



# Frequency-Limiting Effects on Speech and Environmental Sound Identification for Cochlear Implant and Normal Hearing Listeners

Son-A Chang<sup>1</sup>, Jong Ho Won<sup>2</sup>, HyangHee Kim<sup>3</sup>, Seung-Ha Oh<sup>1</sup>,  
Richard S. Tyler<sup>4</sup>, and Chang Hyun Cho<sup>5</sup>

<sup>1</sup>Department of Otolaryngology-Head and Neck Surgery, Seoul National University Hospital, Seoul, Korea

<sup>2</sup>Department of Audiology and Speech Pathology, University of Tennessee Health Science Center, Knoxville, TN, USA

<sup>3</sup>Graduate Program of Speech and Language Pathology, Department and Research Institute of Rehabilitation Medicine, Yonsei University College of Medicine, Seoul, Korea

<sup>4</sup>Department of Otolaryngology-Head and Neck Surgery, University of Iowa Hospitals and Clinics, Iowa City, IA, USA

<sup>5</sup>Department of Otolaryngology-Head and Neck Surgery, Gachon University Gil Medical Center, Incheon, Korea

**Received** July 19, 2017  
**Revised** October 9, 2017  
**Accepted** October 17, 2017

## Address for correspondence

Jong Ho Won, PhD  
Department of Audiology and  
Speech Pathology,  
University of Tennessee  
Health Science Center,  
Knoxville, TN 37996, USA  
**Tel** +1-865-974-5219  
**Fax** +1-865-974-1539  
**E-mail** jhwon15@gmail.com

**Background and Objectives:** It is important to understand the frequency region of cues used, and not used, by cochlear implant (CI) recipients. Speech and environmental sound recognition by individuals with CI and normal-hearing (NH) was measured. Gradients were also computed to evaluate the pattern of change in identification performance with respect to the low-pass filtering or high-pass filtering cutoff frequencies. **Subjects and Methods:** Frequency-limiting effects were implemented in the acoustic waveforms by passing the signals through low-pass filters (LPFs) or high-pass filters (HPFs) with seven different cutoff frequencies. Identification of Korean vowels and consonants produced by a male and female speaker and environmental sounds was measured. Crossover frequencies were determined for each identification test, where the LPF and HPF conditions show the identical identification scores. **Results:** CI and NH subjects showed changes in identification performance in a similar manner as a function of cutoff frequency for the LPF and HPF conditions, suggesting that the degraded spectral information in the acoustic signals may similarly constraint the identification performance for both subject groups. However, CI subjects were generally less efficient than NH subjects in using the limited spectral information for speech and environmental sound identification due to the inefficient coding of acoustic cues through the CI sound processors. **Conclusions:** This finding will provide vital information in Korean for understanding how different the frequency information is in receiving speech and environmental sounds by CI processor from normal hearing. **J Audiol Otol 2018;22(1):28-38**

**KEY WORDS:** Frequency information · Cross-over frequency · Korean speech sound · Environmental sound · Cochlear implant.

## Introduction

One of the ecological properties of sounds, either speech or non-speech, is the rich and dynamic representation of spectral information over the frequency range that is important for hu-

man sound perception. It is well documented that listeners with normal-hearing (NH) show robust speech intelligibility with partial speech information. For example, NH listeners can show remarkable speech identification performance with temporal envelope or temporal fine structure (TFS) information over few frequency bands [1,2]. When speech signals are passed through low-pass filters (LPFs) or high-pass filters (HPFs) to restrict the speech information in the limited frequency range, NH listeners can still integrate the partial speech information

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

for accurate identification [3-6]. The extent to which listeners can identify speech depends on the available spectrum range usually controlled by varying cutoff frequencies for LPFs and HPFs, and the degree of available sound information such as temporal envelope or TFS. Such information is useful to understand the importance of frequency bands for speech intelligibility for different clinical populations.

One of such populations is cochlear implant (CI) users. CI sound processors analyze the spectrum of the acoustic sound using a bank of band-pass filters (BPFs) or the fast Fourier transform. The number of analysis filters is generally associated with the number of physical stimulating electrodes. When the sound is passed through the set of BPFs, envelope components of sub-band signals are defined and compressed to compensate for the reduced dynamic range found in listeners with severe-to-profound hearing loss. The envelopes are then modulated with electrical pulse trains and sent to the corresponding stimulating electrode. Therefore, the CI signal processing itself presents significantly different frequency coding from the physiological processing found for NH listeners. Furthermore, CI sound processors do not efficiently transmit acoustic TFS cues, in which rich spectral information is embedded even though some CI products insist that they are transmitting some TFSs through several electrodes. Ardoint and Lorenzi [5] measured identification of consonants with stimuli that preserve temporal envelope alone, or TFS alone, or both information. Stimuli were passed through either LPFs or HPFs with varying cutoff frequencies to control the amount of possible phonetic cues in the restricted frequency regions. They demonstrated that both envelope and TFS cues convey distinct phonetic information for NH subjects. In particular, the frequency region between 1–2.5 kHz was found to present the most important phonetic cues.

Relevant questions about the use of restricted acoustic cues for speech perception have been addressed by previous studies. For example, Boothroyd, et al. [7] showed that vowels were more susceptible to the effects of spectral smearing than consonants for NH subjects. Xu and Pfungst [8] showed that vowel and consonant recognition for NH subjects listening to vocoder simulations can reach the maximum performance around 8–12 spectral channels. Nie, et al. [9] demonstrated that CI users may rely more on temporal cues for consonant identification, but spectral cues were found to be more important for vowel identification.

