Synchronization between overt speech envelope and EEG oscillations during imagined speech

Hiroki **Watanabe**[1], Hiroki **Tanaka**[2], Sakriani **Sakti**[2,3], and Satoshi **Nakamura**[2,3]

[1]Graduate School of Information Science, Nara Institute of Science and Technology, 8916-5 Takayama-cho, Ikoma-shi, Nara 630–0192, Japan

[2]Graduate School of Science and Technology, Nara Institute of Science and Technology, 8916-5 Takayama-cho, Ikoma-shi, Nara 630–0192, Japan

[3]Center for Advanced Intelligence Project AIP, RIKEN, 8916-5 Takayama-cho, Ikoma-shi, Nara 630–0192, Japan

Corresponding author: Hiroki **WATANABE** (E-mail: *watanabe.hiroki.vx6@is.naist.jp*)

Total number of pages in main text: 18

Author contributions: Hiroki Watanabe performed EEG data collection and data analysis. All authors were involved in the planning of the entire EEG experimental design, data analysis, and discussion of the results. Hiroki Watanabe wrote the manuscript. All authors reviewed the manuscript and agreed to the submission of the current manuscript.

Declaration of interests: none

Funding: Part of this work was supported by JSPS KAKENHI Grant Numbers JP16K16172, JP17H06101, JP17K00237, and JP18J14871.

Abstract

Neural oscillations synchronize with the periodicity of external stimuli such as the rhythm of the speech amplitude envelope. This synchronization induces a speech-specific, replicable neural phase pattern across trials and enables perceived speech to be classified. In this study, we hypothesized that neural oscillations during articulatory imagination of speech could also synchronize with the rhythm of speech imagery. To validate the hypothesis, after replacing the imagined speech with overt speech due to the physically unobservable nature of imagined speech, we investigated (1) whether the EEG-based regressed speech envelopes during the imagined speech correlate with the overt speech envelope and (2) whether the imagined EEG can classify speech stimuli with different envelopes imagined by several participants. The variability of the duration of the imagined speech across trials was corrected using dynamic time warping. The classification was based on the distance between a test data and a template waveform of each class. Results showed a significant correlation between the EEG-based regressed envelope and the overt speech envelope. The average classification accuracy was 38.5%, which is significantly above the rate of chance (33.3%). These results demonstrate the synchronization between EEG during the imagined speech and the envelope of the overt counterpart.

Keywords:

phase synchronization, EEG, imagined speech, neural oscillations, speech envelope

Main text

Introduction

Inner speech, i.e., the subjective experience of hearing one's own voice in the mind, functions in various cognitive processes such as verbal rehearsal (Alderson-Day and Fernyhough, 2015). Recent studies have revealed neural correlations of inner speech generation: namely, a similar network as the speech production including the left inferior frontal region and the left premotor cortex (Price, 2012). Tian and Poeppel (2010) proposed an internal forward model during imagined speech, where a motor efference copy is sent from the motor planning region to the parietal cortex and a further efference copy is sent from the parietal cortex to the temporal cortex. These parietal/temporal areas activated by the two types of efference copy generate a kinesthetic feeling and auditory perceptual feelings, respectively.

In contrast to such research on neural correlations of imagined speech, the modulation of neural oscillations during imagined speech has not received much attention. Rather than imagined speech, so far, intensive research on neural dynamics has been conducted in the neurophysiological speech perception field. Many studies have shown that theta oscillations (4–8 Hz) in the auditory cortex match their phases to the amplitude envelope of speech during speech processing (see Peelle and Davis, 2012; Meyer, 2017 for recent reviews). It has been suggested that this synchronization is based on endogenous theta oscillations in Heschl's gyrus in the right hemisphere dominantly (Giraud et al., 2007).

A similar endogenous fluctuation in theta has also been observed in the part of the ventral premotor cortex related to control of the mouth (Giraud et al., 2007). Given that the amplitude envelope of speech is mainly derived from vowels voiced by the mouth opening, the speech envelope could conceivably be related to an oscillatory rhythm in the mouth premotor cortex in a similar way to synchronization in speech processing during perception. Thus, by considering this together with the fact that motor-related regions such as the premotor cortex can be activated by motor imagery (Gerardin et al., 2000), it can be hypothesized that neural oscillations during speech imagery synchronize with speech rhythms generated by the imagery. Although it is difficult to observe inner speech directly, overt speech can be regarded as a counterpart of imagined speech because the phonetic features of imagined speech are similar to those of overt speech (Filik and Barber, 2011).

In the current research, we partially focus on whether EEG during imagined speech is able to classify speech stimuli with different speech envelopes. Previous M/EEG research on speech perception demonstrated that neural oscillation phases during synchronization with perceived speech enables speech stimuli to be classified (Suppes et al., 1998; Luo and Poeppel, 2007; Howard and Poeppel, 2010; Watanabe et al., 2019) because the neural synchronization induces replicable and stimulus-specific phase patterns of oscillations across trials. Conversely, reliable EEG-based classification of speech stimuli with different speech envelopes suggests that a replicable and stimulus-specific neural phase pattern is induced by imagining the articulatory movements of the speech. Thus, above-chance level accuracies in the classification provide further evidence for EEG synchronization during imagined speech. In addition, such neural decoding of imagined speech can be applied to the neural prosthesis of speech production to help patients with locked-in syndrome communicate. Thus, classification analysis also sheds light on the potential of a brain-computer interface (BCI) for speech communication.

In sum, our research aimed to answer the following research questions: (1) whether EEG oscillations during imagined speech synchronize with the speech envelope of the overt counterpart and (2) whether EEG oscillations during imagined speech can classify speech stimuli with different amplitude envelopes. To this end, we regressed the overt speech envelope using EEG and calculated correlation coefficients between the EEG-based regressed envelope and the overt speech envelope. The classification was based on the distance between a test data and a template waveform of each class. Since the duration of the imagined speech was expected to vary across trials, we used dynamic time warping (DTW) to correct the durational variability for the classification analysis. To the best of our knowledge, this is the first study to investigate synchronization between the overt speech envelope and EEG oscillations during imagined speech.

Materials and methods

Participants

Eighteen right-handed L1 Japanese speakers participated in the experiment (6 female, 12 male, mean age: 23.8, SD = 1.7). They all gave written informed consent to their participation. No participants reported a history of hearing impairment or neurological disorders. The experiment was approved by the ethical review board of the Nara Institute of Science and Technology.

