Decoding Brain Dynamics in Speech Perception Based on EEG Microstates Decomposed by Multivariate Gaussian Hidden Markov Model

NGUYEN THANH DUC(iD), **(Member, IEEE), AND BOREOM LEE**(iD)**, (Member, IEEE)**
Department of Biomedical Science and Engineering, Institute of Integrated Technology, Gwangju Institute of Science and Technology, Gwangju 61005, South Korea

Corresponding author: Boreom Lee (leebr@gist.ac.kr)

ABSTRACT This study aims to reveal dynamic brain networks during speech perception. All male subjects were presented five English vowel [a], [e], [i], [o], and [u] stimuli. Brain dynamics were decoded using multivariate Gaussian hidden Markov model (MGHMM), which trained on spatiotemporal patterns of broadband multivariate event-related potential amplitudes to identify distinct broadband EEG microstates (MS), microstate source imaging, and microstate functional connectivity (μFC). Obtained results showed fluctuated cortical generators and μFC in eight microstates throughout the perception. Microstate source imaging revealed involvements of bilateral (left-side dominance) posterior superior temporal cortex (TC), inferior frontal gyrus (IFG), and supramarginal regions in perception. Precentral cortex where primary motor cortex located was also significantly activated. These regions were early appeared at 96-151 ms (left-side dominance) and at 186-246 ms (left hemisphere only) after the stimuli onset. Results from μFC revealed significant increases in delta (2.5-4.5 Hz), theta (4.5-8.5 Hz), alpha (12.5-14.5 Hz), beta (22.5-24.5 Hz), low gamma (30.5-32.5, 38.5-40.5 Hz) but decreases in high gamma (42.5-46.5 Hz) bands in perception. Increased FC were observed mainly at; (1) microstate segments 34-95 ms (MS2) and 96-151 ms (MS3) in early stages, (2) microstate intervals 186-246 ms (MS5) and 297-449 ms (MS6) in subsequent stages of perception. We found that stronger statistical FC differences in perception at TCs, with respect to left IFG (Broca' area), left TC, and precentral areas. Furthermore, by conducting a comparative protocol measuring FC distinction degree, we showed performance improvements of 8.01% (p-value = 0.0162), 14.41% (p-value = 0.006) when compared MGHMM to well established Lehmann-based modified K-means, Atomize and Agglomerative Hierarchical Clustering and 8.791% (p-value = 0.0097) over the combination of K-means and sliding window methods, respectively. This study indicates the usefulness of EEG microstates to investigate broadband brain dynamics in speech perception. The current findings based on male subjects would be generalized more by future studies with a larger appropriate sample size including female subjects.

INDEX TERMS Speech perception, EEG microstate, source imaging, microstate functional connectivity, MGHMM.

I. INTRODUCTION

Speech perception is a process by which sounds of the spoken languages are heard, interpreted and understood. Researches in the speech perception aim to reveal dynamic mechanisms recruited in the brain from acquisition, comprehension to production of speeches [1]–[3]. The neural processing of

The associate editor coordinating the review of this manuscript and approving it for publication was Venkata Rajesh Pamula(iD).

the spoken language has been hypothesized to occur in a multistage process coordinated by a chain of the established cortical regions [4].

As processing of speeches represents multimodal units that map what is heard with how to produce the sound, neuronal organizations may not be limited to auditory-perceptual temporal area alone, but instead, may be linked to a set of neuronal circuits [5]–[8]. A neuroimaging review [7] concluded that frontoparietal cortices which include the ventral motor

and somatosensory cortices reflect phonological information during spoken language perception. Extensive reviews [9] indicated associated roles of central and posterior parts of the middle and inferior temporal gyri and the posterior inferior frontal gyrus in retrieval of lexical information in the context of language comprehension. Hemodynamic fMRI [10] and event-related potential (ERP) [11] studies have reported that translation of declarative language information from hearing to long-term memory increased neuronal activity in a large number of areas, including the prefrontal, temporal, anterior cingulate, and cerebellar areas. Recently, the concept of dual-stream, proposed by neurolinguistics and neuroanatomists, where ventral (sound to meaning) and dorsal (sound to action) brain areas process information synergistically, has been used to study speech perception [2], [12]. The dorsal stream, which involves structures in the posterior frontal lobe and the posterior dorsal-most aspects of the temporal lobe and parietal operculum, is responsible for translating acoustics into articulatory representations in the frontal cortex essential for speech production. The ventral stream, which involves structures in the superior, middle, and anterior portions of the temporal cortex as well as the ventrolateral prefrontal cortex, is responsible for comprehension. The dual-stream is perhaps the most influential model for language processing in the literature. Despite the wealth of studies using neuroimaging, neural generators, functional associations, and especially dynamic specifications recruited by perception process are not yet fully understood due to a lack of appropriately analytic methods, especially the lack of sophisticated dynamic models [13] as introduced in this study.

Brain is a dynamic network of discrete sub-regions of the functionally specialized neurons. Hence, one should not only investigate where (cortical sources) such functionality occurs, but also examine when (neuronal dynamics) and in what engagements (functional connectivity) the brain is activated when studying speech perception [8], [11]. ERP, which offers excellent temporal resolution, is a powerful tool to decode rapid dynamic patterns using EEG microstate concept that are transient, quasi-stable (∼80-120 ms) and distinct patterns of broadband EEG. Brain dynamic studies suggest that global pattern of brain can be described as being composed of a time sequence of decomposable EEG microstates [14], [15]. Goals of EEG microstate approach are to provide informative dynamics associated with sequences of discrete information processing evoked by presentations of stimuli [16], offering a novel way to understand brain dynamics in cognitive tasks, especially multistage language processing.

Traditional approaches to study dynamic FC (dFC) in EEG utilize clustering-based segmentations of time-varying FC matrix into 'FC-states' using: (1) sliding windows [17]–[20] and (2) K-means [17], [21]–[25]. However, despite encouraging findings, these methods encounter critical issues; that is the former appears limited since strategies for selecting reasonable windows remain unsolved [17], [20]. The latter utilized K-means to cluster the time-varying FC

matrices [23] or FC graphs [24], [25] and topological features [26], [27] which assume that the network is static in one state at a time and the states vary across times. For instance, Allen *et al.* [23] introduced a data-driven method applied on resting-state fMRI to identify 'FC-states'. Similar clustering-based strategies applied to M/EEG recordings were performed by Jamal *et al.* [28], Dimitriadis *et al* [26], [27], [29]–[33]. Alternatively, in recent studies [27], [30], [32], [33] the neural-gas clustering algorithm has been used to detect the connectivity patterns in M/EEG data and proved to be an effective approach for reproducibility. Also using clustering-based approaches on dynamic FC matrices, Mheich *et al.* [25] (scalp data) and Hassan *et al.* [24] (source data) identified the 'FC-states' using EEG cognitive tasks. Despite the usefulness of the K-means method, it has several critical cons. First, K-means requires appropriate selections of clusters at initialization steps. However, choosing a proper number of clusters can be difficult particularly in case the data is dynamic and prior knowledge is unknown. K-means method also starts with random choices of cluster centers and therefore it may yield different clustering results on different iterations, which leads to inconsistent results.

To avoid abovementioned limitations caused by K-means clustering, various advanced directions to track fast changes of FC networks (FCNs) have recently been introduced. For example, resting-state EEG FCNs were revealed using synchronization likelihood analysis and evolutionary clustering [34], tensor decomposition-based [35], and principal component analysis (PCA)-based [36] approaches. Novel approaches that use ERP segmentation, followed by dynamic time warping and quality threshold clustering, to track dynamic patterns of FC in task-based EEG was introduced [21]. However, dynamics rely on the ERP segmentation process that adopted pre-determined ERP components and therefore the generalization cannot be applied precisely. More importantly, the studies therein did not investigate the relations of EEG microstate connectivity networks and brain network dynamics. Then, it raises the questions of whether or not the EEG microstate networks can be relevant to provide insightful brain dynamics coordinated in cognitive tasks. In addition, what is certainly unavailable in the existing literature is a reliable localization and connectivity of the neuronal sources generating these different EEG microstates, that is, the cortical regions that are synchronized with each other during each of these microstates [37].

In recent work [22], we proposed a novel framework to decode microstate FC (μFC) patterns from broadband ERP data using the multivariate Gaussian hidden Markov model (MGHMM) and this approach has been proved to provide several significant advantages over the traditional ones in extracting the true brain dynamic activities. First, MGHMM method trains the multivariate data using Bayesian variational learning to automatically reveal an optimum clustering number of hidden microstates. This strategy overcomes the limitation of the existing methods, in which the plans for selecting the optimum number of clusters remain ambiguous,
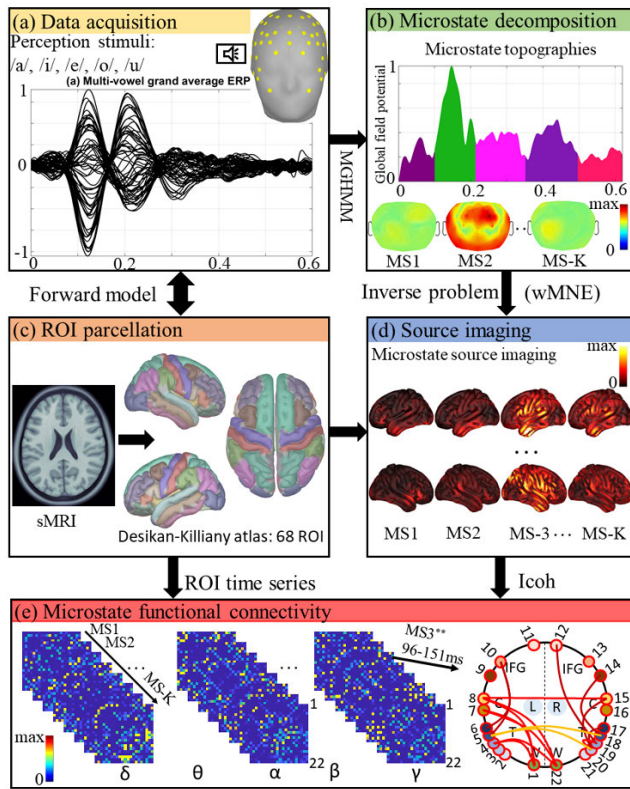
FIGURE 1. Overall framework to study microstate cortical organizations, μFC in speech perception using broadband EEG. (a) ERP response to speech stimuli. (b) EEG microstate decomposed from ERP data using MGHMM. (c) 68-ROI Desikan-Killiany atlas. (d) microstate source imaging using weighted Minimum Norm algorithm (wMNE). (e) μFC using imaginary of coherence (Icoh).

especially for K-means clustering and time sliding window. Second, MGHMM uses Viterbi decoding algorithm learned from the training phase to compute the state sequence to identify the most likely microstate that an observation vector at a time belongs to. This automatic segmentation using multivariate analysis solves the limitation of the pre-determined ERP segmentation performed by univariate analysis method as indicated in [21]. Third, with this new direction, one can be able to investigate the relations of EEG microstate connectivity and brain network dynamics that cannot be studied with existing methods. Fourth, by using EEG simulation and realistic recordings, we proved that MGHMM outperformed the existing approaches in revealing 'true' dynamic FC networks in EEG cognitive task. Lastly, to our knowledge, this is the first study introducing the new concept of microstate FC (μFC) in understanding the brain dynamics in EEG cognition tasks including speech perception.