Identification of environmental sounds also depends on the accurate representations of spectral and temporal information. Warren and Verbrugge [10] demonstrated that a series of discrete impulses over time may be the primary cue for discriminating the sounds of breaking and bouncing events, while

spectral information may not be necessarily required for the perception of such environmental sounds. Gaver [11], however, demonstrated that the impacting wood and metal objects may be discriminated based on the spectral characteristics of the objects. Gygi, et al. [4] suggested that NH listeners may use similar, but not identical, spectro-temporal perceptual mechanisms for identification of speech and non-speech environmental sounds.

The goal of the present study was to evaluate the frequency-limiting effects on sound perception in CI and NH subjects. Comparison of speech identification performance between CI and NH subjects may provide a direct indication of the role of acoustic TFS in speech intelligibility. In addition, the present study employed both speech and non-speech stimuli. For speech stimuli, vowels and consonants of Korean were used and environmental sounds identification was also administered. Identification performance was measured with 7 different cutoff frequencies for LPFs and HPFs without further signal processing. Therefore, NH subjects received both envelope and TFS cues, whereas CI subjects mainly received envelope cues due to the inefficient coding of acoustic TFS through the CI sound processors. The use of multiple sound types is desirable to systematically explore the interaction between the sound type and frequency-limiting effects for CI and NH subjects.

## Subjects and Methods

### Subjects

Ten CI users (2 females, 8 males) participated in this study. The selection criteria for CI subjects included: 1) the onset of post-lingual profound hearing loss which could preclude any effects of language and developmental factors on sound recognition ability, 2) a pure tone threshold better than 40 dBHL in the frequency range between 250 Hz and 6,000 Hz to ensure proper functioning of CI devices, 3) 95% or greater open-set sentence recognition score, and 4) at least one year of duration of CI use prior to this study. These selection criteria generally provide high performing CI subjects with tightly controlled etiological and cognitive factors. Table 1 shows the demographic information for CI subjects. All CI subjects had fully inserted electrodes and used all available channels in their speech coding strategies. During testing, CI subjects used the most recent and adapted clinical maps that had been used for more than 6 months. No subject used a hearing aid in the contralateral ear during the experiment.

Ten NH subjects also participated in this study. NH subjects met the selection criteria of hearing thresholds better than 20 dB HL in each ear at 250, 500, 1,000, 2,000, 4,000, and 6,000 Hz. The mean age of NH subjects was 21 years with



was used to measure identification scores. For consonant identification, three different sets of testing were performed separately for the pre-vocalic consonants with an 18-AFC procedure, the inter-vocalic consonants with an 18-AFC procedure, and the post-vocalic consonants with a 7-AFC procedure. Each vowel and consonant was presented once for male and female speaker. Lastly, environmental sound identification was measured using a 9-AFC procedure. The order of the experimental conditions (vowels, consonants, environmental sounds) was randomized within and across subjects. For each stimulus type, types of filtering (original, full-band, low-pass, or high-pass filtering) and cutoff frequencies were also administered in random order within and across subjects.

### Data analysis

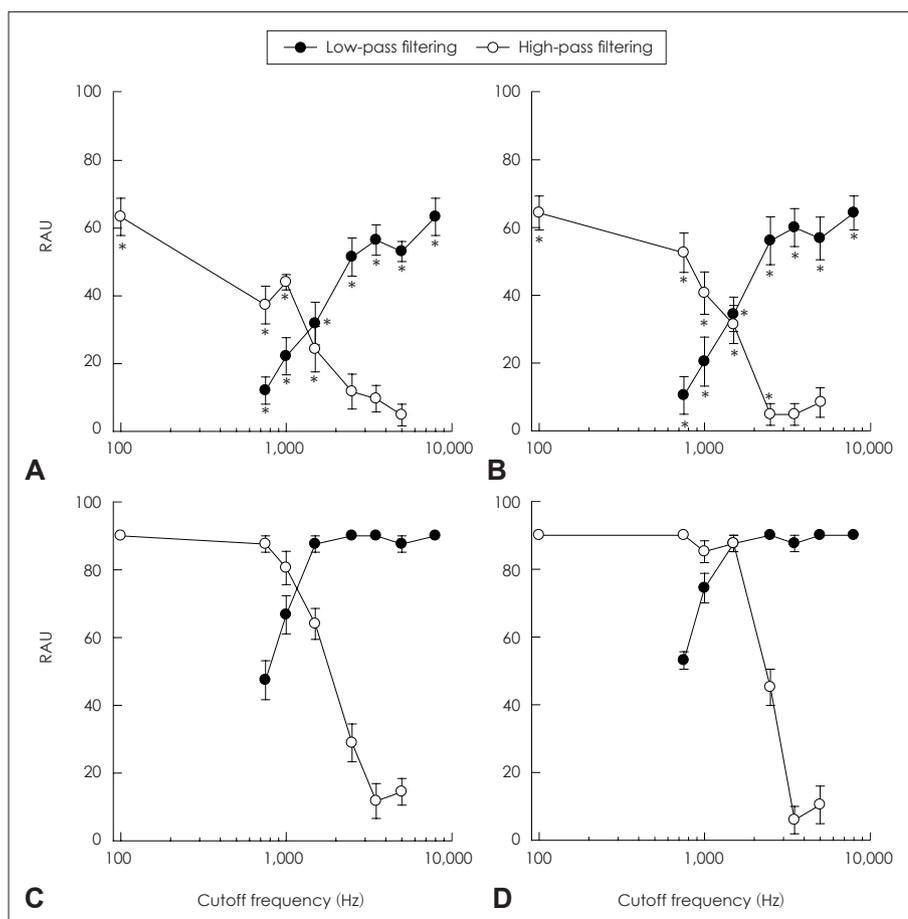
The total number of stimulus presentation for each subject was 1,526, creating a total number of 30,520 responses from all NH and CI subjects. Percent correct scores for identification tests were corrected for chance level and then transformed to rationalized arcsine units (RAUs). The rationalized arcsine transform [14] was applied to the proportions of correct responses in order to make the identification scores follow

a normal distribution. Therefore, a RAU value of 0 in this paper represents performance at chance level. Note that three different sets of consonant identification were performed separately. Identification scores (in chance-corrected RAU) for these three conditions were averaged to compute the final consonant identification score for each individual subject. All statistical analyses were performed using SAS (SAS Institute, Cary, NC, USA) based on the chance-corrected RAU scores.