Speech stimuli

We recorded three speech stimuli from a female L1 Japanese speaker in a sound-attenuated chamber (44.1 kHz/16 bit). All speech stimuli were nonsense sounds because we wanted to avoid the effect of semantic processing of speech on the synchronization analysis. All stimuli consisted of three [ba] and two [ba:] with a prolonged vowel at different positions to differentiate speech

envelopes (stim. 1: [ba] [ba] [ba] [ba:] [ba:], stim. 2: [ba:] [ba:] [ba] [ba] [ba], stim. 3: [ba:] [ba] [ba] [ba] [ba:], Fig. 1). The duration of each stimulus was adjusted to 1,800 ms and the volume was normalized. The pitch height was adjusted to 200 Hz by using Praat (Boersma, 2002).

— [Figure 1 is around here] —

Experimental procedure

EEGs were recorded with an amplifier (BrainAmp DC, Brain Products GmbH., Germany) from 32 Ag/AgCl electrodes (actiCAP, Brain Products GmbH., Germany). The impedance of the electrodes was kept below 10 kΩ. The EEGs were online-filtered with a 0.016-Hz high-pass and 250-Hz low-pass filter. The sampling rate was 1,000 Hz. An FCz electrode and FPz electrode were used for the reference and ground, respectively. The experiment was controlled using Presentation software (Neurobehavioral Systems, Inc., U.S.A).

Participants sat on a comfortable chair in a dimly lit sound-attenuated chamber. A display monitor, keyboard, and microphone were placed on a desk in front of the chair. Participants were instructed to familiarize themselves with the speech stimuli and memorize them before the EEG recording. One trial consisted of three tasks: listening, speaking, and imagining speech. During the trial, they were instructed to stay as still as possible. In the imagined speech task, participants were forbidden to move any articulators such as their mouth or lips and were instructed to imagine the articulatory movements of the speech stimulus without actually making those movements.

The experiments consisted of a practice and three main blocks, where each block consisted of 21 and 20 trials. Each stimulus was presented to each participant 20 times in a randomized order in the main blocks. The procedure of a trial did not vary between the practice and the main blocks.

Each trial started from the appearance of “Ready?” on the display. Participants initiated a trial by pushing the space key on the keyboard, and then “LISTEN” was displayed for indicating the task type for 2,000 ms. After a countdown (from 3 to 1) to the start of task execution, a fixation mark (+) appeared, and at the exact same time, a speech stimulus was played via headphones. In the speaking task, the task indication (“SPEAK”) was followed by a countdown to task execution. After the countdown, participants uttered the speech stimulus that was presented in the listening task at the same speech rate. To reduce variation in the duration of uttered speech across trials, we controlled the timing of the start and the end by means of a progress bar. The progress bar appeared on the display immediately after the countdown, gradually extended horizontally during the imagination task, and stopped at 1,800 ms, which was the same duration as the speech stimuli, relative to the appearance of the progress bar. By having the participants initiate and finish their utterances at the same time as the appearance and stop of the progress bar, respectively, we were able to mitigate duration variabilities across trials. Participants' utterances were recorded using a microphone. In the imagined speech task, after the task indication (“IMAGINE”), the procedure was exactly the same as the speaking task except that participants imagined articulatory movements of the speech stimulus. Variabilities of the duration of the imagined speech were also adjusted by using the progress bar. After every imagination task, the participants were asked to press a button to indicate whether they had been able to imagine the speech (success: F key, failure: J key). The procedure of one trial is summarized in Fig. 2. The experiment continued for about 45 minutes.

— [Figure 2 is around here] —

Preprocessing of EEG data

We analyzed synchronization and its topographical patterns in both perceived speech and imagined speech: synchronization (1) between the amplitude envelope of the speech the participants listened to and EEGs during the time they perceived it and (2) between the amplitude envelope of the speech the participants uttered and the EEGs during imagined speech.

We used the FieldTrip toolbox (Oostenveld et al., 2011) for MATLAB (The MathWorks, Inc., U.S.A) for the EEG data analysis. To remove slow drift artifacts, a 4096th order finite impulse response (FIR) one-pass zero phase high-pass filter at 0.5 Hz (Hamming window) was applied to the continuous data. EEGs were re-referenced to an average of both mastoids and epoched from –1,000 ms to 3,000 ms relative to the task onset. The task onset was set to the appearance of the fixation mark (in the listening task) and the progress bar (in the imagined speech task). Epochs with large amplitudes exceeding $\pm$ 200 μv were rejected. Data from FP1 and FP2 were exempted from this rejection because they included large eye-related artifacts that were later removed during the independent component analysis (ICA). Epochs contaminated by large muscle artifacts were identified using an automatic detection method based on z-value of the data distribution (cutoff = 15) and visual inspection. ICA was used to correct the eye-related artifacts and remaining muscle artifacts. Candidates of eye-related independent components (ICs) were searched on the basis of the average Pearson correlation coefficients between the FP1/FP2 data and ICs. ICs to be removed were identified by visually inspecting their waveforms and spatial distributions.

We separated the EEG datasets on a per condition basis. In the imagined speech dataset, we excluded trials in which participants reported that they had not successfully imagined the speech and trials in which participants uttered the speech incorrectly, such as through a slip of the tongue

in the preceding speaking task. The incorrect utterances were annotated manually. One participant was removed from the analysis because of a large total number of rejected trials across conditions (above 30%). As a result, the average total of rejected trials across conditions was 8.7% (SD = 5.0). A one-way repeated ANOVA test showed no significant differences in the number of rejected trials between speech stimuli ($F(2, 32) = 1.47$, $p = 0.25$).

Speech envelope extraction and band-pass filtering

The amplitude envelope of the speech was extracted using a Hilbert transformation. A two-pass infinite impulse response (IIR) Butterworth band-pass filter of the 8th order (1–7 Hz) was applied to both the speech envelope and the preprocessed EEG, which was further extracted from 0 to 3,000 ms. The frequency ranges of the band-pass filter were decided in order to extract the low-frequency modulation (see Fig. 1; peaks in the frequency domain were observed around 2 and 5 Hz). To avoid filter artifacts, the flipped data were concatenated to the beginning and the end of the data and were removed after the filtering procedure. The speech envelope was downsampled with the same sampling rate as EEG (i.e., 1,000 Hz).

