

Published in final edited form as:

*J Acoust Soc Am.* 2008 April ; 123(4): 2295–2309. doi:10.1121/1.2831738.

## Effects of upper-frequency boundary and spectral warping on speech intelligibility in electrical stimulation

Matthew J. Goupell<sup>a)</sup>, Bernhard Laback, and Piotr Majdak

Acoustics Research Institute, Austrian Academy of Sciences, Wohllebengasse 12-14, A-1040 Vienna, Austria

Wolf-Dieter Baumgartner

ENT-Department, Vienna University Hospital, Währinger Gürtel 18-20, A-1097 Vienna, Austria

### Abstract

Speech understanding was tested for seven listeners using 12-electrode Med-El cochlear implants (CIs) and six normal-hearing listeners using a CI simulation. Eighteen different types of processing were evaluated, which varied the frequency-to-tonotopic place mapping and the upper boundary of the frequency and stimulation range. Spectrally unwarped and warped conditions were included. Unlike previous studies on this topic, the lower boundary of the frequency and stimulation range was fixed while the upper boundary was varied. For the unwarped conditions, only eight to ten channels were needed in both quiet and noise to achieve no significant degradation in speech understanding compared to the normal 12-electrode speech processing. The unwarped conditions were often the best conditions for understanding speech; however, small changes in frequency-to-place mapping ( $<0.77$  octaves for the most basal electrode) yielded no significant degradation in performance from the nearest unwarped condition. A second experiment measured the effect of feedback training for both the unwarped and warped conditions. Improvements were found for the unwarped and frequency-expanded conditions, but not for the compressed condition. These results have implications for new CI processing strategies, such as the inclusion of spectral localization cues.

### I. INTRODUCTION

Large differences are seen in the population of cochlear-implant (CI) listeners for speech understanding, with high-performing CI listeners often being comparable to normal-hearing (NH) listeners using CI simulations in quiet conditions (Friesen *et al.*, 2001). Probable reasons for the large variability in CI performance for understanding speech are a reduced number of functional auditory nerve fibers (dead regions), spread of electrical current fields in the cochlea, and the positioning of the electrodes in relation to the tonotopic place of stimulation, hence the characteristic frequencies of the stimulated auditory nerve fibers. Greenwood's (1990) organ of Corti map has been widely used to relate tonotopic place in a CI to characteristic frequency, although a recent study shows that a spiral ganglion map may be a better model for perimodiolar electrodes (Sridhar *et al.*, 2006). Typically, the most apical electrode is inserted 20–30 mm into the cochlea (Ketten *et al.*, 1998; Gstöettner *et al.*, 1999). According to the Greenwood map, for a 35-mm cochlea, a 30-mm insertion depth corresponds to a 160-Hz center frequency, while a 20-mm insertion depth corresponds to an 1150-Hz center frequency. The lowest frequency electrode often receives frequency information around 300-Hz, yielding a large frequency-to-place mismatch for a 20-mm

insertion depth. Even larger mismatches have been reported; Ketten *et al.* showed four patients to have their most apical electrode at a cochlear place tuned to greater than 1.4 kHz.

To address the large variability of speech understanding in CI listeners, different frequency-to-place mappings have been studied to optimize speech understanding in CI listeners with the hope of increasing understanding to a level comparable to NH listeners. A number of studies by Fu and Shannon (1999a, b, c, d) have shown the importance of well-matched frequency-to-place maps for both CI listeners and NH listeners using a CI simulation. Ba Kent and Shannon showed matched conditions were the best for NH listeners (2003) and CI listeners (2004) when compared to frequency-expanded or frequency-compressed speech. However, frequency-to-place maps need not be perfectly matched. The effect of insertion depth was studied by Dorman *et al.* (1997) with NH simulations. The best performance was seen for a well-matched map, but apical shifts of 0.24 octaves could be tolerated without a significant decrease in speech understanding.

To support the case for unmatched maps, Faulkner *et al.* (2003), from NH simulation data, argued that necessary speech information below 1 kHz is lost if frequencies are matched in the case of shallow insertion depths. Ba Kent and Shannon (2005) also showed with CI listeners the importance of low-frequency information in a study that held the upper-frequency boundary fixed while varying the lower-frequency boundary. This is one of the few studies that shows an improvement over the matched frequency-to-place condition; for CI listeners with shallower insertions, a mild amount of compression was better than truncating the frequency range to match frequency and tonotopic place.

Despite the variability in the effectiveness of a CI due to the uncertainty of the frequency-to-place map, it turns out that CI listeners can adapt fairly well to frequency mismatches that might be due to shallow insertions or dead regions, hence the successful nature of the field of cochlear implantation as a whole. Most of the previously cited studies on CI speech understanding were acute studies, meaning that there was no long-term training. Rosen *et al.* (1999), Fu *et al.* (2002), McKay and Henshall (2002), and Smith and Faulkner (2006) found that short-term experiments can exaggerate the long-term consequences of spectral shifts.

We studied the frequency-to-place mapping of speech in CI listeners by systematically varying the upper-frequency boundary and range of electrodes at different signal-to-noise ratios (SNRs). We decreased the number of electrodes used from the standard 12 electrodes in the Med-El Combi 40+ or Pulsar implants and we varied the upper-frequency boundary from the standard 8.5 kHz while holding the lower-frequency boundary fixed. Hence, unlike some frequency-to-place mapping studies, which attempt to improve speech understanding scores, we determined the amount of “extra” speech information presented to CI listeners in current strategies and attempted to not appreciably decrease speech scores. Determining the amount of extra speech information presented to CI listeners has implications for new CI processing strategies; namely the inclusion of acoustic information other than speech that is presently neglected.

It is unclear if a compression, expansion, or simple low-pass filtering of the speech information will cause significant changes in speech understanding. Skinner *et al.* (1995) argued that improved vowel discrimination in a group of CI listeners was due to assigning more electrodes to important lower-frequency regions. That is to say it may be possible to increase speech understanding if there is more spectral resolution over the first two formants. A study by McKay and Henshall (2002) tested a similar idea with CI listeners. They tested two frequency-to-place mappings in their study: one that evenly spread ten electrodes over a frequency range of 200–10,000 Hz and one that used nine electrodes for 200–2600 Hz and the last electrode for frequencies 2600–10,000 Hz. They found that

assigning more electrodes to lower frequencies helped vowel recognition and sentence recognition, but hurt consonant recognition. Also, a recent study by Ba Kent and Shannon (2007) showed that there is a greater degradation of speech understanding for low-frequency spectral shifts than high-frequency spectral shifts. Therefore our manipulation of varying only the upper-frequency boundary and upper-stimulation boundary is a practically useful manipulation because it attempts to preserve the spectral information for the low-frequency electrodes.

## II. EXPERIMENT 1

### A. Methods

**1. Stimuli**—The Oldenburg Sentence Test (OLSA) (Wagener *et al.*, 1999a, b, c) was used to test speech intelligibility of words under different mapping conditions. The OLSA Test consists of five-word, nonsense sentences. Each sentence is constructed of a name, a verb, a number, an adjective, and an object in that order. Ten different German words comprise a set for each word category. Each word has equal intelligibility. The noise used to mask the sentences has a spectrum that is similar to the spectrum of the speech. The sentences were recorded at a sampling rate of 44.1 kHz.

**2. Signal processing**—For the electrical stimuli, the sentences were processed using the Continuous Interleaved Sampling (CIS) strategy (Wilson *et al.*, 1991). First, pre-emphasis was applied to the signal using a first-order, high-pass Butterworth filter with a corner frequency at 1200 Hz. Next, the signal was filtered into frequency bands by using fourth-order Butterworth filters. The speech information up to the upper-frequency boundary (determined by  $M$ ) was mapped to the number of channels ( $N$ ). The number  $M$  is the frequency-boundary index that represents the upper-frequency boundary ( $f_{\text{upper}}$ ) determined by logarithmic spacing of the bands; hence this number enumerates the spectral content of the stimuli. Figure 1 shows how the speech spectrum is subdivided and enumerated by  $M$ . Different experimental conditions were created in which the upper-frequency boundary of the speech stimuli was set at either 8500 Hz ( $M=12$ ), 4868 Hz ( $M=10$ ), 2788 Hz ( $M=8$ ), or 1597 Hz ( $M=6$ ). The lower-frequency cutoff was always 300 Hz. For each filter configuration, the stimuli were band-pass filtered into either  $N=6, 8, 10$ , or 12 contiguous logarithmically-spaced analysis bands, thus creating 16 different filter sets. In conditions that utilized 6, 8, or 10 analysis bands, the filters were respectively mapped to the 6, 8, or 10 most apical electrodes of the CI listeners' implanted arrays. Conditions utilizing 12 analysis bands mapped the filters to the full 12-electrode array. Therefore matched (unwarped) frequency-to-place maps that vary only the stimulation range were tested. Also, unmatched (warped) maps were tested. See Fig. 1 for an illustration of the types of frequency-to-place mappings. Table I shows the center frequency, bandwidth, and shift for each channel for each combination of  $M$  and  $N$  used in this experiment. The baseline condition,  $M=12$  and  $N=12$ , is labeled  $M_{12}N_{12}$  and corresponds to the normal clinical settings of the CI. The other conditions are labeled similarly. Besides the 16 different filter sets, there were two additional conditions, the  $M_4N_4$  case ( $f_{\text{upper}}=915$  Hz) and an extended frequency range case to be called  $M_{14}N_{12}$ . The extended frequency range had  $f_{\text{upper}}=16,000$  Hz. An illustrative matrix of all the testing conditions can be seen in Fig. 2.<sup>1</sup>

<sup>1</sup>Two CI listeners, CI16 and CI18, had a lower-frequency boundary of 200 Hz to better match their clinical processor settings. One listener, CI17, had a lower-frequency boundary of 250 Hz. Therefore their center frequencies are shifted downward compared to those found in Table I. For the  $M_6N_{10}$  and  $M_6N_{12}$  conditions, an unstable filter would be used when the lower-frequency boundary is less than 300 Hz. Therefore, these two conditions had  $f_{\text{lower}}=300$  Hz for all 13 listeners, even though this differs from the clinical settings for CI16, CI17, and CI18. The three subjects reported no perceptual abnormality for these two conditions compared to the other conditions in this experiment.

