

Published in final edited form as:

*Int J Audiol.* 2011 August ; 50(8): 498–502. doi:10.3109/14992027.2011.576274.

## Voice gender discrimination provides a measure of more than pitch-related perception in cochlear implant users

Tianhao Li and Qian-Jie Fu

Division of Communication and Auditory Neuroscience, House Ear Institute, Los Angeles, USA

### Abstract

**Objectives**—(1) To investigate whether voice gender discrimination (VGD) could be a useful indicator of the spectral and temporal processing abilities of individual cochlear implant (CI) users; (2) To examine the relationship between VGD and speech recognition with CI when comparable acoustic cues are used for both perception processes.

**Design**—VGD was measured using two talker sets with different inter-gender fundamental frequencies (F0), as well as different acoustic CI simulations. Vowel and consonant recognition in quiet and noise were also measured and compared with VGD performance.

**Study sample**—Eleven postlingually deaf CI users.

**Results**—The results showed that (1) mean VGD performance differed for different stimulus sets, (2) VGD and speech recognition performance varied among individual CI users, and (3) individual VGD performance was significantly correlated with speech recognition performance under certain conditions.

**Conclusions**—VGD measured with selected stimulus sets might be useful for assessing not only pitch-related perception, but also spectral and temporal processing by individual CI users. In addition to improvements in spectral resolution and modulation detection, the improvement in higher modulation frequency discrimination might be particularly important for CI users in noisy environments.

### Keywords

Voice gender discrimination; Cochlear implant

---

## INTRODUCTION

Natural speech utterances convey not only linguistic meaning, but also indexical information, such as the talker's gender or identity. Previous studies of normal-hearing (NH) listeners showed that fundamental frequency (F0) and formant structure are major factors for voice gender discrimination (VGD), and that F0 is somewhat more important than formant structure (Fellowes et al, 1997; Bachorowski & Owren, 1999; Hillenbrand & Clark, 2009). NH listeners can almost perfectly identify voice gender of natural speech. However, due to limited spectral and temporal resolution provided by cochlear implants (CIs), VGD performance of CI users is largely affected by the inter-gender F0 difference. For example, mean VGD performance by CI users was ~94% correct when mean F0 difference between

---

Correspondence: Tianhao Li, Division of Communication and Auditory Neuroscience, House Ear Institute, 2100 West Third Street, Los Angeles, CA 90057, USA. tianhaol@gmail.com.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

male and female talkers was 100 Hz, and ~68% correct when mean F0 difference was 10 Hz (Fu et al, 2005). Furthermore, studies using NH subjects listening to acoustic CI simulations, which may provide different spectral and temporal cues, showed that both spectral and temporal cues support VGD (Fu et al, 2004, 2005; Gonzalez & Oliver, 2005). Temporal cues provided enough information when the inter-gender F0 difference was large, and more than ten spectral channels could support good VGD without temporal cues. However, individual CI users differ in their spectral and temporal processing abilities (i.e. functional spectral resolution, spectral distortion, temporal modulation detection), so individual VGD performance may vary for different acoustic CI simulations. Thus, VGD performance of CI users tested with different talker sets and acoustic CI simulations may be useful for assessing individual performance, not only in pitch-related perception, but also in spectral and temporal processing.

For CI users, spectral and temporal processing abilities were shown to correlate with speech recognition in quiet and noise (Fu, 2002; Henry et al, 2005; Litvak et al, 2007; Won et al, 2007); music perception was also shown to correlate with speech recognition in noise (Gfeller et al, 2007; Won et al, 2010). Should VGD be a good indicator of CI users' pitch-related perception, and spectral and temporal processing abilities, then VGD may also correlate with speech recognition for CI users. However, previous studies showed dissociation between speech recognition and perception of indexical information in CI users (Fu et al, 2004; Vongphoe & Zeng, 2005; Dorman et al, 2008). For example, Vongphoe and Zeng found that CI users achieved good performance in vowel recognition, but poor performance in talker identification. However, VGD performance of CI users was better than speech recognition performance when the F0 s of male and female talkers were well separated. The primary reason for this dissociation is that, in contrast to speech recognition, different acoustic cues (i.e. spectro-temporal fine structure) are needed for talker identification, and optional acoustic cues (i.e. temporal cues) are available for VGD. It is still unknown whether speech recognition and VGD are correlated in CI users when comparable acoustic cues (i.e. spectral envelope, spectro-temporal envelope) are used for both perceptual processes.

