Auris Nasus Larynx 47 (2020) 727–733

Contents lists available at ScienceDirect

Auris Nasus Larynx

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

Central processing of speech sounds and non-speech sounds with similar spectral distribution: An auditory evoked potential study

Shinsuke Kaneshiro, Harukazu Hiraumi*, Hiroaki Sato

Department of Otolaryngology - Head and Neck Surgery, Iwate Medical University, 19-1, Uchimaru, Morioka, Iwate, Japan

ARTICLE INFO

Article history:
Received 13 September 2019
Accepted 9 February 2020
Available online 24 February 2020

Keywords:
Speech
Spectrum
Vowel
Auditory evoked potential
P2

ABSTRACT

Objective: The purpose of this study was to measure the auditory evoked potentials for speech and non-speech sounds with similar spectral distributions.

Methods: We developed two types of sounds, comprising naturally spoken vowels (natural speech sounds) and complex synthesized sounds (synthesized sounds). Natural speech sounds consisted of 5 Japanese vowels. Synthesized sounds consisted of a fundamental frequency and its second to fifteenth harmonics equivalent to those of natural speech sounds. The synthesized sound was filtered to have a similar spectral distribution to that of each natural speech sound. These sounds were low-pass filtered at 2000 Hz. The auditory evoked potential elicited by the natural speech sound /o/ and synthesized counterpart for /o/ were measured in 10 right-handed healthy adults with normal hearing.

Results: The natural speech sounds were significantly highly recognized as speech compared to the synthesized sounds (74.4% v.s. 13.8%, $p < 0.01$). The natural speech and synthesized sounds for the vowel /o/ contrasted strongly for speech perception (96.9% vs. 9.4%, $p < 0.01$). However, the vowel /i/ and its counterpart were barely recognized as speech (4.7 v.s. 3.1%, $p = 1.00$). The N1 peak amplitudes and latencies evoked by the natural speech sound /o/ were not different from those evoked by the synthesized sound ($p = 0.58$ and $p = 0.28$, respectively). The P2 amplitudes evoked by the natural speech sound /o/ were not different from those evoked by the synthesized sound ($p = 0.51$). The P2 latencies evoked by the natural speech sound /o/ were significantly shorter than those evoked by the synthesized sound ($p < 0.01$). This modulation was not observed in a control study using the vowel /i/ and its counterpart ($p = 0.29$).

Conclusion: The early P2 observed may reflect central auditory processing of the 'speechness' of complex sounds.

1. Introduction

Hearing loss is the most frequent disability worldwide. Hearing loss is associated with communication difficulties, cognitive disorders, and depression and is recognized as a serious social problem. Recent progress with cochlear implants has enabled patients with hearing loss to sense sounds and understand speech during speech recognition tests [1]. How-

* Corresponding author: Harukazu Hiraumi, Department of Otolaryngology - Head and Neck Surgery, Iwate Medical University, 19-1, Uchimaru, Morioka, Iwate, 020-8505, Japan.
E-mail addresses: hhiraumi@iwate-med.ac.jp, hhiraumi@ent.kuhp.kyoto-u.ac.jp (H. Hiraumi).

https://doi.org/10.1016/j.anl.2020.02.008

ever, even with the improvements, the cochlear implant recipients still encounter problems in challenging auditory conditions. For example, these patients experience difficulty in perceiving unexpected speech. This is partially due to a limited understanding of the central mechanisms that differentiate speech from non-speech sounds without attention. Previous studies have suggested that speech sounds are recognized in a special way called speech mode [2]. Once speech mode is engaged, complex sounds tend to be recognized as speech. Instructing a listener can help to evoke speech mode. Even without instruction, certain acoustic properties of speech, or the 'speechness' of sounds, can trigger speech mode [2]. The coding of this 'speechness' in the cochlear implants is supposed to contribute to the better speech recognition in unexpected conditions. However, little is understood which auditory characteristics contribute to the 'speechness' and how the 'speechness' is reflected in the central auditory systems. Only the psychoacoustic experiments are not enough to reveal the 'speechness', since the attention modulate the recognition of sounds as speech. To develop a speech coding strategy that processes the 'speechness' in the cochlear implant recipients, an objective evaluation of 'speechness' is needed.