After the signals were filtered into  $N$  analysis bands, the Hilbert envelope of each of the bands was calculated and then was compressed by a logarithmic map law. Compression factors, individual electrode threshold levels, and most comfortable levels were set to match the CI listeners' everyday clinical settings. The sentences were presented at a comfortable level, which was chosen by each CI listener.<sup>2</sup>

Like the electrical stimuli, the sentences for the acoustic stimuli for the NH listeners were pre-emphasized. The same procedure was used to select corner frequencies for matched and unmatched frequency-to-place conditions. The signal was filtered into  $N$  bands by fourth-order Butterworth filters and the Hilbert envelope of the bands was calculated. Each speech envelope was multiplied by a narrowband white noise. The noises had a bandwidth and center frequency corresponding to the assumed logarithmic spacing of channels (see Table I,  $M_{12}N_{12}$ ). The  $N$  speech-envelope-modulated noise bands were then summed over the bands into an acoustic stimulus. The sampling rate of the vocoded sentences was 44.1 kHz. The acoustic stimuli were normalized to have approximately the same energy. Sentences were presented at approximately 75 dB (A) SPL.<sup>3</sup> Because the noise vocoder introduced random noise to the stimuli, the level of the stimuli was not precisely controlled. The level varied by as much as  $\pm 0.5$  dB on a given sentence.

**3. Procedure**—The task of the listener was to verbally repeat the five words of the OLSA sentences in blocks where the upper-frequency boundary index,  $M$ , and number of channels,  $N$ , were fixed. No feedback was provided and listeners were encouraged to guess if they were not sure about the correct words, although they were cautioned not to provide the same response for each guess. An experimenter, in the same room as the listener, recorded the number of correct words. At the beginning of each block, ten sentences were presented in quiet so that listeners could adapt to stimulus characteristics that changed from the stimulus processing (e.g., different frequency information or differently warped spectral content). After the initial ten sentences, listeners showed no or insignificant learning over the block. The effects of learning from feedback training over several blocks are explored in experiment 2. Within each block, sentences were presented at four SNRs in random order. The sentences were presented in quiet and at SNR=+10, +5, and 0 dB. Twenty repetitions were presented for each SNR. Thus, including the ten sentence adaptation period, 90 sentences were presented in each block. The subjects had their own randomly generated sentences, meaning that sentences were different from trial to trial and listener to listener. The percent correct ( $P_c$ ) was calculated from the number of correct words in each five-word sentence from the last 80 sentences. A single block took approximately 15–20 min to complete.

Listeners were procedurally trained without feedback on the  $M_{12}N_{12}$  condition until stable performance was achieved and the procedural training period was finished (4–14 blocks). After the  $M_{12}N_{12}$  condition was completed, listeners were presented the other 17 conditions in random order. The randomization of condition order also helped minimize any learning effects for the experiment.

<sup>2</sup>Each CI listener adjusted the stimulus to a comfortable level for each value of  $N$ . There was no systematic difference in the adjusted level across different  $N$  values. Therefore, for each listener a common level was used for all  $N$  values.

<sup>3</sup>For the NH listeners, a preliminary experiment to balance the loudness of the processed sentences was attempted to eliminate possible loudness summation effects from using sentences with different bandwidths. A single sentence was presented processed in two ways:  $M_{12}N_{12}$  and  $M_XN_X$  where  $X=4, 6, 8, 10, \text{ or } 12$ . For  $X=8, 10, \text{ or } 12$ , the measured loudness difference was at most  $\pm 1.5$  dB. For  $X=4$  or  $6$ , the measured loudness difference was at most  $\pm 3.5$  dB. There was a large variability between listeners in the level balancing results, which was probably due to the difficulty of matching the loudness of two sentences with different frequency content. When the loudness balancing results were applied to the sentences in pilot tests, results did not significantly change. In the end, we decided to present the sentences at a nominally equal level of 75 dB (A) SPL.

**4. Listeners and equipment**—Seven CI listeners were tested in this experiment and all were native German speakers. The etiologies of the CI listeners are shown in Table II. All were unilaterally implanted, except CI13, who was bilaterally implanted. All seven listeners used Med-El 12-electrode implants: four CI listeners used the C40+ implant and three CI listeners used the Pulsar implant. In everyday listening, the C40+ users wear the Tempo +processor with the CIS+strategy and the Pulsar users wear the OPUS processor with the FSP strategy. In the experiments, all the listeners were presented with the CIS strategy implemented on a laboratory computer. More details on the Med-EL C40+ can be found in Zierhofer *et al.* (1997).

The stimuli were presented via a Research Interface Box, developed at the University of Innsbruck, Austria. For the one bilateral CI user, CI13, the stimuli were presented to the better ear, which was the left. One listener, CI15, had only the 11 most-apical electrodes activated in her clinical setting. In her case, all the conditions with  $M=12$  or  $N=12$  were replaced by  $M=11$  or  $N=11$  respectively, but her data is analyzed and plotted as it were the  $M=12$  or  $N=12$  condition.

Six NH listeners were tested in this experiment. All the listeners were native German speakers except two of the NH listeners. These two listeners were fluent in German and showed no difference in performance compared to the native German speakers. Two NH listeners (NH2, a native German speaker, and NH10, a native English speaker) were the authors of the paper. They had extensive training with the OLSA material for several different processing conditions before data were taken. All six NH listeners had normal hearing according to standard audiometric tests. Since none of the NH listeners had an asymmetric hearing loss, the right ear was used for all of them. They were presented sentences over headphones (Sennheiser HDA200), after amplification from a headphone amplifier (TDT HB6) and passing a programmable attenuator (TDT PA4). The subjects were seated in a double-walled sound booth.

## B. Matched conditions

**1. Results**—Figure 3 shows the mean  $P_c$  scores of the CI and NH listeners for the matched conditions. For these conditions, the upper boundary of the matched stimulation range was varied by changing the number of channels (electrodes for CIs and analysis bands for NHs). The horizontal axis shows the number of matched channels, where  $N=M$ , and is referred to as  $N'$ . The additional horizontal axis at the top shows the corresponding upper-frequency boundary. Note that our definition of “changing the number of channels” does not change the inter-channel distance as the phrase is usually applied in other studies (e.g. Shannon *et al.*, 1995). The CI listeners were split into two groups: a high-performance group and a low-performance group. The high-performance CI group (upper-left panel) consists of five CI listeners and each had  $P_c > 90\%$  for  $M_{12}N_{12}$  in quiet, comparable to the NH listeners (upper-right panel). The low-performance CI group (lower two panels) consists of listeners CI17 and CI19 and had  $P_c \approx 75\%$  for  $M_{12}N_{12}$  in quiet. The error bars are two standard deviations in overall length.

The performance improved with increasing  $N'$  for both the CI and NH listeners for every SNR. The effects of the factors  $N'$  and SNR were analyzed by a two-way repeated-measures analysis of variance (RM ANOVA), performed separately for the group of CI listeners ( $n=7$ ) and the group of NH listeners ( $n=6$ ). For all analyses based on ANOVA described in this paper, the  $P_c$  scores were transformed using the rationalized arcsine transform proposed by Studebaker (1985) to not violate the homogeneity of variance assumption required for an ANOVA.

For the CI listeners, the main effects were highly significant for  $N'$  [ $F(4,114)=74.4$ ;  $p<0.0001$ ] and SNR [ $F(3,114)=98.8$ ;  $p<0.0001$ ], but there was no interaction between the two factors [ $F(12,114)=0.57$ ;  $p=0.86$ ]. For the NH listeners, the main effects were also highly significant for  $N'$  [ $F(4,95)=129.9$ ;  $p<0.0001$ ] and SNR [ $F(3,95)=214.6$ ;  $p<0.0001$ ]. In contrast to the CI listeners, there was a highly significant interaction between the two factors [ $F(12,95)=4.94$ ;  $p<0.0001$ ], reflecting the smaller improvement with increasing  $N'$  at the higher SNR, which may indicate a ceiling effect.

Helmert contrasts (Chambers and Hastie, 1993) were calculated to determine the level of the factor  $N'$  above which no further improvement in performance occurs. Helmert contrasts test the difference of the performance between a given  $N'$  and the average performance of all the higher  $N'$ . For example, the percent correct for  $N'=6$  [ $P_c(N'_6)$ ] can be compared to the average percent correct for  $N'=8, 10$ , and  $12$  [ $\bar{P}_c(N'_{>6})$ ]. This method has the advantage of a greater statistical power when compared to adjacent  $N'$  comparisons. The contrasts were calculated separately for each SNR.

For all seven CI listeners, increasing  $N'$  resulted in a significant improvement of performance up to  $N'=8$  in quiet and at the +5-dB SNR [ $P_c(N'_6)$  vs  $\bar{P}_c(N'_{>6})$ : largest  $p=0.0001$ ;  $P_c(N'_8)$  vs  $\bar{P}_c(N'_{>8})$ : smallest  $p=0.061$ ]. At the +10- and 0-dB SNRs, performance improved up to  $N'=10$  [ $P_c(N'_8)$  vs  $\bar{P}_c(N'_{>8})$ : largest  $p=0.013$ ;  $P_c(N'_{10})$  vs  $P_c(N'_{12})$ : smallest  $p=0.44$ ]. Said differently, the upper-frequency boundary can be reduced without a significant decrease in performance from 12 to 8 or 10 channels, depending on the SNR. For the five high-performance CI listeners at the three larger SNRs, reducing  $N'$  from 12 to 8 resulted in an average decrease in  $P_c$  of 4.9%. For the two low-performance CI listeners, the performance either dropped exceptionally quickly (CI17) or slowly (CI19) when the number of channels decreased below ten.

