Abstract

Human speech processing, either for listening or oral reading, requires dynamic cortical activities that are not only driven by sensory stimuli externally but also influenced by semantic knowledge and speech planning goals internally. Each of these functions has been known to accompany specific rhythmic oscillations and be localized in distributed networks. The question is how the brain organizes these spatially and spectrally distinct functional networks in such a temporal precision that endows us with incredible speech abilities. For clarification, this study conducted an oral reading task with natural sentences and collected simultaneously the involved brain waves, eye movements, and speech signals with high-density EEG and eye movement equipment. By examining the regional oscillatory spectral perturbation and modeling the frequency-specific interregional connections, our results revealed a hierarchical oscillatory mechanism, in which gamma oscillation entrains with the fine-structured sensory input while beta oscillation modulated the sensory output. Alpha oscillation mediated between sensory perception and cognitive function via selective suppression. Theta oscillation synchronized local networks for large-scale coordination. Differing from a single function-frequency-correspondence, the coexistence of multi-frequency oscillations was found to be critical for local regions to communicate remotely and diversely in a larger network.

Index Terms: neural oscillation, brain network dynamics, hierarchical organization, sentence processing

1. Introduction

The anatomical and functional complexity of brain networks endows humans with incredible speech abilities, while renders itself arduous to be illuminated. Electrophysiological studies on speech processing using electro- and magnetoencephalography (EEG/MEG) have revealed the sequential activation of brain networks with high temporal precision [1]. However, the nature of the functional interactions that enable efficient information flow across networks has yet to be elucidated [2]. The cortical localization of different functional networks involved in speech processing could find abundant evidence from studies using functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) with high spatial resolution [3], [4]. Commonly accepted functional regions include the visual-related occipital cortex, the auditory-related superior temporal cortex, and sensorimotor regions around the central sulcus. Besides, higher cognitive activities were found in the prefrontal lobe (PFL) for executive and control functions, the parietal lobe (PL) for multisensory integration, and the anterior temporal lobe (ATL) for feature combination [1].

On the other hand, recent findings from the perspective of neural oscillation provided insightful functional relevance of rhythmic synchronization [5]. Frequently explored rhythmic oscillations include theta (4–7 Hz), alpha (8–12 Hz), beta (13–30 Hz) and gamma (30-50 Hz) waves [6]. By the nature of neural oscillations [7], there exists an inverse relationship between the scale of integration and the frequency of interaction. Specifically, higher frequency oscillations (e.g. gamma) were quite limited in long-term synchronization due to offset though, they are quite suitable for localized fine-structure (e.g. visual details and phonetic segments) processing within a short time span. Comparatively, lower frequency resonances are more suitable for long-range synchronization and at larger time scales [8]. Theta modulation, for example, closely correlates with the acoustic speech envelope at the syllabic rate, thus suitable for word processing. Because of the collective functions of the theta and gamma oscillations in packaging phonetic segments into word units with appropriate temporal granularity [9], [10], theta-gamma coupling was currently thought to be a foundational mechanism in speech and language processing.

Another prominent and also earliest finding on ‘idling rhythm of alpha was now thought not only for restoring state in the absence of sensory inputs but also serve as a mechanism for attentional mediation by selective suppression of distracting information [11] and disengagement of task-irrelevant brain areas [12], [13]. Less controversial was the beta band suppression that has been conventionally attributed to sensorimotor functions [14].

Although all these findings seem highly inspiring, a major confusion needs to be addressed. Despite the functional relevance of rhythmic oscillation is validly proven, it is unlikely that a cognitive function could be associated in a unique and direct way with a single frequency band. For various spectrally different functional networks to cooperate, a bridging mechanism at least needs to be in place. How such a mediation mechanism works calls for clarification. This study aims to address these issues in the following steps: (i) Adopt a sentence oral reading task and use a multi-modal data acquisition system to record the brain waves, eye movements, and speech signals simultaneously; (ii) segment out cognitive stages with the index from the eye-movement trajectory and speech segmentation; (iii) transform scalp EEG signals to cortical locations and cluster into local networks; (iv) examine the oscillatory patterns of the local networks based on event-related spectrum perturbation; and (v) constructing frequency-specific effective connectivity for the analysis of rhythmic interaction and their functional relevance.

