NeuroImage 229 (2021) 117724

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Neural representations of imagined speech revealed by frequency-tagged magnetoencephalography responses

Lingxi Lu[a,b,c], Jingwei Sheng[b,d], Zhaowei Liu[b,e], Jia-Hong Gao[a,b,f,*]

[a] PKU-IDG/McGovern Institute for Brain Research, Peking University, Beijing, 100871 China
[b] Center for MRI Research, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, 100871 China
[c] Center for the Cognitive Science of Language, Beijing Language and Culture University, Beijing, 100083 China
[d] Beijing Quanmag Healthcare, Beijing, 100195 China
[e] Center for Excellence in Brain Science and Intelligence Technology (Institute of Neuroscience), Chinese Academy of Science, Shanghai, 200031 China
[f] Beijing City Key Lab for Medical Physics and Engineering, Institute of Heavy Ion Physics, School of Physics, Peking University, Beijing, 100871, China

ARTICLE INFO

Keywords:
Frequency tagging
MEG
Speech mental imagery
Speech perception

ABSTRACT

Speech mental imagery is a quasi-perceptual experience that occurs in the absence of real speech stimulation. How imagined speech with higher-order structures such as words, phrases and sentences is rapidly organized and internally constructed remains elusive. To address this issue, subjects were tasked with imagining and perceiving poems along with a sequence of reference sounds with a presentation rate of 4 Hz while magnetoencephalography (MEG) recording was conducted. Giving that a sentence in a traditional Chinese poem is five syllables, a sentential rhythm was generated at a distinctive frequency of 0.8 Hz. Using the frequency tagging we concurrently tracked the neural processing timescale to the top-down generation of rhythmic constructs embedded in speech mental imagery and the bottom-up sensory-driven activity that were precisely tagged at the sentence-level rate of 0.8 Hz and a stimulus-level rate of 4 Hz, respectively. We found similar neural responses induced by the internal construction of sentences from syllables with both imagined and perceived poems and further revealed shared and distinct cohorts of cortical areas corresponding to the sentence-level rhythm in imagery and perception. This study supports the view of a common mechanism between imagery and perception by illustrating the neural representations of higher-order rhythmic structures embedded in imagined and perceived speech.

1. Introduction

Mental imagery is a quasi-perceptual experience that can be internally represented in the absence of external stimulation (Kosslyn et al., 2001). The subjective experience of speech mental imagery is ubiquitous in humans, such as speaking or singing in someone's mind. Previous functional magnetic resonance imaging (fMRI) studies have demonstrated that speech mental imagery involves some of the same neural machinery as auditory perception in the temporal region (McGuire et al., 1996; Shergill et al., 2001; Aleman et al., 2005) and also recruits brain regions outside the traditional auditory cortex, such as the inferior frontal gyrus (Broca's area), which is associated with speech production (Aleman et al., 2005; Kleber et al., 2007; Papoutsi et al., 2009; Price, 2012; Rueckert et al., 1994; Tian et al., 2016) and the temporoparietal junction that is related to memory storage and retrieval (Kleber et al., 2007; Rueckert et al., 1994; Tian et al., 2016) However, the temporal resolution of fMRI is not optimal to adequately characterize the rapid neural dynamics underlying the internal construction of imagined speech.

Recently, magnetoencephalography (MEG) and electroencephalography (EEG) recordings with higher temporal resolution have been used to investigate the dynamic neural representations of imagined speech. Neurophysiological evidence consistent with the view of common mechanisms between speech imagery and perception has been obtained that shows that acoustic features of imagined speech can be reconstructed based on the computational model built from actual speech (Martin et al., 2014) and can be decoded from neural activity in the superior temporal gyrus (STG) with additional contributions from the frontal cortex and sensorimotor cortex (Martin et al. 2016). Imagined speech also causes high gamma activity changes in the superior temporal lobe and the temporoparietal junction (Pei et al., 2011) and shares a phonological processing network with overt speech (Brumberg et al., 2016). The close relation between speech imagery and perception is further demonstrated by studies that show an early-stage interaction between the top-down generation of speech mental imagery and bottom-up stimulus-driven perception by using the imagery-perception repetition paradigm (Ylinen et al., 2015; Whitford et al., 2017; Tian et al., 2018).

* Corresponding author.
E-mail address: jgao@pku.edu.cn (J.-H. Gao).

https://doi.org/10.1016/j.neuroimage.2021.117724
Received 27 July 2020; Received in revised form 25 December 2020; Accepted 3 January 2021
Available online 7 January 2021

The frequency-tagging paradigm allows for the detection of neural signals that change periodically over time. It is a useful tool to investigate the rhythm of internally constructed structures, for example, music rhythm perception (Nozaradan et al., 2011; Nozaradan et al., 2018; Lenc et al., 2018) and speech linguistic construction (Ding et al., 2016), and has recently been applied to imagery studies (Lu et al., 2019). Specifically, Lu et al. (2019) tracked the cyclical neural responses induced by a mental imagery task and discovered frequency-tagged MEG responses to the rhythmic mental operation of silent counting. However, the following considerations should be noted. (i) In real-life situations, the subjective experience of speech mental imagery is more than counting numbers in the mind. Speech processing includes multiple processing stages, from the spectrotemporal analysis of low-level acoustic features and phonetic and categorical encoding to higher-level semantic and linguistic processing (Hickok and Poeppel, 2007). How imagined speech with higher-level structures, such as imagined words, phrases and sentences, is internally unitized and rapidly represented remains elusive. (ii) Furthermore, given the close relationship between speech imagery and perception, it is still unclear whether the internal construction of imagined speech with higher-order structures shares a common neural mechanism with that of perceived speech. Filling this research gap is critical for understanding the top-down brain function involved in mental imagery construction and will contribute to developing speech neuroprosthetic devices that turn internal speech into external signals in applications of brain-machine interfaces.

In the present study, we aim to determine the neural representations of imagined speech with higher-order structures and their relationship with speech perception. Our work is inspired by a recently established metric for tagging the rhythms of phrases and sentences during speech comprehension recently established (Ding et al., 2016; Sheng et al., 2019), which has yielded valuable insights about internally constructed organization in perceived speech when the bottom-up input of continuous speech is parsed into hierarchically embedded structures over distinct timescales. Here, we selected traditional Chinese poetry (named *Jueju*), which contains four sentences (lines) of five syllables, as the experimental material to achieve our research goal. There are two reasons why traditional Chinese poems were used in our study. First, Mandarin-speaking subjects are familiar with traditional Chinese poetry and are able to rapidly generate the mental imagery of a Chinese poem based on their memory without any external cues, allowing for neural tracking of imagined speech in the absence of external stimulation. Second, the rhythm of a traditional Chinese poem renders it suitable for capturing the periodical neural responses to syllables and sentences of the poem at tagged frequencies. We propose that the neural representation of rhythmic constructs in imagined speech (i.e., syllables and sentences in an imagined poem) share a common mechanism with that in speech perception.

2. Materials and Methods

2.1. Participants

Twenty-four young participants (14 females; mean age: 22.7, standard deviation: 3.7) took part in this experiment. All participants were right-handed, with no hearing loss or mental disorders based on their self-reports. The participants gave their informed consent before the experiment and were paid a modest stipend for their participation. The Peking University Institutional Review Board approved this study.

2.2. Stimuli

We generated a pure tone stimulus with a duration of 50 ms and a frequency of 440 Hz in Adobe Audition software (CS6, Adobe Systems Inc., San Jose, California, USA). The sound was sampled at 16 kHz. A sequence of 80 pure tones was prepared as the reference sound, and the onset-to-onset interval of the pure tones was set to 250 ms. The reference sound lasted for 20 s.