## Results

Fig. 1 shows the mean vowel identification scores across CI subjects (upper panels) and across NH subjects (lower panels) as a function of cutoff frequency for male (left panels) and female (right panels) speakers. For the LPF conditions, identification performance increased as the cutoff frequency increased. For the HPF conditions, identification performance decreased as the cutoff frequency increased. Overall, NH subjects showed better identification performance than CI subjects. In the multiple independent samples t-tests between CI and NH subjects, the criterion for statistical significance was divided by 7 (i.e.,  $\alpha=0.007$ ) to correct for the number of com-

**Fig. 1.** Mean vowel identification scores across subjects in chance-corrected, RAUs. Filled and open circles represent the low-pass filtering and high-pass filtering conditions, respectively. Error bars represent  $\pm$  one standard errors about the mean across subjects. The asterisks under the scores in the upper panels show the significant difference in the performance ( $\alpha=0.007$ ) between CI and NH subjects based on multiple independent samples t-tests. A: CI subjects, male speaker. B: CI subjects, female speaker. C: NH subjects, male speaker. D: NH subjects, female speaker. CI: cochlear implant, NH: normal hearing, RAU: rationalized arcsine unit.



parisons. NH subjects showed significantly better identification performance than CI subjects which are shown by the asterisks under the data points except for the few testing conditions. NH subjects showed nearly perfect identification performance for the full-band condition. Here, 90 RAU indicates perfect identification performance. In addition, NH subjects showed nearly perfect identification at cutoff frequencies above 1,500 Hz for the LPF conditions, whereas CI subjects showed an improvement in identification performance beyond 1,500 Hz cutoff frequency. Likewise, for the HPF conditions, NH subjects showed nearly perfect identification (particularly for fe-

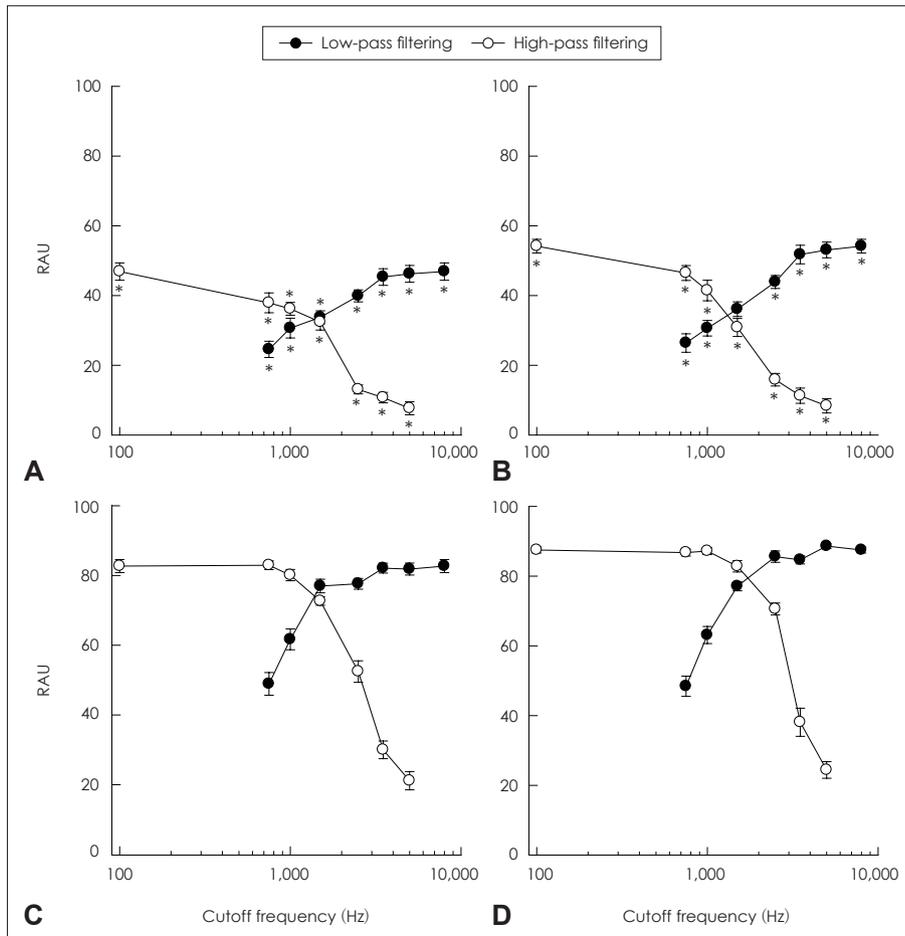
male speaker) up to 1,500 Hz cutoff frequency; and beyond it, identification performance dropped.

