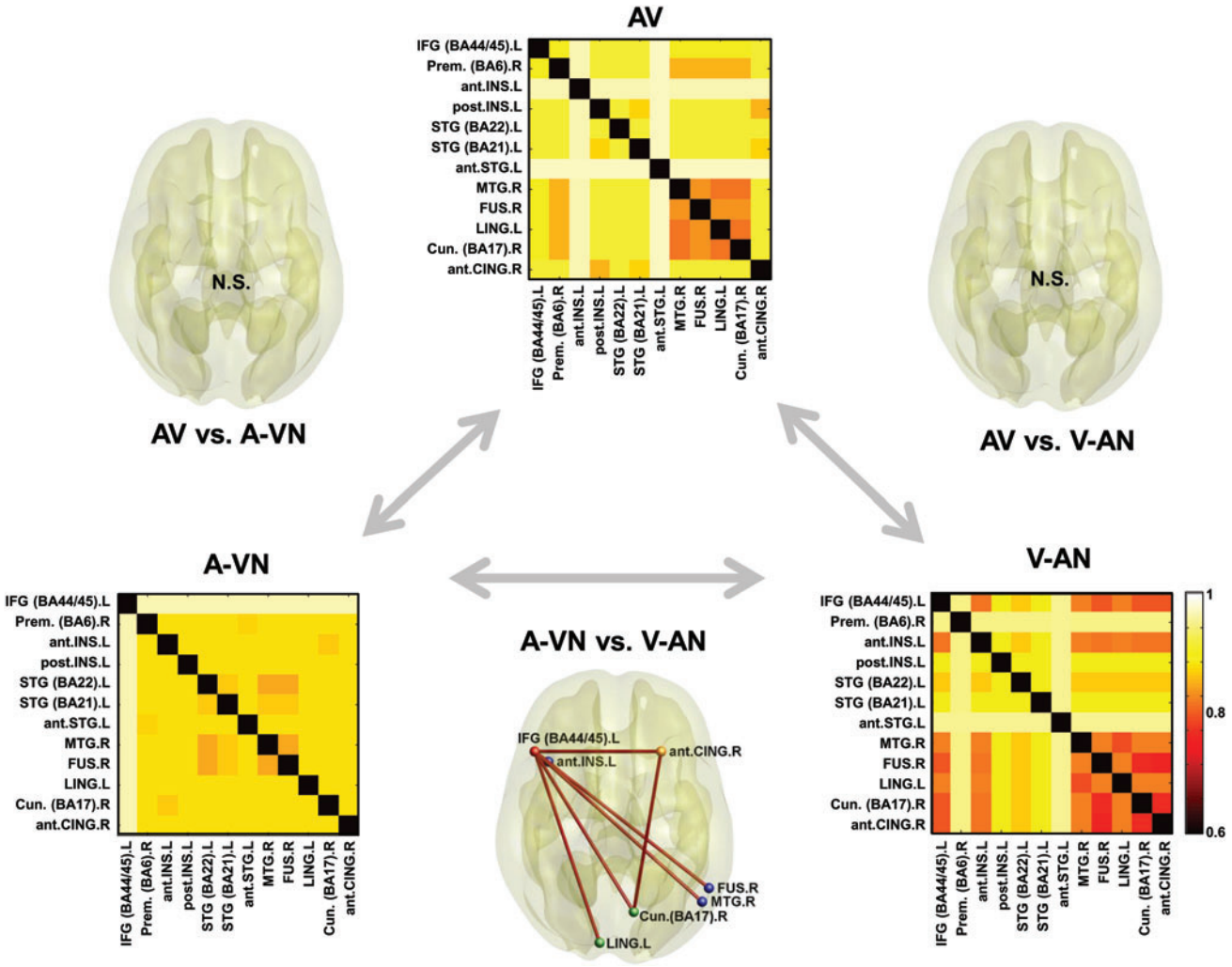

FIG. 9. Differences between single linkage matrices of speech perception conditions in negative coupling. We performed nonparametric permutation test at $p < 0.01$. Single linkage matrices are shown at the top for the AV condition, at the left for the A-VN condition, and at the right for the V-AN condition. The darker the color is, the stronger (tighter and earlier merge during filtration) the connection is. In 3D-rendered brain images, significantly different connections between these speech perception conditions are displayed: AV versus A-VN in the top-left, AV versus V-AN in the top-right, and A-VN versus V-AN in the bottom-center. Color images available online at www.liebertpub.com/brain