The speech stimuli were three traditional Chinese poems (entitled “Sympathy for the peasants”, “A tranquil night” and “Spring morning”). Each poem contained 20 syllables with every five syllables forming a sentence (line). Here, the rhythmic constructs in a poem were defined as periodically combining syllables into sentences. In the perception condition, we synthesized the poem using the male speaker Liang of Neospeech synthesizer (http://www.neospeech.com/). The duration of each syllable was adjusted to 250 ms by padding a silence or removing the end of a syllable with a 25 ms $\cos^2$ falling ramp. An actual poem lasted for 5 s and was repeated 4 times to match the 20 s reference sound.

2.3. Procedure

In the experiment, the participants sat inside a dimly lighted two-layer magnetically shielded room (Vacuumschmelze GmbH, Germany). A projection screen was placed in front of the participants at a distance of 1 m. The sound stimuli were presented via MEG-compatible insert earphones (CareFusion, Germany) at a comfortable sound level for the listeners. There were three blocks (conditions) in the experiment (Fig. 1): (1) the imagery condition, (2) the perception condition, and (3) the control condition. Each condition contained 15 trials. In the imagery and perception conditions, one target poem was selected in a trial, and each of the three Chinese poems was selected as the target poem 5 times for a total of 15 trials. The presentation order of three conditions was arranged using a Latin square design across subjects, and the presentation order of the 15 trials in a condition was randomized.

At the beginning of each trial, the instructions were shown on the screen. In the imagery condition, the instruction read “Please imagine the following poem loudly in mind following the pure tones”, and the content of the target poem was presented on the screen. The participant pressed a response button using his/her index finger of the right hand after he/she was ready. Then, the visual content of the target poem disappeared, and a fixation point appeared at the centre of the screen. After a random interval of 1–1.5 s, a sequence of pure tones was bilaterally presented to the subjects as the reference sound, with a 250 ms onset-to-onset interval between the pure tones. Therefore, the stimulus-presentation rate was tagged at 4 Hz. In the meantime, the subjects were required to imagine the target poem in the mind four times following the reference sound, leading to the mental construction of 80 syllables following the 80 pure tones. Critically, 16 sentences were formed from the 80 syllables with every 5 syllables combined as a sentence; thus, the rhythm of sentences was tagged at 0.8 Hz. It was noteworthy that, by applying a frequency-tagging paradigm, we were able to track the rhythmic neural signals induced by the internal construction of imagined speech without requiring overt articulation in the imagery condition. In the perception condition, the instruction read “Please listen carefully to the following poem”, and the content of the target poem was also displayed on the screen. The trial structure was the same as that in the imagery condition, except that real speech stimuli of the target poems were presented along with the reference sound and the participants were asked to listen to the actual poems instead of forming speech mental imagery. The sentence-level rhythm in a perceived poem was also tagged at 0.8 Hz, and the stimulus-/syllable-level rhythm was tagged at 4 Hz. In the control condition, the instruction read “Please count freely in your mind”, and the participants counted numbers in their minds without strictly following the presentation of the pure tones until the sound sequence terminated. The free counting task in the control condition was designed to maintain participants’ attention and to control for the stimulus-level responses at 4 Hz caused by the presentation of pure tones.

Before the formal experiment, the participants were required to fluently recite the three traditional Chinese poems. Then, they received a training session to ensure they understood the procedure of the

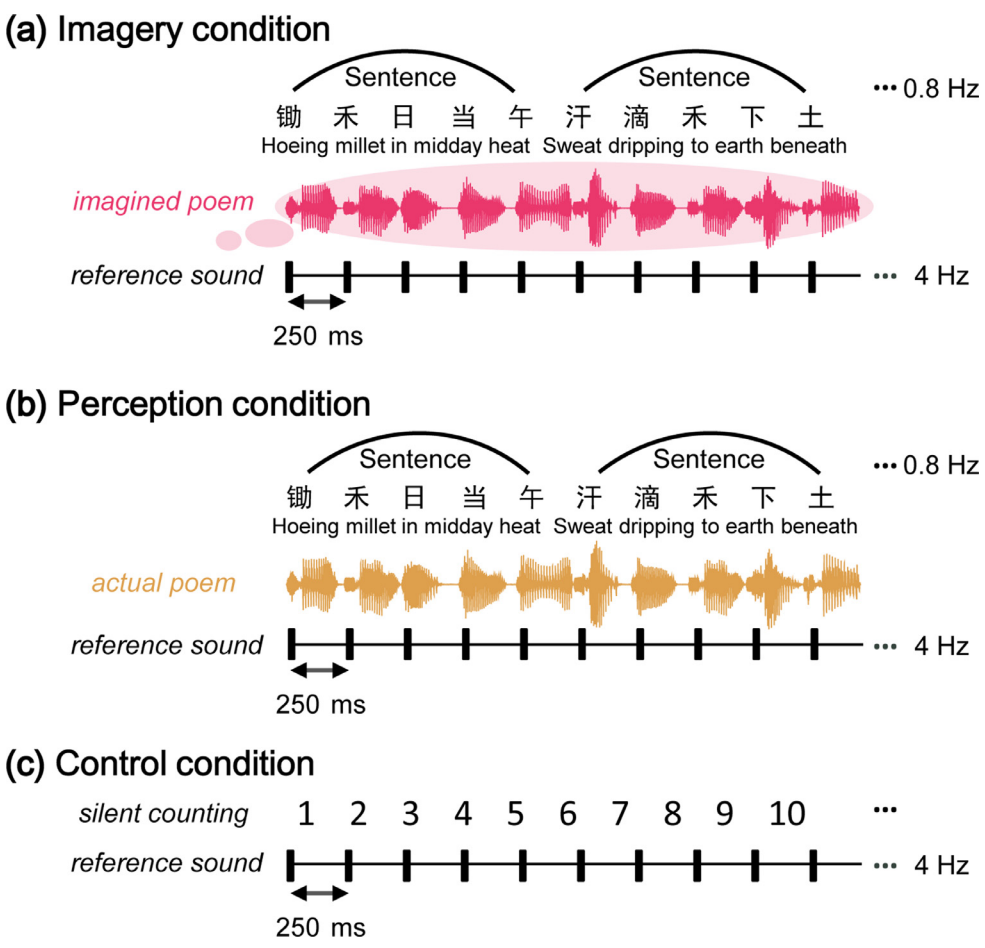

Fig. 1. Experimental materials for the three conditions. (a) In the imagery condition, mental imagery of a target Chinese poem was constructed (red) while following a sequence of reference sounds with a 250 ms onset-to-onset interval (presentation rate = 4 Hz). As every five syllables in the imagined poem formed a sentence, the rhythm of the speech mental imagery was tagged at 0.8 Hz. The English sense-for-sense translations of each sentence are presented. (b) In the perception condition, actual poem stimuli (yellow) were delivered along with the reference sounds, generating a stimulus-rate frequency of 4 Hz and a sentence-level frequency of 0.8 Hz. (c) In the control condition, the participants silently counted numbers without following the pace of the reference sounds.

experiment and could correctly perform this task. The participants were all capable of generating the imagined speech of the Chinese poems following the reference sounds according to their oral report after training.

2.4. MEG and MRI collection

The MEG data were recorded using the Neuromag TRIUX^{TM} whole-head MEG system (MEGIN Oy, Helsinki, Finland) with 306 channels at Peking University. Two electrooculogram (EOG) electrodes were applied to monitor the vertical and horizontal eye movements and blinks, which were attached inferior to the left eye and superior to the right eye. The positions of four head-position indicator coils were detected and recorded at the beginning of each block. The MEG signals were sampled at 1000 Hz and filtered online with a bandpass filter of 0.1–300 Hz.

The structural MRI data of the subjects' heads were collected on a 3T MR scanner (Discovery MR750, GE Healthcare, Wisconsin, USA). A three-dimensional fast spoiled gradient echo (FSPGR) sequence was applied to obtain T1-weighted anatomical images with $1 \times 1 \times 1$ mm^3 resolution. The MEG head position was co-registered to the structural MRI image according to three anatomical landmarks (the nasion and the bilateral preauricular points) in each subject's brain and over 150 digital points on each individual's scalp recorded using the Probe Position Identification system (Polhemus, Colchester, Vermont, USA).

