

## Speech Perception by Adults

MICHAEL F. DORMAN

Approximately half of the population of adult patients fit with the current generation of cochlear implants achieve scores of 80% to 100% correct on tests of sentence recognition when the sentences are presented by sound alone. In functional terms, this outcome means that more than half of the adult patients can work in an environment that demands the use of a telephone. Thus, the majority of adult implant patients can function in the workplace with only a few restrictions—listening in a noisy environment remains difficult for implant patients, as it is difficult for “normally” hearing-impaired individuals. The aim of this chapter is to describe how far implants have come since the landmark paper by Djourno and Eyries in 1957, and to describe how far we have to go in order to say that implants transmit as much information as they can, given the number of channels of stimulation provided by a given device.

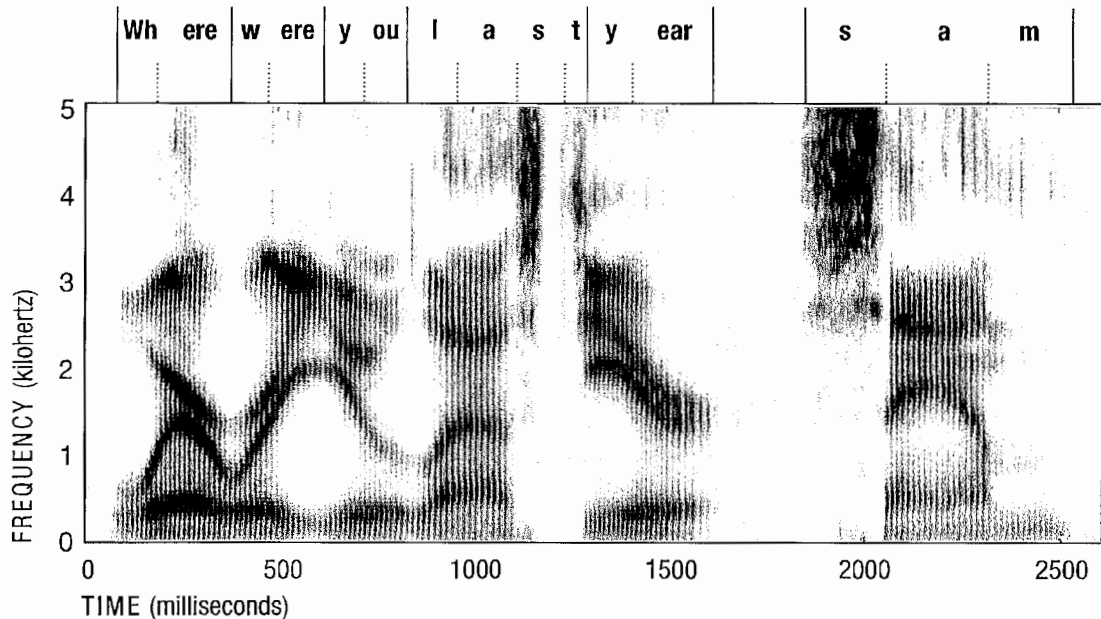
### ■ The Signal

In order to understand the design of a cochlear implant, it is necessary to appreciate the nature of the speech signals that implants are designed to transmit. As shown in Figure 16-1 in the spectrogram of the utterance, “Where were you last year, Sam?” speech is composed of frequencies that can range from approximately 125 Hz, the pitch of a male voice, to 4 to 5 kHz, the frequencies that signal the presence of “s” in *Sam*. The upper range of frequencies in female and children’s speech can be, for “s,” as high as 7 to 9 kHz. Thus, the band pass filter at the “front end” of an implant must pass a wide range of frequencies. Lawson et al. (1996), for example, found

better speech recognition for implant patients with input filters set to an upper limit of 9.5 kHz than with filters set to an upper limit of 5.5 kHz.

It is convenient to divide the information in the speech signal into two categories. In one category is the information about the identity of a sound that exists in the amplitude envelope of the signal and can be coded by a single electrode. In the other category is the information about the identity of a sound that exists in the frequency domain above several hundred hertz and can be coded best by multiple electrodes. \*The amplitude envelopes associated with consonants in a vowel-consonants-vowel (VCV) environment are shown in Figure 16-2. The shapes of the amplitude envelopes are different for stop consonants, semivowels, and nasals (i.e., for /ada/, /ata/, /ana/, and /ala/). The amplitude envelopes of the voiced and voiceless stop consonants /ada/ and /ata/ differ in terms of the presence and/or absence of noise between the burst release and vowel onset. The presence of long-duration, aperiodic noise signals the voiceless fricative in /asa/. These observations can be summarized by the statement that the shape and excitation (noise versus voiced) of amplitude envelopes can provide information about consonant “manner” (e.g., /d/ versus /l/) and consonant “voicing” (e.g., /d/ versus /t/) (see Rosen, 1992, for a discussion of envelope features with respect to speech perception by means of cochlear implants). Envelope

\*This account ignores rate pitch, which may excite frequencies to 1 kHz (Hochmair-Desoyer et al., 1983; Townshend et al., 1987). The devices discussed in this chapter generally use a low-pass filter at 400 Hz or less in the process of envelope detection. This operation minimizes the possibility of using rate pitch cues to frequency.



**FIGURE 16-1.** Spectrogram of the sentence, “Where were you last year, Sam.” Darkness indicates intensity.

features, however, provide little information about consonant “place of articulation” (e.g., /b/ versus /d/, /m/ versus /n/, or /s/ versus /f/) or about vowel identity. The cues for these sounds of speech reside in the frequency domain above several hundred hertz. The cues are shown in the spectrogram in Figure 16-1.

Figure 16-1 illustrates that, in the frequency domain, the sounds of speech are characterized by multiple concentrations of energy, or formants. In order to identify a vowel sound, an implant must be able to resolve, and then specify, where concentrations of energy occur in the frequency domain. The vowel “ou” in *you* is characterized by energy at 301 Hz and at 980 Hz. The vowel “a” in *last* has concentrations of energy at 592 Hz and 1357 Hz. In the case of vowel sounds that are very similar, such as the vowels in *bit* and *bet*, an implant must be able to resolve, at the level of the input filters, differences in formant frequencies of 100 to 200 Hz. These differences in input frequency must then be encoded by which electrode is stimulated in the cochlea, or by the pattern of activity among adjacent electrodes in the cochlea.

Because each of the sounds of speech has a different acoustic signature, the formant frequencies of the speech signal will change from moment to moment. For example, as shown in Figure 16-1, the “y” in *you* is characterized by a falling second formant that changes in frequency over a period of approximately 135 msec. Other sounds change more quickly. For example, the formant frequencies in “b,” “d,” and “g” (not shown in Fig. 16-1) change over a period of 10 to 50 msec. Thus, an implant must be

able to resolve where the formant frequencies are at a given moment in time and must be able to resolve rapid changes in formant frequencies over time. Given these requirements, we should suppose that the more input filters, and the more electrodes, the better an implant would function.

