The Harris Science Review of Doshisha University, Vol. 57, No. 4 January 2017

Infrared Neural Stimulation Evokes Auditory Brain Stem Responses following the Acoustic Feature of Speech Sounds

Yuta TAMAI*, Yuka SHINPO*, Kensuke HORINOUCHI*, Shizuko HIRYU*, and Kohta I. KOBAYASI*

(Received October 20, 2016)

Compound action potentials can be evoked by irradiating the cochlea with an infrared laser. We investigated whether a pulsed infrared laser could produce speech perception using a human and a Mongolian gerbil (*Meriones unguiculatus*) as subjects. In a previous study, we developed a speech-encoding scheme for single-channel cochlear stimulation. Pulsed laser speech sound (PLS) was created following the scheme; a click-moderated speech sound (CMS) was used to simulate the perception of PLS. Click sounds and the pulsed laser were used to compose CMS and PLS, respectively, the pitch of which followed the first formant frequencies (F1) of an original speech sound. We presented CMS and PLS to the gerbil, and compound action potentials were recorded from the round window. The CMS was presented to a human subject, and frequency-following responses were recorded with a scalp electrode. Human and gerbil recordings showed clear harmonic structures corresponding to the F1 of the original speech, and the frequency spectra of the recordings resembled each other. These findings suggest that the pulsed laser can generate a similar compound action potential to click-modulated speech in the gerbil, opening the possibility of creating intelligible speech perception by irradiating the human cochlea with an infrared laser.

Key words : infrared laser, distorted speech sound, formant frequency, amplitude envelope

1. Introduction

Severe hearing loss or deafness prevents the translation of sound information into neural electrical activity; however a cochlear implant may allow the hearing-impaired to reconstruct a sense of hearing. A cochlear implant is a hearing aid that provides a sense of sound by stimulating cochlear nerves directly. However, cochlear implant surgery involves the insertion of an electrode into the cochlea; as such, invasive surgery is required for this procedure, with loss of the remaining hearing as a possible consequence.

A previous study revealed that action potentials could be evoked by irradiating neurons *in vivo* with an infrared laser [1)]. Because an infrared laser can stimulate nerves without contacting the tissue, such ‘optical’ stimulation has gained attention as a possible substitute for electrical stimulation. There are several previous reports regarding the application of an infrared laser for a hearing aid. Izzo and colleagues demonstrated that neural activity was generated by irradiating cochlear nerves with an infrared laser through the round window [2)]. Another study with cats assessed the influence of auditory perception created by irradiating the cochlea with an infrared laser [3)].

Our previous study indicated that a train of click sounds, the repetition rate of which followed the first

---

*Neuroethology and Bioengineering Laboratory, Department of Biomedical Information, Doshisha University, Kyo-tanabe, Kyoto, 610-0321
Telephone: +81-774-65-6439, E-mail: dmq1041@mail4.doshisha.ac.jp

formant frequency transition of a speech sound, was at least partially intelligible [4]; the sound was referred to as a click-modulated speech sound (CMS). Additionally, an extra-cochlear stimulating system may create a perception resembling a click sound, because the system evokes action potentials from all cochlear nerves simultaneously. Thus, CMS may be a simulated sound for an extra-cochlear stimulation scheme. Our previous study revealed that similar compound action potentials were observed by infrared laser irradiation through the round window as those by click sounds [4]. We expect that the laser stimulation was encoded in the same way as synthesizing CMS and could create a perception similar to CMS. The encoded laser stimulation is referred to as pulsed laser speech sound (PLS). A perception resembling a CMS could be created by presenting PLS to a human subject.

The purpose of this study was to assess the possibility that PLS would induce the same neural activity as speech does in the human auditory brain stem. Because the invasiveness of the laser stimulation was unclear, PLS was first assessed with Mongolian gerbils (*Meriones unguiculatus*), not humans. The evoked compound action potentials were compared to those from CMS to validate CMS as a simulation of PLS. Moreover, human neural activities evoked by CMS were recorded; the neural response was compared to that of the gerbil to evaluate how similarly (or differently) our auditory pathway processes PLS and CMS. Thus, we sought to understand how humans perceive PLS.

2. Materials and Methods

2.1 Subject

A Mongolian gerbil (78 g) and a native Japanese speaker were used as study subjects. The human subject passed a hearing screening at 25 dB HL at frequencies of 0.5, 1, 2, and 4 kHz.