Several previous studies have reported on the central processing of speech sounds. In a review of positron emission tomography and functional magnetic resonance imaging studies, speech-selective auditory responses included associations of the left posterior superior temporal cortex with sound familiarity, left anterior superior temporal gyrus with speech complexity, and left inferior frontal and premotor areas with auditory categorization and phonological discrimination tasks [3]. However, this review did not indicate when functional associations occurred following sound presentation. Magnetoencephalography (MEG) has also been used to detect differential processing of speech and non-speech sounds [4–7]. MEG studies have measured the latencies and the amplitude of the auditory evoked responses. After the auditory stimuli, several deflections are observed. Deflections peaking around 100 ms after the onset of a sound stimulus were called N1, and the deflections peaking around 200 ms were called P2. These studies reported that the N1 latency evoked by a monosyllabic speech sound stimulus is longer than that evoked by a pure tone stimulus [6,5]. Diesch and Luce compared auditory evoked responses to monosyllabic speech sound stimuli and synthesized sound stimuli composed of two pure tones corresponding to the first and second formant. These authors found that the N1 latencies for speech sound stimuli were longer than those for synthesized sound stimuli [7]. However, the prolonged N1 latencies observed were not necessarily caused by the 'speechness' of the stimuli. The human cochlea divides sound signals into frequency bands. Therefore, it is possible that sound signals covering a wide frequency range undergo different peripheral modulation compared to pure tones. Indeed, sound complexity has been reported to affect the latency of the auditory evoked response [8,9]. This problem can be more serious when evaluating cochlear implant recipients since cochlear implants divide sound into multiple frequency bands, which are independently modulated. Therefore, to understand the central processing of 'speechness' and to import the 'speechness' into the cochlear implant recipients, non-speech sounds with similar spectral distribution to speech sounds should be examined. Moreover, the imaging modality used should be applicable to patients with cochlear implants containing magnets. Electroencephalography (EEG) is appropriate for patients with cochlear implants, and a commercially accessible EEG device compatible with cochlear implants has recently been developed [10,11].

In the present study, we developed non-speech sounds with similar spectral distributions to speech sounds, which were not recognized as speech, and measured the auditory evoked potentials (AEPs) evoked by speech and non-speech sound stimuli.

Table 1
The first, second and third formants of five Japanese vowels (/a/, /e/, /i/, /o/ and /u/).

	/a/	/e/	/i/	/o/	/u/
1st formant (Hz)	964	515	324	494	354
2nd formant (Hz)	1312	2194	2429	678	1244
3rd formant (Hz)	2654	2681	3305	2970	2344

2. Material and methods

All the participated in the present study gave informed written consent, which was approved by the Ethics Committee of Iwate Medical University (Protocol No. H29-25), in accordance with the Declaration of Helsinki.

2.1. Stimuli

We developed two types of sounds, comprising naturally spoken vowels (natural speech sounds) and complex synthesized sounds (synthesized sounds). Natural speech sounds consisted of five Japanese vowels (/a/, /e/, /i/, /o/, and /u/) spoken by a Japanese male professional announcer. He was asked to pronounce at a constant pitch and similar length. The formants of these vowels are shown in Table 1. The average fundamental frequency of the five vowels was 170 Hz. The duration was between 210 and 220 ms. Synthesized sounds consisted of the fundamental frequency (170 Hz) and its second to fifteenth harmonics. Each synthesized sound was filtered to have a similar spectral distribution to those of the steady-state component of each natural speech sound. The rise-fall pattern was also matched to that of each natural speech sound. All natural speech and synthesized sounds were matched for equal loudness. Finally, natural speech and synthesized sounds were low-pass filtered with a cut-off frequency of 2000 Hz, since our pilot psychoacoustic study suggested that this filtering most enhanced the different recognition of the two kinds of stimuli. The waveforms and spectral distributions are shown in Fig. 1. The amplitude and the frequency of the natural speech sound are less constant than those of the synthesized sound, which are thought to contribute to the 'speechness' [12]. Following filtering, vowel /i/ completely lost the second formant and the second formant of vowel /e/ was reduced. Vowels /a/, /o/, and /u/ lost the third and higher formants;

