# Spatiotemporal Characteristics of Cortical Activities Associated with Articulation of Speech Perception

Naoki Saga[1], Hajime Yano[1], Tetsuya Takiguchi[2], Yoshiharu Soeta[3] and Seiji Nakagawa[4]

*Abstract*— **Recently, brain computer interface (BCI) technologies that control external devices with human brain signals have been developed. However, most of the BCI systems, such as P300-speller, can only discriminate among options that have been given in advance. Therefore, the ability to decode the state of a person's perception and recognition, as well as that person's fundamental intention and emotions, from cortical activity is needed to develop a more general-use BCI system. In this study, two experiments were conducted. First, articulations were measured for Japanese monosyllabic utterances masked by several levels of noise. Second, auditory brain magnetic fields evoked by the monosyllable stimuli used in the first experiment were recorded, and neuronal current sources were localized in regions associated with speech perception and recognition — the auditory cortex (BA41), the Wernicke's area (posterior part of BA22), Broca's area (BA22), motor (BA4), and premotor (BA6) areas. Although the source intensity did not systematically change with SNR, the peak latency changed along SNR in the posterior superior temporal gyrus in the right hemisphere. The results suggest that the information associated with articulation is processed in this area.**

## I. INTRODUCTION

The improvement of brain computer interface (BCI) technologies has been remarkable in the past decade. A BCI provides direct communication between the human brain and external devices. One of the most popular BCI systems uses the P300 response, which is typically elicited by a low-probability target stimulus mixed with high-probability standard stimulus. However, the P300-BCI only discriminates among options given in advance, and, at this point in time, it does not enable people to express their intrinsic intentions through the computer. In order to develop more versatile BCI systems, it is desirable to decode brain activities associated with speech perception and recognition.

N1m response is the most prominent component observed in the auditory cortex. N1m response is affected by cognitive process to some extent, but mainly reflects the physical properties of the stimulus. In addition, it appears about 100 ms after the onset of the stimulus, so it is hard to say if it strongly reflects higher processing, such as the linguistic process [1].

On the other hand, it is assumed that the linguistic information in speech sounds is processed after the acoustic properties of sounds are analyzed in the auditory cortex. As representative examples of brain regions involved in language information processing, Wernicke's area, the angular gyrus, and the supramarginal gyrus are widely known [2]. In addition, recent studies suggest that brain areas associated with speech production are also involved in speech perception. For instance, Pulvermüller *et al.* reported that the area used to articulate a phone in the premotor cortex is also activated by the perception of the phone [3]. Besides this, Broca's area, which has been said to control speech production, also takes part in perceiving speech sounds [4]. Hickok and Rauschecker proposed the dual-stream processing model, which assumes there are two pathways in perceiving speech sounds [5][6]. One is the dorsal pathway, which regulates phonological information 'where / how' in audition from the auditory cortex to the premotor cortex via the Wernicke' area, angular gyrus, and supramarginal gyrus. The other is the ventral pathway, which regulates 'what' in audition from the auditory cortex to Broca's area.

In terms of brain activities in speech perception under noisy environments, Wang *et al.* reported that the activity in the superior temporal gyrus in both hemispheres increases when articulation decreases [7], but this experiment compares changes of two extreme conditions. The spatial characteristics of brain activity associated with gradual changes of articulation and its latency have not yet been clarified.

This study was carried out to clarify the neural mechanisms involved with processing the articulation of speech perception. We attempted to observe the spatiotemporal properties of brain activities correlated with speech articulation using magnetoencephalography (MEG). First, articulations were measured for Japanese monosyllable utterances masked by several levels of noise. Second, auditory brain magnetic fields evoked by the monosyllable stimuli used in the first experiment were recorded and neuronal current sources were estimated. Then, the peak intensity and peak latency of the sources were compared over the cortical area associated with speech perception and recognition.

