

Optimizing Frequency-to-Electrode Allocation in Cochlear Implants

Jaime R. Leigh*

Katherine R. Henshall*

Colette M. McKay*

Abstract

This study tested the hypothesis that speech perception of cochlear implant users could be improved by increasing the number of electrodes allocated to frequencies below 2.6 kHz, with correspondingly fewer electrodes allocated above this point. Eight users of the Nucleus CI22 implant participated and wore experimental maps in which the widths of the analysis filters were altered to provide either two or three extra low-frequency channels. Speech perception was tested following periods of take-home experience. Information transmission analysis of vowel and consonant confusions appear to support our hypothesis, showing that vowel first formant information was significantly better perceived with the experimental map and that consonant information was not significantly different. The results of CNC word testing appear to contradict this result, showing that the experimental map degraded both vowel and consonant perception. Overall, the experimental map had a small and variable effect on speech perception. These results do not support our hypothesis and indicate that further research is needed to investigate the possible effects of narrowing the filter bandwidth in low frequencies, as done in this experiment.

Key Words: Cochlear implants, speech perception

Abbreviations: CNC = Consonant-Nucleus-Consonant; CUNY sentences = City University of New York sentences; SNR = signal-to-noise ratio; SPEAR = Sound Processor for Electric and Acoustic Research

Sumario

Este estudio evaluó la hipótesis de que la percepción del lenguaje en los sujetos con implante coclear podía mejorarse incrementando el número de electrodos asignados a las frecuencias por debajo de 2.6 kHz, con menos electrodos asignados por encima de ese punto. Ocho usuarios del implante Nucleus CI22 participaron y usaron mapas experimentales, en los cuales se alteraron los anchos de los filtros de análisis para lograr dos o tres canales más de baja frecuencia. Se evaluó la percepción del lenguaje luego de períodos de uso hogareño. La confusión en el análisis de la transmisión de información de vocales y consonantes parece apoyar esta hipótesis, mostrando que la información sobre el primer formante de las vocales era significativamente mejor percibida con el mapa experimental, y que la información de las consonantes no era significativamente diferente. Los resultados en la evaluación de palabras CNC entran en contradicción con estos resultados, mostrando que el mapa experimental degradó tanto la percepción de las vocales como

*University of Melbourne, Department of Otolaryngology

Jaime Leigh, Cochlear Implant Clinic, Royal Victorian Eye and Ear Hospital, 32 Gisbourne Street, East Melbourne, 3002, Australia; Phone: 61 3 9929 8624; Fax: 61 3 9929 8625; E-mail: jleigh@bionicear.org

This work was supported by the Australian National Health and Medical Research Council of Australia (project grants 114226 and 208927) and the Garnett Passe and Rodney Williams Memorial Foundation.

de las consonantes. Globalmente, el mapa experimental tiene un pequeño pero variable efecto sobre la percepción del lenguaje. Los resultados no apoyan nuestra hipótesis e indican que se requiere de más investigación para explorar los posibles efectos de una reducción en el ancho de banda de los filtros en las frecuencias graves, como se hizo en este experimento.

Palabras Clave: Implante coclear, percepción del lenguaje

Abreviaturas: CNC = consonante-núcleo-consonante; Frases CUNY = Frases de la Universidad de la Ciudad de Nueva York; SNR = tasa señal-ruido; SPEAR = procesador de sonido para investigación acústica y eléctrica

The primary aim of the present study was to manipulate the frequency-to-electrode allocation in the cochlear implant speech processor, in order to investigate the potential benefits of allocating fewer electrodes to the frequency regions above 2600 Hz, thus allowing a greater number of electrode positions to be allocated to frequencies below 2600 Hz. Based on a study by McKay and Henshall (2002), it was hypothesized that vowel perception of users of the 22-channel implant would be improved by the allocation of fewer electrode positions to frequency regions above 2600 Hz (3–4 in contrast to 5–7 in the subject's clinic map), without a detrimental effect on consonant perception (as seen in McKay and Henshall, 2002). In the present study, a comparison of two maps, the clinical map and the experimental map, was used to test this hypothesis. All aspects of the subject's map (i.e., number of electrodes, number of analysis bands, mode of stimulation) remained constant, other than frequency allocation, thus minimizing secondary effects in the experiment.

The SPEAK strategy, used with the Nucleus CI22 (Cochlear Ltd., 1994), is a filterbank strategy, consisting of a set of bandpass filters covering the input frequency range of the speech spectrum. For a given number of active electrodes, the clinician can choose one from among several sets of fixed bandpass filters. A particular filterbank is described by the number of filterbands and a "frequency table," which defines the frequency boundaries of the individual bands.

The 20 channel default case is known as Frequency Table 9 (range 150 to 10,823), although Table 7 (range 120 to 8658) is now recommended (Skinner et al, 1995; Skinner et al, 1997). The filter bandwidths in all frequency tables are linearly distributed below approximately 1000 Hz, and logarithmically distributed above 1000 Hz.

