JARO

The Role of Spectral and Temporal Cues in Voice Gender Discrimination by Normal-Hearing Listeners and Cochlear Implant Users

QIAN-JIE FU, SHEROL CHINCHILLA, AND JOHN J. GALVIN

Department of Auditory Implants and Perceptions, House Ear Institute, Los Angeles, CA 90057, USA

ABSTRACT

The present study investigated the relative importance of temporal and spectral cues in voice gender discrimination and vowel recognition by normal-hearing subjects listening to an acoustic simulation of cochlear implant speech processing and by cochlear implant users. In the simulation, the number of speech processing channels ranged from 4 to 32, thereby varying the spectral resolution; the cutoff frequencies of the channels' envelope filters ranged from 20 to 320 Hz, thereby manipulating the available temporal cues. For normal-hearing subjects, results showed that both voice gender discrimination and vowel recognition scores improved as the number of spectral channels was increased. When only 4 spectral channels were available, voice gender discrimination significantly improved as the envelope filter cutoff frequency was increased from 20 to 320 Hz. For all spectral conditions, increasing the amount of temporal information had no significant effect on vowel recognition. Both voice gender discrimination and vowel recognition scores were highly variable among implant users. The performance of cochlear implant listeners was similar to that of normal-hearing subjects listening to comparable speech processing (4–8 spectral channels). The results suggest that both spectral and temporal cues contribute to voice gender discrimination and that temporal cues are especially important for cochlear implant users to identify the voice gender when there is reduced spectral resolution.

Keywords: Cochlear implants, voice gender, spectral cues, temporal cues

INTRODUCTION

Although cochlear implants (CIs) are able to transmit only limited spectral and temporal information, they have been remarkably successful in restoring hearing sensation and speech recognition to profoundly deafened patients. Because spectral and temporal cues can be processed independently with the implant device, the CI has become a powerful research tool with which to study the relative contributions of these cues to speech recognition. With a typical CI speech processor, the spectral representation of sound is encoded by the number of stimulated electrodes, while temporal resolution is constrained by the cutoff frequency of the envelope filters and the rate of stimulation. For CI users, spectral and temporal resolution can be further limited by the health and proximity of neural populations relative to the stimulating electrodes. In general, greater numbers of stimulated electrodes (channels) provide better spectral representation of speech sounds, while higher stimulation rates improve the temporal sampling of the speech signal.

The effects of spectral resolution on speech recognition performance has been explored in both CI users and normal-hearing (NH) subjects listening to an acoustic simulation of CI speech processing

Correspondence to: Qian-Jie Fu · Department of Auditory Implants and Perception · House Ear Institute · 2100 West Third St. · Los Angeles, CA 90057. Telephone: (213) 273-8036; fax: (213) 413-0950; email: qfu@hei.org

(Dorman et al. 1998; Friesen et al. 2001; Fu et al. 1998a; Shannon et al. 1995). These studies consistently showed that speech recognition improved as the number of channels was increased. However, the spectral resolution required for asymptotic performance levels depended on the speech materials and/or the difficulty of the listening task. For example, as few as 4 channels was sufficient for recognition of simple sentences (Shannon et al. 1995), while as many as 16 or more spectral channels were necessary to understand speech in noise (Fu et al. 1998a; Dorman et al. 1998). Similar trends in speech recognition performance were observed between CI and NH listeners, although CI users achieved asymptotic performance levels with 8 or fewer channels, regardless of the testing materials (Friesen et al. 2001).

The contribution of temporal cues to speech recognition has also been explored in both NH listeners and CI users. In many experiments, the temporal resolution has been manipulated by either varying the cutoff frequency of envelope extraction (Fu and Shannon 2000; Rosen 1992; Van Tasell et al. 1987; Drullman et al. 1994; Wilson et al. 1999) or by varying the rate of stimulation (Brill et al. 1997; Friesen et al. 2004; Fu and Shannon 2000; Loizou et al. 2000; Wilson et al. 1999). In general, results from these cutoff frequency studies showed that temporal cues made only a moderate contribution to listeners' speech recognition. The improvement was not significant for envelope filter cutoff frequencies above 20 Hz. However, mixed results have been reported in the stimulation rate studies; some studies showed continual improvement in recognition with increasing stimulation rate, while the others showed no significant effect of the stimulation rate on speech performance. It is possible (but unlikely) that any benefits from high stimulation rates were due to improved temporal sampling of the speech signal. The combined results from these studies suggest that temporal cues make only limited contributions to speech recognition; these limited contributions were consistent across phoneme and sentence recognition tasks, in quiet and in noise.

