Speech perception created by noninvasive direct cochlear stimulation

Yuta TAMAI† , Yuka SHINPO†, Kensuke HORINOUCHI†, Makoto ARIMURA†, Sizuko HIRYU†‡ and Kohta I. KOBAYASI†

†Neuroethology and bioengineering laboratory, Doshisha University, 1-3 Tatara Miyakodani, Kyotanabe, Kyoto, Japan
‡JST PREST
E-mail: †dmq1041@mail4.doshihsa..ac.jp,

Abstract Action potentials are evoked by irradiating infrared laser to a cochlea. We investigated whether pulsed infrared laser could produce speech perception. Our previous our study developed a speech encoding scheme for a single channel cochlear stimulation. Mongolian gerbil (Meriones unguiculatus) and Japanese native speaker were used as subjects. Click modulated speech sound (CMS) and pulsed laser speech sound (PLS) were presented to the gerbil. Sound pressure level was fixed at 80 dB SPL. Radiation exposure was fixed at 20 μJ. Neural responses were measured from round window. The CMS were fixed at 70 dB SPL and proposed to human subject. The responses were recorded from scalp electrode. As results, coherence of correlation in all signals indicated high value. The result suggests that PLS might generate the similar neural responses as CMS from auditory brainstem in human. The result opens up the possibility of creating intelligible speech perception by irradiating infrared laser to the cochlea.

Keywords infrared laser, distorted speech sound, formant frequency, amplitude envelope

1. Introduction

Cochlear implants are widely used to compensate for sensorineural hearing loss. The devices restore the hearing of an otherwise deaf individual. In the cochlea, low-frequency sounds activate neurons in the apex of cochlea, and high-frequency sounds stimulate the basal part; this organization is known as tonotopicity. A multi-channel electrode is inserted into the cochlea to restore the tonotopic responses found in the normal, acoustically stimulated cochlea [1][2]. By restoring tonotopicity, thecochlear implant can produce detailed frequency information, and many cochlear implant users perceive electric stimulation produced by the system as speech sounds.

One of the largest drawbacks of a cochlear implant is that it requires major surgical intervention. From the early stages of its development, an extra-cochlear implant has been considered as a possible alternative as a multi-channel implant. A pioneering study by Fourcin and colleagues (1979) showed that a single electrode placed at the round window was able to produce various acoustic features of speech such as intonation andvoiced-voiceless information [3]. Other studies revealed that the extra-cochlear single-channel implant improved lip-reading ability [4].

Despite these early successes, an extra cochlear implant has not been fully implemented clinically. Because the extra-cochlear single-channel system stimulates all cochlear nerve fibers simultaneously, the system cannot replicate the fine frequency structure, and is regarded as less capable of restoring speech perception than the multi-channel system. The extra-cochlear implant has thus far been primarily used as speech-reading aid.

It is well known that action potentials can be triggered by irradiating infrared laser stimulation. Wells and his colleagues reported that pulsed infrared laser evokes neural activities in mammalian peripheral nerve *in vivo* [5]. Other studies tried to apply the optic stimulation to cochlear implant [6][7]. In contrast to electrical stimulation, optical stimulation has several appearing features. For example, optical stimulation can stimulate nerves without contacting tissues, spatial selectivity of stimulation is improved, and stimulation artifact in not generated by irradiating infrared laser to neurons. Optic stimulation has been much attention for stimulating nerves.

In this study, we tried to confirm that speech sound is restored by irradiating infrared laser to cochlear nerves. Our previous study revealed that pulsed infrared laser can stimulate cochlear nerves through tympanic membrane [8]. As mentioned earlier, it is difficult to stimulate cochlear nerves separately to replicate tonotopicity with single-channel stimulation. We therefore assumed that single channel stimulation would create a perception similar to a click sound because the system stimulates all nerves simultaneously. We synthesized a click modulated speech sound as a simulated sound of pulsed laser. Our previous study quantified the intelligibility of simulated sounds of extra-cochlear single-channel implant [unpublished]. However, the availability of restoring speech sounds by irradiating pulsed laser has not been investigated. We synthesized pulsed laser speech sound (PLS) and hypothesize if the similar responses were observed by proposing CMS and PLS, same information could reach cochlea. Thus, we reasonably expected that CMS and PLS could evoke same neural responses from different subject.