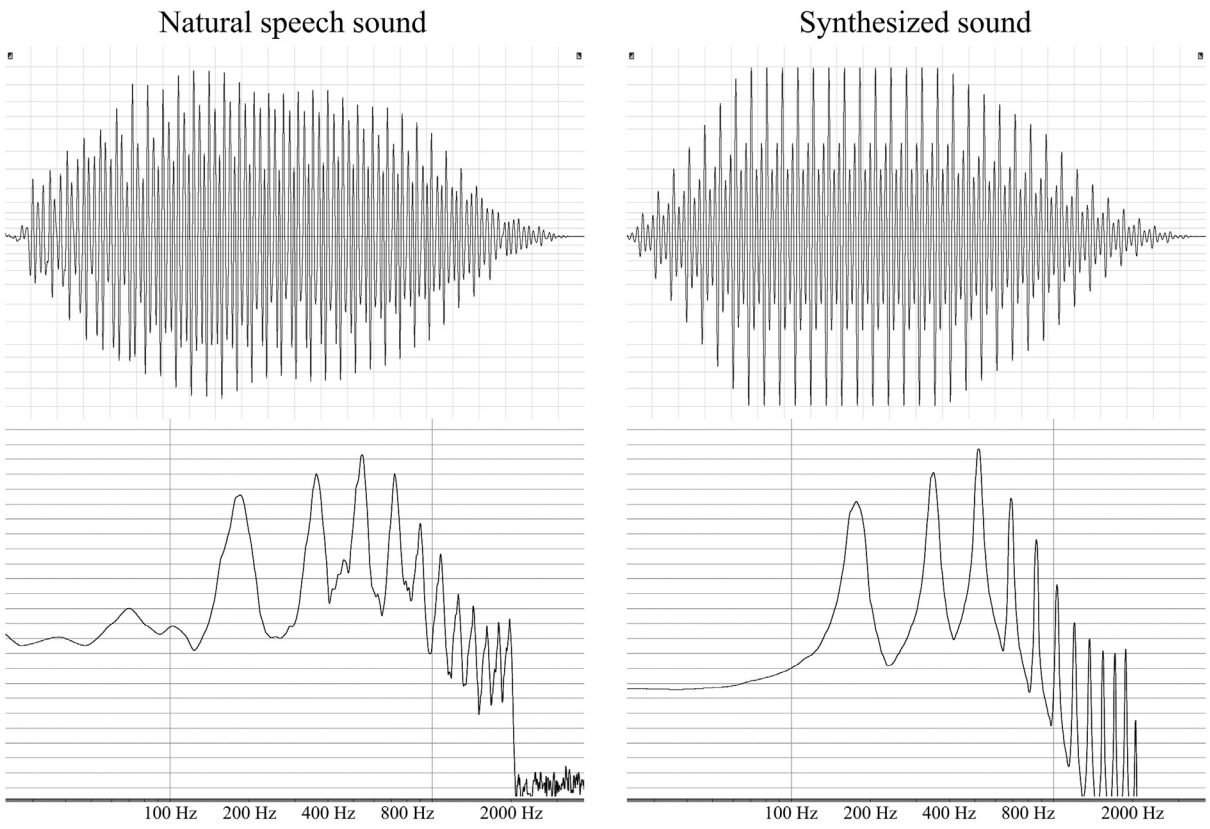

Fig. 1. The waveforms (top) and frequency spectra (bottom) of the natural speech sound /o/ (left) and synthesized sound /o/ (right) are shown. Note the spectral similarity between natural speech and synthesized sounds. The natural speech sound fluctuates in the frequency and in the amplitude.

however, the first and second formants were preserved. In total, 10 stimuli (i.e., five natural speech sounds and five synthesized sounds) were created. All of the sounds were prepared using Adobe Audition (Adobe, San Jose, CA).

2.2. Psychoacoustic experiment

Eight right-handed healthy adults with normal hearing (four men and four women; aged 23–28 years) participated in this study. No subjects had a history of hearing loss or neurological disorders. In this experiment, we used a two-alternative forced-choice task. The subjects sat in an upright position in an acoustically shielded room. The 10 stimuli (i.e., five natural speech sounds and five synthesized sounds) were randomly presented with an inter-stimulus interval of 3500 ms. Each stimulus was presented eight times. In total, 80 stimuli were presented in two blocks. The sounds were delivered binaurally through a pair of headphones (MDR-Z900; Sony, Tokyo, Japan) at the most comfortable level for each subject. The subjects were instructed to answer whether each stimulus was a speech or non-speech sound by writing on an answer sheet. No further information was given to the subjects. The frequency of stimuli perceived as a speech was calculated in each sound independently. The obtained data were analyzed with Fisher's exact test using IBM SPSS Statistics software (version 22; IBM Corporation, Armonk, NY). A p-value <0.05 was set as the level of statistical significance.

2.3. EEG experiment

Ten right-handed healthy adults with normal hearing (seven men and three women; aged 23–47 years) participated in this study. No subjects had a history of hearing or neurological disorders. None of these subjects participated in the psychoacoustic experiment, since our pilot study suggested that the participants to the psychoacoustic experiment tended to concentrate on the stimulus sound during the EEG experiment. In this experiment, the natural speech sound /o/ and synthesized counterpart for /o/ (i.e., synthesized sound /o/) were presented as stimuli. Subjects sat in an upright position with a backrest in an acoustically and electrically shielded room, and were alternately presented with natural speech and synthesized sounds at an inter-stimulus interval of 2000 ms. The stimuli were presented binaurally through a pair of headphones (MDR-Z900) at the most comfortable level for each subject. The subjects watched a silent movie of their choice and were instructed to ignore the presented sounds. The AEPs were recorded using a MEB-9400 system (Nihon Kohden Corporation, Tokyo, Japan) externally triggered with Presentation software (Neurobehavioral Systems, Inc., Albany, CA). The recording bandpass was 0.1–20 Hz, and the sampling rate was 2000 Hz. Over 100 responses to each stimulus were collected and averaged online. Epochs exceeding 150 uV were excluded from online averaging. The period of analysis was 1000 ms, including a pre-stimulus baseline of 100 ms. The AEP was recorded at the vertex (Cz) with a reference electrode positioned on the left mastoid and ground electrode

Table 2
The percentage of stimuli perceived as speech.