While temporal cues may not contribute strongly to recognition of English speech, they may contribute more strongly to recognition of tonal languages such as Mandarin Chinese (Fu et al. 1998b; Fu and Zeng 2000; Luo and Fu 2004). For example, Fu at al. 1998b found that, while NH listeners' Chinese phoneme recognition improved as the number of channels in a noise-band processor was increased from 1 to 4, increased temporal cues had no effect on vowel and consonant recognition. However, Mandarin tone recognition did benefit from the additional temporal cues, especially periodicity cues. Fu et al. (2004) also measured Chinese tone recognition in CI patients

with different speech processing strategies. Tone recognition did not significantly improve with 4 or more channels; however, tone recognition did significantly improve when the stimulation rate was increased from 250 to 900 Hz per channel. These results suggest that temporal cues may be more important for discriminating suprasegmental than segmental information. Thus, the impact of temporal cues may not have been fully realized in the above-mentioned English speech recognition tasks because the temporal cues were not critically important for English speech intelligibility.

Besides suprasegmental information and speech quality, voice gender perception is another aspect of natural speech perception to which temporal cues may contribute. The acoustic differences between male and female voices have been studied extensively. Differences in speakers' ''breathiness'' (Klatt and Klatt 1990), fundamental frequency (F_0) (Whiteside 1998), and formant structure (Mury and Sigh 1980) help listeners to discriminate voice gender. Difference in speakers' F_0 is perhaps the most important cue for voice gender discrimination. The F_0 of a female voice is typically about one octave higher than that of a male's voice (Linke 1973); the mean values are near 125 and 225 Hz for males and females, respectively. Physical differences between males and females, such as the length and thickness of the vocal folds, are primarily responsible for F_0 differences between the male and female voice (Titze 1987, 1989).

Recent studies have shown that voice gender identification is possible in conditions of reduced spectral resolution, such as 3-sinewave replicas of natural speech (Remez et al. 1997; Sheffert et al. 2002). A few studies have also explored gender and speaker discrimination ability in CI listeners (Cleary and Pisoni 2002; McDonald et al. 2003; Spahr et al. 2003). Cleary and Pisoni (2002) tested whether children who had used a CI for at least 4 years could discriminate differences between two female voices. In general, the results suggested that speaker discrimination was difficult for CI patients. McDonald et al. (2003) examined CI users' ability to discriminate talker identity as a function of the linguistic content of the stimuli and the talker's gender. In the McDonald study, two speech stimuli were presented to the subject, each spoken by either the same talker or different talkers; subjects were asked to identify the talker from among a closed set. CI subjects' best performance was observed when the same talker produced both tokens in the stimulus pair. When different talkers produced the tokens in the stimulus pair, performance was significantly better for male–female talker contrasts than for within-gender talker contrasts. These results suggest that, while CI patients

TABLE 1

Relevant information for cochlear implant subjects who participated in the experiment

Subject	Age	Etiology	Prosthesis	Age of onset hearing loss	Years with prosthesis	Speech strategy
S1	52	Unknown	Nucleus 22	44	9	SPEAK
S2	61	Hereditary	Nucleus 22	47	12	SPEAK
S3	45	Trauma	Nucleus 22	35	11	SPEAK
S4	61	Trauma/Unknown	Nucleus 22	45	13	SPEAK
S5	70	Unknown	Nucleus 24	55	3	ACE
S6	39	Unknown	Med-el	7	3	CIS
S7	49	Unknown	Clarion II	35	5	HiRes
S8	54	Unknown	Clarion II	N/A	2	HiRes
S9	65	Otosclerosis	Clarion II	69	Reimplanted 5 years ago	HiRes
S10	70	Unknown	Clarion II	65	Reimplanted 6 months ago	HiRes
S11	58	Unknown	Clarion II	58	5	HiRes

may have difficulty in identifying talkers, they are somewhat capable of voice gender discrimination.

