Combined Effects of Frequency Compression-Expansion and Shift on Speech Recognition

Deniz Başkent and Robert V. Shannon

***Objective:* To explore combined acute effects of frequency shift and compression-expansion on speech recognition, using noiseband vocoder processing.**

***Design:* Recognition of vowels and consonants, processed with a noiseband vocoder, was measured with five normal-hearing subjects, between the ages of 27 and 35 yr. The speech signal was filtered into 8 or 16 analysis bands and the envelopes were extracted from each band. The carrier noise bands were modulated by the envelopes and resynthesized to produce the processed speech. In the baseline matched condition, the frequency ranges of the corresponding analysis and carrier bands were the same. In the shift only condition, the frequency ranges of the carrier bands were shifted up or down relative to the analysis bands. In the compression and expansion only conditions, the analysis band range was made larger or smaller, respectively, than the carrier band range. By applying the shift to carrier bands and compression or expansion to analysis bands simultaneously, the combined effects of the two spectral distortions on speech recognition were explored.**

***Results:* When the spectral distortions of compression-expansion or shift were applied separately, the performance was reduced from the baseline matched condition. However, when the two spectral degradations were applied simultaneously, a compensatory effect was observed; the reduction in performance was smaller for some combinations compared to the reduction observed for each distortion individually.**

***Conclusions:* The results of the present study are consistent with previous vocoder studies with normal-hearing subjects that showed a negative effect of spectral mismatch between analysis and carrier bands on speech recognition. The present results further show that matching the frequency ranges of 1 to 2 kHz, which contain important speech information, can be more beneficial for speech recognition than matching the overall frequency ranges, in certain conditions.**

(Ear & Hearing 2007;28;277–289)

In the human auditory system, the individual frequencies of a complex sound can be perceived separately. In normal hearing, the frequency selectivity is further enhanced by nonlinear cochlear mechanisms (Oxenham & Bacon, 2003). In sensorineural hearing loss, the cochlear nonlinearity might be impaired, and the frequency selectivity may be reduced (Henry et al., 2005). In cochlear implants, frequency selectivity is mainly limited by the number of electrodes, typically 16 to 22, and a further reduction may occur due to channel interactions. A reduction in the spectral resolution can have negative effects on speech recognition, especially in background noise (Stelmachowicz et al., 1985; Başkent, 2006).

In hearing aids and cochlear implants, the signal processing can cause spectral distortions that may additionally affect speech intelligibility. Frequency transposition, for example, was suggested for hearing aid users with steeply sloping high-frequency hearing loss. This feature lowers the high frequency speech components to the frequency range of residual hearing of the patient. Even though frequency transposition provides more speech information within the region of functional hearing, earlier studies have not shown improvement in speech intelligibility compared to simple amplification (Braida et al., 1979). Newer studies that attempted to preserve important acoustic cues for speech in the processing, such as the ratio of the formants (Turner & Hurtig, 1999), showed benefit for some subjects.

In cochlear implants, the input acoustic frequency range assigned onto the electrodes does not always match the tonotopic stimulation range; there might be a spectral shift between the acoustic input frequencies and the cochlear stimulation range. Fu & Shannon (1999) measured the effect of frequency shift on recognition of vowels by implant users. Similar conditions were simulated with normal-hearing subjects using a noiseband vocoder. In the vocoder simulations the carrier noise bands could be interpreted as a representation of the cochlear stimulation while the analysis envelope bands could be interpreted as a representation of the acoustic input. Frequency shift was produced by lowering or raising the carrier band range relative to the analysis band

---

Department of Biomedical Engineering (D.B., R.V.S.), University of Southern California, Los Angeles; and the Department of Auditory Implants (D.B., R.V.S.), House Ear Institute, Los Angeles, California.

0196/0202/07/2803-0277/0 • Ear & Hearing • Copyright © 2007 by Lippincott Williams & Wilkins • Printed in the U.S.A.

range. The performance by both subject groups decreased significantly for shifts larger than 40 to 60%.

Spectral distortions can also be caused by the compressive or expansive mapping of the acoustic input spectrum onto the stimulation range of the electrodes. Başkent & Shannon (2003) simulated compressed and expanded frequency-place maps using noiseband vocoder processing, and measured speech recognition by normal-hearing listeners as a function of the acoustic input range. The performance was best with the matched map where the speech spectrum was truncated to match the simulated stimulation range, and performance deteriorated when the acoustic input range was made wider or narrower. A follow-up study (Başkent & Shannon, 2004) showed similar degradation in speech recognition by implant users when the acoustic input frequency range was manipulated.

The aforementioned studies showed the detrimental effects of spectral distortions due to compressed-expanded or shifted maps on speech recognition separately. However, in implant patients, the two types of spectral distortions may occur at the same time (Fu & Shannon, 1999). The present study explores such combined effects on speech perception by normal-hearing subjects. Using a noiseband vocoder, the amount and the type of the spectral distortion applied to speech was systematically changed to observe the acute effects of the two distortion types, individually and combined, on speech intelligibility.

METHODS

Subjects

Five normal-hearing listeners between the ages of 27 to 35 participated in the study. All subjects were native speakers of American English, and had pure tone thresholds better than 20 dB HL at audiometric frequencies between 250 and 8000 Hz.

Stimuli

Recordings of phonemes with medial vowels and consonants were used for the speech recognition task. Vowel stimuli (Hillenbrand et al., 1995) consisted of 12 medial vowels (10 monophthongs and 2 diphthongs) presented in /h/-vowel-/d/ context (heed, hid, head, had, hod, hawed, hood, who'd, hud, heard, hayed, hoed) and spoken by 5 male and 5 female talkers. Chance level on this test was 8.33% correct. Consonant stimuli (Fu et al., 1998; Turner et al., 1999) consisted of 14 medial consonants /b d f g k m n p s ʃ t θ v z/, presented in /a/-consonant-/a/ context and spoken by 3 male and 3 female talkers. Chance performance level for this test was 7.14% correct.

Experimental Procedures

The stimuli were presented in free field at 70 dBA. The subjects were seated in a sound-proof booth, approximately at one meter distance from the speaker. The stimuli, processed off-line prior to the experiment, were presented by custom software (Robert, 1998). The presentation order of the conditions and the presentation order of stimuli in a specific condition were randomized. During testing, a menu on the screen showed the list of all possible phonemes. The subject identified the phoneme that was presented by selecting the appropriate entry in the menu using the mouse. All subjects were tested once with all experimental conditions.

The subjects who participated in the present study had extensive experience with noiseband vocoder processing due to their involvement in previous vocoder experiments. The study explored the acute effects and no training was provided for any experimental condition. However, a preview of the processed phonemes was allowed prior to testing for each condition.