Given this gap in the existing literature, in this study we investigate EEG microstates, microstate-guided source localizations and microstate-guided μFC patterns dynamically underlying the speech perception using broadband EEG. Our work is based on the assumption that the broadband source localization and connectivity are oriented by the estimated EEG microstates at the sensor level representing that broadband EEG. Fig. 1 depicts the overview framework of

this study. First, EEG signals were recorded from the male healthy subjects when they heard the vowel (perception condition) stimuli and mute (zero volume) sound (baseline control condition). The number of broadband EEG microstates underlying the perception process was decoded from multivariate grand-averaged ERP amplitude data using MGHMM method. We then employed weighted Minimum Norm Estimate (wMNE) to specifically map each of estimated broadband microstate topographies into each of source-space representations. FCs were estimated using imaginary part of coherence (Icoh) which was introduced to overcome volume conduction effects and reveal the 'true' connectivity. μFC were obtained by averaging the FC across time segments where the existence of corresponding broadband EEG microstates on the sensor level, which are estimated by MGHMM method, is expected. To validate our approach on studying the brain dynamics, we constructed a comparative protocol that we compare sensor-based connectivity distinction degrees estimated by MGHMM with that of well-established methods including K-means and sliding window combination that estimates the dynamic 'FC-states'. The comparisons also included Lehmann-based Atomize and Agglomerative Hierarchical Clustering (AAHC) and modified K-means methods that estimate microstate-guided functional connectivity [14], [38]. This research is among the first reports, we believe, to investigate dynamic cortical organizations and functional interactions on the basis of broadband EEG microstates that are revealed in the speech perception.

II. MATERIAL AND METHODS
A. EEG DATA ACQUISITION
1) PARTICIPANTS
Eleven healthy native Korean volunteers (all males; graduate students; mean age: 28.25$\pm$2.71; range: 26-32) who were well known to English language participated in this experiment. Their English proficient levels, in overall, were above either TOEIC score of 750 or IELTS score of 6.5 with recognized certificates. None of the participants had major neurological or hearing diseases, and all were right-handed naïve to the experiment. The clinical evaluation was performed by two experienced psychiatrists in our research group. All participants gave the written informed consent, and the experimental paradigm was approved by the Institutional Review Board (IRB) of Gwangju Institute of Science and Technology (GIST), Gwangju, South Korea. The approval process of the IRB complied with the declaration of Helsinki.

2) EXPERIMENT PROTOCOL
Fig. 2 describes schematic sequences of the experiment paradigm. The participants were each seated in a comfortable armchair and wore earphones (Etymotic research, Inc., IL 60007, U.S.A) providing speech stimuli. For perception condition, five English vowel stimuli; /a/, /e/, /i/, /o/, and /u/ were randomly and equally presented for one second while for baseline control condition a mute sound
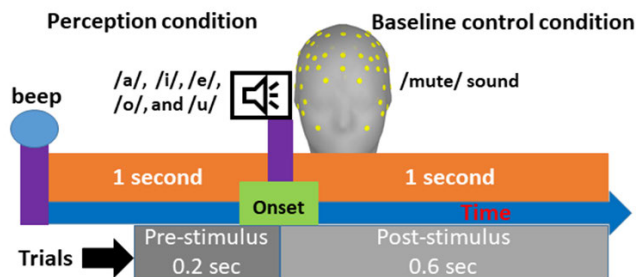
FIGURE 2. Schematic sequence of experiment protocol. The EEG data were acquired for two conditions: 1) perception condition when the subjects heard five English vowels /a/, /e/, /i/, /o/, and /u/, and 2) baseline control condition when the subjects were at resting-state.

FIGURE 3. Hierarchical architecture of the MGHMM consists of state-space HMM model (a) and multivariate Gaussian observation model (b).

was provided. Above five stimuli were recorded using Gold-ware software (GoldWave Inc., St. John's, Newfoundland, Canada), while the source audio was acquired online from Oddcast's (www.oddcast.com/home/demos/tts/).

At beginning of each trial, a one-second beep sound was presented to prepare the participants for perception, which was followed by the onset of the target stimulus. A 2 sec fixed inter-stimulus interval (ISI) was set between trials and a cross mark appeared on screen at exact time when the participant was to perceive speech stimuli. The experimental paradigm was designed with e-Prime 2.0 software (Psychology Software Tools, Inc., Sharpsburg, PA, USA). The EEG data were continuously recorded using a HydroCel Geodesic Sensor Net with 64 electrodes, Net Amps 300 amplifiers, and Net Station software package (ver. 4.5.6, Electrical Geodesics, Inc., Eugene, OR, USA). The sampling rate was set at 1000 Hz. Each subject performed five sessions; each session consisted of 10 trials per stimulus, resulting in 60 trials for each stimulus and 300 trials for each subject. A 1-minute resting time was provided between sessions and the total time required for each subject was approximately 15 min.

3) PREPROCESSING
Preprocessing of EEG data was performed with Field-trip [39], a toolbox for EEG analysis compatible with MATLAB (2017b; MathWorks, Inc., Natick, MA, USA). First, an IIR notch filter (Butterworth; order: 4; bandwidth: 59-61 Hz) was applied to remove power line noise. The raw continuous EEG signals were then filtered by a 0.5–46.5 Hz zero-phase shift band-pass filter to remove baseline drift and high-frequency artifacts before the EEG data were averagely re-referenced. Afterwards, EEG were segmented into 0.8 sec time-locked epochs with 0.2 sec pre-stimulus baseline plus 0.6 sec post-stimulus. Four ocular channels monitored vertical and horizontal eye movements/blinks from the outer canthi and left infraorbital ridges and were excluded before further analysis. To remove electro-oculography (EoG) and electro-cardiography (ECG) artifacts, the independent component analysis (ICA) using the 'runICA' algorithm which comes as a built-in tool in Fieldtrip was applied to the preprocessed data. Artifact epochs were removed manually
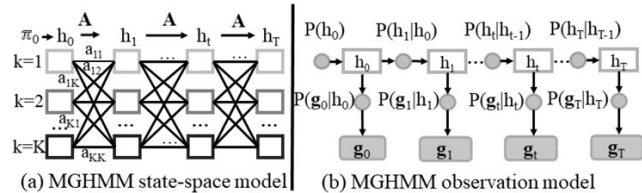
after visual inspections of each component by an expert neu-ropsychiatrist. Next, we reconstructed the time-series using the artefact-free independent components. In addition, the epochs whose signals amplitudes exceeded ± 100 μV were excluded. After this process, a total of 2164 out of 3300 trials remained for further analysis. Next, the grand-averaged ERPs for speech perception condition were calculated. Finally, we normalized the ERP data to remove the effects of pre-stimulus interval (0.2 sec).

4) REGIONS OF INTEREST
An average 15000-voxel cortex was parcellated into a Desikan-Killiany atlas [40] with 68 anatomical regions of interest (ROI) in Brainstorm [41] (Fig. 1c). These ROIs are considered to be the source-space nodes for computing FC. Detail of language-related ROIs in Desikan-Killiany atlas is provided in Appendix Table 2.

B. EEG MICROSTATES DECOMPOSED BY MGHMM
Our proposed MGHMM method (Fig. 3) was used to train patterns of multivariate ERP amplitude data for microstate decomposition [22]. An observation $\boldsymbol{g}_t$ representing multivariate $N_{sensor} \times 1$ ERP at time t is generated by a hidden state h_t which gets one of K possible states. $P(h_t|h_{t-1})$ is the probability of achieving the state h_t, which is conditionally dependent only on its immediate previous state h_{t-1}. Observed $\boldsymbol{g}_t$ is dependent only on the current state h_t but independent of other observations, denoted as $P\left(\boldsymbol{g}_t \mid h_t\right)$. Collectively, the true posterior distribution of T hidden states $\boldsymbol{h} = \{h_0, \cdots, h_t, \cdots, h_T\}$ and observations $\boldsymbol{G} = \{\boldsymbol{g}_0, \cdots, \boldsymbol{g}_t, \cdots, \boldsymbol{g}_T$ can be presented as follows:

$$P\left(\boldsymbol{h}, \boldsymbol{G}, \boldsymbol{\theta}, \boldsymbol{A}\right)$$
$$= P\left(h_0 \mid \boldsymbol{\pi}_0\right) \prod_{t=1}^{T} [P(h_t|h_{t-1}, \boldsymbol{A})P(\boldsymbol{g}_t|h_t, \boldsymbol{\theta})]P(\boldsymbol{\theta})P(\boldsymbol{A}) \tag{1}$$

In (1), $\boldsymbol{A}$ is $K \times K$ state transition probability matrix. $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_k, \cdots, \boldsymbol{\theta}_K\}$ with $\boldsymbol{\theta}_k = \{\boldsymbol{\mu}_k, \sum_k\}$ characterizes the observation model for the state k which is a multivariate Gaussian distribution expressed as $P\left(\boldsymbol{g}_t \mid h_t = k, \boldsymbol{\theta} = \boldsymbol{\theta}_k\right) = P\left(\boldsymbol{g}_t \mid \boldsymbol{\mu}_k, \sum_k\right)$, where $\boldsymbol{\mu}_k$ is the mean and $\sum_k$ is the $K \times K$ covariance matrix of the Gaussian distribution. The initial state probability vector, $\boldsymbol{\pi}_0$, has elements $\pi_{0k} = P(h_0 = k)$ that satisfy $\sum_k \pi_{0k} = 1$. In addition, $P(\boldsymbol{\theta})$ and $P(\boldsymbol{A})$ are non-informative priors.

An unsupervised Bayesian variational inference method was performed to learn model parameters. Principles of Bayesian learning is based on iterative minimization of Kullback-Leibler divergence (KL-divergence) between the true posterior probability in (1) and an analytical approximation [42], [43]. Conceptually, KL-divergence minimization process with respect to the model dimension leads to the most probable model size.

EEG microstate decomposition from ERP data using MGHMM was illustrated with three subsequence steps [22]. First, a strategy for the model dimension selection revealed an optimum number of the hidden microstates. Second, once the state dimension was determined, the converged KL-divergence at the optimal state dimension was estimated. We decoded the state sequence (s_t) using the Viterbi algorithm to identify the most likely state (k) that a multivariate vector, $g_t \in G$ (G: observation matrix), belongs to as follows:

$$s_t = \max P(h_t = \max_k k | g_t) \quad (2)$$

Third, we smoothed microstate topologies with temporal window of 30 ms by eliminating microstates that occurred continuously for less than 30 ms and reassigning them to one of the neighboring microstates, which obtained highest spatial correlation [38]. Further details of EEG microstate decomposition on spatiotemporal multivariate ERP amplitude data using MGHMM as well as the mathematic explanations of above equations can be referred to our recent paper [22].

C. EEG SOURCE IMAGING

Let $G = \{g_1, \cdots, g_t, \cdots, g_T$ be a $N_{sensor} \times T\}$ matrix that represents the multivariate EEG signals at N_{sensor} electrodes over all T samples. The broadband EEG potential is expressed as a linear combination of a lead field matrix $L = \{l_1, \cdots, l_p, \cdots, l_P\}$ and the current dipole source time-series $= \{s_1, \cdots, s_t, \cdots, s_T\}$ as follows:

$$G = L \cdot S + E \quad (3)$$

In (3), L represents an $N_{sensor} \times P$ matrix obtained by solving a forward problem, given the distributed source model. P (15000 voxels) is the number of active sources with fixed positions and orientations. E is the $N_{sensor} \times T$ measurement noise corrupted to EEG data.