Optimizing the delay in synchronization

We corrected the delay in synchronization between the EEGs and the speech envelope because a certain delay can be expected (e.g., a number of milliseconds for participants to recognize the appearance of the progress bar and begin imagining the speech). The cross-correlation coefficients between all concatenated trials of the band-pass filtered EEG and speech envelope were calculated per EEG electrode and then averaged across electrodes. Before calculating the coefficients, both data were demeaned. We searched for the peak showing the highest coefficients within the lag range [10,

200 ms] for perception data and [180, 500 ms] for imagined speech data. A narrower lag range was adopted for perception data because less variability in the delay was expected than for the imagined data, where participants needed to respond to the appearance of the progress bar to initiate the imagined speech. In contrast, the relatively wide range of the lag in the imagined speech was set because larger variability of the reaction times across participants was expected in the imagined speech. The lag range of the imagined speech was chosen with consideration of the fact that human simple reaction time is generally around 200 ms (Ulrich et al., 1998).

After each EEG trial was shifted in the time domain by this delay per participant data, EEG data from the new onset time point to 1,800 ms, which was the same duration as the speech stimulus, were used for further analysis of the perceived data. In the case of the imagined speech data, data from the new onset time point to the duration of the corresponding overt speech (i.e., overt speech data in the speaking task immediately before the imagined speech) were extracted under the assumption that the durations of the overt speech and imagined speech were similar.

Synchronization analysis

A multiple linear regression method (Bayraktaroglu, et al., 2011) was used for analyzing the synchronization between the EEG and the speech envelope. Specifically, we used a concatenated, delay-optimized, band-pass filtered EEG matrix $M^{n \times m}$ across trials, where n is the total number of data points and m is the number of electrodes. Each electrode data was demeaned by subtracting the electrode average value. First, the number of dimensions was reduced using PCA in order to avoid collinearity. M was projected into the space spanned by eigenvectors of the covariance matrix of M covering 99% of the variance. The projected M and the space spanned by eigenvectors were expressed by M_k and P_k, respectively. After the concatenated speech envelope across trials, denoted

by s, was demeaned, the envelope was modeled by a linear combination of the columns of M_k and noise ϵ:

$$s = M_k b + \epsilon. \qquad (1)$$

The optimal coefficients b were estimated using the least-squares method. The regressed speech envelope $\hat{s}$ is expressed by

$$\hat{s} = M_k b. \qquad (2)$$

The spatial pattern of b on the topography (p_{eeg}) was calculated on the basis of Bayraktaroglu et al. (2011) and Parra et al. (2002). First, the pattern p_{pca} was calculated as

$$p_{pca} = \frac{M_k^T \hat{s}}{\hat{s}^T \hat{s}}. \qquad (3)$$

From Eq. (3), p_{pca} can be regarded as the coupling between the regressed speech envelope $\hat{s}$ and the projected EEG data M_k (Bayraktaroglu et al., 2011). Finally, p_{eeg} was calculated using the previous PCA space P_k:

$$p_{eeg} = P_k p_{pca}. \qquad (4)$$

In terms of visualizing the topography, the absolute value of each participant's p_{eeg} was normalized to have a certain range [0, 1] (1 represents the maximum coefficients of the synchronization). We calculated the coefficients of the Spearman rank correlation (Spearman's rho) between the EEG-based regressed envelope ($\hat{s}$) and speech envelope (s) as an index of the synchronization. Spearman's rho was converted using the Fisher-Z transform to approximate a normal distribution.

Classification analysis

We performed a classification analysis to investigate whether EEG during speech perception and speech imagery included a signature of the envelope of the perceived speech and of the overt speech, respectively. Classifiers were trained using the delay-optimized, band-pass filtered EEG data in a speech perception and imagined speech task. The classification method was similar to Zhang et al. (2012)'s electrocorticography (ECoG)-based classification of overt speech. Since we expected the duration of the imagined speech to vary across trials regardless of the duration variability mitigation afforded by the progress bar, each EEG data in the imagined speech task was realigned using the DTW method, which is an algorithm to find a path that minimizes the distance between two signals. The performance of the classifier was evaluated by leave-one-out cross-validation.

The classification was based on the Euclidean distance between a test data and a template waveform of each class. The templates were constructed by using the training data. To this end, first, training data were separated on the basis of the class labels. In the imagined speech data, each training data was realigned to the envelope of the speech stimulus (see Fig. 1) corresponding to the class label using the DTW in order to correct duration variability. In the perception data, realignment was not performed because less durational variability was expected. EEG data and the envelope of the speech stimulus data were standardized by using z-score in order for all data to take values in a similar scale. Each template waveform of the class label was constructed by averaging the training data belonging to the class in the time domain. The class label of the template waveform showing the least square Euclidean distance to a test data was considered the prediction result. In the case of the imagined data, each test data was also realigned to each template using the DTW before the

classification and the distance between the realigned test data and the template were calculated because the duration of the test data also differed across trials.

For classification, we used the five electrodes showing the highest absolute values of the coefficients of the EEG pattern among electrodes positioned in the frontal and central region (i.e., Fz, F3, F4, F7, F8, FC5, FC6, FCz, FC1, FC2, Cz, C3, C4, CP5, CP6, CP1, and CP2), as the synchronization was mainly observed in these regions (see Results). The final prediction of the speech stimuli was decided by a voting system across the results of the five electrodes.

There is a possibility that the classification was performed only on the basis of event-related responses to the perceived speech or the task onset, not on the basis of the neural synchronization with the slow modulations of the perceived speech and imagined speech. In order to exclude this possibility, we performed an additional classification using data in which the first part of the waveform was excluded. In this classification, when calculating the distances between the test data and each template, the first 150 ms of data in the waveform were ignored.

Results

First, we plotted the histograms of all participants' estimated delays of perceived and imagined speech EEG data (Fig. 3). The average delays across participants were 126 ms (SD = 49) and 364 ms (SD = 77) for the perceived and imagined speech, respectively.

— [Figure 3 is around here] —

As for the synchronization analysis, the Spearman's rho averaged across participants was 0.15 (SD = 0.04) and 0.10 (SD = 0.03) for perceived and imagined speech, respectively. One sample t-test

revealed that both Spearman's rhos significantly differed from zero (perceived: t (16) = 14.3, p<0.01, imagined: t (16) = 15.0, p<0.01). Box plots of the Spearman's rho in the perceived and imagined speech are provided in Fig. 4(A). We also plotted an example of synchronization between the EEG-based regressed envelope and a corresponding speech envelope (Fig. 4(B)). Both envelopes were standardized using z-score for visualization.