Mapping Conditions

Eight- and 16-band noiseband vocoder processing (Shannon et al., 1995) was implemented in Matlab. When the vocoder is used to simulate cochlear implant processing, the carrier bands simulate the range for place of stimulation in the cochlea. Therefore, for the carrier bands, the low and high ends of the overall range and the partitioning of the bands were determined in cochlear distance (in mm) from the apex. The overall range was 16 mm and two distances were used for the low-frequency (apical) end; 10 and 15 mm from the apex. Assuming an average length of 35 mm for the human cochlea, these distances simulate insertions of 25 and 20 mm from the basal end, respectively. The analysis bands of the vocoder represent the acoustic input range, which is normally expressed in frequency (in Hz). However, for consistency, the partitioning of the analysis bands was also determined in cochlear distance (in mm). The cut-off frequencies of individual bands were determined by converting the band edges in mm to the frequency domain using Greenwood's frequency-place mapping function (1990):

$$f(x) = A(10^{ax} - k) \qquad (1)$$

In Eq. (1), x represents the cochlear distance in mm from the apex. For the human cochlea, the constants of $A = 165.4$, $a = 0.06$, and $k = 0.88$ were used. Similar to critical bands, the equation shows a relatively linear relationship between cochlear distance and frequency for low frequencies, and a logarithmic relationship for higher frequencies.

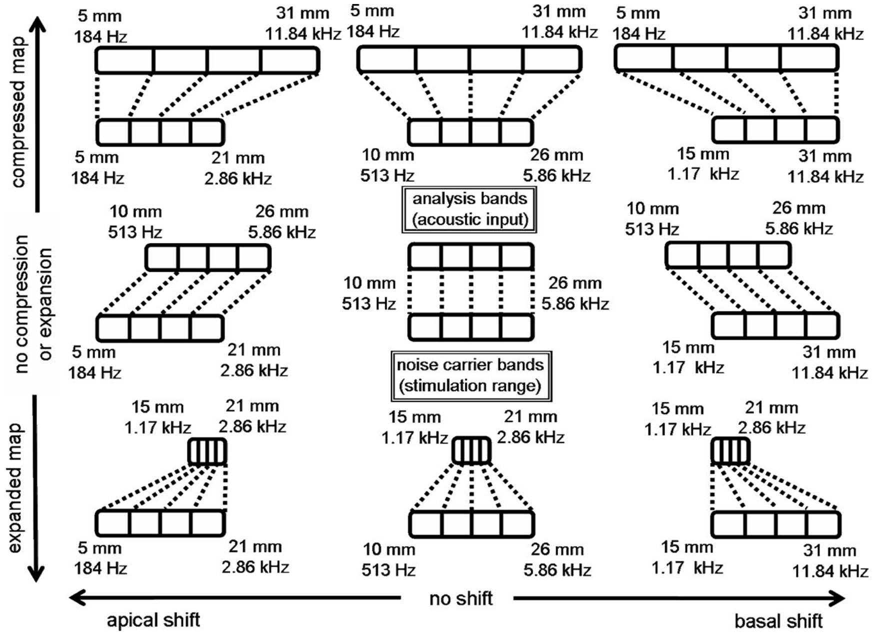

Fig. 1. Frequency-place maps used in the experiment, shown for an example 4-band vocoder processor. The map in the middle shows the baseline matched map where the cut-off frequencies were the same for the analysis and carrier bands. The middle row shows the shift only map with no compression or expansion, where the carrier bands were shifted apically to lower frequencies (shown at left) or basally to higher frequencies (shown at right). The middle column shows the compression and expansion maps with no shift, where the analysis band range was wider (shown on top) or narrower (shown at bottom) than the carrier band range, respectively. The corners show the combined maps; the maps in the top left and bottom left corners are the apically shifted maps combined with compression or expansion, respectively. The maps in the top right and bottom right corners are the basally shifted maps combined with compression or expansion, respectively.

In the baseline matched map the frequency ranges of the corresponding carrier and analysis bands were the same. The middle part of Figure 1 shows the matched map for a carrier band range with the low frequency end at 513 Hz, simulating a cochlear distance of 10 mm from the apex. The overall carrier band range of 16 mm was partitioned into the appropriate number of bands in equal cochlear distances. Once the cut-off frequencies were determined, the phonemes were band-pass filtered into analysis bands, using a set of 6th-order Butterworth filters with the low- and high-frequency filter slopes of –18 dB/octave. The speech envelope was extracted from each band by half-wave rectification and low-pass filtering using a 3rd-order Butterworth filter with a cut-off frequency of 160 Hz at –3 dB and a filter slope of –18 dB/octave. The carrier bands were produced by band-pass filtering white noise with the same set of band-pass filters used for the analysis bands. The carrier bands were modulated by the extracted envelopes, and the modulated carrier bands were summed to produce the processed phoneme. The amplitude levels were adjusted such that the original and processed output had the same overall RMS levels.

The maps representing shifting without compression or expansion were produced by manipulating the carrier bands only; the carrier band range was shifted apically, to lower frequencies, or basally, to higher frequencies. The middle row of Figure 1 illustrates the shifted maps, before compression or expansion was added. The compressed map without shift was produced by manipulating the analysis bands only; the analysis band range was made wider by 5 mm in cochlear distance at each end. As the carrier bands were not changed, this manipulation resulted in a compressive mapping of the analysis bands over the carrier bands. The middle of the top row in Figure 1 illustrates the compressed map without a shift in the carrier bands. The expanded map without shift was produced by making the analysis band range narrower by 5 mm in cochlear distance at each end. The middle of the bottom row in Figure 1 illustrates the expanded map without a shift in the carrier bands. The combinations of the two spectral distortions were produced by applying compression or expansion to the analysis bands and applying a shift to the carrier bands in the same map. The combination maps are shown in the corners in Figure 1.

The mapping conditions are summarized in Tables 1 through 4. Tables 1 and 3 show the band ranges in simulated cochlear distances (in mm) and Tables 2 and 4 show the band ranges in frequency

TABLE 1. Analysis and carrier band ranges shown in cochlear position (in mm)

		Shift in carrier bands						
	Compression-expansion	−5 mm apical shift: Lower frequencies	−3 mm apical shift: Lower frequencies	−1 mm apical shift: Lower frequencies	0 mm No shift	+1 mm basal shift: Higher frequencies	+3 mm basal shift: Higher frequencies	+5 mm basal shift: Higher frequencies
Analysis/Carrier band range (in mm)	+5 mm Compression: Wider analysis band range	5–31 5–21	5–31 7–23	5–31 9–25	5–31 10–26	5–31 11–27	5–31 13–29	5–31 15–31
	0 mm No compression-expansion	10–26 5–21	10–26 7–23	10–26 9–25	Matched map: 10–26 10–26	10–26 11–27	10–26 13–29	10–26 15–31
	−5 mm Expansion: Narrower analysis band range	15–21 5–21	15–21 7–23	15–21 9–25	15–21 10–26	15–21 11–27	15–21 13–29	15–21 15–31

In the baseline matched map, shown in the middle entry, the apical end of the carrier band range was at a simulated distance of 10 mm from the apex, and the carrier and analysis band ranges were the same. The entries in the middle row are the ranges for the map with shift only, as shown in the middle row of Figure 1. The entries in the middle column are the ranges for the map with compression-expansion only, as shown in the middle column of Figure 1. The top and bottom rows show the ranges for the combined maps where the shift was applied simultaneously with compression or expansion, respectively, as shown in the corner portions of Figure 1. Figures 2 and 4 show the results with the maps shown in this table.