A 2×7 repeated measures analysis of variance (ANOVA) with the main factors of speakers (male vs. female) and cutoff frequencies (7 cutoffs for each LPF and HPF condition) was performed on CI and NH subjects' RAU scores (Table 2). This ANOVA analysis was performed separately for the LPF and HPF conditions, because the frequency range for each cutoff frequency differs for the two testing conditions (e.g., when a cutoff frequency was set to 1,000 Hz, the LPF condition covers 100–1,000 Hz, whereas the HPF condition covers 1,000–

**Table 2.** Results for the 2×7 repeated measures analysis of variance for vowel identification

Subject group	Main factor	Low-pass filtering		High-pass filtering	
		F-values	p-values	F-values	p-values
CI subjects	Speaker	0.34	0.57	0.4	0.54
	Cutoff	27.51	<0.001	46.59	<0.001
	Speaker×cutoff	0.14	0.99	1.34	0.25
NH subjects	Speaker	1.42	0.26	22.54	0.001
	Cutoff	57.58	<0.001	146.20	<0.001
	Speaker×cutoff	0.92	0.49	8.35	<0.001

CI: cochlear implant, NH: normal-hearing



**Fig. 2.** Mean consonant identification scores across subjects. Data format is identical to that of Fig. 1. The asterisks under the scores in the upper panels show the significant difference in the performance ( $\alpha=0.007$ ) between CI and NH subjects based on multiple independent samples t-tests. A: CI subjects, male speaker. B: CI subjects, female speaker. C: NH subjects, male speaker. D: NH subjects, female speaker. CI: cochlear implant, NH: normal-hearing, RAU: rationalized arcsine unit.

8,000 Hz). For both CI and NH subjects, and for both the LPF and HPF conditions, the main effect of cutoff frequency reached significance, indicating that vowel identification performance changed significantly with different frequency range available for listeners. The effect of speakers did not reach significance except for the HPF condition in NH subjects.

In consonant identification, overall, NH subjects also showed better identification performance than CI subjects. For the HPF conditions, identification performance was consistently high for cutoff frequencies below 1,000 Hz in NH subjects, but CI subjects showed further improvement for cutoff frequencies below 1,000 Hz. For the LPF conditions, identification performance appears to be maxed out at 3,500 Hz for both CI and NH subjects (Fig. 2). For both CI and NH subjects and for both the LPF and HPF conditions, the effects of speaker and cutoff frequency reached significance in consonant identification (Table 3).

Fig. 3 shows the mean environmental sound identification results. As with the vowel and consonant identification, environmental sound identification performance changed as a function of cutoff frequency (Table 4 for ANOVA results). For the HPF conditions, NH subjects showed perfect identifi-

cation performance for cutoff frequencies below 1,000 Hz. Above 1,000 Hz cutoff frequency, NH subjects showed nearly perfect identification performance for the LPF conditions. CI subjects showed a gradual change in identification performance for the HPF conditions as the cutoff frequency increased. For the LPF conditions, CI subjects showed the peak at 3,500 Hz, but performance for cutoff frequencies above 2,500 Hz was not significantly different each other (paired t-test,  $p>0.05$ ).

Pearson correlation coefficients were computed between vowel/consonant identification (averaged across male and female speakers) and environmental sound identification (Fig. 4). Significant correlations were found between speech and environmental sound identification for CI subjects ( $r=0.87, p<0.001$ ) and for NH subjects ( $r=0.88, p<0.001$ ).

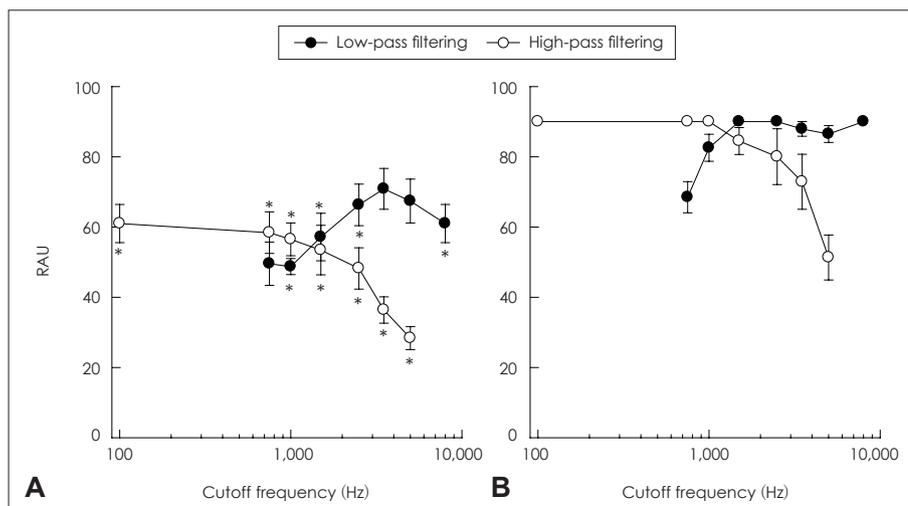
Two different psychometric curves were derived for the LPF and HPF conditions for each individual subject. The function was generated using the nonparametric local linear approach [15] with inputs of identification scores (in percent correct) and the base-2 logarithm of the cutoff frequencies, as used by Whitmal and DeRoy [6]. The intersection point between the psychometric functions for the LPF and HPF conditions

**Table 3.** Results for the 2×7 repeated measures analysis of variance for consonant identification

Subject group	Main factors	Low-pass filtering		High-pass filtering	
		F-values	p-values	F-values	p-values
CI subjects	Speaker	11.37	0.008	33.29	<0.001
	Cutoff	50.56	<0.001	114.68	<0.001
	Speaker×cutoff	2.37	0.042	70.86	0.02
NH subjects	Speaker	8.88	0.015	119.52	<0.001
	Cutoff	88.71	<0.001	287.56	<0.001
	Speaker×cutoff	2.39	0.041	6.14	<0.001

CI: cochlear implant, NH: normal-hearing

**Fig. 3.** Mean environmental sound identification scores across subjects. Data format is identical to that of Fig. 1, 2. The asterisks under the scores in the upper panels show the significant difference in the performance ( $\alpha=0.007$ ) between CI and NH subjects based on multiple independent samples t-tests. A: CI subjects. B: NH subjects. CI: cochlear implant, NH: normal-hearing, RAU: rationalized arcsine unit.



was assessed to determine the “crossover frequency.” The crossover frequency describes the frequency boundary, where above or below it, 50% of intelligible phonetic information is transmitted [16,17]. The crossover frequency is generally speech material dependent. Table 5 shows the crossover frequencies for vowel, consonant, and environmental sound identification.