One popular neuroimaging approach to solve ill-posed ($P \gg N$) inverse problem is the weighted Minimum Norm Estimate (wMNE). Basically, the wMNE algorithm compensates for the tendency of classical MNEs to favor weak and surface sources [24]. This is done by introducing a $P \times P$ weighting matrix W_S to the estimated sources:

$$\hat{S}_{wMNE} = (L^T W_S L + \alpha I)^{-1} L^T W_S G \quad (4)$$

where matrix W_S regulates the solution by diminishing the bias inherent to MNE solutions. Typical selection of W_S is a diagonal matrix derived from the lead field matrix L with non-zero elements inversely proportional to the norm of the lead

field vectors. In addition, the value α is calculated relative to signal to noise ratio between the post-stimulus and pre-stimulus intervals [24].

D. FUNCTIONAL CONNECTIVITY METRIC: ICOH

We employed imaginary part of the coherence (Icoh) method to assert brain FC. The complex coherence between two time series can be defined as the cross spectrum divided by the product of the two power spectra. Its mean over all frequencies can alternatively be computed via the mean over time of the corresponding analytical signals [44]:

$$C = \frac{\langle A_1 A_2 e^{i\Delta\theta} \rangle}{\sqrt{\langle A_1^2 \rangle \langle A_2^2 \rangle}} \quad (5)$$

where A_1 and A_2 are amplitudes, and $\Delta\theta$ is the instantaneous phase difference (Hilbert transform) of the two time series. Absolute coherency is Magnitude-squared Coherence (MsC). MsC is a well-established metric of linear association between two signals at certain frequencies. However, MsC is known to be sensitive to volume conduction that leads to spurious zero-lag interactions not attributed to neural sources [45]. Imaginary part of coherency (Icoh) is derived as:

$$Im\{C\} = \frac{\langle A_1 A_2 \sin\Delta\theta \rangle}{\sqrt{\langle A_1^2 \rangle \langle A_2^2 \rangle}} \quad (6)$$

As Icoh was introduced to address volume conduction issue [46] by capturing the part of synchronization that occurs with a non-zero time lag, it is only sensitive to signals that are time-lagged to each other and therefore isolates the part of coherency which necessarily reflects true interaction. FC measures were performed and analyzed using Matlab (The MathWorks Inc., Natich, MA) on the 68 scout Desikan-Killiany time series for computation efficiency since FCs on 15000-voxel cortex demand enormous computations. Before FC computation, each ROI time series was obtained by averaging all voxels, which laid in the corresponding ROI. FCs were employed on each EEG frequency between 0.5 and 46.5 Hz (23 segments of 2 Hz bandwidth with 50% overlap). μFC were obtained by averaging the FC across time intervals where the existence of corresponding EEG microstates was decomposed. Values for Icoh range from 0 to 1. A zero Icoh value indicates a complete absence of synchrony between two brain time series while a value of 1 specifies perfect synchronization.

E. STATISTICAL ANALYSIS

Statistical analyses to compare significant differences of (1) EEG microstates, (2) microstate cortical imaging, and (3) microstate FC between two conditions (perception- and baseline control conditions) were assessed by two-sample t-test via Monte Carlo permutation in a nonparametric model in Matlab. Specifically, the statistical differences between the two conditions for each EEG microstate and microstate source imaging intervals were performed on microstate
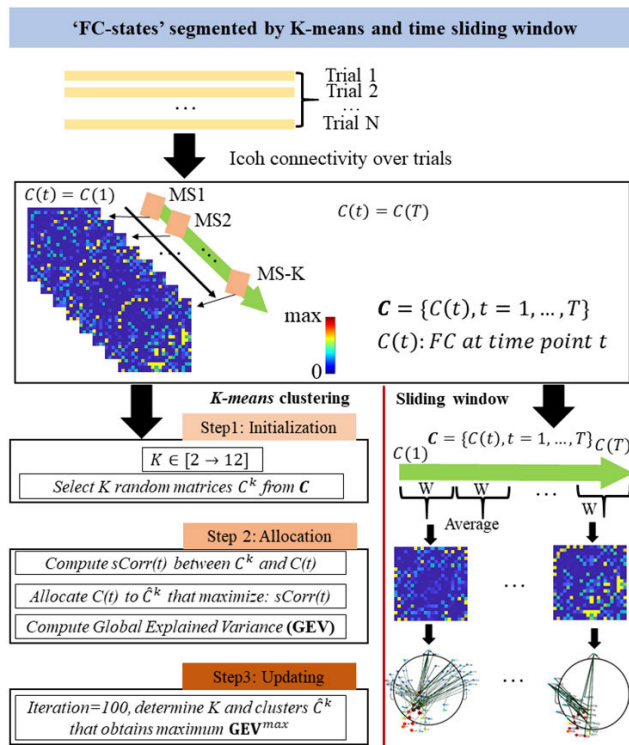
FIGURE 4. Illustrations of the clustering-based approaches using K-means and sliding window to retrieve the dynamic 'FC-states'.

topographies (electrode pairs) and on 15000-voxel cortex (voxel pairs), respectively, while the statistical differences of microstate FC were performed on 68 cortical ROIs at each frequency bins (23 bins). We employed the nonparametric permutation t-test (randomization test) for statistical analysis as it does not require data to be normally distributed and allows testing for multiple data types. The number of randomizations for permutation test was set to 1000 times. For all of inputs, the corrected significant level (corrected p-values) were obtained using false discovery rate (FDR, p <0.05) method for multiple comparison [41].

F. COMPARATIVE METHODS

To validate the efficacy of MGHMM approach on studying brain dynamics, we compare it with long-established methods using K-means clustering and time sliding window as illustrated in Fig. 4 and presented in the followings [24], [25] (see [22] for further details of the comparative protocol):

- K-means clustering: This method aims to cluster the T connectivity matrices $C = \{C(t), t = 1, \ldots, T\}$ into a number of static 'FC-states' with the following three-step process. First step is *initialization* which randomly determines K clusters $C^k = \{\bar{C}^k, k = 1, \ldots, K\}$ from C. The greedy-search algorithm was performed to identify the most optimal K within the range $K = [212], k \in [1, K]$. Second step is *allocation* which computes the spatial correlation $sCorr(t)$ between $C(t)$ with selected matrix $\bar{C}^k$. $sCorr(t)$ ranges from 0

(uncorrelated) to 1 (fully-correlated). Thus, each of FC matrix $C(t)$ was allocated to the cluster $\bar{C}^k$ for which the spatial correlation $sCorr(t)$ was maximized as: $\hat{C}^k = \{C(t) : sCorr^k_{C(t),\bar{C}^k} \geq sCorr^h_{C(t),\bar{C}^h} \forall h, k \in [1, K]\}$. We adapted global explain variance (GEV) [38] to allocate $C(t)$: $GEV^K = \sum_{k=1}^{K} GEV^k$ with $GEV^k = \sum_{t=1}^{T} sCorr_{C(t),\bar{C}^k} \cdot \lambda_{C(t),\bar{C}^k}$, where $\lambda_{C(t),\hat{C}^k} = 1$ if $C(t) \in \hat{C}^k$, otherwise if $C(t) \notin \hat{C}^k$, then $\lambda_{C(t)0,\hat{C}^k} = 0$. Last step is *updating* which means at each iteration ($t \in [1, \ldots, T]$), the new set of clusters $\bar{C}^k$ are recomputed and updated. For each $K \in [2, 12]$, *allocation* and *updating* steps are repeated 100 times, the highest GEV leads to the optimum cluster set ($K = 8$).

- Time sliding window: Given a series of dynamic FC matrices $C = \{C(t), t = 1, \ldots, T\}$ across times, this method utilized a temporal sliding-window to cluster the time-varying patterns. We used a window size of 80 time points (80 ms with the sample rate of 1 kHz) and there are no strides between consecutive windows. We then averaged the FC matrices within a specific window to obtain a FC matrix that represents for a functional stationary 'FC-state'. Collectively, $M = 8$ functional networks, same as those identified by MGHMM and K-means, were investigated in the speech perception.

The comparative protocol also includes microstate-guided functional connectivity based on classical EEG microstate analysis algorithms. EEG microstate analysis was performed using EEGLAB microstate plugin, which can be found at *www.thomaskoenig.ch*. EEGLAB microstate plugin is a free, graphical user interface designed toolkit and developed by Koenig *et al.* [47] for the identification and quantifications of EEG microstates in time-series EEG data. In this work, we used the two classical methods that appear in most of the current literature [48], *i.e.,* modified K-means clustering and Atomize and Agglomerative Hierarchical Clustering (AAHC) [14], [38], to extract a microstate sequence representing for the broadband EEG data. The EEG microstate analysis approach is based on the observation that the Global Field Potential (GFP) periodically achieves local maxima and that EEG microstates are defined at these maxima [48]. The optimal number of EEG microstate clusters is determined through a cross-validation strategy with Global Explained Variance (GEV) criterion. For each algorithm, we set the number of clusters from 2 to 12 and the optimal number of clusters is the one that best explains the variances in each cluster. The minimum peak distance between microstates is set to 30 ms. Using these algorithms, EEG topographies can be assigned with microstate labels by competitive fitting based on spatial correlation, leading to a microstate sequence to represent the EEG data. Further details of AAHC, modified K-means algorithms, and experiment setups can be further provided in [14], [38], [48] and at *www.thomaskoenig.ch*, respectively. The temporal windows for each microstate FC are set to be aligned with the time intervals, where the existence of the estimated EEG microstate is expected.
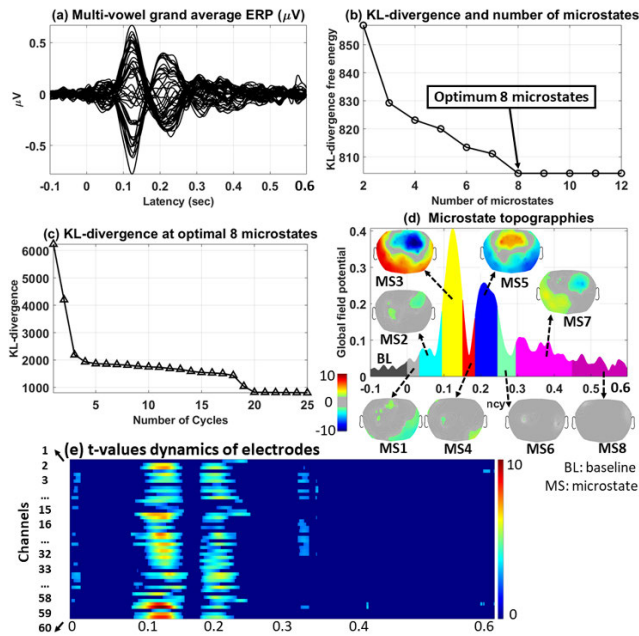
FIGURE 5. The results of microstate segmentation in speech perception EEG data using MGHMM algorithm. (a) the grand-averaged ERP responded to speech stimuli (perception condition). (b) model-dimension selecting strategy. (c) KL-divergence convergence at the optimum eight-microstate. (d) Viterbi sequence decoding and projected t-values on topographies showing statistical differences between perception condition and baseline control condition at each EEG microstates. (e) corrected t-value dynamics of electrodes.