— [Figure 4 is around here] —

The grand averages of EEG patterns across participants are shown in Fig. 5(A). The EEG pattern of the perceived data showed the synchronization at the electrodes in a central region. In contrast, the pattern of imagined data was distributed in a more frontal region. This indicates that the neural generator differs across perceived and imagined speech.

In the classification analysis, mean accuracies across participants were 54.7% (SD = 10.8) for perceived speech and 38.5% (SD = 5.3) for imagined speech (Fig. 5(B)). One sample t-test revealed that the accuracies were significantly above the 33.3% chance rate (perceived: t (16) = 7.9, p<0.01, imagined: t (16) = 3.9, p<0.01). The accuracies of the classification based on perceived data significantly outperformed the ones based on imagined speech (paired sample t-test: t (16) = 5.2, p<0.01).

— [Figure 5 is around here] —

As for the classification using the data from which the first 150 ms were excluded, data in both the perceived and imagined speech showed similar results to the previous classifications: 53.8% (SD = 9.3) for perceived speech and 37.9% (SD = 6.0) for imagined speech. Both accuracies significantly

outperformed the chance rate (one-sample t-test, perceived: $t(16) = 8.8$, $p<0.01$, imagined: $t(16) = 3.0$, $p<0.01$).

Discussion

The purpose of this research was to investigate the synchronization of imagined speech and neural oscillations during generated speech imagination. To this end, we substituted participants' overt speech for imagined speech because it is impossible to observe inner speech physically. Multiple regression-based analysis (Bayraktaroglu et al., 2011) revealed a significant correlation between the EEG-based regressed speech envelope and the overt speech envelope as well as the perceived data. In addition, the classification performances of speech stimuli with different envelopes achieved the accuracy of 38.5% using the EEG during the imagined speech. These results indicate that EEG oscillations during the imagined speech contain the signature of the speech envelope of the overt counterpart of the imagined speech.

The EEG patterns of synchronization in the perceived and imagined data were differently distributed over the scalp: to the central region and the frontal region, respectively. Although it is difficult to identify the generator by using EEG due to the low signal-to-noise ratio and volume conduction, the difference in the topographies, at least, indicates that the neural generators of the synchronization differ from each other. In the case of speech perception, the neural source of the synchronization is the auditory cortex (Giraud et al., 2000; Kubanek et al., 2013), while the generator of the synchronization during the imagined speech seems to be in the more frontal parts, for example, in the frontal lobe including the motor-related area.

As mentioned in the Introduction, one candidate of the generator of the imagined speech synchronization is the ventral premotor cortex, as Giraud et al. (2007) revealed the endogenous fluctuation at the theta frequency band in the ventral premotor cortex of the mouth. Considering that the waveform of the speech envelope is formed on the basis of the cycling of the mouth opening, conceivably the endogenous theta oscillations in the region are modulated depending on the timing of the cycling of the mouth opening during the imagined speech.

Assuming that the generator is the ventral premotor cortex, one question is raised: what is the functional role of the phase synchronization in the premotor cortex during imagined speech? We know that during the periodical auditory stimulus process in the brain, neural oscillations show phase-locked responses to the periodicity of the stimulus (Will and Berg, 2007). The functional role of the phase-locked responses to the external periodicity is to enable the brain to predict the timing of future input stimuli and process them at the state of high excitability (see Arnal and Giraud, 2012 for a review) because the neural oscillation phases are related to neuronal excitability (Lakatos et al., 2005) and the state of the ongoing oscillatory activity in the pre-stimulus period affects the processing of sensory stimuli (Arieli et al. 1996; Romei et al., 2010). This phase-locked response is also widely observed during speech perception: synchronization between the speech amplitude envelope, which features pseudo-periodical fluctuations around 4–8 Hz, and the neural oscillation phase in the theta frequency band (Ahissar et al., 2001; Luo and Poeppel, 2007; Howard and Poeppel, 2010; Luo and Poeppel, 2012; Peelle et al., 2012). Because syllabic information is dominant in the amplitude envelope of speech, it has been suggested that the syllabic information is sampled and segmented via this neural phase synchronization (see Meyer, 2017 for a review). As evidence to support this notation, Hyafil et al. (2015)'s computational model demonstrated that the syllable boundaries can be reliably predicted from neural oscillations during speech perception.

Considering that the phase synchronization during speech perception may be related to the syllable duration, phase synchronization during imagined speech might be also related. The Directions Into Velocities of Articulators (DIVA) model (Guenther, 2006; Guenther et al., 2006) defines the role of the ventral premotor cortex during speech production as a speech sound map that stores a repository of the motor programs of frequently observed production units such as syllables. Together with the DIVA model and the current result, we speculate that during syllable production, a motor program of the syllable is read from the ventral premotor region and the syllable duration (i.e., how long the syllable is pronounced) is encoded in the timing of the neural oscillation phases in the region. Alternatively, Zhang et al (2012) revealed that the ECoG-based classification of overt sentences, which features a classification method similar to the current research, was successful at an electrode corresponding to the Broca area. This result suggests that the Broca region also might show synchronization with the imagined or overt speech envelope. In order to further investigate the neural source and the functional role of the neural synchronization during imagined speech, we aim to localize the neuronal source using other brain imaging techniques such as functional magnetic resonance imaging (fMRI) in the future.

In this study, we used only two types of syllable ([ba] and [ba:]) for the purpose of controlling the segmental information across speech stimuli. However, even if we used different syllables such as [ku] and [ku:] or [de] and [de:] with the same sequence as the current research, we expect a similar result would be observed. This is because we analyzed the similarity between the band-pass filtered EEG and the envelope pattern of overt speech in both synchronization and classification analysis (synchronization analysis: correlation, classification analysis: Euclidean distances) and did not use the information related to the segmental information itself. Thus, if we had used different syllables from [ba] and [ba:], the synchronization would probably be observed as long as the stimuli include

a rhythm of the speech envelope. Supporting this, Deng et al. (2010) performed EEG-based classification of trials with different timings of syllable imagination (3 different timings × 2 syllables: [ba] and [ku]), showing significant accuracies of three-class of the syllable imagination timing using EEG data of both [ba] and [ku].