## II. ARTICULATION TEST

### A. Subjects

Seven adults, who are native Japanese speakers and right-handed, participated in this study (four males and three females). Their ages ranged from 20 to 39 years and the

---

*This work was not supported by any organization

[1] Naoki Saga and [1] Hajime Yano are with Kobe University, Kobe, Hyogo and National Institute of Advanced Industrial Science and Technology (AIST), Ikeda, Osaka, Japan. { `saga, pi314yano` } `@me.cs.scitec.kobe-u.ac.jp`

[2] Tetsuya Takiguchi is with Kobe University, Kobe, Hyogo, Japan. `takigu@kobe-u.ac.jp`

[3] Yoshiharu Soeta is with the National Institute of Advanced Industrial Science and Technology (AIST), Ikeda, Osaka. `y.soeta@aist.go.jp`

[4] Seiji Nakagawa is with Chiba University, Chiba, and National Institute of Advanced Industrial Science and Technology (AIST), Ikeda, Osaka, Japan. `s-nakagawa@chiba-u.jp`

978-1-5386-3646-6/18/$31.00 ©2018 IEEE

averaged age was 24.4 years. All of them had normal hearing and no history of neurological disorders. Each participant gave informed consent prior to commencing the experiment involved with this study.

### B. *Stimuli*

Auditory stimuli were 100 Japanese monosyllable utterances taken from a commercially-available database (NTT-AT FW03), produced by a female native Japanese speaker. These speech sounds were calibrated so that the loudness of each syllable was equal based on the report of Nagatani *et al.* [8]. Each syllable was embedded in white noise at a specific SNR. The levels of noise were 5 phases: SNR -10, -5, 0, 7, 15 dB. These SNR values were decided based on the results of the preliminary experiment so that the articulation scores would be 25, 40, 55, 70, 85 %. The length of a total stimulus was set to 0.4716 seconds. Only white noise was presented in the first 25 ms of the stimulus and in the section from the end of speech sound to the end of stimulus. In addition, the sound volume was adjusted for each listener to sound clear and feel comfortable to listen to.

### C. *Methods*

The experiment was carried out in a soundproof room and the stimuli were presented bilaterally using earphones. The stimuli were presented every 5 seconds, and the participants were asked to fill in the answer sheets as heard during interval. First, as a training session, the participants listened to 100 monosyllables without any background noise. Then the listening test was conducted for each SNR condition in random order.

### D. *Results*

Fig. 1 shows the results of the articulation test. A tendency toward decreased articulation with decreasing SNR was observed across all participants. However, considering the correlation between SNR and articulation for each syllable, there were some syllables in which articulation was always low regardless of the SNR levels.

## III. MEG RECORDING

### A. *Methods*

The same subjects used in the articulation test participated in the MEG recording. MEG data were recorded using a 122-channel, whole-head neuromagnetometer (Neuromag-122$^{\mathrm{TM}}$, Neuromag, Ltd., Helsinki, Finland) in a magnetically-shielded room. Among the mono-syllabic utterances used in the articulation test, the magnetic field was measured for the syllables (/pe/ /kyo/ /shu/) whose correlation between SNR and articulation was the largest. Stimulus sounds were presented with insertable earphones in the same way as the articulation test. The interval of the stimuli was 2 ms because the participants did not have to write anything down, unlike the articulation test. First, as a training session, participants were asked to listen to 100 syllables without background noise. And then highly correlated syllables, which were determined by the articulation test, were presented in random order with the SNR adjusted at -10, -5, 0, 7, 15 dB. In order to help keep the participants concentrated on listening, they were instructed to respond by pressing a button when to the heard the target sounds (/a/ i/ /u/), which were inserted at a certain interval.