Then, sensor-based FC associated with each broadband EEG microstate is computed by averaging the FC across temporal windows that correspond to the estimated EEG microstates on the sensor level.

III. RESULTS

In what follows, we demonstrate usefulness of our framework in decoding neuronal dynamics including EEG microstates (MS), microstate source imaging, and microstate μFC networks that were commonly engaged in speech perception process.

A. EEG MICROSTATES IN SPEECH PERCEPTION

Fig. 5 shows analytical results of microstate decomposition applied on the perception condition grand-averaged ERP (Fig. 5a). We iteratively recorded the converged KL-divergence computed with microstate dimensions from 2 to 12 (Fig. 5b) and an optimal number of eight microstates (KL-divergence: 804.16; converged cycles: 25, Fig. 5c) was found; namely MS1, MS2... and MS8. Afterward, we performed Viterbi decoding on multivariate time series (Fig. 5d). We emphasize that MGHMM algorithm with temporal smoothing (size = 30 ms) decoded the perception-condition ERP into eight optimum EEG microstates. Fig. 5d shows t-value maps of corrected statistical differences between perception- and baseline control conditions at each EEG microstates. The highest t-values were obtained in

TABLE 1. Characteristics of decomposed eeg microstates.

| Statistics	Durations (ms)	Mean (µV)	STD	Correlation
Microstate 1 (MS1)	1:33	0.038	0.026	0.522
Microstate 2 (MS2)	34:95	0.068	0.051	0.467
Microstate 3 (MS3)	**96:151**	**0.309**	**0.169**	**0.932**
Microstate 4 (MS4)	152:185	0.101	0.068	0.446
Microstate 5 (MS5)	**186:246**	**0.239**	**0.097**	**0.803**
Microstate 6 (MS6)	247:296	0.080	0.061	0.443
Microstate 7 (MS7)	297:449	0.062	0.045	0.431
Microstate 8 (MS8)	450:600	0.032	0.021	0.032

Microstates in bold (**MS3** and **MS5**) indicate the highest statistical

MS3 which lasts from 95 ms to 151 ms and in MS5 which lasts from 186 ms to 246 ms. The t-value maps clearly indicated significant activations of bilateral temporal areas (left-side dominance) at those microstates. Central and inferior frontal areas were also significantly activated. By contrast, statistical analysis results determined that there were no statistical differences on MS6 (247-296 ms) and MS8 (450-600 ms) between perception- and baseline control conditions. Fig. 5e shows the corrected statistical t-value of each electrode at each time point while grand-averaged numerical characteristics of decomposed EEG microstates are provided in Table 1.

B. MICROSTATE CORTICAL REPRESENTATION IN SPEECH PERCEPTION

Fig. 6a-h show cortical mappings at the basis of EEG microstates, while Fig. 6i presents the macrostate cortical representation, which was averaged at all-time instants. We projected voxel-wise corrected significant t-values on the cortex. In general, as presented in Fig. 6i, we demonstrate that for mapping the acoustic spoken speech onto conceptual and semantic representations, multiple language-relevant regions were recruited. Firstly, we emphasize involvements of the bilateral posterior superior temporal gyri (pSTG) and inferior supramarginal gyrus (SG) with left-side dominance (Wernicke's area). Secondly, we observed the ventral processing induced by the bilateral ventral prefrontal lobes and anterior temporal gyrus (ATG includes temporal pole-TP). Third, an important region incorporated during the task was the bilateral inferior frontal gyrus (IFG) with left-side dominance where the Broca's area is located. Finally, our results revealed involvements of motor cortices supporting the task.

We have seen language-relevant regions elicited by spoken vowels at macroscale. However, questions have arisen as to where and when neuronal organizations involved in such tasks. Dynamic activities are revealed in Fig. 6a-h, which present the source imaging of individual decomposed EEG microstates (from MS1 to MS8), indicating transient fluctuations of multi-stage activities from hearing to perceiving and articulating. The earliest cortical stage (MS2: 34-95 ms) of speech perception primarily took place in the left anterior middle and inferior temporal gyri (aMTG
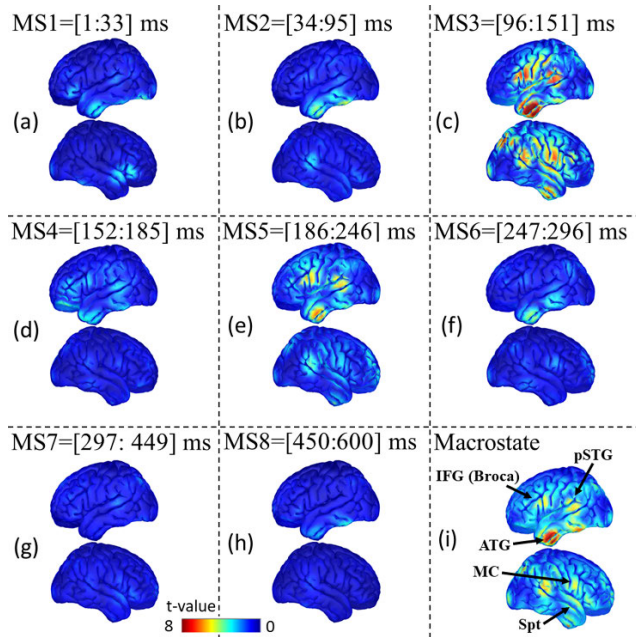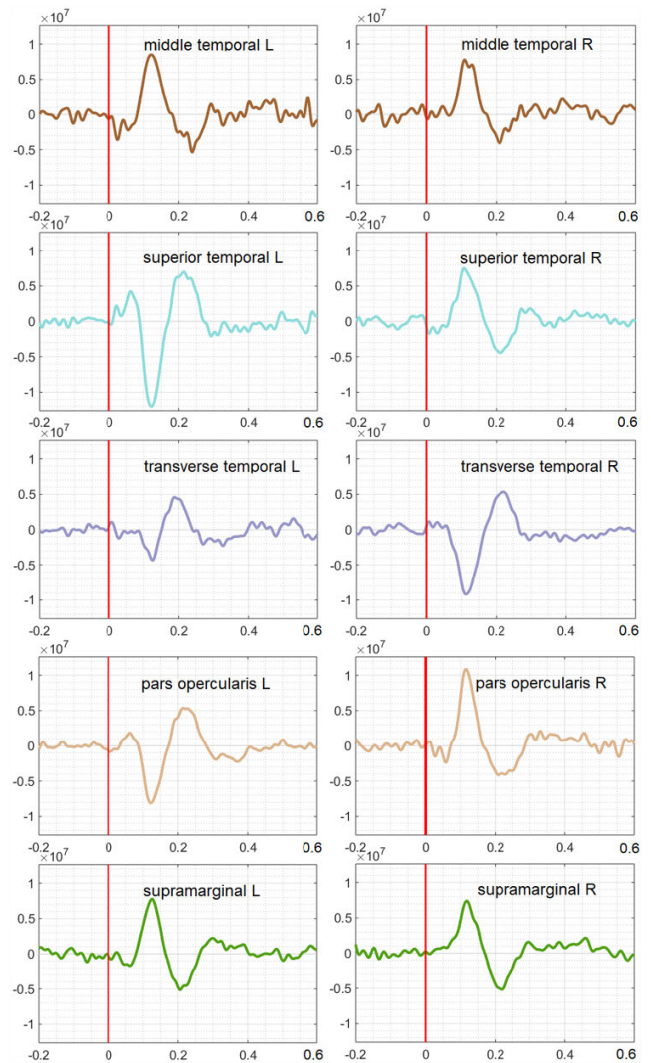
FIGURE 6. Results of microstate (a-h) and macrostate (i) cortical presentations. The corrected statistical t-values were computed between perception- and baseline control conditions at each microstate intervals and projected on the cortex. IFG: inferior frontal gyrus. pSTG: posterior of superior temporal gyrus. ATG: anterior temporal gyrus. MC: motor cortex. Spt: Sylvian parieto-temporal.

FIGURE 7. Grand-averaged typical language-relevant ROI time series. Each ROI time series was obtained by averaging all voxels, which laid in the corresponding ROI. L = left hemisphere; R = right hemisphere.
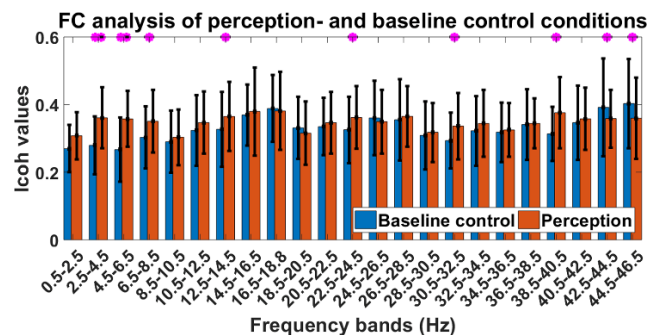
and aITG), followed by the strongest activation (largest t-values) interval in MS3 (96-151 ms) where the bilateral pSTG (Wernicke' area) and IFG (Broca's area) regions with left-side dominances. Our findings also revealed involvements of bilateral motor cortices as well as aMTG and aITG at MS3 segment from 96 to 151 ms after the onset of speech stimuli. Next, we observed an inactivation interval at MS4 (152-185 ms) where the brain goes to the idle stage in which no activated regions were shown and it remained inactivated at subsequent microstate intervals (MS7 from 297-449 ms and MS8 from 450 to 600 ms). In between, we observed recurrences of pSTG, IFG, and aMTG with left-side dominances at MS5 interval (186-246 ms). Constructed time series of language-related regions, which showed significant differences in perception as compared to baseline control, are shown in Fig. 7. These significant areas are bilateral middle temporal, superior temporal, transverse temporal, pars opercularis, supramarginal ROIs.

C. MICROSTATE FUNCTIONAL CONNECTIVITY IN SPEECH PERCEPTION

We present ROI-space microstate FC analyses coordinated in speech perception. Statistical differences in Icoh-based FC between perception- and baseline control conditions in frequency bins are summarized in Fig. 8. Separate frequency bins ranging from delta (0.5-4.5 Hz), theta (4.5-8.5 Hz), alpha (8.5-14.5 Hz), beta (14.5-24.5 Hz), and gamma (24.5-46.5 Hz) were analyzed. Results in Fig. 8 reveal significant connectivity increases in delta (δ: 2.5-4.5 Hz), theta

FIGURE 8. Statistical analysis of Icoh-based FC between perception and baseline control conditions in all ranges of frequency bands. Significant differences are marked by stars where a '∗' indicates 0.01 < p-values ≤ 0.05, and '∗∗' indicates p-values ≤ 0.01.