(in Hz). In Tables 1 and 2, the low-frequency end of the carrier band range in the matched condition is 513 Hz, simulating a cochlear distance of 10 mm from the apex. In Tables 3 and 4, the low-frequency end of the carrier band range in the matched condition is 1168 Hz, simulating a cochlear distance of 15 mm from the apex.

RESULTS

Figures 2 and 3 show the percent correct scores for vowel recognition, averaged across subjects and corrected for chance, as a function of the shift in carrier bands. Figures 4 and 5 show the scores for consonant recognition. The simulated cochlear position of the most apical end of the carrier band range, with reference to the apex, is also shown at the top of the figures.

The baseline matched map with no shift and no compression-expansion is shown by the open circle at the 0 mm shift condition. In general, highest scores were obtained with the matched map and the performance generally decreased when the mapping was distorted with frequency shift only (open circles), compression-expansion only (filled triangles and squares at the 0 mm shift condition), or the combinations of the two (filled triangles and squares). The effects of the shift only map and compression-expansion only maps were similar to results reported in previous studies (Fu & Shannon, 1999; Başkent & Shannon, 2003). In the present study, compression and expansion were additionally applied to the shifted maps to observe the combined effects.

The scores with the shifted map combined with compression are shown by the filled triangles in Figures 2 through 5. These scores were compared to the scores with the shift only map (open circles), to explore the effect of adding compression to the shifted map. The statistical significance was determined using a two-factor repeated-measures analysis of variance (RM ANOVA), with the main factors of the shift in the carrier bands and the added compression, and the interaction of the two factors. The F and p values of the statistical analysis are shown in Tables 5 and 6. A posthoc multiple comparisons Tukey test was used to compare pairs of conditions (Roberts & Russo, 1999). In Figures 2 through 5, the stars above the data show statistical significance ($p < 0.05$) by the Tukey test when the combined shifted and compressed map was compared with the shift only map for a particular condition.

There was a significant main effect of shift ($p < 0.001$; Table 4) and a significant main effect of added compression (generally $p < 0.01$; Table 4) on recog-

TABLE 2. Analysis and carrier band ranges shown in frequency (in Hz)

		Shift in carrier bands						
	Compression-expansion	−5 mm apical shift: Lower frequencies	−3 mm apical shift: Lower frequencies	−1 mm apical shift: Lower frequencies	0 mm No shift	+1 mm basal shift: Higher frequencies	+3 mm basal shift: Higher frequencies	+5 mm basal shift: Higher frequencies
Analysis/Carrier band range (in Hz)	+5 mm	184–11837	184–11837	184–11837	184–11837	184–11837	184–11837	184–11837
	Compression: Wider analysis band range	184–2864	290–3822	428–5085	513–5860	610–6750	851–8944	1168–11837
	0 mm	513–5860	513–5860	513–5860	Matched map:	513–5860	513–5860	513–5860
	No compression-expansion	184–2864	290–3822	428–5085	513–5860 513–5860	610–6750	851–8944	1168–11837
	−5 mm	1168–2864	1168–2864	1168–2864	1168–2864	1168–2864	1168–2864	1168–2864
	Expansion: Narrower analysis band range	184–2864	290–3822	428–5085	513–5860	610–6750	851–8944	1168–11837

Values are converted from cochlear positions presented in Table 1, using Greenwood's frequency-place mapping equation. The baseline matched map is shown in the middle entry. The middle row shows the map with shifting only. The middle column shows the map with compression-expansion only. The top and bottom rows show the combined maps where the shift was applied simultaneously with compression or expansion, respectively.

nition of vowels. When the shift only and combined shift-compression maps were compared in pairs for individual conditions, it was observed that introducing compression to the shifted maps generally resulted in an additional reduction in the performance (Figures 2 and 3). However, a compensatory effect was observed for the conditions where the carrier bands were shifted apically at least 3 mm, as shown by the –5 mm and –3 mm shift conditions; the performance with the shifted and compressed map was better than the shift only map ($p < 0.001$ for the –5 mm condition and $p < 0.05$ for the –3 mm condition for the simulated stimulation range at 10 mm from the apex, determined by the Tukey test). With consonants, there was a significant main effect of shift ($p < 0.001$; Table 5) but there was no significant main effect of added compression. However, the same compensatory effect was observed for consonants at the apical shift of –5 mm ($p < 0.001$, determined by the Tukey test). The compensatory effect is also likely to be the main reason for the significant interaction between the two main factors of the RM ANOVA, observed for many conditions ($p < 0.001$; Tables 5 and 6). If adding compression to the shift only map had a purely subtractive effect, the performance curve would be reduced for all conditions and the performance curves with the shift only map and the combined map would still be parallel. This was only observed with consonant recognition with the 16-channel vocoder where the unshifted carrier bands were at the simulated distance of 10 mm from the apex (right panel, Figure 4). Adding compression to the map lowered the scores significantly ($p < 0.01$, Table 6) but there was no significant interaction. For the other conditions, however, there was a significant interaction ($p < 0.001$), implying that the added compression produced better performance than the shift only map in some conditions. For instance, note that the performance with the combined map generally peaked for shifts around –3 mm and –1 mm. In contrast, performance for the shift only map peaked around 0 mm.

The scores with the shifted map combined with expansion are shown by the filled squares. The results were analyzed by comparing these scores with those of the shift only map and using a two-factor RM ANOVA with the main factors of the shift in the carrier bands and the added expansion, and the interaction of the two factors (Tables 5 and 6). In Figures 2 through 5, the stars below the data show statistical significance ($p < 0.05$) by the Tukey test when the performance with the combined shifted and expanded map was compared to the performance with the shift only map for a particular condition.