2.5. Data pre-processing

We first applied the signal space separation in a spatio-temporal approach (tSSS) (Taulu and Simola, 2006) using Maxfilter to suppress the magnetic artefacts of the raw MEG data. Then, bad channels with large fluctuations or excessive noise (no more than six channels per subject) were manually identified and discarded before conducting further analyses. The MEG data were registered to the head position of the first block by MaxMove (MEGIN Oy, Helsinki, Finland). Independent component analysis (ICA) was used to measure eye movement- and eye blink-related artefacts, and then the component showing the highest correlation to the EOG signal was removed. After applying a bandpass filter of 0.2–60 Hz and a notch filter at 50 Hz, the data in the time window from −1 to 20 s relative to the onset of the stimulus were epoched, and 15 trials in each condition were averaged to obtain the event-related field (ERF) responses.

We used FreeSurfer recon-all pipeline (http://surfer.nmr.mgh.harvard.edu/) to reconstruct and segment each subject's cortex based on their T1-weighted brain images. To calculate the forward problem, we employed a realistic boundary element method model and used a 5 mm cubic grid in the source space for source estimation.

2.6. Sensor-level activity analysis

To eliminate the transient response at the beginning of the ERF response, the first 1.25 s of the 20 s ERF was excluded. Thus, the duration of ERF was modified to 18.75 s, resulting in a frequency resolution of 0.053 Hz. This frequency resolution produces sharp spectral responses and allows for good detection of cyclical neural activities at tagged frequencies. Then, a noise covariance matrix was obtained from the baseline time window of 1 s, and it was used for pre-whitening data when the magnetometer and gradiometer were combined in a single estimate (Hämäläinen et al., 2010). After data pre-whitening, the signal in the time domain was converted to the frequency domain by fast Fourier transformation (FFT) to obtain evoked powers at each channel. We calculated the averaged response power across all channels for the

estimation of the rhythmic neural activities in the frequency domain under each condition.

2.7. Source Estimation Method

A minimum L1-norm source estimation method from the recently developed VESTAL-family methods (Huang et al., 2014; Huang et al., 2016; Sheng et al., 2019; Lu et al., 2019) was applied to obtain the source level neural activities at the tagged frequencies. First, we transformed the $m \times t$ sensor waveform data matrix into the frequency domain matrix K using FFT. Here, m is the number of MEG channels and t is the time point. K includes both the real and imaginary parts. Next, we retrieved the single frequency bin K_f, in which f represents the selected frequency bin, for source estimation aiming to concentrate on the neural responses at the tagged frequency. Finally, we applied a convex second-order cone programming (SOCP) method to obtain the solution for the minimum L1-norm problem by minimizing the bias towards the coordinate axes as below (Ou et al., 2009):

$$\begin{aligned} &\min \sum_{i=1}^{n} w_i \sqrt{\left(\omega_{i,real}^{\theta}\right)^2 + \left(\omega_{i,imag}^{\theta}\right)^2 + \left(\omega_{i,real}^{\phi}\right)^2 + \left(\omega_{i,imag}^{\phi}\right)^2} \\ &s.t. \quad \mathbf{K}_f = \mathbf{G}\boldsymbol{\Omega}_f \end{aligned} \tag{1}$$

in which w represents the depth weighting vector, G represents the $m \times 2N$ lead-field matrix, i refers to the index of the source grid, and n refers to the number of source grids. θ and ϕ represent the dominant source directions extracted from the lead-field using singular value decomposition. Ω_f is the solution of the optimization problem with a dimension of $n \times 1$ for a certain frequency bin f. Note that with the SOCP method, we combined the real and imaginary parts with different principle axes derived from the same active source sites to adequately suppress noise (i.e., reduce false-positive artefacts) in the L1-norm solvers.

2.8. Statistical analysis

For sensor-level activities, in order to identify whether a significant peak occurred, the peak power at a certain frequency bin was compared with the averaged power of its two neighbouring frequency bins using the one-tailed paired t test. The null hypothesis for this test was that the power at a certain frequency bin was not higher than the average of its adjacent neighbours. We applied this test to all of the frequency bins ranging from 0.5 to 4.5 Hz in each condition, with the false discovery rate (FDR) correction for multiple comparisons. Note that the null hypothesis here is that the power of a single frequency bin is not larger than the average of its two neighbours. We assume that when the rhythm of the imagined and perceived speech is tagged at certain frequencies, the neural responses will change periodically at the corresponding frequencies, causing rising power at a single frequency bin compared to its neighbours.

Next, to compare the frequency-tagged spectral responses between conditions, the response powers were obtained by retrieving the peak power at a chosen frequency bin (i.e., 0.8 Hz and 4 Hz) relative to the average of its two neighbouring frequency bins. Repeated measures ANOVA was conducted to compare the response powers among the three conditions, followed by post hoc comparisons with Bonferroni correction between pairs of two conditions to identify the differences among the imagery, perception and control conditions.

For source-level activities, each subject's MEG responses at a single frequency bin were calculated by combining real and imaginary parts using root mean square (RMS). A corresponding control state was obtained by calculating the averaged MEG responses at the two neighbouring frequency bins. Then, each subject's MEG images were converted into Montreal Neurological Institute (MNI) space with $2 \times 2 \times 2$ mm^3 resolution using the FSL software package. The normalized image was spatially smoothed with a 5 mm full-width at half-maximum (FWHM) Gaussian kernel and log-transformed to reduce data skew (Huang et al., 2016; Lu et al., 2019) for further analysis. To measure the brain activation map at a target frequency (i.e., 0.8 Hz or 4 Hz), the neural responses at the chosen frequency bin were compared to the neighbouring control states using a one-tailed paired t test. The null hypothesis for this test was that the neural responses at the chosen frequency bin were not stronger compared to its neighbouring control states. The results were considered significant at a voxelwise threshold of $p < 0.001$ and a cluster-level familywise error (FWE) correction for multiple comparisons of $p < 0.01$. The clusters were defined as corner-connected. To make contrasts between conditions, the brain activations at a target frequency relative to their neighbouring control states were retrieved and compared between conditions using two-tailed paired t-tests (voxel-level $p < 0.001$, cluster-level FWE-corrected $p < 0.01$). The null hypothesis for this test was that the brain activations at the chosen frequency bin relative to their neighbouring control states were not different between conditions. The peak voxels of the activated brain regions were localized by finding the local maximum of the significant neural clusters, and their labels were defined based on the AAL atlas (Tzourio-Mazoyer et al., 2002). Regions of interest (ROIs) were identified by finding the local maximum of the significant neural clusters, and a 5 mm cubic was applied to retrieve the brain activation value from each ROI.