inferior frontal and ventral motor regions during V-AN reminded us of reports that the left Broca's area is activated by hand or mouth movements as well as during speech perception (Rizzolatti and Craighero, 2004) and that neurons respond to actions in the left inferior frontal region as well as in the left superior temporal sulcus. Interestingly, it has been reported that individuals who are incapable of lip-reading utilize a strategy to mimic the speaker's lip movement, which activates the left ventral inferior frontal region (Nishitani and Hari, 2002). The same strategy is also used for guessing ambiguous words from unrecognizable sentences during V-AN, unlike during the A-VN condition. To identify a speech cue, it is possible that a negative coupling is needed to exert a more attentional and inhibitory process. The anterior cingulate was negatively connected to left inferior frontal gyrus and visual regions, which may be associated with conflict monitoring for audio-visual stimuli (Wang et al., 2005). The semantic judgment of auditory speech in AV and A-VN conditions is very easy; however, during visual speech cues with white noise, most subjects having had no lip-reading training suffered difficulty in speech perception. Although a negative coupling does not mean inhibitory circuits are involved in speech perception, there might be an intrinsic inhibitory mechanism of nonspeech visual input to aid in successful lip-reading. Moreover, not only noisy auditory input but also nonspeech visual cues, that is, face perception or other visual motion perception could have an inhibitory influence on lip-reading. It has been shown that middle temporal, visual motion areas are activated during nonlexical lip-reading (Paulesu et al., 2003). Thus, the changing pattern of connectivity during filtration would have been affected by the inhibitory anti-correlative characteristics of this condition.

In this study, we compared single linkage matrices of speech perception of bimodal or unimodal speech cues with counterpart irrelevant noise using the permutation test, which disclosed the differences between connectivity patterns of speech conditions. Connected components (and

their numbers) are the simplest topological features of networks represented by barcodes or dendrograms equivalent to the 0th Betti number chain complex of topological data representations (Lee et al., 2012). Graph filtration along varying thresholds of distance for a single linkage matrix allowed us to discover hierarchical structures of correlation matrices between speech-related areas. By observing the changing pattern of these matrices during filtration, we could identify which regions were more tightly coupled during specific cognitive tasks. Thus, this method could help elucidate network characteristics along all threshold distances between task-related brain regions.

Thresholds are chosen arbitrarily and thresholding favors strong edges; thus, interpretation is biased to prominently coupled brain regions (Bassett et al., 2012). Moreover, information within the weighted matrix will not be fully used in the characterization of activated networks during a specific cognitive task. This is especially the case if the number of network edges are different; comparisons become harder as the calculated network measures get affected by the number of edges (van den Heuvel et al., 2008). Using our graph filtration method, the network was not thresholded, allowing us to examine the changing pattern of brain connectivity over all thresholds.

We interpret later merging areas engaged in the audiovisual speech perception as less tightly coupled, but as a definitely involved inter-correlated network. Here, the term inter-correlated network represents connected components during filtration. We need to pay attention to the fact that earliness or lateness of merging during filtration does not mean the temporal priority or posterity of neural interaction. Although the connected component represents one type of topological feature, which we do not assume is the best metric, the examination of other Betti number, holes or voids would be a future study. This is the first study of the application of the persistent homological framework to speech-related brain networks.

Understanding the connectivity among regions involved in speech perception helps explain how processing of speech information is handled in brain networks. This study provides an example that connectivity analyzed using the topological framework of persistent homology avails itself for the elucidation of mechanisms underlying auditory or visual speech perception.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2011-0030815) and also supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2013R1A1A2057782).

Author Disclosure Statement

No competing financial interests exist.

References

Abrams DA, Ryali S, Chen T, Balaban E, Levitin DJ, Menon V. 2013. Multivariate activation and connectivity patterns discriminate speech intelligibility in Wernicke's, Broca's, and Geschwind's areas. Cereb Cortex 23:1703–1714.

Achard S, Salvador R, Whitcher B, Suckling J, Bullmore E. 2006. A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. J Neurosci 26:63–72.

Ackermann H, Riecker A. 2004. The contribution of the insula to motor aspects of speech production: a review and a hypothesis. Brain Lang 89:320–328.

Ackermann H, Riecker A. 2010. The contribution(s) of the insula to speech production: a review of the clinical and functional imaging literature. Brain Struct Funct 214:419–433.

Alexander-Bloch AF, Gogtay N, Meunier D, Birn R, Clasen L, Lalonde F, Lenroot R, Giedd J, Bullmore ET. 2010. Disrupted modularity and local connectivity of brain functional networks in childhood-onset schizophrenia. Front Syst Neurosci 4:147.

Bartoli E, D'Ausilio A, Berry J, Badino L, Bever T, Fadiga L. 2013. Listener-speaker perceived distance predicts the degree of motor contribution to speech perception. Cereb Cortex [Epub ahead of print] DOI: 10.1093/cercor/bht257

Bassett DS, Bullmore E. 2006. Small-world brain networks. Neuroscientist 12:512–523.

Bassett DS, Nelson BG, Mueller BA, Camchong J, Lim KO. 2012. Altered resting state complexity in schizophrenia. Neuroimage 59:2196–2207.

Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SC, McGuire PK, Woodruff PW, Iversen SD, David AS. 1997. Activation of auditory cortex during silent lipreading. Science 276:593–596.

Carlsson G, Memoli F. 2010. Characterization, stability and convergence of hierarchical clustering methods. J Mach Learn Res 11:1425–1470.

Chu YH, Lin FH, Chou YJ, Tsai KW, Kuo WJ, Jaaskelainen IP. 2013. Effective cerebral connectivity during silent speech reading revealed by functional magnetic resonance imaging. PLoS One 8:e80265.