The present study measured VGD performance by CI users, using two talker sets with different inter-gender F0 differences, as well as different acoustic CI simulations; and compared VGD performance with vowel and consonant recognition performance in quiet and in noise. The results may provide insights into the feasibility of using VGD to assess the spectral and temporal processing abilities of individual CI users. Furthermore, the relationship between speech recognition performance and VGD performance might provide a better understanding of speech perception and pitch-related perception in CI users.

## Methods

### Subjects

The subjects included eleven CI users (Table 1). All subjects were native speakers of American English and postlingually deaf CI users. All subjects gave informed consent before the study. During testing, subjects were seated in a sound-treated booth and listened to the speech tokens via a loudspeaker. The presentation level of speech tokens was 65 dBA. Subjects wore their regular, clinically assigned, speech processors. Three subjects who also wore hearing aids (HA) in another ear in daily life were not allowed to use the HA during the testing. Their unaided pure-tone average (PTA) were 70 dB HL. Voice gender discrimination

Speech stimuli for VGD measurements were twelve medial vowels presented in the format of h-/V/-d (i.e. had, head, hod, hawed, hayed, heard, hid, heed, hud, hood, who'd, hoed).

Two talker sets (five males and five females in each set), adopted from Fu et al (2005), were used for VGD measurements. In talker set 1, the difference between mean F0 s of male and female talkers was 100 Hz; in talker set 2, the difference was 10 Hz (see Fu et al, 2005, for details). Vowel tokens were selected from the database recorded by Hillenbrand (1995). The inter-gender F0 difference is ~90 Hz across all vowels in this database.

VGD by each subject was measured with four vowel stimulus sets: (1) unprocessed talker set 1; (2) talker set 1 processed by a 16-channel sine-wave vocoder with a 50 Hz temporal envelope cutoff frequency, removing temporal periodicity cues; (3) talker set 1 processed similarly to (2), but with a 200 Hz temporal envelope cutoff frequency, preserving temporal periodicity cues; (4) unprocessed talker set 2. The structure of the 16-channel sine-wave vocoder was similar to that used by Li et al (2009), but with different temporal envelope cutoff frequencies. The acoustic input frequency range of the vocoder was 200–7000 Hz. The corner frequencies of the 16 analysis filters were selected to give each channel similar spatial extension in the cochlea, based on Greenwood's function (1990). The center frequencies/carriers were calculated as the arithmetic center of the analysis bands. The CI users did not subjectively perceive any obvious difference in sound quality between unprocessed and vocoder-processed speech tokens.

VGD was measured using the 2-alternative forced choice (2-AFC) paradigm. In each trial, a vowel token from the stimulus set (12 vowels  $\times$  10 talkers = 120 vowel tokens) was randomly chosen and presented without repetition. The subject responded by clicking on one of the two response buttons, labeled 'male' and 'female'. No feedback was provided. The presentation order of all conditions was randomized.

### Speech recognition

Speech recognition test materials included vowels and consonants. The twelve medial vowels were the same as those used in VGD, but generated by eight other talkers (4 males and 4 females). There were 96 vowel tokens (12 vowels  $\times$  8 talkers) in each trial. The consonant stimuli consisted of 20 consonants presented in the context of a-/C/-a (i.e. aba, aga, ada, aza, asha, aya, ala, ata, aja, ava, awa, acha, atha, afa, apa, aka, ara, ama, ana, asa). There were 100 total consonant tokens in each trial, all digitalized natural productions of three female and two male talkers. The consonant tokens were from the database recorded by Shannon et al (1999). Vowel and consonant recognition by each subject was measured in quiet and in speech shaped static noise at a 5 dB signal-to-noise ratio (SNR). The SNR was the difference between the long-term RMS levels of the speech signal and the noise. The background noise commenced 100 ms before and ended 100 ms after the speech signal. In each trial, a test token from the stimulus set was randomly chosen and presented to subjects without repetition. Subjects responded by clicking on one of twelve (for vowels) or twenty (for consonants) response buttons, so that vowel recognition was measured by a 12-AFC paradigm, and consonant recognition was measured with a 20-AFC paradigm. The response buttons were labeled with 'h/V/d' for vowel recognition test, or 'a/C/a' for consonant recognition test. No feedback was provided. The presentation order of all conditions was randomized.