For the NH listeners, increasing  $N'$  resulted in a significant improvement of performance up to  $N'=8$  in quiet [ $P_c(N'_6)$  vs  $\bar{P}_c(N'_{>6})$ : largest  $p=0.004$ ;  $P_c(N'_8)$  vs  $\bar{P}_c(N'_{>8})$ : smallest  $p=0.89$ ] and up to  $N'=10$  for all conditions with noise [ $P_c(N'_8)$  vs  $\bar{P}_c(N'_{>8})$ : largest  $p=0.011$ ;  $P_c(N'_{10})$  vs  $P_c(N'_{12})$ : smallest  $p=0.63$ ].

**2. Discussion**—Speech understanding of CI listeners was found to improve significantly when the number of matched channels increased up to eight to ten, which corresponds to an upper boundary of the stimulation range of about 2.8 to 4.9 kHz. A different way of interpreting this result is that speech understanding did not decrease substantially when the number of matched channels was reduced to the eight to ten apical-most channels depending on the SNR. All CI listeners showed decreasing performance with decreasing SNR, but the effect of the stimulation range was similar at each SNR. The NH listeners, listening to vocoded speech, showed improvements in performance up to eight channels in quiet and up to ten channels for the conditions with noise. Considering the similar improvements in  $P_c$  above  $N'$  equal eight for the CI and NH listeners, the “critical” upper boundary of the stimulation range above which the performance saturates was quite similar for the two groups. This comparison is of course based on the assumption that the NH results at higher SNRs were not affected by the ceiling.

Most studies on the effect of the number of channels on speech understanding with CIs involve varying the number of electrodes over a fixed stimulation range. Thus, the CI listeners' results for the matched condition can be compared only with NH results. The NH listeners' performance as a function of  $N$  is roughly in line with predictions of the speech intelligibility index (SII) model (ANSI standard, 1997), using frequency importance functions for either average speech or CID-22 words. We avoided, however, a quantitative comparison with the SII model since an appropriate frequency-importance function seems to be unavailable for the OLSA sentences. Furthermore, the comparison may be complicated by the fact that the SII is based on normal rather than vocoded speech, even though Faulkner *et al.* (2003) have argued that the relative (but not the absolute) values of the SII weights as a function of frequency should not depend on the degree of frequency selectivity. Interestingly, the SII model predictions are not higher than the performance obtained for vocoded speech. This may be surprising, considering that vocoder processing represents spectral smearing, a manipulation well known to hinder speech understanding (Boothroyd *et al.*, 1996; ter Keurs *et al.*, 1992, 1993; Baer and Moore, 1993, 1994). However, differences in the overall intelligibility of the sentence materials should be taken into account. For unprocessed speech at the 0-dB SNR, NH listeners score near 100% in case of the OLSA test (Wagener *et al.*, 1999c) and around 90% in case of the CID-22 test (Studebaker and Sherbecoe, 1991). Thus, the higher intelligibility of the OLSA material may have compensated for potentially detrimental effects of channel vocoder processing. In summary, the effect of varying the upper boundary of the matched stimulation range on speech understanding was found to be quite comparable between our CI listeners and our NH listeners listening to an acoustic CI simulation, which is in rough agreement to the SII.

### C. Unmatched conditions

**1. Results**—Figure 4 compares conditions vertically in our matrix of conditions seen in Fig. 2. It shows  $\Delta P_c$ , the change in  $P_c$  from the matched conditions, as a function of number of channels in each panel. Therefore, the matched condition, which is indicated by the dotted vertical line in each panel, has exactly zero  $\Delta P_c$  and no error bars. The data are plotted as a difference to better identify relative trends and to combine the high-performance and low-performance CI listeners. Note that on this plot the points to the left of the dotted line represent conditions with less channels for a fixed upper-frequency boundary, hence a compressed frequency-to-place map. Points to the right of the dotted line represent conditions with increased frequency resolution, namely with more channels, hence an expanded map. Arcsine-transformed  $P_c$  values (not  $\Delta P_c$ , the points plotted in Fig. 4) were analyzed with a three-way RM ANOVA (factors  $N$ ,  $M$ , and SNR) including Tukey's post-hoc tests. In Fig. 4, open symbols show no significant difference from the matched condition with the post-hoc tests; closed symbols show a significant difference at the 0.05 level.

Figure 4 shows that the matched condition was often the best, although there were several conditions that were not significantly different from the matched condition. Significant decreases from the matched condition occurred only for conditions that were more than plus or minus two channels relative to the matched condition. Furthermore, conditions where  $M=12$  and 10 had many more significant decreases than  $M=8$  and 6.

For  $M=12$ , there was no significant interaction between  $N$  and SNR for the CI listeners in a two-way RM ANOVA [ $F(9,96)=0.56$ ;  $p=0.83$ ]. However, there was a significant interaction for the NH listeners [ $F(9,80)=3.59$ ;  $p=0.001$ ]. Similarly, for  $M=10$ , there was no significant interaction between  $N$  and SNR for the CI listeners [ $F(9,96)=0.29$ ;  $p=0.98$ ] but there was for the NH listeners [ $F(9,80)=4.53$ ;  $p<0.0001$ ]. It could not be determined whether this interaction for the NH listeners was a ceiling effect or a true asymptote. For  $M=8$ , both CI [ $F(9,96)=0.16$ ;  $p=1$ ] and NH [ $F(9,80)=1.49$ ;  $p=0.17$ ] listeners showed no significant

interaction between  $N$  and SNR. Similarly, for  $M=6$ , there was no significant interaction between  $N$  and SNR for CI [ $F(9,96)=0.43$ ;  $p=0.92$ ] and NH listeners [ $F(9,80)=0.47$ ;  $p=0.83$ ].

Figure 5 shows  $\Delta P_c$  as a function of  $M$ , comparing conditions horizontally on our matrix of conditions seen in Fig. 2. Note that on this plot the points to the left of the dotted line represent conditions with a decreased frequency range compared to the matched condition (spectral expansion) and points to the right of the dotted line represent conditions with an increased frequency range (spectral compression).

Again, significant decreases from the matched condition occurred only for conditions that were more than plus or minus two  $M$  relative to the matched condition. This corresponds to a frequency shift of 0.77 octaves, plus or minus, for the most basal electrode used in presenting the speech information. There was a significant interaction between  $M$  and SNR for  $N=12$  for the NH listeners [ $F(9,80)=3.71$ ;  $p=0.001$ ] but not for the CI listeners [ $F(9,96)=0.56$ ;  $p=0.83$ ]. No other values of  $N$  showed a significant interaction.

To directly compare the baseline condition  $M_{12}N_{12}$  to conditions that have  $M < 12$  and  $N < 12$  electrodes, and are unmatched ( $N \neq M$ ), differences between the baseline condition and all 18 conditions can be seen in Table III. The difference between  $M_{12}N_{12}$  and all other conditions was averaged over all four SNRs. Negative values show an average decrease from the baseline condition. The  $p$  values were determined by Tukey's post-hoc comparisons. For both CI and NH listeners, conditions  $M_{10}N_{12}$ ,  $M_{12}N_{10}$ ,  $M_{10}N_{10}$ , and  $M_8N_{10}$  were not different from  $M_{12}N_{12}$ . An additional condition,  $M_{14}N_{12}$ , was not different for the NH listeners only and will be discussed in Sec. II D. To summarize Table III, even though several of these conditions showed no significant decrease from the nearest matched condition (e.g.,  $M_{10}N_8$  is not significantly different than  $M_{10}N_{10}$  and  $M_8N_8$  for all SNRs), most of them showed a decrease from the baseline  $M_{12}N_{12}$  condition.

Lastly, we observed the interaction of  $N$  and  $M$  for fixed SNRs. There were highly significant effects for each SNR for the single factors  $N$  and  $M$ , and also the interactions between  $N$  and  $M$  (largest  $p=0.001$ ).

**2. Discussion**—There was considerable agreement in the relative trends of the unmatched conditions between the three groups of listeners (high-performance CIs, low-performance CIs, and NHs). The CI listeners showed most of the trends that were seen for the NH listeners. However, there was one notable difference between the NH and CI listeners for fixed  $M=10$  or 12 or fixed  $N=10$  or 12. For these conditions, at the larger SNRs, the NH listeners had relatively larger values of  $\Delta P_c$  than the CI listeners. Either NH listeners were not hindered as much as CI listeners for large spectral shifts and high SNRs, which can be seen in the top two panels of Figs. 4 and 5, or this difference was due to an effect of the ceiling of  $P_c$ .

Our results for both the CI and NH groups showed that varying the number of channels ( $N$ ) by two caused no significant decrease in speech understanding performance compared to the nearest matched condition for a fixed upper-frequency boundary or fixed spectral content. This also held true for applying a gradually increasing spectral shift for a fixed number of channels, either by spectral compression or expansion. However, this was only possible while the upper-frequency boundary was shifted by no more than 0.77 octaves. Note that it was not possible to combine this result with the results from the matched conditions. For example, changing  $M_{10}N_{10}$ , a condition that showed no difference in the matched analysis to  $M_{12}N_{12}$ , to  $M_{10}N_8$  caused significant decreases in performance compared to  $M_{12}N_{12}$ . Therefore, there were only three conditions that were not different from the baseline for the unmatched conditions,  $M_{12}N_{10}$ ,  $M_{10}N_{12}$ , and  $M_8N_{10}$ .



The auditory system seemed to be able to fully accommodate small frequency shifts when the task is to understand speech. This result is similar to other experiments on frequency-warped speech material. For example, Ba Kent and Shannon (2003, 2004) used expansions and compressions where the frequency information at the center electrode was fixed and had a range of shifts from 0.25 octaves to 1 octave for the most apical and basal electrodes. Only the smallest shifts of 0.25 octaves did not significantly decrease sentence recognition scores. Therefore it seems that subjects can tolerate our spectral warping with higher octave shifts of 0.77 octaves because only the high-frequency information was shifted by such a large amount.