Another question is whether the observed synchronization in the current research is specific to language processing or not. It is possible that the synchronization and reliable classification performances were obtained when the participants imagined the rhythms of non-linguistic content, such as beating a drum or blowing a whistle. We have stated that the syllable duration information might be encoded in the low-frequency oscillations in the ventral premotor cortex related to mouth control. If so, it is predicted that the synchronization would not be observed when non-linguistic content, which is not related to controlling mouth movements, is imagined. Alternatively, it might be possible that the synchronization to non-linguistic rhythms might be observed in other motor-related brain regions. Comparing the results of tasks where participants imagine the rhythm of producing syllables and non-linguistic rhythms such as the rhythm of a beat will lead to further understanding of the functional role of synchronization during imagined speech.

We obtained classification accuracies (38.5%) significantly above the level of chance in classifying speech stimuli with different envelopes using EEG during the imagined speech. This result partially supports the neural synchronization with the overt speech envelope because it indicates that EEG oscillations during imagined speech induce a stimulus-specific, replicable oscillation pattern. While the neural synchronization with speech envelope during speech perception has been applied to M/EEG-based sentence classification (Luo and Poeppel, 2007; Howard and Poeppel, 2010; Watanabe et al., 2019), the current research applied it to the classification of imagined speech. From

the viewpoint of neural engineering, such neural decoding of speech without movements functions as a non-invasive method of uttering speech for people with locked-in syndrome. So far, there have been many studies on noninvasive neural decoding of imagined speech performed on the classification of vowels (DaSalla et al., 2009), syllables (D'zmura et al., 2009), and words (Salama et al., 2014) and the rhythms of syllable imagination timing (Deng et al., 2010). However, so far, the neurophysiological mechanisms enabling the neural decoding of imagined speech have not been sufficiently investigated. The current research provides a novel, neurophysiologically motivated feature for neural decoding of imagined speech.

One of the difficulties in the neural decoding of imagined speech stems from the variable duration of imagined speech, which cannot be observable directly. We demonstrated that duration control using a progress bar and the DTW-based correction inhibited this variability across trials and enabled a reliable classification performance, at least, above the chance level. However, because the classification performances based on the EEG during the imagined speech were significantly lower than those during the perceived data, further considerations to improve the classification accuracy—at least, in order to achieve a similar performance level to the perceived data—are required.

As an alternative possible explanation of the above-chance level classification performances, one might argue that the trained classifier relied on the speech onset, as opposed to the dynamics of the low-frequency modulation of the speech stimuli or imagined speech, because there is a possibility that the difference in the segmental information of the onset of the speech stimuli (e.g., [ba] and [ba:]) evoked different responses to the event onset. However, when we performed classification using the EEG data with the first 150 ms omitted, the classification performances remained

significantly above the level of chance. Thus, in the current classification, we tend to conclude that the whole waveform pattern of EEG, not the onset information, contributed to the classification of speech stimuli with different envelope patterns.

Conclusion

This is the first study to reveal synchronization between EEG oscillations during imagined speech and the speech envelope of the overt counterpart. The topographical pattern of the synchronization showed that this synchronization during imagined speech was observed at electrodes around the frontal region on the topography. The classification analysis of stimulus with different envelopes demonstrated that the average classification accuracy, 38.5%, was significantly above the level of chance, indicating the synchronization between EEG during imagined speech and the envelope of the overt counterpart, and that it induces a replicable, stimulus-specific oscillatory pattern.

Acknowledgements

We thank Dr. Lars Meyer (Max Planck Institute for Human Cognitive and Brain Sciences) for his insightful comments about the synchronization analysis. Part of this work was supported by JSPS KAKENHI Grant Numbers JP16K16172, JP17H06101, JP17K00237, and JP18J14871.

References

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., Merzenich, M. M. 2001. Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. Proc. Natl. Acad. Sci. U.S.A. 98, 13367-13372.

Alderson-Day, B., Fernyhough, C. 2015. Inner speech: development cognitive functions, phenomenology, and neurobiology. Psychol. Bull. 141, 931–965.

Arieli, A., Sterkin, A., Grinvald, A. Aertsen, A. D. 1996. Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. Sci, 273, 1868-1871.

Arnal, L. H., Giraud, A. L. 2012. Cortical oscillations and sensory predictions. Trends Cogn. Sci. 16, 390–398.

Bayraktaroglu, Z., von Carlowitz-Ghori, K., Losch, F., Nolte, G., Curio, G., Nikulin, V. V. 2011. Optimal imaging of cortico-muscular coherence through a novel regression technique based on multi-channel EEG and un-rectified EMG. Neuroimage. 57, 1059–1067.

Boersma, P. 2002. Praat, a system for doing phonetics by computer. Glot. Int. 5, 341-345.

Deng, S., Srinivasan, R., Lappas, T., D'Zmura, M. 2010. EEG classification of imagined syllable rhythm using Hilbert spectrum methods. J. N.a. Eng. 7, 046006.

DaSalla, C. S., Kambara, H., Sato, M., Koike, Y. 2009. Single-trial classification of vowel speech imagery using common spatial patterns. N.a. Netw. 22, 1334-1339.

D'Zmura, M., Deng, S., Lappas, T., Thorpe, S., Srinivasan, R. 2009. Toward EEG sensing of imagined speech. Proc. Int. Conf. Hum. Comput. Interact. 40-48, Berlin, Germany.

Filik, R., Barber, E. 2011. Inner speech during silent reading reflects the reader's regional accent. PLoS One. 6, e25782.

Gerardin, E., Sirigu, A., Lehéricy, S., Poline, J. B., Gaymard, B., Marsault, C., Agid, Y., Le Bihan, D. 2000. Partially overlapping neural networks for real and imagined hand movements. Cereb. Cortex. 10, 1093–1104.

Giraud, A. L., Kleinschmidt, A., Poeppel, D., Lund, T. E., Frackowiak. R. S. J., Laufs, H. 2007. Endogenous cortical rhythms determine cerebral specialization for speech perception and production. Neuron. 56, 1127–1134.

Giraud, A. L., Lorenzi, C., Ashburner, J., Wable, J., Johnsrude, I., Frackowiak, R., Kleinschmidt, A. 2000. Representation of the temporal envelope of sounds in the human brain. J. Neurophysiol, 84, 1588-1598.

Guenther, F. H. 2006. Cortical interactions underlying the production of speech sounds. J. Commun. Disord. 39, 350-365.

Guenther, F. H., Ghosh, S. S., Tourville, J. A. 2006. Neural modeling and imaging of the cortical interactions underlying syllable production. Brain Lang. 96, 280-301.

Howard, M. F., Poeppel, D. 2010. Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. J. Neurophysiol. 104, 2500-2511.

