

Original Article

Relative Contributions of Spectral and Temporal Cues for Speech Recognition in Patients with Sensorineural Hearing Loss

XU Li, ZHOU Ning, Rebecca Brashears, Katherine Rife

School of Hearing, Speech and Language Sciences, Ohio University, Athens, Ohio 45701, USA

Abstract The present study was designed to examine speech recognition in patients with sensorineural hearing loss when the temporal and spectral information in the speech signals were co-varied. Four subjects with mild to moderate sensorineural hearing loss were recruited to participate in consonant and vowel recognition tests that used speech stimuli processed through a noise-excited vocoder. The number of channels was varied between 2 and 32, which defined spectral information. The lowpass cutoff frequency of the temporal envelope extractor was varied from 1 to 512 Hz, which defined temporal information. Results indicate that performance of subjects with sensorineural hearing loss varied tremendously among the subjects. For consonant recognition, patterns of relative contributions of spectral and temporal information were similar to those in normal-hearing subjects. The utility of temporal envelope information appeared to be normal in the hearing-impaired listeners. For vowel recognition, which depended predominately on spectral information, the performance plateau was achieved with numbers of channels as high as 16–24, much higher than expected, given that the frequency selectivity in patients with sensorineural hearing loss might be compromised. In order to understand the mechanisms on how hearing-impaired listeners utilize spectral and temporal cues for speech recognition, future studies that involve a large sample of patients with sensorineural hearing loss will be necessary to elucidate the relationship between frequency selectivity as well as central processing capability and speech recognition performance using vocoded signals.

Key words spectral, temporal, speech recognition, hearing loss

Introduction

In current multichannel cochlear implant technology, the incoming acoustic signal is divided into a number of frequency bands through a bank of bandpass filters. Information in each band is delivered to a particular location in the cochlea via a stimulating electrode. The number of bands, or frequency channels, constitutes the spectral cues that we study in the present study. In each frequency band, the envelope of the bandpassed signal is extracted using a full- or half-wave rectifica-

tion followed by lowpass filtering. The envelope is then used to modulate the electric pulse trains. Thus, the amount of temporal envelope information, determined by the lowpass cutoff frequency of the lowpass filter, constitutes the temporal cues that we study in the present study^[11, 12].

Previous research has extensively examined the contributions of temporal cues and spectral cues to speech recognition in normal-hearing listeners^[3-9]. Typically, a noise-excited vocoder is used to control the amount of spectral and temporal cues in the signal and speech recognition performance is measured in normal-hearing subjects. There is a consensus in the literature that speech recognition can be achieved with as few as 4 spectral channels and with temporal envelope as low as

Correspondence to: Dr. Xu Li, School of Hearing, Speech and Language Sciences, Ohio University, Athens, Ohio 45701, USA.
E-mail address: xul@ohio.edu

20 Hz. More recent research has shown that the number of spectral channels required for good speech recognition might be larger than the minimum of 4, depending on the level of challenge of the listening conditions (e.g., signal-to-noise ratio)^[10, 11]. We have further shown that the temporal envelope required for vowel and consonant recognition is 4 and 16 to 32 Hz, respectively^[11, 12].

In a series of studies where we used the vocoder technique to examine the relative contributions of temporal and spectral cues to speech recognition, we have shown that speech recognition in quiet and in noise depends on both temporal and spectral cues when the two cues were co-varied^[8, 11, 12]. In addition, there is a tradeoff relation between the temporal and spectral cues. That is, enriched spectral information can compensate for diminished temporal information in speech recognition, and vice versa^[2].