2. Experiments and Methods

2.1. Subjects and stimuli

Twenty Mandarin speakers (age: 22.2 years, SD = 3.4) were recruited in this study. All the subjects are right-handed [15], and reported with normal or corrected-to-normal vision, normal hearing and speaking abilities, and no psychological disorders. Ethical approval for this experiment was obtained from the Local Research Ethics Committee.
The text material used in this experiment consists of 60 structurally-consistent sentences. Each sentence is composed of 16 characters/syllables, and every two characters construct a semantic word (8 two-character words).

2.2. Equipment and Data Acquisition

The oral reading task was conducted in a soundproof, electro-magnetically shielded room. For each experimental trial, one sentence was presented on a screen with all 16 characters evenly spaced (with no slash as exemplified above, thus there is no obvious cue for subjects to mark out the two-character word units). Each of the 60 sentences was repeated 3 times (total 180 trials, separated into three sessions) and presented randomly. The subjects were asked to scan the sentence and read it aloud at a natural rate.

EEG signals were recorded from the scalp of each subject with a 128-channel Quik-Cap (Neuroscan, USA) placed in accordance with the extended 10-5 system [16]. The sampling rate was 1000 Hz, and the electrode impedance was maintained below 5 k throughout the acquisition. The eye movements of the subjects were recorded via a monocular pupil tracking system (Eyelink 1000, SR Research Ltd., Mississauga, Canada) at the sampling rate of 100 Hz. To avoid large movement, a headrest was placed for subjects to fix their forehead. Before the start of each session, a three-point calibration was conducted for the accuracy of eye-tracking (Gaze accuracy deviation <0.50º). Speech signals were recorded simultaneously using an electret-condenser microphone (SONY ECM MS957) at 44100 Hz. Usually, the production of the whole sentence could be finished within 4500 ms. Details of the experimental setup are described elsewhere [17].

2.3. Behavioral analyses

Behavioral analyses were conducted on the eye movement and speech signals by segmenting the ocular motion onsets and offsets, separated into three sessions) and presented randomly. The averaged time latency (ms) and standard deviations were listed in Table 1.

Table 1: Averaged time latencies (ms) and deviations of the eye and speech onsets and offsets for the 8 words.

<table>
<thead>
<tr>
<th></th>
<th>eye onset</th>
<th>eye offset</th>
<th>speech onset</th>
<th>speech offset</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>361 (125)</td>
<td>771 (142)</td>
<td>999 (154)</td>
<td>1428 (242)</td>
</tr>
<tr>
<td>2</td>
<td>796 (141)</td>
<td>1158 (164)</td>
<td>1409 (222)</td>
<td>1801 (284)</td>
</tr>
<tr>
<td>3</td>
<td>1183 (164)</td>
<td>1619 (248)</td>
<td>1775 (273)</td>
<td>2225 (328)</td>
</tr>
<tr>
<td>4</td>
<td>1644 (208)</td>
<td>2062 (275)</td>
<td>2188 (318)</td>
<td>2670 (375)</td>
</tr>
<tr>
<td>5</td>
<td>2087 (255)</td>
<td>2502 (326)</td>
<td>2625 (355)</td>
<td>3108 (416)</td>
</tr>
<tr>
<td>6</td>
<td>2527 (296)</td>
<td>2977 (362)</td>
<td>3055 (396)</td>
<td>3534 (468)</td>
</tr>
<tr>
<td>7</td>
<td>3002 (306)</td>
<td>3477 (397)</td>
<td>3486 (408)</td>
<td>3967 (516)</td>
</tr>
<tr>
<td>8</td>
<td>3502 (367)</td>
<td>3752 (443)</td>
<td>3917 (476)</td>
<td>4414 (573)</td>
</tr>
</tbody>
</table>

2.4. EEG processing

2.4.1. Artifact reduction and source reconstruction

Preprocessing of raw EEG data was firstly performed individually as follows: (1) down-sample to 250 Hz; (2) filter at a bandwidth of 1-50 Hz; (3) remove bad channels with over 10% of non-stationary high variance signals; (4) re-reference to average leading; and (5) cut trials into epochs ranging from -1000 ms to 5000 ms around each sentence presentation onset, with the 1000-ms pre-onset period as a baseline for correction. After preprocessing, independent component analysis (ICA) [18], [19] was performed on the concatenated epochs to separate brain sources from biological artifacts that arise from currents in the heart, eyes, and muscles. Then, the equivalent current dipole (ECD) model for the effective components was computed using a boundary element model (BEM) [20] which helps identify the dipole locations and cluster them based on their 3D coordinates and spectral properties.