3. Results

3.1. Rhythmic neural responses induced by imagined and perceived speech

Sensor-level MEG tracked the rhythmic neural activities that were precisely tagged at the stimulus rate of 4 Hz and sentence rate of 0.8 Hz (Fig. 2). In the imagery condition, we observed significant spectral peaks at 0.8 Hz ($t_{23} = 3.79$, $p = 0.017$, Cohen's d = 0.77, FDR corrected) and its harmonics (at 1.6 Hz: $t_{23} = 3.31$, $p = 0.028$, Cohen's d = 0.68; at 3.2 Hz: $t_{23} = 3.35$, $p = 0.028$, Cohen's d = 0.68, both FDR corrected). Additionally, a robust response was found at 4 Hz, corresponding to the presentation rate of pure tones ($t_{23} = 21.92$, $p = 4.11 \times 10^{-15}$, Cohen's d = 4.48, FDR corrected). In the perception condition, we found significant spectral peaks not only at 0.8 Hz ($t_{23} = 6.11$, $p < 0.001$, Cohen's d = 1.25, FDR corrected) and its harmonics (at 1.6 Hz, 2.4 Hz and 3.2 Hz, all $p < 0.001$, FDR corrected) but also at other frequencies from 1.4 Hz to 4.4 Hz with an interleave of 0.2 Hz (all $p < 0.05$, FDR corrected). The 0.2 Hz interleaved spectral peaks were caused by the repeated presentation of the same poem in a single trial under the perception condition (see Supplementary Materials, Fig. S1). A robust spectral peak was observed at 4 Hz ($t_{23} = 20.14$, $p = 1.64 \times 10^{-14}$, Cohen's d = 4.11, FDR corrected). In the control condition, a significant peak was detected only at the stimulus rate of 4 Hz ($t_{23} = 21.26$, $p = 4.11 \times 10^{-15}$, Cohen's d = 4.34, FDR corrected). The topological distribution of the response power showed bilateral responses at both the 4 Hz stimulus level and the 0.8 Hz sentence level. In particular, right-lateralized power responses were observed at 4 Hz in both the imagery condition ($t_{23} = 2.57$, $p = 0.017$, Cohen's d = 0.53) and the control condition ($t_{23} = 3.28$, $p = 0.003$, Cohen's d = 0.67) but not in the perception condition ($t_{23} = 1.35$, $p = 0.192$, Cohen's d = 0.27). However, at the sentential rhythm of 0.8 Hz, there was no hemispheric lateralization for either imagined speech ($t_{23} = 0.46$, $p = 0.654$, Cohen's d = 0.09) or perceived speech ($t_{23} = 1.01$, $p = 0.323$, Cohen's d = 0.21).

Having captured the rhythmic MEG responses in each condition, we then compared the spectral responses among conditions using repeated measures ANOVA (Fig. 3). At the stimulus level of 4 Hz, the main effect of condition was significant ($F_{2,46} = 15.62$, $p < 0.001$, $\eta p^2 = 0.40$). Post hoc comparisons (Bonferroni corrected) revealed enhanced peak power in the perception condition compared to the control condition ($t_{23} = 5.10$, $p < 0.001$, Cohen's d = 1.04). The peak power in the imagery condition was smaller than that in the perception condition ($t_{23} = -3.77$, $p = 0.003$, Cohen's d = 0.77) and no different from that in the control condition ($t_{23} = 1.64$, $p = 0.346$, Cohen's d = 0.33). At the sentence-level of 0.8 Hz, the main effect of condition was also significant ($F_{2,46} = 7.44$, $p = 0.002$, $\eta p^2 = 0.24$). Interestingly, we found stronger peak power

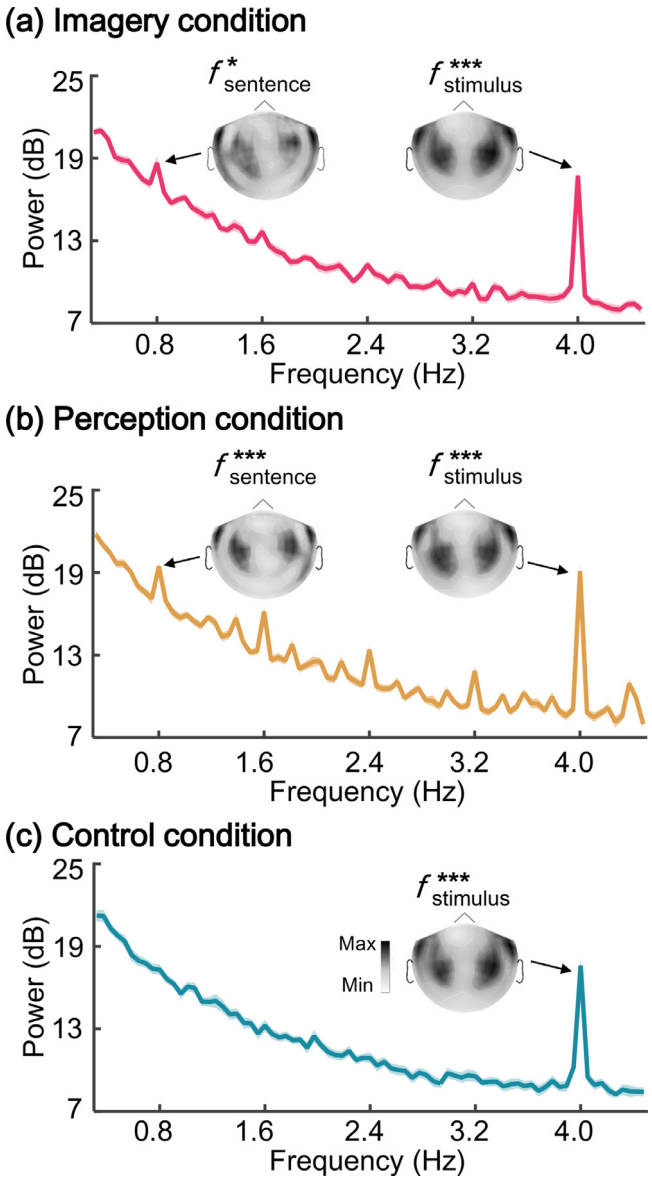

Fig. 2. Sensor-level MEG responses in neural tracking of imagined and perceived speech. Significant spectral peaks at the stimulus-level frequency of 4 Hz were observed in each condition, corresponding to the presentation rate of pure tones or syllables (paired one-sided *t*-tests, with FDR correction). Furthermore, there were significant spectral peaks at the sentence-level frequency of 0.8 Hz under both the imagery (red) and perception (yellow) conditions, reflecting the internal construction of sentences from syllables in the poems for both imagined and perceived speech. Topographic distribution was right-lateralized at 4 Hz for the imagery and control conditions in which only pure tones were presented. $^{*}p < 0.05$, $^{***}p < 0.001$.

during the speech imagination ($t_{23} = 2.63$, $p = 0.045$, Cohen's d = 054) and speech perception ($t_{23} = 4.47$, $p = 0.001$, Cohen's d = 0.91) conditions than during the control condition, and no significant difference was found between the imagery and the perception conditions ($t_{23} = 0.97$, $p = 0.99$, Cohen's d = 0.20). We also detected the neural time scales corresponding to the tagged frequencies (see Supplementary Materials, Fig. S2). Taken together, these results showed that rhythmic neural responses at the sentence-rate frequency were induced by both imagined and perceived speech, and similar stimulus-rate neural responses were elicited by the identical auditory input of pure tones in the imagery and control conditions.

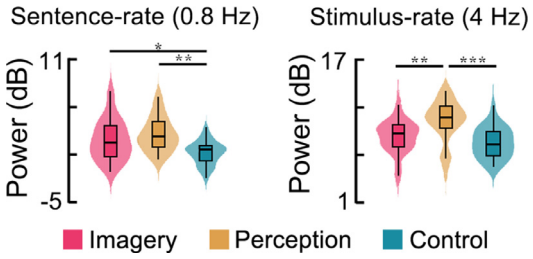

Fig. 3. Comparisons of sensor-level MEG responses among conditions. Power at 4 Hz showed no difference between the imagery (red) and control (blue) conditions when pure tones were presented and was enhanced when actual poem stimuli were played along with the pure tones in the perception condition (yellow). With identical bottom-up input, the top-down generation of poem imagery induced stronger rhythmic activity at the sentence rate of 0.8 Hz (red) compared to the control condition (blue) and was comparable to that in the perception condition (yellow). $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table 1
Major brain regions activated at the sentence rate of 0.8 Hz