TABLE 3. Analysis and carrier band ranges shown in cochlear position (in mm)

		Shift in carrier bands						
	Compression-expansion	–5 mm apical shift: Lower frequencies	–3 mm apical shift: Lower frequencies	–1 mm apical shift: Lower frequencies	0 mm No shift	+1 mm basal shift: Higher frequencies	+3 mm basal shift: Higher frequencies	+5 mm basal shift: Higher frequencies
Analysis/Carrier band range (in mm)	+5 mm	10–35	10–35	10–35	10–35	10–35	10–35	10–35
	Compression: Wider analysis band range	10–26	12–28	14–30	15–31	16–32	18–34	20–35
	0 mm	15–31	15–31	15–31	Matched map:	15–31	15–31	15–31
	No compression-expansion	10–26	12–28	14–30	15–31	16–32	18–34	20–35
					15–31			
	–5 mm	20–26	20–26	20–26	20–26	20–26	20–26	20–26
	Expansion: Narrower analysis band range	10–26	12–28	14–31	15–31	16–32	18–34	20–35

The apical end of the carrier band range in the matched map was at a simulated distance of 15 mm from the apex. The baseline matched map is shown in the middle entry. The middle row shows the map with shifting only. The middle column shows the map with compression-expansion only. The top and bottom rows show the combined maps where the shift was applied simultaneously with compression or expansion, respectively. In the compressed maps, shown in the top row, the most basal distance of the analysis band range would theoretically be 36 mm. Similarly, in the basal shift condition of +5 mm, the most basal end of the carrier band range would theoretically be 36 mm. Since the apical distance of 36 mm is larger than the assumed cochlear length of 35 mm used in the present study, the most basal end of the band ranges were truncated to 35 mm. Figures 3 and 5 show the results with the maps presented in this table.

There was a significant main effect of shift ($p < 0.001$) and a significant main effect of added expansion ($p < 0.001$) on recognition of vowels and consonants. There was also a significant interaction ($p < 0.001$; Tables 5 and 6), implying that the reduction in the scores due to added expansion was not purely subtractive; the performance with the shifted and expanded map generally peaked around the +1 mm and +3 mm shift conditions. Again, note that this is different from the shift only condition where the peak in performance occurred at 0 mm. In most conditions, adding expansion to the shifted map caused a further reduction in performance ($p < 0.001$ for most conditions, determined by the Tukey test). The reduction in scores due to added expansion was more drastic compared to the reduction due to added compression (Figures 2 through 5). In the expanded maps the bandwidth of the acoustic input into the map was heavily limited, which most likely contributed to poor performance (ANSI. S35, 1997; Başkent, 2003). Shifted and expanded maps almost always resulted in poorer performance than the shift only maps. However, for vowels, for the basally shifted carrier bands, namely the +3 mm and +5 mm shift conditions, the opposite was observed; despite the more restricted acoustic input range, the shifted and expanded map produced better performance compared to the shifted and compressed map (paired t-test: $p < 0.001$ and $p < 0.05$ for the simulated cochlear positions of 10 and 15 mm, respectively).

DISCUSSION

The noiseband vocoder has previously been used to systematically manipulate speech and explore the effects of temporal and spectral parameters on speech perception by normal-hearing listeners. Fu & Shannon (1999) produced a spectral shift in the frequency-place mapping by moving the acoustic input range to higher or lower frequencies while the simulated stimulation range remained the same. Başkent & Shannon (2003) produced compressed and expanded frequency-place maps where the acoustic input range was made wider or narrower, respectively, than the simulated stimulation range. In both studies, the acute effects with no training were observed. The common finding was that speech recognition was best with frequency-place maps where the acoustic input range matched the simulated stimulation range, whereas introducing spectral distortions in the mapping in the form of a shift or compression-expansion generally decreased the performance.

The present study explored the combined effects of the two spectral distortions, shift and compres-

TABLE 4. Analysis and carrier band ranges shown in frequency (in Hz)

		Shift in carrier bands						
	Compression-expansion	−5 mm apical shift: Lower frequencies	−3 mm apical shift: Lower frequencies	−1 mm apical shift: Lower frequencies	0 mm No shift	+1 mm basal shift: Higher frequencies	+3 mm basal shift: Higher frequencies	+5 mm basal shift: Higher frequencies
Analysis/Carrier band range (in Hz)	+5 mm	513–20677	513–20677	513–20677	513–20677	513–20677	513–20677	513–20677
	Compression: Wider analysis band range	513–5860	722–7771	999–10290	1168–11837	1363–13612	1843–17990	2476–20677
	0 mm	1168–11837	1168–11837	1168–11837	Matched map: 1168–11837	1168–11837	1168–11837	1168–11837
	No compression-expansion	513–5860	722–7771	999–10290	1168–11837	1363–13612	1843–17990	2476–20677
	−5 mm Expansion:	2476–5860	2476–5860	2476–5860	2476–5860	2476–5860	2476–5860	2476–5860
	Narrower analysis band range	513–5860	722–7771	999–10290	1168–11837	1363–13612	1843–17990	2476–20677

Values are converted from cochlear positions presented in Table 3, using Greenwood's frequency-place mapping equation. The baseline matched map is shown in the middle entry. The middle row shows the map with shifting only. The middle column shows the map with compression-expansion only. The top and bottom rows show the combined maps where the shift was applied simultaneously with compression or expansion, respectively.

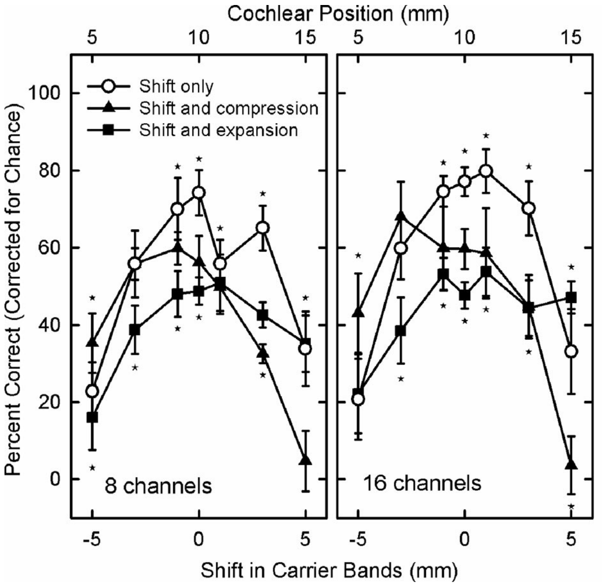

Fig. 2. Percent correct scores, averaged across subjects and corrected for chance, shown for vowel recognition as a function of the shift in the carrier bands. The results with the 8- and 16-channel processors are presented in the left and right panels, respectively. The performance with the baseline matched map is shown by the open circle at the 0 mm shift condition. The unshifted carrier bands in the matched map simulated a cochlear stimulation range with the most apical end at a distance of 10 mm from the apex. The open circles show the scores with the map where the carrier bands were shifted without compression or expansion, similar to the conditions presented in the middle rows of Figure 1 and Tables 1 and 2. The filled triangles and squares show the scores when the shift was combined with compression or expansion, respectively. These conditions were presented in the top and bottom rows, respectively, in Figure 1 and Tables 1 and 2. The error bars show one standard deviation. The stars above the scores show the shift conditions where adding the compression changed the performance significantly ($p < 0.05$), determined by a posthoc Tukey multiple comparisons test, applied following a two-factor Repeated Measures ANOVA. Similarly, the stars below the scores show the shift conditions where adding the expansion changed the performance significantly ($p < 0.05$).