CI subjects showed a minimal change in the mean crossover frequency between vowels and consonants or between male and female speakers. However, NH subjects showed larger differences in the mean crossover frequency between vowels and consonants or between male and female speakers. This pattern is confirmed by a 2×2 repeated measures ANOVA with factors of stimulus type (vowels and consonants) and speaker (male and female) performed separately for CI and NH subjects. For CI subjects, the effects of both factors did not reach significance at the 0.05 level [ $F(1,9)=1.61, p=0.24$  for stimulus type;  $F(1,9)=0.43, p=0.53$  for speaker]. However, NH subjects showed significant main effects of both factors on the crossover frequency [ $F(1,9)=24.73, p=0.001$  for stimulus type;  $F(1,9)=25.43, p=0.001$  for speaker].

Figs. 1–3 show that there was a change in identification performance between two adjacent cutoff frequencies. Here, gradients refer to the change in performance with respect to cutoff frequency. To examine the contribution of a frequency band to identification performance, gradients were computed for group data for the LPF and HPF conditions. For example, at the cutoff frequency 1,000 Hz, the gradient for the LPF

condition can be computed by subtracting the identification score at 750 Hz cutoff frequency (i.e., available frequency range: 100–750 Hz) from the identification score at 1,000 Hz (i.e., available frequency range: 100–1,000 Hz). Therefore, the gradient at 1,000 Hz for the LPF condition estimates the contribution of a frequency band between 750–1,000 Hz to the identification performance. All gradient values were computed based on RAU scores and they were normalized by the difference between two adjacent cutoff frequencies (on an equivalent rectangular bandwidth, or ERB<sub>N</sub> scale).

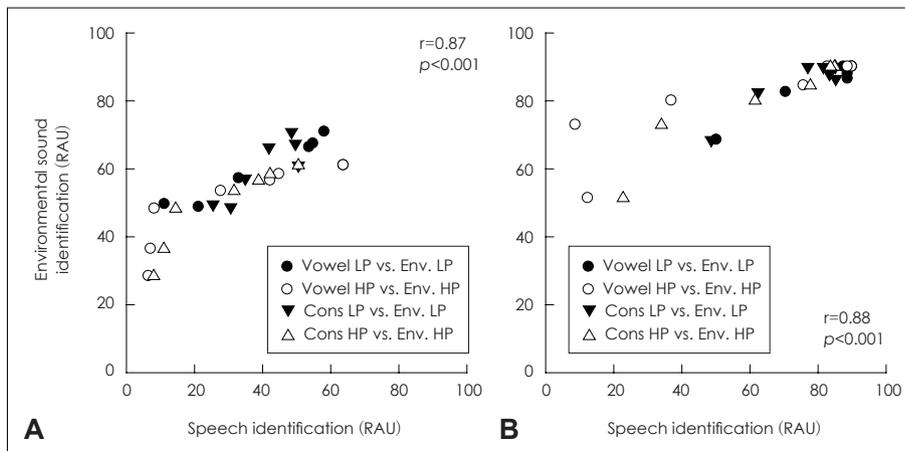
In Fig. 5, gradients for vowel, consonant, and environmental sound identification are plotted as a function of cutoff frequency. In general, NH subjects showed greater gradients than CI subjects, suggesting that NH subjects showed greater performance changes between two adjacent cutoff frequencies. For the LPF condition, the distributions of gradients were skewed to the low frequency side regardless of stimulus type. For the HPF condition, the peak of the distribution of gradients for CI subjects is located in the middle frequency side except for the environmental sounds. NH subjects, however, showed the peak of the distribution of gradients for the HPF conditions at frequencies around 2.5–3.5 kHz for vowels, 3.5 kHz for consonants, and 8 kHz for environmental sounds.

To capture the precise effects of each factor on the gradients, a 3×2×7 repeated measures ANOVA (3 stimulus types; 2 filtering conditions, 7 cutoff frequencies) was performed for CI and NH subjects separately. For CI subjects, all three factors had significant effects [ $F(2,18)=5.0, p=0.019$  for stimulus

**Table 4.** Results for the 2×7 repeated measures analysis of variance for environmental sound identification

Subject group	Main factor	Low-pass filtering		High-pass filtering	
		F-values	p-values	F-values	p-values
CI subjects	Cutoff	5.67	<0.001	11.54	<0.001
NH subjects	Cutoff	11.77	<0.001	12.53	<0.001