2. Materia and methods

Subject

A Mongolian gerbil (*Meriones unguiculatus*) with 78 g was used as a subject and a native Japanese speaker was participated. The human subject passed a hearing screening at 25 dB HL with frequencies of 0.5, 1, 2, and 4 kHz.

Surgery

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

Stimuli

Click-modulated speech sound (CMS)

The CMS simulates the perception that is evoked by single-channel stimulation of a cochlear nerve bundle. The sound is a click train, whose pitch (repetition rate) follows the first formant center frequency of an original speech sound. The pulse width was 100 μs. Specifically, formant frequencies were extracted from the original sounds by linear predictive coding (LPC) and fast Fourier transforms (FFTs) at 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). Original speech sound was four mora Japanese word (‘a’, ‘ma’, ‘gu’, ‘mo’). The word was voiced by a female speaker.

Pulsed laser speech sound (PLS)

The stimulation is a pulsed laser, whose repetition rate is the same rate as CMS. The wavelength of pulsed laser was 1871 nm. The pulse width is 100 μs.

Experimental environment

Mongolian gerbil

The experiment using 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 by saline as reference. Acoustic stimulus was presented at 10 cm from the subject, and error rage was about 5 cm. Optic stimulation was irradiated to cochlear nerves through the round window.

Human

The experiment using human subject was conducted in an acoustically and electrically shielded room. Subject closes his eyes and reclined on the 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 4000 Hz. Common ground was placed on the left mastoid served as the common ground. The impedances between the electrodes were calibrated below 3000 Ohms (MaP811, Measurement and Processing). Stimuli were delivered at 40 cm from the subject, and error range was about 10 cm.

Experimental procedure

Mongolian gerbil

Neural responses evoked by acoustic and optic stimulation were recorded. Acoustic stimuli were proposed 100 times by dome tweeter (FT28D, FOSTEX).

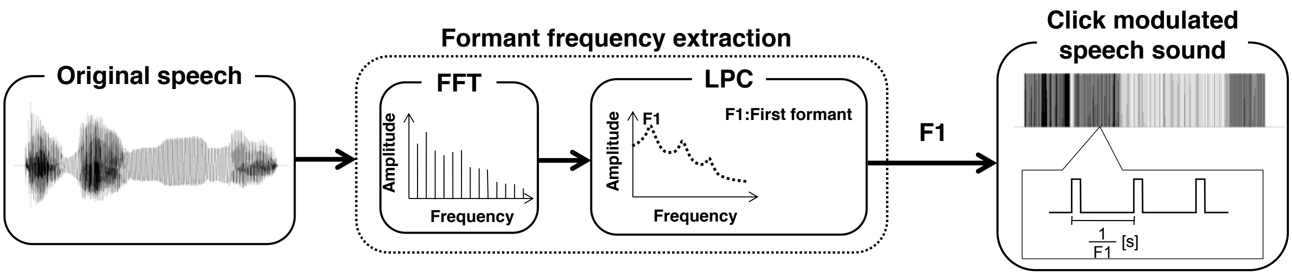

Fig. 1 Encoding process of the click-moderated speech sound (CMS). Schematic diagram depicting the process of analyzing the speech signal and synthesizing the CMS. See text for details.

