Temporal-Spatial-Spectral Investigation of Brain Network Dynamics in Human Speech Perception and Production

Bin Zhao[1, 2]

[1] Japan Advanced Institute of Science and Technology, Japan
[2] College of Intelligence and Computing, Tianjin University, China
zhaobeiyi@tju.edu.cn

1. Introduction

Speech functions, as an incredible manifestation of human intelligence, entail intricate coordination of brain network dynamics across temporal, spatial, and spectral scales [1]. Current brain investigation techniques such as functional magnetic resonance imaging (fMRI), positron emission tomography (PET), electro- and magnetoencephalography (EEG/MEG) have revealed multi-facets of brain activities and supported our understanding of speech processing from different perspectives [2, 3]. However, a systematic framework incorporating all three of temporal, spatial, and spectral aspects as well as relate them to cognitive-behavioral responses is still yet to be constructed.

1.1. Motivation of research

To provide a comprehensive view of the whole-brain, whole-range network dynamics during speech processing from the temporal, spatial, and spectral aspects, this study is designed to combine different modalities, such as brain imaging, electrophysical and behavioral data with advanced brain network analysis and multimodal data integration methods in various speech perception and production tasks. The systematic revelation and the insights from speech neuroscience are promising for the rehabilitation of patients with speech disorders caused by brain injury, but also for the development of human-computer interaction technology and the progress of artificial intelligence with speech communication.

1.2. Key issues identified

To track the transient spatiotemporal brain dynamics during speech processing is a highly challenging task. The difficulties lie not only in the unsatisfactory spatial and temporal precision limited by different modalities of brain investigation techniques but also in the complexity of integrating those different modalities of data into a systematic framework, letting along relating the neurophysiological data with cognitive behaviors. To date, the most prevailing speech processing model – the dual-stream model indicated the regional functionality and information flow pathways. However, temporal dynamics and frequency-specific oscillatory patterns are still unclear.

Moreover, speech processing is not supposed to be a simple feedforward procedure where incoming linguistic information gradually transforms from a string of linguistic segments into a comprehensive concept. Rather, contextual information from long-term memory such as semantics and syntactic can also provide feedback information, thus affecting the perception of sensory input [4, 5]. Therefore, we need to clarify how such bidirectional interactions operate on neural substance and oscillatory patterns for smooth communication.

2. Methods

2.1. Data acquisition

This study utilized a multi-modal data acquisition system to simultaneously record brain waves, eye movement, and speech signals while participants perform speech perception and production tasks. Figure 1 demonstrated the procedure of a natural reading process. EEG signals were recorded from the scalp of the participants with a 128-channel Quik-Cap (Neuroscan, USA) at 1000 Hz. Eye movements were recorded via a monocular pupil tracking system (Eyelink 1000, SR Research Ltd., Mississauga, Canada) at 100 Hz. The speech signals from articulation were recorded using a microphone (SONY ECM MS957) at 44100 Hz.

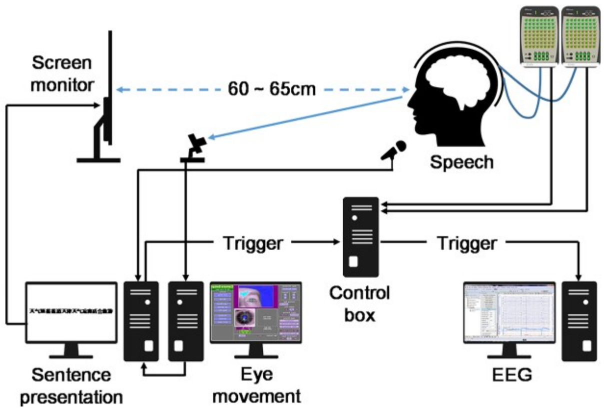

Figure 1: *Schematic diagram of data acquisition.*

2.2. Behavioral analysis

Behavioral analyses were conducted on the eye movement and speech signals by segmenting the eye onset and offset as well as speech onset and offset. The time indexes of these on- and off-sets are indicative of different stages of cognitive processing and provide references for the subsequent neural investigation.

2.3. EEG source reconstruction

The raw EEG signals recorded directly from the scalp is a summed electrical field potential from cortical neurons blurred with non-brain artifacts, it suffers from a high risk of false positives from volume conduction and difficulty in anatomical interpretability [6]. Thus, we performed independent component analysis (ICA) [7] to separate brain sources from those biological artifacts. Equivalent current dipoles (ECD) of the effective components were then computed using a boundary element model (BEM) [8] for cortical localization.