Brain Region	Peak MNI Coordinate			*t* value
	x	*y*	*z*	
Imagery				
L inferior frontal gyrus	−43	21	3	5.37
R supramarginal gyrus	55	−29	23	6.68
Perception				
L Heschl's gyrus	−47	−16	9	5.13
L precuneus	−1	−72	42	5.96
L superior temporal gyrus	−51	−1	−9	6.93
R superior temporal gyrus	56	−28	10	5.53
R middle temporal gyrus	59	−16	−13	4.27
R supramarginal gyrus	49	−36	26	4.27
R inferior parietal lobule	48	−49	40	4.65

3.2. Source estimation for the brain activities at tagged frequencies

Having captured the tagged neural responses in imagined and perceived speech at the sensor level, a subsequent question arose regarding which brain network generated the tagged neural activities. We applied the L1-norm source estimation method to address this question and localized the brain regions activated at a single frequency bin in each condition. As a result, we observed two critical neural clusters involved in poem imagination (voxel-level $p < 0.001$, cluster-level FWE-corrected $p < 0.01$). Specifically, the left opercular and triangular parts of the inferior frontal gyrus (IFG) and the right supramarginal gyrus (SMG) were significantly activated at 0.8 Hz corresponding to the sentence-level rhythm in the poems. In the perception condition, when subjects listened to the poem, brain regions including the bilateral STG, the left Heschl's gyrus (HG), the left precuneus (PrC), the right SMG and the right inferior parietal lobule (IPL) were activated at 0.8 Hz (Fig. 4a, Table 1). More extensive brain regions were involved at the stimulus rate of 4 Hz, extending from the bilateral auditory cortex to distributed cortical networks (Fig. 4b, Table 2).

We identified the overlapping brain regions responsive to both imagined and perceived speech at the sentence rate of 0.8 Hz (Fig. 5a), which was localized in the right SMG. The activation of the left IFG was induced only by imagined but not perceived speech, while additional brain regions, including the bilateral temporal cortex and right inferior parietal lobe, were involved only with perceived but not imagined speech. These results implied that neural tracking of the sentence-level rhythm in imagined and perceived speech relied on brain networks with shared and distinct cohorts of cortical areas.

To further distinguish the rhythmic responses induced by imagined and perceived speech, we compared, in additional analyses, the source activities at 0.8 Hz and 4 Hz among conditions using whole-brain paired

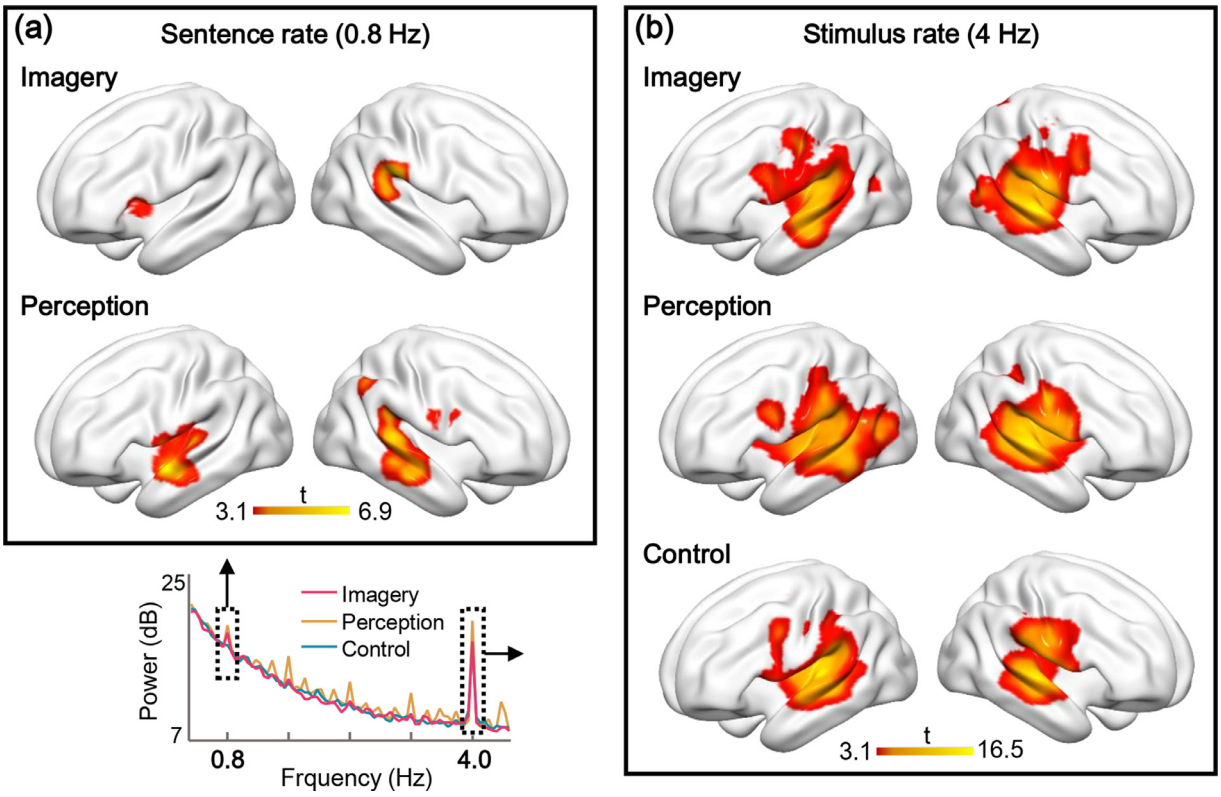

Fig. 4. Source estimation for the spectral peaks at tagged frequencies. (a) At the sentence-rate frequency of 0.8 Hz, the left IFG and right SMG were activated by sentence-level construction in the imagined speech condition, while more extensive areas including the bilateral temporal cortex, left PrG and right temporoparietal junction were involved in the perceived speech condition (voxel-level $p < 0.001$, cluster-level FWE-corrected $p < 0.01$). (b) At the stimulus-rate frequency of 4 Hz, brain regions centring at the bilateral temporal cortex and extending to the sensorimotor and parietal areas were activated in all conditions.

t-tests (Fig. 5b). This analysis attenuated the common neural activations in speech imagery and perception and highlighted the differences between them. We found stronger cortical activation induced by perceived speech than imagined speech in the left precentral gyrus (PrG) ([−42, −10, 39], t = 4.63) at 0.8 Hz and in the left middle frontal gyrus (MFG) ([−25, 51, −16], t = −3.64) and left STG ([−43, −22, −2], t = −4.48) at 4 Hz. This analysis suggested that these areas play an important role in processing the bottom-up input of actual speech but not the top-down organization of imagined speech. Moreover, the contrast between the neural activities at 4 Hz between the imagery and control conditions revealed significant differences in the right SPL ([14, −78, 53], t = 5.21), indicating imagery-induced parietal deactivation in the dorsal pathway.

4. Discussion

This study tracked MEG responses to the sentence-level rhythm in imagined poems and compared the neural activities induced by sentence-level chunking between speech imagery and perception. Similar neural tracking of the sentence-rate rhythm in imagined and perceived speech was captured, which relied on partially overlapping brain networks in the temporoparietal junction. Our data provide the first evidence on the neural dynamics underlying the high-order rhythmic construction of sentences in speech mental imagery.

In our study, the rhythmic constructs in imagined speech were defined as periodically combining syllables into sentences in an imagined poem, which enabled the neural tracking of spectral responses at tagged frequencies corresponding to the presentation rate of syllables and sentences. It should be noted that the sentence-level rhythm in a poem is not only caused by syntactic- or semantic-based chunking (Ding et al. 2016) but also related to a rhyme scheme that structures a poem periodically in time (Obermeier et al., 2013) with phonologically matching vowels of the last word in the sentences causing recursive patterning in poetry. The rhyme scheme also induces prosodic expectations that modulate early phonological processing in speech recognition (Chen et al. 2016) and cause predictive speech segmentation (Teng et al. 2020). Therefore, the sentence-level rhythmic neural activity captured in this study was, in essence, induced by multiple components based on phonological, syntactical and semantic details of the content in a poem. The core finding of the present study is that we are able to track sentence-level rhythmic neural activity in imagined speech and further localize the neural clusters responsive for such internal speech construction. It will be important to determine the contributions of the phonological, syntactical or semantic components in forming internal speech in future studies.