Note that there were exceptions for individuals to our spectral warping results. Compared to the matched case, some CI listeners showed significant decreases<sup>4</sup> when a 0.77 octave shift was applied to the spectral information, either by changing the number of electrodes or amount of spectral content via the upper-frequency boundary. However, there were also cases of significantly *increased* speech understanding for deviations from the matched condition. Numerous significant increases were seen for CI18 when the frequency resolution was increased ( $N > M$ ) with an expanded frequency-to-place mapping. In fact, CI18's best speech understanding occurred for  $M_{10}N_{12}$ , not the matched  $M_{12}N_{12}$  condition. On average across the entire population of listeners,  $M_{10}N_{12}$  appeared to be a slightly better map than  $M_{12}N_{12}$  as shown by the small but insignificant gains seen in Table III.

It was shown that there was a significant interaction between  $N$  and  $M$  for fixed SNRs. This is contrary to the results of Fu and Shannon (1999a) that showed no interaction between frequency resolution and frequency allocation. It appears that a constant tonotopic shift along the basilar membrane (in mm) is a different manipulation from the spectral warping that we performed. Therefore, unlike Fu and Shannon's, our spectral manipulation appears not to have orthogonal bases.

Figures 4 and 5 show asymmetric decreases in performance when the spectral range or the number of channels was changed. Decreasing the spectral range or number of channels was more detrimental to understanding speech than increasing the spectral range or number of channels, as shown in the number and magnitude of significant decreases from the matched conditions. Said another way, removing information reduced scores more than adding information despite having similar octave shifts. For example, in Fig. 5, the top panels show that decreasing the upper-frequency boundary and removing spectral content from  $M=12$  to 6 when  $N=12$  decreases  $P_c$  by up to 50%, whereas the bottom panels show that adding spectral content from  $M=6$  to 12 when  $N=6$  decreases  $P_c$  by only 25% or less. This is a promising result for the extended frequency range condition, which adds spectral content to include frequencies above 8.5 kHz.

#### D. Extended frequency range

An additional condition, referred to as the extended frequency range condition ( $M_{14}N_{12}$ ), was included in this experiment. This condition maps the frequencies ranging from  $f_{\text{lower}}=300$  Hz<sup>1</sup> to  $f_{\text{upper}}=16,000$  Hz to the full range of available channels. See Table I for corner frequencies and octave shifts for this condition. It was tested to examine the effects of the inclusion of higher frequencies that contain no important speech cues but are known to be important for sound source localization in the vertical planes for NH listeners. It was hypothesized that listeners would have no degradation in performance due to spectral warping because the most basal electrode was shifted by only  $-0.88$  octaves, close to  $-0.77$

<sup>4</sup>Significant differences were determined using the binomial model proposed by Thorton and Raffin (1978) and were significant at the 0.05 level. The sample size for one condition in a single block was 100, since 20 sentences consisting of five independent words each were presented.

octave shift found to be acceptable from the previous section. The next largest warping tested in this experiment was  $-1.55$  octaves and was found to cause significant decreases in speech understanding.

The results comparing  $M_{14}N_{12}$  to  $M_{12}N_{12}$  can be seen in Table IV. Significance values were determined from a one-way RM ANOVA for each SNR. For the CI listeners, the performance for  $M_{14}N_{12}$  decreased significantly compared to  $M_{12}N_{12}$  in quiet (mean decrease of 11.3%) and for the +10-dB SNR (mean decrease of 15.7%). There were decreases of 15.0% and 20.4% for the +5- and 0-dB SNR conditions, respectively, but they were not significant. The reason that these decreases were not significant was the large variance of the average performance for the group of CI listeners in the noisier conditions, which can be seen in the standard deviations in Table IV. For the NH listeners, the performance for  $M_{14}N_{12}$  was not significantly different from  $M_{12}N_{12}$  for all four SNRs.

The difference between the CI and NH listeners could be due to the ceiling effect that may be affecting the NH listeners, which has been hinted at in the data of the matched and unmatched sections. However, the CI listeners did not show this possible ceiling effect. Therefore, if the speech material were more difficult, the NH listeners might have also shown significant decreases for the extended frequency range condition compared to the baseline. Another reason for this discrepancy between the groups of listeners could be that CI listeners are less resistant to spectral warping than NH listeners as seen in Figs. 4 and 5. All the  $p$  values in Table IV were on the verge of significance, implying that a spectral shift of  $-0.88$  octaves was just enough to cause a significant difference. It may be the case that a slightly smaller shift closer to  $-0.77$  octaves would not cause a substantial decrease in this experiment. Thus it may be possible to obtain a condition without a significant decrease in performance if frequencies up to 14.5 kHz, which corresponds to a  $-0.74$  octave shift for the most basal electrode, were used.

### III. EXPERIMENT 2

As mentioned in the Introduction, the importance of training with different frequency-to-place maps is well known (Rosen *et al.*, 1999; Fu *et al.*, 2002; Fu *et al.*, 2005). For example, Ba kent and Shannon (2006) showed no improvement in speech understanding when maps were altered to reassign information around dead regions in NH simulations. However, in a subsequent study, Smith and Faulkner (2006) showed significant learning with extended training, reducing the problem of holes in hearing and coming to the opposite conclusion as Ba kent and Shannon. Furthermore, the results of experiment 1 could have been confounded by the possibility that subjects require more perceptual learning to obtain an asymptote in performance for conditions involving spectral shift than for conditions involving no shift. For example, the matched condition  $M_6N_6$  involves no spectral warping from the clinical setting, whereas the unmatched conditions  $M_6N_{12}$  and  $M_{12}N_6$  involve large shifts of spectral information for the most basal electrodes. This experiment studied the time course of perceptual learning with three selected conditions from experiment 1, one without spectral warping and two with warping. While previous studies have shown a stronger learning effect for shifted-speech stimuli compared to unshifted stimuli, these results may not be applicable to the unmatched conditions of experiment 1. A major difference is that these studies applied a constant shift across the signal spectrum, whereas we apply a shift that gradually increases with frequency starting at approximately zero shift at the lower frequency end. Thus, there is less shift in the lower spectral region, which is considered to be most important for speech perception. The main goal of the present experiment is to estimate to what extent the relative differences between the results obtained for the unmatched and the matched conditions may change if the subjects are provided with feedback training over several sessions.

## A. Methods

Three of the normal-hearing listeners (NH2, NH5, and NH10) that participated in experiment 1 performed this experiment. Three mapping conditions were tested, namely  $M_6N_6$ ,  $M_6N_{12}$ , and  $M_{12}N_6$ —conditions representative of the matched, expanded, and compressed conditions, respectively (see Table I). The condition without spectral warping was included to separate the adaptation to vocoded speech in general from the adaptation to the spectral warping. While all three conditions may involve vocoded-speech adaptation, only the unmatched conditions involve spectral-warping adaptation.

We trained and tested two different conditions in a single test. We felt this was justified by Dorman and Dahlstrom (2004) who showed listeners probably using two speech maps simultaneously and Hofman *et al.* (1998) who showed that at least two localization maps could be mutually utilized by subjects. First, the  $M_6N_6$  condition was tested with the  $M_6N_{12}$  condition. Each subject completed a total of eight experimental sessions. After data were taken for these two conditions, the  $M_6N_6$  condition was tested with the  $M_{12}N_6$  condition. Within one session, feedback training and subsequent testing was performed for both signal conditions. The order of matched and unmatched blocks was randomly chosen for each session. Each block consisted of ten sentences in quiet for the initial listener adaptation, followed by 80 sentences (the four SNRs used in experiment 1 presented in random order). The sentences were different from trial to trial, but, in contrast to experiment 1, they were not different from listener to listener. The reason for this was to determine if certain words affect the results of the sentence tests. In the feedback training runs, each sentence was displayed on a computer screen, followed by a gap of 500 ms with subsequent acoustic presentation of the sentence. The subjects initiated each successive trial by pressing a button. All other aspects of the experiment were the same as in experiment 1.

## B. Results

Figure 6 shows the results of experiment 2 averaged over the three listeners. Each plot depicts  $P_c$  as a function of session number. The baseline session, labeled as 0, represents the acute data from these three listeners taken in experiment 1. Only eight sessions for the matched  $M_6N_6$  condition (taken from the first test group of  $M_6N_6$  and  $M_6N_{12}$ ) are shown because learning had saturated after the first eight sessions, which is supported below. Qualitatively, it seems that there was a training effect and more learning seems possible for  $M_6N_{12}$  (left panel), the learning seems to have saturated for  $M_6N_6$  (middle panel), and that there was no learning for  $M_{12}N_6$  (right panel). To determine the significance of the learning effect, we compared the acute measurements from experiment 1 and the average of the last three measurements in a two-way RM ANOVA [factors SNR and session (four)]. Two conditions showed significant learning:  $M_6N_6$  ( $p < 0.0001$ ) and  $M_6N_{12}$  ( $p < 0.0001$ ). However, the  $M_{12}N_6$  condition did not show significant learning ( $p = 0.098$ ). This agrees with the interaction analysis in a three-way RM ANOVA [factors SNR, session (all nine), and condition]. The interaction between conditions  $M_6N_6$  and  $M_6N_{12}$ , the conditions with learning, and session number, was insignificant ( $p = 0.069$ ). However, the interaction between conditions  $M_6N_6$  and  $M_{12}N_6$  and session number was significant ( $p = 0.011$ ). The  $M_6N_6$  condition was tested twice, eight sessions with  $M_6N_{12}$  and eight sessions with  $M_{12}N_6$ . The last three sessions of each were compared and the difference between them was insignificant ( $p = 0.51$ ). Therefore, learning had truly saturated for  $M_6N_6$  after the first eight sessions.

For the two conditions with learning, the session number where significant learning occurs was determined by Helmert contrasts. For the matched condition, significant learning was seen up to the second session (session 1 vs later:  $p = 0.002$ , session 2 vs later:  $p = 0.36$ ). For the expanded condition, significant learning was seen up to the third session (session 2 vs later:  $p = 0.007$ , session 3 vs later:  $p = 0.17$ ).