### Results

Figure 1(A) shows individual and mean VGD performance for different stimulus sets. There was a large inter-subject variability under all test conditions. Mean VGD performance was highest for unprocessed talker set 1, was lower for the vocoder-processed condition with a 200 Hz cutoff frequency, and was even lower for the vocoder-processed condition with a 50 Hz cutoff frequency. Variability among subjects was higher for the vocoder-processed conditions than for unprocessed talker set 1. Mean VGD performance was lowest for

unprocessed talker set 2. One-way repeated measures (RM) ANOVA revealed that VGD performance varied significantly among different test conditions [ $F(3, 30)=51.728$ ,  $p<0.001$ ]. Post-hoc Bonferroni t-tests showed significant differences between any two conditions except between vocoder-processed talker set 1 with a 50 Hz cutoff frequency and unprocessed talker set 2 ( $p=0.263$ ). Figure 1(B) shows vowel and consonant recognition scores in quiet and noise. Similar to VGD, there was a large inter-subject variability in speech recognition performance under all test conditions, especially under noise backgrounds. One-way RM ANOVA revealed that vowel and consonant recognition differed significantly between quiet and noise backgrounds ( $[F(1, 10)=16.078$ ,  $p=0.002]$  for vowels;  $[F(1, 10)=77.935$ ,  $p<0.001]$  for consonants). Correlations between VGD and speech recognition performance by individual subjects were examined using Pearson's correlation coefficients and statistical significance was accepted at  $p=0.0032$  after Bonferroni multiple testing corrections. The overall significance level was 0.05. Table 2 shows the correlation coefficients and p-values between VGD for different stimulus sets and vowel or consonant recognition. Significant correlations were found in several comparisons. Vowel recognition in quiet was significantly correlated with VGD tested with vocoder-processed talker set 1 with a 50 Hz cutoff frequency ( $p=0.0025$ ); consonant recognition in quiet was significantly correlated with VGD tested with vocoder-processed talker set 1 with a 200 Hz cutoff frequency ( $p=0.0016$ ); and vowel and consonant recognition in noise were strongly correlated with VGD tested with unprocessed talker set 2 ( $p<0.001$  for vowels;  $p=0.0016$  for consonants).

## Discussion

Due to the limited spectral resolution provided by CI, temporal modulation processing (i.e. modulation detection and modulation frequency discrimination) is important for VGD by CI users. Most of the tested CI users achieved a high level of VGD performance with unprocessed talker set 1, in which the  $F_0$ s ranged from 100 to 150 Hz across five male talkers, and from 200 to 250 Hz across five female talkers (Fu et al, 2005). The result is not surprising, given that mean Weber fractions in CI users were less than ~20% at the temporal modulation frequencies of 100–250 Hz (Chatterjee & Peng, 2008). In contrast, the  $F_0$ s of males and females in talker set 2 overlapped within the 150–200 Hz range, and the inter-gender  $F_0$  difference was only 10 Hz (Fu et al, 2005). Mean  $F_0$  is ~130 Hz for natural male talkers and ~220 Hz for natural female talkers, with a median ~170–180 Hz. In addition to temporal envelope cues, spectral cues or formant structures are critical for recognizing female talkers with  $F_0=180$  Hz, and male talkers with  $F_0=170$  Hz (Li & Fu, 2010). In the present study, therefore, both spectral processing and temporal modulation processing were important for VGD with talker set 2. VGD performance was significantly lower with unprocessed talker set 2 than with unprocessed talker set 1. There was also a larger range among subjects with talker set 2, suggesting that individual CI users differ in their spectral processing abilities and/or temporal modulation processing abilities, especially in temporal modulation frequency discrimination abilities within the range of 150–200 Hz.