2.2 Surgery

The Mongolian gerbil was anesthetized with ketamine (50 mg/kg) and xylazine (10 mg/kg). Half doses were repeated every 30 min as necessary. Head skin and muscles were removed and a hole was made in the skull. The tympanic bulla was exposed by incision from the shoulder to the jaw and two holes were made for the electrode and laser fiber pathway. An electrode was hooked onto the bony rim of the round window to record neural responses evoked by stimuli.

2.3 Stimuli

Click-modulated speech sound (CMS)

CMS simulates the perception that is evoked by single-channel stimulation of a cochlear nerve bundle. The sound is a click train, the pitch (repetition rate) of which follows the first formant center frequency of an original speech sound. The pulse width was 100 μs. More specifically, formant frequencies were extracted from the original sounds by linear predictive coding (LPC) and fast Fourier transforms (FFTs) at a 48-kHz sampling rate and 1024-point FFT length. LPC was calculated every 15 ms over 30-ms Hamming-windowed segments. All signal processing was performed using Matlab (MathWorks; Fig. 1). An example of an original

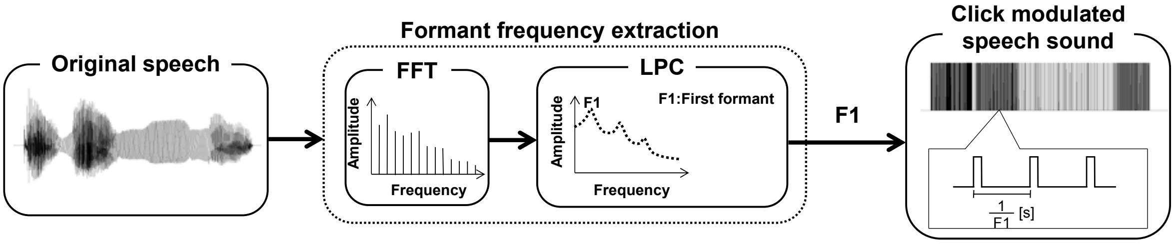

Fig. 1. Procedure for encoding click-modulated speech sound (CMS). The schematic diagram shows how to analyze the speech signal and synthesize the CMS.

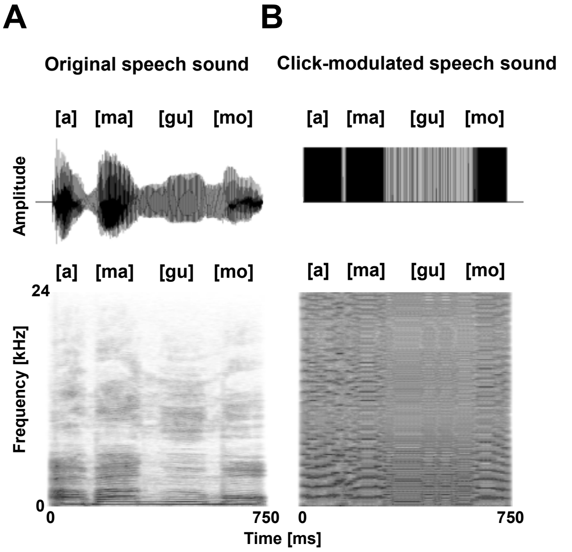

Fig. 2. An example of an original speech sound and a click-modulated speech sound. Upper figures show the waveform. Lower figures show spectrograms. The speech sound was processed by fast Fourier transform (FFT; sampling rate: 48,000 Hz, FFT length: 1024 points). (A) Original speech sound “[a], [ma], [gu], and [mo]”. (B) Click-modulated speech sound synthesized from the original speech sound.

speech sound and a click-modulated speech sound was shown in Fig.2. The original speech sound was four mora Japanese words (‘a’, ‘ma’, ‘gu’, and ‘mo’). The words were voiced by a female speaker.

Pulsed laser speech sound (PLS)

The stimulation used a pulsed laser, with the same repetition rate as for CMS. The wavelength of the pulsed laser was 1871 nm, and the pulse width was 100 μs.

2.4 Experimental environment

Mongolian gerbil

The experiment with the Mongolian gerbil was conducted in an acoustically and electrically shielded box. Neural responses were recorded from the electrode placed on the round window with body skin wet with saline as a reference. The acoustic stimulus was presented at 10 cm from the subject. Optical stimulation was provided to the cochlear nerves by irradiation through the round window.