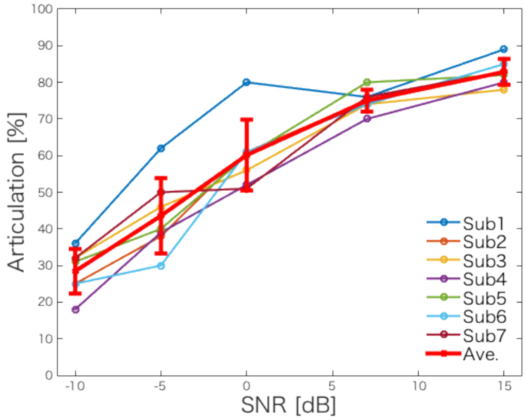
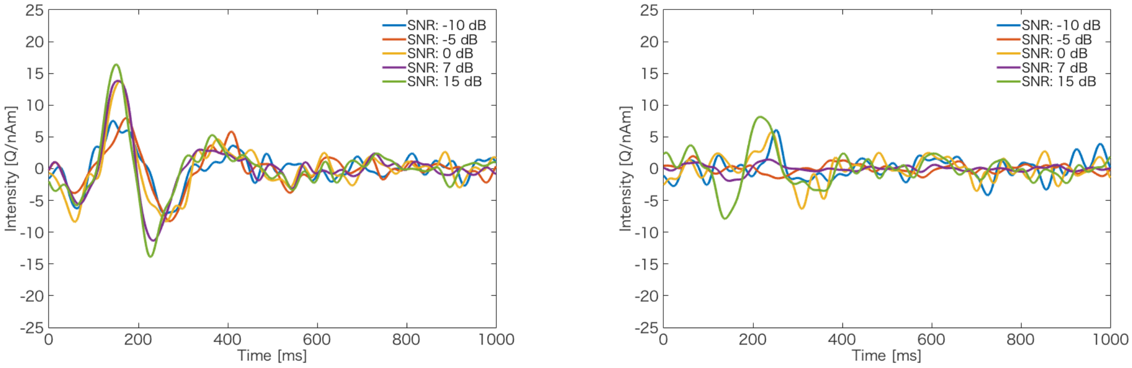


Fig. 1. *The relationship between SNR and articulation.*

MEG data were sampled at 400 Hz, and preprocessed by an analog band-pass filter between 0.03-100 Hz. Since magnetic signals exceeding 3,000 fT/cm were considered as artifacts, the epochs coinciding with it were rejected from any further analysis. In each SNR, the evoked responses for each syllable were added 70 times or more, and a digital band pass filter of 2 - 30 Hz was applied to the data obtained from the added average data.

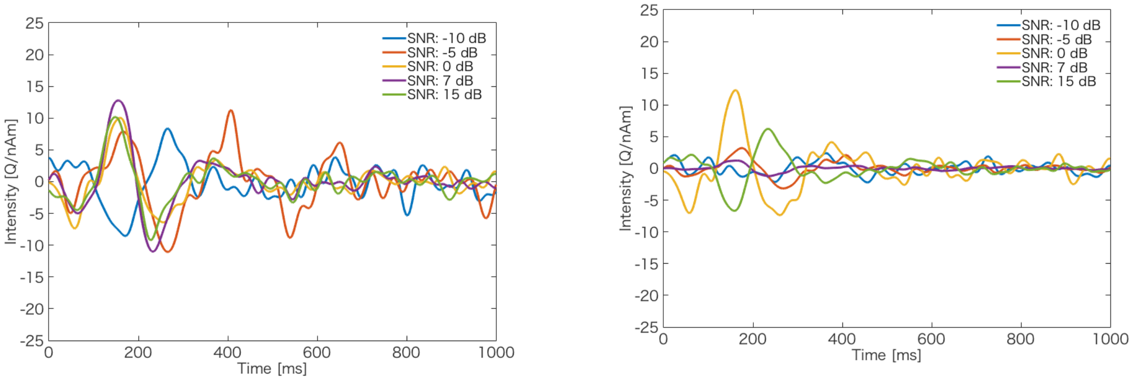
### B. *Data analysis*

Initially, the source of the magnetic fields was modeled as equivalent current dipoles (ECDs). Single dipoles were estimated in the cortex areas that are said to be associated with speech perception: (i) the auditory cortex, (ii) the pars opercularis and the pars triangularis in/around the inferior frontal gyrus (Broca's area), (iii) the posterior superior temporal gyrus (Wernicke's area), the angular gyrus and the supramarginal gyrus, and (iv) the motor and premotor cortex. The number of channels used in the source localization were (i) 18, (ii) 14 , (iii) 16, (iv) 16, respectively (it was felt that this number of channels would cover each area). Dipoles were estimated using the data from these channels taken every 2.5 ms from the onset of the stimulus sound presentation to 1,000 ms. Two criteria, "goodness of fit" and "confidence volume", were used for screening the estimated results. Only ECDs with a confidence volume $<1$ cm$^3$ and the highest goodness-of-fit value were adopted as a representative source. Based on the source, the direction and coordinate were determined, and only the amplitude was estimated [9]. After obtaining the time series of the source activity, the relationship between SNR and the peak latency were investigated and also the relationship between SNR and the peak value, which is larger than the average (including standard error) was also investigated.