Hyafil, A., Fontolan, L., Kabdebon, C., Gutkin, B., Giraud, A. L. 2015. Speech encoding by coupled cortical theta and gamma oscillations. Elife. 4, e06213.

Kubanek, J., Brunner, P., Gunduz, A., Poeppel, D., Schalk, G. 2013. The tracking of speech envelope in the human cortex. PloS One. 8, e53398.

Lakatos, P., Shah, A. S., Knuth, K. H., Ulbert, I., Karmos, G., Schroeder, C. E. 2005. An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. J. Neurophysiol, 94, 1904-1911.

Luo, H., Poeppel, D. 2007. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron. 54, 1001–1010.

Luo, H., Poeppel, D. 2012. Cortical oscillations in auditory perception and speech: evidence for two temporal windows in human auditory cortex. Front. Psychol, 3, 170.

Meyer, L. 2017. The neural oscillations of speech processing and language comprehension: state of the art and emerging mechanisms. Eur. J. Neurosci. 48, 2609-2621.

Oostenveld, R., Fries, P., Maris, E., Schoffelen, J. M. 2011. FieldTrip: open source software for advanced analysis of MEG, EEG and invasive electrophysiological data. Comput. Intell. Neurosci. 2011, 1.

Parra, L., Alvino, C., Tang, A., Pearlmutter, B., Yeung, N., Osman, A., Sajda, P. 2002. Linear spatial integration for single-trial detection in encephalography. Neuroimage. 17, 223–230.

Peelle, J. E., Davis, M. H. 2012. Neural oscillations carry speech rhythm through to comprehension. Front. Psychol, 320.

Peelle, J. E., Gross, J., Davis, M. H. 2012. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. Cereb. Cortex, 23, 1378-1387.

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.

Romei, V., Gross, J., Thut, G. 2010. On the role of prestimulus alpha rhythms over occipito-parietal areas in visual input regulation: correlation or causation?. J. Neurosci, 30, 8692-8697.

Salama, M., ElSherif, L., Lashin, H., Gamal, T. 2014. Recognition of unspoken words using electrode electroencephalograhic signals, Proc. Int. Conf. Adv. Cogn. Technol. Appl, 51-55, Venice, Italy.

Suppes, P., Han, B., Lu, Z. L. 1998. Brain-wave recognition of sentences. Proc. Natl. Acad. Sci. U.S.A. 95, 15861-15866.

Tian, X., Poeppel, D. 2010. Mental imagery of speech and movement implicates the dynamics of internal forward models. Front. Psychol. 1, 166.

Ulrich, R., Rinkenauer, G., Miller, J. 1998. Effects of stimulus duration and intensity on simple reaction time and response force. J. Exp. Psychol. Hum. Percept Perform. 24, 915-928.

Watanabe, H., Tanaka, H., Sakti, S., Nakamura, S. 2019. Neural oscillation-based classification of Japanese spoken sentences during speech perception, IEICE Trans. Inf. Syst. E102-D, 383-391.

Will, U., Berg, E. 2007. Brain wave synchronization and entrainment to periodic acoustic stimuli. Neurosci. Lett. 424, 55-60.

Zhang, D., Gong, E., Wu, W., Lin, J., Zhou, W., Hong, B. 2012. Spoken sentences decoding based on intracranial high gamma response using dynamic time warping. Proc. Int. Conf. IEEE. Eng. Med. Biol. Soc. 3292-3295, San Diego, U.S.A.

Figure Captions

Figure 1.

(Left) Waveforms and spectrogram of speech stimuli. (Right) Amplitude spectrum of speech stimuli as a function of frequency.

Figure 2.

Procedure of an experimental trial. In the listening task, the stimulus was played after the task indication followed by a countdown to task execution. In the speaking task, participants uttered the speech stimuli into a microphone. In the imagined speech task, they imagined the articulatory movement of speech stimuli without making movements. After the imagined speech task, participants reported whether they had successfully imagined the speech by pressing a button.

Figure 3.

Histograms of estimated delays in perceived and imagined speech. Filled curves represent the densities of the distributions.

Figure 4.

(A) Box plots of Spearman's rho between EEG-based regressed speech envelope and speech envelopes per condition. (B) An example of the EEG-based regressed envelope and the corresponding speech envelope from subject 03 in perceived and imagined speech.

Figure 5.

(A) Grand averaged synchronization patterns across participants in perceived and imagined speech. (B) Box plots of accuracies in EEG-based classification of speech stimuli with different amplitude

envelopes in perceived and imagined speech. Dotted horizontal line represents the level of chance (33.3%)

FIGURES

Figure 1.

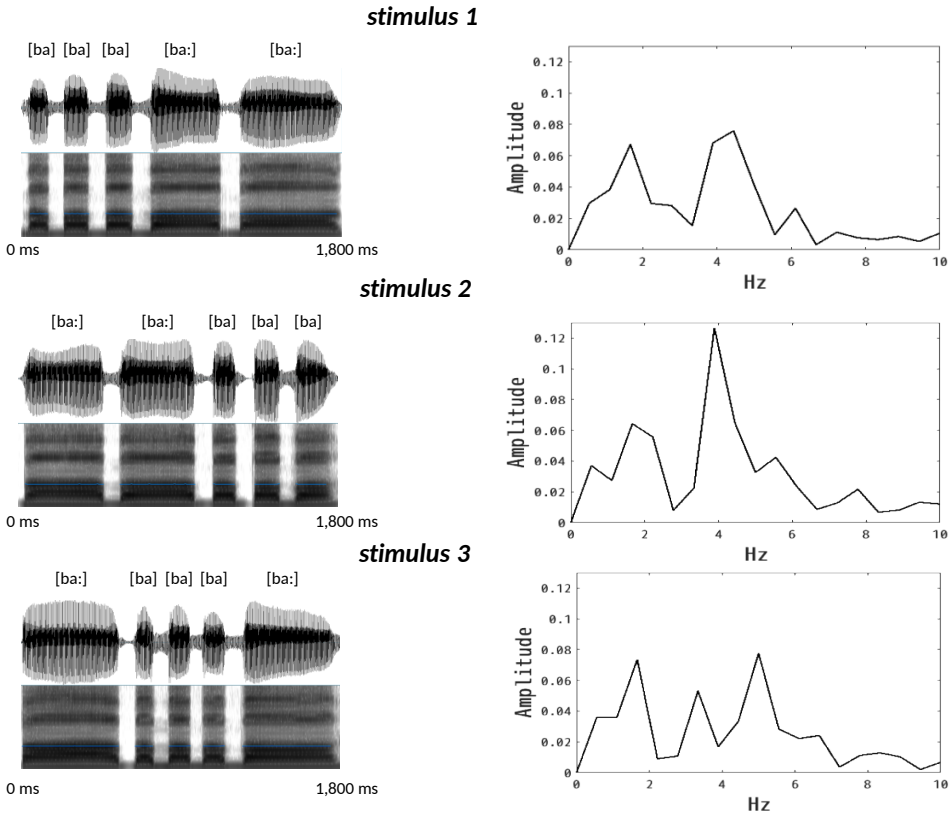


Figure 2.