Sensorineural hearing loss affects millions of individuals today causing difficulties in understanding speech. Over the years, the reasons behind poor speech recognition in patients with sensorineural hearing loss have caused considerable controversy and provoked numerous research studies. Some researchers have suggested that these speech recognition difficulties arise primarily from reduced audibility^[13-15], while others^[16-20] believe there is a reduced ability to discriminate suprathreshold sounds^[21]. Nonetheless, few studies have examined the relative importance of spectral and temporal cues for speech recognition in patients with sensorineural hearing loss. Turner et al.^[22] found that, despite the poor speech recognition performance in general in patients with moderate to severe sensorineural hearing loss, the temporal (non-spectral) acuity for speech recognition was comparable to that of the normal-hearing controls, when audibility of the stimuli is compensated for in speech signals. In another study, consonant recognition showed continuous improvement in both normal-hearing and hearing-impaired subjects as the number of channels was increased from 1 to 8^[23]. Baskent^[24] examined the use of spectral cues for consonant and vowel recognition in patients with moderate sensorineural hearing loss. Baskent showed that in quiet and low background noise (signal-to-noise ratio at 10 dB), performance by both normal-hearing and hearing-impaired subjects

saturates at 8 channels. In higher background noise (signal-to-noise ratio at 0 and -5 dB), performance by hearing-impaired subjects saturates at 8 channels, while performance by normal-hearing subjects continued to benefit from more channels up to 12-16 with vowels, and 10-12 with consonants. It appears that patients with sensorineural hearing loss have limited spectral resolution to utilize spectral information in the speech signal.

Relative contributions of spectral and temporal cues for speech recognition in patients with sensorineural hearing loss have not been studied. Studies by Turner et al.^[22, 23] indicated that even though hearing-impaired listeners were able to make use of temporal cues with one-channel processing, they had difficulties to combine information from multiple channels. Thus, it is important to examine how spectral and temporal cues interact and their effects on speech recognition when both cues are co-varied. The present study is designed to extend our previous observations on the relative contributions of spectral and temporal cues for speech recognition in normal-hearing population to patients with sensorineural hearing loss. It is hypothesized that a tradeoff exists between the spectral and temporal cues for speech recognition in patients with sensorineural hearing loss. While the use of temporal cues might be similar to that in normal-hearing listeners, the number of channels needed to reach performance plateau for hearing-impaired listeners may be lower than that reported in normal-hearing listeners. The data obtained from the present study will not only provide us with better understanding of the use of spectral and temporal information in the impaired auditory systems but also guide us in future design of speech processing strategies used in hearing aids and cochlear implants.

Methods

Subjects

Four subjects (S1, S2, S3, and S4) with mild to moderately severe sensorineural hearing loss were recruited to participate in the present study. There were two males (S2 and S4) and two females (S1 and S3). The ages were between 19 and 81 years. Detailed demographic data are provided in Table 1. The audiometric results for the four subjects are shown in Fig. 1. All

subjects used hearing aids in their daily life. Subject S1 had a progressive congenital sensorineural hearing loss in the left ear and a severe-to-profound mixed hearing loss in the right ear that was diagnosed at three years of age. Subject S2 had a bilateral mild-to-moderate sensorineural hearing loss due to presbycusis and noise exposure. Subject S3 had a bilateral progressive moderate-to-severe sensorineural hearing loss that was acquired around 71 years of age due to presbycusis. Subject S4 had a bilateral high-frequency sloping mild-to-moderately-severe sensorineural hearing loss.

Table 1 Demographic data of the subjects

Subject	Age (yrs)	Gender	Ear tested	Consonant score *	Vowel score *
S1	19	f	Left	83.33	85.00
S2	67	m	Left	87.50	81.25
S3	81	f	Left	67.50	56.25
S4	40	m	Left	77.50	93.75

* Percent correct using original, unprocessed speech signals

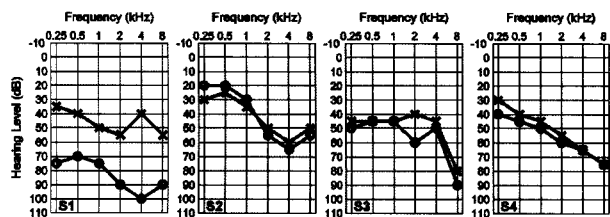


Fig. 1 Audiometric results for the four subjects with sensorineural hearing loss. Subject S1: Moderate sensorineural hearing loss in the left ear and a severe-to-profound mixed hearing loss in the right ear. Subject S2: Moderate sensorineural hearing loss sloping to severe at 4000 Hz bilaterally. Subject S3: Mild-to-moderate sensorineural hearing loss bilaterally. Subject S4: Bilateral high-frequency sloping mild-to-moderately-severe sensorineural hearing loss.