	/a/	/e/	/i/	/o/	/u/	all
Natural speech sound (%)	98.4	79.7	4.7	96.9	92.2	74.4
Synthesized sound (%)	54.7	25.0	3.1	9.4	21.9	13.8

Except for /i/, natural speech sounds were significantly and frequently recognized as speech ($p < 0.001$, Fisher's exact test). Note that the natural speech and synthesized sounds for the vowel /o/ contrasted strongly for speech perception. All stimuli were low-pass filtered at 2000 Hz.

positioned on the nose using nonpolarizable Ag/AgCl electrodes. The impedance value of the electrodes was less than 5.0 kOhm. The electrode placement used is that which is minimally required to measure the N1-P2 complex and is commonly used in patients with cochlear implants [11,13]. The negative deflections peaking between 75 and 125 ms were defined as N1, and the positive deflections peaking between 150 and 275 ms after the onset of sound stimulus were defined as P2. The peak amplitude and the latencies of these responses were calculated. Differences between the response latencies and amplitudes of natural speech and synthesized sounds were examined with Wilcoxon signed-rank test using IBM SPSS Statistics software (version 22; IBM Corporation, Armonk, NY). A p-value <0.05 was set as the level of statistical significance.

As a control experiment, the AEPs evoked by the natural speech sound /i/ and synthesized sound /i/ were recorded in 10 right-handed healthy adults with normal hearing (eight men and two women; aged 23–36 years). The acquisition parameters and the conditions are identical with those described above.

3. Results

3.1. Psychoacoustic experiment

All of the subjects tolerated and completed the experiment. Table 2 shows the results of the psychoacoustic experiment. Overall, natural speech sounds were significantly and highly recognized as speech and synthesized sounds were recognized as non-speech ($p < 0.001$, Fisher's exact test). Except for the vowel /i/ and its synthesized counterpart, natural speech sounds were almost always recognized as speech (79.7–98.4%) and synthesized sounds tended to be recognized as non-speech (9.4–54.7%). The difference between each vowel and its counterpart was statistically significant ($p < 0.001$, Fisher's exact test). However, the vowel /i/ and its counterpart were barely recognized as speech (4.7 and 3.1%, respectively), and there was no significant difference between them ($p = 1.00$, Fisher's exact test). The natural speech and synthesized sounds for the vowel /o/ contrasted strongly for speech perception. Consequently, these stimuli were used in the EEG experiment. In addition, we adopted the natural speech and synthesized sounds for the vowel /i/ as control, which were recognized similarly.

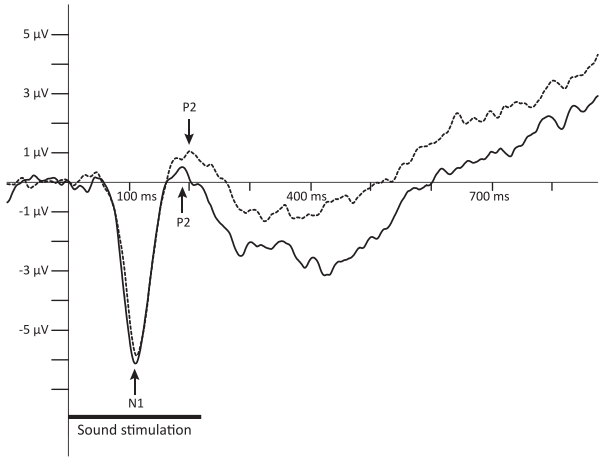

Fig. 2. The grand-averaged waveform ($n = 10$ subjects) evoked by natural speech (solid line) and synthesized (dashed line) sounds /o/.

Table 3
Mean and standard error of the mean of N1 and P2 peak latencies and amplitudes evoked by natural speech /o/ and synthesized sounds /o/.