### ■ Strategies to Transmit Speech by Means of a Cochlear Implant

At present there are two, quite different strategies to transmit the information in speech by a cochlear implant [for reviews of the speech perception abilities of patients fit with older signal-processing strategies see, for example, Clark et al. (1990), Cohen et al. (1993), Dorman (1993), and Tyler et al. (1996)]. In one strategy, the “m-of-n” strategy, the signal, in the most common commercial implementation of the device, is divided into a relatively large number of channels (e.g.,  $n = 20$ ) in order to achieve a reasonably high degree of frequency resolution, and every 4 msec or so the 6 to 10 channels with the largest energy ( $m$ ) are determined and the electrodes associated with those channels are stimulated. Wilson et al. described this strategy in a generic form for cochlear implants in 1988. This type of strategy is implemented as the SPEAK strategy, run on the Spectra processor by Cochlear Corporation (McDermott and McKay, 1992; Skinner et al., 1994) and is implemented as a processing op-

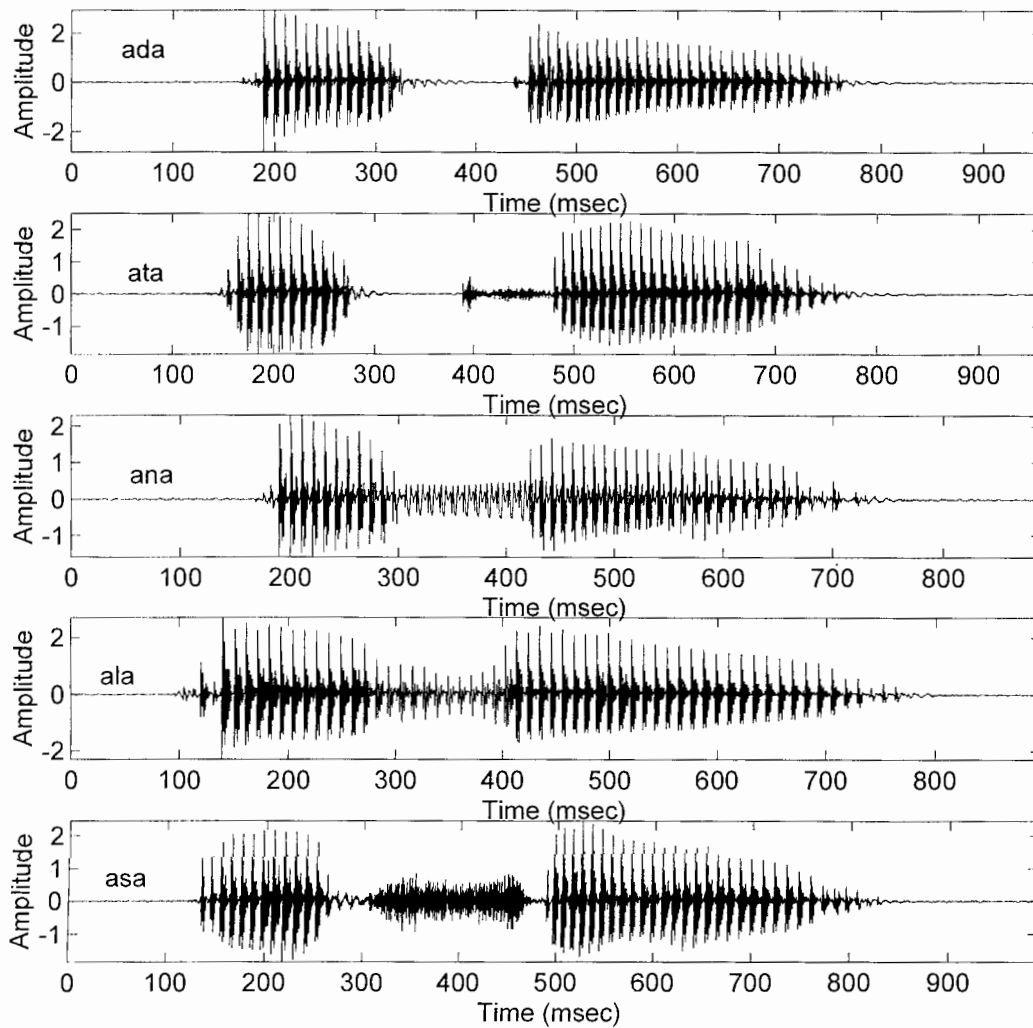


FIGURE 16-2. Amplitude envelopes of consonants in VCV syllables.

tion for Med El's Combi 40+ processor. The "m-of-n" strategy is an excellent way to transmit speech because only the peaks in the spectrum, or the formants, are transmitted, and as the formant frequencies change, so do the electrodes being stimulated. The change in location of the stimulated electrode mimics the normal change in location of spectral peaks along the basilar membrane. Experiments with normal-hearing listeners dating to the late 1940s and early 1950s have shown that speech can be transmitted with a high level of intelligibility if only information about the location of the formant peaks is transmitted [see, for example, the very first experiments on the Pattern Playback speech synthesizer by Cooper et al. (1950) and experiments on peak-picking vocoders by Peterson and Cooper (1957) as cited in Flanagan (1972)].

The other strategy to transmit speech by means of an implant is to divide the signal into a relatively small number of bands (e.g., 4–12) and to transmit the energy in all of the bands at each processor update cycle. This is one

of the strategies used in the Advanced Bionics Corporation's Clarion device, in the Med El Combi 40 devices, and in the Bionic Systems' LAURA device. This strategy, at first blush, seems an unlikely choice for speech transmission because frequency resolution is relatively poor, due to the small number of channels being stimulated (relative to the 20 "n" channels in the most common clinical implementation of the m-of-n strategy). Moreover, in contrast to the m-of-n strategy where the stimulated electrodes change as the formant frequencies change, in the fixed channel strategy all of the electrodes are stimulated on each update cycle and the location of the formant peaks must be inferred from the relative energy delivered to adjacent electrodes (Dorman et al., 1997a). In spite of these seeming drawbacks, experiments dating back to Dudley (1939) have shown that speech can be understood by normal-hearing individuals with a high degree of accuracy when transmitted by relatively few (7–10) fixed channels (e.g., Halsey and Swaffield, 1948; Hill et al., 1968).

## ■ Experiments on the Potential of m-of-n and Fixed-Channel Systems in Quiet

Several groups have conducted experiments in which speech signals were processed in the manner of m-of-n and fixed-channel, cochlear implant signal processors and were presented to normal-hearing listeners for identification. These experiments differed from the earlier experiments cited above in terms of both the motivation for the experiments (the previous experiments were designed to principally investigate sound quality for speech transmission systems) and in the details of the implementations of the signal processors (in the more recent experiments, signal processing has been based on cochlear implant signal processors, instead of speech transmission systems). The experiments have proven useful on at least three grounds. First, by presenting signals acoustically to normal-hearing individuals we can establish how well implant patients would perform if electrode arrays were able to reproduce, by artificial electrical stimulation, the stimulation produced by acoustic signals. This data can serve as a baseline to assess how close, or how far away, we are from recreating, with artificial electrical stimulation, the stimulation provided by acoustic stimulation of the cochlea. Second, experiments with normal-hearing listeners allow factors, normally confounded in implant patients, to be examined in isolation. For example, it is difficult to assess the effects of depth of electrode insertion on speech understanding because in patients with deep insertions there may be no viable neural elements near the electrode, whereas in patients with shallow insertions there may be viable neural elements. Of course, patients differ in many other ways, all of which are uncontrollable in a given experiment. Third, the experiments indicate how many channels are necessary to achieve a high level of speech understanding for both m-of-n and fixed-channel systems, a question that is central to the design of cochlear implants.

In experiments with fixed-channel processors, signals are first processed through a preemphasis filter (e.g., low-pass below 1200 Hz,  $-6$  dB per octave) and then bandpassed into  $N$  logarithmic frequency bands using, for example, sixth-order Butterworth filters. The envelope of the signal is extracted by full-wave rectification and low-pass filtering (e.g., second-order Butterworth with a 400-Hz cutoff frequency). Two types of outputs have been used. In some experiments sinusoids are generated with amplitudes equal to the root-mean-square (rms) energy of the envelopes (e.g., computed every 4 msec) and frequencies equal to the center frequencies of the bandpass filters. In other experiments, noise bands the width of the input filters are generated. The sinusoids or noise bands are finally summed and presented to listeners at a comfortable level. Signal processing for m-of-n

processors is similar in broad measure to that for fixed-channel processors. However, instead of directing the signals to each “ $n$ ” channel, an algorithm chooses the “ $m$ ” channels with the highest energy and directs signals to those channels at some update rate [see Kendall et al. (1997) for a detailed digital signal-processing simulation of the Spectra-22 speech processor].