(*a*) *The intensity of source localization in/around the auditory cortex in the left hemisphere.*

(*b*) *The intensity of source localization in/around Broca's area in the left hemisphere.*



(*c*) *The intensity of source localization in/around Wernicke's area in the left hemisphere.*

(*d*) *The intensity of source localization in/around the motor and the premotor cortex in the left hemisphere.*

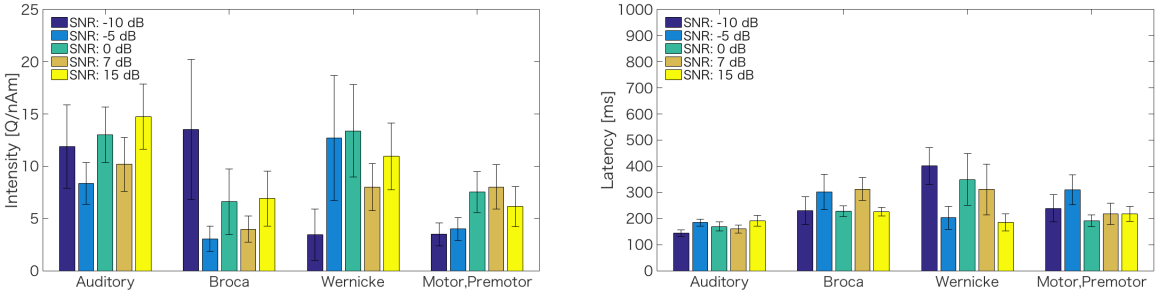
Fig. 2. *Time series of the source intensity.*

### C. *Results*

Fig. 2 shows the time series of the source intensity in each area (i)∼(iv) for one participant. Fig. 3 shows the relationship between SNR and the peak intensity and latency of source activity across all participants. Two-way ANOVA was carried out using the design cortex areas and SNR conditions. The peak intensity in/around the auditory cortex was larger ($p < 0.05$) than that in/around the Broca's area and the motor and premotor cortex in the left hemisphere. The peak latency in/around the auditory cortex was earlier than that in/around the other three cortex areas. There was no significant difference, with SNR as the main effect, in peak intensity and latency. In the right hemisphere, the peak intensity in/around the auditory cortex was larger than that in/around the inferior frontal gyrus, and the peak intensity in/around the motor and premotor cortex was smaller than that in/around the other three cortex areas. The peak latency in/around the auditory cortex was earlier than that in/around the posterior superior temporal gyrus and the motor/premotor cortex. There was a marginal difference ($p = 0.054$) between the auditory cortex and the inferior frontal gyrus. The peak latency in/around the posterior superior temporal gyrus was significantly different from the SNR conditions. Fig. 4 shows the relationship between the peak latency and articulation in/around the posterior superior temporal gyrus.