Human

The experiment with the human subject was conducted in an acoustically and electrically shielded room. The subject closed his eyes and reclined in a chair. Neural responses were recorded between the electrode placed on the midline of the forehead and the seventh cervical vertebra (C7 location) using a sampling rate of 8000 Hz. A common ground was placed on the left mastoid. Impedances between the electrodes were calibrated below 3000 Ω (MaP811, Measurement and Processing). Stimuli were delivered at 40 cm from the subject.

2.5 Experimental procedure

Mongolian gerbil

Neural responses evoked by acoustic and optic stimulation were recorded. Acoustic stimuli were presented 100 times using a dome tweeter (FT28D, Fostex). The stimuli were calibrated at 80 dB SPL with a microphone (Type 1, ACO Pacific). Optic stimuli were presented 100 times with a diode laser stimulation system (BWF-OEM, B&W). The stimuli were calibrated at 20 μJ per laser pulse with an actinometer (Thorabs GmbH).

Human

Neural responses evoked by the acoustic stimulus were recorded. Acoustic stimuli were delivered 5000 times (EMC2.0-USB, Diamond Audio Technology). The stimuli were calibrated at 70 dB SPL (ER-7C Series B, Etymotic Research).

2.6 Electrophysiological analysis

Electrical signals from the Mongolian gerbil and human were amplified, 1,000- and 20,000-fold, respectively, with a low-cut filter (cut-off: 0.08 Hz; MEG-1200, Nihon Kohden). In the gerbil experiment, the signals were averaged 100 times. In the human experiment, the signals were averaged 5,000 times. The averaged signal was extracted per 8 ms and processed using a FFT at an 8000-Hz sampling rate and 64-point FFT length. The correlation coefficient was measured between the extracted spectra. All signal processing was performed using Matlab (MathWorks).

3. Results

Figure 3A shows a CMS waveform recorded with a microphone. Figures 3B and 3C show the compound action potentials of cochlear nerves evoked by CMS and by PLS in the Mongolian gerbil, respectively, and Fig. 3D shows the compound action potentials from the auditory brainstem evoked by CMS in the human. These neural activities (Figs. 3B–D) all resembled the CMS signal (Fig. 3A) with respect to both waveform and spectrum, while the amplitude of the compound action potentials evoked by CMS in the Mongolian gerbil (Fig.3B) was larger than that in the human auditory brainstem (Fig. 3D).

The positive peak in the cross-correlation coefficient between the waveform of the compound action potentials of the Mongolian gerbil's cochlear nerves evoked by CMS (Fig. 3B) and that by PLS (Fig. 3C) was measured at a 250-μs time lag. The peak in cross-correlation between the waveform of neural activities in the Mongolian gerbil evoked by CMS and