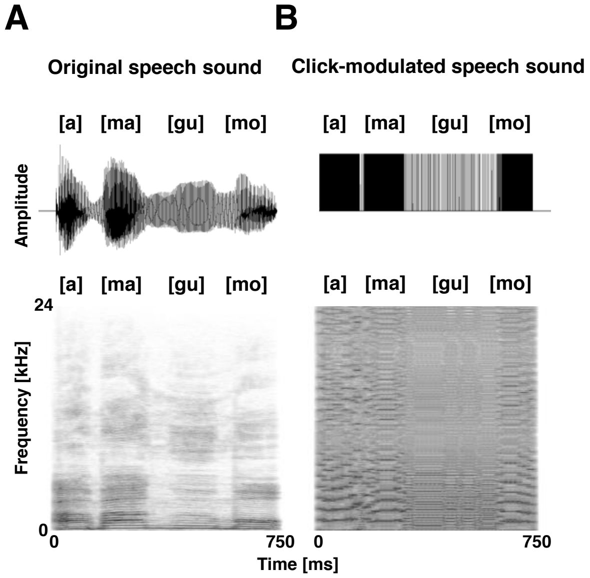

Fig. 2 An example of the stimulus. Upper figures show the waveform, and lower figures show spectrograms. The signals were processed by fast Fourier transforms at 48000 Hz sampling rate and 1024-point FFT length. (A) Original speech sound. A Japanese word "[a][ma][gu][mo]." (B) Click-modulated speech sound synthesized from the original sound (A).

The stimuli were calibrated at 80 dB SPL by microphone (Type1, ACO pacific).Optic stimuli were proposed 100 times by diode laser stimulation system (BWF-OEM, B&W). The stimuli were calibrated at 20 μJ per pulsed laser by actinometer (GmbH, THORABS).

Human

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

Electrophysiological analysis

The electric signals were amplified by 20000 times with low cut filter (cutoff: 0.08 Hz) (MEG-1200, NIHON KOHDEN).In Mongolian gerbil experiment, the signals were averaged 100 times. In human experiment, the signals were averaged 5000 times. The averaged signal was extracted per 8 ms and processed by fast Fourier transforms (FFTs) at 8000 Hz sampling rate and 64-point FFT length. Coherence of correlations was measured between these extracted spectrums. All signal processing was performed using Matlab (MathWorks).

3. Results

Fig.3-A and Fig.3-B shows neural responses evoked by CMS and PLS in Mongolian gerbi, respectivelyl. This figure indicates in both temporal waveforms and spectrums, similar responses were obse rved by proposed CMS and PLS to Mongolian gerbil. Fig.3 shows that brain steam response evoked by CMS resembles neural responses recorded from gerbil and human i except for amplitude.

Fig.4 shows the result of correlation estimate between neural responses evoked by CMS and that by PLS (Fig.4-A) and between neural responses evoked by CMS and brain steam response evoked by CMS (Fig.4-B). Fig.4-A indicates that all coefficient of correlation is higher than chance level. Fig.4-A also shows that the average of correlation indicates 0.85. Fig.4-B shows that the correlation was higher than chance level in the mora of 'a', 'ma', 'mo'. The values in the mora of 'u' was almost same value as chance level. Fig.4-B also shows that the average of correlation indicates 0.55. A crude estimate of chance level can be defined as 0.18, because the value was indicated by comparing different mora such as comparing 'a' with 'mo'.