		Natural speech sound /o/	Synthesized sound /o/	p-value
Latency (ms)	N1	111.5 ± 1.4	113.9 ± 1.9	0.284
	P2	185.4 ± 6.8	220.7 ± 0.8	0.007*
Amplitude (uV)	N1	−6.19 ± 0.90	−5.98 ± 0.90	0.575
	P2	1.34 ± 0.53	2.06 ± 0.75	0.508

The P2 latency evoked by natural speech sound /o/ is significantly shorter than that evoked by synthesized sound /o/.
* Wilcoxon signed-rank test.

3.2. EEG experiment

Clearly identifiable N1 and P2 responses were obtained for all subjects. Fig. 2 shows a grand averaged waveform among all subjects. The peak amplitudes and latencies of all subjects are summarized in Table 3. The N1 peak amplitudes and latencies evoked by natural speech sounds /o/ were not different from those evoked by synthesized sounds /o/ ($p = 0.575$ and $p = 0.284$, respectively, Wilcoxon signed-rank test). The P2 amplitudes evoked by natural speech sounds /o/ were not different from those evoked by synthesized sounds /o/ ($p = 0.508$, Wilcoxon signed-rank test). The P2 latencies evoked by natural speech sounds /o/ were significantly shorter than those evoked by synthesized sounds /o/ ($p = 0.007$, Wilcoxon signed-rank test). The N1 and P2 peak latencies obtained for each subject are shown in Fig. 3. The inter-peak latency and amplitude between N1 and P2 had similar results to those from the P2 analysis. The amplitude difference between N1 and P2 was not statistically significant ($p = 0.333$, Wilcoxon signed-rank test). The inter-peak latency between N1 and P2 was significantly shorter for natural speech sounds /o/ than those for synthesized sounds /o/ ($p = 0.007$, Wilcoxon signed-rank test).

In the control study, the N1 peak amplitudes and latencies evoked by natural speech sounds /i/ were not different from those evoked by synthesized sounds /i/ ($p = 0.445$ and $p = 0.683$, respectively, Wilcoxon signed-rank test). The P2

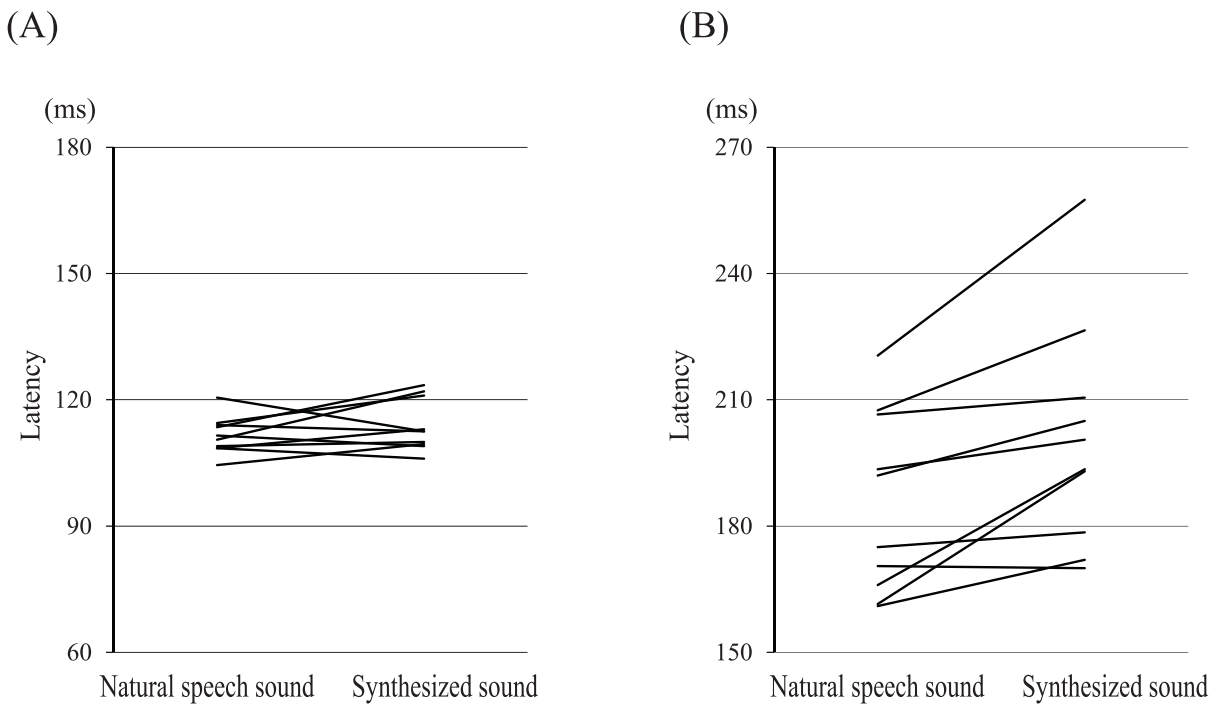

Fig. 3. The N1 (A) and P2 (B) latencies for each subject are shown. The N1 latencies are not different between natural speech and synthesized sounds /o/. The P2 latencies evoked by natural speech /o/ were significantly shorter than those evoked by synthesized sound /o/.