2.4.2. Event-related spectral perturbation

To investigate the full-spectrum event-related brain dynamics in response to the oral reading task, an event-related spectral perturbation (ERSP) analysis [21] was performed. The average dynamic changes in amplitude of the broadband EEG frequency spectra was measured as a function of time. Firstly, baseline spectra were calculated from the baseline period as selected earlier. The post-stimulus epoch was then undertaken by moving the average of the amplitude spectra of overlapping windows. Each of these spectral transforms of the individual epoch was normalized by dividing their respective mean baseline spectrum and averaged to produce the final results.

2.4.3. Effective connectivity analysis

The effective connectivity analysis is capable of indicating the information flow within brain networks based on the estimated causal influences among each pair of network components [22]. Granger-Causality (GC)-based methods in this field have been increasingly exploited. It is based on the minimization of prediction errors of autoregressive models to fit the component waveforms. The estimator implemented in this study was the direct directed transfer function (dDTF), which is the product of a full-frequency directed transfer function (which allows a more interpretable comparison of information flow at different frequencies) and the partial coherence (a conditional coherence that could not be explained by a linear coherence between components) [23]. A segmentation-based adaptive multivariate auto-regressive model was constructed with a 500-ms sliding window and a 25-ms step size. The model order of 10 was selected based on the Vieira-Morf lattice algorithm [23]. After checking the stability and residual whiteness, statistical analysis was implemented using bootstrap resampling and phase randomization methods, and finally significant connections (p < 0.05) were visualized in Figure 2 with the BrainNet viewer [24].

3. Data Analyses and Results

3.1. Frequency oscillation of local regions

3.1.1. ICA clustering and localization

From ICA decomposition and clustering, 13 spatially constrained local regions were identified. Figure 1 plotted their spatial localization on fMRI templates, followed by their ERSP results as the relative spectral log amplitude on a time-by-frequency plane. The scales of the time range, frequency, and spectral perturbation are illustrated at the bottom plot. From bottom up, the plots were arranged in line with the oral reading task flow, namely from visual input (sensory) to comprehension (ventral), cognition, and finally articulation (dorsal). The notation of “ventral” and “dorsal” conforms to the well-known dual stream model for speech processing [25], where speech comprehension was thought to be conducted ventrally along the
temporal lobe while articulation was conducted dorsally along the frontal lobe.

3.1.2. regional even-related spectral perturbation

The plot-by-plot description for the ERSP results was given below. The bottom “Occ” plot stands for the occipital cortex where visual processing is initiated. In the ERSP result of Occ, theta synchronization (red area in the range), alpha desynchronization (blue area in the range), and gamma synchronization (red area in the range) were the most significant oscillatory profiles. On the “ventral” panel, the “lTP” plot depicted the neural oscillation in the left temporoparietal junction, where visual word form was thought to be transferred to phonological form or vice versa [26]. We could find similar spectral patterns of the “lTP” plot as in the “Occ” plot, which suggested a resonance mechanism for a flow of information. The “rTP” plot in the right hemisphere stood for a location related to prosody, especially in Mandarin [27]. Different from the left counterpart, this area recruited stronger theta-gamma synchronization more than alpha desynchronization. The left anterior temporal lobe (lATL), the semantic convergence center, also appeared with stronger theta-gamma synchronization. Note the clear segmented blocks in the theta band of lATL, which closely aligned with the 8-word sequences. This strongly suggested the involvement of semantic constraints from higher-order cognitive functions. The right superior temporal gyrus (rSTG) where auditory information was processed showed gamma synchronization after the articulatory onset, which may function as the auditory feedback. The “cognitive” panel included a number of frontal and parietal components that are thought to be a domain- and process-general [28], [29], [30]. Among them, the left prefrontal (lPFL) activity activated broadband theta-gamma synchronization and alpha-beta desynchronization, while little activation was found in the right prefrontal (rPFL). This might be explained by the functional lateralization of lPFL in general cognitive control and syntactic processing [31]. In contrast, rPFL is more specific for processing complex words, which is not the case for our word materials [32]. In the parietal lobe (PL), two oscillatory patterns were found. The two inner plots for the left and right parietal lobe (lPLi and rPLi) evolved with parallel desynchronization profiles in the alpha and beta bands. The two outer ones (lPLo and rPLo) were more dominant with theta synchronization and alpha desynchronization. In the “dorsal” panel, the two sensorimotor regions (lSM and rSM) are responsible for motor planning and execution. They showed a decrease in alpha and beta perturbation, which is in line with many studies on the motor response, also known as the mu rhythm [33], [34]. To sum up, high-frequency gamma oscillation mainly occurs in the sensory input areas (e.g. the visual and auditory cortices) and the general cognitive region (e.g. the left prefrontal, parietal, and anterior temporal lobe). Beta oscillation was particularly significant in the sensory output region (e.g. the sensorimotor and multisensory integration regions). Alpha suppression could be found in almost all the involved regions except for the bilateral auditory areas. Theta oscillation often occurs in parallel with gamma oscillation, dominating the higher cognitive regions and sensory regions. As can be seen, from low-level sensory input and output to higher-order multisensory integration and cognitive control, there is an oscillatory hierarchy configuring in the reverse order from high frequency to low frequency. This hierarchy also corresponds to the inverse relation between oscillatory rhythms and network scales, where the local region recruits high frequency for processing fine structure and low frequency carries a long-range connection for large-scale network association.
3.1.3. Frequency-specific effective connectivity analysis