Comparisons between matched and unmatched conditions are most important for this study. First, comparisons between the matched  $M_6N_6$  condition and the unmatched conditions were done. For the acute measurements, there was a significant decrease between the matched condition and each of the compressed and expanded conditions (see, Figs. 4 and 5 or Table III). For the three measurements at the end of the training, there continued to be a significant decrease between the matched and expanded conditions ( $p < 0.0001$ ) and between the matched and compressed conditions ( $p < 0.0001$ ). This is to be expected between  $M_6N_6$  and  $M_{12}N_6$  because the compressed condition showed no significant learning. However, it is noteworthy that the  $M_6N_{12}$  condition, although having showed learning, did not improve enough to remove the difference from the  $M_6N_6$  condition.

Second, comparisons between the baseline condition,  $M_{12}N_{12}$ , and the three conditions tested in experiment 2 were done. Note that all listeners were trained on  $M_{12}N_{12}$  until stable performance was achieved. For the three NH listeners used in this experiment, there was always a significant difference between  $M_{12}N_{12}$  and the tested conditions in the acute testing. After the training, all three conditions still showed significant decreases from  $M_{12}N_{12}$  ( $p < 0.0001$  for all three comparisons).

As a final note, one difference between experiments 2 and 1 is that experiment 2 used the same sentences for all three listeners. It was found that listeners'  $P_c$  values were not significantly correlated over the session number. Therefore, certain words or configurations of words in the OLSA sentences did not significantly affect the outcome of the test.

### C. Discussion

After two or three training sessions with feedback, the listeners obtained scores that were only marginally lower (on average across the listeners) than those after eight training sessions with feedback. Qualitatively, the improvement for the  $M_6N_6$  condition appears to have saturated. The  $M_{12}N_6$  condition showed no significant improvement. However, the  $M_6N_{12}$  condition may still have been improving. We cannot say whether there would be further improvement over months of training as suggested by Fu *et al.* (2002). Nevertheless, we can say that feedback training seems to be very helpful before testing different frequency-to-place mappings (matched and unmatched), even if it is only a couple of sessions. The data of Fu *et al.* (2002) showed that the biggest gains were within the first few days of training, similar to our data. A major difference between these two studies is that the CI listeners in Fu *et al.* used a shifted-spectral map all day long, whereas our NH listeners heard warped sentences for an hour a day. Rosen *et al.* (1999) and Fu *et al.* (2005) also showed the largest improvements within the first few testing sessions for spectrally-shifted material.

There was no evidence that listeners learn differently between the matched  $M_6N_6$  condition and the unmatched  $M_6N_{12}$  condition. This is contrary to the studies of Rosen *et al.* (1999) and Fu *et al.* (2005) who found that there was less learning for the matched condition than for the unmatched condition. A major difference between those studies and our study is the type of shift. We held the lower-frequency boundary constant for each listener, which involves less shift at lower frequencies that are more important for speech understanding. Interestingly, the compressed condition,  $M_{12}N_6$ , was resistant to learning for our listeners. The reason for the difference between the compressed and expanded conditions might be that there is improved resolution of the formants for the expanded condition, which can be adapted to even if the tonotopic place of formant presentation has been altered. However, reducing the resolution of the formants cannot be adapted to when the speech is represented in six channels. Also note that in Table I the octave shifts for the lowest channels are slightly larger for the compressed condition,  $M_{12}N_6$ , which may have also contributed to the difference in learning.

Even though there was a difference in learning between the compressed condition and the matched and expanded conditions, both of the latter conditions were still significantly lower than the baseline condition. Hence, the results from experiment 2 give insight to adaptation of speech understanding under different matched, expanded, and compressed frequency-to-place maps. Nonetheless, we need to be cautious about these results. First, only three NH listeners were used, no CI listeners. Next, eight sessions was possibly not a sufficient training period for the expanded  $M_6N_{12}$  condition; longer testing periods should be attempted in the future. Last, the unmatched conditions chosen for this experiment had the most extreme spectral warpings of those tested in experiment 1. It may be that milder warpings could show changes in the relative differences to the baseline condition, especially conditions with sufficient formant resolution. However, since our goal is to find conditions that are not different from the baseline, this would actually be a positive finding for our study on spectral warping and using different frequency-to-place mappings to implement new CI strategies.

## IV. GENERAL DISCUSSION

### A. Effects of spectral warpings

The purpose of this study was to test a wide range of spectral mappings and find the effect of varying the upper boundary of the frequency and stimulation range and spectral warpings on speech understanding for CI users, with the hope of gaining insight into the possibility of implementing new processing strategies for CIs. Specifically, we were trying to determine if there is extra speech information currently presented to CI users. The type of spectral warping that we implemented was unique because it kept the lower boundary of the frequency and stimulation range fixed, and introduced an increasing amount of shift towards higher channels. This allowed for testing of both matched frequency-to-place conditions, where speech material was only low-pass filtered, and unmatched conditions.

For the matched conditions, where the spectral content was matched to the place of presentation, it was found that speech understanding in CI listeners saturates at eight to ten electrodes, which does not present frequency information above 2.8 to 4.9 kHz, respectively. There was no significant difference between the twelve and eight or ten channel conditions, depending on the SNR. Care should be taken extrapolating these results to different speech material or different maskers like single- or multi-talker interference. As always, because of the highly individualistic nature of CI users, care should also be taken to derive general conclusions about the entire population.

Combining our results with those obtained by Garnham *et al.* (2002) and by Ba kent and Shannon (2005), who used the same type of electrode array and thus the same spacing of electrodes, reveals an interesting observation. Increasing the number of matched channels beyond eight does not improve speech understanding substantially, irrespective if the channels are distributed across a constant stimulation range (Garnham *et al.* 2002), if the lower-frequency boundary of the matched stimulation range is varied (Ba kent and Shannon, 2005), or if the higher-frequency boundary of the matched stimulation range is varied (present study). In the study by Garnham *et al.*, this limit appears to be related to the restriction of the number of independent channels as a result of channel interactions (e.g., Shannon, 1983). However, for the Ba kent and Shannon study and our experiment, the saturation at eight to ten channels can be understood by a different mechanism. The channel interactions are held constant in both studies. The limit on the number of channels observed in these studies can be explained by the frequency range important for speech perception. This corresponds to approximately 300–5000 Hz, below and above which the SII shows minimal importance of the speech information for NH listeners. In fact, this also explains why there was no difference between the CI and NH listeners for the saturation of speech

understanding in quiet and in noise. NH listeners can benefit more than CI listeners from a larger number of channels if they are distributed within the frequency range important to speech perception, particularly at low SNRs (e.g., Dorman *et al.*, 1998).

It was found that it is possible to shift the spectral mapping slightly without decreasing speech understanding. For the 12-electrode Med-El implant, it is possible to change the amount of frequency information by 0.77 octaves for the most basal electrode and have no decrease in speech understanding. This can be done either by varying the upper-frequency boundary and spectral content for a fixed number of electrodes or by changing the number of channels by two for a fixed amount of spectral content. As was mentioned before, only eight to ten channels with frequency content matched to tonotopic place (as assumed by the typical clinical settings) are necessary for no significant decrease in speech understanding. However, it is not possible to then apply a shift of 0.77 octaves to eight matched channels without causing significant decreases from the matched twelve channel condition. In summary, the conditions where speech understanding is not significantly different from the  $M_{12}N_{12}$  case are:  $M_{10}N_{12}$ ,  $M_{12}N_{10}$ ,  $M_{10}N_{10}$ , and  $M_8N_{10}$ .

A second experiment was performed to test the adaptation of listeners to matched, expanded, and compressed frequency-to-place maps using feedback training. It was found that listeners' speech understanding improved for the matched and expanded maps, but not for the compressed map. The reason for this may be that the compressed map had insufficient spectral resolution for adequate formant recognition. The expanded map could be adapted to apparently because of the formants could be resolved. This explanation may also be supported by the small but insignificant overall gains seen in Table III for the  $M_{10}N_{12}$  condition that also has improved resolution compared to the baseline condition. Similarly, McKay and Henshall (2002) found that after two weeks experience, CI listeners could fully adapt to a "mildly" expanded map, where nine electrodes were assigned to frequencies below 2600 Hz, compared to the normal five electrodes.

A recent study by Ba kent and Shannon (2007) studied the effect of combining spectral shift and frequency expansion and compression using noise-vocoded speech with acute measurements. They found a compensatory effect for shifted and compressed maps, hence a relative gain in speech understanding, which may seem contradictory to the poor performance of the compressed condition in experiment 2. The difference between these two studies is that a shift and compression is not the same as a spectral warping where the lower-frequency boundary is fixed. Our lower-frequency boundary always contained frequencies as low as 300 Hz. The stimuli in Ba kent and Shannon's study had frequency content as low as 184 Hz for a compressed map, 513 Hz for a matched map, and 1170 for an expanded map. If the low-frequency content was kept fixed and training occurred, then we believe that a compensatory effect may be seen for the expanded maps also.

## B. In the context of sound localization

While there are great advances in the area of speech understanding with CIs, there are other areas that are still severely lacking. For example, an area that needs attention is sound localization. For localization in the vertical plane (front-back and up-down), spectral peaks and notches are needed. These spectral cues are normally introduced by reflections from the pinna, head, and shoulders. The findings of this study are a preliminary step in realizing a CI spatialization strategy, focusing mainly on the possibility of including detectable spectral cues for vertical plane localization. Introducing these cues may have a detrimental effect on speech understanding, especially for the CI listeners with their restricted number of channels. A primary goal in developing a CI spatialization strategy is to have no or very little loss of speech understanding, the main purpose for implantation of a prosthetic hearing device. Any localization strategy that hinders a CI's major reason for existence is not

particularly useful. Besides improving the localization abilities of CI users, such a spatialization strategy may someday help speech understanding in noisy situations, an extremely difficult situation for CI users.