Before the data for normal-hearing listeners are described, it is important to note that the normal-hearing listeners in these experiments had, most generally, only limited experience (generally between 2 and 10 hours) in listening to signals processed in  $N$  channels. Thus, the levels of performance obtained by the listeners are a conservative estimate of the levels of performance they might achieve if they had not hours, but years, of experience with the signals (as do implant patients).

### *Experiments with m-of-n Processors in Quiet*

Dorman et al. (in press, a) processed the single-syllable NU6 words through simulations of m-of-n processors when “ $n$ ” was set to 20 and “ $m$ ” was 2, 4, and 6. The 2 of 20 processor allowed a mean score of 73% correct, the 4 of 20 processor allowed a mean score of 87% correct, and the 6 of 20 processor allowed a mean score of 90% correct. Loizou and Tu (unpublished observations) processed the difficult, and sometimes quirky, sentences from the multiple-talker TIMIT data base (Lamel et al., 1987) and recorded scores of 70% correct for a 2 of 16 processor, 86% correct for a 4 of 16 processor, and 97% correct for a 6 of 16 processor. The results of these experiments are consistent with the much older literature on “peak-picking” channel vocoders (e.g., Peterson and Cooper, 1958) and demonstrate that a high level of speech intelligibility can be achieved when picking only a small handful of high-amplitude channels from a set of 16 to 20 channels [see Blamey et al. (1985) for acoustic simulations of early formant-tracking, speech-encoding strategies for cochlear implants; see also Blamey et al. (1987) for results for patients using the formant-tracking speech processor]. These data suggest that implant patients who use the Cochlear Corporation’s SPEAK strategy (which is a 6 to 10 of 20 strategy) could, if each channel were used effectively, achieve very high scores on tests of speech understanding. In a later section we will see if this is the case.

### *Experiments with Fixed-Channel Processors in Quiet*

In the experiment that prompted the recent interest in simulations of cochlear implant signal processors, Shannon et al. (1995) processed sentences into one, two, three and four bands, modulated narrow-band noise the width

of the filters by the energy in each band, and presented the summed bands to normal-hearing listeners for identification. Noise bands were used as output signals so that there would be no frequency-specific information within each band. On the test of sentence identification, two bands allowed mean scores of 25% correct; three bands allowed scores of 85% correct; and four bands allowed scores of 95% correct. Dorman et al. (1997a) substituted sine waves at the center of the filter frequencies for noise bands and replicated the Shannon et al. (1995) outcome recording mean scores of 43% for a two-channel processor, 72% correct for a three-channel processor, and 92% correct for a four-channel processor. The Shannon et al. (1995) and Dorman et al. (1997a) data converge to demonstrate that sentences, when processed in the manner of a cochlear implant signal processor, can be transmitted with a high degree of intelligibility with as few as four channels of stimulation.

The high degree of intelligibility for sentences transmitted with only a few channels of stimulation appears to arise from a combination of two effects. One is that the consonant features of *voicing* and *manner* can be well recognized with only one or two channels of stimulation (Dorman et al., 1997a; Shannon et al., 1995; Van Tassel et al., 1987) because information about voicing and manner can be conveyed by amplitude envelopes. If only information about manner and voicing are made available to a listener, the manner of word choices for an unknown signal is greatly reduced. Zue and colleagues report that if each of the segments of an unknown signal is described only in terms of six manner features (stop, weak fricative, strong fricative, semi-vowel, nasal, vowel), then the number of word matches in a 20,000 word vocabulary is reduced to approximately two possibilities (Huttenlocher and Zue, 1984; Pisoni et al., 1995; Zue, 1985). The other factor is “top-down” linguistic knowledge. Many experiments have shown that words are much easier to identify in noise when presented in sentence context than in isolation (e.g., Miller et al., 1951). In this instance, tacit knowledge of word order and semantics aids in word identification. Thus, the information in amplitude envelopes, in combination with only a little frequency-specific information, can, with the aid of top-down linguistic knowledge, lead to a high level of sentence recognition.

Other signals, for which there is less top-down linguistic information, require more channels to be well recognized. Synthetic vowels that differ only in formant frequencies (and not in vowel duration), naturally produced vowels from multiple talkers (Hillenbrand et al., 1994), and the consonant feature *place of articulation* (e.g., whether the consonant is from the set [“bdg”] or from the set [“wrlj”]) are identified with a high degree of accuracy only when the signals are processed into eight

or more channels (Dorman et al., 1997a; Shannon et al., 1995). Given this data, isolated words should also require a relatively large number of channels. This is indeed the case. Dorman et al. (in press, a) processed the NU-6 words through processors with 4, 6, 8, and 12 channels and found mean word scores of 48% correct, 72% correct, 87% correct, and 89% correct, respectively. The phoneme scores for the four conditions were 67% correct, 84% correct, 94% correct, and 96% correct. These data suggest that the Med El continuous interleaved sampling (CIS) device which drives 6 channels, Advanced Bionics Clarion device, which drives 8 channels, and the Med El devices with 8 and 12 channels could provide essentially perfect sentence understanding and greater than 70% NU-6 word understanding, if artificial electrical stimulation were to replicate the effects of the acoustic stimulation used in the experiments with normal-hearing listeners, and if patients possessed neural structures capable of transmitting the electrical stimulation. In the next section we will see if this is the case.

### ■ Speech Recognition by Adults with Different Signal-Processing Strategies and Different Numbers of Electrodes

In the previous two sections, conservative estimates of the levels of performance possible with *m* of *n* and fixed-channel processors were documented. In this section the performance of patients fit with cochlear implants is described and compared with the potential levels of performance that might be attained by the patients.

#### **Scores on Tests of Sentence Recognition**

As noted above, tests of word recognition in sentence context are relatively easy. Normal-hearing listeners, presented 4, 6, and 8-channel versions of the Hearing in Noise Test (HINT) (Nilsson et al., 1994) achieved mean scores of 92% correct, 99% correct, and 100% correct (Dorman et al., 1997a). If patients can extract as little as four channels of information from their processor, then patients' scores should fall in the 80% to 100% correct range. Unpublished data indicates that 16 of 21 patients fit with 5- and 6-channel Med El CIS processors posted scores between 88% and 100% correct. Helms et al. (1997) reported, for 41, 8-channel Med El patients tested 6 months after fitting, a median score of 90% correct with an interquartile range of 66% to 95% correct. Kessler (personal communication) reported, for 62, 8-channel Clarion patients tested 6 months after fitting, a median score of 84% correct with an interquartile range of 48% to 97% correct. Skinner et al. (1994) reported for 61 patients, who represented the top 75% of the distribution

of patients and who were switched from the Multi-Peak (MPEAK) to SPEAK strategy, a median score of 86% correct with an interquartile range of 67% to 96% correct (estimated from Fig. 4 of Skinner et al., 1994).

It is not possible, from the foregoing data, to compare the merit of one device versus another because of differences in how the patient samples were chosen and because of differences in test materials. However, overall, the results are what we should expect because patients need only to extract a few channels of information to achieve a high score on a test of sentence identification. A useful summary statement is that with the current signal processors, slightly more than half of the adult patients achieve a score between 80% and 100% correct. This is a level of performance that allows use of a telephone and enables patients to return to the workforce (see, for example, Mawman et al. in the papers following this chapter). From this point of view, cochlear implants in adults should be considered an overwhelming success. On the other hand, the data indicates that many patients do not receive the equivalent of even four channels of information. This issue is explored in the next section.