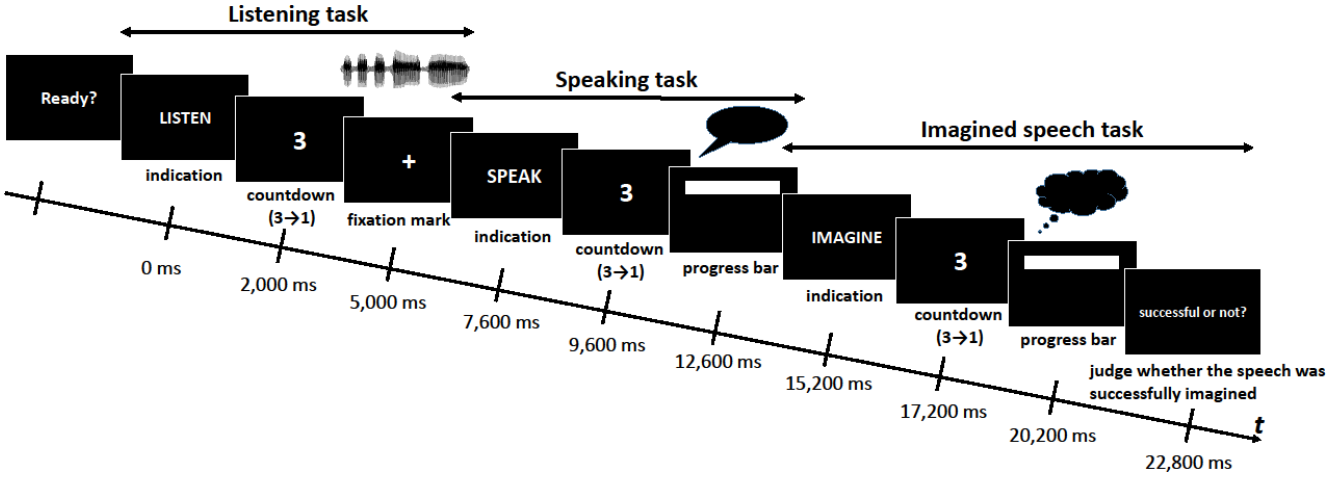

Listening task
Speaking task
Imagined speech task
Ready?
LISTEN
indication
3
countdown
(3→1)
+
fixation mark
SPEAK
indication
3
countdown
(3→1)
progress bar
IMAGINE
indication
3
countdown
(3→1)
progress bar
successful or not?
judge whether the speech was successfully imagined
0 ms
2,000 ms
5,000 ms
7,600 ms
9,600 ms
12,600 ms
15,200 ms
17,200 ms
20,200 ms
22,800 ms
t

Figure 3.

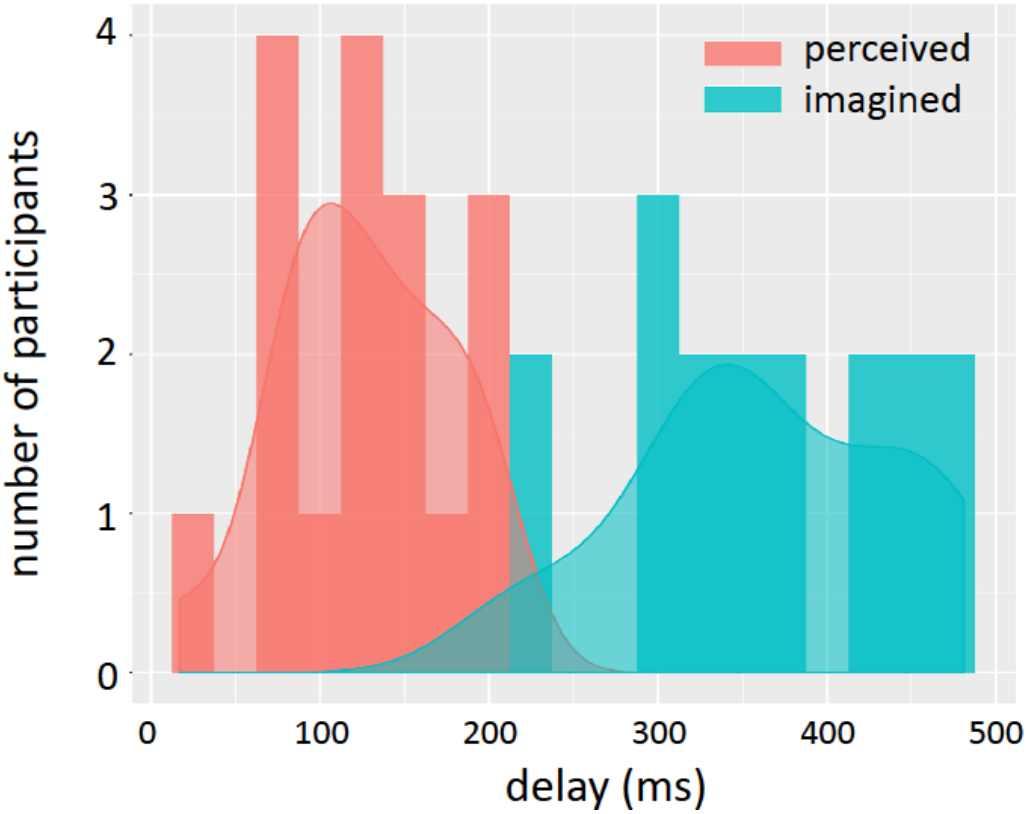

perceived
imagined
number of participants
0
1
2
3
4
0
100
200
300
400
500
delay (ms)

Figure 4.