Research requiring manipulation of individual filter bandwidths has been limited in the past because the bandwidth for individual filters could not be manipulated. The Sound Processor for Electric and Acoustic Research (SPEAR), developed by the Cooperative Research Centre for Cochlear Implant and Hearing Aid Innovation to facilitate research with acoustic hearing aids as well as with cochlear implants, utilizes digital signal processing exclusively to maximize flexibility. The benefit of using the SPEAR processor compared to the Spectra22 in order to investigate the frequency-to-electrode allocation and speech perception with cochlear implants is that the SPEAR allows flexibility in selecting parameter values for the sound-processing scheme. In the SPEAR, parameters such as the rate of stimulation, the number and shapes of filters used to analyze input signals, the number of outputs selected for stimulation, and the temporal order in which stimulation pulses are delivered to the implanted electrodes are programmable. In previous studies using the Spectra22 (Fishman et al, 1997; Fu and Shannon, 1999a; Fu and Shannon, 1999b; McKay and Henshall, 2002), it has been necessary to reduce the number of electrodes

or frequency ranges used in order to manipulate the frequency-to-electrode allocation, as the filters are fixed in the Spectra. The flexibility of the SPEAR allows investigation into the effect of manipulating frequency-to-electrode allocations on speech perception without the secondary effects associated with reducing the number of electrodes or the frequency range used.

The effects of altering the pattern of frequency-to-electrode mapping on speech perception has had some attention in research. A study by Fu and Shannon (1999a) compared a series of four-electrode maps. The maps were constructed by assigning the output of more than one filter to each electrode. In each of the four-electrode conditions, speech sounds were still analyzed into 20 frequency bands and processed according to the SPEAK processing strategy. With all other factors kept constant, the location and spacing of the four stimulated electrode pairs was varied systematically and the effect measured. Recognition of medial vowels and consonants was measured in three subjects using the custom four electrode maps. Overall, vowel and consonant recognition was found to be best when the match between the characteristic frequency difference between the four electrodes and the frequency differences across the corresponding analysis filters was closest and for the most apical electrode locations. Consonant recognition was relatively robust to alterations in electrode location and spacing. The best vowel scores were about 10% lower with the four-electrode speech processors than scores obtained with the full 20-electrode speech processors. However, the best consonant scores with four-electrode speech processors were similar to those obtained with the full 20-electrode speech processors. Information transmission analysis showed that voicing and manner (largely temporal envelope cues) were not strongly affected by changes in electrode location and spacing, whereas vowel recognition and consonant place of articulation (largely spectral cues) were strongly affected by the degree of what the authors refer to as "tonotopic warping," created by altering both the location and spacing of the activated electrodes. The authors concluded that, overall, the experiment showed that electrode location was more important for vowel perception than for consonant perception.

In a second study by Fu and Shannon (1999b), vowel discrimination ability of five users of the Nucleus CI22 implant and SPEAK strategy were tested using an experimental four-electrode speech processor with continuous interleaved sampling (CIS) speech processing strategy. In this study, the range of acoustic frequencies allocated to the four channels/electrodes was shifted systematically. The results showed that recognition of multitalker vowels was highly dependent on frequency allocation for all electrode configurations, that speech recognition is highly sensitive to the match between frequency information and electrode location and that vowel discrimination is best when the frequency allocation was most similar to that of the subject's normal processor.

The correlation between the ability to distinguish electrodes and speech recognition ability has been investigated by Nelson et al (1995) and by Henry et al (2000). Nelson found that the ability of implantees, using the Mini Speech Processor (MSP)/MultiPeak (MPEAK) strategy, to perceive consonants was correlated with their ability to discriminate electrodes. In the later study by Henry et al (2000) the amount of speech information in different frequency bands for implantees using the SPEAK strategy was compared to that for normal-hearing listeners. The study showed that the speech perception of implantees was correlated to their ability to discriminate adjacent electrodes. The authors measured the amount of transferred speech information in different frequency regions and correlated this with the subjects' ability to discriminate electrodes allocated to the same frequency regions, in the presence of random loudness variation. They found that, relative to normal-hearing listeners, speech information was, on average, significantly more reduced in the frequency regions between 170 Hz and 2680 Hz than in the region between 2680 Hz and 5744 Hz. There was a significant correlation between electrode discrimination ability and the amount of speech information perceived in the regions between 170–2680 Hz, which was not found for the region 2680–5744 Hz. This would therefore imply that spectral discrimination may not be as important for frequencies greater than 2680 Hz and may be relatively more important in the vowel format regions. If the correlations reflect a causal