### **Scores on Tests of Word Recognition**

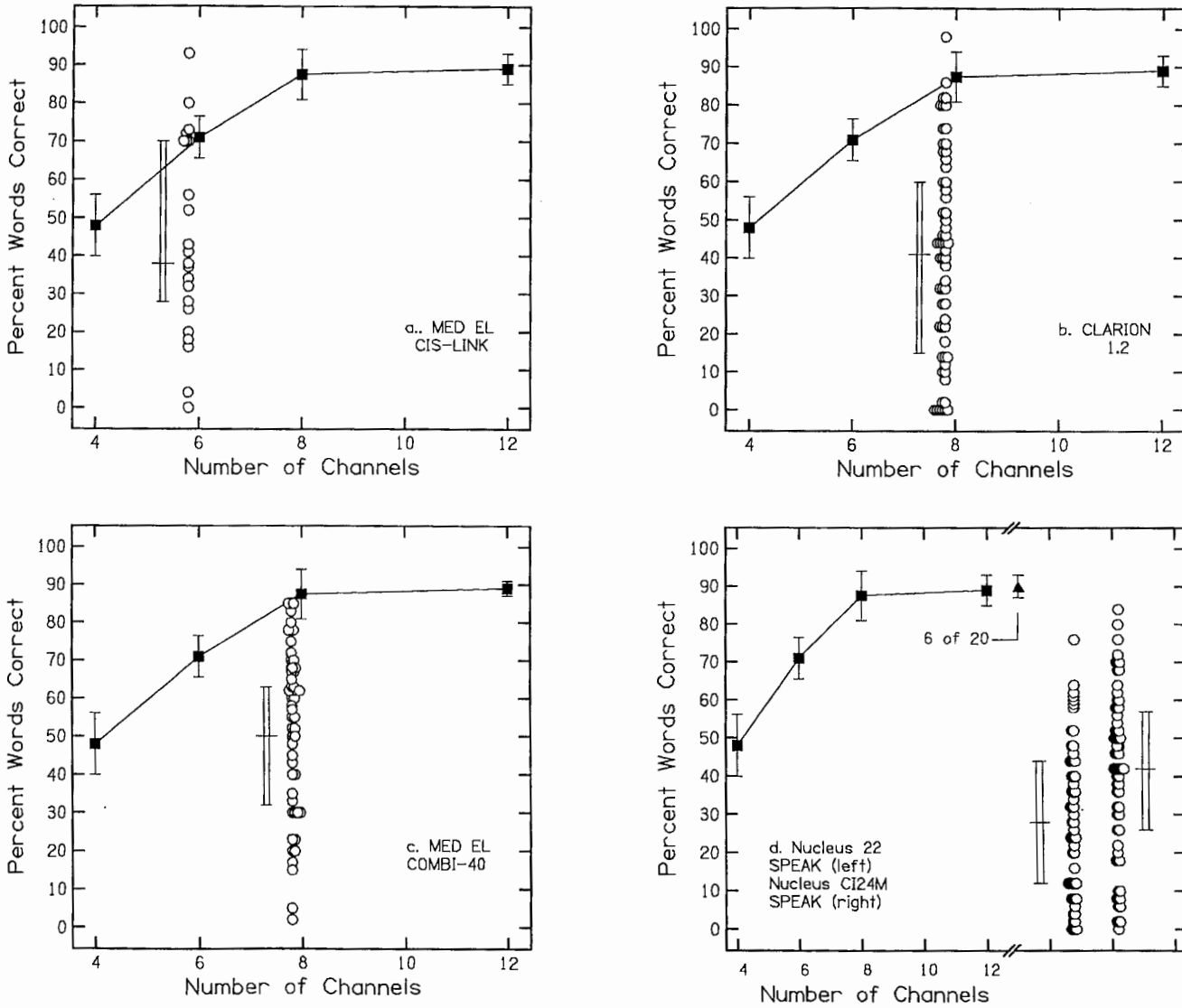
As pointed out above, tests of sentence understanding are easy. Indeed they are so easy that we cannot infer from the results of the tests how many channels of information patients receive through their processors. To answer this question, more difficult tests (e.g., monosyllabic word tests) are necessary. Figure 16-3 shows the level of performance achieved by normal-hearing subjects listening to the NU-6 word lists processed into 4, 6, and 8 channels and the performance of different groups of implant patients. The mean scores for the normal-hearing listeners were 48% correct, 72% correct, and 87% correct for the 4-, 6-, and 8-channel processors, respectively. Consider first, in Figure 16-3A, the small, and not necessarily representative, group of subjects who use 5- and 6-channel CIS processors. The mean score was 49% correct. The scores of five of the patients fell within, or just above,  $\pm 1$  standard deviation of scores for normal-hearing listeners (i.e., between 65% and 77% correct). It appears that these individuals were able to extract as much information from the electrically delivered signals as normal-hearing listeners were able to extract from the acoustically delivered signals.

Consider next the scores for two groups of patients fit with 8-channel CIS processor: the Med El Combi 40 and the Clarion version 1.2. The mean score for normal-hearing listeners with an 8-channel processor is 87% correct with  $\pm 1$  standard deviation limits at 81% correct and 94% correct. As shown in Figure 16-3B, the mean score for the

patients fit with the Clarion device is 40% correct (scores provided by D. Kessler). As shown in Figure 16-3B, the mean score for patients fit with the Med El device is 48% correct (scores provided by I. Hochmair-Desoyer from monosyllabic word test data collected in Europe). The median scores are 41% correct and 50% correct, respectively, for the Clarion and Med El devices. As was the case for sentence material, those data cannot be used to compare the relative merit of different processors because of differences in test materials, in time since implantation, and in length of deafness. For both sets of patients, scores range from near zero to very high. A few patients have scores within  $\pm 1$  standard deviation of the scores for normal-hearing listeners using an 8-channel processor. The majority of the “better” patients have scores within the range of scores for a 6-channel processor. A large number of patients have scores within, or below, the standard deviation of scores for a 4-channel processor. The single score at the high end of the continuum of scores—98% correct—is interesting because the patient is not an experienced test taker; indeed, he had never before been tested with the words. The high score, albeit beyond the  $\pm 1$  standard deviation points for normal-hearing listeners, is within the range of scores achieved by normal-hearing listeners.

Consider finally the scores for patients who use the Cochlear Corporation’s SPEAK strategy. The N22 data were from patient samples at the U. of Michigan, the House Ear Institute and the U. of Arkansas. The N24 sample was provided by Cochlear Corporation. As shown in Figure 16-3D, a 6 of 20 processor allows a mean score of 90% correct for normal-hearing listeners. The mean score for patients using the N22 SPEAK strategy was 29% correct. A similar score was reported by Battmer and Lenarz (1995) for patients in Germany. The mean score for patients using the N24 SPEAK strategy was 46% correct. The higher level of performance for the N24 patients was due, most likely, to better residual hearing for patients in that group (Rubinstein et al., 1999). As was the case for patients using other processors, there is a wide range of performance with the best patients achieving scores at the 6–8 channel level and a larger number of patients performing at, or below, the four channel level.