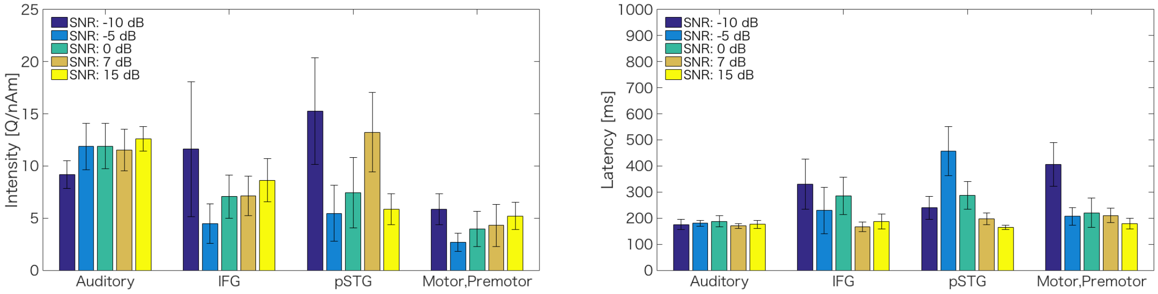
## IV. DISCUSSION

The response with the large peak was observed between 100 ms and 200 ms in the auditory cortex. The components observed at a latency of 100 ms were considered to be the N1m response because there was no difference among any SNR conditions. The peak intensity in/around the Wernicke's area was large as well as the one in the auditory cortex. On the other hand, the peak latency is later than auditory cortex's. It is possible that the speech perception process transfers from the auditory cortex to Wernicke's area. The peak latency of the Broca's area and the motor and premotor cortex were later than the auditory cortex, but it is hard to say if the activity of the Wernicke's area transfers to these areas. In the right hemisphere, it is also possible that there is a transition from the auditory cortex to other areas but there no tendency to transfer among these areas. In addition, the activity in the motor and premotor cortex was smallest among these areas so it is thought that participation in speech perception and understanding in this area is relatively weak. No significant difference among the peak intensities when changing the SNR could be found, but a significant difference among the latencies when changing the SNR could be found in the posterior superior temporal gyrus in the right hemisphere. According to the relationship between articulation and the peak latency of source activity, the



(*a*) *The peak intensity of brain activity in the left hemisphere.*

(*b*) *The peak latency of brain activity in the left hemisphere.*



(*c*) *The peak intensity of brain activity in the right hemisphere.*

(*d*) *The peak latency of brain activity in the right hemisphere.*

Fig. 3. *The relationship between SNR and the features of the source localization in both hemispheres. IFG means inferior frontal gyrus and pSTG means posterior superior temporal gyrus.*

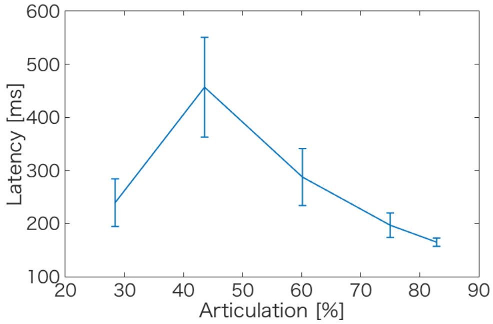


Fig. 4. *The relationship between articulation and latency in the posterior superior temporal gyrus in the right hemisphere.*

latency is latest when the SNR is -5 dB, and the tendency that the latency becomes earlier as the SNR improves can be confirmed. From this fact, it is conceivable that the response may be delayed when the noise is large to some extent and a load to identify the speech sounds is applied. When the SNR is -10 dB, the latency is earlier, but this is due to too much noise and the fact that the load for identifying the speech sounds decreases.

From this result, the systematic changes between the articulation and peak latency of source activity in the posterior superior temporal gyrus in the right hemisphere can be suggested, but it was not observed in the left hemisphere, which is said to govern the speech center. The stream from the posterior superior temporal gyrus to the premotor cortex could not be seen, so more advanced analysis or an experiment method that reflects the articulation more clearly may be needed to clarify the relationship between articulation of speech perception and the corresponding brain responses.

## REFERENCES

[1] R. Hari *et al.*, Auditory Evoked Transient and Sustained Magnetic Fields of the Human Brain, Experimental Brain Research, 40, 237-240, 1980.

[2] M. S. Gazzaniga *et al.*, "Cognitive Neuroscience: The Biology of the mind (4th Ed)," W. W. Norton & Company, 2014.

[3] F. Pulvermüller *et al.*, Motor Cortex Maps Articulatory Features of Speech Sounds, PNAS, 103, 7865-8770, 2006.

[4] K. Watkins *et al.*, Modulation of Motor Excitability During Speech Perception: The Role of Broca's Area, Journal of Cognitive Neuroscience, 16, 978-987, 2004.

[5] G. Hickok *et al.*, The Cortical Organization of Speech Processing, Nature Neuroscience, 8, 393-402, 2007.

[6] J. R. Rauschecker *et al.*, Maps and Streams in the Auditory Cortex: Nonhuman Primates Illuminate Human Speech Processing, Nature Neuroscience, 12 (6), 718-724, 2009.

[7] P. C. Wang *et al.*, Cortical Mechanisms of Speech Perception in Noise, Journal of Speech, Language, and Hearing Research, 51, 1026-1041, 2008.

[8] Y. Nagatani *et al.*, Loudness calibration of monosyllabic speech sounds in FW03, The Journal of the Acoustical Society of Japan, 64, 647-649, 2008.

[9] N. Nishitani and R. Hari, Temporal Dynamics of Cortical Representation for Action, PANS, 97, 913-918, 2000.