CI: cochlear implant, NH: normal-hearing



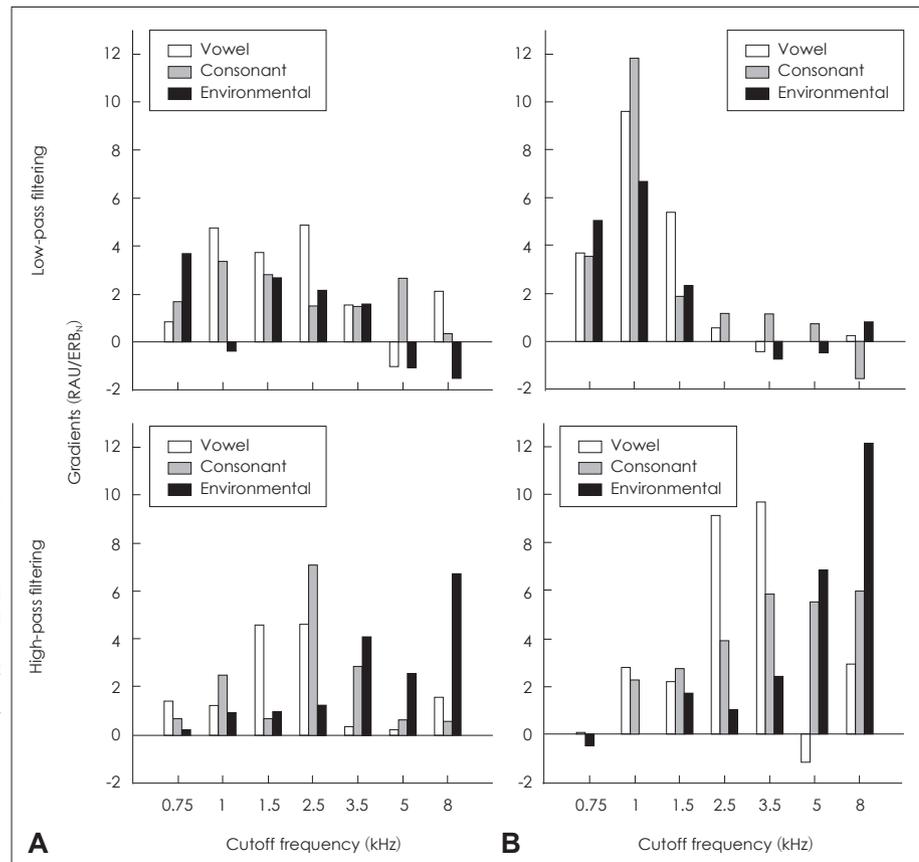
**Fig. 4.** Environmental sound identification scores plotted against speech identification scores (mean across vowel and consonant identification). Each symbol corresponds to the speech type and filtering conditions. The Pearson correlation coefficient is computed including all data points in each figure. A: CI subjects. B: NH subjects. CI: cochlear implant, NH: normal-hearing, RAU: rationalized arcsine unit.

**Table 5.** Crossover frequencies (in Hz) for vowel, consonant, and environmental sound identification

Subject group	Vowel identification				Consonant identification				Environmental sound identification	
	Male speaker		Female speaker		Male speaker		Female speaker		Range	Mean
	Range	Mean	Range	Mean	Range	Mean	Range	Mean		
CI subjects	841–2,005	1,283	1,058–1,905	1,363	759–1,621	1,209	768–1,816	1,248	1,024–2,019	1,295
NH subjects	768–1,501	1,109	1,118–1,838	1,515	1,305–1,681	1,498	1,408–2,179	1,716	1,024–2,150	1,505

CI: cochlear implant, NH: normal-hearing

**Fig. 5.** Mean gradients across subjects for the low-pass filtering (top panels) and high-pass filtering (bottom panels) filtering conditions. In each figure, gradients are plotted as a function of cutoff frequency for vowel (open bars), consonant (light-hatched bars), and environmental sound (dark-hatched bars) identification. Gradient values are shown in chance-corrected, RAUs, normalized by the  $ERB_N$  for the corresponding bandwidth between two adjacent cutoff frequencies. A: CI subjects. B: NH subjects. CI: cochlear implant, NH: normal-hearing, RAU: rationalized arcsine unit,  $ERB_N$ : equivalent rectangular bandwidth.



type;  $F(1,9)=10.27$ ,  $p=0.011$  for filtering conditions;  $F(6,54)=3.50$ ,  $p=0.005$  for cutoff frequency]. All two-way interactions also reached significance at the 0.05 level, but three-way interaction did not reach significance [ $F(12,108)=1.35$ ,  $p=0.20$ ]. For NH subjects, all three factors also had significant effects [ $F(2,18)=7.59$ ,  $p=0.004$  for stimulus type;  $F(1,9)=134.47$ ,  $p<0.001$  for filtering conditions;  $F(6,54)=4.85$ ,  $p=0.001$  for cutoff frequency]. All two-way as well as three-way interactions reached significance.

## Discussion

In the present study, both CI and NH subjects were presented with sounds that were passed through either low-pass or HPFs. The filtered signals were not further processed using any vocoder processing. Therefore, NH subjects received

both temporal envelope and fine structure information in the restricted frequency region. CI subjects, however, mainly received the temporal envelope information while the encoding of acoustic TFS information was very limited due to the use of fixed-rate, pulsatile stimulation. For the most conditions in the present study, CI subjects showed significantly worse identification scores than NH subjects. There were just few conditions where there was not a significant difference in identification performance between the two groups. For example, when vowel formants in the low frequency region were discarded by the HPF with cutoff frequency above 2,500 Hz, both groups equally performed worse.

Ardoit and Lorenzi [5] also demonstrated differences in consonant identification between the stimuli having both envelope and TFS information (i.e., intact speech condition in their study) and stimuli preserving envelope information over

16 frequency bands (i.e., E-speech condition in their study by using 16-band tone envelope vocoders). The magnitude of difference in performance between CI and NH subjects in the present study appears to be greater than the magnitude of difference in performance between the intact and E-speech conditions in Ardoint and Lorenzi [5]. This may be partly due to the fact that the baseline performance for the full-band condition was better for NH subjects presented with E-speech stimuli in Ardoint and Lorenzi [5] compared to CI subjects in the present study. That is, at the full-band condition, NH subjects in Ardoint and Lorenzi [5] showed nearly perfect identification performance with E-speech stimuli for a male speaker, whereas CI subjects in the current study showed 55% of identification performance for consonants produced by a male speaker.

Shafiro [18] tested NH subjects on an environmental sound identification test in a 60-alternative forced choice paradigm, and Shafiro, et al. [19] administered the same environmental sound identification test for CI subjects. CI subjects in Shafiro, et al. [19] showed average identification performance of 45.3%. When NH subjects were presented with spectrally degraded stimuli (8-, 16-, and 24-channel noise vocoders), they showed average performance between 60–70% [17], which is apparently higher performance than CI subjects tested on the same environmental identification task. Taken together, these previous studies along with the present study highlight the difference in speech and environmental sound identification between CI and NH subjects. The difference in performance between the two subject groups may be attributed to multiple factors such as the inefficient coding of acoustic TFS information and poor spectral resolution due to the limited numbers of stimulating electrodes and channel interactions.