One feature stands out in Figure 16-3A—the 93% correct score for a patient fit with a 6-channel, CIS processor. This score is beyond the range of scores found for normal-hearing listeners. This subject has been a test subject in many hundreds of hours of testing and is one of the most tested, and psychophysically sophisticated, subjects in the entire implant population. If other 6-channel patients were able to achieve scores in the 90% correct range, then the data for normal-hearing listeners should be seen as far too conservative. However, this is



**FIGURE 16-3.** Monosyllabic word recognition as a function of the number of channels in a signal processor for normal-hearing listeners (filled squares). Patient performance is shown by open circles. The open rectangles indicate the interquartile range of performance. Horizontal bars indicate median scores.

not the case. The best patients achieve similar scores in the 70% to 80% correct range, just where they should be relative to the scores of normal-hearing listeners. Thus, the performance of this patient on this test is not representative of other 6-channel patients. This situation is similar to the situation of the single-channel patient C.K., who was able to score 98% correct on tests of sentence identification and 52% correct on a test of word identification—scores far beyond the range of scores achieved by the majority of other single-channel patients (Hochmair-Desoyer et al., 1985). The abilities of patients like the two just described, who are extreme outliers, are difficult to explain.

**■ Speech Recognition by Patients When the Number of Channels in a Processor is Reduced**

The data described above suggest that only a relatively small proportion of patients extract the equivalent of 6 or 8 channels of information from devices with 6, 8, and 10 channels. If this is the case, it should be possible to significantly reduce the number of channels implemented in most patients' processors and not reduce the patients' scores on tests of speech intelligibility. Several experiments indicate that this is the case.

Wilson (1997) tested five patients fit with a Nucleus 22 electrode array and a percutaneous connector (like that of the Ineraid). The percutaneous connector allowed different processing strategies (m of n and fixed-channel) to be implemented by means of a research processor for each patient. In no case did CIS processors with 8, 11, or 21 channels allow better performance than processors with 6 channels. Wilson also reported that the best 6-channel CIS processors allowed higher scores than the patients' SPEAK strategy running in monopolar mode.

Fishman et al. (1997) created, for patients using the SPEAK strategy, devices with 1, 2, 4, 7, and 20 electrodes by directing outputs from multiple filters to electrode pairs. For consonants and sentences, performance reached asymptote with four channels of stimulation. For vowels and monosyllabic words, performance reached asymptote with seven channels of stimulation. The data from Wilson (1997) and Fishman et al. (1997) fit well with the conclusion reached in the previous section—that most patients cannot take advantage of the information from more than a relatively small number of channels of stimulation (see, also, Kiefer et al., in the papers following this chapter, for results with the 8- and 12-channel Med El processors).

## ■ The Recognition of Frequency-Shifted Speech Signals

In the experiments with normal-hearing listeners described above, the sine-wave or noise-band output signals were delivered to the place in the cochlea corresponding either to the frequencies of the sine waves or the noise bands (i.e., the signals were delivered to the “correct” place in the cochlea). For implant patients this is not the case; signals are not commonly delivered to the correct place in the cochlea. Ketten et al. (1998) have reported, from analysis of ultra-high resolution computed tomography imaging of electrodes in the cochlea, that the deepest electrode in 20 Nucleus and 10 Ineraid patients was between 13.5 and 24.5 mm from the round window. The average insertion depth was 20.2 mm. This place of stimulation corresponds to a frequency of approximately 900 Hz, assuming a 33-mm canal length. This frequency is higher than the approximately 350-Hz center frequency of the first filter in, for example, an 8-channel signal processor. The amount of upshift for the other channels in an implant will depend on the distance between adjacent electrodes. For the extreme case of Ineraid electrodes 4 mm apart, given a 2 mm longer-than-average insertion depth of 22 mm, the output from channels 1 to 6, with center frequencies of 393, 639, 1037, 1685, 2736, and 4443 Hz, will be delivered to the 643, 1264, 2380, 4386, 7991, and 14,469 Hz places along the cochlea. This

magnitude of frequency upshift should cause great difficulty in speech understanding since the acoustic cues to phonetic contrasts fall into frequency regions far from the regions in which they normally fall (see Fig. 16-1).

To assess the effects of frequency upshifting on speech understanding, Dorman et al. (1997b) presented normal-hearing subjects with signals from a 5-channel processor that had been upshifted in frequency corresponding to 25- to 22-mm insertions into the cochlea with electrodes separated by 4 mm. For sentence material, a simulated 25-mm insertion allowed scores (94% correct) that did not differ significantly from scores collected when the stimuli were presented to the correct place in the cochlea (98% correct). However, the simulations of 22-, 23-, and 24-mm insertions produced significantly poorer performance (43% correct, 73% correct, and 85% correct, respectively). Thus, the decrease in intelligibility that should accompany a large upshift in frequency does, in fact, occur [see also Fu (1997), Fu and Shannon (1999a) and Rosen et al. (1997)].

Dorman et al. (1997b) noted that in implant patients the effect of insertion depth may be realized in performance on two different measures. One is the time taken to reach asymptotic performance. Many implant patients report that for a period after the implant is activated, speech sounds high pitched or unnatural in a number of ways. Over time, patients usually report that speech sounds more normal and that speech is more intelligible even though the parameters of stimulation have not changed. It is likely that shallow insertion depths are at least partially responsible for the reports of unnatural of high-pitched speech at the time of processor activation and for poor initial intelligibility. It is not likely that patients are sufficiently sophisticated to be able to tell the difference between an abnormally high pitch per se and upshifted formant frequencies. If shallow insertion depth is responsible for the initial percepts described by the patients, then patients can learn to hear through, or compensate for, the distortion in frequency representation because over time speech becomes more intelligible (e.g., Tye-Murray et al., 1992). Indeed, as shown in Figure 16-3 some patients with 6- and 8-channel processors, including patients fit with the Ineraid electrode array, can achieve a level of performance equivalent to that of normal-hearing subjects listening to speech processed into six and eight channels and presented to the correct place in the cochlea.

Rosen et al. (1997) have shown that adaptation to frequency-shifted signals can take place rapidly. When normal-hearing listeners were presented signals processed into four channels and were upshifted by the equivalent of a 6.5 mm basalward shift in frequency, performance decreased from 64% correct to 1% correct for words in sentences, from 42% to 5% correct for vowels in

lbVdl format, and from 57% to 33% correct for VCV syllables. After only 3 hours of training, performance on the vowel and sentence tests increased to about half of that of the unshifted conditions. Performance on the consonant test was not different from that in the unshifted condition, due most likely to the temporal cues available for manner and voicing, but the recognition of place of articulation (cued by information in the frequency domain) remained poorer than in the unshifted condition. It is not clear how long normal-hearing listeners would take, if ever, to completely adapt to the shifts in frequency occasioned by large basalward shifts in cochlear place of stimulation [see, for example, Kileny et al. (1992) for data from implant patients]. The outcome of such an experiment would be important because it would give important information about the flexibility of the cortical pattern recognition routines involved in speech recognition.

The second measure that could be affected by cochlear place of stimulation is the level of terminal performance. As shown in Figure 16-3, the majority of patients do not achieve the level of performance of normal-hearing subjects listening to signals filtered into six or eight channels. Distortion in the mapping of input frequency to cochlear place of stimulation may be one of many factors responsible for low levels of terminal performance for some implant patients. On this view, patients may be able to compensate for only distortions of a modest magnitude in the mapping of input frequency to cochlear place of stimulation. The data from patients who have a short electrode insertion speaks to this issue. The results of several studies indicate that insertion depth does have a significant effect on performance (Bredberg and Lindstrom, 1995; Hartrampf et al., 1995; Kileny et al., 1992; Marsh et al., 1993). However, note the performance of the CIS patients using the Ineraid electrode array who achieved scores within the range of normal-hearing listeners on the NU-6 test. These data suggest that, over time, cortical processing routines can adapt to at least some large changes in the absolute location of frequency information. An important difference between these patients and the patients with short electrode insertion is that the deepest electrode is located at a more basal location for patients with short insertions. Fu and Shannon (in press, a) reported that performance for implant patients decreases as the deepest electrode is moved to a more basal location. This issue is revisited in a later section.