The fact that full adaptation to compressed frequency-to-place maps may not be possible is an important fact to consider when designing a CI spatialization strategy. The most important spectral cues used for vertical plane localization occur between 4 and 16 kHz (Hebrank and Wright, 1974; Langendijk and Bronkhorst, 2002). Current processing strategies that use spectral information up to 8.5 or 10 kHz (depending on processor type) would already be able to implement peaks and notches in the lower half of the important spectral region. However, it was seen that CI listeners had significantly lower speech understanding scores, when acutely measured, when the upper-frequency boundary was increased to 16 kHz. A strict interpretation of this result combined with the results from the learning experiment suggests that the mapping of spectral peaks and notches from a straightforward frequency compression may not be the solution to gain access to the localization cues above 10 kHz while preserving adequate speech understanding. This of course is assuming that the lack of learning result for the severely compressed maps applies to the slightly compressed maps, which is probably not the case. Unfortunately, this study did not determine how much spectral compression retains sufficient spectral resolution of the formants to yield sufficient long-term learning of a compressed map. Whether or not a mildly compressed map can be used, a likely solution to implement a CI localization strategy would be to match frequency-to-place up to about eight channels and then apply a novel mapping of higher frequency regions to the remaining channels, depending on the importance of spectral features in these regions. Afterwards, results from this study can be combined with data on the sensitivity of CI listeners to spectral peaks and notches to determine if spectral sound localization cues can be used by CI listeners. Because it is possible for listeners to adapt to new spectral localization cues (Hofman *et al.*, 1998), we believe the auditory system has enough plasticity to make use of a CI localization strategy.

## Acknowledgments

We would like to thank M. Mihocic for running experiments, our listeners who performed a total of 250 h in these experiments, and the Med-El Corp. for providing the equipment for direct electrical stimulation. We would also like to thank the associate editor, Andrew Oxenham, and two anonymous reviewers for improving the quality of this manuscript. This study was funded by the Austrian Science Fund (Project No. P18401-B15) and the Austrian Academy of Sciences.

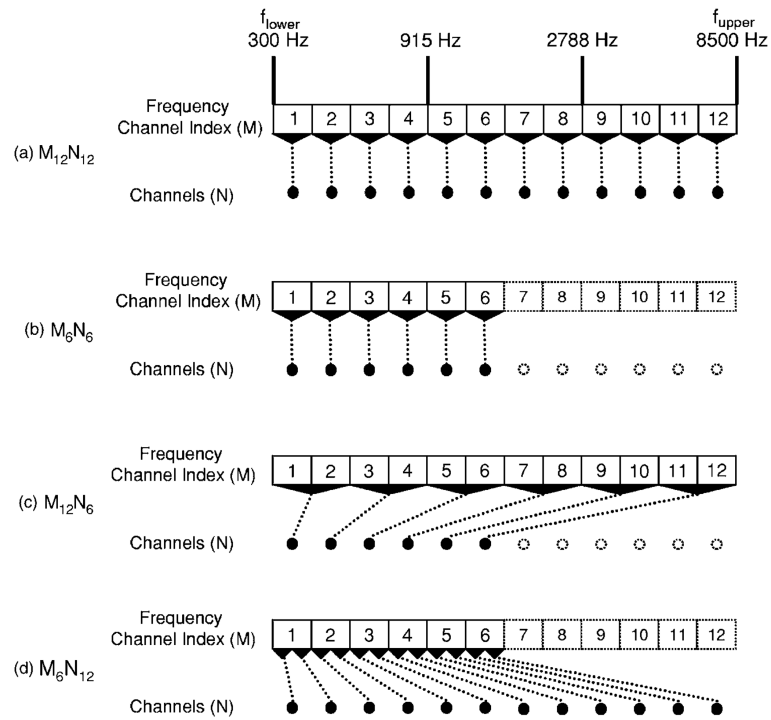
## References

- ANSI. American National Standard Methods for the Calculation of the Speech Intelligibility Index. American National Standards Institute; New York: 1997. ANSI S3.5-1997
- Baer T, Moore BCJ. Effects of spectral smearing on the intelligibility of sentences in noise. *J. Acoust. Soc. Am.* 1993; 94:1229–1241.
- Baer T, Moore BCJ. Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech. *J. Acoust. Soc. Am.* 1994; 95:2277–2280. [PubMed: 8201124]
- Ba kent D, Shannon RV. Speech recognition under conditions of frequency-to-place compression and expansion. *J. Acoust. Soc. Am.* 2003; 113:2064–2076. [PubMed: 12703717]
- Ba kent D, Shannon RV. Frequency-to-place compression and expansion in cochlear-implant listeners. *J. Acoust. Soc. Am.* 2004; 116:3130–3140. [PubMed: 15603158]
- Ba kent D, Shannon RV. Interactions between cochlear implant electrode insertion depth and frequency-to-place mapping. *J. Acoust. Soc. Am.* 2005; 117:1405–1416. [PubMed: 15807028]
- Ba kent D, Shannon RV. Frequency transposition around dead regions simulated with a noiseband vocoder. *J. Acoust. Soc. Am.* 2006; 119:1156–1163. [PubMed: 16521776]

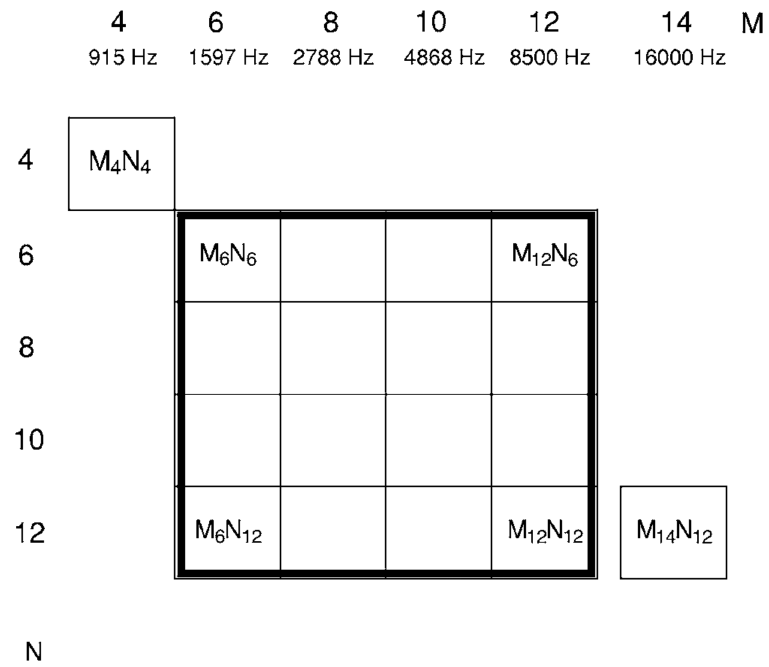
- Ba kent D, Shannon RV. Combined effects of frequency compression-expansion and shift on speech recognition. *Ear Hear.* 2007; 28:277–289. [PubMed: 17485977]
- Boothroyd A, Mulheam B, Gong J, Ostroff J. Effects of spectral smearing on phoneme and word recognition. *J. Acoust. Soc. Am.* 1996; 100:1807–1818. [PubMed: 8817914]
- Chambers, J.; Hastie, TJ. *Statistical Models.* Chapman and Hall; London: 1993.
- Dorman MF, Dahlstrom L. Speech understanding by cochlear-implant patients with different left- and right-ear electrode arrays. *Ear Hear.* 2004; 25:191–194. [PubMed: 15064664]
- Dorman MF, Loizou PC, Rainey D. Simulating the effect of cochlear-implant electrode depth on speech understanding. *J. Acoust. Soc. Am.* 1997; 102:2993–2996. [PubMed: 9373986]
- Dorman MF, Loizou PC, Fitzke J, Tu Z. 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.* 1998; 104:3583–3585. [PubMed: 9857516]
- Faulkner A, Rosen S, Stanton D. Simulations of tonotopically mapped speech processors for cochlear implant electrodes varying in insertion depth. *J. Acoust. Soc. Am.* 2003; 113:1073–1080. [PubMed: 12597200]
- Friesen LM, Shannon RV, Ba kent D, Wang X. Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *J. Acoust. Soc. Am.* 2001; 110:1150–1163. [PubMed: 11519582]
- Fu Q-J, Shannon RV. Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing. *J. Acoust. Soc. Am.* 1999a; 105:1889–1900. [PubMed: 10089611]
- Fu Q-J, Shannon RV. Effects of electrode location and spacing on phoneme recognition with the Nucleus-22 cochlear implant. *Ear Hear.* 1999b; 20:321–331. [PubMed: 10466568]
- Fu Q-J, Shannon RV. Effects of electrode configuration and frequency allocation on vowel recognition with the Nucleus-22 cochlear implant. *Ear Hear.* 1999c; 20:332–344. [PubMed: 10466569]
- Fu Q-J, Shannon RV. Effect of acoustic dynamic range on phoneme recognition in quiet and noise by cochlear-implant listeners. *J. Acoust. Soc. Am.* 1999d; 106:L65–L70. [PubMed: 10615712]
- Fu Q-J, Shannon RV, Galvin JJ III. Perceptual learning following changes in the frequency-to-electrode assignment with the Nucleus-22 cochlear implant. *J. Acoust. Soc. Am.* 2002; 112:1664–1674. [PubMed: 12398471]
- Fu Q-J, Nogaki G, Galvin JJ III. Auditory training with spectrally shifted speech: Implications for cochlear implant patient auditory rehabilitation. *J. Assoc. Res. Otolaryngol.* 2005; 6:180–189. [PubMed: 15952053]
- Garnham C, O'Driscoll M, Ramsden R, Saeed S. Speech understanding in noise with a Med-El COMBI 40+ cochlear implant using reduced channel sets. *Ear Hear.* 2002; 23:540–552. [PubMed: 12476091]
- Greenwood DD. A cochlear frequency-position function for several species—29 years later. *J. Acoust. Soc. Am.* 1990; 87:2592–2605. [PubMed: 2373794]
- Gstoettner W, Franz P, Hamzavi J, Plenk H, Baumgartner W, Czerny C. Intracochlear position of cochlear implant electrodes. *Acta Oto-Laryngol.* 1999; 119:229–233.
- Hebrank J, Wright D. Spectral cues used in the localization of sound sources on the median plane. *J. Acoust. Soc. Am.* 1974; 56:1829–1834. [PubMed: 4443482]
- Hofman PM, Van Riswick JGA, Van Opstal AJ. Re-learning sound localization with new ears. *Nat. Neurosci.* 1998; 1:417–421. [PubMed: 10196533]
- Ketten DR, Vannier MW, Skinner MW, Gates GA, Wang G, Neely JG. *In vivo* measures of cochlear length and insertion depth of nucleus cochlear implant electrode arrays. *Ann. Otol. Rhinol. Laryngol.* 1998; 107:1–16. [PubMed: 9439380]
- Langendijk EHA, Bronkhorst AW. Contribution of spectral cues to human sound localization. *J. Acoust. Soc. Am.* 2002; 112:1583–1596. [PubMed: 12398464]
- McKay CM, Henshall KR. Frequency-to-electrode allocation and speech perception with cochlear implants. *J. Acoust. Soc. Am.* 2002; 111:1036–1044. [PubMed: 11863160]
- Rosen S, Faulkner A, Wilkinson L. Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants. *J. Acoust. Soc. Am.* 1999; 106:3629–3636. [PubMed: 10615701]