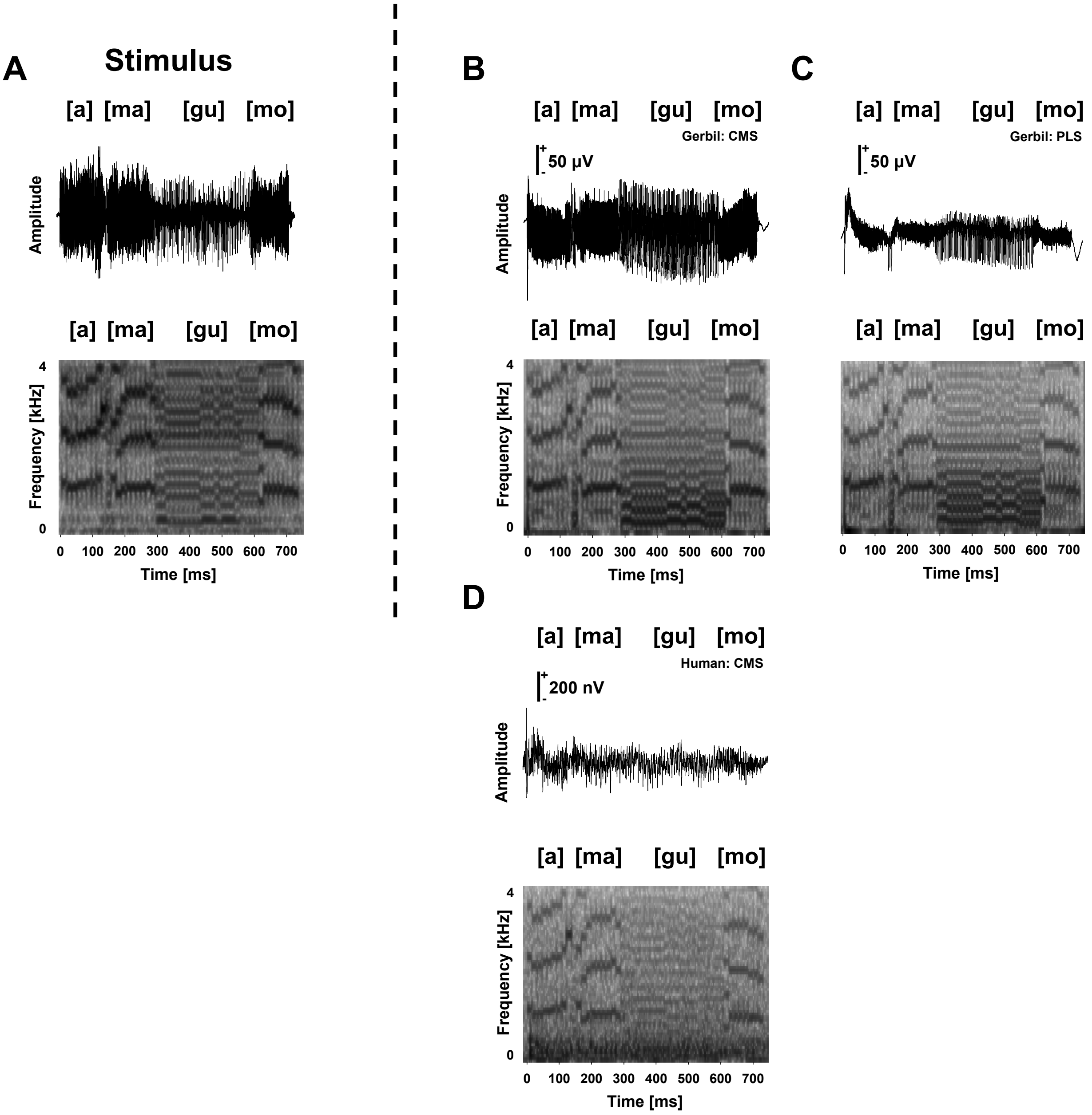

Fig. 3. A recorded click-modulated speech sound (CMS) and auditory brainstem responses. Upper figures show the waveform. Lower figures show spectrograms. These signals were processed by FFT with a 8000-Hz sampling rate and 64-point FFT length. (A) A CMS signal recorded with a microphone. (B) Cochlear nerve responses evoked by CMS in the Mongolian gerbil. (C) Cochlear nerve responses evoked by a pulsed laser speech sound (PLS) in the Mongolian gerbil. (D) Auditory brainstem responses evoked by CMS in the human subject.

that in the human was recorded with an 875-µs time lag.

Figure 4 shows the amplitude spectrum for each mora. The amplitude spectra of the neural activities evoked by CMS in the Mongolian gerbil and that by PLS had harmonic structures with clear peaks ($Q_{10\ dB}$: >1.3) except for the mora 'gu' (Fig. 4A). Indeed, the peak frequencies of the harmonics were almost the same; the difference between the two was less than 129 Hz. As frequency increased, the amplitude of the harmonics of the spectrum evoked by CMS decreased, but not by as much as that by PLS. A clear harmonic structure was also observed in the amplitude spectra of neural activities from the human auditory brainstem (Fig. 4B). Comparing the amplitude spectra of CMS-evoked neural activity in the Mongolian gerbil with that in the human, the peaks of the amplitude spectra were comparable, with the exception of a low frequency peak in the human (see arrows in Fig. 4B).