relationship between electrode discrimination and speech information transfer, it can be proposed that speech perception would be improved if discrimination of electrodes allocated to the low-to-mid frequencies were improved. This hypothesis was tested by McKay and Henshall (2002). They compared two 10-electrode maps for users of the Nucleus CI22 implant; one in which 18 analysis filters were distributed evenly over the 10 electrodes, and one in which the first 9 of the 18 filters (up to 2600 Hz) were allocated to 9 of the 10 electrodes. Prior to the test sessions, each subject wore each map for two weeks away from the laboratory. Test sessions involved assessment of perception of words in quiet and sentences in noise. The authors found that the 10-electrode map that allocated low-frequency information to 9 electrodes provided better speech perception in noise and better vowel perception in quiet compared to the even-allocation 10-electrode map, whereas consonant perception in quiet was degraded in this map for some subjects. Overall, the results of this study suggest that speech perception of users of the 22-channel implant might be improved by the allocation of fewer electrode positions to frequency regions above 2600 Hz (2–5 in contrast to the standard 8–9), thus allowing a greater number of discriminable electrode positions to be allocated to frequencies below 2600 Hz. The above findings form the basis of the experiment below.

This study aimed to determine whether the speech recognition of cochlear implantees could be improved by allocating fewer electrodes to the frequency regions above 2600 Hz, thus allowing a greater number of electrode positions to be allocated to frequencies below 2600 Hz. This change in frequency-to-electrode allocation was achieved by manipulating the filter bandwidths, while keeping all other map parameters constant (i.e., number of electrodes, frequency range, rate, threshold and comfortably-loud levels).

METHODS

Eight subjects implanted with the Nucleus CI22 implant participated in the experiment. Relevant background information about the subjects is provided in Table 1. The only criteria applied when selecting subjects to participate in the study was their ability to attend weekly sessions over a period of 12 weeks. All subjects volunteered their time and received no payment other than reimbursement for batteries and travel expenses.

Two maps were constructed for each subject, both implementing the spectral peak (SPEAK) speech processing strategy (McDermott et al, 1993; Skinner et al, 1994) and implemented in the SPEAR v.3 processor. The first, the reference map, was constructed to be as similar to the subject's own clinic map as possible. Due to technical limitations of the

Table 1. Subject Details

Subject	Age (years)	Years of profound deafness	Years of implant experience	Etiology	Number of electrodes in use	Device type (prior to experiment)	Speech processing strategy
1	64	16	13	Otosclerosis	17	Spear v.2	SPEAK
2	66	2	9	Genetic, Progressive	20	ESPrIt 22	SPEAK
3	55	5	13	Familial, Progressive	17	Spectra	SPEAK
4	52	2	5	Unknown, Progressive	15	Spectra	SPEAK
5	48	5	13	Unknown, Sudden	20	ESPrIt 22	SPEAK
6	65	40	7	Otosclerosis	19	ESPrIt 22	SPEAK
7	79	9	18	Unknown, Progressive	16	ESPrIt 22	SPEAK
8	47	12	10	Otosclerosis	18	Spectra	SPEAK

SPEAR v.3 processor, an exact replication of the frequency boundaries for four of the eight subjects' clinically used frequency tables could not be achieved. Frequency boundaries of the bands in the reference and clinical maps were all within 30 Hz, with the exception of the highest frequency band of subjects 2 and 5. This band was 658 Hz narrower in the reference map compared to the subject's own clinic map for both subjects. This occurred because 8 kHz is the maximum frequency analyzable by the digital filters (defined by the sampling rate of the digital filters). This was not expected to produce a significant perceptual difference in the signal heard by subjects 2 and 5, as the frequency response of the speech processor microphone has a significant drop-off above 5 kHz. It was also demonstrated by Henshall and McKay (2001) that extending the high-frequency range of the subject's map above 6 kHz does not improve speech perception. All other parameters in the clinical map (mode of stimulation, pulse rate, pulse duration, interphase gap, and number of maxima) could be replicated in the reference map for all subjects.

The second map, the experimental map, was custom designed to allocate fewer electrodes to the high frequencies (3–4 in contrast to 5–7 in the subjects' clinical map) and correspondingly more to the lower frequencies. For a typical filterbank setup, the frequency bounds of the bandpass filters are roughly linearly spaced below 1 kHz and logarithmically (ratio) spaced above 1 kHz. This distribution of the filters in the subject's reference map formed the basis of the calculations to determine the frequency bounds of the analysis filters in the experimental map. A break point that represented a divide of low frequencies from high frequencies was chosen to be as close to 2600 Hz as possible, as it was below 2600 Hz that electrode discrimination was deemed to be most important (Henry et al, 2000). The reduced bandwidth for the linear filters in the experimental map (compared to the reference map) was determined by the following formula:

$$L = \frac{(a \times c)}{b}$$

L = bandwidth of linear filters in the experimental map

a = number of bands allocated below the break point (~2600 Hz) in the reference map

b = number of bands allocated below the break point (~2600 Hz) in the experimental map

c = bandwidth of the linear filters in the reference map

A ratio frequency step was then calculated for the bands above the break point using the following formula:

$$R^e = f^d$$

R = ratio steps for bands above the break point in the experimental map

d = number of bands allocated to frequencies above the break point in reference map

e = number of bands allocated to frequencies above the break point in experimental map

f = ratio step for bands above the break point in reference map

The ratio step size used for the bands that occurred above the linearly spaced bands but below the break point was calculated so as to smoothly divide the frequencies in this midregion over the required number of filters. The number of electrodes allocated to the high frequencies in the experimental map was dependent on the total number of active electrodes and the number of electrodes allocated to the high frequencies in the subjects reference map and was typically 2–3 (range 1–3) less than the reference map. All map parameters other than filter boundaries (including threshold [T] and comfortably loud [C] levels) were the same in both maps. Thus the sound-field thresholds were the same for each map (these were not measured, however).