Speech Test Materials

The consonant recognition test used 20 syllables in consonant-/a/ format from Shannon’s consonant stimulus set^[25]. One male (#3) and one female (#3) speakers each produced the stimulus set that consisted of the following 20 syllables: ba, cha, da, fa, ga, ja, ka, la, ma, na, pa, ra, sa, sha, ta, tha, va, wa, ya, and za. The vowel recognition test used 12 syllables in /h/-vowel-/d/ format from Hillenbrand’s vowel stimulus set^[26].

Two male(#48 and #49) and two female(#39 and #44) speakers each produced the stimulus set that consisted of the following 12 words: had, hayed, hawed, head, heard, heed, hid, hod, hoed, hood, hud, and who’d.

Signal Processing

A noise-excited vocoder was used to process the speech signals. Technical details of the signal processing are available in Xu et al.^[2, 8, 11, 12, 27, 28] Briefly, the speech signals in the frequency range of 150 and 5500 Hz were bandpass filtered through a bank of filters. The number of bands (i.e., number of channels) was varied and included 2, 4, 6, 8, 12, 16, 24, and 32. The corner frequencies of the bandpass filters were determined based on Greenwood formula^[29]. The temporal envelope of each band was extracted via half-wave rectification and lowpass filtering. The corner frequency of the low-pass filter was varied between 1 and 512 Hz in octave steps (i.e., 1, 2, 4, 8, 16, 32, 64, 128, 256, and 512 Hz). The temporal envelope of each band was then used to modulate a white noise that was bandpassed through the same bank of filters used earlier to filter original speech signals. Finally, the modulated noise bands were summed and the resultant signals were stored on a computer hard disk for presentation. The signal processing was realized in MATLAB programming environment.

Procedures

No hearing aids were used in the experiment. The speech signals were presented monaurally through a circumaural headphone (Sennheiser, HD 265) in a sound booth to the better ear of the subjects. The intensity was adjusted to the most comfortable level for each subject. Two separate graphical user interfaces (GUIs) were created in MATLAB to present the consonants and vowels. In the GUIs, the phonemic symbols (20 initial consonants or 12 hVd vowels) were represented in alphabetical order in a grid on a computer screen. The consonant or vowel stimuli were presented in random orders. The subjects were instructed to use a computer mouse to select the button that shows the phonemic symbol of the consonant or vowel that they had heard.

Prior to the beginning of the experiment, performance of the unprocessed, original speech stimuli was obtained for each subject (Table 1). A training session

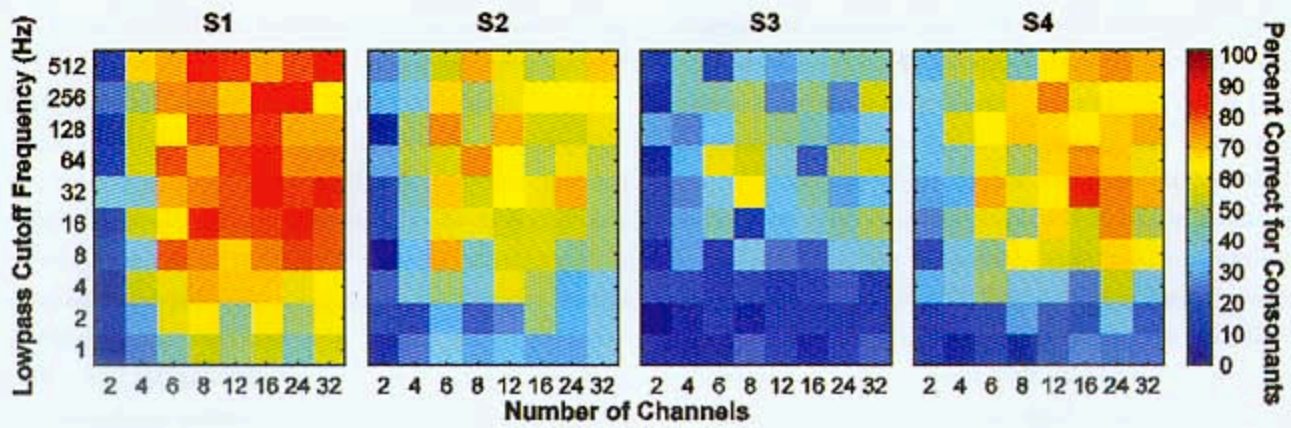


Fig. 2 Consonant recognition performance of the four hearing-impaired subjects. Panels from left to right are for subjects S1 through S4. The abscissa represents the 8 different numbers of channels whereas the ordinate represents the 10 different lowpass cutoff frequencies. The color in each small square represents the consonant recognition performance as indicated by the color bar on the right.