Fig.5-A shows amplitude spectrum of neural responses

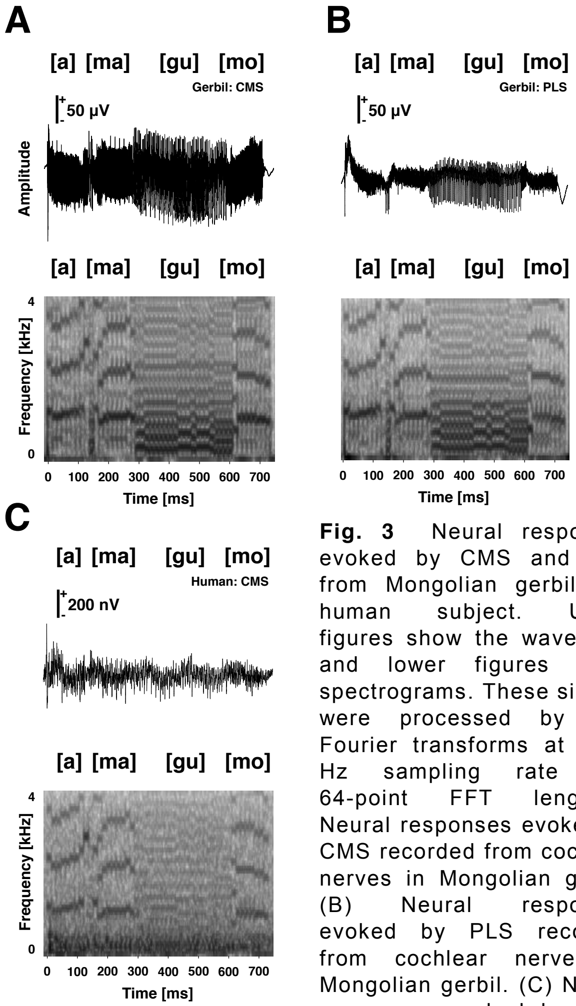

Fig. 3 Neural responses evoked by CMS and PLS from Mongolian gerbil and human subject. Upper figures show the waveform, and lower figures show spectrograms. These signals were processed by fast Fourier transforms at 8000 Hz sampling rate and 64-point FFT length(A) Neural responses evoked by CMS recorded from cochlear nerves in Mongolian gerbil. (B) Neural responses evoked by PLS recorded from cochlear nerves in Mongolian gerbil. (C) Neural responses evoked by CMS recorded from auditory brainstem in human subject.

in Mongolian gerbil evoked by CMS and PLS in each mora (‘a’, ‘ma’, ‘gu’, ‘mo’). These figures show that the specific peaks of spectrum of ‘a’, ‘ma’, ‘mo’ were observed upper 1000 Hz. Specific peak of ‘u’ spectrum were observed below 1000 Hz. Fig.5-B shows amplitude spectrum of cochlear nerve’s responses in Mongolian gerbil and auditory brainstem responses in human evoked by CMS. These figures also show that specific peaks of spectrum were observed upper 1000 Hz except for the mora of ‘gu’.

Fig.6-A shows that the results of cross correlation estimate between neural responses evoked by CMS and PLS in Mongolian gerbil. The maximum value of correlation was at 250 μs. Fig.6-B shows that the results of cross correlation estimate between neural responses in Mongolian gerbil and auditory brainsteam responses in human evoked by CMS. The maximum value of correlation was at 875 μs.

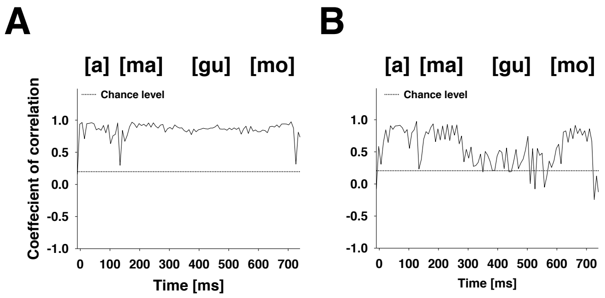

Fig.4 Time variation of coefficient of correlation in neural responses. (A) correlation estimate between neural responses evoked by CMS from cochlear nerves in Mongolian gerbil. (B) correlation estimate between neural responses evoked by CMS from cochlear nerves in Mongolian gerbil and from auditory stream in human subject. The horizontal dotted line indicates chance level.

4. Discussion

The results of Fig.3-A and Fig.3-B show that PLS evokes the similar neural responses as CMS. Our previous studies revealed that pulsed laser evokes similar neural responses as that by proposing click sound [9]. This means that the PLS have the similar properties as that composed of click sounds. The results indicate that PLS produces similar neural responses as CMS.

The results of Fig.3-A and Fig.3.C show that CMS evokes similar neural responses in both Mongolian gerbil and human. However, the amplitude of these two signals widely different. As previous studies have demonstrated, frequency following responses recorded from round window were about μV degree [10][11][12]. The responses recorded from scalp electrode were about nV degree [13][14]. Previous study indicated that the amplitude of response were different depending on recording place and recording from subjects. [15]. The results of Fig.3-A and Fig.3.C reveal that similar neural responses were observed in different subject in different recording site by proposed CMS.