Each subject was required to attend nine sessions. The first session involved providing each subject with the SPEAR v.3 processor containing the reference map, checking the T and C levels, and checking that the electrodes were loudness-balanced. Speech

perception data was collected in the remaining eight sessions.

The test sessions followed an ABBBBBAA pattern, where A was periods with the reference map and B was periods with the experimental map. The subjects each used the experimental map at home for a minimum of three weeks prior to speech perception testing in the first period B. This period was considered important to allow subjects to adapt to the frequency allocation change, as recommended by McKay and Henshall (2002). Test sessions were approximately one week apart after this period.

Vowel perception was assessed using a closed-set list of the 11 pure vowels of Australian English. Consonant perception was assessed using 12 of the 24 consonants of Australian English (/b, d, f, g, k, m, n, p, s, t, v, z/), as used by Dowell et al (1982). Presentation was from an audio recording of the vowel in the /h/-vowel-/d/context and consonants in an /a/-consonant-/a/ context, of a male speaker with an average Australian accent. There was one token of each vowel and consonant in the respective recordings. Vowels and consonants were presented at 70 dB SPL in randomized lists of 55 and 48 respectively, and percentage correct was determined.

Information transmission analysis (Miller and Nicely, 1955) was performed on the results of the vowel and consonant confusion tests in order to compare the types of error patterns made with each map. Vowel features analyzed included durational cues (long vs. short), and first and second formant frequency (low, mid or high). Consonant features included voicing cues (yes or no), place of articulation (three categories), manner of articulation (three categories), and amplitude envelope (five categories).

The Consonant-Nucleus-Consonant (CNC) word lists were used to assess speech perception of real words. Twenty-two lists, randomly selected from a choice of 30 phonetically balanced lists, were presented at 70 dB SPL (eight lists per map were used in the analysis for each subject). The lists were recorded on compact disc by a female speaker with an average Australian accent. Each list constituted a set of 50 words, based on the phoneme set of Peterson and Lehiste (1962) with some adaptations for Australian English. Response mode was by verbal repetition of the perceived word and scored by percentage phonemes correct. Analysis of the list results

was further broken down into percentage of vowel and consonants correct, allowing the analysis of vowel and consonant scores independently.

Two different adapted versions of the City University of New York (CUNY) sentence lists (Boothroyd et al, 1985) were used in this experiment. The first was a rerecorded version as used by Skinner et al (1994). These lists maintain the original construction of 12 sentences, scored by number of words correct out of a total of 102, but were randomized with the topic removed. Lists were presented in quiet, all words were scored, and a percentage correct score was obtained for each presented list. Sentence length varied from 3 to 14 words and was presented in quiet at 70 dB SPL. This test was only performed in the first session, and its purpose was to derive a subject-dependent target score for the adaptive sentence test in noise described below.

An adaptive sentence test in noise utilized another adaptation of the CUNY sentence lists. The sentence lists, renamed "CUNY-like sentences," comprised 20 sentences from the original lists, which had a length of between 3 and 7 words. The computer program randomized sentence order, and response mode was by verbal repetition of the perceived sentence. The adaptive procedure was based on a method described by Levitt (1978) and was used by Henshall and McKay (2001). Presentation level of the speech stimulus was 68 dB SPL. The first sentence was presented in speech masking noise at a signal-to-noise ratio (SNR) of 20 dB. The percentage correct score for that sentence was calculated and compared to the target score (70% of the subject's own CUNY score in quiet). Where the actual score was lower than the target score, the noise level was reduced, and where it was higher, the noise level was increased. This process was repeated for each subsequent sentence. Noise step sizes were 3 dB, and the test was completed when eight reversals were recorded or all 20 sentences had been presented. The SNR was averaged for the last six turning points.

A similar adaptive test in noise was followed for subject 4, using a closed-set test of 16 spondees rather than CUNY-like sentences. This subject found the sentence test in noise to be too difficult. The speech stimulus was set so that average level of the

spondees was 70 dB SPL. Training in identification of the spondees in quiet was conducted until subject 4 was able to identify the correct spondee 100% of the time. The target percentage correct for use in the adaptive test was 50%. Spondees were selected randomly, and the first spondee was presented in speech masking noise at an SNR of 20 dB. If the response was incorrect, the noise level was reduced, and if correct, the noise level was increased. This process was repeated for each subsequent spondee. Noise level step sizes were 3 dB, and the test was completed when 10 reversals were recorded. The SNR was averaged for the last six turning points.