Mean VGD performance with vocoder-processed talker set 1 with a 50 Hz cutoff frequency was relatively high, but significantly lower than that with unprocessed talker set 1, suggesting that most of the tested CI users appeared to be able to use spectral cues for VGD, but were limited by their spectral processing ability. Therefore, VGD performance with vocoder-processed talker set 1 with a 50 Hz cutoff frequency could be used to measure the spectral processing ability of individual CI users. Mean VGD performance was significantly higher for vocoder-processed talker set 1 with a 200 Hz cutoff frequency, indicating that CI users' VGD improved with more temporal periodicity cues. However, the improvement varied among subjects. For example, VGD performance of subject S4 improved from 63% to 87% correct, while VGD performance of subject S6 only improved from 85% to 90%

correct. Individual differences in improvement may reflect differences in the ability to use spectral and temporal cues. Mean VGD performance with vocoder-processed talker set 1 with a 200 Hz cutoff frequency was significantly lower than that with unprocessed talker set 1 and avoided the ceiling effect. Analysis based on individual talkers showed that the major performance difference between the two conditions came from five female talkers. The low-pass filter with a 200 Hz cutoff frequency might have compressed the temporal envelope modulation depth of the five female talkers, affecting temporal modulation detection. Thus, VGD performance measured with vocoder-processed talker set 1 with 50 and 200 Hz temporal cutoff frequencies could provide a measure of the spectral processing and temporal modulation detection abilities of individual CI users.

In agreement with previous studies, results of the present study showed that there was large inter-subject variability in speech recognition. The range of phoneme recognition performance in quiet was 66 to 96% correct. The decline in performance between speech recognition in quiet and in noise also showed large inter-subject variability. For example, subjects S2, S5, S7, S9, and S10 showed little difference in performance, especially for vowel recognition, while speech recognition performance of subjects S1 and S8 declined more than 30% in noise. These results suggest that speech recognition by CI users is affected by more factors in noise than in quiet and/or that more spectro-temporal fine information is needed for speech recognition in noise.

Despite the high inter-subject variability, VGD and speech recognition were strongly correlated under certain conditions. Vowel recognition in quiet was significantly correlated with VGD tested with vocoder-processed talker set 1 with a 50 Hz cutoff frequency, because both processes are primarily related to spectral processing. Consonant recognition in quiet was significantly correlated with VGD tested with vocoder-processed talker set 1 with a 200 Hz cutoff frequency; consonant recognition depends heavily on temporal envelope cues, and VGD under these conditions also relies on temporal modulation detection. Modulation detection was correlated with consonant recognition in quiet by English-speaking and Mandarin-speaking CI users (Fu, 2002; Luo et al, 2008).

While pitch discrimination and speech perception in noise involve different neural processes, results of the present study surprisingly showed that VGD performance by the eleven subjects tested with talker set 2 was significantly correlated with all phoneme recognition performances in noise. The correlation can be contributed to several factors.

First, with the small inter-gender F0 difference in talker set 2, spectral processing abilities become more important for VGD by CI users, especially functional spectral resolution. A previous study (Fu et al, 2005) using an acoustic CI simulation has showed that, without temporal cues, VGD performance with talker set 2 increased from the chance level to ~80% correct when the simulated spectral channels increased from 4 to 16. Similarly, vowel/consonant recognition in noise improved significantly when the number of spectral channels increased from 4 to ~12–16 (Xu & Zheng, 2007). Thus, individual CI users differ in their functional spectral resolution, resulting in a strong correlation. However, functional spectral resolution was not the only factor contributing to the strong correlation, because phoneme recognition in noise and VGD performance tested with vocoder-processed talker set 1 with a 50 Hz cutoff frequency were not strongly correlated.