(θ_1: 4.5-6.5 Hz, θ_2: 6.5-8.5 Hz), alpha (α: 12.5-14.5 Hz), beta (β: 22.5-24.5 Hz), low gamma (low γ_1: 30.5-32.5, low γ_2:
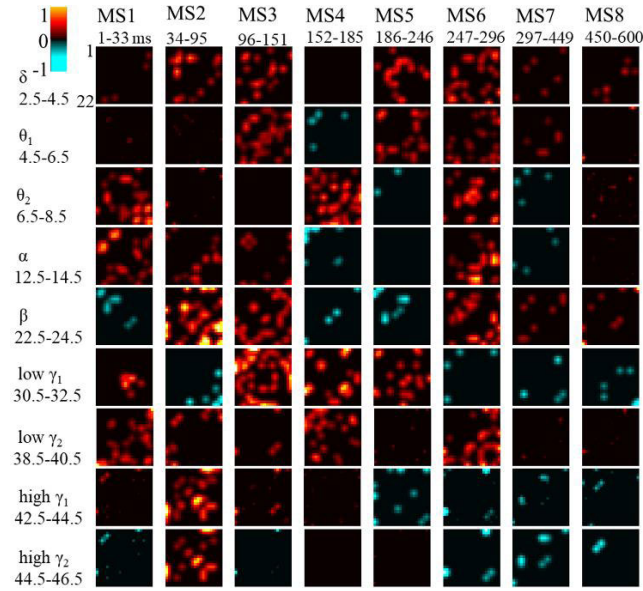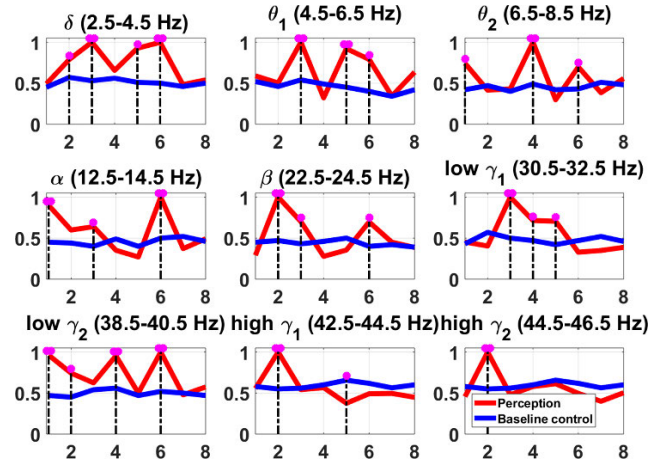
FIGURE 9. Statistical analyses of μFC matrices between perception- and baseline control conditions in all significant-difference frequency bands. For clarity, non-significant Icoh are set to zero in the maps. Hot colormap presents μFC increases while cool colormap represents μFC decreases in perception.

FIGURE 10. Statistical analyses of μFC between perception and baseline control conditions at eight segmented microstate intervals. The analyses were performed on only frequency bins at which the significant differences in FC in speech perception were found in Fig. 8. Significant differences are marked by stars where a '*' indicates $0.01 < p$-values ≤ 0.05, and '**' indicates p-values ≤ 0.01. Numbers at x-axis represents for subsequence microstates (MS1 to MS8).

38.5-40.5 Hz) and significant decreases in high gamma (high γ_1: 42.5-44.5, high γ_2: 44.5-46.5 Hz) bands in the speech perception process.

Regarding dFC analyses, Fig. 9 shows dynamical significant connectivity between 22 language-relevant ROIs for eight EEG microstate intervals decomposed by MGHHM (see Fig. 5 and Table 1) in perception as compared to baseline control conditions. dFC analyses were performed only on frequency bands which led to significant differences between perception- and baseline control conditions as revealed in Fig. 8. Only language-related ROIs (see details in Appendix Table 2) that were significantly activated during the speech perception (as found in section III.B) were used for dFC analyses.

μFC statistical analyses between the perception- and baseline control conditions at the eight decomposed EEG microstate segments in the nine significant-difference frequency bins (see Fig. 8) are given in Fig. 10. First, it is clearly showing that perceiving the speech stimuli significantly increased FC at early temporal segments at MS1 (1-33 ms) in θ_2 (6.5-8.5 Hz), α (12.5-14.5 Hz), low γ_2 (38.5-40.5 Hz) and especially at MS2 (34-95 ms) in δ (2.5-4.5 Hz), β (22.5-24.5 Hz), low gamma (low γ_2: 38.5-40.5 Hz), and high gamma (high γ_1: 42.5-44.5, high γ_2: 44.5-46.5 Hz) frequency band. In addition, we found that speech perception process significantly increased functional connectivity at MS3 interval (96-151 ms) in δ (2.5-4.5 Hz), θ_1(4.5-6.5 Hz), α (12.5-14.5 Hz), and low γ_1 (30.5-32.5 Hz) bands.

The increase FC were also found the significant increases in perception at MS4 (152-185 ms) in θ_2 (6.5-8.5 Hz), low γ_1 (30.5-32.5 Hz), low γ_2 (38.5-40.5 Hz), at MS5 (186-

246 ms) in δ, θ_1, low γ_1 bands and at MS6 (247-296 ms) in δ (2.5-4.5 Hz), θ_1(4.5-6.5 Hz), θ_2(6.5-8.5 Hz), α (12.5-14.5 Hz), β (22.5-24.5 Hz) and low γ_2 (38.5-40.5 Hz) for subsequent perception process. In addition, there were no significant FC differences between two conditions in MS7 (297-449 ms) and MS8 (450-600 ms) in all frequency bands.

Illustrations of dynamic interactions in significant-difference frequency bands at eight MS between 22 language-related ROIs when subjects heard speech stimuli are provided in Fig. 11. It is obvious that the FC fluctuated and the brain recruited different connected regions during the perception process. We found that connectivity was mainly distributed at temporal regions for all microstate segments. Interactions also existed between left temporal areas and left inferior frontal gyrus; left temporal areas and precentral areas where primary motor cortex is laid on. In addition, we found connections between supramarginal regions with precentral regions.

D. PERFORMANCE COMPARISON RESULTS

KL-divergence was used as a comparative indicator to evaluate the dynamic FC performance between μFC-based MGHMM approach with the aforementioned 'FC-states'-based K-means and time sliding window approaches. KL-divergence of two distributions, $\aleph_0$ and $\aleph_1$, is expressed as follows:

$$D_{KL}(\aleph_0 \mid \aleph_1) = 0.5[tr(\sum\nolimits_{\aleph_1}^{-1} \sum\nolimits_{\aleph_0}) + (\boldsymbol{\mu}_{\aleph_1} - \boldsymbol{\mu}_{\aleph_0})^T$$
$$\times \sum\nolimits_{\aleph_1}^{-1} (\boldsymbol{\mu}_{\aleph_1} - \boldsymbol{\mu}_{\aleph_0}) + \ln(\frac{det \sum_{\aleph_1}}{det \sum_{\aleph_0}})] \quad (7)$$

where $\boldsymbol{\mu}_{\aleph_0}$ and $\boldsymbol{\mu}_{\aleph_1}$ are means of the distributions $\aleph_0$ and $\aleph_1$, while $\sum_{\aleph_0}$ and $\sum_{\aleph_1}$ represent their corresponding covariance matrices. In principle, higher values of $D_{KL}(\aleph_0 \mid \aleph_1)$
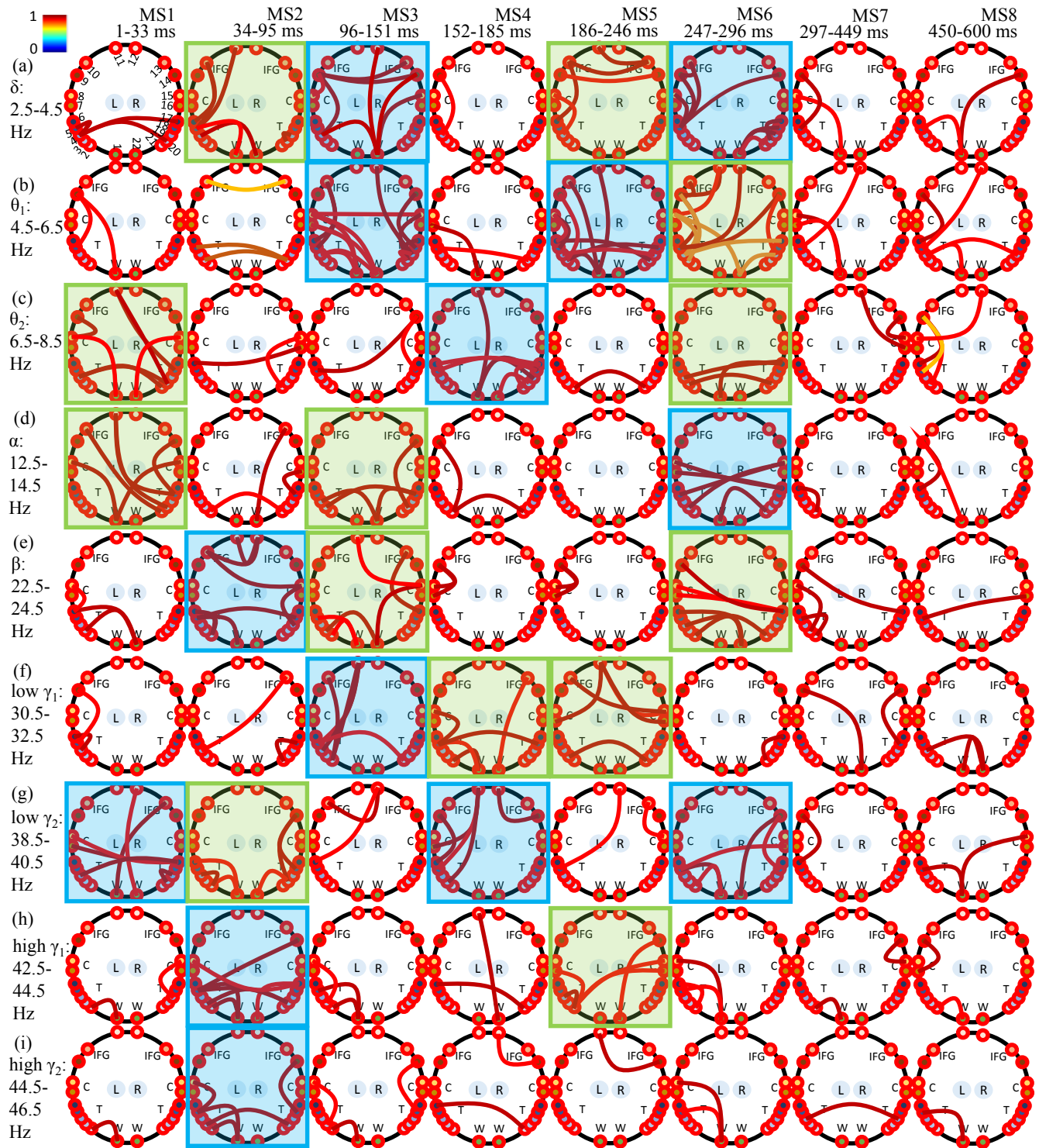
FIGURE 11. Connectograms of microstate FC in: (a) delta (δ: 2.5-4.5 Hz), (b) theta (θ1: 4.5-6.5 Hz), (c) theta (θ2: 6.5-8.5 Hz), (d) alpha (α: 12.5-14.5 Hz), (e) beta (β: 22.5-24.5 Hz, (f) low gamma (low δ1: 30.5-32.5 Hz), (g) low gamma (low δ2: 38.5-40.5 Hz), (h) high gamma (high δ1: 42.5-44.5 Hz), and (i) high gamma (high δ2: 44.5-46.5 Hz) bands. These frequency bands were shown significant differences in perception. Significant segments in perception presented in Fig. 10 are grouped by green rectangles (0.01 < *p*-values ≤ 0.05) and blue rectangles (*p*-values ≤ 0.01). Regions by numbers are provided in Appendix Table 2. L = left hemisphere; R = right hemisphere; T = temporal region; W = Wernicke's region; IFG = inferior frontal gyrus; C = central region.

mean greater statistical independences in covariance of FC matrices, thus represent higher distinctive degrees of unique dynamic FC patterns [22].