It is important to note that cortical pattern recognition routines for speech recognition must be inherently flexible because men, women, and children do not produce the same set of formant frequencies when producing the same phonetic segment (due to differences in vocal tract dimensions). For example, the first two formant frequen-

cies for the vowel /ae/ are, on average, at 660 and 1720 Hz for men, 860 and 2050 Hz for women, and 1010 and 2320 Hz for children (Peterson and Barney, 1952). Thus, pattern recognition routines must, in the normal course of events, be able to extract an unvarying phonetic description of physically different signals. Perhaps it is this inherent flexibility, found even in young infants (Kuhl, 1979) and probably a part of our biological endowment, that allows frequency-shifted signals to be heard appropriately. Yet there are, no doubt, limits to the flexibility of the pattern recognition routines for frequency-shifted signals. Perhaps if the frequency shifting is linear along some scale, then recognition routines can adapt (consider the equal log spacing of the Ineraid's electrodes). However, if frequency space is severely warped or distorted, then recognition may break down. It may be that poor performers are patients who have populations of surviving spiral ganglion cells that are remote from one or more of the intracochlear electrodes and create a warped map of frequency. Indeed, Fu and Shannon (1999a) suggest that this factor may be more important to speech intelligibility than other factors, such as current interaction among electrodes, which have long been thought to characterize patients with poor scores on tests of speech identification [see Blamey et al. (1996) for a comprehensive review of factors that may account for differences in patient performance].

A reasonable signal-processing solution to the problem of upshifted frequencies is to match the place of stimulation with the appropriate input filter. Fu and Shannon (1999b) created 4-channel processors for patients using the 20-electrode Nucleus system and systematically mapped the first channel of the 4-channel systems to more basal electrodes. As noted above, as the first (most apical) electrode of the 4-electrode set became more basal, speech intelligibility decreased. Fu and Shannon then paired each of the more basal electrode maps with 10 different frequency allocation tables. The maps had a significant effect on performance, demonstrating that matching the frequency allocation table to the location of the electrodes can improve performance. However, there will most likely be a limit to the benefit achieved by such matching. If low frequencies in the speech signal are completely left out of the input signal, in an attempt to match frequencies with a basal electrode location, then speech understanding will suffer (Fletcher, 1953; Miller and Nicely, 1955). Thus, there will be a trade-off between the benefits due to matching input filters to electrode location and the losses due to high-pass filtering.

Another solution to the problem of the mismatch between analysis filters and electrode location is to develop electrode arrays that can be inserted more apically into the cochlea. An array that has achieved greater than

30-mm insertion depths is in use with the Med El Combi 40+ system.

## ■ Speech Perception in Noise

In previous sections, the performance of normal-hearing listeners and implant patients was assessed in quiet. In the real world outside of a sound booth, speech signals are likely to occur against a background of environmental noise. It is possible that the conclusions reached above (i.e., that a small number of channels are sufficient to reach a high level of performance) will change for signals in noise.

### *The Potential of m-of-n Processors in Noise*

The level of performance allowed by a processor with a given number of channels will, of course, decrease as the signal-to-noise ratio (SNR) becomes poorer. For example, in tests with normal-hearing listeners, Kendall et al. (1997) reported that a 4 of 7 processor that allows a score of 97% correct on a test of sentence recognition in quiet, allows 74% correct at +12 dB SNR, and allows 59% correct at +6 dB SNR. In order to achieve a high score in noise, either more “m” or “n” channels are needed. Loizou et al. (in press) reported, for HINT sentences at +2 dB SNR, scores of 46% correct, 61% correct, 83% correct, and 90% correct, respectively, for 2, 4, 6, and 8 of 12 processors. Thus, increasing the number of “m” channels improves performance. So does increasing the number of “n” channels. Loizou et al. (in press) reported scores of 60% correct for a 6 of 8 processor and 77% correct for a 6 of 16 processor (i.e., processor with twice as many “n” channels). Kendall et al. (1997) reported, for BKB sentences at +6 dB SNR, scores of 59% correct for a 4 of 7 processor and scores of 82% for a 4 of 20 processor. Thus, increasing the number of “n” channels increases performance in noise. However, even a processor with 20 channels does not allow recognition at the level of unprocessed speech when “m” is small. Kendall et al. reported a score of 86% correct for a 6 of 20 processor at +6 dB SNR, but a score of 98% correct for unprocessed speech at the same SNR.

### *The Potential of Fixed-Channel Processors in Noise*

The results for speech recognition in noise for m of n processors generally demonstrate that more channels are needed in noise than in quiet to achieve a high level of speech recognition. A similar outcome obtains for fixed-channel processors. Dorman et al. (1998b) presented the HINT sentences processed into 2 to 20 channels in quiet

and at +2 dB and -2 dB SNR to normal-hearing listeners for identification. Scores for an 8-channel processor in quiet, at +2 and at -2 dB SNR were 100% correct, 55% correct and 15% correct, respectively. In quiet, performance was limited by a ceiling effect (97% correct) with five channels of stimulation. At +2 dB SNR, performance reached asymptote (85% correct) with 12 channels of stimulation. At -2 dB SNR performance reached asymptote (63% correct) with 20 channels of stimulation. Thus, as noted above, more channels are needed in noise than in quiet to achieve a high level of speech recognition (Eddington et al., 1997; Fu, 1997). It is also the case for fixed-channel processors in noise, as for m-of-n processors, that processed signals are less intelligible than unprocessed signals. Fu et al. (1998) reported that 16 band processors, which allowed scores similar to those for unprocessed speech in quiet, allowed much poorer scores at +12 to 0 dB SNR.

### *Patient Performance in Noise*

Kiefer et al. (1996) compared the performance in noise of patients who used the N22 SPEAK strategy and patients who used a high rate CIS strategy (the Med EL Combi 40). For a relatively easy test of sentence understanding, Kiefer et al. (1996) reported, for patients using the SPEAK strategy, scores of 79% correct in quiet, 66% correct at +15 SNR, and 56% correct at +10 dB SNR. The scores for patients using the CIS strategy were 88% correct, 81% correct, and 72% correct, respectively. The patients also were tested with more difficult sentences. On these tests the scores with the SPEAK strategy were 23% correct at +15 dB SNR and 12% correct at 10 dB SNR. With the CIS strategy the scores were 45% correct and 28% correct, respectively [see also Battmer and Lenarz (1995)]. The scores for the two processors differed significantly. As always, definitive answers to questions about group differences as a function of processor design must wait until much larger patient samples have been examined.

## ■ A Look at Some Assumptions

Over much of the history of cochlear implants, the following statements were widely assumed to be correct. More channels are better than fewer because more channels provide better frequency resolution. Bipolar electrodes are better than monopolar electrodes because bipolar electrodes limit current spread. Nonsimultaneous stimulation across channels is better than simultaneous stimulation because channel interactions are kept to a minimum with nonsimultaneous stimulation. Some of

these assumptions have not proved viable and others are being questioned.