2.4. Brain network analysis

Granger-Causality (GC)-based effective connectivity analysis is used to calculate the information flow within brain networks based on the analysis of prediction errors of autoregressive models. The estimator implemented in this study is the direct directed transfer function (dDTF) which is the product of a full-frequency directed transfer function (enabling more interpretable comparisons of information flow at different frequencies) and the partial coherence (a conditional coherence that cannot be explained by a linear coherence between components) [9]. A segmentation-based adaptive multivariate auto-regressive model was constructed with a 500-ms sliding window and a step size of 25 ms. The model order of 10 was selected based on the Vieira-Morf lattice algorithm.

Frequency-specific network analyses were also conducted to examine functional specifications for brain network dynamics in different frequency bands. Moreover, phase-amplitude coupling (PAC) [10] and intersite phase clustering (ISPC) [11] were performed to investigate the cross-frequency coupling phenomenon and bidirectional information interaction which were carried by different rhythmic oscillations.

2.5. Multimodal integration

To relate different modalities of brain activities, we introduced an fMRI-based Morphological and Connectomic Atlas of Human Brain Functions into a representational similarity analysis [12]. Each fMRI template corresponds to a network distribution for a specific function. By calculating the similarity between the fMRI network distribution and each frame of the previously constructed EEG-based connectivity by means of correlation coefficient, we could estimate the activity strengths of that fMRI-defined functional network that presented in our real-time EEG signal at that time point. And with the point-by-point time series of such activity strengths, we could also compare the temporal significance of a specific functional network along the whole speech perception procedure.

3. Discussion of results

By examining the regional multi-frequency oscillatory patterns, frequency-specific brain network dynamics, and cross-frequency coupling characteristics, our results revealed a hierarchical cortical organization and frequency-specific bidirectional information flow patterns. These functional networks progressed parallelly for the completion of the complex speech perception and production processes.

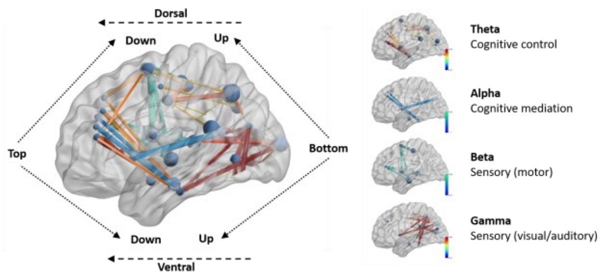

Figure 2: *Hierarchical organization of bidirectional interaction in speech processing.*

An overview of our results is illustrated in Figure 2. Two significant oscillatory patterns appeared in the frequency-resolved network dynamics. The first is the oscillatory hierarchy of the cortical framework. Consistent with convergent evidence [13-18], there is an inverse relationship exists between interaction scale and oscillation frequency. Specifically, high-frequency oscillations (e.g., gamma and beta) with fast cycles are most suitable for binding together fine structures within a small patch. High gamma activity was found in the visual and auditory cortices for processing sensory details that fit into the gamma cycles. Beta oscillations, usually in the form of suppression, were commonly observed in the sensorimotor regions during the execution of motor output. In comparison, low-frequency resonances (e.g., alpha and theta) have the advantage of long-range synchronization at larger time scales. So, it is optimal for grouping sub-phenetic fragments into a meaningful concept at the lexical and phrase levels. Moreover, our results showed that each functional stage is not dominated by a single frequency band. Instead, they benefit from the coexistence of rhythmic oscillation at different frequency bands. The co-existence multi-frequencies across areas is especially beneficial for exchanging information in both feedforward and feedback directions.

The second pattern is the bidirectional interaction of dual-stream processing via different frequency channels. Our results verified parallel processing in the dual streams regarding its spatial functionality, i.e., the ventral stream for speech comprehension and the dorsal stream for speech production. We further expanded the dual-stream model with additional temporal and spectral details. Based on the temporal-spectral dynamics, we discovered that neither of the dual streams flowed in a one-way direction but were subject to both bottom-up integration and top-down regulation. The goal-directed, sensory-guided behavior relies on both feedforward and feedback interactions between brain regions. These large-scale interactions are reflected by the phase coherence and amplitude correlation of oscillations between brain regions in different frequency bands.