To evaluate MGHMM method on speech perception task without ground truth, we computed KL-divergence between μFC-based distributions as a metric to measure connectivity
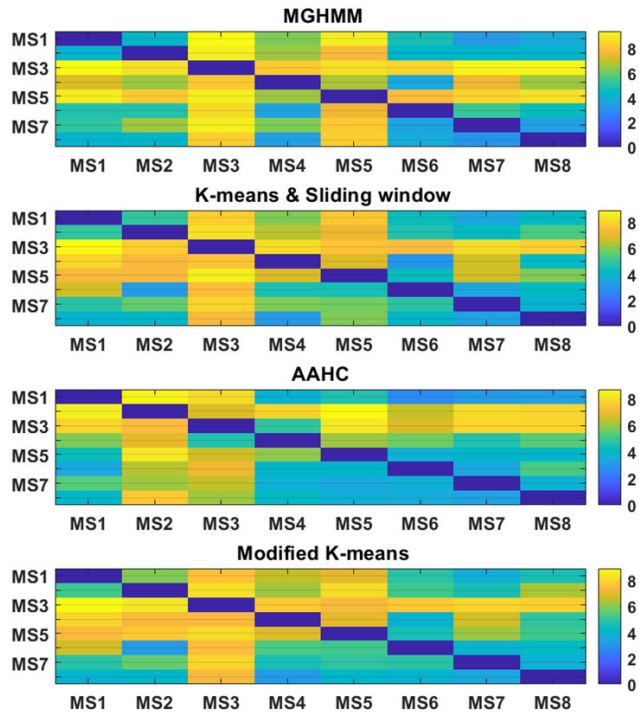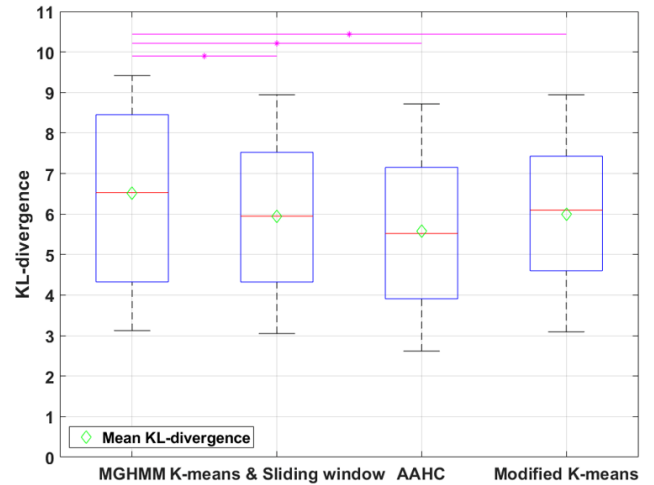
FIGURE 12. KL-divergence presentations of four comparative approaches. An element in each matrix represents the distinction degree between two inferred μFC shown in MGHMM, AAHC and modified K-means and 'FC-states' shown in K-means and sliding window combination methods. Light yellow elements imply high distinction levels while dark blue ones indicate low distinction degrees.

FIGURE 13. Distribution presentations of FC KL-divergence. In average, MGHMM provides the largest distinction degree between estimated μFC among four methods, reflected by the highest mean KL-divergence (green dots) compared to those of K-means clustering and sliding window combination computing 'FC-states' as well as those of Lehmann-based AAHC and modified K-means methods. By using the two-sample t-test, the results also show that KL-divergence distribution of MGHMM method is significantly larger than those of the comparative methods (magenta star: $0.0162 > p$-value > 0.0097).

distinction between sets of estimated microstates (Fig. 12). Theoretically, a larger KL-divergence indicates a greater difference between two μFC patterns, which qualifies our purpose. Overall, in terms of dFC segmentation qualification shown in Fig. 13, we present mean and standard deviations of KL-divergences of four comparative methods as follows: 6.518 ± 2.091 for MGHMM, 5.945 ± 1.722 for K-means and sliding window combination, 5.5794 ± 1.806 for Lehmann-based AAHC, and 5.996 ± 1.643 for Lehmann-based modified K-means approach. These indicators prove that our MGHMM approach outperforms the well-examined K-means clustering combined with sliding window (significant improvement with p-value $= 0.0097$ using two sample t-test, 8.791% improvement) in terms of FC distinction qualification. In terms of microstate-guided functional connectivity comparisons, our results show that the performance by MGHMM approach is superior to those of Lehmann's approaches, particularly, 14.41% improvement (p-value $= 0.006$) over AAHC, and 8.01% improvement (p-value $= 0.0162$) over modified K-means methods.

IV. DISCUSSIONS
A. BRAIN DYNAMICS IN SPEECH PERCEPTION
State-of-the-art neuroimaging and machine learning in computational neurosciences have offered novel strategies to study brain mechanisms [49]–[52]. This article introduced a new framework which, when applied to multivariate high

temporal resolution EEG, revealed microstate source generators and functional connectomics coordinated in speech perception. Our microstate decomposition algorithm, MGHMM, adopted Bayesian variational learning, revealed optimal eight EEG microstate segments in which each microstate represents for unique dynamic patterns (see Fig. 5 and Table 1).

We also revealed microstate language-relevant source generators in perception process. This is among first EEG studies to demonstrate microstate alternations of cortical clusters during speech perception, as most existing source-space approaches assumed stationarity. Unsupervised classification of the multivariate ERP into quasi-stable intervals with MGHMM has insights compared to any stationary source reconstruction methods. Here, such insights were achieved by showing microstate segments when language-relevant regions appear (Fig. 6a-h) which we cannot see in macrostate scope (Fig. 6i). Generally, as shown in Fig. 6, strongest activations (highest t-values) of bilateral pSTG, IFG, and anterior MTG and ITG with left-side dominances can be observed during MS3 (96-151 ms). The results are consistent with previous studies and functionally referred to phonological and lexical-semantic processing. Furthermore, we found significant involvements of bilateral precentral areas where the primary motor cortices are located. This involvement may be due to the reflections from sounds to articulation and the semantic action representation in speech.

We found that perceiving vowel stimuli significantly increased connectivity in delta (2.5-4.5 Hz), theta (4.5-6.5, 6.5-8.5 Hz), alpha (12.5-14.5 Hz), beta (22.5-24.5 Hz), low gamma (30.5-32.5, 38.5-40.5 Hz) but decreased connectivity in high gamma (42.5-46.5 Hz) bands (see Fig. 8). In parallel, increased FC were observed mainly in; (1) microstate

segments 34-95 ms (MS2) and 96-151 ms (MS3) in early stages of perception as well as (2) microstate intervals 186-246 ms (MS5) and 247-296 ms (MS6) in late perception stages (Fig. 10).

As shown in Fig. 11, different brain regions were connected and the dynamic connections were observed throughout the perception at different frequency patterns. For instance, at beginning of speech perception from the onset of stimuli to 33 ms (MS1), delta band intra-connections between temporal regions at left hemisphere and connections between bilateral temporal regions were found. At MS2 (34-95 ms), there was an information flow between left temporal cortex and left inferior frontal cortex and information flowed between the left temporal cortex and bilateral supramarginal areas. Then from 96 ms to 151 ms at MS3, the connectivity between left precentral and left inferior frontal areas as well as connections between bilateral inferior frontal regions appeared. After that, we observed the connectivity between the left temporal and left inferior frontal areas in MS4 from 152 ms to 185 ms. The bilateral inferior frontal connections, also the information flow between left temporal and precentral region can be seen in the time segment at 186-246 ms (MS5). As shown at MS6 interval, the connections between left temporal and left inferior frontal cortices were found. In addition, we observed associations between right temporal and right supramarginal regions.

The fundamental principle of dynamic FC studies in neuroscience is to develop a method that perfectly partitions time-varying FC patterns into a set of unique truth 'FC-states' in which a 'FC-state' represents for a unique brain network incorporated the task. Without the ground truth FC patterns, no one can be able to assure perfections of their methods. In this study, we defined a comparative protocol that measures the distinction levels of the independency between extracted networks. We used KL-divergence as a metric to measure such distinction degrees. As shown in the results, our MGHMM method was able to detect the 'true' dynamic networks in the speech perception as it significantly outperformed the two long-established K-means and sliding window methods which are used to estimate the FC-states. Particularly, we showed performance improvements of 8.791% (p-value = 0.0097) when compared to K-means and time sliding window (see results in Fig. 12 and Fig. 13). In addition, as seen in Fig. 12, our MGHMM significantly outperformed the other two Lehmann's based AAHC (14.41% improvement with p-value = 0.006) and modified K-means (8.01% improvement with p-value = 0.0162) approaches in estimation of microstate-guided functional connectivity.

B. DYNAMIC FUNCTIONAL CONNECTIVITY VERSUS SOURCE IMAGING

In this work, there is a critical concern toward differences between microstate source localization (Fig. 6) and microstate FC networks (Fig. 11) coordinated in speech perception. In general, both approaches led to identical results

regarding the significant involvement of bilateral temporal and inferior frontal as well as primary motor cortices during speech perception process. Regarding to dynamic findings, both FC and source imaging results indicated that for speech perception the brain mainly coordinated at microstate segments 96-151 ms (MS3) and 186-246 ms (MS5) as well as no significant activities (neither source activations nor connectivity) at microstate segments 297-449 ms (MS7) and 450-600 ms (MS8) after the onset of speech stimuli. However, our results revealed that, in both cases, the information elicited from high temporal EEG were significantly different. Intuitively, the fundamental difference is that cortical generators determined regions with high activations, but completely omitted functional associations between them. In contrast, the brain FC approach only revealed regions that were highly interacted regardless of their amplitudes. For instance, microstate source localization showed modest activations in microstate segments 34-95 ms (MS2) and 247-296 ms (MS6) but microstate FC revealed significant connectivity in these segments.

C. METHODOLOGICAL LIMITATIONS AND FUTURE STUDIES

Our framework, the backbone of which is the EEG microstate, offers a new direction to study dynamic characteristics in speech perception. To achieve this, there are four core components of the framework that warrant discussion: (1) broadband EEG microstate decomposition – MGHMM, (2) source mapping – wMNE, (3) FC metric – Icoh, and (4) brain dynamic functional connectivity oriented by the estimated broadband EEG microstates representing the broadband EEG data.

First, regarding EEG microstate decomposition method, the MGHMM was selected based on its successful performance in previous task-based M/EEG cognitive studies [42], [43]. It should be borne in mind that three assumptions were made to successfully develop the algorithm: i) independent Markov hidden variables, ii) multivariate Gaussian observation model, and iii) independent hierarchical structures. To train model parameters, we employed KL–divergence Bayesian variational learning. These assumptions, however, might be too unsophisticated to fully reveal complexities underlying the brain dynamics. Therefore, more advanced methods could be developed for elegantly modelling complex time-varying EEG with excellent accuracies and provide appropriate results for clustering ERP data into quasi-stable EEG microstates and thus can be thought for future studies [38].