The range of mean crossover frequencies for vowel and consonant identification for NH subjects in this study were generally consistent with previously reported values using English [20–22], Cantonese [23], or Korean [24]. It is an interesting finding that the crossover frequency in the environmental sound identification is similar to that in the speech sound identification in this study. The significant effect of speaker on the crossover frequency for NH subjects is consistent with the results of Ardoint and Lorenzi [5] who also reported similar significant effects of speaker (male vs. female) on the crossover frequency for NH subjects listening to intact or vocoded consonants. The non-significant effects of stimulus type and speaker on the crossover frequency for CI subjects may suggest that CI subjects are generally less sensitive to the changes in acoustic stimuli. In addition, the difference in the mean crossover frequencies between CI and NH subjects may reflect the fact that the frequency-place map for electric and acoustic hearing

differ due to multiple factors such as the frequency allocation of BPFs associated with stimulating electrodes, the actual position of electrodes in the cochlea, and spectral smearing effects by the spread of excitation [25–28].

Vowels were more susceptible for frequency information than consonants in Korean, which is consistent to the results of Boothroyd, et al. [7] For vowels and consonants, 1–1.5 kHz is the crucial area for providing phonetic information in NH subjects. Both group performed similarly poor on vowel identification with the HPF for cutoff frequencies above 2.5 kHz. CI subjects, however, were able to utilize frequency information mainly up to 2.5 kHz especially for consonant identification. This is interesting because CI subjects showed significantly poor performances with cutoff frequencies above 2.5 kHz compared to NH subjects in consonant identification. This means that NH subjects still utilize the acoustic information above 2.5 kHz for consonant identification, but this was not the case for the CI subjects although they need higher frequency information to identify consonants.

The significant relationship between speech and non-speech environmental sound identification shown in Fig. 4 suggests that speech and environmental sounds may share similar pattern of perception and acoustic features [4,29]. For example, both speech and environmental sound generally require fine spectral resolution for better identification outcomes particularly for CI users [8,9,19,30]. Shafiro [18] demonstrated that environmental sound identification improved with increasing spectral resolution in NH subjects listening to vocoded stimuli. Using 17 CI users, Shafiro, et al. [31] reported significant correlations between environmental sound identification and sentence, consonant, monosyllabic word recognition in quiet, as well as sentence recognition in background noise. Therefore, on the one hand, a similar perceptual mechanism might drive identification performance for speech and non-speech environmental sounds. This might be one of the reasons that the crossover frequency is similar between speech sounds and environmental sounds.

On the other hand, however, gradients were significantly different across stimulus type (vowels, consonants, environmental sounds), filtering conditions (LPF vs. HPF), and cutoff frequencies both for CI and NH subjects. This may suggest that the frequency-limiting effects may be differential across listening tasks. From the acoustic perspective, this is due to the fact that as a function of frequency band, the amount of useful acoustic information for identification task may differ across vowels, consonants, and environmental sounds. For example, as the cutoff frequency for the HPF increased, important linguistic information such as formants was not transmitted to subjects that are important for vowel identification; hence,

vowel identification performance markedly decreased.

However, listeners can utilize both temporal and spectral acoustic information for environmental sound identification. In fact, previous studies have suggested that listeners may put more emphasis on the temporal features of environmental sounds such as attack and decay characteristics [4,10,29], thus the net effect of frequency-limiting may be less for environmental sound identification than for vowel or consonant identification. This is supported by the fact that the magnitude of difference in identification performance between the lowest and highest cutoff frequencies was smaller for environmental sound identification than for speech identification (CI subjects: 11.3 RAU and 32.6 RAU differences for the LPF and HPF conditions, respectively; NH subjects: 21.5 RAU and 38.7 RAU differences for the LPF and HPF conditions, respectively). In addition, performance at the most extreme cutoff frequencies such as 750 Hz for the LPF, and 5,000 Hz for the HPF, environmental sound identification performance was still fairly good for both CI (50 RAU at 750 Hz cutoff for the LPF; 28 RAU at 5,000 Hz cutoff for the HPF) and NH (68 RAU at 750 Hz cutoff for the LPF; 51 RAU at 5,000 Hz cutoff for the HPF) subjects, whereas speech identification at these extreme cutoff frequencies were very low, particularly in CI subjects. These results suggest that, although the overall pattern of identification performance as a function of cutoff frequency for the LPF and HPF conditions changed in a similar manner for speech and non-speech environmental sounds, listeners showed more robust identification performance for environmental sounds than for speech.

The effects by spectral and temporal features in sound need to be investigated further. This study also suggested that environmental sound discrimination would be considered more to evaluate perception status of the CI users and/or to bridge the sound perception training stages in more discrete steps to get better speech perception. There were not many findings demonstrating Korean-specific speech perception characteristics in the current study so that there should be more studies to reveal language-specific perceptual features which might improve sound processing for the CI users.

In conclusion, The present study demonstrated substantial differences were found in overall identification performance, crossover frequencies, and gradients between CI and NH subjects, indicating the alteration of acoustic information through the CI processing. The approach in this study may provide more detailed analysis of the effects of novel CI sound processing strategies or mapping parameters on speech or environmental sound identification, because it enables to reveal the potential advantages or disadvantages of CI sound processing strategies in restricted frequency bands. Typical speech

identification tests, where the entire frequency information is presented altogether, are not feasible to provide such information. This point may be particularly important when evaluating CI performance for different languages.

### Acknowledgments

This study is a part of doctoral dissertation of the first author. It is dedicated to the late Dr. Won-Sang Lee who were a keen and considerate scientist as well as a passionate doctor and had enormously contributed to this study.