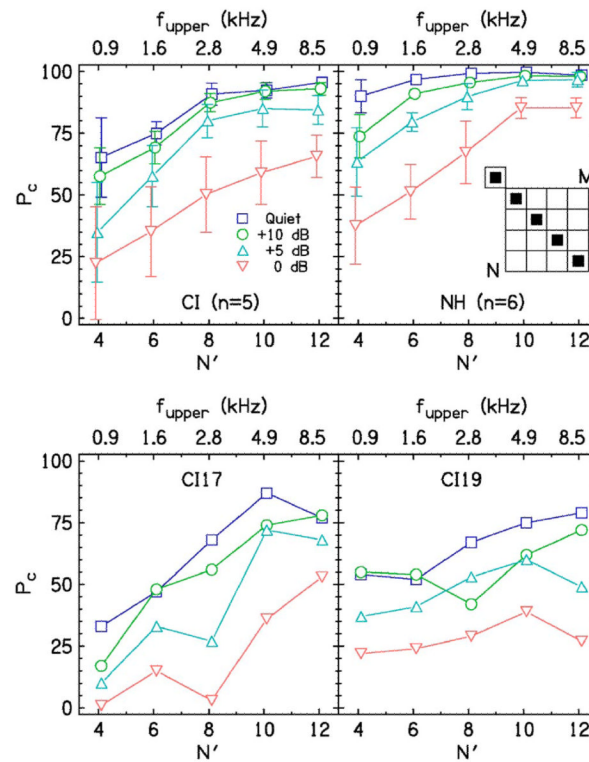
- Shannon RV. Multichannel electrical stimulation of the auditory nerve in man. II. Channel interaction. *Hear. Res.* 1983; 12:1–16. [PubMed: 6689326]
- Shannon RV, Zeng F-G, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. *Science.* 1995; 270:303–304. [PubMed: 7569981]
- Skinner MW, Fourakis MS, Holden TA. Effect of frequency boundary assignment on speech recognition with the SPEAK speech-coding strategy. *Ann. Otol. Rhinol. Laryngol. Suppl.* 1995; 166:307–311. [PubMed: 7668683]
- Smith M, Faulkner A. Perceptual adaptation by normally hearing listeners to a simulated ‘hole’ in hearing. *J. Acoust. Soc. Am.* 2006; 120:4019–4030. [PubMed: 17225428]
- Sridhar D, Stakhovskaya O, Leake PA. A frequency-position function for the human cochlear spiral ganglion. *Audiol. NeuroOtol.* 2006; 11(Suppl 1):16–20. [PubMed: 17063006]
- Studebaker GA. A ‘rationalized’ arcsine transform. *J. Speech Hear. Res.* 1985; 28:455–462. [PubMed: 4046587]
- Studebaker GA, Sherbecoe RL. Frequency-importance and transfer functions for recorded CID W-22 word lists. *J. Speech Hear. Res.* 1991; 34:427–438. [PubMed: 2046367]
- Thornton AR, Raffin MJ. Speech-discrimination scores modeled as a binomial variable. *J. Speech Hear. Res.* 1978; 21:507–518. [PubMed: 713519]
- ter Keurs M, Festen JM, Plomp R. Effect of spectral envelope smearing on speech reception. I. *J. Acoust. Soc. Am.* 1992; 91:2872–2880. [PubMed: 1629480]
- ter Keurs M, Festen JM, Plomp R. Effect of spectral envelope smearing on speech reception. II. *J. Acoust. Soc. Am.* 1993; 93:1547–1552. [PubMed: 8473608]
- Wagener K, Brand T, Kühnel V, Kollmeier B. “Development and evaluation of a German sentence test I: Design of the Oldenburg sentence test,” (in German). *Zeitschrift für Audiologie/Audiological Acoustics.* 1999a; 38:4–14.
- Wagener K, Brand T, Kühnel V, Kollmeier B. “Development and evaluation of a German sentence test II: Optimization of the Oldenburg sentence test,” (in German). *Zeitschrift für Audiologie/Audiological Acoustics.* 1999b; 38:44–56.
- Wagener K, Brand T, Kühnel V, Kollmeier B. “Development and evaluation of a German sentence test III: Evaluation of the Oldenburg sentence test,” (in German). *Zeitschrift für Audiologie/Audiological Acoustics.* 1999c; 38:86–95.
- Wilson BS, Finley CC, Lawson DT, Wolford RD, Eddington DK, Rabinowitz WM. New levels of speech recognition with cochlear implants. *Nature (London).* 1991; 352:236–238. [PubMed: 1857418]
- Zierhofer CM, Hochmair IJ, Hochmair ES. The advanced Combi 40+ cochlear implant. *Am. J. Otol.* 1997; 18(Suppl):S37–38. [PubMed: 9391589]

**FIG. 1.**

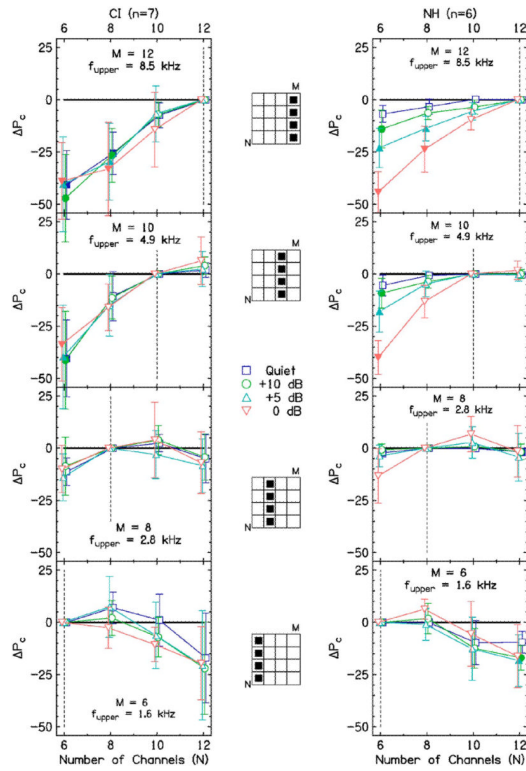
Different frequency-to-tonotopic place mappings used in this experiment. Panel (a) shows the baseline condition  $M_{12}N_{12}$ , where frequencies 300–8500 Hz of the speech signal are mapped to 12 channels, essentially the same as the normal clinical mapping. Panel (b) shows the  $M_6N_6$  condition, a matched condition where the spectral content is low-pass filtered, but unwarped from the normal clinical mapping. Panel (c) shows a compression of full spectral content to a limited set of 6 channels. Panel (d) shows an expansion of limited spectral content to a full set of 12 channels.

**FIG. 2.**

An illustrative matrix of conditions tested in experiment 1. In total there are 18 different conditions tested, which vary the upper-frequency boundary index ( $M$ ) or the number of channels ( $N$ ).

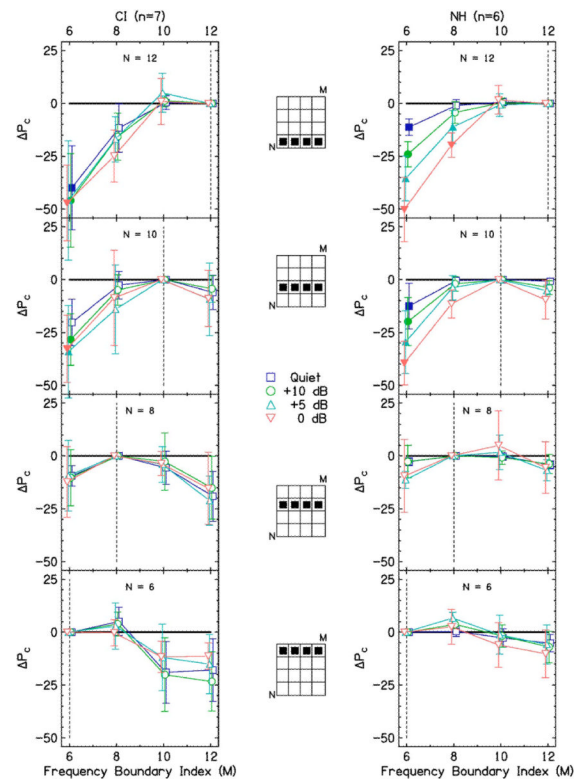


**FIG. 3.** (Color online) Percent correct scores as a function of matched channels  $N'$  ( $N = M$ ) for five high-performance CI listeners (upper-left panel), six NH listeners (upper-right panel), and two low-performance CI listeners (bottom panels). The upper panels have errors bars that are two standard deviations in length.

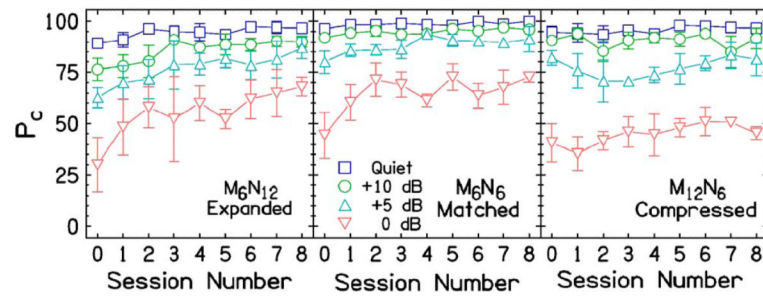


**FIG. 4.**

(Color online) Percent correct scores relative to the matched conditions (shown by the vertical dotted line) vs the number of channels ( $N$ ) for a fixed upper-frequency boundary and spectral content ( $M$ ). Open symbols show no significant difference from the matched case. Closed symbols show a significant difference from the matched case ( $p < 0.05$ ). The matched case was often the best case. Significant decreases were seen for changes of more than two channels from the matched cases.