Previous studies have shown that speech comprehension requires neural integration of hierarchal linguistic structures of different sizes, such as syllables, words and sentences (Ding et al., 2016; Sheng et al., 2019). Such internal structure-building operations were also observed in our study, in which the neural activities of the embedded structures were tagged at the sentence-level and syllable-/stimulus-level rhythms while perceiving real poems. More importantly, we found similar rhythmic constructs at the sentence-level rate during the construction of the imagined poem when there was no bottom-up input of speech stimuli, implying the existence of a common mechanism underlying speech imagery and perception. Recent work has revealed that the low-level acoustic features of imagined speech can be reconstructed from a computational model built from real speech (Martin et al., 2014), and the processing of low-level perceptual attributes of imagined speech, such as loudness, interacts with auditory perception at an early processing stage in the auditory cortex (Tian et al., 2018). Our study thus extends the view of a common mechanism underlying speech imagery and perception from the lower-level acoustic representations to higher-level language processing, demonstrating that the formation of imagined speech requires

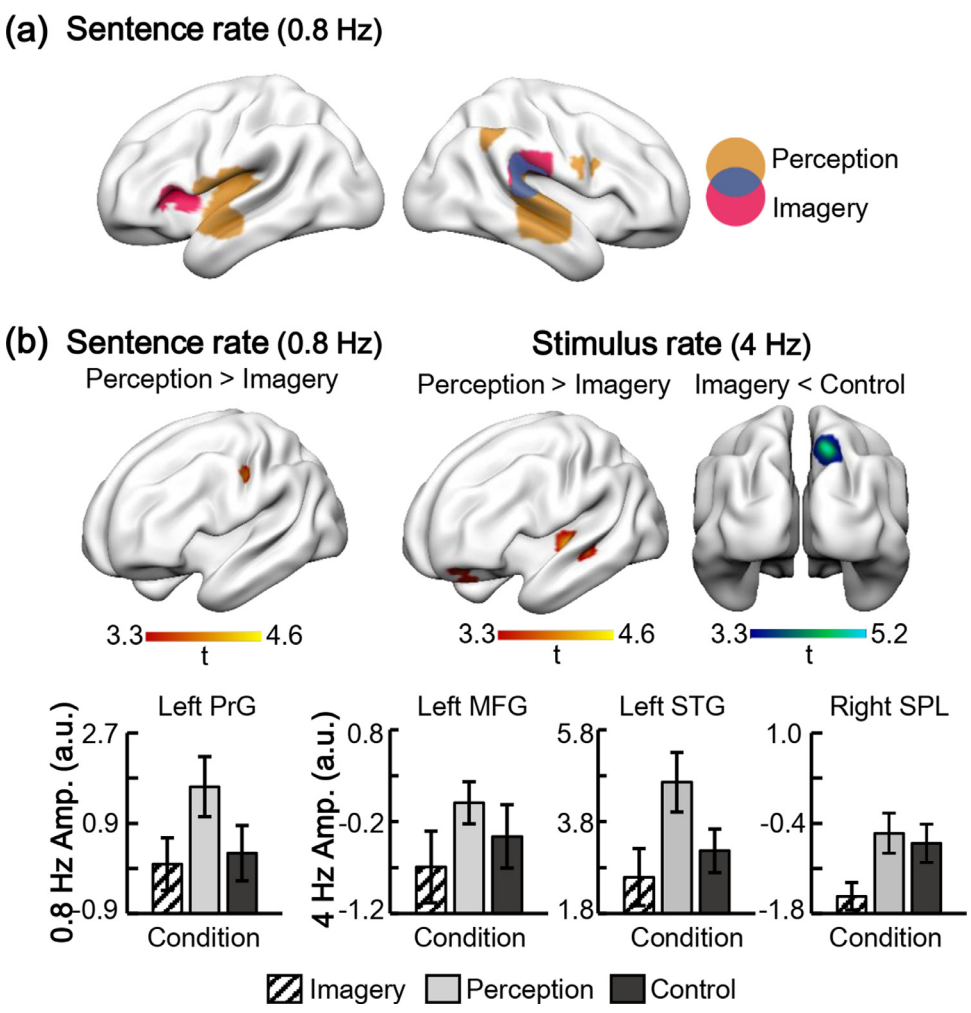

Fig. 5. Comparisons of source neural activities among conditions. (a) Imagined and perceived speech shared a common region in the right SMG that was activated at the sentential rhythm, whereas the left IFG was activated by only imagined but not perceived speech. (b) Whole-brain paired *t*-tests among conditions (voxel-level $p < 0.001$, cluster-level FWE-corrected $p < 0.01$) revealed stronger activation induced by perceived speech than imagined speech in the left PrG at the sentential rhythm and in the left MFG and STG at the stimulus rhythm. The right IPL showed deactivation in the imagined speech condition compared to the control condition. ROI activation of the significant neural clusters is displayed in the lower panel. The error bars represent the standard error (ER).

Table 2
Major brain regions activated at the stimulus rate of 4 Hz

Brain Region	Peak MNI Coordinate			*t* value
	x	*y*	*z*	
Imagery				
L inferior frontal gyrus	−58	18	4	3.65
R middle frontal gyrus	37	6	40	4.75
R Heschl's gyrus	58	−10	6	12.80
R superior parietal gyrus	22	−52	62	3.53
L postcentral gyrus	−44	−13	42	4.97
L precentral gyrus	−40	−5	35	4.13
R precentral gyrus	37	−14	48	3.62
L precuneus	−8	−61	38	4.00
R precuneus	2	−54	43	4.38
L Rolandic operculum	−47	−22	13	12.35
L middle temporal gyrus	−58	−14	−8	9.50
R middle temporal gyrus	52	−16	8	4.68
Perception				
L inferior frontal gyrus	−38	14	14	4.50
L middle occipital gyrus	−42	−72	12	5.10
L postcentral gyrus	−52	−16	24	8.27
R precuneus	7	−41	40	4.66
R Rolandic operculum	47	−13	22	15.29
L middle temporal gyrus	−52	−29	4	7.76
L superior temporal gyrus	−46	−16	1	8.04
Control				
L inferior frontal gyrus	−47	13	6	3.90
L precentral gyrus	−40	2	35	16.46
R Rolandic operculum	52	−12	9	11.04
L superior temporal gyrus	−48	−23	7	16.46
R superior temporal gyrus	57	−12	6	11.34

the mental organization of embedded structures (i.e., combining syllables into sentences) similar to that required for perceived speech. In other words, grouping words into multiple-word chunks in language processing requires sharing of brain function during both speech perception and speech mental imagery.

In our study, the left inferior frontal cortex, which has been reported to be involved in the silent articulation of speech (Tian et al., 2016; Papoutsi et al., 2009; Aleman et al., 2005; Price, 2012; Rueckert et al., 1994; Kleber et al., 2007), and the right temporoparietal junction, which has been closely associated with memory storage and retrieval (Kleber et al., 2007; Tian et al., 2016) as well as auditory working memory (Paulesu et al., 1993), were found to be activated at the sentence-level rhythm during poem imagination, which might be explained by participants' preparation for generating phonological sequences as they were trying to produce poems in their minds. This finding agreed with the previous study by Lu et al. (2019) showing that the left IFG and right SMG contributed to organizing numbers into mental groups during a rhythmic inner counting task, suggesting that these brain regions play crucial roles in the top-down induction mechanism to build imagined speech. Note that in our study we did not ask participants to strictly distinguish between hearing imagery and articulatory imagery in the poem imagination task. The finding of neural activation in the left IFG, which is a crucial region for covert articulation planning (Tian et al., 2016; Price, 2012), suggests that the articulatory preparation for inner speech production is probably recruited in the present imagination task. Considering the challenge to track an individual's inner subjective representations in time without external signals to identify imagery-related neural activities, the frequency-tagging paradigm will be a promising

tool to investigate the neural dynamics underlying silent articulation or covert speech production without overt articulatory movements. In future studies, whether the neural representations of imagined speech uncovered in our paradigm can be extended to a richer account of inner speech should be considered. Moreover, we found that the right SMG was activated at the sentence-level rate not only in the imagery condition but also in the perception condition. This result demonstrated that the common mechanism of building higher-order rhythmic constructs in imagined and perceived speech might be rooted in the right temporoparietal junction, which is complementary to previous findings that the temporoparietal junction is the nexus area of multiple higher-order brain functions, including the memory and language processing streams (Carter and Huettel, 2013).