As noted above, more electrodes, or channels, are better than fewer, especially in noise, although four channels allow very high levels of sentence understanding in quiet. Unfortunately, it has proven difficult to provide patients with more than four to eight effective channels of stimulation. The presumed advantage for bipolar electrodes has failed to materialize. Patients using Advanced Bionics' Clarion devices, and Mel El's Combi 40 devices, all of which drive monopolar electrodes, perform well relative to patients with bipolar electrode configurations (Figure 16-3). Zwolan et al. (1996) reported that Nucleus patients programmed in monopolar mode perform as well, or better, than patients programmed in bipolar mode (see also Battmer et al., 1993; von Wallenberg et al., 1994). The view that nonsimultaneous stimulation should allow better speech understanding than simultaneous stimulation seemed to be on very firm ground. Wilson et al. (1991) reported improvement, and in some cases dramatic improvement, in speech understanding for Ineraid patients fit with a CIS strategy when compared with the performance of the same subjects using the simultaneous, compressed analog strategy. The large advantage of CIS over simultaneous, compressed analog stimulation has been replicated many times (Boex et al., 1994; Dorman and Loizou, 1997). Given this history, it will be of interest to view the results of Advanced Bionics' field trial of a simultaneous stimulation scheme (the SAS strategy). The motivation for the trial is data collected in Europe indicating that about half of the patient sample preferred the SAS strategy to a CIS strategy. It remains to be seen whether speech understanding, for some patients, is better with the SAS strategy than with a CIS strategy. If it is, then we should look to the possibility that details of the low-frequency spectrum, in the frequency domain of the first formant in speech, are better coded by rate pitch in the SAS scheme than by place pitch in the CIS scheme.\* Electrode design also differs between the Ineraid and Clarion devices.

## ■ Summary

The current generation of cochlear implants allow a majority of patients to achieve high scores, greater than 80%

\* It may be that the place frequency maps created using the Greenwood (1990) equation are misleading in the region of the apex for implant patients because the elements stimulated in implant patients appear to be, in the main, cell bodies in the spiral ganglion. Because the spiral ganglion does not extend as far apically as the basilar membrane, a clump of cell bodies with low-frequency input in the spiral ganglion near the apex may respond to the most apically inserted electrode, even if the electrode is not inserted to the "appropriate" place by reference to the basilar membrane.

correct, on tests of sentence recognition and allow mean scores of about half that level on tests of monosyllabic word recognition. The high level of performance on the tests of sentence recognition is possible because only four channels of information need to be transmitted to achieve a high level of sentence understanding. More channels need to be transmitted in order to achieve high scores on tests of monosyllabic word recognition. Only a few patients achieve the level of performance possible with eight channels of stimulation (i.e., 80%–90% recognition). The "better" patients achieve scores in the range of a 6-channel processor for normal-hearing listeners (i.e., 65%–75% recognition). The median score for patients is at the level of a 4-channel processor for normal-hearing listeners (i.e., 40%–55% recognition).

More channels are needed in noise than in quiet to achieve high scores on tests of sentence identification. The number of channels necessary to achieve a given level of performance varies, of course, with the SNR. Twelve channels allow a high level of performance down to near 0 dB SNR. Given the data cited above, it will be some time before patients will achieve high levels of performance in environments characterized by modest amounts of noise.

## ■ Acknowledgment

The writing of this chapter was supported by a grant from The National Institute on Deafness and Other Communication Disorders (R01-DC-000654-7).

## REFERENCES

- Battmer, R., Gnadeberg, D., von Wallenberg, E., Lenhardt, E., Allum, D. (1993). A study of monopolar and bipolar stimulation modes with a modified Nucleus mini-22 cochlear implant. *Adv Otorhinolaryngol* 48:9–16.
- Battmer, R., Lenarz, T. (1995). The performance of the Nucleus Spectra 22 and the Clarion CIS devices in quiet and in noise. Abstracts of the IIIrd Paris International Cochlear Implant Conference, p. 330.
- Blamey, P., Dowell, R., Brown, A., Clark, G. (1987). Vowel and consonant recognition of cochlear implant patients using formant-estimating speech processors. *J Acoust Soc Am* 82:48–57.
- Blamey, P., Arndt, P., Bergeron, F., et al. (1996). Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants. *Audiol Neurootol* 1:293–306.
- Blamey, P., Martin, L., Clark, G. (1985). A comparison of three speech coding strategies using an acoustic model of a cochlear implant. *J Acoust Soc Am* 77:209–217.
- Boex, C., Pellizzone, M., Montandon, P. (1994). Improvements in speech understanding with the CIS strategy for the Ineraid multichannel cochlear implant. In Hochmair-Desoyer, I., Hochmair, E., eds. *Advances in Cochlear Implants*, pp. 136–140. Vienna: International Interscience Seminars.
- Bredberg, G., Lindstrom, B. (1995). Insertion length of electrode array and its relation to speech communication performance and nonaudi-