Table 4
Mean and standard error of the mean of N1 and P2 peak latencies and amplitudes evoked by low-pass filtered spoken vowel /i/ and synthesized sounds with similar spectral distribution.

		Natural speech sound /i/	Synthesized sound /i/	*p*-value
Latency (ms)	N1	119.8 ± 2.4	120.9 ± 3.6	0.683
	P2	184.4 ± 6.8	188.8 ± 7.4	0.285
Amplitude (uV)	N1	−4.77 ± 0.80	−5.47 ± 1.08	0.445
	P2	1.62 ± 0.97	1.40 ± 0.62	0.445

amplitudes and latencies evoked by natural speech sounds /i/ were not different from those evoked by synthesized sounds /i/ ($p = 0.445$ and $p = 0.285$, respectively, Wilcoxon signed-rank test). The results are summarized in Table 4.

The modification of latency (time difference between the AEPs evoked by natural speech sounds and synthesized sounds) were compared between /o/ and /i/. The modulation of P2 latency evoked by /o/ was significantly larger than that evoked by /i/ ($p = 0.035$, Mann-Whitney U test).

4. Discussion

In this study, we found that the P2 latency evoked by natural speech sounds /o/ were significantly shorter than that evoked by synthesized sounds /o/. This modulation of P2 latency was not observed in the control experiment using natural speech stimuli /i/ and the synthesized counterpart, which were both barely recognized as speech. The N1 latencies and amplitudes evoked by speech and non-speech sounds were not different.

The central processing of sound 'speechness' has been explored using EEG and MEG. Since stimulus complexity affects the auditory cortical response [9], non-speech stimuli with the same complexity as speech stimuli are needed to differentiate speech and non-speech processing. However, this is a real challenge [14]. Diesch and Luce used multiple tones with frequencies corresponding to the formants [7], however, such sounds lack harmonic components and do not have the same complexity as speech. Spectrally rotated signals, which preserve the temporal and spectral complexity of speech, have also been used as speech stimuli counterparts [14]. However, with these signals, the spectral distribution is different from that of the original speech stimuli. In the present study, we developed multiple tones with harmonic components and speech-like spectral distributions as non-speech stimuli. In our pilot study, these multiple tones were recognized as speech since the subjects tended to listen out for speech sound. To eliminate the 'speechness' of these signals, we applied a low-pass filter at 2000 Hz informed by a prior pilot study. Our psychoacoustic experiment showed that low-pass filtered natural speech sounds were almost always recognized as speech. In contrast, synthesized sounds, which lack fluctuation both in the amplitude and in the frequency, were barely recognized as speech. The acoustic similarity of the natural speech and synthesized sounds was partially proven by the EEG experiment. The N1 changes according to the physical property of the sound. The latency and amplitude of the N1 component evoked by the speech and non-speech stimuli were completely identical, meaning that the properties of the two stimuli were very close. The N1 has multiple generators in primary and secondary auditory cortex [13]. Our result suggests that the natural speech and synthesized sounds

undergo similar processing at the level of the primary auditory cortex. Despite their acoustic similarity, the two sounds were recognized differently, which suggests that these sound stimuli can be used to differentiate speech and non-speech processing.

In the EEG experiment, we found that the P2 latency evoked by natural speech /o/ was significantly shorter than that evoked by synthesized sound /o/. The sound stimuli used in the control experiment were developed in the same manner with the natural speech /o/ and synthesized counterpart. In the control experiment, the modulation of P2 latency was not observed, meaning that the origin of short P2 latency can be attributed to the 'speechness' of the natural speech. P2 is thought to have multiple generators located in multiple auditory areas but the significance of P2 is still not understood [13]. One hypothesis for the short P2 latency evoked by the natural speech sound is the activation of some neurons specific to the 'speechness.' In a study using rhesus monkeys, Rauschecker and his colleagues reported that neurons surrounding primary auditory cortex showed greater responses to the complex sounds than to the pure tones. Many of those neurons showed a preference for species-specific communication calls over others [15]. Similar preference to the species-specific vocalization was reported in other animals [16,17]. Wang and Kadia recorded neuronal responses evoked by natural and time-reversed marmoset vocalization in the primary auditory cortex of the marmosets and cats. The primary auditory cortices of marmoset showed higher firing rate to the natural vocalization, whereas the primary auditory cortices of cat did not show preference to the natural vocalization [18]. It is possible that the human also have similar neurons preferring human vocalization in or around the primary auditory cortex, and recruitment of such neurons resulted in the early P2 evoked by the natural speech sound.