Second concern relates to source localization method – weighted Minimum Norm. wMNE is a popular method of inverse solution and has been demonstrated previously to be particularly useful [24] in reconstructions of cortical signals. The wMNE compensates for the tendency of the MNE, which is based on a search for a solution with minimum energy using L2 norms to regularize the inverse problem, to favor cortical sources closer to the sensors. However, we do not

rule out other inverse solutions, which could be substituted for wMNE, and they would likely provide similar results.

Third, we used Icoh to measure FC between ROIs. This method has been successful in elucidating FC networks at each time instant. An advantage property of Icoh is that its (non-vanishing) finite value cannot be caused by a linear mixing of uncorrelated sources ('volume conduction') and thus reflects 'true' interactions between two sources. However, Icoh is not yet a perfect measure of coupling since it depends upon the strength of the coupling and the magnitude of the phase difference [44]. Therefore, other approaches such as phase lag index, Granger causality and phase locking value can provide different interpretations of functional connectivity.

Fourth, in this work we assumed that the brain dynamic functional connectivity are guided by the segmented broadband EEG microstates on the sensor level. On the source level, we thus estimated functional brain networks by taking the time points aligned with the temporal segments of every microstate extracted by the multivariate MGHMM method. For the comparative methods using Lehmann's approaches (AAHC and modified K-means) the process to compute brain networks was similar, however, the time points corresponding to each estimated EEG microstate are different. This assumption, however, might be too simple to reveal the 'true' dynamics of the underlying brain networks. Therefore, subsequent studies should be focused on estimation of brain dynamic connectivity with the assumption that source activity orients brain connectivity.

We demonstrated that common dynamic brain networks induced by speech perception of all vowel stimuli may be resolved into a set of stationary μFC patterns and the framework was performed on the averaged data; i.e., grand mean ERP (mean of ERPs). The advantage of this averaged ERP is that it preserves common networks to all stimuli and minimizes inter-vowel variability [11]. Indeed, previous studies (speech perception using M/EEG) have noted that the ERPs and source generators of vowels have commonalities [4], [5], [8]. That is, ERP waveforms of each stimulus are comparatively similar and the source generators are also relatively close. Therefore, it is reasonable that the grand mean ERP can preserve common EEG microstates; for instance, significant activations in MS3 (95-151 ms). Note that in this work, dynamics responding separately to each vowel were not studied. There could be variations (ERPs' amplitudes and delays) in responses for individual vowels [11], which may lead to different microstate cortical generators and μFC modulated by individual vowels. Therefore, considerably more effort must be expended in this area, in order to comprehend and interpret the dynamic specificities separately evoked by each speech stimulus, which could be a potential direction of our future studies. In addition, it is worth noting that the experiment designed with only vowels (one of the two principal classes of speech sounds) performed in this work can be too simple to elucidate the complete characteristics of speech perception. Subsequent studies can also be extended

by including more sophisticated human speech conditions, thus, the full mechanism of human brain when perceiving the human speech can be exposed.

It should be emphasized that we assume: i) there is an anatomical and functional correspondence between subjects (a template 3D cortex used as a source model) and ii) transitions between decomposed EEG microstates occur in a repeatable manner across subjects. However, we are fully aware that there may be variability in the timecourse of reconstructed sources and connectivity across subjects. Given inter-individual variations in stationary source reconstructions and connectivity [53], it is not surprising that dynamic functional representations of neural oscillations also exhibit relatively limited inter-individual differences. Ultimately, in order to solve inter-subject variability, future works must be done.

In this work, we only focused on analyzing the dynamics of brain networks at the broadband manner. However, the brain dynamics at bandpass level is apparently of interest to fully reveal the neural mechanisms underlying the cognitive task. Thus, in next studies we plan to estimate source-level frequency-dependent functional brain networks using MGHMM. These could be done by multiple approaches. One approach is to source-localize the broadband EEG and run the MGHMM on band-limited frequency source activity and then estimate the dynamics of brain networks with the assumption that the activity orients brain connectivity. Second approach is to estimate source time-series in frequency band of interests independently and run the MGHMM on these bandpass source time-series, and then estimate brain network dynamics with the assumption that activity orients brain connectivity independently at each frequency limit.

Last but not least, generalizability of the findings in this study could be less powerful due to the small sample size (eleven male subjects). In neuroscience community, it has been claimed that small sample sizes reduce the replicability of drawn conclusions. Thus, the current results need to be generalized with a larger appropriate sample size in subsequence studies. Moreover, we only studied the time-varying neural activities of the male subjects. However, the neural-responsed differences between males and females have been reported at multiple levels [54]. Therefore, it would not be surprised if results of female brain dynamics for speech perception were different from the findings found in this study. This leads to our future studies that would be directed toward understanding human brain dynamics including both male and female subjects as well as gender differences of neural dynamic activities in speech perceptions.

V. CONCLUSION

In this work, we introduce a novel direction based on EEG microstate to study brain dynamics during speech perception by taking advantage of high temporal resolution of multivariate patterns in ERP data. Results suggested that the brain coordinated eight distinct EEG microstates, upon which microstate source localizations and functional networks were

based, for sequential process of speech stimuli from hearing to meaning and articulating. The results showed that bilateral temporal, IFG, supramarginal, and precentral gyri (left-side dominance) are not only significantly activated, but also have information exchanges throughout speech perception. Importantly, our results indicated the main processing occurred in microstate segments from 95 to 151 ms and 186 to 296 ms after onset of stimuli. We also revealed the importance of delta (2.5-4.5 Hz), theta (4.5-6.5, 6.5-8.5 Hz), alpha (12.5-14.5 Hz), beta (22.5-24.5 Hz), low gamma (30.5-32.5, 38.5-40.5 Hz), and high gamma (42.5-46.5 Hz) band connectivity in speech perception for the early perception process at MS1, MS2, and MS3 and late perception process at MS4, MS5 and MS6. Our quantitative comparisons showed the significant outperformance (8.791% improvement, p-values<0.0097) of MGHMM method over existing K-means and sliding window combination method. Furthermore, the findings show that our MGHMM significantly outperformed the other two Lehmann's based AAHC (14.41% improvement with p-value $= 0.006$) and modified K-means (8.01% improvement with p-value $= 0.0162$) approaches in estimation of microstate-guided functional connectivity. However, it should be noted that these current results obtained with a modest sample size should be extended in future with a larger appropriate number of subjects that should include female participants in order to achieve universal findings. Thus, this work could be useful to provide a neurophysiological reference for clinical studies related to deficits in speech perception.

APPENDIX
See Table 2 here.

TABLE 2. 22 Language-relevant roi numbers (#) and names used in microstate FC analyses in Fig. 11.

#	ROI name – Left side	#	ROI name – Right side
1	Supramarginal gyrus-L	22	Supramarginal gyrus-R
2	Inferior temporal gyrus-L	21	Inferior temporal gyrus-R
3	Middle temporal gyrus-L	20	Middle temporal gyrus-R
4	Transverse temporal-L	19	Transverse temporal-R
5	Superior temporal gyrus-L	18	Superior temporal gyrus-R
6	Temporal pole-L	17	Temporal pole-R
7	Post central gyrus-L	16	Post central gyrus-R
8	Precentral gyrus-L	15	Precentral gyrus-Right
9	Pars opercularis gyrus-L	14	Pars opercularis gyrus-R
10	Pars triangularis gyrus- L	13	Pars triangularis gyrus-R
11	Pars orbitalis gyrus-L	12	Pars orbitalis gyrus-R

REFERENCES

[1] J. Brauer, A. Anwander, D. Perani, and A. D. Friederici, "Dorsal and ventral pathways in language development," *Brain Lang.*, vol. 127, no. 2, pp. 289–295, Nov. 2013.

[2] J. P. Rauschecker and S. K. Scott, "Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing," *Nature Neurosci.*, vol. 12, no. 6, pp. 718–724, Jun. 2009.

[3] J. Obleser and F. Eisner, "Pre-lexical abstraction of speech in the auditory cortex," *Trends Cognit. Sci.*, vol. 13, no. 1, pp. 14–19, Jan. 2009.

[4] E. Diesch, C. Eulitz, S. Hampson, and B. Ross, "The neurotopography of vowels as mirrored by evoked magnetic field measurements," *Brain Lang.*, vol. 53, no. 2, pp. 143–168, 1996.

[5] J.-L. Schwartz, A. Basirat, L. Ménard, and M. Sato, "The Perception-for-Action-Control theory (PACT): A perceptuo-motor theory of speech perception," *J. Neurolinguistics*, vol. 25, no. 5, pp. 336–354, Sep. 2012.

[6] K. Strijkers and A. Costa, "The cortical dynamics of speaking: Present shortcomings and future avenues," *Lang., Cognition Neurosci.*, vol. 31, no. 4, pp. 484–503, Apr. 2016.

[7] M. R. Schomers and F. Pulvermäller, "Is the sensorimotor cortex relevant for speech perception and understanding? An integrative review," *Frontiers Hum. Neurosci.*, vol. 10, p. 435, Sep. 2016.

[8] N. Mainy, J. Jung, M. Baciu, P. Kahane, B. Schoendorff, L. Minotti, D. Hoffmann, O. Bertrand, and J.-P. Lachaux, "Cortical dynamics of word recognition," *Hum. Brain Mapping*, vol. 29, no. 11, pp. 1215–1230, Nov. 2008.

[9] P. Indefrey and A. Cutler, "Prelexical and lexical processing in listening," in *The Cognitive Neurosciences*. Cambridge, MA, USA: MIT Press, 2004, pp. 759–774.

[10] R. Cabeza and L. Nyberg, "Imaging cognition II: An empirical review of 275 PET and fMRI studies," *J. Cognit. Neurosci.*, vol. 12, no. 1, pp. 1–47, Jan. 2000.

[11] M. Bastiaansen and P. Hagoort, "Oscillatory neuronal dynamics during language comprehension," *Prog. Brain Res.*, vol. 159, pp. 179–196, 2006.

[12] G. Hickok and D. Poeppel, "The cortical organization of speech processing," *Nature Rev. Neurosci.*, vol. 8, no. 5, pp. 393–402, May 2007.

[13] A. D. Friederici, "The cortical language circuit: From auditory perception to sentence comprehension," *Trends Cognit. Sci.*, vol. 16, no. 5, pp. 262–268, 2012.

[14] C. M. Michel and T. Koenig, "EEG microstates as a tool for studying the temporal dynamics of whole-brain neuronal networks: A review," *NeuroImage*, vol. 180, pp. 577–593, Oct. 2018.

[15] S. Zhang, S. Tian, M. R. Chattun, H. Tang, R. Yan, K. Bi, Z. Yao, and Q. Lu, "A supplementary functional connectivity microstate attached to the default mode network in depression revealed by resting-state magnetoencephalography," *Prog. Neuro-Psychopharmacol. Biol. Psychiatry*, vol. 83, pp. 76–85, Apr. 2018.

[16] S. Cacioppo and J. T. Cacioppo, "Dynamic spatiotemporal brain analyses using high-performance electrical neuroimaging, Part II: A step-by-step tutorial," *J. Neurosci. Methods*, vol. 256, pp. 184–197, Dec. 2015.