As seen in Fig.4, neural responses recorded in the experiment have the high correlation each other. However, the correlation of the mora of ‘gu’ in Fig.4-B is different from other morae. As seen in Fig.5-B, amplitude spectrum of each mora shows that obvious difference between each spectrum was not observed compared to other morae. In many cases, the correlation easily reaches high values if these signals have specific peaks and the peaks are corresponding to each other. Previous study said that the first formant frequency of ‘u’ was lower than other vowels (below 1000 Hz) [16][17] and demonstrated that hearing sensitivity at lower frequencies such as 500 Hz were mrkedly different between Mongolian gerbil and human (ca. 10 dB SPL) [18]. In this experiment, recording site and experimental equipment such as electrode were different. The experimental procedure makes a large difference in environmental noise. These artifacts at lower frequencies can reduce the coherence of correlation about ‘u’.

Electrophysiological recording could measure electromagnetic wave. This is because speaker that we used in the experiment was droved electrically though the stimulator was shielded by faraday cage. Cross-correlation analysis is widely used for determining time delay between two signals [19]. We used this analysis for delay estimation between neural responses evoked by CMS and PLS and between those evoked by CMS from different subject (between Mongolian gerbil and human). The results suggest that time delay was 250 μs between neural responses evoked by CMS and PLS. The latency of action potentials evoked by acoustic stimulation was longer than that evoked by Optic stimulation. There were two reasons. At first, optic stimulation is faster than acoustic stimulation. Second, infrared laser was irradiated near the round window through there was some distance between speaker and subject in acoustic stimulation. In this experiment, delay time between neural responses evoked by CMS and that evoked by PLS was estimated in the range of 147 – 441 μs because the distance from speaker to subject’ ear was 10 cm (±5 cm). The result of cross-correlation analysis was within the range. In addition, delay time between neural responses evoked by CMS from Mongolian gerbil and that from human was estimated within the range of 294 – 882 μs because the distance from speaker to subject' ear was 30 cm (± 10 cm). The result of cross-correlation analysis wasin also in the range. Theses result mean that the latency of neural responses were long in the order of that by CMS from human, that by CMS from Mongolian gerbil, that by PLS from Mongolian gerbil. The result that delay times are different in three experimental conditions indicates that the signals we recorded are not caused by electromagnetic wave.

The high coherence of correlation values between these signals we measured suggests that PLS could produce a similar neural response like CMS evoking produces in human subjects. Previous studies reveal that the rapid increase in temperature of water in nerves causes action potentials [20], and the heating changes the electrical capacitance of the plasma membrane [21]. Other study said that TRPV4 channels mediate the neural responses [22][23]. However, detail mechanism of neural firing evoked by infrared laser irradiation has been a still unexplained. In addition, the risk caused by irradiating infrared laser to neurons has not been revealed. Hence, the experiment of infrared neural stimulation to human subjects have not been conducted yet. We hypothesize if the similar neural responses were observed in neural responses evoked by CMS, PLS from Mongolian gerbil, and that evoked by CMS from human, the similar responses could be measured by irradiating PLS to human. Krishnan and his colleagues show that human frequency following response includes some acoustic feature of speech sound [13][14]. Our previous study demonstrates that CMS is at least partially intelligible to human subjects [unpublished]. In all, our results demonstrate that human subject could perceive PLS as like CMS.