All testing took place in a sound-controlled booth meeting the standards outlined in ANSI S3.1-1991. The vowel and consonant confusions, adaptive CUNY-like sentences, adaptive spondee tests, and masking noise were played from a PC sound card transduced through a single loudspeaker, and the CNC word lists were played from a compact disc player transduced through a single loudspeaker. Participants were seated one and a half meters from the loudspeaker. The sensitivities of their speech processors were held constant throughout the experiment.

In each test session, one vowel, one consonant confusion, and two lists of CNC words were presented in quiet for all subjects. Two measures of the SNR were obtained for the adaptive (CUNY-like) sentence test for subjects 1-3 and 5-8 or, for spondees in the case of subject 4, in each subsequent test session.

RESULTS

Raw scores for all eight subjects for each map were analyzed using a two-way analysis of variance (ANOVA), with subject as a random factor and map as a fixed factor, utilizing the General Linear Model.

CNC word list results were separated into vowel and consonant results to allow evaluation of vowel and consonant scores separately.

Standardized residuals from general linear model testing were recorded and tested for normality using a Ryan-Joiner test. This test looks at the probability that any deviation observed in the data is due to something other than random error. A $p > 0.100$ indicates

there is no evidence to suggest that the distribution of the standardized residuals is non-normal. Transformed scores were used in the analysis for CNC vowel scores and the adaptive test scores, as these did not pass the normality test. Once transformed, the standard residuals were again tested for normality and were found to be normal.

Figure 1 shows the results of the vowel and consonant confusions with the reference and experimental maps for each subject, and the mean results for all eight subjects. A two-way ANOVA conducted on the vowel confusion results showed a significant subject effect ($F = 23.46$, $p < 0.001$), no significant map effect ($F = 2.35$, $p = 0.155$), and no significant interaction effect ($F = 0.84$, $p = 0.561$). A two-way ANOVA conducted on the consonant

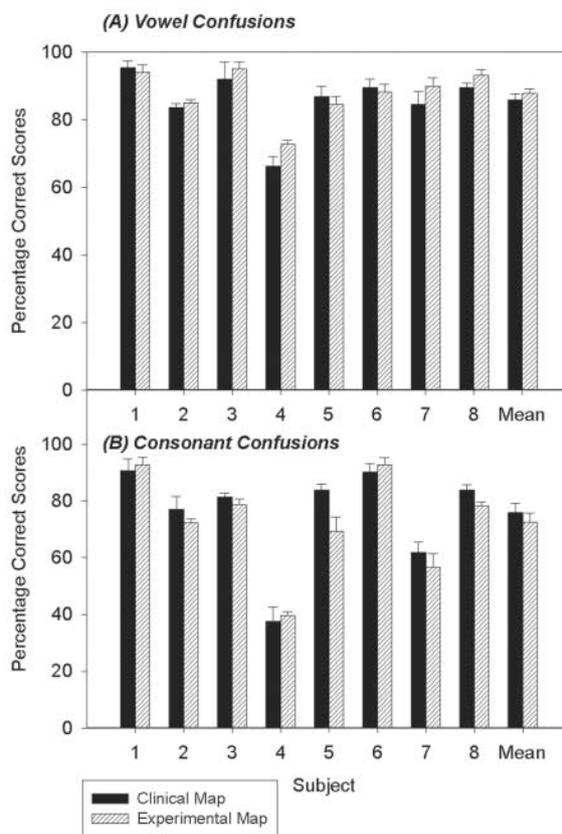


Figure 1. Individual and pooled mean vowel (A) and consonant (B) confusion scores (with standard errors) for eight subjects comparing their reference and experimental maps.

confusion results showed a significant subject effect ($F = 37.80, p < 0.001$), no significant map effect ($F = 2.64, p = 0.148$), and no significant interaction effect ($F = 1.60, p = 0.158$).

Figure 2 shows the results for the information transmission analysis for percentage of information transmission for the pooled data (data for all subjects summed) for both the vowel and consonant confusions. Results of the analysis showed that the experimental map yielded significantly higher scores for percentage of information transmission for total information transmission (T -value = $-3.58, p = 0.009$), duration ($T = -2.53, p = 0.039$), and F1 information ($T = -7.15, p < 0.001$). These differences were found to be statistically significant. No statistically significant differences were found for any of the consonant features.