### Conflicts of interest

The authors have no financial conflicts of interest.

### REFERENCES

- 1) Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. *Science* 1995;270:303-4.
- 2) Lorenzi C, Husson M, Ardoit M, Debruelle X. Speech masking release in listeners with flat hearing loss: effects of masker fluctuation rate on identification scores and phonetic feature reception. *Int J Audiol* 2006;45:487-95.
- 3) Studebaker GA, Sherbecoe RL. Frequency-importance and transfer functions for recorded CID W-22 word lists. *J Speech Hear Res* 1991; 34:427-38.
- 4) Gygi B, Kidd GR, Watson CS. Spectral-temporal factors in the identification of environmental sounds. *J Acoust Soc Am* 2004;115:1252-65.
- 5) Ardoit M, Lorenzi C. Effects of lowpass and highpass filtering on the intelligibility of speech based on temporal fine structure or envelope cues. *Hear Res* 2010;260:89-95.
- 6) Whitmal NA 3rd, DeRoy K. Adaptive bandwidth measurements of importance functions for speech intelligibility prediction. *J Acoust Soc Am* 2011;130:4032-43.
- 7) Boothroyd A, Mulhearn B, Gong J, Ostroff J. Effects of spectral smearing on phoneme and word recognition. *J Acoust Soc Am* 1996; 100:1807-18.
- 8) Xu L, Pfingst BE. Spectral and temporal cues for speech recognition: implications for auditory prostheses. *Hear Res* 2008;242:132-40.
- 9) Nie K, Barco A, Zeng FG. Spectral and temporal cues in cochlear implant speech perception. *Ear Hear* 2006;27:208-17.
- 10) Warren WH Jr, Verbrugge RR. Auditory perception of breaking and bouncing events: a case study in ecological acoustics. *J Exp Psychol Hum Percept Perform* 1984;10:704-12.
- 11) Gaver WW. Auditory icons: using sound in computer interfaces. *Hum Comput Interact* 1986;2:167-77.
- 12) Kim BJ, Chang SA, Yang J, Oh SH, Xu L. Relative contributions of spectral and temporal cues to Korean phoneme recognition. *PLoS One* 2015;10:e0131807.
- 13) Chang SA, Tyler RS, Dunn CC, Ji H, Witt SA, Gantz B, et al. Performance over time on adults with simultaneous bilateral cochlear implants. *J Am Acad Audiol* 2010;21:35-43.
- 14) Studebaker GA. A "rationalized" arcsine transform. *J Speech Hear Res* 1985;28:455-62.
- 15) Zychaluk K, Foster DH. Model-free estimation of the psychometric function. *Atten Percept Psychophys* 2009;71:1414-25.
- 16) Hirsh IJ, Reynolds EG, Joseph M. Intelligibility of different speech materials. *J Acoust Soc Am* 1954;26:530-537.
- 17) Studebaker G, Sherbecoe R. Frequency importance functions for speech recognition. In: Studebaker GA, Hochberg I, editors. *Acoustical factors affecting hearing aid performance*. 2nd ed. Needham Heights, MA: Allyn & Bacon;1993. p.185-204.
- 18) Shafiro V. Development of a large-item environmental sound test and the effects of short-term training with spectrally-degraded stim-

- uli. *Ear Hear* 2008;29:775-90.
- 19) Shafiro V, Gygi B, Cheng MY, Vachhani J, Mulvey M. Factors in the perception of environmental sounds by patients with cochlear implants. [cited 2010 Oct 20]. Available from: <https://www.audiology-online.com/articles/factors-in-perception-environmental-sounds-879>.
  - 20) French NR, Steinberg JC. Factors governing the intelligibility of speech sounds. *J Acoust Soc Am* 1947;19:90-119.
  - 21) Warren RM, Riener KR, Bashford JA Jr, Brubaker BS. Spectral redundancy: intelligibility of sentences heard through narrow spectral slits. *Percept Psychophys* 1995;57:175-82.
  - 22) Studebaker GA, Pavlovic CV, Sherbecoe RL. A frequency importance function for continuous discourse. *J Acoust Soc Am* 1987;81:1130-8.
  - 23) Wong LL, Ho AH, Chua EW, Soli SD. Development of the Cantonese speech intelligibility index. *J Acoust Soc Am* 2007;121:2350-61.
  - 24) Kim YW, Choi DL, Lee SH, Lee YJ. Perceptual characteristics of Korean vowels distorted by the frequency band limitation. *Phon Speech Sciences* 2014;6:85-93.
  - 25) Fu QJ, Shannon RV. Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing. *J Acoust Soc Am* 1999;105:1889-900.
  - 26) Baskent D, Shannon RV. Speech recognition under conditions of frequency-place compression and expansion. *J Acoust Soc Am* 2003;113(4 Pt 1):2064-76.
  - 27) Bařkent D, Shannon RV. Frequency-place compression and expansion in cochlear implant listeners. *J Acoust Soc Am* 2004;116:3130-40.
  - 28) Bařkent D, Shannon RV. Interactions between cochlear implant electrode insertion depth and frequency-place mapping. *J Acoust Soc Am* 2005;117(3 Pt 1):1405-16.
  - 29) Reed CM, Delhorne LA. Reception of environmental sounds through cochlear implants. *Ear Hear* 2005;26:48-61.
  - 30) Goupell MJ, Laback B, Majdak P, Baumgartner WD. Effects of upper-frequency boundary and spectral warping on speech intelligibility in electrical stimulation. *J Acoust Soc Am* 2008;123:2295-309.
  - 31) Shafiro V, Sheft S, Risley R. Perception of interrupted speech: effects of dual-rate gating on the intelligibility of words and sentences. *J Acoust Soc Am* 2011;130:2076-87.