sion-expansion, on speech recognition. Similar to previous studies, the matched frequency-place map usually produced the best performance, and applying either the shift or the compression-expansion resulted in lower scores. However, when the distortions were applied simultaneously, a compensatory effect was observed for some of the conditions; the reduction in the scores with the combined maps was smaller than the reduction caused by each map separately. For almost all apical shift conditions of –5 mm and for some of the shift conditions of –3 mm, where the carrier band spectrum was shifted to lower frequencies, the shifted and compressed map produced better performance than the shift only

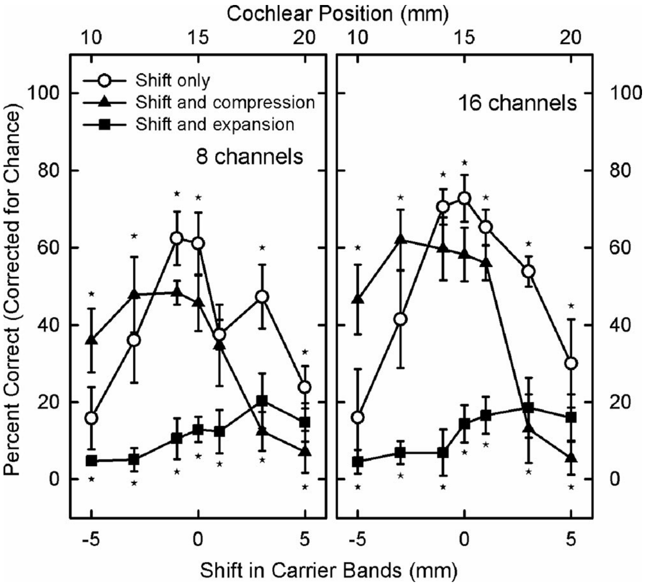

Fig. 3. Vowel recognition scores, averaged across subjects and corrected for chance, shown as a function of the shift in the carrier bands. Compared with Figure 2, the cochlear position of the simulated stimulation range was more basal; the most apical end of the unshifted carrier bands was at a simulated distance of 15 mm from the apex. The analysis and carrier band ranges of the maps are shown in Tables 3 and 4. Similar to Figure 2, the stars above and below the data show the conditions where adding compression or expansion, respectively, to the shift only map changed the performance significantly ($p < 0.05$).

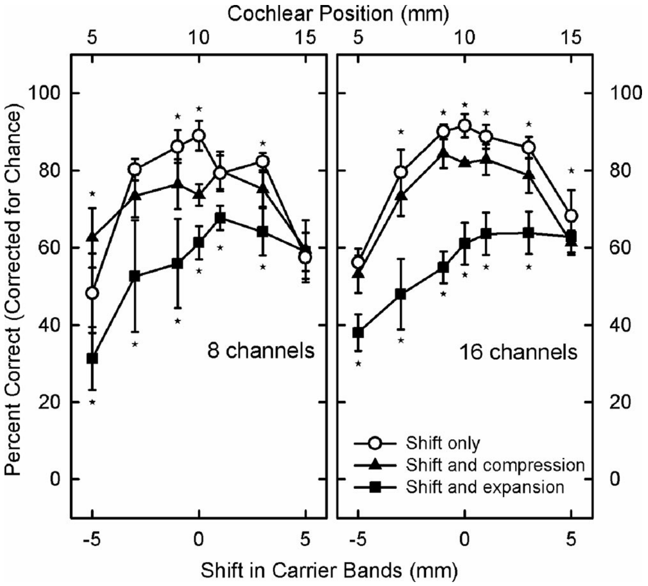

Fig. 4. Similar to Figure 2, except consonant recognition scores, averaged across subjects and corrected for chance, shown as a function of the shift in the carrier bands. The most apical end of the unshifted carrier bands was at a simulated distance of 10 mm from the apex.

map. A comparison was made between the individual analysis and carrier bands of these maps in an attempt to determine important frequency ranges for speech recognition. An example of this analysis is shown in Figure 6, where the absolute frequency shift in octaves is plotted for individual analysis bands as a function of the center frequencies of the bands. The maps shown were for the 8-channel vocoder with a simulated cochlear position of 15 mm from the apex, the same processor presented in Figure 3, left panel. In each panel, open squares show the baseline matched map, included as a reference. The open circles show the maps with the frequency shift only. In panels (a) to (d), the shift in carrier bands is –5 mm, –3 mm, +3 mm, and +5 mm, respectively. In the shift only conditions, the overall frequency ranges were the same for the analysis and carrier bands and the individual bands were uniformly shifted. In the ±3 mm shift only conditions the individual bands were shifted to a smaller degree than the ±5 mm shift only conditions. This is consistent with the higher scores observed with a ±3 mm shift compared to a ±5 mm shift (Figure 3, left panel, open circles). The filled circles in Figure 6 show the shift for individual analysis bands with the shifted and compressed map. In panels (a) and (b) frequencies lower than 4000 Hz were better matched than the high frequencies. The performance with these conditions was shown with the filled triangles at the –5 and –3 mm shift in Figure 3, left panel. Panels (c) and (d) show that with the compressed and shifted maps the high

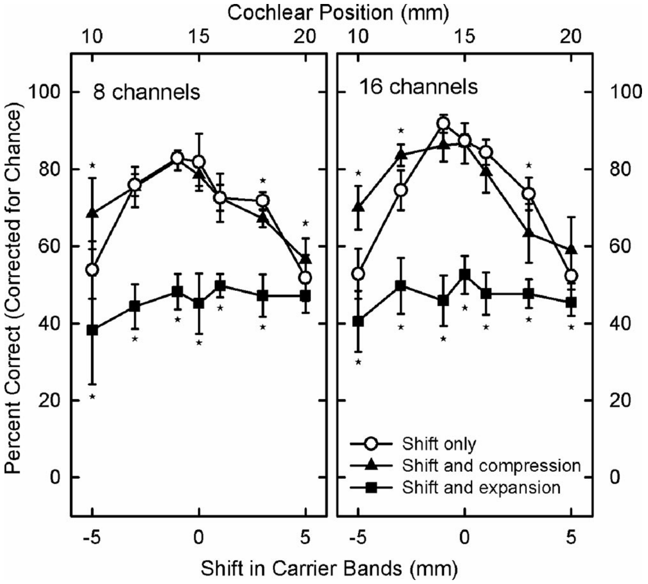

Fig. 5. Consonant recognition scores, averaged across subjects and corrected for chance, shown as a function of the shift in the carrier bands. Compared with Figure 4, the cochlear position of the simulated stimulation range was more basal; the most apical end of the unshifted carrier bands was at a simulated distance of 15 mm from the apex.