Figure 3 shows the results of the CNC

vowel and consonant tests with the reference and experimental maps for each subject, and the mean results for all eight subjects. A two-way ANOVA was used to determine if an overall difference existed between the two maps for all subjects. For the CNC vowel results, it was found that there was a significant effect for subject ($F = 9.20, p = 0.019$) and map ($F = 16.03, p = 0.001$), with the reference map being better than the experimental map. There was also a significant interaction effect ($F = 3.21, p = 0.004$). The interaction can be explained by looking at the individual subject results for the CNC vowel scores. It is illustrated in Figure 3 that all subjects performed better with the reference map but that the increase in performance was much greater for subjects 5 and 6. CNC consonant results showed a significant subject ($F = 14.65, p = 0.006$) and

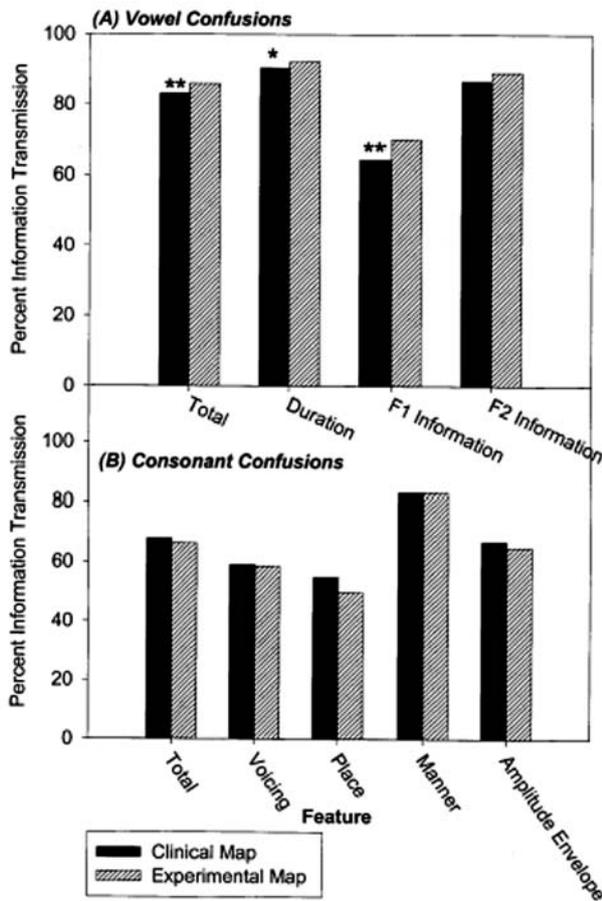


Figure 2. Results of the information transmission analysis for the percentage of information transmission for the pooled data for vowel (A) and consonant (B) confusions. The significant results are shown by an asterisk.

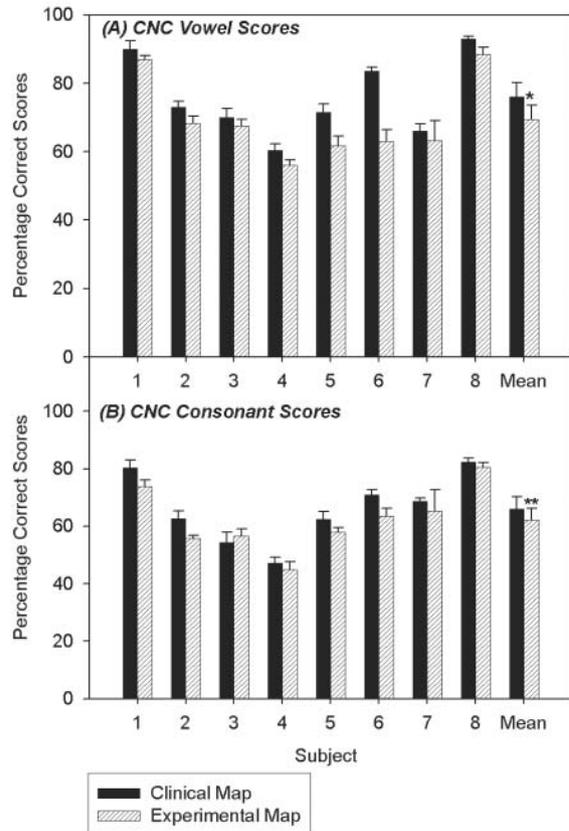


Figure 3. Individual and pooled mean CNC vowel (A) and CNC consonant (B) scores (with standard errors) for eight subjects comparing their reference and experimental maps. The significant results for the overall means are shown by an asterisk.

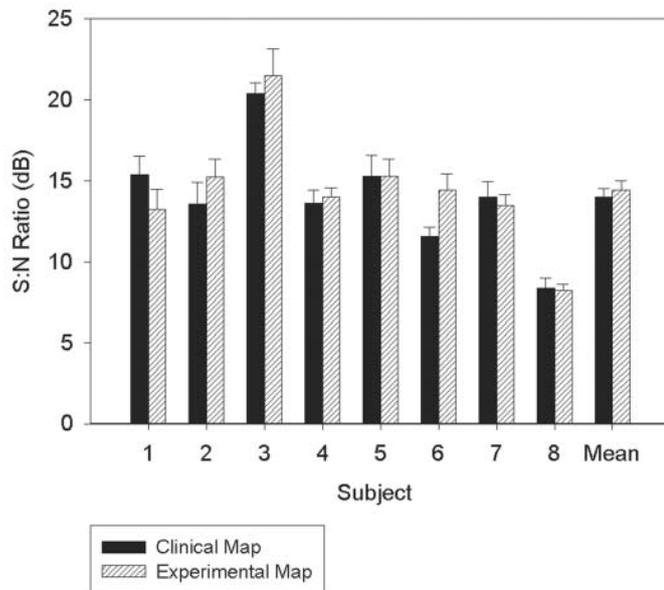


Figure 4. Individual and pooled mean adaptive test results (with standard errors) for eight subjects comparing their reference and experimental maps.