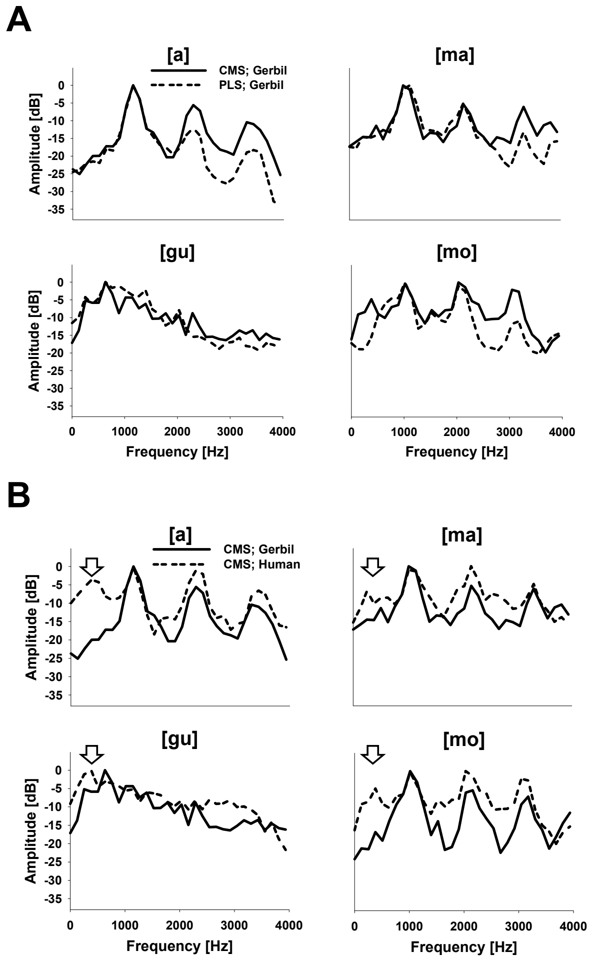

Fig. 4. Amplitude spectra for each mora in auditory stem responses. (A) Amplitude spectra of different mora signals evoked by CMS and PLS in the Mongolian gerbil. (B) Amplitude spectra of different mora signals in the Mongolian gerbil and human subject evoked by CMS. Arrow shows a low frequency peak in the spectrum of human subject.

Figure 5A shows the time variance of the correlation coefficient between the spectra of neural activities evoked by CMS and by PLS. The correlation coefficients stayed positive and was always higher than the statistically significant level ($r = 0.24$; $p < 0.05$). The time variance of the correlation coefficient between the spectra of compound action potentials evoked by CMS in the Mongolian gerbil and in the human was shown in Fig. 5B. The correlation coefficients stayed positive, while these were between 0 and 0.9.

4. Discussion

The compound action potentials of cochlear nerves evoked by PLS (Fig 3C) resembled those by CMS (Fig. 3B). Weinberger and colleagues revealed that an alternating current (AC) signal that represents the summation of phase-locked cochlear nerve activity was recorded by a presenting low-frequency tone [5)]. The response was termed an auditory nerve neurophonic (ANN) [6-8)]. The cochlear response evoked by CMS (Fig. 3A) is considered to be an ANN, because our experimental paradigm followed previous studies recording ANNs in Mongolian gerbils [6-8)]. Our own previous research with the Mongolian gerbil revealed that compound action potentials of cochlear nerves evoked with an infrared laser resembled that evoked by click sounds [9)]. Because PLS was apparently a similar stimulus to CMS, except that the pulsed infrared laser substituted for the click sounds, the PLS-evoked neural response we observed could also be an ANN; both CMS and PLS resembled the first formant frequency transition of an original human speech sound (Figs. 3B–D).

The results in Figs. 3B and 3D revealed that presenting CMS to a human produced similar neural activities as that of the Mongolian gerbil. As previous studies have demonstrated, an AC signal that encodes speech-specific information can be observed by presenting a complex sound to a human using a scalp electrode. The neural responses were referred to frequency following responses (FFRs) [10-12]. Snyder and colleagues revealed that the neural responses recorded at the scalp had several common features, apart from some cochlear microphonic contamination, as that recorded at the cochlear nerves in a cat [13]. The study also demonstrated that the amplitude of FFR was smaller than that of ANN. The results of our study are comparable to the previous study. Thus, CMS could produce similar neural activities from the cochlear nerves of the Mongolian gerbil and the auditory brainstem of the human subject.

We conducted a cross-correlation analysis between all neural activities to estimate the latency difference. Our recording could be contaminated with electromagnetic waves produced by the speaker system, and if we primarily recorded the electromagnetic interference, not the electrophysiological response, the latency of all the data could be the same. However, if we recorded neural activities, the latencies of the recorded compound action potentials would be expected to differ with each stimulation method. Knapp and colleagues indicated that the time delay between two signals could be estimated by a cross-correlation analysis [14]. Indeed, cross-correlation analysis indicated that the latency of neural responses evoked by CMS was longer than that evoked by PLS (difference: 250 μs). The same analysis showed that neural activities evoked by CMS in the Mongolian gerbil were shorter than that in the human (difference: 875 μs). These results suggest that latencies of neural activities are longer, by about 1100 μs; the prolongation roughly matches the sum of the sound propagation, cochlear delay, and neural response latency. Taken together, this correspondence validates our experimental settings.