While CI users may be able to distinguish voice gender, it is not clear how spectral and temporal cues contribute to voice gender discrimination. The present study examined the effects of spectral and temporal cues on vowel recognition and voice gender discrimination by NH subjects listening to a CI simulation and by CI listeners. Vowel recognition and voice gender discrimination were measured in quiet. For the NH subjects listening to the CI simulation, the spectral resolution and the amount of temporal information were parametrically manipulated.

METHODS

Subjects

Six NH subjects (3 males and 3 females, 22–30 years old) and 11 CI users participated in the study. All NH subjects had pure tone threshold better than 15 dB HL at octave frequencies from 250 to 8000 Hz; CI subject details are listed in Table 1. All subjects were native English speakers. All subjects were paid for their participation.

Test materials

The tokens used for closed-set vowel recognition and voice gender discrimination tests were digitized natural productions from 5 males and 5 females, drawn from speech samples collected by Hillenbrand et al. (1995). There were 12 phonemes in the stimulus set, including 10 monophthongs (/i ɪ ɛ æ u ʊ ɑ ʌ ɔ ɝ/) and 2 diphthongs (/o e/), presented in a /h/vowel/d/ context (heed, hid, head, had, who'd, hood, hod, hud, hawed, heard, hoed, hayed). Each test block included 120 tokens (12 vowels × 10 talkers) for both vowel recognition and voice gender discrimination

tests. All stimuli were normalized to have the same long-term root-mean-square (RMS) values.

Signal processing

For NH listeners, a sinewave processor was used to simulate a CI speech processor fitted with the Continuously Interleaved Sampling (CIS) strategy. The processor was implemented as follows. The signal was first processed through a pre-emphasis filter (highpass with a cutoff frequency of 1200 Hz and a slope of 6 dB/octave). An input frequency range (200–7000 Hz) was bandpassed into a number of frequency analysis bands (4, 8, 16, or 32 bands) using 4th order Butterworth filters; the distribution of the analysis filters was according to Greenwood's (1990) formula. The corner frequencies (3 dB down) of the filters are listed in Table 2. The temporal envelope was extracted from each frequency band by half-wave rectification and low-pass filtering (the cutoff frequency of the envelope filter was varied according to the temporal envelope experimental conditions). For each channel, a sinusoidal carrier was generated; the frequency of the sinewave was equal to the center frequency of the analysis filter. The extracted temporal envelope from each band was used to modulate the corresponding sinusoidal carrier. The amplitude of the modulated sinewave was adjusted to match the RMS energy of the temporal envelopes. Finally, the modulated carriers of each band were summed and the overall level was adjusted to be the same level as the original speech.

Note that sinewave carriers, rather than noise-band carriers, were used in the present study's cochlear implant simulation. Before beginning the present study, pilot data showed that NH listeners' voice gender discrimination was nearly at chance-level performance for 1- and 4-channel acoustic processors, using noise-band carriers in the simulation.

TABLE 2

Analysis and carrier band cutoff frequencies used in noise-vocoder cochlear implant simulation

Cutoff frequency (Hz)			Cutoff frequency (Hz)			Cutoff frequency (Hz)			Cutoff frequency (Hz)		
Channel	Lower	Upper	Channel	Lower	Upper	Channel	Lower	Upper	Channel	Lower	Upper
1	200	234	1	200	272	1	200	359	1	200	591
2	234	272									
3	272	313	2	272	359						
4	313	359									
5	359	409	3	359	464	2	359	591			
6	409	464									
7	464	525	4	464	591						
8	525	591									
9	591	664	5	591	745	3	591	931	2	591	1426
10	664	745									
11	745	833	6	745	931						
12	833	931									
13	931	1037	7	931	1155	4	931	1426			
14	1037	1155									
15	1155	1284	8	1155	1426						
16	1284	1426									
17	1426	1582	9	1426	1753	5	1426	2149	3	1426	3205
18	1582	1753									
19	1753	1942	10	1753	2149						
20	1942	2149									
21	2149	2377	11	2149	2627	6	2149	3205			
22	2377	2627									
23	2627	2903	12	2627	3205						
24	2903	3205									
25	3205	3538	13	3205	3904	7	3205	4748	4	3205	7000
26	3538	3904									
27	3904	4306	14	3904	4748						
28	4306	4748									
29	4748	5233	15	4748	5768	8	4748	7000			
30	5233	5768									
31	5768	6355	16	5768	7000						
32	6355	7000									