TABLE 5. F and *p* values of the two-factor RM ANOVA, shown for vowel recognition scores

	8 channels		16 channels	
	Shift and compression	Shift and expansion	Shift and compression	Shift and expansion
Matched carrier bands simulated at 10 mm apical distance	Effect of shift F(6,24) = 52.07 $p < 0.001$	Effect of shift F(6,24) = 75.82 $p < 0.001$	Effect of shift F(6,24) = 67.5 $p < 0.001$	Effect of shift F(6,24) = 88.86 $p < 0.001$
	Effect of added compression F(1,4) = 42.21 $p < 0.01$	Effect of added expansion F(1,4) = 57.20 $p < 0.01$	Effect of added compression F(1,4) = 42.21 $p < 0.01$	Effect of added expansion F(1,4) = 120.29 $p < 0.001$
	Interaction F(6,24) = 50.61 $p < 0.001$	Interaction F(6,24) = 13.03 $p < 0.001$	Interaction F(6,24) = 50.61 $p < 0.001$	Interaction F(6,24) = 39.17 $p < 0.001$
Matched carrier bands simulated at 15 mm apical distance	Effect of shift F(6,24) = 62.16 $p < 0.001$	Effect of shift F(6,24) = 42.59 $p < 0.001$	Effect of shift F(6,24) = 143.38 $p < 0.001$	Effect of shift F(6,24) = 61.61 $p < 0.001$
	Effect of added compression F(1,4) = 15.21 $p < 0.05$	Effect of added expansion F(1,4) = 483.94 $p < 0.001$	Effect of added compression F(1,4) = 30.40 $p < 0.01$	Effect of added expansion F(1,4) = 597.46 $p < 0.001$
	Interaction F(6,24) = 20.18 $p < 0.001$	Interaction F(6,24) = 19.81 $p < 0.001$	Interaction F(6,24) = 44.75 $p < 0.001$	Interaction F(6,24) = 35.49 $p < 0.001$

Numbers in parentheses show the degrees of freedom for the main factor of shift and the error term within this main factor (6 and 24, respectively), the main factor of added compression or expansion and the error term within this main factor (1 and 4, respectively), and the interaction and the error term within interaction (6 and 24, respectively).

frequencies, around 8000 Hz and higher, were better matched. Performance was higher for maps where the low frequencies were better matched (–3 and –5 mm) as compared to maps where the high frequencies were better matched (+3 and +5 mm). Let us compare the shift only map with the shifted and compressed map in each panel. In panels (a) and (b), the low frequencies, lower than around 4000 Hz, were better matched with the shifted and compressed map than the shift only map. The results from Figure 3, left panel, shows that the performance was better in these conditions with the compressed and shifted map, as shown by the filled triangles and open circles at the –5 and –3 mm shift conditions. In panels (c) and (d), the low frequencies were shifted more with the compressed and shifted map compared to the shift only map. With these conditions, the performance with the compressed and shifted map was worse than the shift only map, as it was shown in Figure 3, left panel, with the filled triangles and open circles at the +3 and +5 mm shift. This analysis shows that the compensatory effect was more commonly observed for the maps where the combination of the shift and compression resulted in the alignment of the low frequencies, especially around 1 to 2 kHz, even when the rest of the frequencies did not necessarily match. These frequencies have previously been shown to be most important for speech perception in studies where the intelligibility of narrowband speech with varying spectral content was measured (Warren et al., 1995; ANSI S3.5, 1997).

The carrier band ranges used in the matched conditions of the present study were at simulated cochlear positions of 10 and 15 mm from the apex, determined by Greenwood's mapping function. When the noiseband vocoder is used for simulating cochlear implants, the frequencies of the carrier bands represent the positions of the electrodes. If the average human cochlear length is assumed to be 35 mm, the carrier bands at the apical distances of 10 and 15 mm would simulate electrode arrays inserted to 25 and 20 mm, respectively, from the round window. The previous studies (Fu & Shannon, 1999; Başkent & Shannon, 2003) and the present study show the advantage of a matched frequency-place map for speech recognition, for electrodes simulated to be located deeper than 20 mm. However, Ketten et al. (1998) observed insertions as shallow as 14 mm from the round window when CT scans from implant users were used to estimate the electrode array insertion depth. Faulkner et al. (2003) simulated shallow insertions of cochlear implant electrode arrays using the noiseband vocoder. For simulated insertions less than 19 mm, there was a significant reduction in speech recognition with

TABLE 6. F and *p* values of the two-factor RM ANOVA, shown for consonant recognition scores

	8 channels		16 channels	
	Shift and compression	Shift and expansion	Shift and compression	Shift and expansion
Matched carrier bands simulated at 10 mm apical distance	Effect of shift F(6,24) = 42.04 $p < 0.001$	Effect of shift F(6,24) = 31.97 $p < 0.001$	Effect of shift F(6,24) = 80.71 $p < 0.001$	Effect of shift F(6,24) = 39.42 $p < 0.001$
	Effect of added compression F(1,4) = 4.26 $p = 0.11$	Effect of added expansion F(1,4) = 60.95 $p < 0.001$	Effect of added compression F(1,4) = 34.77 $p < 0.01$	Effect of added expansion F(1,4) = 141.37 $p < 0.001$
	Interaction F(6,24) = 8.58 $p < 0.001$	Interaction F(6,24) = 8.75 $p < 0.001$	Interaction F(6,24) = 8.83 $p = 0.56$	Interaction F(6,24) = 12.77 $p < 0.001$
Matched carrier bands simulated at 15 mm apical distance	Effect of shift F(6,24) = 51.83 $p < 0.001$	Effect of shift F(6,24) = 16.25 $p < 0.001$	Effect of shift F(6,24) = 61.12 $p < 0.001$	Effect of shift F(6,24) = 23.38 $p < 0.001$
	Effect of added compression F(1,4) = 4.51 $p = 0.10$	Effect of added expansion F(1,4) = 988.05 $p < 0.001$	Effect of added compression F(1,4) = 1.99 $p = 0.23$	Effect of added expansion F(1,4) = 735.95 $p < 0.001$
	Interaction F(6,24) = 8.00 $p < 0.001$	Interaction F(6,24) = 17.16 $p < 0.001$	Interaction F(6,24) = 8.63 $p < 0.001$	Interaction F(6,24) = 30.95 $p < 0.001$

the matched map. With such shallow insertions, when the acoustic input range is truncated to fully match the stimulation range, important low frequencies for speech, i.e. less than 2 kHz, would be excluded from the map. Similar results were observed by Başkent and Shannon (2005) who simulated shallow insertions by turning off apical electrodes of implant users. For insertions from 17 to 19 mm, optimal speech recognition was observed with a moderately compressed map; the matched map that missed important low-frequency information and the fully compressed map that introduced substantial spectral distortion both resulted in poorer performance. The present study complements these results and shows that when there is substantial spectral distortion, matching the frequencies that are important for speech perception might be more beneficial than matching the overall acoustic input range.