Since P2 latency and amplitude often change with N1 latency and amplitude, little are known about the independent modulation of P2 latency. In the present study, N1 did not differ between speech and non-speech stimuli but the P2 latencies were different. Some papers have reported prolonged P2 latency without N1 modulation [19–21]. Du et al. measured the auditory evoked potential in adolescents with the attention deficit/hyperactivity disorder (ADHD) and the conduct disorders (CD). In subjects with ADHD + CD, the latencies of P2 and the later components were prolonged, whereas the latency of N1 was comparable to that of the control group. Since the attention happens where stimulus attracts our mental consideration [21], the prolonged P2 latency may be related with the deficits during automatic processing of passive auditory stimuli. Huang et al. measured cortical potentials evoked by a break in the correlation between a direct sound and a delayed sound. The break in the correlation evoked an N1 and a P2. An increase of the delay resulted in the increase of the P2 latency but not an increase of the N1 latency. The authors speculated that the P2 is closely related to listeners' perception of the fusion of direct wave and its reflections [19], suggesting that the natural sound evokes P2 with short latency. Interestingly, older subjects have reported to show delayed P2 evoked by speech [20]. Tremblay and his colleagues recorded AEPs in younger and older subjects with normal hearing. The older subjects elicited prolonged P2 latencies in response to /ba/ (0 ms voice onset time), whereas their N1 latencies were comparable to those of younger subjects [20]. In a pure tone stimulus study, no age-related changes were observed for N1 and P2 latencies [22]. These findings suggest that the P2 delay observed in older subjects is attributable to age-related modulation of central speech processing. Moreover, older subjects have deteriorated temporal resolution. One of the differences between the natural speech and synthesized sounds used in the present study was temporal fluctuation. A small temporal fluctuation of sound (i.e., micromodulation) contributes to 'speechness' [12]. It can be hypothesized that older subjects with poor temporal resolution were not able to detect the temporal fluctuation of speech, which resulted in a prolonged P2 latency. To prove this hypothesis, further studies comparing younger and older subjects are needed.

In the present study, the amplitude of the AEPs did not differ between speech and non-speech stimuli. In the previous studies reporting prolonged P2 latency without N1 modulation, the amplitude of P2 did not changed [20,21] or changed along with N1 amplitude [19]. Since recruitment of additional cortical response is supposed to change the amplitude of the AEPs, the modulated P2 latency may be derived from the engagement of different neural pathway, and not from the recruitment of additional cortical activation. The previous psychoacoustic experiment hypothesizes that the 'speechness' of sounds switch on a distinct neural pathway called speech mode [2]. It is possible that the short P2 latency is the consequence of the activation of neural pathway specific to the speech mode.

This study has some limitations. One limitation is that we used low-pass filtered vowels. Complex sounds within a speech-like spectrum are easily perceived as speech when the subjects are instructed or intentionally trying to recognize a sound as speech. To minimize the 'speechness' of the stimuli, we applied a low-pass filter at 2000 Hz. Low-pass filtering reduces the naturalness of speech, while ensuring its intelligibility [14], which suggests that our natural speech sounds may not be natural. To reduce the effect of the low-pass filter, we only used vowels since high-frequency components are especially important for consonants. In our psychoacoustic experiment, most of the subjects perceived the low-pass filtered vowels as speech. Thus, we believe the naturalness of the stimuli was preserved. The other limitation is that we only measured the AEPs at the Cz. Simultaneous recording from multiple electrode locations is recommended to disentangle overlapping event-related potential components [23]. Our final goal is to develop a clinically available tool for measuring the 'speechness' of sounds. Therefore, we adopted settings from a commercially available AEP recording system, which can measure the cortical response evoked by a short voice in cochlear implant recipients [11]. The two hemispheres are known to work differently in central processing of speech sounds. Further EEG study covering whole skull or MEG study may prove hemispheric differences in speech and non-speech perception.

5. Conclusion

The auditory P2 latency evoked by natural speech vowels /o/ was significantly shorter than that evoked by multiple tones with similar spectral distributions to the natural speech vowels /o/. The early P2 observed may reflect central auditory processing of the 'speechness' of complex sounds.

Declaration of Competing Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

Compliance with Ethical Standards

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent: Informed consent was obtained from all individual participants included in the study.

Funding

This work was supported by the JSPS KAKENHI(grant number JP17K11341)

References

[1] Hiraumi H, Tsuji J, Kanemaru S, Fujino K, Ito J. Cochlear implants in post-lingually deafened patients. Acta Otolaryngol Suppl 2007(557):17–21. doi:10.1080/03655230601065225.