These data were not consistent with the performance observed in some CI patients who could easily distinguish male and female voices, even with a single-channel processor. A detailed waveform analysis indicates that noise-band carriers may disrupt the fine temporal structure embedded in the envelopes with high modulation frequencies; sinewave carriers can better preserve this fine temporal structure. A detailed comparison between noise-band and sinewave carriers was presented at the 2002 ARO Mid-Winter Meeting. The pilot data and waveform analysis suggest that sinewave carriers may be more appropriate to compare the tradeoff between spectral and temporal cues in the voice gender discrimination.

All CI subjects were tested using their clinically assigned speech processors (details of CI subjects' speech processors are shown in Table 1). Subjects were instructed to use their normal volume and sensitivity settings and not to use any noise suppression settings. Depending on the implanted device, four different speech processing strategies were used by the CI subjects in their clinically assigned speech processors. Four Nucleus-22 patients were fitted with the SPEAK strategy (Seligman and McDermott 1995). One Nucleus-24 patient was fitted with the ACE strategy (Arndt et al. 1999). One Med-El patient was fitted with the CIS strategy (Wilson et al. 1991). Five Clarion II patients all were fitted with the high-resolution speech coding strategy (HiRes).

Procedure

Vowel recognition was measured using a closed-set, 12-alternative, forced-choice (12AFC) identification paradigm. For each trial, a stimulus token was chosen randomly, without replacement, from the vowel set and presented to the subject. The subject responded by clicking on one of the 12 response choices displayed on a computer screen; the response buttons were labeled in a /h/-vowel-/d/ context. If subjects were unsure of

the correct answer, they were instructed to guess as best that they could. No feedback was provided to the subjects. For NH listeners, vowel recognition was measured for 4 spectral resolution conditions (4-, 8-, 16-, and 32-channel processors), combined with 3 different temporal envelope conditions (envelope filter cutoff frequencies of 40, 80, and 320 Hz).

Voice gender recognition was measured using a closed-set, 2AFC identification paradigm. For each trial, a stimulus token was chosen randomly, without replacement, from the vowel set and presented to the subject. The subject responded by clicking on one of 2 response choices displayed on a computer screen; the response buttons were labeled ''male'' or ''female.'' If subjects were unsure of the correct answer, they were instructed to guess as best that they could. No feedback was provided to the subjects. Again, CI listeners were tested using their clinically assigned speech processors. For NH listeners, voice gender recognition was measured for 4 spectral resolution conditions (4-, 8-, 16-, and 32-channel processors) combined with 5 different temporal envelope conditions (envelope filter cutoff frequencies of 20, 40, 80, 160, and 320 Hz).

For both NH and CI listeners, vowel and voice gender identification tests were administered in free field in a sound-treated booth (IAC). Stimuli were presented over a single loudspeaker (Tarrnoy Reveal) at 70 dBA (long-term average level).
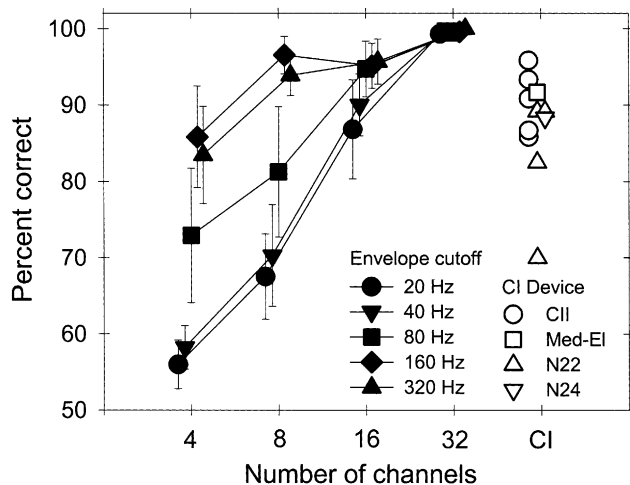
RESULTS