Second, in addition to differences in functional spectral resolution, modulation detection and modulation frequency discrimination abilities within the 150–200 Hz range may also have influenced individual VGD performance with talker set 2. Because the correlation between phoneme recognition in noise and VGD performance, tested with vocoder-processed talker set 1 with a 200 Hz cutoff frequency, was relatively weaker, hence, the other important

factor might be the modulation frequency discrimination ability. Moreover, individual CI users showed a large range of frequency difference limens relative to the 10 Hz inter-gender difference at the temporal modulation frequencies between 150–200 Hz (Chatterjee & Peng, 2008). While temporal modulation frequencies beyond 16 Hz would not provide more acoustic information for vowel/consonant recognition in noise (Xu & Zheng, 2007), the temporal processing mechanisms underlying higher temporal modulation frequencies discrimination may affect speech recognition in noise in electrical hearing. In fact, recent evidence demonstrated the role of higher modulation frequencies in vocoded-speech recognition in noise (Stone et al, 2008; Chen & Loizou, 2011).

Third, the relationship between VGD with talker set 2 and speech recognition in noise might also be affected by individual differences in higher-level central patterns. In electric hearing, VGD and speech recognition are strongly affected by acoustic cues available via CI, however, pattern recognition processes also mediate VGD and speech recognition performances. All eleven subjects in the present study were postlingually deaf CI users, and might have formulated their central patterns before they had significant hearing loss. The effect of central patterns formulated with different experiences might have been trivial for the present study. Finally, speech processor strategies used by different CI devices may provide different acoustic cues. For example, two of the subjects used CI devices with HiRes strategy, which might provide more temporal cues with high-rate stimulation, facilitating VGD. However, the effect might have made a minor contribution to individual differences in the present study.

In summary, VGD measured with different talker sets and acoustic CI simulations appears to be useful for assessing not only pitch-related perception of CI users, but also individual ability to process spectral and temporal information. Furthermore, significant correlations between VGD performance and speech recognition under certain conditions suggest that the improvements of spectral resolution, modulation detection, and higher modulation frequency discrimination are important for CI users.

## Acknowledgments

The authors thank all subjects for their time and attention, and Sandy Oba and Jintao Jiang for their help in recruiting CI subjects. The authors thank the editors and two anonymous reviewers for their insightful comments on the earlier version of this paper. NIH/NIDCD (R01-DC004993 and R01-DC004792) supported the present project.

## Abbreviations

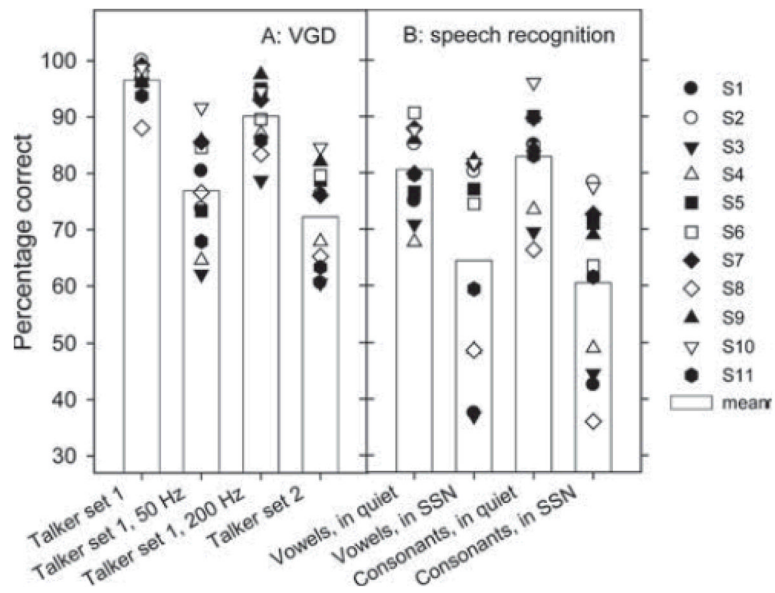
<b>ACE</b>	Advanced combination encoder
<b>AFC</b>	Alternative forced choice
<b>CI</b>	Cochlear implant
<b>CIS</b>	Continuous interleaved sampling
<b>NH</b>	Normal-hearing
<b>SPEAK</b>	Spectral peak
<b>VGD</b>	Voice gender discrimination