In addition, the STG and MFG showed stronger activation levels at the stimulus rate during speech perception than during speech imagery. These results can be explained from the perspective of the dual stream prediction model (Tian and Poeppel, 2012, 2013; Tian et al., 2016) that perceiving real speech stimuli largely required bottom-up auditory representation in the superior temporal cortex and top-down memory networks in the frontal cortex. In addition, we also found high levels of activation in the motor cortex at the sentence rate during speech perception, which is consistent with the common observations of motor cortex activation in speech comprehension (Hickok and Poeppel, 2007; Si et al., 2017; Morillon & Baillet, 2017).

The non-linear minimum L1-norm source estimation method from the VESTAL-family methods (Huang et al., 2014; Huang et al., 2016; Sheng et al., 2019; Lu et al., 2019) was applied in the present study to obtain focal source activity based on the assumption that a small number of focal brain regions were involved in the stimulus-/sentence-rate rhythm. Compared with linear methods (e.g., L2-minimum-norm-type methods and linearly constrained beamformers) that produce smoother source distributions and are less affected by noise (Bertero et al., 1988; Menke, 1989; Hauk et al., 2019), the classic L1-minimum-norm-type method is less robust to noise and suffers more from instability in spatial reconstruction (Fuchs et al., 1999; Uutela et al., 1999), and the focality might be reduced when applying spatial smoothing before group analysis. Although advanced L1-penalized models such as the mixed-norm estimation (MxNE) methods (Gramfort et al., 2013; Strohmeier et al., 2016) and VESTAL-family methods (Huang et al., 2014; Huang et al., 2016) have addressed some limitations of the L1-norm-based source models, care should be taken when interpreting the brain source result, which relies on the assumption of the underlying source distribution. Verification using direct recording (e.g., source localization confirmed by intracranial EEG, Lu et al., 2019) and combination of MEG-EEG to improve spatial resolution (Hauk et al., 2019; Sharon et al., 2007) should be considered in future studies.

There are some limitations in the present study that need to be addressed. In the frequency-tagged MEG responses, we found similar spectral peaks at a sentence rate of 0.8 Hz and a stimulus rate at 4 Hz (Fig. 2, Fig. 3). Thus, we attempted to analyse the MEG sensor-level data in the temporal domain with corresponding time windows of 1.25 s (0.8 Hz) and 0.25 s (4 Hz) to examine the patterns of change of neural time scales that might underlie the frequency-tagged activities. We found statistically significant patterns of change of ERF amplitudes in speech imagery and perception at a sensor in the left frontotemporal region (Fig. S2) that is close to activated neural cluster at 0.8 Hz in the imagery condition according to MEG source estimation (Fig. 4). However, the decreasing trends every 1.25 s are observed only at this representative sensor with the strongest power response at 0.8 Hz and not at the other sensors with relatively weaker power responses at 0.8 Hz. Therefore, care should be taken when interpreting the sensor-level activities in the temporal domain. In the future, the temporal dynamics related to the frequency-tagged spectral responses during the generation of imagined speech should be examined. Moreover, visual prompts might be applied as the reference stimuli in future studies to further eliminate the influence of tone-related brain responses on the speech-/imagery-induced neural activations. Additionally, in future imagery research, it will be important to monitor articulatory movements (e.g., using an electromyogram sensor) to exclude contamination to the MEG data caused by muscle movements.

5. Conclusions

In summary, we concurrently tracked the neural activities induced by the generation of rhythmic constructs embedded in speech imagery and the stimulus-driven processing that were precisely frequency-tagged. We observed similar neural tracking of imagined and perceived speech at the sentential rhythm and further localized the overlapping and distinct neural cohorts responsive to the rhythmic constructs in imagined and perceived speech. We found that the left IFG and right SMG were activated at the sentential rhythm during the construction of speech mental imagery, with a shared activation of the temporoparietal junction in the perception condition. Our findings support the view of a common mechanism between imagery and perception by illustrating the neural representations of higher-order rhythmic structures embedded in speech mental imagery.

Declaration of Competing Interest

The authors have no financial or competing interests to declare.

Credit authorship contribution statement

Lingxi Lu: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing - original draft, Writing - review & editing. **Jingwei Sheng:** Methodology, Software, Formal analysis, Writing - review & editing. **Zhaowei Liu:** Software, Writing - review & editing. **Jia-Hong Gao:** Conceptualization, Project administration, Resources, Funding acquisition, Methodology, Writing - original draft, Writing - review & editing, Supervision.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (81790650, 81790651, 81727808, 31421003 and 81627901), the National Key Research and Development Program of China (2017YFC0108900), the Beijing Brain Initiative of Beijing Municipal Science & Technology Commission (Z181100001518003), the Guangdong Key Basic Research Grant (2018B030332001) and Guangdong Pearl River Talents Plan (2016ZT06S220). We thank the National Center for Protein Sciences at Peking University in Beijing, China, for assistance with data collection.

Data and code availability

Data supporting the findings in this study and the computer codes used for the data analyses are available from the corresponding author upon written request.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2021.117724.

References

Aleman, A., Formisano, E., Koppenhagen, H., Hagoort, P., de Haan, E.H.F., Kahn, R.S., 2005. The functional neuroanatomy of metrical stress evaluation of perceived and imagined spoken words. Cereb Cortex 15, 221–228.

Bertero, M., Mol, C.D., Pike, E.R., 1988. Linear inverse problems with discrete data: II. stability and regularisation. Inverse Problems 4 (3), 573–594.

Brumberg, J.S., Krusienski, D.J., Chakrabarti, S., Gunduz, A., Brunner, P., Ritaccio, A.L., Schalk, G., 2016. Spatio-temporal progression of cortical activity related to continuous overt and covert speech production in a reading task. PLOS ONE 11, e0166872.