Figure 1 shows vowel recognition results for NH and CI listeners. NH listeners' results are shown for all envelope filter cutoff frequencies as a function of the number of spectral channels; CI listeners' results are shown for each subject's clinically assigned speech processor. For NH listeners, vowel recognition scores improved as the number of spectral channels was increased, for all envelope filter conditions. A two-way repeated-measures analysis of variance (ANOVA) revealed a main effect for the number of channels $[F(3,60) = 143.72, \ p < 0.001]$. However, the main effect of envelope filter cutoff frequency failed to reach statistical significance $[F(2,60) = 0.963, \ p = 0.387]$. Nor was there a significant interaction between the number of channels and the envelope filter cutoff frequency $[F(6,60) = 0.374, \ p = 0.893]$. CI users' vowel recognition performance was considerably variable, even among users of the same type of implant device. The performance range of the CI users was similar to that of NH subjects listening to between 4 and 8 spectral channels, while the best-performing CI users' performance was comparable to that of NH subjects listening to 8 spectral channels.
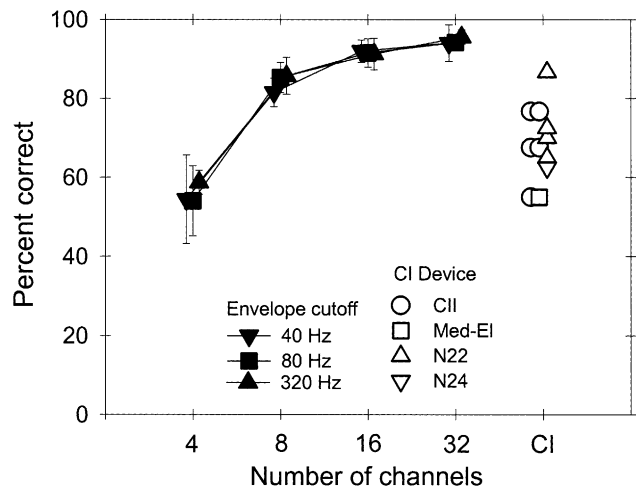
FIG. 1. Mean vowel recognition scores as a function of the number of channels; the parameter is the envelope filter cutoff frequency. Filled symbols show NH listeners' results for different envelope filters; the open symbols show CI listeners' results for different implant devices. Error bars indicate ±1 standard deviation. The chance performance level for vowel recognition is 8.33 correct. Note that the symbols have been slightly offset for presentation clarity.

Figure 2 shows voice gender identification results for NH and CI listeners. NH listeners' results are shown for all envelope filter cutoff frequencies as a function of the number of spectral channels; CI listeners' results are shown for each subject's clinically assigned speech processor. Similar to vowel recognition, NH subjects' voice gender identification improved as the number of spectral channels was increased, for all temporal envelope conditions. Performance with relatively low temporal envelope cutoff frequencies (20, 40, and 80 Hz) sharply increased as more spectral channels were added. Different from vowel recognition performance, voice gender recognition with relatively low spectral resolution (4 and 8 channels) was sharply improved as the amount of temporal information was increased. A two-way ANOVA revealed a main effect for the number of channels on voice gender discrimination $[F(3,100) = 213.654, \ p < 0.001]$. In contrast to vowel recognition, there was a main effect of envelope cutoff frequency $[F(4,100) = 65.516, \ p < 0.001]$. Also, there was a significant interaction between spectral resolution and the amount of temporal information $[F(12,100) = 13.296, \ p < 0.001]$. As with vowel recognition scores, there was great variability in CI users' voice gender discrimination performance. Most CI users performed comparably to NH subjects listening to 4–8 spectral channels with greater amounts of temporal information (160– 320-Hz envelope cutoff frequencies).

FIG. 2. Mean voice gender identification scores as a function of the number of channels; the parameter is the envelope filter cutoff frequency. Filled symbols show NH listeners' results for different envelope filters; the open symbols show CI listeners' results for different implant devices. Error bars indicate ±1 standard deviation. The chance performance level for voice gender identification is 50 correct. Note that the symbols have been slightly offset for presentation clarity.

DISCUSSION

NH listeners' performance with multichannel CI simulations showed that the number of spectral channels significantly affected vowel recognition, consistent with results from previous studies (Shannon et al. 1995; Dorman et al. 1998; Fu et al. 1998a). Voice gender discrimination was also significantly affected by the number of channels; it steadily improved as more spectral channels were added. These results suggest that spectral cues are important for recognizing both linguistic content and suprasegmental information.