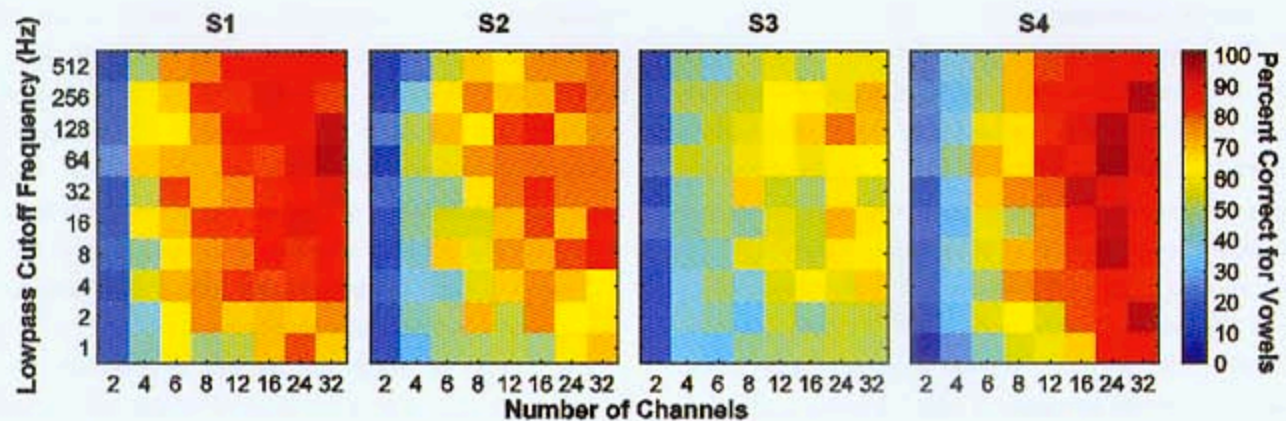


Fig. 3 Vowel recognition performance of the four hearing-impaired subjects. Panels from left to right are for subjects S1 through S4. The abscissa represents the 8 different numbers of channels whereas the ordinate represents the 10 different lowpass cutoff frequencies. The color in each small square represents the vowel recognition performance as indicated by the color bar on the right.

was followed during which each subject listened for about 2–3 h to the processed speech materials to familiarize themselves with the experiment. Eight combinations of the two variables (i.e., number of channels and lowpass cutoff frequency) were used during training. They appeared in the following order, presumably from easy to more difficult combinations: (1) 32 channels, 512 Hz lowpass cutoff; (2) 24 channels, 256 Hz lowpass cutoff; (3) 16 channels, 64 Hz lowpass cutoff; (4) 12 channels, 32 Hz lowpass cutoff; (5) 8 channels, 16 Hz lowpass cutoff; (6) 6 channels, 4 Hz lowpass cutoff; (7) 4 channels, 2 Hz lowpass cutoff; and (8) 2 channels, 1 Hz lowpass cutoff. Feedback was provided during training.

During the experiment, 80 combinations of number of channels and lowpass cutoff frequency (8 numbers of

channels \times 10 lowpass cutoff frequencies) were randomized. For each combination, the consonants were tested first followed by the vowel test. A total of 3200 tokens (20 tokens \times 2 talkers \times 80 combinations) were presented to each subject in the consonant tests. A total of 3840 tokens (12 tokens \times 4 talkers \times 80 combinations) were presented to each subject in the vowel tests. No feedback was provided during the experiment. The experiment was scheduled in blocks of 1 to 3 h, and it took 9–10 h for each subject to complete the experiment.