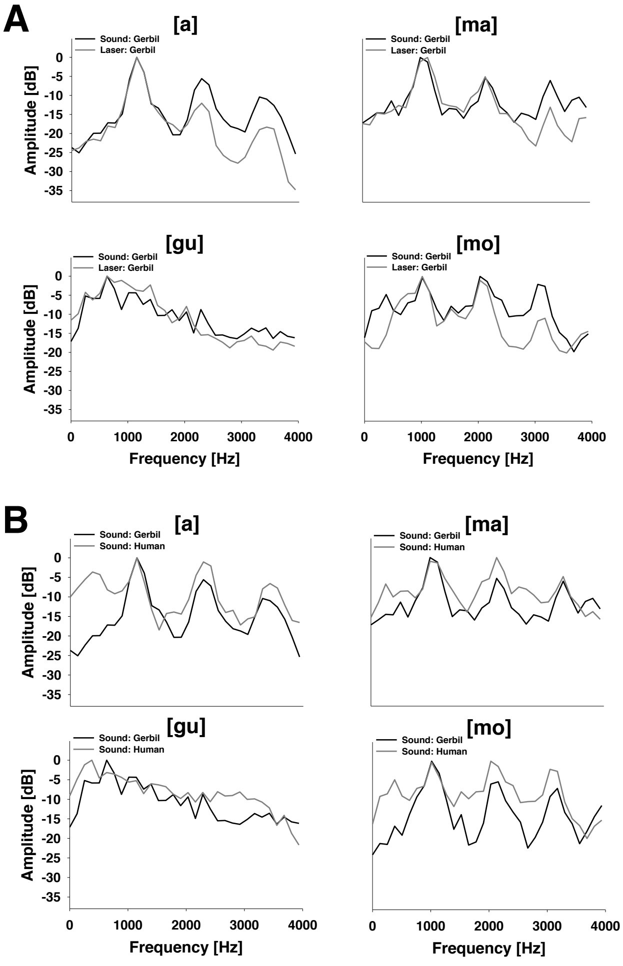

Fig.5 Amplitude spectrum of each mora in neural responses. (A) A comparison between spectrum of neural responses evoked by CMS and that by PLS from cochlear nerves in Mongolian gerbil. (B) A comparison between spectrum of neural responses evoked by CMS from cochlear nerves in Mongolian gerbil and from auditory stream in human subject.

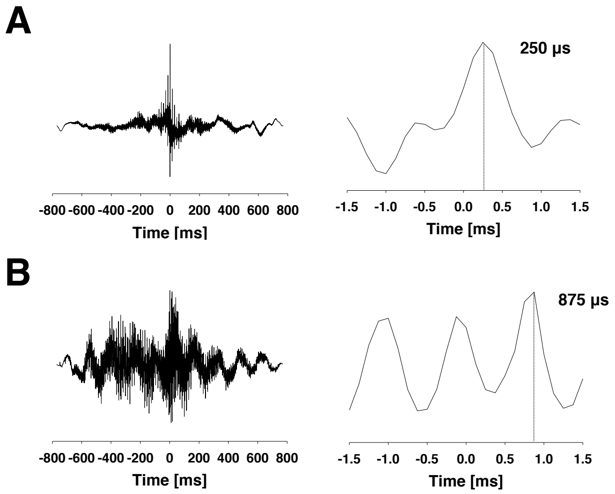

Fig.6 Coss-correlation of neural responses. Left figure shows entire function of cross-correlation. Right figure shows around the maximum point. (A) Correlation between neural responses evoked by CMS and PLS from cochlear nerves in Mongolian gerbil. (B) Correlation between neural responses evoked by CMS from cochlear nerves in Mongolian gerbil and from auditory steam in human subject.

References

[1] N. Y. Kiang and E. C. Moxon, "Physiological considerations in artificial stimulation of the inner ear.," *Ann. Otol. Rhinol. Laryngol.*, vol. 81, no. 5, pp. 714–730, Oct. 1972.

[2] J. E. Rose, J. F. Brugge, D. J. Anderson, and J. E. Hind, "Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey.," *J. Neurophysiol.*, vol. 30, no. 4, pp. 769–93, Jul. 1967.