To further verify the frequency hierarchy hypothesis, Figure 2 visualized the time- and frequency-dependent interactive information flow between localized EEG sources. For illustration, one period of processing for the first word from eye-onset to speech offset was depicted. At the eye onset, visual activation in the occipital lobe was firstly observed in the gamma band. Meanwhile, in the theta band, the prefrontal and parietal areas were connected, suggesting an early higher cognitive regulation. Alpha oscillation served as a mediation, bridging together the prefrontal lobe, posterior occipital lobe, and temporal lobe. During the second time period, gamma activity in the visual area moved ventrally to the anterior temporal lobe and prefrontal lobe. These two areas are considered to be the semantic association area and syntactic integration center, respectively. Meanwhile, in the theta and alpha bands, the interregional connection between the visual sensory area, the parietal multisensory integration center, and the prefrontal control area also became strengthened. This formed a dorsal path from bottom-up. The beta band activities didn’t show active participation until the speech onset. As we can find in the third time window, beta oscillation occurred as a clear executive chain, linking the prefrontal executive center with Broca’s area for phonetic program, then to the premotor and primary motor areas for articulation. During this period, network connections carried by theta and alpha band oscillation focused more specifically to the prefrontal, parietal, and motor areas in support of the articulatory process. The gamma-band activity in this time slot showed an increase in the right hemisphere where prosody is assumed to be processed. This also makes sense, especially in Mandarin, prosody as a lexical factor should be carried on each character or syllable for speech production. At the end of speech offset, the activity is attenuated, remaining some weak connections between the prefrontal and temporal lobes for auditory feedback, as well as weak connections between the prefrontal and parietal lobes in the theta band for high-level cognitive control.

4. Discussion

This study investigated the neural mechanism of rhythmic oscillation and network interaction by examining both regional multi-frequency oscillatory patterns and interregional frequency-specific connection patterns. Our results support the hypothesis of a hierarchical oscillatory system where rhythmic synchronization of neural discharges in the high gamma and beta bands may provide the necessary spatial and temporal links that bind together the processing in primary sensorimotor areas. While oscillatory rhythms in the low theta and alpha bands may synchronize distributed neural networks and build dynamic assemblies for specific perceptual and cognitive functions. However, different from traditional views that tend to attribute a single function to a specific frequency, this study revealed that each functional network is actually working on many frequency bands. And the cortico-cortical interactions are carried on different frequency oscillations depending on the specific functionality, the interactive target, and the information flow direction. Such co-existence of high and low frequencies in both local and associative areas is both anatomically and operationally plausible and is promising to explain the cooperation of the large-scale networks [8], [35].

5. Conclusions

Cognitive network interaction for sentence processing is most likely to be mediated by a hierarchical bidirectional organization over multiple frequency bands. Local network synchronization for fine-structure processing is carried in high gamma and beta frequencies whereas long-range network is mediated via low theta and alpha frequencies. These network interactions are supposed to be conducted bidirectionally, where the external sensory input and internal cognitive control function simultaneously via the coexistence of common rhythmic oscillations. The proposed bidirectional oscillatory hierarchy reflects a generic neural oscillatory mechanism that allows for dynamic routing of information in a network of task-relevant brain regions, and hold the promise to extend to other cognitive domains.

6. Acknowledgements

This study is supported in part by JSPS KAKENHI Grant (20K11883), and in part by National Natural Science Foundation of China (No.61876126).
7. References