Results

The percent correct scores of consonant recognition for the 80 combinations of number of channels and lowpass cutoff frequency are shown in Fig. 2. The ab-

scissa represents the 8 different numbers of channels whereas the ordinate represents the 10 different low-pass cutoff frequencies. The color in each small square represents the consonant recognition performance as indicated by the color bar on the right. All hearing-impaired subjects demonstrated consonant recognition performance that was dependent on both number of channels and lowpass cutoff frequency. There were remarkable individual differences in consonant recognition performance in the hearing-impaired subjects. Subject S1 showed fairly high recognition performance, similar to the mean performance of the normal-hearing subjects [11, 12]. The other three subjects showed lower recognition performance with subject S3 being the poorest in performance.

The percent correct scores of vowel recognition for the 80 combinations of number of channels and low-pass cutoff frequency are shown in Fig. 3. The figure is plotted in the same format as Fig. 2. Again, there were large individual differences in vowel recognition performance as in consonant recognition. Subjects S1 and S4 showed high performance that was comparable to that of normal-hearing subjects [11, 12]. The other two subjects (S2 and S3) showed relatively poor performance in vowel recognition with S3 being particularly poor. All hearing-impaired subjects demonstrated vowel recognition performance that was highly dependent

on the number of channels but much less so on the lowpass cutoff frequency.

Fig. 4 shows the mean performance for consonant (left panel) and vowel (right panel) recognition across the four subjects. The figure is plotted in a contour-plot format in which particular combinations of number of channels and lowpass cutoff frequency that produced similar recognition performance are grouped together as areas. The color in each area represents the recognition scores as indicated by the color bar on the right. Note that the abscissa represents the number of channels in logarithmic scale. The mean consonant recognition scores indicate that both spectral cues (number of channels) and temporal cues (lowpass cutoffs) contributed to the performance (Fig. 4, left). Consonant recognition improved as a function of number of channels up to 6 channels. Further increase in number of channels did not lead to improvement in consonant recognition performance. Consonant recognition also improved as a function of lowpass cutoff frequencies up to 16–32 Hz. Further increase in lowpass cutoff frequencies did not contribute to any further improvement in consonant recognition performance.

The pattern in which the number of channels and lowpass cutoff frequency contributed to vowel recognition was quite different from that for consonant recognition (Fig. 4, right). The mean vowel recognition per-

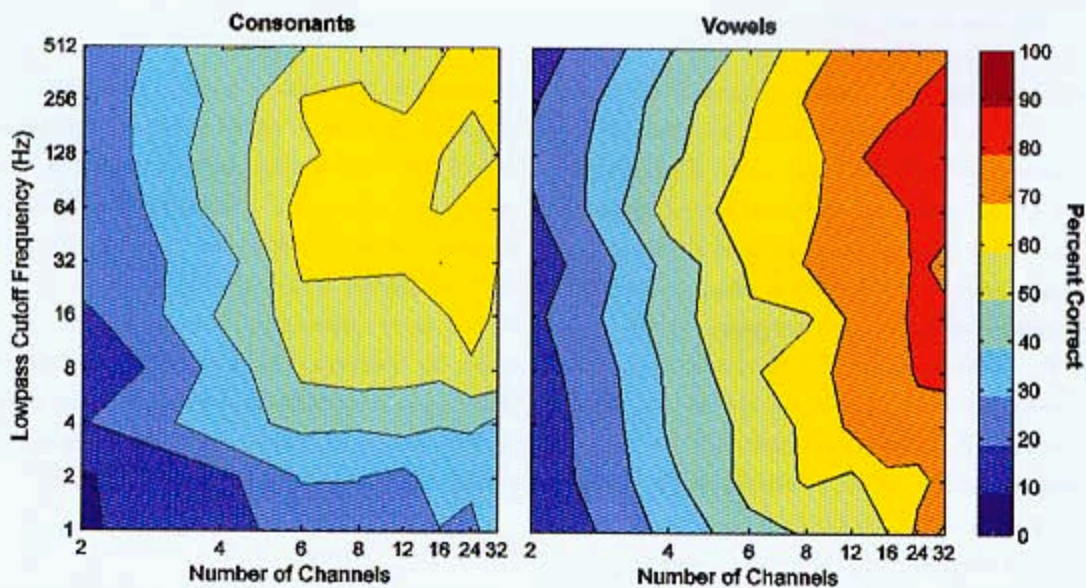


Fig. 4 Mean consonant (left) and vowel (right) recognition performance across the four hearing-impaired subjects. In each contour plot, the area that is filled with a particular color represents the phoneme recognition score (see color bar on the right) for a given number of channels (abscissa) and lowpass cutoff frequency (ordinate).