Temporal cues also contributed to NH listeners' performance, but more toward the recognition of suprasegmental information. Vowel recognition was not significantly affected by different amounts of temporal information (40-, 80-, and 320-Hz envelope filter cutoff frequencies). However, voice gender discrimination did significantly improve as more temporal cues were added, especially when few spectral channels (4–8) were available (Fig. 2). When only 4 spectral channels were available, the effect of increasing the filter cutoff frequency from 40 to 80 Hz was comparable to that of doubling the number of spectral channels. For example, mean voice gender discrimination with 4 channels improved by 14% as the envelope filter cutoff frequency increased from 40 to 80 Hz. If the temporal information was limited to 40 Hz, performance only improved by 5% as the number of channels was increased from 4 to 8. Similarly, increasing the envelope filter cutoff fre-

quency from 80 Hz to 160 Hz improved mean voice gender discrimination by 10%. Limiting the temporal information to 80 Hz and increasing the number of channels from 4 to 8 improved performance by only 8%.

Rosen (1992) divides speech information in the time/amplitude domain into three distinct categories: (1) envelope cues (temporal fluctuations less than 50 Hz), (2) periodicity cues (temporal fluctuations between 50 and 500 Hz), and (3) fine-structure cues (temporal fluctuations beyond 500 Hz). In the present study, when the envelope filter cutoff frequency was 40 Hz, only envelope cues were preserved; both periodicity and fine-structure cues were lost. Because there was no significant difference in vowel recognition between the 40-, 80-, and 320-Hz envelope filters (for all spectral resolution conditions), the data suggest that the amplitude variations across frequency (gross spectral cues) contribute most strongly to vowel identification and that the fine temporal details within a spectral band are less important, consistent with data from previous studies (Drullman et al. 1994; Shannon et al. 1995; Van Tasell et al. 1987, 1995; Fu and Shannon 2000).

In contrast to vowel recognition, increasing the cutoff frequency of the envelope filter significantly improved voice gender discrimination; however, there was also some interaction between the amounts of available temporal and spectral information. When the envelope filter cutoff frequency was relatively low (20–40 Hz), preserving only envelope cues, vowel recognition and voice gender discrimination exhibited similar patterns: Performance improved sharply, then more gradually as the number of spectral channels was increased. When the envelope filter cutoff frequency was relatively high (160–320 Hz), preserving both envelope and periodicity cues, the vowel recognition and voice gender discrimination patterns were quite different. Vowel recognition again improved sharply, then more gradually as more spectral channels were added. However, with the higher envelope cutoff frequencies (160–320 Hz), voice gender discrimination only slightly improved when the spectral resolution was increased; mean performance with only 4 channels was already 86% correct. This difference between vowel and voice gender recognition performances suggests that, while envelope cues may be adequate for identifying vowels, periodicity cues are important for the perception of suprasegmental information under conditions of severely reduced spectral resolution.

Because of the acoustic differences between male and female voices, it is not surprising that periodicity cues are important for voice gender discrimination. Differences in speakers' F_0 are perhaps the most important cues for voice gender discrimination. The

F_0 of a female voice is typically about one octave higher than that of a male's voice (Linke 1973). The F_0 of a female voice may be as high as 400 Hz, while the F_0 of a male voice may be only as high as 200 Hz. F_0 information, while not explicitly coded in modern CI speech processors (real or simulated), is well-represented by time waveform in each channel, making voice gender discrimination possible in the absence of explicit place coding of F_0 information.

Spectral cues also contribute strongly to voice gender discrimination. When temporal cues were limited (<80-Hz envelope filter), NH listeners' performance improved steadily as the number of spectral channels was increased from 4 to 16; increased spectral cues were especially helpful when periodicity cues were not available (20- and 40-Hz envelope filters). Near-perfect recognition was achieved with a 32-channel processor regardless of the temporal resolution. One possible explanation is that, for the 32-band processor, the spectral cues were preserved well enough to enable subjects to "decode" voice source spectrum and average formant frequency differences. However, these differences are somewhat subtle across male and female talkers, compared with the more robust F_0 differences; NH listeners may not be able to hear these differences under conditions of reduced spectral resolution (relative to normal). More probably, the spectral resolution in the 32-channel processor allowed listeners to reconstruct the "missing" fundamental (which, for male talkers, fell below the 200-Hz minimum acoustic input frequency) and thereby differentiate the voice gender.