Spectra analyses of compound action potentials showed that high-frequency responses were more prominent in the response to CMS than to PLS (Fig. 4A). This difference could have occurred because the stimulating point differed between the infrared laser and the click sound. The click sound had a broadband frequency component; the sound could stimulate cochlear nerves over a wide frequency range, including the high-frequency portion. However, the laser may have stimulated a relatively limited frequency part of the cochlear nerve, because it stimulated nerves through the round window. A previous study in gerbils suggested that the cochlea has a place frequency map, and the map had a slope of less than 1.4 mm/octave [15,16]. Additionally, Matic and colleagues argued that it was difficult to stimulate the very base of the cochlea nerves, the high-frequency region, using an optic fiber through the round window, due to limitations associated with probe orientation [17].

A stronger low-frequency response was observed in the spectrum of neural activities in the human than the gerbil (Fig. 4B). One possible reason is the difference in hearing sensitivity between the Mongolian gerbil and humans. Ryan reported that hearing sensitivity below 1000 Hz in the Mongolian gerbil was lower than in humans by 10 dB at 500 Hz [18]. This hearing sensitivity difference could produce the low-frequency peak in the human subject.

Although the correlation coefficients in Fig. 5A were always high, those in Fig. 5B were relatively

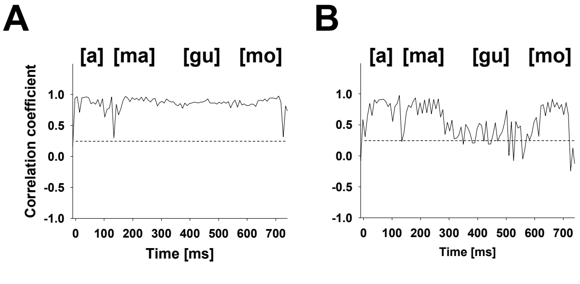

Fig. 5. Correlation estimate of auditory brainstem responses. (A) Time variance of correlation coefficient between cochlear nerve responses evoked by CMS and that by PLS in the Mongolian gerbil. (B) Time variance of correlation coefficient between auditory stem response in the human subject and cochlear nerve responses in the Mongolian gerbil evoked by CMS. Dotted line shows statistically significant level ($r = 0.24$; $p < 0.05$).

variable and lower at for the mora of ‘gu’ than for the other morae. This difference could at least partially be the result of the species difference in auditory sensitivity, as above; also, the difference could be attributable to masking of the stimulus sound by low-frequency environmental noise. Compound action potentials evoked by CMS preserved the first formant frequency of an original speech sound; the F1 of mora of ‘u’ was lower than those of ‘a’ and ‘o’ 19,20). Thus, low-frequency noise may have significantly lowered the correlation coefficient for the mora of ‘gu’ than the other morae (Fig. 5B).

In this study, we investigated whether PLS could produce speech perception in a human subject. Our results with the Mongolian gerbil demonstrated that PLS evoked similar compound action potentials to CMS. The results suggested that both stimuli, PLS and CMS, created similar neural activity in the auditory pathway in the gerbil. Additionally, the experiment with the human subject suggested that similar compound action potentials were evoked by CMS in the Mongolian gerbil and human, suggesting CMS was processed in the auditory pathways of both species in a similar manner. Taken together, PLS could evoke similar neural activity to CMS in a human subject. Our previous study revealed that CMS was at least partially intelligible as a speech sound 4); thus, our findings suggest that PLS could be at least partially intelligible as speech sounds.

The authors thank Hidetaka Yashiro for technical support and valuable discussion. This research was supported by the JSPS KAKENHI (Grant Number: 15K12069).

References

1) J. Wells, C. Kao, K. Mariappan, J. Albea, E. D. Jansen, P. Konrad, and A. Mahadevan-Jansen, “Optical Stimulation of Neural Tissue in vivo.”, *Opt. Lett.*, **30** [5], 504–506 (2005).