formance continued to improve as the number of channels increased from 2 to 16 or 24. On the other hand, the performance barely benefited from the increase of lowpass cutoff frequency greater than 4 Hz.

Discussion

One of the findings in the present study was that the performance of individuals with sensorineural hearing loss was on average lower than that of normal-hearing individuals [11, 12]. This was particularly true for consonant recognition (Fig. 2). Turner et al. [23] found that consonant recognition in six hearing-impaired subjects was lower than the normal-hearing controls with number of channels between 2 and 8. Similar findings for consonant recognition were reported by Baskent [24] in four hearing-impaired subjects with number of channels between 2 and 40. In our group of subjects, only one (S1) showed a consonant recognition performance comparable to that of the normal-hearing listeners. It is puzzling though that even in the 2-channel condition, consonant recognition performance in hearing-impaired patients was lower than that of the normal-hearing listeners. Patients with sensorineural hearing loss apparently have normal or close-to-normal temporal processing capability for speech perception as shown in Turner et al. [22] and in the present study (Fig. 4, see discussion below). Nonetheless, reduced spectral resolution of the damaged cochlea cannot fully explain the decreased performance in consonant recognition either, because patients with moderate to severe sensorineural hearing loss should have at least 2 channels to operate. Two reasons have been proposed as explanations by Turner et al. [23]. One is audibility. This can well be the case in some of our subjects. For example, S2 had a notch at 4 kHz whereas S4 had high-frequency sloping loss. Perhaps some of our subjects had issues with audibility in some of the high frequencies. The other is central processing mechanisms. The impaired auditory system cannot combine temporal information across multiple channels. The subjects in our sample who performed the most poorly were the older subjects (S2 and S3) (Figs. 2 and 3). However, our sample size is too small to draw any conclusions about the effect of age on the speech perception with vocoded signals.

Vowel recognition performance in the hearing-impaired subjects was only slightly lower than that in normal-hearing listeners reported in the literature [11, 12, 24]. Two of the four subjects (S1 and S4) had vowel recognition scores comparable to those of normal-hearing subjects (Fig. 3). Baskent [24] showed that vowel recognition in quiet and in high signal-to-noise ratio (+10 dB) did not differ between the hearing-impaired and normal-hearing groups with number of channels between 2 and 40. Only at lower signal-to-noise ratio (0 and -5 dB) did the performance of the hearing-impaired subjects fall short of that of the normal-hearing subjects. It is worth noting that individual differences in vowel recognition performance were just as large as in consonant recognition. Obviously some of the hearing-impaired subjects failed to reach a performance level comparable to that of the normal-hearing listeners. Perhaps the differences in frequency resolution in the patients can account for the differences in performance.

Even though there were large individual differences in both consonant and vowel recognition performance, the mean data did provide valuable information on how temporal and spectral information interact in providing cues for speech recognition (see Fig. 4). The patterns of interaction for both consonant and vowel recognition in the hearing-impaired subjects were remarkably similar to those found in normal-hearing listeners [11, 12]. Consonant recognition depended on temporal cues up to 16-32 Hz of the lowpass cutoffs of the envelope extractor. A tradeoff between the spectral and temporal cues existed for consonant recognition but was much less so for vowel recognition. The temporal resolution appeared intact for the hearing-impaired patients to use temporal envelope cues up to at least 16-32 Hz, which was sufficient for speech recognition. Thus, the use of temporal information by the hearing-impaired listeners is considered normal. It would be interesting to test whether the use of temporal envelope information, such as the periodicity cues (50-500 Hz) [30], is compromised in hearing-impaired subjects or not. Bacon and Gleitman [31] demonstrated that the temporal modulation transfer function in patients with sensorineural hearing loss was similar to that of the normal-hearing subjects. Recent studies indicate that

the temporal fine structure (500–10000 Hz) processing ability in hearing-impaired listeners is dramatically reduced^[32].