Ciftci K. 2011. Minimum spanning tree reflects the alterations of the default mode network during Alzheimer's disease. Ann Biomed Eng 39:1493–1504.

Edelsbrunner H, Harer J, Mascarenhas A, Pascucci V, Snoeyink J. 2008. Time-varying Reeb graphs for continuous space-time data. Comput Geom 41:149–166.

Friston KJ, Worsley KJ, Frackowiak RS, Mazziotta JC, Evans AC. 1994. Assessing the significance of focal activations using their spatial extent. Hum Brain Mapp 1:210–220.

Grant KW, Seitz PF. 2000. The use of visible speech cues for improving auditory detection of spoken sentences. J Acoust Soc Am 108:1197–1208.

Hickok G, Poeppel D. 2007. Opinion—the cortical organization of speech processing. Nate Rev Neurosci 8:393–402.

Kang E, Lee DS, Kang HJ, Hwang CH, Oh SH, Kim CS, Chung JK, Lee MC. 2006. The neural correlates of cross-modal interaction in speech perception during a semantic decision task on sentences: a PET study. Neuroimage 32:423–431.

Krieger-Redwood K, Gaskell MG, Lindsay S, Jefferies E. 2013. The selective role of premotor cortex in speech perception: a contribution to phoneme judgements but not speech comprehension. J Cogn Neurosci 25:2179–2188.

Lancaster JL, Tordesillas-Gutierrez D, Martinez M, Salinas F, Evans A, Zilles K, Mazziotta JC, Fox PT. 2007. Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template. Hum Brain Mapp 28:1194–1205.

Lee H, Kang H, Chung M, Kim B, Lee D. 2012. Persistent brain network homology from the perspective of dendrogram. IEEE Trans Med Imaging 31:2267–2277.

Lee UV, Kim S, Jung KY. 2006. Classification of epilepsy types through global network analysis of scalp electroencephalograms. Phys Rev E Stat Nonlin Soft Matter Phys 73(Pt 1): 041920.

McGettigan C, Faulkner A, Altarelli I, Obleser J, Baverstock H, Scott SK. 2012. Speech comprehension aided by multiple modalities: behavioural and neural interactions. Neuropsychologia 50:762–776.

Nishitani N, Hari R. 2002. Viewing lip forms: cortical dynamics. Neuron 36:1211–1220.

Obleser J, Wise RJ, Dresner MA, Scott SK. 2007. Functional integration across brain regions improves speech perception under adverse listening conditions. J Neurosci 27:2283–2289.

Osnes B, Hugdahl K, Specht K. 2011. Effective connectivity analysis demonstrates involvement of premotor cortex during speech perception. Neuroimage 54:2437–2445.

Paulesu E, Perani D, Blasi V, Silani G, Borghese NA, De Giovanni U, Sensolo S, Fazio F. 2003. A functional-anatomical model for lipreading. J Neurophysiol 90:2005–2013.

Reijneveld JC, Ponten SC, Berendse HW, Stam CJ. 2007. The application of graph theoretical analysis to complex networks in the brain. Clin Neurophysiol 118:2317–2331.

Rizzolatti G, Craighero L. 2004. The mirror-neuron system. Annu Rev Neurosci 27:169–192.

Schall S, von Kriegstein K. 2014. Functional connectivity between face-movement and speech-intelligibility areas during auditory-only speech perception. PLoS One 9:e86325.

Scott SK, Blank CC, Rosen S, Wise RJ. 2000. Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123(Pt 12):2400–2406.

Singh G, Memoli F, Ishkhanov T, Sapiro G, Carlsson G, Ringach DL. 2008. Topological analysis of population activity in visual cortex. J Vis 8:11.1–18.

Spitsyna G, Warren JE, Scott SK, Turkheimer FE, Wise RJ. 2006. Converging language streams in the human temporal lobe. J Neurosci 26:7328–7336.

Sporns O. 2011. The human connectome: a complex network. Ann N Y Acad Sci 1224:109–125.

Sporns O, Zwi JD. 2004. The small world of the cerebral cortex. Neuroinformatics 2:145–162.

van den Heuvel MP, Stam CJ, Boersma M, Hulshoff Pol HE. 2008. Small-world and scale-free organization of voxel-based resting-state functional connectivity in the human brain. Neuroimage 43:528–539.

Wang C, Ulbert I, Schomer DL, Marinkovic K, Halgren E. 2005. Responses of human anterior cingulate cortex microdomains to error detection, conflict monitoring, stimulus-response mapping, familiarity, and orienting. J Neurosci 25:604–613.

Yue Q, Zhang L, Xu G, Shu H, Li P. 2013. Task-modulated activation and functional connectivity of the temporal and frontal areas during speech comprehension. Neuroscience 237:87–95.

Address correspondence to:
Dong Soo Lee
Department of Nuclear Medicine
College of Medicine
Seoul National University
28 Yeongeon-Dong
Jongno-Gu
Seoul 110-744
Korea

E-mail: dsl@snu.ac.kr

Hyejin Kang
Department of Nuclear Medicine
College of Medicine
Seoul National University
28 Yeongeon-Dong
Jongno-Gu
Seoul 110-744
Korea

E-mail: hkang211@snu.ac.kr