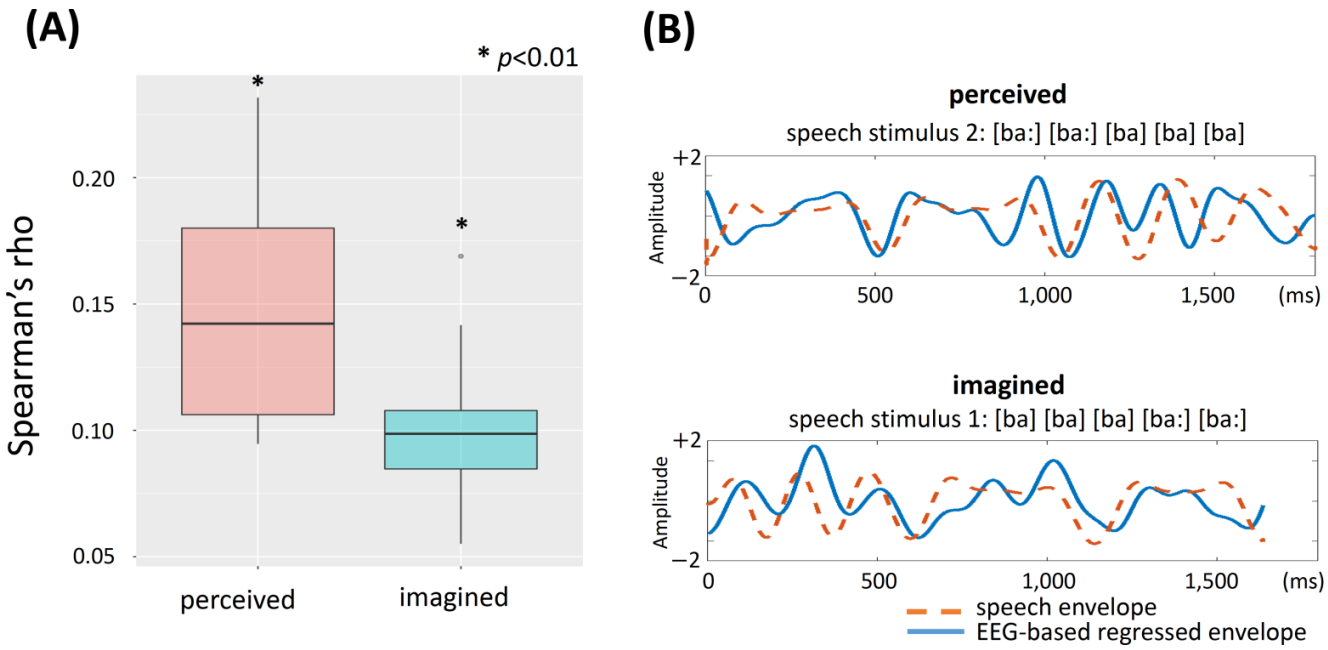
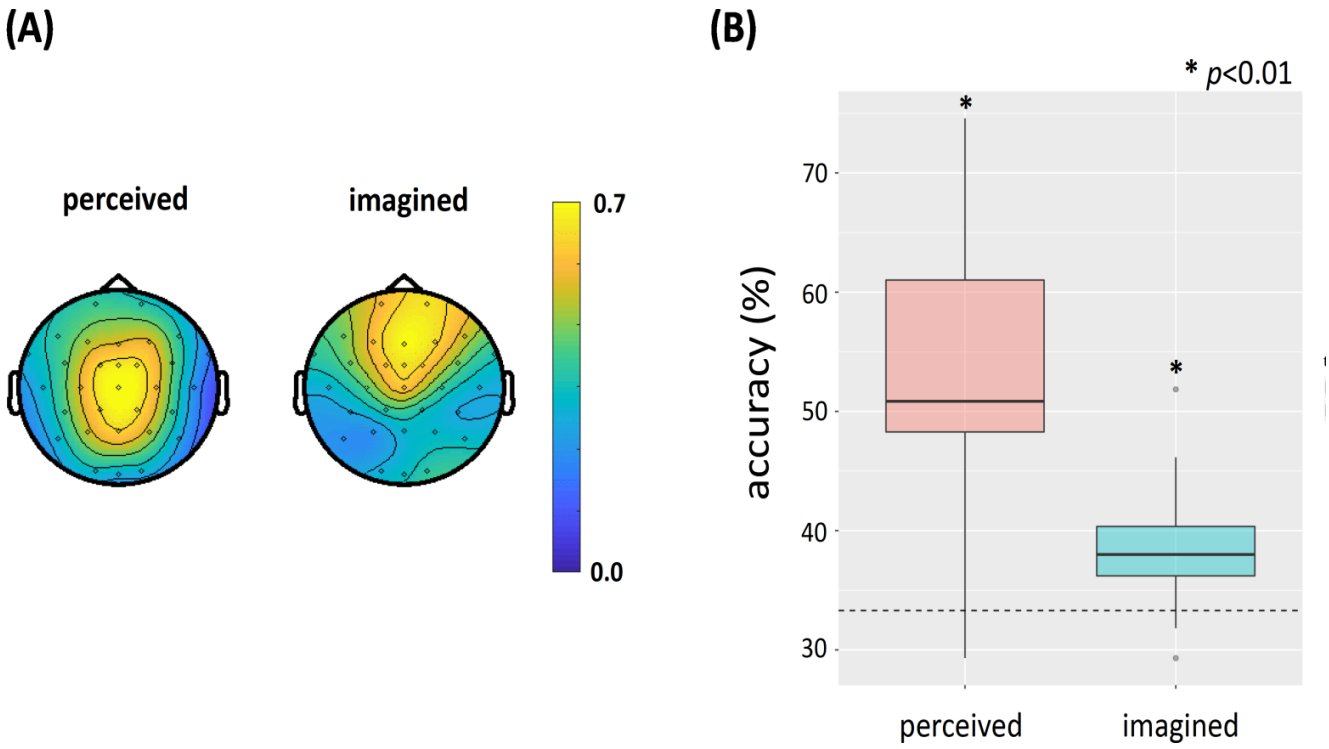

(A)
* p<0.01
Spearman's rho
0.20
0.15
0.10
0.05
perceived
imagined
(B)
perceived
speech stimulus 2: [ba:] [ba:] [ba] [ba] [ba]
Amplitude
+2
−2
0
500
1,000
1,500
(ms)
imagined
speech stimulus 1: [ba] [ba] [ba] [ba:] [ba:]
speech envelope
EEG-based regressed envelope

Figure 5.


(A)
perceived
imagined
0.7
0.0
(B)
* p<0.01
accuracy (%)
70
60
50
40
30
perceived
imagined