[2] Moore BC. Speech perception. An introduction to the psychology of hearing. 6th edn. Leiden: Brill; 2012.

[3] Price CJ. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. Neuroimage 2012;62(2):816–47. doi:10.1016/j.neuroimage.2012.04.062.

[4] Roberts TP, Poeppel D. Latency of auditory evoked M100 as a function of tone frequency. Neuroreport 1996;7(6):1138–40.

[5] Tiitinen H, Sivonen P, Alku P, Virtanen J, Naatanen R. Electromagnetic recordings reveal latency differences in speech and tone processing in humans. Brain Res Cogn Brain Res 1999;8(3):355–63.

[6] Lin YY, Chen WT, Liao KK, Yeh TC, Wu ZA, Ho LT. Hemispheric balance in coding speech and non-speech sounds in Chinese participants. Neuroreport 2005;16(5):469–73.

[7] Diesch E, Luce T. Magnetic fields elicited by tones and vowel formants reveal tonotopy and nonlinear summation of cortical activation. Psychophysiology 1997;34(5):501–10.

[8] Cansino S, Ducorps A, Ragot R. Tonotopic cortical representation of periodic complex sounds. Hum Brain Mapp 2003;20(2):71–81. doi:10.1002/hbm.10132.

[9] Bardy F, Van Dun B, Dillon H. Bigger is better: increasing cortical auditory response amplitude via stimulus spectral complexity. Ear Hear 2015;36(6):677–87. doi:10.1097/AUD.0000000000000183.

[10] Legris E, Galvin J, Roux S, Gomot M, Aoustin JM, Marx M, et al. Cortical reorganization after cochlear implantation for adults with single-sided deafness. PLoS ONE 2018;13(9):e0204402. doi:10.1371/journal.pone.0204402.

[11] Kosaner J, Van Dun B, Yigit O, Gultekin M, Bayguzina S. Clinically recorded cortical auditory evoked potentials from paediatric cochlear implant users fitted with electrically elicited stapedius reflex thresholds. Int J Pediatr Otorhinolaryngol 2018;108:100–12. doi:10.1016/j.ijporl.2018.02.033.

[12] Bregman AS. Auditory Scene Analysis. Cambridge, MA: MIT Press; 1990.

[13] Martin BA, Tremblay KL, Korczak P. Speech evoked potentials: from the laboratory to the clinic. Ear Hear 2008;29(3):285–313. doi:10.1097/AUD.0b013e3181662c0e.

[14] Christmann CA, Berti S, Steinbrink C, Lachmann T. Differences in sensory processing of German vowels and physically matched non-speech sounds as revealed by the mismatch negativity (MMN) of the human event-related brain potential (ERP). Brain Lang 2014;136:8–18. doi:10.1016/j.bandl.2014.07.004.

[15] Rauschecker JP, Tian B, Hauser M. Processing of complex sounds in the macaque nonprimary auditory cortex. Science 1995;268(5207):111–14.

[16] Wang X, Merzenich MM, Beitel R, Schreiner CE. Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. J Neurophysiol 1995;74(6):2685–706. doi:10.1152/jn.1995.74.6.2685.

[17] Horpel SG, Firzlaff U. Processing of fast amplitude modulations in bat auditory cortex matches communication call-specific sound features. J Neurophysiol 2019;121(4):1501–12. doi:10.1152/jn.00748.2018.

[18] Wang X, Kadia SC. Differential representation of species-specific primate vocalizations in the auditory cortices of marmoset and cat. J Neurophysiol 2001;86(5):2616–20. doi:10.1152/jn.2001.86.5.2616.

[19] Huang Y, Lu H, Li L. Human scalp evoked potentials related to the fusion between a sound source and its simulated reflection. PLoS ONE 2019;14(1):e0209173. doi:10.1371/journal.pone.0209173.

[20] Tremblay KL, Piskosz M, Souza P. Aging alters the neural representation of speech cues. Neuroreport 2002;13(15):1865–70.

[21] Du J, Li JM, Wang Y, Jiang QJ, Livesley WJ, Jang KL, et al. Event-related potentials in adolescents with combined ADHD and CD disorder: a single stimulus paradigm. Brain Cognition 2006;60(1):70–5. doi:10.1016/j.bandc.2005.09.015.

[22] Amenedo E, Diaz F. Ageing-related changes in the processing of attended and unattended standard stimuli. Neuroreport 1999;10(11):2383–8.

[23] Picton TW, Bentin S, Berg P, Donchin E, Hillyard SA, Johnson R Jr, et al. Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. Psychophysiology 2000;37(2):127–52.