4. Future plans and road map

In our future work, we would like to incorporate more functional modalities (e.g. electrocorticography (ECoG) and magnetoencephalography (MEG), etc.) to complement our current study with more precise temporal and spatial resolutions. Speech tasks will also be gradually complexified to explore different levels of speech processing on the cortical hierarchy. So far, our brain investigation of speech perception and production functions is conducted in a relatively independent manner, which is intended to clarify their separate functionality. But indeed, speech perception and production are an evolved coherent process that is neither efficient nor likely to work independently. So, our next investigation step will integrate perception and production in a more natural conversation scenario and examine how the speech chain is rolled out between listeners and speakers in perception and production.

5. Acknowledgements

I would like to express my sincere gratitude towards my supervisor Prof. Jianwu Dang, Gaoyan Zhang, and Kiyoshi Honda who have led me all along the road of exploration. I would also like to thank all other mentors and researchers in the speech and neuroscience community who has provided valuable instructions and guidance for the progression of young students. This study is supported in part by JSPS KAKENHI Grant (20K11883), and in part by the National Natural Science Foundation of China (No.61876126).

6. References

[1] Hagoort, P., *The neurobiology of language beyond single-word processing.* Science, 2019. **366**(6461): p. 55-58.

[2] Price, C., *A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading.* Neuroimage, 2012. **65**: p. 816-847.

[3] Munding, D., A.S. Dubarry, and F.X. Alario, *On the cortical dynamics of word production: a review of the MEG evidence.* Language Cognition & Neuroscience, 2017(4): p. 1-22.

[4] Schoffelen, J.-M., et al., *Frequency-specific directed interactions in the human brain network for language.* Proceedings of the National Academy of Sciences, 2017. **114**(30): p. 8083-8088.

[5] Shuai, L. and T. Gong, *Temporal relation between top-down and bottom-up processing in lexical tone perception.* Front Behav Neurosci., 2014. **25**.

[6] Bell, A.J. and T.J. Sejnowski, *An information-maximization approach to blind separation and blind deconvolution*. Neural Comput, 1995. 7(6): p. 1129-59.

[7] Jung, T.P., et al., *Independent Component Analysis of Electroencephalographic and Event-Related Potential Data*. Advances in Neural Information Processing Systems, 1996. 8(8): p. 1548-1551 vol.2.

[8] Oostenveld, R. and T.F. Oostendorp, *Validating the boundary element method for forward and inverse EEG computations in the presence of a hole in the skull*. Human Brain Mapping, 2002. 17(3): p. 179.

[9] Mullen, T., et al., *An Electrophysiological Information Flow Toolbox for EEGLAB.* Biological Cybernetics, 2010.

[10] Jensen, O. and L.L. Colgin, *Cross-frequency coupling between neuronal oscillations.* Trends in Cognitive Sciences, 2007. **11**(7): p. 267-269.

[11] Cohen, M., *Analyzing Neural Time Series Data.* MIT Press, 2014.

[12] Kriegeskorte, N., Mur, M., and Bandettini, P., *Representational Similarity Analysis – Connecting the Branches of Systems Neuroscience*. Frontiers in Systems Neuroscience, vol. 2, p. 4, 2008

[13] Pylkkänen, L., *The neural basis of combinatory syntax and semantics.* Science, 2019. **366**(6461): p. 62-66.

[14] Schoffelen, J.-M., et al., *Frequency-specific directed interactions in the human brain network for language.* Proceedings of the National Academy of Sciences, 2017. **114**(30): p. 8083-8088.

[15] Ding, N., et al., *Cortical Tracking of Hierarchical Linguistic Structures in Connected Speech.* Nature Neuroscience, 2016. **19**(1): p. 158-164.

[16] Bonhage, C.E., et al., *Oscillatory EEG dynamics underlying automatic chunking during sentence processing.* Neuroimage, 2017. **152**: p. 647-657.

[17] Meyer, L., *The Neural Oscillations of Speech Processing and Language Comprehension: State of the Art and Emerging Mechanisms.* European Journal of Neuroscience, 2017. **48**(7).

[18] Lam, N.H.L., et al., *Neural activity during sentence processing as reflected in theta, alpha, beta, and gamma oscillations.* NeuroImage, 2016. **142**: p. 43-54.