Note that the compensatory effect was observed for a small portion of the conditions tested in the present study. Adding expansion to the shift only map produced a significant reduction in the scores for all shift conditions other than the basal shift of +5 mm. This finding was expected as the bandwidth of the acoustic input to the vocoder was limited in the expansion conditions. Başkent and Shannon (2003) showed that, under similar conditions, narrowing overall bandwidth of the analysis and carrier bands decreased the performance; adding expansion by widening the carrier band range caused further reduction. When compression was added to the shift only map, more of the speech spectrum was included in the analysis band range. Adding compression to the shift only map reduced the consonant recognition scores only by a small amount; however, the additional reduction in the vowel recognition scores was substantial for all shift conditions other than the apical shifts of –3 and –5 mm. This finding implies that if the insertion depth of an electrode array could be determined with an accuracy of a few mm, a frequency-place map that is determined using the estimated insertion depth could still be beneficial to implant users.

One should be cautious about applying the results from simulations with normal-hearing listeners to actual implant users, as there are numerous differences between acoustic and electric hearing. In electric stimulation, there are many parameters such as the exact position of the electrode array or the length of the cochlea that are not known precisely unless imaging techniques such as CT scans are used (Ketten et al., 1998). Factors such as the proximity of electrodes to the modiolus (Saunders et al., 2002) and the nerve survival pattern (Khan et al., 2005) might affect speech perception. Most im-

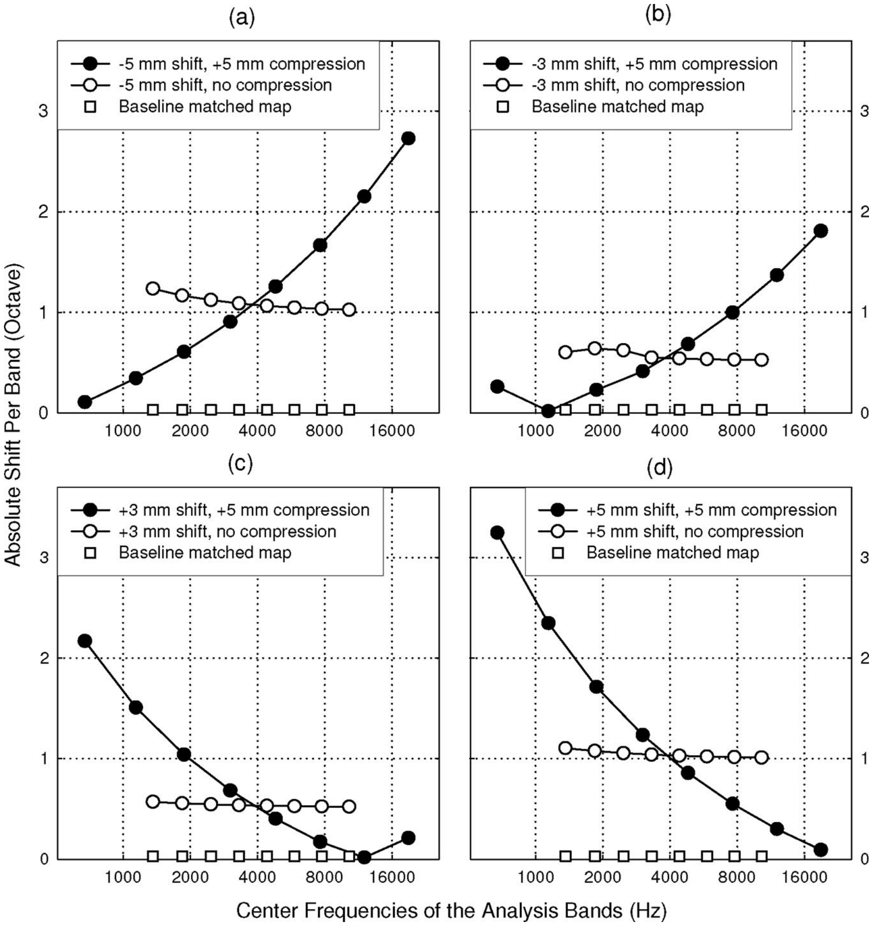

Fig. 6. Absolute frequency shift for the individual analysis bands shown in octaves as a function of the center frequencies of the analysis bands. The processing was the 8-channel vocoder with a simulated stimulation range at the cochlear position of 15 mm. In panels (a) to (d) mapping conditions are compared for a shift of –5 mm, –3 mm, +3 mm, and +5 mm in carrier bands, respectively. In each panel, open squares show the baseline matched map, the open circles show the shift only map, and the filled circles show the shifted and compressed map.

portantly, the Greenwood equation, which was used in determining the acoustic input frequencies in the experimental conditions of the present study, might not be applicable to electrical stimulation of the auditory nerve because it was originally formulated for the frequency response along the organ of Corti in the healthy ear. Recent studies indicated that the electrode-pitch function of electric hearing might, in fact, be different than the frequency-place mapping of acoustic hearing. Baumann & Nobbe (2006), for example, used acoustic tones as stimuli and asked implant users with residual acoustic hearing in the contralateral ear to match the pitch between the two ears. Many subjects reported lower pitch for the stimulation of basal electrodes than the values estimated by Greenwood's equation, and the electrode-pitch function was estimated to be more linear than the frequency-place function. Therefore, the findings of the simulation studies should be perceived as preliminary data and be cautiously interpreted for real-life applications with implant users.

The results of the present study were acute effects observed immediately following the changes in the frequency-place maps. The subjects were all experienced from previous participation in similar experiments; however, they were not provided an adaptation period, other than a preview of the processed stimuli, nor a special training for any of the maps used in the present study. Recent studies have shown strong learning effects with normal-hearing subjects. Rosen et al. (1999) used connected discourse tracking as an interactive training method with a shifted map. After nine sessions of 20 minutes, there was a significant improvement in speech recognition. However, the improved performance with the shifted map at the end of the experiment was lower than the matched map; it is not clear if a complete adaptation would be possible. Fu et al. (2002) programmed shifted maps in the processors of implant users. At the end of an adaptation period of 3 mo, the scores with the shifted map were lower than the baseline scores that were obtained with the user's clinical map. Studies by Svirsky et al. (2004) and Fu et al. (2005) showed that adaptation as a result of daily exposure could in fact be a very lengthy process.

These studies that show the acute effects of spectral distortions on speech recognition, including the present study, can be used as guidelines for optimal fitting of cochlear implants in initial sessions, which could facilitate an easier adaptation subsequently. In a simulation study, Svirsky et al. (2003) showed that the initial performance was better with a smaller frequency shift, and this advantage persisted for several 1-hr sessions. It is also possible that adaptation to a shifted map could be faster and easier for the patient if the shift was introduced in smaller steps gradually. With this idea, implant users could initially be fit with a map that is more similar to the normal mapping of acoustic hearing and has minimal spectral distortions. Such a map might also provide better sound quality. The acoustic input range can then be extended, to provide more acoustic input to the listener, and in gradual steps, for a possibly easier adaptation. A targeted training in each step, as suggested by Fu et al. (2005), could further help the implant user.