Vowel recognition relies minimally on the temporal cues, but heavily on the spectral cues. Interestingly, vowel recognition performance continued to improve as the number of channels increased up to 16–24. Surprisingly, this finding does not support our original hypothesis that patients with sensorineural hearing loss might reach performance plateau at a low number of channels due to their reduced frequency resolution. This finding is also not consistent with Baskent^[24] study in which the author showed that hearing-impaired listeners reached performance plateau for vowel recognition at a number of channel of 8. A close inspection of the line plots in Baskent^[24] reveals that the vowel recognition performance by hearing-impaired subjects did show further steady improvement between 8 and 24 channels. It is likely that the small sample size (N=4) in that study prevented the author from finding significant differences among the channel conditions. It has been shown in normal-hearing listeners that more than 30 spectral channels can be utilized^[10, 33]. Nevertheless, we still do not understand how hearing-impaired subjects can utilize as many as 24 channels. Direct psychoacoustic measures of frequency resolution in these patients may provide valuable information as to how many independent channels each of the damaged cochleae can sustain.

In summary, the present study shows that the contributions of the spectral and temporal cues to speech recognition in hearing-impaired subjects are similar to those in normal-hearing subjects. Future studies that involve a large sample of patients with sensorineural hearing loss will be necessary to confirm these findings. Several issues behind these findings remain to be addressed on a theoretic basis. One is whether the frequency resolution ability of hearing-impaired individuals can predict speech recognition performance using the vocoder speech. Patients with sensorineural hearing loss have different etiologies and pathologies which can affect frequency processing even at the level of the cochlea. Future studies should include psychoacoustic measures of frequency selectivity in the tests. However, ter Keurs et al.^[34] found that psychoacoustic measures

of frequency resolution in patients with sensorineural hearing loss was not correlated with the degree of spectral smearing required to degrade speech recognition in noise. On the other hand, future studies should also include assessment of central auditory processing abilities in patients with sensorineural hearing loss. Central processing deficits are a potential important component in disorders such as presbycusis. This approach is likely to provide valuable insights on the ability of patients to integrate information across spectral channels.

Acknowledgement

The authors acknowledge the technique assistance by Heather Schultz and Marisol Szymanski. The study was supported in part by NIH/NIDCD grants R03 – DC006161 and R15–DC009504.

References

- 1 Loizou PC. Speech-processing in vocoder-centric cochlear implants. *Adv Otorhinolaryngol*, 2006, 64: 109–143.
- 2 Xu L, Pfungst BE. Spectral and temporal cues for speech recognition: Implications for auditory prostheses. *Hear Res*, 2008, 242: 132–140.
- 3 Van Tasell DJ, Soli SD, Kirby VM, Widin GP. Speech waveform envelope cues for consonant recognition. *J Acoust Soc Am*, 1987, 82: 1152–1161.
- 4 Drullman R, Fensten JM, Plomp R. Effect of temporal envelope smearing on speech perception. *J Acoust Soc Am*, 1994, 95: 1053–1064.
- 5 Shannon RV, Zeng F–G, Kamath V, et al. Speech recognition with primarily temporal cues. *Science*, 1995, 270: 303–304.
- 6 Dorman MF, Loizou PC, Rainey D. Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. *J Acoust Soc Am*, 1997, 102: 2403–2411.
- 7 Loizou PC, Dorman M, Tu Z. On the number of channels needed to understand speech. *J Acoust Soc Am*, 1999, 106: 2097–2103.
- 8 Xu L, Tsai YJ, Pfungst BE. Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses. *J Acoust Soc Am*, 2002, 112: 247–258.
- 9 Zeng FG, Nie K, Stickney GS, et al. Speech recognition with amplitude and frequency modulations. *Proc Natl Acad Sci USA*, 2005, 102: 2293–2298.
- 10 Shannon RV, Fu QJ, Galvin J. The number of spectral channels required for speech recognition depends on the difficulty of the listening situation. *Acta Otolaryngol*, 2004,