[3] A. Fourcin and S. Rosen, "External electrical

stimulation of the cochlea: clinical, psychophysical, speech-perceptual and histological findings," *Br. J. Audiol.*, vol. 13, no. 3, pp. 85–107, 1979.
[4] S. Rosen and V. Ball, "Speech perception with the Vienna extra-cochlear single-channel implant: a comparison of two approaches to speech coding.," *Br. J. Audiol.*, vol. 20, no. 1, pp. 61–83, Feb. 1986.
[5] 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.*, vol. 30, no. 5, pp. 504–506, 2005.
[6] A. D. Izzo, J. T. Walsh, E. D. Jansen, M. Bendett, J. Webb, H. Ralph, and C. P. Richter, "Optical parameter variability in laser nerve stimulation: A study of pulse duration, repetition rate, and wavelength," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 6, pp. 1108–1114, 2007.
[7] C. P. Richter, R. Bayon, A. D. Izzo, M. Otting, E. Suh, S. Goyal, J. Hotaling, and J. T. Walsh, "Optical stimulation of auditory neurons: Effects of acute and chronic deafening," *Hear. Res.*, vol. 242, no. 1–2, pp. 42–51, 2008.
[8] Y. Tamai, S. Matsui, K. I. Kobayasi, and H. Riquimaroux, "Optic method for stimulating cochlear nerve in Mongolian gerbil," *Proc. Audit. Res. Meet.*, vol. 45, no. 3, pp. 135–139, 2015.
[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.*, vol. 43, no. 8, pp. 621–624, 2013.
[10] K. R. Henry, "Auditory nerve neurophonic recorded from the round window of the Mongolian gerbil," *Hear. Res.*, vol. 90, no. 1–2, pp. 176–184, 1995.
[11] K. R. Henry, "Tuning curves of the difference tone auditory nerve neurophonic," *Hear. Res.*, vol. 99, no. 1–2, pp. 160–167, 1996.
[12] K. R. Henry, "Sharply tuned cochlear nerve ensemble periodicity responses to sonic and ultrasonic frequencies," *J. Comp. Physiol. - A Sensory, Neural, Behav. Physiol.*, vol. 181, no. 3, pp. 239–246, 1997.
[13] A. Krishnan, "Human frequency-following responses: Representation of steady-state synthetic vowels," *Hear. Res.*, vol. 166, no. 1–2, pp. 192–201, 2002.
[14] A. Krishnan, Y. Xu, J. T. Gandour, and P. A. Cariani, "Human frequency-following response: Representation of pitch contours in Chinese tones," *Hear. Res.*, vol. 189, no. 1–2, pp. 1–12, 2004.
[15] D. S. Barth, N. Goldberg, B. Brett, and S. Di, "The spatiotemporal organization of auditory, visual, and auditory-visual evoked potentials in rat cortex," *Brain Res.*, vol. 678, no. 1–2, pp. 177–190, 1995.
[16] G. Peterson and H. L. Barney, "Control Methods Used in a Study of the Vowels," *Jarnal Acoust. Soc. Am.*, vol. 24, no. 2, pp. 175–184, 1952.
[17] J. Hillenbrand, L. A. Getty, M. J. Clark, and K. Wheeler, "Acoustic characteristics of American English vowels.," *J. Acoust. Soc. Am.*, vol. 97, no. 5 Pt 1, pp. 3099–111, May 1995.
[18] A. Ryan, "Hearing sensitivity of the mongolian gerbil, Meriones unguiculatis.," *J. Acoust. Soc. Am.*, vol. 59, no. 5, pp. 1222–1226, 1976.
[19] C. H. Knapp and G. C. Carter, "The Generalized Correlation Method for Estimation of Time Delay," *IEEE Trans. Acoust.*, vol. 24, no. 4, pp. 320–327, 1976.
[20] J. Wells, C. Kao, P. Konrad, T. Milner, J. Kim, A. Mahadevan-Jansen, and E. D. Jansen, "Biophysical mechanisms of transient optical stimulation of peripheral nerve.," *Biophys. J.*, vol. 93, no. 7, pp. 2567–80, 2007.
[21] M. G. Shapiro, K. Homma, S. Villarreal, C.-P. Richter, and F. Bezanilla, "Infrared light excites cells by changing their electrical capacitance.," *Nat. Commun.*, vol. 3, p. 736, 2012.
[22] E. S. Albert, J. M. Bec, G. Desmadryl, K. Chekroud, C. Travo, S. Gaboyard, F. Bardin, I. Marc, M. Dumas, G. Lenaers, C. Hamel, a. Muller, and C. Chabbert, "TRPV4 channels mediate the infrared laser-evoked response in sensory neurons," *J. Neurophysiol.*, vol. 107, no. 12, pp. 3227–3234, 2012.
[23] J. Yao, B. Liu, and F. Qin, "Rapid temperature jump by infrared diode laser irradiation for patch-clamp studies," *Biophys. J.*, vol. 96, no. 9, pp. 3611–3619, 2009.