[17] R. M. Hutchison, T. Womelsdorf, E. A. Allen, P. A. Bandettini, V. D. Calhoun, M. Corbetta, S. Della Penna, J. H. Duyn, G. H. Glover, J. Gonzalez-Castillo, D. A. Handwerker, S. Keilholz, V. Kiviniemi, D. A. Leopold, F. de Pasquale, O. Sporns, M. Walter, and C. Chang, "Dynamic functional connectivity: Promise, issues, and interpretations," *NeuroImage*, vol. 80, pp. 360–378, Oct. 2013.

[18] R. Hindriks, M. H. Adhikari, Y. Murayama, M. Ganzetti, D. Mantini, N. K. Logothetis, and G. Deco, "Can sliding-window correlations reveal dynamic functional connectivity in resting-state fMRI?" *NeuroImage*, vol. 127, pp. 242–256, Feb. 2016.

[19] M. G. Preti, T. A. Bolton, and D. Van De Ville, "The dynamic functional connectome: State-of-the-art and perspectives," *NeuroImage*, vol. 160, pp. 41–54, Oct. 2017.

[20] S. Shakil, C.-H. Lee, and S. D. Keilholz, "Evaluation of sliding window correlation performance for characterizing dynamic functional connectivity and brain states," *NeuroImage*, vol. 133, pp. 111–128, Jun. 2016.

[21] N. Karamzadeh, A. Medvedev, A. Azari, A. Gandjbakhche, and L. Najafizadeh, "Capturing dynamic patterns of task-based functional connectivity with EEG," *NeuroImage*, vol. 66, pp. 311–317, 2013.

[22] N. T. Duc and B. Lee, "Microstate functional connectivity in EEG cognitive tasks revealed by a multivariate Gaussian hidden Markov model with phase locking value," *J. Neural Eng.*, vol. 16, no. 2, Apr. 2019, Art. no. 026033.

[23] E. A. Allen, E. Damaraju, S. M. Plis, E. B. Erhardt, T. Eichele, and V. D. Calhoun, "Tracking whole-brain connectivity dynamics in the resting state," *Cerebral Cortex*, vol. 24, no. 3, pp. 663–676, Mar. 2014.

[24] M. Hassan, P. Benquet, A. Biraben, C. Berrou, O. Dufor, and F. Wendling, "Dynamic reorganization of functional brain networks during picture naming," *Cortex*, vol. 73, pp. 276–288, Dec. 2015.

[25] A. Mheich, M. Hassan, M. Khalil, C. Berrou, and F. Wendling, "A new algorithm for spatiotemporal analysis of brain functional connectivity," *J. Neurosci. Methods*, vol. 242, pp. 77–81, Mar. 2015.

[26] S. I. Dimitriadis, N. A. Laskaris, V. Tsirka, M. Vourkas, S. Micheloyannis, and S. Fotopoulos, "Tracking brain dynamics via time-dependent network analysis," *J. Neurosci. Methods*, vol. 193, no. 1, pp. 55–145, Oct. 2010.

[27] S. I. Dimitriadis, N. A. Laskaris, and A. Tzelepi, "On the quantization of time-varying phase synchrony patterns into distinct functional connectivity microstates (FCmustates) in a multi-trial visual ERP paradigm," *Brain Topogr*, vol. 26, no. 3, pp. 397–409, Jul. 2013.

[28] W. Jamal, S. Das, and K. Maharatna, "Existence of millisecond-order stable states in time-varying phase synchronization measure in EEG signals," in *Proc. 35th Annu. Int. Conf. Eng. Med. Biol. Soc. (EMBC)*, Oct. 2013, pp. 2539–2542.

[29] S. I. Dimitriadis, N. A. Laskaris, Y. Del Rio-Portilla, and G. Koudounis, "Characterizing dynamic functional connectivity across sleep stages from EEG," *Brain Topogr*, vol. 22, no. 2, pp. 33–119, Sep. 2009.

[30] S. I. Dimitriadis, N. A. Laskaris, and S. Micheloyannis, "Transition dynamics of EEG-based network microstates during mental arithmetic and resting wakefulness reflects task-related modulations and developmental changes," *Cognit. Neurodyn.*, vol. 9, no. 4, pp. 371–387, Aug. 2015.

[31] S. I. Dimitriadis, N. A. Laskaris, V. Tsirka, M. Vourkas, and S. Micheloyannis, "An EEG study of brain connectivity dynamics at the resting state," *Nonlinear Dyn. Psychol Life Sci*, vol. 16, no. 1, pp. 5–22, Jan. 2012.

[32] S. I. Dimitriadis, B. Routley, D. E. Linden, and K. D. Singh, "Reliability of static and dynamic network metrics in the resting-state: A MEG-beamformed connectivity analysis," *Frontiers Neurosci.*, vol. 12, p. 506, Aug. 2018.

[33] S. I. Dimitriadis, "Altered temporal correlations in resting-state connectivity fluctuations in children with reading difficulties detected via MEG," *NeuroImage*, vol. 83, pp. 307–317, Dec. 2013.

[34] R. F. Betzel, M. A. Erickson, M. Abell, B. F. O'Donnell, W. P. Hetrick, and O. Sporns, "Synchronization dynamics and evidence for a repertoire of network states in resting EEG," *Frontiers Comput. Neurosci.*, vol. 6, p. 74, Oct. 2012.

[35] A. G. Mahyari, D. M. Zoltowski, E. M. Bernat, and S. Aviyente, "A tensor decomposition-based approach for detecting dynamic network states from EEG," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 1, pp. 225–237, Jan. 2017.

[36] N. Leonardi, J. Richiardi, M. Gschwind, S. Simioni, J.-M. Annoni, M. Schluep, P. Vuilleumier, and D. Van De Ville, "Principal components of functional connectivity: A new approach to study dynamic brain connectivity during rest," *NeuroImage*, vol. 83, pp. 937–950, Dec. 2013.

[37] A. Custo, D. Van De Ville, W. M. Wells, M. I. Tomescu, D. Brunet, and C. M. Michel, "Electroencephalographic resting-state networks: Source localization of microstates," *Brain Connectivity*, vol. 7, no. 10, pp. 671–682, Dec. 2017.

[38] M. M. Murray, D. Brunet, and C. M. Michel, "Topographic ERP analyses: A Step-by-Step tutorial review," *Brain Topography*, vol. 20, no. 4, pp. 249–264, Jun. 2008.

[39] R. Oostenveld, P. Fries, E. Maris, and J.-M. Schoffelen, "FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data," *Comput. Intell. Neurosci.*, vol. 2011, pp. 1–7, May 2011.

[40] R. S. Desikan, F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, M. S. Albert, and R. J. Killiany, "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest," *NeuroImage*, vol. 31, no. 3, pp. 968–980, Jul. 2006.

[41] F. Tadel, S. Baillet, J. C. Mosher, D. Pantazis, and R. M. Leahy, "Brainstorm: A user-friendly application for MEG/EEG analysis," *Comput. Intell. Neurosci.*, vol. 2011, Apr. 2011, Art. no. 879716.

[42] M. W. Woolrich, A. Baker, H. Luckhoo, H. Mohseni, G. Barnes, M. Brookes, and I. Rezek, "Dynamic state allocation for MEG source reconstruction," *NeuroImage*, vol. 77, pp. 77–92, Aug. 2013.

[43] D. Vidaurre, A. J. Quinn, A. P. Baker, D. Dupret, A. Tejero-Cantero, and M. W. Woolrich, "Spectrally resolved fast transient brain states in electrophysiological data," *NeuroImage*, vol. 126, pp. 81–95, Feb. 2016.

[44] C. J. Stam, G. Nolte, and A. Daffertshofer, "Phase lag index: Assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources," *Human Brain Mapping*, vol. 28, no. 11, pp. 1178–1193, 2007.

[45] M.-È. Desjardins, J. Carrier, J.-M. Lina, M. Fortin, N. Gosselin, J. Montplaisir, and A. Zadra, "EEG functional connectivity prior to sleepwalking: Evidence of interplay between sleep and wakefulness," *Sleep*, vol. 40, no. 4, p. 2, Apr. 2017.

[46] G. Nolte, O. Bai, L. Wheaton, Z. Mari, S. Vorbach, and M. Hallett, "Identifying true brain interaction from EEG data using the imaginary part of coherency," *Clin. Neurophysiol.*, vol. 115, no. 10, pp. 2292–2307, Oct. 2004.

[47] T. Koenig, "Millisecond by millisecond, year by year: Normative EEG microstates and developmental stages," *NeuroImage*, vol. 16, no. 1, pp. 8–41, May 2002.

[48] F. von Wegner, P. Knaut, and H. Laufs, "EEG microstate sequences from different clustering algorithms are information-theoretically invariant," *Frontiers Comput. Neurosci.*, vol. 12, p. 70, Aug. 2018.

[49] D. T. Nguyen, S. Ryu, M. N. I. Qureshi, M. Choi, K. H. Lee, and B. Lee, "Hybrid multivariate pattern analysis combined with extreme learning machine for Alzheimer's dementia diagnosis using multi-measure rs-fMRI spatial patterns," *PLoS ONE*, vol. 14, no. 2, Feb. 2019, Art. no. e0212582.

[50] N. T. Duc, S. Ryu, M. N. I. Qureshi, M. Choi, K. H. Lee, and B. Lee, "3D-deep learning based automatic diagnosis of Alzheimer's disease with joint MMSE prediction using resting-state fMRI," *Neuroinformatics*, vol. 18, no. 1, pp. 71–86, Jan. 2020.

[51] M. N. I. Qureshi, B. Min, H.-J. Park, D. Cho, W. Choi, and B. Lee, "Multiclass classification of word imagination speech with hybrid connectivity features," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 10, pp. 2168–2177, Oct. 2018.

[52] J. Kim, S.-K. Lee, and B. Lee, "EEG classification in a single-trial basis for vowel speech perception using multivariate empirical mode decomposition," *J. Neural Eng.*, vol. 11, no. 3, Jun. 2014, Art. no. 036010.

[53] G. C. O'Neill, P. K. Tewarie, G. L. Colclough, L. E. Gascoyne, B. A. E. Hunt, P. G. Morris, M. W. Woolrich, and M. J. Brookes, "Measurement of dynamic task related functional networks using MEG," *NeuroImage*, vol. 146, pp. 667–678, Feb. 2017.

[54] E. Choleris, L. A. Galea, F. Sohrabji, and K. M. Frick, "Sex differences in the brain: Implications for behavioral and biomedical research," *Neurosci. Biobehav. Rev.*, vol. 85, pp. 126–145, Dec. 2018.

NGUYEN THANH DUC (Member, IEEE) received the B.S. degree in electrical and electronic engineering from Vietnam National University–Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam, in 2010, the M.Sc. degree from the Tampere University of Technology, Tampere, Finland, in 2013, and the Ph.D. degree from the Department of Biomedical of Science and Engineering, Gwangju Institute of Science and Technology, Gwangju, South Korea, in 2020. His current research interests include biomedical signal and brain image processing, and machine learning applications in neuroimaging.

BOREOM LEE (Member, IEEE) graduated from the College of Medicine, Seoul National University, Seoul, South Korea. He received the M.D. and Ph.D. degrees in biomedical engineering from Seoul National University, in 1998 and 2007, respectively.

He joined the faculty of the Gwangju Institute of Science and Technology (GIST), in 2011, where he is currently an Associate Professor with the Department of Biomedical Science and Engineering (BMSE). His current research interests include brain connectomics, healthcare systems, biomedical signal processing and instrumentation, and machine learning application in medical research areas.

• • •