map effect ($F = 61.40$, $p < 0.001$) with the reference map giving a significantly higher score than the experimental map, and no significant interaction effect ($F = 0.80$, $p = 0.589$).

Figure 4 shows the results of the adaptive tests with the reference and experimental maps for each subject, and the mean results for all eight subjects. The y-axis is signal-to-noise ratio, and therefore lower bars indicate better performance. A two-way ANOVA was used to determine if an overall difference existed between the two maps. It was found that there was a significant subject effect ($F = 24.41$, $p < 0.001$), and no significant map ($F = 0.66$, $p = 0.442$) or interaction effect ($F = 1.32$, $p = 0.24$).

DISCUSSION

This experiment tested the hypothesis that increasing the number of low-frequency filters (by reducing their bandwidth) and decreasing the number of high-frequency filters (by increasing their bandwidth) would improve speech perception. In particular, we hypothesized (based on McKay and Henshall, 2002) that vowel perception would be improved, without a significant degradation of consonant information. In general, the results of this experiment show that the

manipulations in the experimental map had very minimal effect on speech perception. The results of the analysis of information transmission appear to support our initial hypothesis, showing that vowel first formant information was significantly better perceived with the experimental map, and that consonant information was not significantly different. However, the results of the CNC word test appear to contradict this result, showing that both vowel and consonant perception were degraded by the experimental map.

One reason for these contradictory results may be that there is in fact no effect of the experimental map, and the significant analyses are the result of type I error (i.e., the results are significant by chance alone). A perhaps more likely reason is that the experimental map may have had different effects for the different speech tests. There are several differences in the phoneme confusion tests and the CNC tests that could potentially affect the way that the experimental map altered perception. First, the word test contains lexical and co-articulation cues that are absent in the phoneme-confusion tests. This would seem unlikely to result in the pattern of results found in this experiment, however. Presumably, the use of lexical cues in the CNC words would be the same for both maps, and the co-articulation cues would be expected to be improved in the CNC words if the vowel formant information is improved in the vowel-confusion test.

A second difference between the phoneme-confusion and CNC word tests was that the former used a male speaker and the latter used a female speaker. This may be a significant difference when the processing by the reference and experimental maps is considered. In both cases, the acoustic signal is passed through a series of bandpass filters and the 6 electrodes corresponding to largest filter outputs are selected for stimulation. In the normal operation of this process the individual harmonics of the speaker are not resolved by the analysis, and the filters with the largest outputs correspond to the vowel formant frequencies. However, if the filters used are narrow and the harmonics are far apart (as for women and children), it is possible that the electrodes selected will correspond not to the formant frequencies but to the harmonic frequencies. This might lead to errors of vowel perception for the

implantees. In the case of this experiment, the lower filters are certainly narrow enough for this effect to occur with the female speaker. The bandwidths of the linear filters (up to 2856 Hz) in the experimental maps were as low as 128 Hz, whereas the harmonic spacing for the female speaker was 200–230 Hz. This combination will certainly allow individual harmonics to be represented in the output of SPEAK.

Another study, by Fu et al (2002), also found a speaker-gender effect when the bandwidths in the vowel regions were reduced. In that experiment, three users of the Nucleus-22 implant and using the SPEAK strategy wore an experimental map for three months, in which the frequency-to-electrode allocation was altered by using a frequency table that shifted the center frequencies of all filters downward by about an octave, also halving the filter bandwidths (to 100 Hz in the first eight filters). Two of the three subjects were quickly able to fully adjust to the frequency allocation shift when listening to men's voices. However, for female voices, and even more strikingly for children's voices, vowel perception was severely affected over the whole three-month period for all three subjects. Although the authors interpreted this result as evidence for inability to adjust to the frequency-shift, it is likely, given the extremely strong interaction of the effect with speaker gender, that the narrow filters, combined with the electrode selection of the SPEAK strategy, contributed to the continued degradation of vowel perception, via the mechanism described above.