Most CI users were able to achieve good vowel recognition and voice gender discrimination with their clinical speech processor. There was great variability in performance within and across subjects and implant devices; the best/worst performer in vowel recognition was not always best/worst performer in voice gender discrimination. The reduced spectral resolution in CI patients may be much more limited than the amount of spectral information transmitted by the speech processor. Although most CI speech processors transmit between 8 and 22 channels of spectral information, data from previous studies suggest that the "effective" number of spectral channels may be fewer then 8 channels due to factors such as electrode interaction, frequency-to-electrode mismatch, and others.

While there may be a "ceiling" to the effective spectral resolution available to CI users, there may also be a limit to CI users' temporal resolution. Given that increased temporal cues allowed NH listeners to improve voice gender discrimination when there was limited spectral resolution (4–8 channels), there might have been significant differences between CI devices as the devices provide different amounts of temporal cues. For example, the Nucleus-22 stimulation rate with the SPEAK strategy is only 250 Hz/channel; the Nucleus-24 stimulation rate with the ACE strategy is 900 Hz/channel. Both devices would readily transmit envelope cues, but the Nucleus-24 device could more readily transmit periodicity cues. Similarly, the Clarion CII and the Med-El devices stimulate at much higher rates than the Nucleus-22 device. However, there is no clear advantage for the higher-rate devices in the voice gender discrimination task, in which the increased temporal cues would be beneficial. Intersubject variability, combined with a small sampling for each implant device, makes it difficult to compare CI users' reception of temporal cues provided by their clinically assigned speech processors. CI users may also differ in their temporal processing capabilities; some CI users' effective temporal resolution may be much better than other CI users'. Further work with implant listeners, in which the amount of available spectral and temporal information is directly controlled, is necessary to understand the relative contributions of spectral and temporal cues to voice gender discrimination.

ACKNOWLEDGMENTS

REFERENCES

ARNDT P, STALLER S, ARCAROLI J, HINES A, EBINGER K. Within-subject comparison of advanced coding strategies in the Nucleus 24 Cochlear Implant. Cochlear Corporation, Englewood, CO, 1999.

BRILL S, GSTOTTNER W, HELMS J, ILBERG CV, BAUMGARTNER W, MULLER J, KIEFER J. Optimization of channel number and stimulation rate for the fast CIS strategy in the COMBI 40+. Am. J. Otol. 18:S104–S106, 1997.

CLEARY M, PISONI DB. Talker discrimination by prelingually deaf children with cochlear implants: preliminary results. Ann. Otol. Rhinol. Laryngol. Suppl. 189:113–118, 2002.

DORMAN M, LOIZOU P, RAINEY D. Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band output. J. Acoust. Soc. Am. 102:2403–2411, 1997.

DORMAN MF, LOIZOU PC, FITZKE J. The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6–20 channels. J. Acoust. Soc. Am. 104:3583–3585, 1998.

DRULLMAN R, FESTEN JM, PLOMP R. Effect of temporal envelope smearing on speech reception. J. coust. Soc. Am. 95:1053–1064, 1994.

FELLOWES JM, REMEZ RE, RUBIN PE. Perceiving sex and identity of a talker without natural voice timbre. Percept. Psycophys. 59(6):839–849, 1997.

Friesen LM, Shannon RV, Baskent D, Wang X. Speech recognition in noise as a function of the number of spectral channels: a comparison of acoustic hearing and cochlear implants. J. Acoust. Soc. Am. 110:1150–1163, 2001.

Friesen LM, Shannon RV, Cruz R. Speech recognition as a function of stimulation rate in Clarion and Nucleus-22 cochlear implants. Audiol. Neurootol. 2004, in press.

Fu Q-J, Shannon RV. Effects of stimulation rate on phoneme recognition in cochlear implant users. J. Acoust. Soc. Am. 107:589–597, 2000.