2) A. D. Izzo, J. T. Walsh, Jr., E. D. Jansen, M. Bendett, J. Webb, H. Ralph, and CP. Richter, “Optical ParameterVariability in Laser Nerve Stimulation: A Study of Pulse Duration, Repetition Rate, and Wavelengt”, *IEEE Trans. Biomed. Eng.*, **54** [6], 1108–1114 (2007).

3) A. I. Matic, A. M. Robinson, H. K. Young, B. Badofsky, S. M. Rajguru, S. Stock, CP. Richter, “Behavioral and Electrophysiological Responses Evoked by Chronic Infrared Neural Stimulation of the Cochlea”, *PLoS One*, **8** [3], (2013).

4) Y. Tamai, S. Hiryu, and K. I. Kobayasi, “How to Create the Intelligible Speech Perception by Cochlear Stimulation with the Laser”, *Soc. Neurosci. Abstr.* 237: 16366, (2016).

5) N. M. Weinberger, L. M. Kitzes, and D. A. Goodman, “Some Characteristics of the ‘Auditory Neurophonic’”, *Experientia*, **26** [1], 46–48 (1970).

6) K. R. Henry, “Auditory Nerve Neurophonic Recorded from the Round Window of the Mongolian Gerbil”, *Hear. Res.*, **90** [1–2], 176–184 (1995).

7) K. R. Henry, “Auditory Nerve Neurophonic Produced by the Frequency Difference of Two Simultaneously Presented Tones”, *Hear. Res.*, **99** [1–2], 151–159 (1996).

8) K. R. Henry, “Tuning Curves of the Difference Tone Auditory Nerve Neurophonic”, *Hear. Res.*, **99** [1–2], 160–167 (1996).

9) S. Matsui, T. Morimoto, S. Boku, K. I. Kobayasi, and H. Riquimaroux, “Compound Action Potentials Generated by Near Infrared Laser Irradiation Recorded at the Round Window”, *Proc. Audit. Res. Meet.*, **43** [8], 621–624 (2013).

10) S. Greenberg, J. T. Marsh, W. S. Brown, and J. C. Smith, “Neural Temporal Coding of Low Pitch. I. Human Frequency-following Responses to Complex Tones”, *Hear. Res.*, **25** [2–3], 91–114 (1987).

11) N. Russo, T. Nicol, G. Musacchia, and N. Kraus, “Brainstem Responses to Speech Syllables”, *Clin. Neurophysiol.*, **115** [9], 2021–2030 (2004).

12) I. Akhoun, S. Gallégo, A. Moulin, M. Ménard, E. Veuillet, C. Berger-Vachon, L. Collet, H. Thai-Van, “The Temporal Relationship between Speech Auditory Brainstem Responses and the Acoustic Pattern of the Phoneme /ba/ in Normal-hearing Adults”, *Clin. Neurophysiol.*, **119** [4], 922–933 (2008).

13) R. L. Snyder and C. E. Schreiner, “The Auditory Neurophonic: Basic Properties”, *Hear. Res.*, **15** [3], 261–280 (1984).

14) C. H. Knapp and G. C. Carter, “The Generalized Correlation Method for Estimation of Time Delay”, *IEEE Trans. Acoust.*, **24** [4], 320–327 (1976).

15) M. Müller, “The Cochlear Place-frequency Map of the Adult and Developing Mongolian Gerbil”, *Hear. Res.*, **94** [1–2], 148–156 (1996).

16) C. P. Richter, R. Edge, D. Z. Z. He, and P. Dallos, “Development of the Gerbil Inner Ear Observed in the Hemicochlea”, *J. Assoc. Res. Otolaryngol.*, **1** [3], 195–210 (2000).

17) A. I. Matic, J. T. Walsh, and C.P. Richter, “Spatial Extent of Cochlear Infrared Neural Stimulation Determined by Tone-on-light Masking”, *J. Biomed. Opt.*, **16** [11], 118002 (2011).

18) A. Ryan, “Hearing Sensitivity of the Mongolian Gerbil, Meriones Unguiculatis”, *J. Acoust. Soc. Am.*, **59** [5], 1222–1226 (1976).

19) G. Peterson and H. L. Barney, “Control Methods Used in a Study of the Vowels”, *Jarnal Acoust. Soc. Am.*, **24** [2], 175–184 (1952).

20) J. Hillenbrand, L. A. Getty, M. J. Clark, and K. Wheeler, “Acoustic Characteristics of American English Vowels”, *J. Acoust. Soc. Am.*, **97** [5 Pt 1], 3099–111 (1995).