In summary, the experiment in the present paper, and that of McKay and Henshall (2002), both attempted to improve vowel perception by spreading out the allocation of vowel frequencies over a wider cochlear spatial extent. In this experiment, the clinical frequency allocation was altered by narrowing the filter bandwidths in the vowel region (leading to a larger number of electrodes for this frequency region), whereas in the previous study, the filters were not altered but spacing of the electrodes was increased (by not using some electrodes). The two experiments produced somewhat different results for vowel perception. The former study showed no effect on vowel perception, whereas the present study showed an improvement of vowel information on one test (vowel confusions, male speaker) and a degradation on another test (CNC words,

female speaker). The inconsistency in the results may reflect the fact that the effects are all rather small in the subjects tested. Alternatively, it is possible that they reflect an effect of speaker gender (or F0) that is apparent when using narrow filters.

CONCLUSION

The effect of increasing the number of low-frequency filters (by narrowing the bandwidths) and decreasing the number of high-frequency filters used in the SPEAK strategy had small and variable effects on speech perception. In particular, it appeared to improve first-formant vowel information for male speakers in a vowel confusion test but decrease vowel and consonant perception for female speakers in words. It is possible that this variable effect may be due to the use of narrow bandwidths in the low-frequency region, in addition to the extra number of filters. That is, when filters are narrow and harmonics are widely spaced (for females and children), the SPEAK strategy may select electrodes corresponding to harmonic frequencies rather than vowel formant frequencies. This possibility needs further investigation. It is likely that this effect would not occur in situations where the filters are wider or all electrodes are selected (as in the CIS strategy, for example).

Acknowledgments. We thank Andrew Vandali for his assistance with programming the SPEAK processors. We also thank two anonymous reviewers for their valuable comments and the cochlear implantees who dedicated their time to this study.

REFERENCES

- Boothroyd A, Hanin L, Hnath T. (1985) *A Sentence Test of Speech Perception: Reliability, Set Equivalence, and Short Term Learning*. New York: City University of New York.
- Cochlear Ltd. (1994) *Technical Reference Manual*. Sydney: Cochlear Ltd.
- Dowell RC, Martin LF, Tong YC, Clark GM, Seligman PM, Patrick JF. (1982) A 12-consonant confusion study on a multiple-channel cochlear implant patient. *J Speech Hear Res* 25:509–516.
- Fishman KE, Shannon RV, Slattery WH. (1997) Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor. *J Speech Hear Res* 40:1201–1215.
- Fu QJ, Shannon RV. (1999a) Effects of electrode location and spacing on phoneme recognition with the Nucleus-22 cochlear implant. *Ear Hear* 20:321–331.

Fu QJ, Shannon RV. (1999b) Effects of electrode configuration and frequency allocation on vowel recognition with the Nucleus-22 cochlear implant. *Ear Hear* 20:332–344.

Fu QJ, Shannon RV, Galvin JJ, 3rd. (2002) Perceptual learning following changes in the frequency-to-electrode assignment with the Nucleus-22 cochlear implant. *J Acoust Soc Am* 112:1664–1674.

Henry BA, McKay CM, McDermott HJ, Clark GM. (2000) The relationship between speech perception and electrode discrimination in cochlear implantees. *J Acoust Soc Am* 108:1269–1280.

Henshall KR, McKay CM. (2001) Optimizing electrode and filter selection in cochlear implant speech processor maps. *J Am Acad Audiol* 12:478–489.

Levitt H. (1978) Adaptive testing in audiology. *Scand Audiol Suppl* 6:241–291.

McDermott HJ, Vandali AE, van Hoesel RJM, McKay CM, Harrison JM, Cohen LT. (1993) A portable programmable digital sound processor for cochlear implant research. *IEEE Trans Rehabil Eng* 1:94–100.

McKay CM, Henshall KR. (2002) Frequency-to-electrode allocation and speech perception with cochlear implants. *J Acoust Soc Am* 111:1036–1044.

Miller GA, Nicely PE. (1955) An analysis of perceptual confusions among some English consonants. *J Acoust Soc Am* 27:301–315.

Nelson DA, Van Tasell DJ, Schroder AC, Soli S, Levine S. (1995) Electrode ranking of “place pitch” and speech recognition in electrical hearing. *J Acoust Soc Am* 98:1987–1999.

Peterson G, Lehiste I. (1962) Revised CNC lists for auditory tests. *J Speech Hear Disord* 27:62–70.

Skinner MW, Clark GM, Whitford LA, Seligman PM, Staller SJ, Shipp DB, Shallop JK, Everingham C, Menapace CM, Arndt PL, Antogenelli T, Brimacombe JA, Pijl S, Daniels P, George CR, McDermott HJ, Beiter AL. (1994) Evaluation of a new spectral peak coding strategy for the Nucleus 22 Channel Cochlear Implant System. *Am J Otol* 15(suppl. 2):15–27.

Skinner MW, Holden LK, Holden TA. (1995) Effect of frequency boundary assignment on speech recognition with the SPEAK speech-coding strategy. *Ann Otol Rhinol Laryngol Suppl* 104:307–311.

Skinner MW, Holden LK, Holden TA. (1997) Parameter selection to optimize speech recognition with the Nucleus implant. *Otolaryngol Head Neck Surg* 117:188–195.