ACKNOWLEDGMENTS

The authors would like to thank John Galvin for help with the experimental setup, Andrew Faulkner, Johan Laneau, and an anonymous reviewer for helpful comments on a previous version of the manuscript, and the subjects for their enthusiastic participation. Funding for the research was provided in part by NIDCD Grant R01-DC-01526 and Contract N01-DC-92100.

Address for correspondence: Dr. Deniz Başkent, Starkey Hearing Research Center, 2150 Shattuck Avenue, Ste. 408, Berkeley CA 90704. E-mail: deniz_baskent@starkey.com.

Received January 6, 2005; accepted December 6, 2006.

REFERENCES

American National Standards Institute (1997). Methods for Calculation of the Speech Intelligibility Index, S35.

Başkent, D. (2003). Speech recognition under conditions of frequency-place compression and expansion. *PhD Thesis*, University of Southern California, CA.

Başkent, D., Shannon, R. V. (2003). Speech recognition under conditions of frequency-place compression and expansion. *Journal of the Acoustical Society of America*, *113*, 2064–2076.

Başkent, D., Shannon, R. V. (2004). Frequency-place compression and expansion in cochlear implant listeners. *Journal of the Acoustical Society of America*, *116*, 3130–3140.

Başkent, D., Shannon, R. V. (2005). Interactions between cochlear implant electrode insertion depth and frequency-place mapping. *Journal of the Acoustical Society of America*, *117*, 1405–1416.

Başkent, D. (2006). Speech recognition in normal hearing and sensorineural hearing loss as a function of spectral channels. *Journal of the Acoustical Society of America*, *120*, 2908–2925.

Baumann, U., Nobbe, A. (2006). The cochlear implant electrode-pitch function. *Hearing Research*, *213*, 34–42.

Braida, L. D., Durlach, N. I., Lippmann, R. P., Hicks, B. L., Rabinowitz, W. M., Reed, C. M. (1979). Hearing aids: A review of past research on linear amplification, amplitude compression, and frequency lowering. *ASHA Monograph No*, *19*, 1–114.

Faulkner, A., Rosen, S., Stanton, D. (2003). Simulations of tonotopically mapped speech processors for cochlear implant electrodes varying in insertion depth. *Journal of the Acoustical Society of America*, *113*, 1073–1080.

Fu, Q.-J., Shannon, R. V., Wang, X. (1998). Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing. *Journal of the Acoustical Society of America*, *104*, 3586–3596.

Fu, Q.-J., Shannon, R. V. (1999). Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing. *Journal of the Acoustical Society of America*, *105*, 1889–1900.

Fu, Q.-J., Shannon, R. V., Galvin, J. J. (2002). Perceptual learning following changes in the frequency-to-electrode assignment with the Nucleus-22 cochlear implant. *Journal of the Acoustical Society of America, 112*,1664–1674.

Fu, Q.-J., Nogaki, G., Galvin, J. J. (2005). Auditory training with spectrally shifted speech: Implications for cochlear implant patient auditory rehabilitation. *Journal of the Association for Research in Otolaryngology, 6*, 180–189.

Greenwood, D. D. (1990). A cochlear frequency-position function for several species—29 years later. *Journal of the Acoustical Society of America*, *87*, 2592–2605.

Henry, B. A., Turner, C. W., Behrens, A. (2005). Spectral peak resolution and speech recognition in quiet: Normal hearing, hearing impaired, and cochlear implant listeners. *Journal of the Acoustical Society of America*, *118*, 1111–1121.

Hillenbrand, J., Getty, L. A., Clark, M. J., Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, *97*, 3099–3111.

Ketten, D. R., Skinner, M. W., Wang, G., Vannier, M. W., Gates, G. A., Neely, J. G. (1998). In vivo measures of cochlear length and insertion depth of Nucleus cochlear implant electrode arrays. *Annals of Otology, Rhinology and Laryngology Supplement*, *175*, 1–16.

Khan, A. M., Whiten, D. M., Nadol, J.B. Jr., Eddington, D. K. (2005). Histopathology of human cochlear implants: Correlation of psychophysical and anatomical measures. *Hearing Research*, *205*, 89–93.

Oxenham, A. J., Bacon, S. P. (2003). Cochlear compression: Perceptual measures and implications for normal and impaired hearing. *Ear and Hearing*, *24*, 352–366.

Robert, M. E. (1998). CONDOR: Documentation for Identification Test Program. House Ear Institute, Los Angeles, CA.

Roberts, M. J., Russo, R. (1999). A student's guide to Analysis of Variance. Routledge, London, U.K.

Rosen, S., Faulkner, A., Wilkinson, L. (1999). Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants. *Journal of the Acoustical Society of America*, *106*, 3629–3636.

Saunders, E., Cohen, L., Aschendorff, A., Shapiro, W., Knight, M., Stecker, M., Richter, B., Waltzman, S., Tykocinski, M., Roland, T., Laszig, R., Cowan, R. (2002). Threshold, comfortable level and impedance changes as a function of electrode-modiolar distance. *Ear and Hearing*, *23*:28S–40S.

Shannon, R. V., Zeng, F-G., Kamath, V., Wygonski, J., Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*, 303–304.

Stelmachowicz, P. G., Jesteadt, W., Gorga, M. P., Mott, J. (1985). Speech perception ability and psychophysical tuning curves in hearing-impaired listeners. *Journal of the Acoustical Society of America*, *77*, 620–627.

Svirsky, M. A., Sinha, S., Neuburger, H. S., Talavage, T. M. (2003). Gradual adaptation to shifts in the peripheral acoustic frequency map. *Twenty-Sixth Midwinter Meeting, Association for Research in Otolaryngology,* Daytona Beach, FL.

Svirsky, M. A., Silveira, A., Neuberger, H., Teoh, S. W., Suarez, H. (2004). Long-term auditory adaptation to a modified peripheral frequency map. *Acta Oto-laryngologica*, *12*, 381–386.

Turner, C. W., Hurtig, R. R. (1999). Proportional frequency compression of speech for listeners with sensorineural hearing loss. *Journal of the Acoustical Society of America*, *106*, 877–886.

Turner, C. W., Chi, S.-L., Flock, S. (1999). Limiting spectral resolution in speech for listeners with sensorineural hearing loss. *Journal of Speech and Hearing Research*, *42*, 773–784.

Warren, R. M., Riener, K. R., Bashford, J. A. Jr., Brubaker, B. S. (1995). Spectral redundancy: Intelligibility of sentences heard through narrow spectral slits. *Perception and Psychophysics*, *57*, 175–182.