Carter, R.M., Huettel, S.A., 2013. A nexus model of the temporal–parietal junction. Trends Cogn. Sci. 17, 328–336.
Chen, Q., Zhang, J., Xu, X., Scheepers, C., Yang, Y., Tanenhaus, M.K., 2016. Prosodic expectations in silent reading: ERP evidence from rhyme scheme and semantic congruence in classic Chinese poems. Cognition 154, 11–21.
Ding, N., Melloni, L., Zhang, H., Tian, X., Poeppel, D., 2016. Cortical tracking of hierarchical linguistic structures in connected speech. Nat Neurosci. 19, 158–164.
Fuchs, M., Wagner, M., Kohler, T., Wischmann, H.A., 1999. Linear and nonlinear current density reconstructions. J. Clin. Neurophysiol. 16 (3), 267–295.
Gramfort, A., Strohmeier, D., Haueisen, J., Hamalainen, M.S., Kowalski, M., 2013. Time-frequency mixed-norm estimates: Sparse M/EEG imaging with non-stationary source activations. Neuroimage 70, 410–422.
Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. Nat Rev Neurosci. 8, 393–402.
Hämäläinen, M.S., Lin, F.-H., Mosher, J.C., 2010. Anatomically and functionally constrained minimum-norm estimates. In: Hansen, P., Kringelbach, M., Salmelin, R. (Eds.), MEG: an introduction to methods, pp. 186–215.
Hauk, O., Stenroos, M., Treder, M., 2019. EEG/MEG source estimation and spatial filtering: the linear toolkit. Magnetoencephalography. Springer Nature Switzerland AG, pp. 167–203.
Huang, C.W., Huang, M.X., Ji, Z., Swan, A.R., Angeles, A.M., Song, T., Huang, J.W., Lee, R.R., 2016. High-resolution MEG source imaging approach to accurately localize Broca's area in patients with brain tumor or epilepsy. Clin. Neurophysiol.: Off. J. Int. Feder. Clin. Neurophysiol. 127, 2308–2316.
Huang, M.-X., Huang, C.W., Robb, A., Angeles, A., Nichols, S.L., Baker, D.G., … Lee, R.R., 2014. MEG source imaging method using fast L1 minimum-norm and its applications to signals with brain noise and human resting-state source amplitude images. Neuroimage 84, 585–604. doi:10.1016/j.neuroimage.2013.09.022.
Kleber, B., Birbaumer, N., Veit, R., Trevorrow, T., Lotze, M., 2007. Overt and imagined singing of an Italian aria. Neuroimage 36, 889–900.
Kosslyn, S.M., Ganis, G., Thompson, W.L., 2001. Neural foundations of imagery. Nat Rev Neurosci. 2, 635–642.
Lenc, T., Keller, P.E., Varlet, M., Nozaradan, S., 2018. Neural tracking of the musical beat is enhanced by low-frequency sounds. Proceed. Natl. Acad. Sci. 115, 8221–8226.
Lu, L., Wang, Q., Sheng, J., Liu, Z., Qin, L., Li, L., Gao, J-H., 2019. Neural tracking of speech mental imagery during rhythmic inner counting. eLife 8, e48971.
Martin, S., Brunner, P., Holdgraf, C., Heinze, H.-J., Crone, N.E., Rieger, J., Schalk, G., Knight, R.T., Pasley, B.N., 2014. Decoding spectrotemporal features of overt and covert speech from the human cortex. Front. Neuroeng. 7.
Martin, S., Brunner, P., Iturrate, I., Millán, J.d.R., Schalk, G., Knight, R.T., Pasley, B.N., 2016. Word pair classification during imagined speech using direct brain recordings. Sci. Rep. 6, 25803.
McGuire, P.K., Silbersweig, D.A., Murray, R.M., David, A.S., Frackowiak, R.S.J., Frith, C.D., 1996. Functional anatomy of inner speech and auditory verbal imagery. Psychol. Med. 26, 29–38.
Menke, W., 1989. Geophysical data analysis: Discrete inverse theory. Academic Press, Inc., San Diego.
Morillon, B., Baillet, S., 2017. Motor origin of temporal predictions in auditory attention. Proceed. Natl. Acad. Sci. 114, E8913–E8921.
Nozaradan, S., Keller, P.E., Rossion, B., Mouraux, A., 2018. EEG Frequency-tagging and input–output comparison in rhythm perception. Brain Topogr. 31, 153–160.
Nozaradan, S., Peretz, I., Missal, M., Mouraux, A., 2011. Tagging the neuronal entrainment to beat and meter. J. Neurosci. 31, 10234–10240.
Obermeier, C., Menninghaus, W., von Koppenfels, M., Raettig, T., Schmidt-Kassow, M., Otterbein, S., Kotz, S., 2013. Aesthetic and emotional effects of meter and rhyme in poetry. Front. Psychol. 4.
Ou, W., Hämäläinen, M.S., Golland, P., 2009. A distributed spatio-temporal EEG/MEG inverse solver. Neuroimage 44, 932–946.
Papoutsi, M.d.Z.J.A., Jansma, J.M., Pickering, M.J., Bednar, J.A., Horwitz, B., 2009. From phonemes to articulatory codes: an fMRI study of the role of Broca's area in speech production. Cereb Cortex 19, 2156–2165.
Paulesu, E., Frith, C.D., Frackowiak, R.S.J., 1993. The neural correlates of the verbal component of working memory. Nature 362, 342–345.
Pei, X., Leuthardt, E.C., Gaona, C.M., Brunner, P., Wolpaw, J.R., Schalk, G., 2011. Spatiotemporal dynamics of electrocorticographic high gamma activity during overt and covert word repetition. Neuroimage 54, 2960–2972.
Price, C.J., 2012. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. Neuroimage 62, 816–847.
Rueckert, L., Appollonio, I., Grafman, J., Jezzard, P., Johnson, R., Le Bihan, D., Turner, R., 1994. Magnetic resonance imaging functional activation of left frontal cortex during covert word production. Journal of Neuroimaging 4 (2), 67–70. doi:10.1111/jon19944267.
Sharon, D., Hämäläinen, M.S., Tootell, R.B., Halgren, E., Belliveau, J.W., 2007. The advantage of combining MEG and EEG: comparison to fMRI in focally stimulated visual cortex. Neuroimage 36 (4), 1225–1235.
Sheng, J., Zheng, L., Lyu, B., Cen, Z., Qin, L., Tan, L.H., Huang, M.-X., Ding, N., Gao, J.-H., 2019. The cortical maps of hierarchical linguistic structures during speech perception. Cereb. Cortex (New York, NY: 1991).
Shergill, S.S., Bullmore, E.T., Brammer, M.J., Williams, S.C.R., Murray, R.M., McGuire, P.K., 2001. A functional study of auditory verbal imagery. Psychol. Med. 31, 241–253.
Si, X., Zhou, W., Hong, B., 2017. Cooperative cortical network for categorical processing of Chinese lexical tone. Proc. Natl. Acad. Sci. U S A. 114, 12303–12308.
Strohmeier, D., Bekhti, Y., Haueisen, J., Gramfort, A., 2016. The iterative reweighted mixed-norm estimate for Spatio-Temporal MEG/EEG source reconstruction. IEEE Trans. Med. Imaging 35 (10), 2218–2228.
Taulu, S., Simola, J., 2006. Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. Phys. Med. Biol. 51, 1759.
Teng, X., Ma, M., Yang, J., Blohm, S., Cai, Q., Tian, X., 2020. Constrained structure of ancient Chinese poetry facilitates speech content grouping. Curr. Biol..
Tian, X., Ding, N., Teng, X.B., Bai, F., Poeppel, D., 2018. Imagined speech influences perceived loudness of sound. Nat. Hum. Behav. 2, 225–234.
Tian, X., Poeppel, D., 2012. Mental imagery of speech: linking motor and perceptual systems through internal simulation and estimation. Front. Hum. Neurosci. 6.
Tian, X., Poeppel, D., 2013. The effect of imagination on stimulation: the functional specificity of Efference copies in speech processing. J. Cogn. Neurosci. 25, 1020–1036.
Tian, X., Zarate, J.M., Poeppel, D., 2016. Mental imagery of speech implicates two mechanisms of perceptual reactivation. Cortex 77, 1–12.
Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15, 273–289.
Uutela, K., Hamalainen, M., Somersalo, E., 1999. Visualization of magnetoencephalographic data using minimum current estimates. Neuroimage 10 (2), 173–180.
Whitford, T.J., Jack, B.N., Pearson, D., Griffiths, O., Luque, D., Harris, A.W.F., Spence, K.M., Le Pelley, M.E., 2017. Neurophysiological evidence of efference copies to inner speech. eLife 6, 23.
Ylinen, S., Nora, A., Leminen, A., Hakala, T., Huotilainen, M., Shtyrov, Y., Makela, J.P., Service, E., 2015. Two distinct auditory-motor circuits for monitoring speech production as revealed by content-specific suppression of auditory cortex. Cereb Cortex 25, 1576–1586.