**FIG. 5.**

(Color online) Percent correct scores relative to the matched condition (shown by the vertical dotted line) vs the amount of spectral content ( $M$ ) for a fixed channel number ( $N$ ). Open symbols show no significant difference from the matched case. Closed symbols show a significant difference from the matched case. The matched case was often the best case. Significant decreases were seen for changes of more than two frequency boundary indices ( $M$ ) from the matched cases.



**FIG. 6.** (Color online) Results from experiment 2 showing percent correct vs session number averaged over three NH listeners. Session 0 contains the baseline data from experiment 1. Sessions 1–8 are from experiment 2.

**TABLE I**  
Center frequency, bandwidth, and octave shift for each band for each combination of  $M$  and  $N$ .

M	N	Channel	1	2	3	4	5	6	7	8	9	10	11	12
12	12	Center (Hz)	348	460	608	804	1062	1403	1854	2449	3236	4276	5651	7467
		BW (Hz)	96	128	168	223	294	388	513	678	896	1184	1565	2067
		Shift (8ve)	...	...	...	...	...	...	...	...	...	...	...	...
10	12	Center (Hz)	339	428	540	681	859	1084	1367	1724	2175	2743	3460	4364
		BW (Hz)	78	99	125	158	198	251	315	399	503	634	799	1009
		Shift (8ve)	0.04	0.11	0.17	0.24	0.31	0.37	0.44	0.51	0.57	0.64	0.71	0.77
8	12	Center (Hz)	331	398	480	578	696	838	1008	1214	1462	1760	2120	2552
		BW (Hz)	61	74	89	107	129	155	186	225	271	326	393	472
		Shift (8ve)	0.07	0.21	0.34	0.48	0.61	0.74	0.88	1.01	1.15	1.28	1.41	1.55
6	12	Center (Hz)	323	371	426	490	563	647	744	856	983	1130	1300	1494
		BW (Hz)	45	51	60	68	78	90	104	119	136	158	181	207
		Shift (8ve)	0.11	0.31	0.51	0.71	0.92	1.12	1.32	1.52	1.72	1.92	2.12	2.32
12	10	Center (Hz)	360	503	702	981	1370	1914	2674	3736	5220	7292		
		BW (Hz)	119	167	232	325	454	634	886	1238	1729	2416		
		Shift (8ve)	-0.05	-0.13	-0.21	-0.29	-0.37	-0.45	-0.53	-0.61	-0.69	-0.77		
10	10	Center (Hz)	348	460	608	804	1062	1403	1854	2449	3236	4276		
		BW (Hz)	96	128	168	223	294	388	513	678	896	1184		
		Shift (8ve)	...	...	...	...	...	...	...	...	...	...		
8	10	Center (Hz)	338	422	528	659	824	1029	1286	1607	2008	2510		
		BW (Hz)	75	94	117	146	183	228	285	357	446	557		
		Shift (8ve)	0.04	0.12	0.20	0.29	0.37	0.45	0.53	0.61	0.69	0.77		
6	10	Center (Hz)	328	387	457	541	639	755	893	1055	1247	1474		
		BW (Hz)	55	64	76	91	106	126	149	176	208	246		
		Shift (8ve)	0.09	0.25	0.41	0.57	0.73	0.89	1.05	1.21	1.38	1.54		
12	8	Center (Hz)	378	574	872	1324	2012	3055	4640	7048				
		BW (Hz)	156	236	359	546	829	1258	1912	2904				
		Shift (8ve)	-0.12	-0.32	-0.52	-0.72	-0.92	-1.12	-1.32	-1.53				
10	8	Center (Hz)	363	514	728	1031	1461	2069	2931	4152				
		BW (Hz)	125	177	251	356	503	714	1010	1432				



M	N	Channel	1	2	3	4	5	6	7	8	9	10	11	12
		Shift (8ve)	-0.06	-0.16	-0.26	-0.36	-0.46	-0.56	-0.66	-0.76				
8	8	Center (Hz)	348	460	608	804	1062	1403	1854	2449				
		BW (Hz)	96	128	168	223	294	388	513	678				
		Shift (8ve)	...	...	...	...	...	...	...	...				
6	8	Center (Hz)	335	413	509	627	773	952	1174	1447				
		BW (Hz)	70	86	106	130	161	198	245	301				
		Shift (8ve)	0.05	0.16	0.26	0.36	0.46	0.56	0.66	0.76				
12	6	Center (Hz)	412	720	1256	2193	3828	6684						
		BW (Hz)	224	391	682	1191	2080	3632						
		Shift (8ve)	-0.24	-0.65	-1.05	-1.45	-1.85	-2.25						
10	6	Center (Hz)	389	619	985	1566	2492	3964						
		BW (Hz)	177	283	449	714	1137	1808						
		Shift (8ve)	-0.16	-0.43	-0.70	-0.96	-1.23	-1.50						
8	6	Center (Hz)	368	533	773	1121	1625	2356						
		BW (Hz)	135	196	284	411	597	865						
		Shift (8ve)	-0.08	-0.21	-0.35	-0.48	-0.61	-0.75						
6	6	Center (Hz)	348	460	608	804	1062	1403						
		BW (Hz)	96	128	168	223	294	388						
		Shift (8ve)	...	...	...	...	...	...						
4	4	Center (Hz)	348	460	608	804								
		BW (Hz)	96	128	168	223								
		Shift (8ve)	...	...	...	...								
14	12	Center (Hz)	359	500	697	970	1351	1882	2622	3652	5086	7084	9867	13744
		BW (Hz)	118	164	229	318	444	618	861	1199	1670	2326	3240	4513
		Shift (8ve)	-0.04	-0.12	-0.20	-0.27	-0.35	-0.42	-0.50	-0.58	-0.65	-0.73	-0.80	-0.88

TABLE II

Bibliographic data of the CI listeners.

Subject	Etiology	Age (yr)	Duration of deafness (yr)	Implant use (yr)	Hearing aid use (yr)	Implant type	Insertion depth (mm)
CI13	Progressive	58	1	8	0.5	C40+	?
CI14	Progressive	67	11	6	15	C40+	30.0
CI15	Unknown	73	30	6	30	C40+	30.0
CI16	Progressive	54	8	1	30	Pulsar	21.0
CI17	Morbus Meniere	59	1	1.5	14	Pulsar	32.0
CI18	Progressive	49	11	2	33	Pulsar	30.0
CI19	Progressive	62	14	5	22	C40+	33.0

TABLE III

Percent correct differences between the baseline  $M_{12}N_{12}$  condition and the other conditions averaged over all four SNRs. The  $p$  values were found from Tukey's post-hoc tests from a three-way RM ANOVA. Values not different from the baseline ( $p > 0.05$ ) are marked in bold.

Condition	CI		NH	
	Difference (%)	$p$	Difference (%)	$p$
$M_{12}N_{12}$	...	...	...	...
$M_{10}N_{12}$	<b>1.89</b>	<b>1</b>	<b>0.50</b>	<b>1</b>
$M_8N_{12}$	-17.25	<0.0001	-9.04	<0.0001
$M_6N_{12}$	-44.68	<0.0001	-30.21	<0.0001
$M_{12}N_{10}$	<b>-8.86</b>	<b>0.152</b>	<b>-4.54</b>	<b>0.052</b>
$M_{10}N_{10}$	<b>-1.75</b>	<b>1</b>	<b>0.29</b>	<b>1</b>
$M_8N_{10}$	<b>-9.29</b>	<b>0.126</b>	<b>-4.17</b>	<b>0.299</b>
$M_6N_{10}$	-30.61	<0.0001	-25.04	<0.0001
$M_{12}N_8$	-28.71	<0.0001	-11.75	<0.0001
$M_{10}N_8$	-15.00	<0.0001	-5.29	0.034
$M_8N_8$	-11.04	0.016	-6.67	0.004
$M_6N_8$	-21.39	<0.0001	-13.21	<0.0001
$M_{12}N_6$	-41.79	<0.0001	-22.00	<0.0001
$M_{10}N_6$	-40.54	<0.0001	-17.87	<0.0001
$M_8N_6$	-22.00	<0.0001	-11.67	<0.0001
$M_6N_6$	-24.89	<0.0001	-14.96	<0.0001
$M_4N_4$	-38.25	<0.0001	-28.50	<0.0001
$M_{14}N_{12}$	-16.00	<0.0001	<b>0.21</b>	<b>1</b>

TABLE IV

Comparison of the performance in percent correct between the baseline condition ( $M_{12}N_{12}$ ) and the extended frequency range condition ( $M_{14}N_{12}$ ) for all four SNRs. The  $p$  values were found by one-way RM ANOVAs.

CI SNR (dB)	$M_{12}N_{12}$		$M_{14}N_{12}$		Difference	$p$
	Average	Stand. Dev.	Average	Stand. Dev.		
Quiet	90.57	8.79	79.29	13.50	-11.29	0.002
+10	87.86	9.19	72.14	15.14	-15.71	0.004
+5	76.86	15.94	61.86	22.14	-15.00	0.064
0	56.86	16.22	36.43	26.44	-20.43	0.054
NH SNR (dB)	$M_{12}N_{12}$		$M_{14}N_{12}$		Difference	$p$
	Average	Stand. Dev.	Average	Stand. Dev.		
Quiet	98.50	1.76	98.67	1.03	0.17	0.72
+10	98.00	1.55	97.83	1.17	-0.17	0.41
+5	96.67	3.01	95.17	4.45	-1.50	0.43
0	85.33	4.08	88.33	3.98	3.00	0.19