- tory side effects in multichannel-implanted patients. *Ann Otolaryngol Rhinol Laryngol* 104 (Suppl 166):256–258.
- Clark, G., Tong, Y., Patrick, J. (1990). *Cochlear Prostheses*. New York: Churchill Livingstone.
- Cohen, N., Waltzman, S., Fisher, S.G. (1993). Department of Veterans Affairs Cochlear Implant Study group: A prospective, randomized study of cochlear implants. *N Engl J Med* 328:233–237.
- Cooper, F., Liberman, A., Borst, J. (1950). Preliminary studies of speech produced by a pattern playback. *J Acoust Soc Am* 22:678.
- Djournio, A., Eyries, C. (1957). Prothese auditive par excitation électrique a distance du nerf sensoriel a l'aide d'un bobbinage inclus a demeure. *Presse Med* 35:14–17.
- Dorman, M. (1993). Speech Perception by Adults. In Tyler, R., ed. *Cochlear Implants*. San Diego: Singular Publishing Group.
- Dorman, M., Loizou, P. (1997). Mechanisms of vowel recognition for Ineraid patients fit with continuous interleaved sampling processors. *J Acoust Soc Am* 102:581–587.
- Dorman, M., Loizou, P., Fitzke, J., Tu, Z. (in press). The recognition of NU-6 words by cochlear-implant patients and by normal-hearing subjects listening to NU-6 words in the manner of CIS and SPEAK strategies. *Ann Otol Rhinol Laryngol*.
- Dorman, M., Loizou, P., Fitzke, J., Tu, Z. (1998). 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*.
- Dorman, M.F., Loizou, P., Rainey, D. (1997a). Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. *J Acoust Soc Am* 102:2403–2411.
- Dorman, M.F., Loizou, P., Rainey, D. (1997b). Simulating the effect of cochlear-implant electrode insertion depth on speech understanding. *J Acoust Soc Am* 102:2993–2996.
- Dudley, H. (1939). Remaking speech. *J Acoust Soc Am* 11:169–177.
- Eddington, D., Rabinowitz, W., Tierney, J., Noel, Whearty, M. (1997). Speech processors for auditory prostheses. Eighth Quarterly Progress Report on NIH Contract N01-DC-6-2100.
- Fishman, K., Shannon, R., Slatery, W. (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.
- Flanagan, J. (1972). *Speech Analysis, Synthesis and Perception*, 2nd ed, pp. 329–331. New York: Springer-Verlag.
- Fletcher, H. (1953). *Speech and Hearing in Communication*, pp. 418–423. Van Nostrand.
- Fu, Q.-J. (1997). Speech perception in acoustic and electric hearing [Ph. D. dissertation]. Los Angeles: University of Southern California.
- Fu, Q.-J., Shannon, R. (1999a). Effects of electrode location and spacing on speech recognition with the Nucleus-22 cochlear implant. *Ear Hear* 20.
- Fu, Q.-J., Shannon, R. (1999b). Effects of electrode configuration and frequency allocation on vowel recognition with the Nucleus 22 cochlear implant. *Ear Hear* 20.
- Fu, Q.-J., Shannon, R., Wang, X. (1998). Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing. *J Acoust Soc Am*, 104:3586–3596.
- Greenwood, D. (1990). A cochlear frequency-position function for several species—29 years later. *J Acoust Soc Am* 87:2592–2605.
- Halsey, R., Swaffield, J. (1948). *Analysis-Synthesis Telephony, with Special Reference to the Vocoder*, pp. 391–411. London: Institute of Electrical Engineers.
- Hartrampf, R., Dahm, M., Battmer, R.D., Gnadenberg, M., Straub-Schier, A., Rost, M., Lenarz, T. (1995). Insertion depth of the nucleus electrode array and relative performance. *Ann Otolaryngol Rhinol Laryngol* 104(Suppl 166):277–279.
- Helms, J., Muller, J., Schon, F., et al. (1997). Evaluation of performance with the Combi 40 cochlear implant in adults: A multicentric clinic study. *Otolaryngol Rhinol Laryngol* 59:23–35.
- Hill, J., McRae, P., McClellan, R. (1968). Speech recognition as a function of channel capacity in a discrete set of channels. *J Acoust Soc Am* 44:13–18.
- Hillenbrand, J., Getty, L., Clark, M., Wheeler, K. (1994). Acoustic characteristics of American English vowels. *J Acoust Soc Am* 97:3099–3111.
- Hochmair-Desoyer, I., Hochmair, E., Burian, K., Stiglbrunner, H. (1983). Percepts from the Vienna cochlear prosthesis. In Parkins, C.O., Anderson, S., eds. *Cochlear Prostheses*, pp. 295–306. New York: New York Academy of Sciences.
- Hochmair-Desoyer, I., Hochmair, E., Stiglbrunner, H. (1985). Psychoacoustic temporal processing and speech understanding in cochlear implant patients. In Schindler, R., Merzenich, M., eds. *Cochlear Implants*, pp. 291–304. New York: Raven.
- Huttenlocher, D., Zue, V. (1994). A model of lexical access based on partial phonetic information. *Proc ICASSP-84* 2.
- Kendall, M., Summerfield, A., Chambers, J. (1997). Digital signal-processing simulation of the Spectra-22 speech processor. Manuscript from the MRC Institute of Hearing Research, Nottingham, UK.
- Ketten, D., Skinner, M., Wang, G., Vannier, M., Gates, G., Neeley, G. (1998). In vivo measures of cochlear length and Nucleus cochlear-implant array insertion depth. *Ann Otol Rhinol Laryngol*.
- Kiefer, J., Muller, J., Pfenningdorf, T.H. (1996). Speech understanding in quiet and in noise with the CIS speech coding strategy compared to the multipeak and spectral peak strategies. *Oto Rhino Laryngol* 58:127–135.
- Kileny, P., Zimmerman-Phillips, S., Zwolan, T., Kemink, J. (1992). Effects of channel number and place of stimulation on performance with the Cochlear Corporation multichannel implant. *Am J Otol* 13:117–123.
- Kuhl, P. (1979). Speech perception in early infancy: perceptual constancy for spectrally dissimilar vowel categories. *J Acoust Soc Am* 66:1668–1679.
- Lamel, L., Kassel, R., Seneff, S. (1986). Speech database development: design and analysis of the acoustic-phonetic corpus. *Proceedings of DARPA Speech Recognition Workshop*, pp. 100–109.
- Lawson, D., Wilson, B., Zerbi, M., Finley, C. (1996). Speech processors for auditory prostheses. Third Quarterly Progress Report, NIH contract N01-DC-5-2103.
- Loizou, R., Dorman, M., Tu, Z. (In press). The recognition of sentences in noise by normal-hearing listeners using simulations of SPEAK-type cochlear implant signal processors. *Ann Otol Rhinol Laryngol*.
- Marsh, M., Xu, J., Blamey, P., et al. (1993). Radiologic evaluation of multichannel intracochlear implant insertion depth. *Am J Otol* 14:386–391.
- McDermott, H., McKay, C. (1992). A new portable sound processor for the University of Melbourne/Nucleus Limited multielectrode cochlear implant. *J Acoust Soc Am* 91:3367–3371.
- Miller, G., Heise, G., Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *J Exp Psychol* 41:329–335.
- Miller, G., Nicely, P. (1955). An acoustic analysis of perceptual confusions among some English consonants. *J Acoust Soc Am* 27:338–346.
- Nilsson, M., Soli, S., Sullivan, J. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and noise. *J Acoust Soc Am* 95:1085–1099.
- Peterson, G., Barney, H. (1952). Control methods used in a study of vowels. *J Acoust Soc Am* 24:175–184.
- Peterson, G., Cooper, F. (1957). Peakpicker: A bandwidth compression device. *J Acoust Soc Am* 29:777(A).
- Pisoni, D., Nusbaum, H., Luce, P., Słowiaczek, L. (1985). Speech perception, word recognition and the structure of the lexicon. *Speech Commun* 4:75–95.

- Rosen, S. (1992). Temporal information in speech and its relevance for cochlear implants. *Philos Trans R Soc Lond* 336:333-367.
- Rosen, S., Faulkner, A., Wilkinson, L. (1997). Perceptual adaptation by normal listeners to upward shifts of spectral information in speech and its relevance for users of cochlear implants. In *Speech, Hearing, and Language: Work in Progress, Phonetics and Linguistic*, Vol. 10. London: University College.
- Rubinstein, J., Parkinson, W., Tyler, R., Gantz, B. (1999). Residual speech recognition and cochlear implant performance: Effects of implantation criteria. *Am J Otol* 20(3):445-452.
- Shannon, R., Zeng, F-G., Kamath, V., Wygonski, J., Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science* 270:303-304.
- Skinner, M., Clark, G., Whitford, L., Seligman, P., Staller, S. Shipp, D., Shallop, J., Everingham, C., Menapace, C., Arndt, P., Antogenelli, T., Brimacombe, J., Pijl, S., Daniels, P., George, C., McDermont, H., Beiter, A. (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.
- Townshend, B., Cotter, N., Compennolle, D., White, R. (1987). Pitch perception by cochlear implant subjects. *J Acoust Soc Am* 82: 106-115.
- Tye-Murray, N., Tyler, R., Woodworth, G., Gantz, B. (1992). Performance over time with a Nucleus or Ineraid cochlear implant. *Ear Hear* 13:200-209.
- Tyler, R., Gantz, B., Woodworth, G., Parkinson, A., Lowder, M., Schum, L. (1996). Initial independent results with the Clarion cochlear implant. *Ear Hear* 17:528-536.
- Van Tassel, D., Soli, S., Kirby, V., Widin, G. (1987). Speech envelope cues for consonant recognition. *J Acoust Soc Am* 82:1152-1161.
- von Wallingberg, E., Battmer, R., Doden, I., Gnadenberg, D., Hautle, K., Lenarz, T. (1994). Place pitch and speech perception measures with bipolar and monopolar electrical stimulation of the cochlea. Abstracts of the International Cochlear Implant, Speech and Hearing Symposium, Melbourne, Australia, October 24-28.
- Wilson, B. (1997). The future of cochlear implants. *Br J Audiol* 31:205-225.
- Wilson, B., Finley, C., Farmer, J., Lawson, D., Weber, B., Wolford, R., Kenan, P., White, M., Merzenich, M., Schindler, R. (1988). Comparative studies of speech processing strategies for cochlear implants. *Laryngoscope* 98:1069-1077.
- Zue, V. (1985). The use of speech knowledge in automatic speech recognition. *Proc IEEE* 73:1602-1615.
- Zwolan, T., Kileny, P., Ashbaugh, C., Telian, S. (1996). Patient performance with the Cochlear Corporation "20+2" implant: bipolar versus monopolar activation. *Am J Otol* 17:717-723.