- 124: 50–54.
- 11 Xu L, Zheng Y. Spectral and temporal cues for phoneme recognition in noise. *J Acoust Soc Am*, 2007, 122: 1758–1764.
 - 12 Xu L, Thompson CS, Pfingst BE. Relative contributions of spectral and temporal cues for phoneme recognition. *J Acoust Soc Am*, 2005, 117: 3255–3267.
 - 13 Humes LE, Dirks DD, Bell TS, et al. Recognition of nonsense syllables by hearing-impaired listeners and by noise-masked normal hearers. *J Acoust Soc Am*, 1987, 81: 765–773.
 - 14 Zurek PM, Delhorne LA. Consonant reception in noise by listeners with mild and moderate sensorineural hearing impairment. *J Acoust Soc Am*, 1987, 82: 1548–1559.
 - 15 Lee LW, Humes LE. Evaluating a speech-reception threshold-model for hearing-impaired listeners. *J Acoust Soc Am*, 1993, 93: 2879–2885.
 - 16 Plomp R. Auditory handicap of hearing impairment and the limited benefit of hearing aids. *J Acoust Soc Am*, 1978, 63: 533–549.
 - 17 Plomp R. A signal-to-noise ratio model for the speech reception threshold of the hearing impaired. *J Speech Hear Res*, 1986, 29: 146–154.
 - 18 Dreschler WA, Plomp R. Relations between psychophysical data and speech perception for hearing-impaired subjects I. *J Acoust Soc Am*, 1980, 68: 1608–1615.
 - 19 Dreschler WA, Plomp R. Relations between psychophysical data and speech perception for hearing-impaired subjects. II. *J Acoust Soc Am*, 1985, 78: 1261–1270.
 - 20 Glasberg B, Moore B. Psychoacoustic abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech. *Scand Audiol*, 1989, 32: 1–25.
 - 21 Moore BCJ. *Cochlear Hearing Loss*. London: Whurr Publishers Ltd, 1998.
 - 22 Turner CW, Souza PE, Forget LN. Use of temporal envelope cues in speech recognition by normal and hearing-impaired listeners. *J Acoust Soc Am*, 1995, 97: 2568–2576.
 - 23 Turner CW, Chi SL, Flock S. Limiting spectral resolution in speech for listeners with sensorineural hearing loss. *J Speech Lang Hear Res*, 1999, 42: 773–784.
 - 24 Baskent D. Speech recognition in normal hearing and sensorineural hearing loss as a function of the number of spectral channels. *J Acoust Soc Am*, 2006, 120: 2908–2925.
 - 25 Shannon RV, Jensvold A, Padilla M, et al. Consonant recordings for speech testing [letter]. *J Acoust Soc Am*, 1999, 106: L71–74.
 - 26 Hillenbrand J, Getty LA, Clark MJ, et al. Acoustic characteristics of American English vowels. *J Acoust Soc Am*, 1995, 97: 3099–3111.
 - 27 Zhou N, Xu L. Lexical tone recognition with spectrally mismatched envelopes. *Hear Res*, 2008, 246: 36–43.
 - 28 Xu L. Temporal and spectral cues for speech perception. *Chinese J Otol*, 2006, 4: 335–342.
 - 29 Greenwood DD. A cochlear frequency-position function for several species – 29 years later. *J Acoust Soc Am*, 1990, 87: 2592–2605.
 - 30 Rosen S. Temporal information in speech: Acoustic, auditory and linguistic aspects. *Philos Trans R Soc London, Ser B* 336: 367–373.
 - 31 Bacon SP, Gleitman RM. Modulation detection in subjects with relatively flat hearing losses. *J Speech Hear Res*, 1992, 35: 642–653.
 - 32 Moore BCJ. The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. *JARO*, 2008, 9: 399–406.
 - 33 Kong YY, Zeng FG. Temporal and spectral cues in Mandarin tone recognition. *J Acoust Soc Am*, 2006, 120: 2830–2840.
 - 34 ter Keurs M, Festen JM, Plomp R. Effect of spectral envelope smearing on speech reception. II. *J Acoust Soc Am*, 1993, 93: 1547–1552.

(Received November 8, 2008)