Fu Q-J, Zeng F-G. Effects of envelope cues on Mandarin Chinese tone recognition. Asia Pac. J. Speech Lang. Hear. 5(1):45–57, 2000.

Fu Q-J, Shannon R, Wang X. Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing. J. Acoust. Soc. Am. 104:3586–3596, 1998a.

Fu Q-J, Zeng F-G, Shannon RV, Soli SD. Importance of tonal envelope cues in Chinese speech recognition. J. Acoust. Soc. Am. 104:505–510, 1998b.

Fu Q-J, Hsu C-J, Horng M-J. Effects of speech processing strategy on Chinese tone recognition by nucleus-24 cochlear implant patients. Ear Hear. 2004, in press.

Greenwood DD. A cochlear frequency–position function for several species—29 years later. J. Acoust. Soc. Am. 87:2592–2605, 1990.

Hillenbrand J, Getty LA, Clark MJ, Wheeler K. Acoustic characteristics of American English vowels. J. Acoust. Soc. Am. 97:3099–3111, 1998.

Klatt DH, Klatt LC. Analysis, synthesis and perception of voice quality variations among female and male talkers. J. Acoust. Soc. Am. 80:820–857, 1990.

Ladefoged P. A Course In Phonetics. 3rd ed. Harcourt Brace College Publishers, City, p 187, 1993.

Linke CE. A study of pitch characteristics of female voices and their relationship to vocal effectiveness. Folia Phoniatr. (Basel). 25(3):173–185, 1973.

Loizou P, Poroy O, Dorman M. The effect of parametric variations of cochlear implant processors on speech understanding. J. Acoust. Soc. Am. 108:790–802, 2000.

Loizou P, Poroy O. Minimal spectral contrast needed for vowel identification by normal hearing and cochlear implant listeners. J. Acoust. Soc. Am. 110:1619–1627, 2001.

Luo X, Fu Q-J. Enhancing Chinese tone recognition by manipulating amplitude contour: Implications for cochlear implants. J. Acoust. Soc. Am. 2004, in press.

McDonald CJ, Kirk KI, Krueger T, Houston D, Sprunger A. Talker Discrimination by Adults with Cochlear Implants. Association for Research in Otolaryngology, 26th Midwinter Meeting, February 2003.

Mury T, Sigh S. Multidimensional analysis of male and female voices. J. Acoust. Soc. Am. 68:1294–1300, 1980.

Remez RE, Fellowes JM, Rubin PE. Talker identification based on phonetic information. J. Exp. Psychol. Hum. Percept. Perform. 23:651–666, 1997.

Rosen S. Temporal information in speech: acoustic, auditory and linguistic aspects. Philos. Trans. R. Soc. Lond. B Biol. Sci. 336:367–373, 1992.

Seligman PM, McDennott HJ. Architecture of the Spectra-22 speech processor. Ann. Otol. Rhinol. Laryngol. 104(Suppl. 166):139–141, 1995.

Shannon RV, Zeng FG, Kammath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. Science. 270:303–304, 1995.

Sheffert SM, Pisoni DB, Fellowes JM, Remez RE. Learning to recognize talkers from natural, sinewave, and reversed speech samples. J. Exp. Psychol. Hum. Percept. Perform. 28:1447–1469, 2002.

Spahr A, Dorman MF, Kirk KI. A comparison of performance among patients fit with the CII Hi-resolution, 3G and TEMP+ processors. 2003 Conference on Implantable Auditory Prostheses, p 161, 2003.

Titze IR. Some technical considerations in voice perturbation measurements. J. Speech Hear Res. 30(2):252–260, 1987.

Titze IR. Physiologic and acoustic differences between male and female voices. J. Acoust. Soc. Am. 85(4):1699–1707, 1989.

Van Tasell DJ, Greenfield DG, Logemann JJ, Nelson DA. Temporal cues for consonant recognition: Training, talker generalization, and use in evaluation of cochlear implants. J. Acoust. Soc. Am. 92:1247–1257, 1992.

Whiteside SP. Identification of a speaker's sex: a study of vowels. Percept. Mot. Skills. 86(2):579–584, 1998.

Wilson B, Lawson D, Zerbi M, Wolford R. Speech Processors for Auditory Prostheses. NIH Project N01-DC-8-2105, Third Quarterly Progress Report, 1999.