## References

- Bachorowski JA, Owren M. Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech. *J Acoust Soc Am.* 1999; 106:1054–1063. [PubMed: 10462810]



- Chatterjee M, Peng SC. Processing F0 with cochlear implants: Modulation frequency discrimination and speech intonation recognition. *Hear Res.* 2008; 235:143–156. [PubMed: 18093766]
- Chen F, Loizou PC. Predicting the intelligibility of vocoded speech. *Ear Hear.* 2011 Epub ahead of print.
- Dorman MF, Gifford RH, Spahr AJ, McKarns SA. The benefits of combining acoustic and electric stimulation for the recognition of speech, voice, and melodies. *Audiol Neurootol.* 2008; 13:105–112. [PubMed: 18057874]
- Fellowes JM, Remez RE, Rubin PE. Perceiving the sex and identity of a talker without natural vocal timbre. *Percept Psychophys.* 1997; 59:839–849. [PubMed: 9270359]
- Fu QJ. Temporal processing and speech recognition in cochlear implant users. *Neuroreport.* 2002; 13:1635–1639. [PubMed: 12352617]
- Fu QJ, Chinchilla S, Galvin JJ. Voice gender discrimination and vowel recognition in normal-hearing and cochlear implant users. *J Assoc Res Otolaryngol.* 2004; 5:253–260. [PubMed: 15492884]
- Fu QJ, Chinchilla S, Nogaki G, Galvin JJ. Voice gender identification by cochlear implant users: The role of spectral and temporal resolution. *J Acoust Soc Am.* 2005; 118:1711–1718. [PubMed: 16240829]
- Gfeller K, Turner C, Oleson J, Zhang X, Gantz B, et al. Accuracy of cochlear implant recipients on pitch perception, melody recognition, and speech reception in noise. *Ear Hear.* 2007; 28:412–423. [PubMed: 17485990]
- Gonzalez J, Oliver JC. Gender and speaker identification as a function of the number of channels in spectrally reduced speech. *J Acoust Soc Am.* 2005; 118:461–470. [PubMed: 16119365]
- Greenwood DD. A cochlear frequency-position function for several species: 29 years later. *J Acoust Soc Am.* 1990; 87:2592–2605. [PubMed: 2373794]
- Henry BA, Turner CW, Behrens A. Spectral peak resolution and speech recognition in quiet: Normal hearing, hearing impaired, and cochlear implant listeners. *J Acoust Soc Am.* 2005; 118:1111–1121. [PubMed: 16158665]
- Hillenbrand JM, Getty LA, Clark MJ, Wheeler K. Acoustic characteristics of American English vowels. *J Acoust Soc Am.* 1995; 97:3099–3111. [PubMed: 7759650]
- Hillenbrand JM, Clark MJ. The role of  $f(0)$  and formant frequencies in distinguishing the voices of men and women. *Attention Percept Psychophys.* 2009; 71:1150–1166.
- Li T, Galvin JJ, Fu QJ. Interactions between unsupervised learning and the degree of spectral mismatch on short-term perceptual adaptation to spectrally shifted speech. *Ear Hear.* 2009; 30:238–249. [PubMed: 19194293]
- Li T, Fu Q-J. Perceptual adaptation of voice gender discrimination with spectrally shifted vowels. *J Speech Lang Hear Res.* 2010 Epub ahead of print.
- Luo X, Fu QJ, Wei CG, Cao KL. Speech recognition and temporal amplitude modulation processing by Mandarin-speaking cochlear implant users. *Ear Hear.* 2008; 29:957–970. [PubMed: 18818548]
- Litvak LM, Spahr AJ, Saoji AA, Fridman GY. Relationship between perception of spectral ripple and speech recognition in cochlear implant and vocoder listeners. *J Acoust Soc Am.* 2007; 122:982–991. [PubMed: 17672646]
- Shannon RV, Jensvold A, Padilla M, Robert ME, Wang X. Consonant recordings for speech testing. *J Acoust Soc Am.* 1999; 106:71–74.
- Stone M, Fullgrabe C, Moore B. Benefit of high-rate envelope cues in vocoder processing: Effects of number of channels and spectral region. *J Acoust Soc Am.* 2008; 124:2272–2282. [PubMed: 19062865]
- Vongphoe M, Zeng FG. Speaker recognition with temporal cues in acoustic and electric hearing. *J Acoust Soc Am.* 2005; 118:1055–1061. [PubMed: 16158660]
- Won JH, Drennan WR, Rubinstein JT. Spectral-ripple resolution correlates with speech reception in noise in cochlear implant users. *J Assoc Res Otolaryngol.* 2007; 8:384–392. [PubMed: 17587137]
- Won JH, Drennan WR, Kang RS, Rubinstein JT. Psychoacoustic abilities associated with music perception in cochlear implant users. *Ear Hear.* 2010; 31:796–805. [PubMed: 20595901]
- Xu L, Zheng Y. Spectral and temporal cues for phoneme recognition in noise. *J Acoust Soc Am.* 2007; 122:1758–1764. [PubMed: 17927435]



**Figure 1.** Individual performance (symbols) and mean performance (bars) of CI subjects under various test conditions: (A) VGD performance with four different stimulus sets. (B) Vowel and consonant recognition performance in quiet and in speech-shaped noise (SSN) at a 5 dB SNR.



**Table 1**

Demographic information, CI type, and experience (R = right ear, L = left ear, ACE = advanced combination encoder, CIS = continuous interleaved sampling, and SPEAK = spectral peak), and etiology of hearing loss of subjects.

Subject	Gender	Age	Device	Strategy	Years of using CI (ear)	Etiology
S1	F	63	Nucleus-24	ACE	19 (R)/7 (L)	Unknown
S2	F	67	Nucleus-24	ACE	6 (R)	Genetic
S3	F	76	Nucleus-24	ACE	9 (R)	Unknown
S4	M	76	CII	HiRes	10 (L)	Unknown
S5	M	72	Freedom	ACE	1 (R)	Noise-induced
S6	M	59	Freedom	ACE	3 (R)	Genetic
S7	F	25	Freedom	ACE	2 (R)	Unknown
S8	M	79	Nucleus-22	SPEAK	14 (L)	Noise-induced
S9	F	61	CI (R)/HiRes 90K (L)	CIS (R)/HiRes (L)	16 (R)/4 (L)	Ototoxicity
S10	F	28	Freedom	ACE	3 (R)/1 (L)	Unknown
S11	M	61	Freedom (R)/Nucleus-22 (L)	ACE (R)/SPEAK (L)	2 (R)/16 (L)	Unknown

**Table 2**

Correlation coefficients with  $p$ -values between VGD performance tested with different stimulus sets and speech recognition performance under different background conditions (SSN: speech-shaped noise). Significant correlations ( $p < 0.0032$ ) after Bonferroni multiple testing corrections are shown in bold fonts.

	Vowels, in quiet	Vowels, in SSN	Consonants, in quiet	Consonants, in SSN
Talker set 1	$r = 0.2024, p = 0.5506$	$r = 0.4068, p = 0.2143$	$r = 0.6186, p = 0.0425$	$r = 0.6218, p = 0.0411$
Talker set 1, 50 Hz	<b><math>r = 0.8110, p = 0.0025</math></b>	$r = 0.6211, p = 0.0414$	$r = 0.6253, p = 0.0396$	$r = 0.4687, p = 0.1459$
Talker set 1, 200 Hz	$r = 0.5070, p = 0.1115$	$r = 0.7067, p = 0.015$	<b><math>r = 0.8284, p = 0.0016</math></b>	$r = 0.6805, p = 0.0212$
Talker set 2	$r = 0.7178, p = 0.0129$	<b><math>r = 0.9289, p = 0.0000</math></b>	$r = 0.6936, p = 0.0179$	<b><math>r = 0.8283, p = 0.0016</math></b>