Speech Processors for Auditory Prostheses

NIH Contract NO1-DC-92100

QPR #4: Oct-Nov-Dec 1999

Dynamic range and speech recognition in cochlear implant listeners

Submitted by

Fan-Gang Zeng and Ginger Grant
Department of Hearing and Speech Science
University of Maryland
College Park, MD 20742

Robert V. Shannon, John Galvin III, and Qian-Jie Fu
House Ear Institute
Los Angeles, CA 90057

Jane Opie and Phil Segel
Advanced Bionics Corporation
Sylmar, CA 91342

January 31, 2000

Table of Contents: **PAGE**

ABSTRACT

In this quarterly progress report, we update research on dynamic range and speech recognition in cochlear implant listeners. Our objective is to study how to optimally set the acoustic dynamic range and map it into the electric dynamic range in cochlear-implant listeners. Here we present empirical data on acoustic dynamic range, electric dynamic range, and electrode interactions. We also report how these dynamic ranges and electrode interactions affect speech recognition in Clarion and Nucleus implant users. Acoustic analyses of phoneme tokens produced by 5-female and 5-male talkers showed a 50 dB speech dynamic range, much wider than the commonly assumed 30 dB range. Psychophysical data collected over a large set of parameters including monopolar and bipolar modes, sinusoidal and pulsatile waveforms, stimulus frequency, and pulse duration showed that electric dynamic range rarely exceeds 30 dB. Modeling and experimental data indicated that electrode interactions should be taken into account when fitting speech processors of modern multi-electrode cochlear implants. Direct electrical field summation across electrodes reduces the effective number of independent channels, while loudness summation reduces the effective dynamic range under multiple electrode stimulation. Corresponding speech recognition experiments also yielded interesting results. An optimal performance was achieved when the acoustic dynamic range was set at 50 dB, consistent with the acoustic analysis results. Reducing electric dynamic range produced little effects on Nucleus users of the SPEAK strategy but decreased speech recognition in Clarion users of the CIS and SAS strategies. When electrode interactions were taken into account in a new “speech-adjusted” fitting strategy, noticeable improvement in speech recognition was observed in two Clarion users. The present data indicate that both acoustic and electric dynamic ranges need to be optimized in cochlear implants and different optimizations may be needed for different processing strategies. Furthermore, the present results suggest that electrode interactions must be taken into account in speech processor fittings and new fitting protocols should be developed to use wide-band, dynamic speech to fit cochlear implants under realistic listening situations. These new fitting protocols may be particularly useful in children who use cochlear implants.

Advances in speech processing strategies have allowed most modern multi-electrode cochlear implant users to enjoy a high level of speech recognition including telephone use. However, there is a great deal of variability among individual users and still a significant gap in performance between cochlear-implant and normal-hearing listeners under realistic listening situations such as noise (Skinner et al., 1994; Kessler et al., 1997; Fu and Shannon, 1999; Zeng and Galvin, 1999). While the individual variability and the performance gap may have an origin in etiology, deafness duration, linguistic and other cognitive capabilities, they are also likely due to not-yet-optimized parameters in the speech processor. In this report, we consider the relationship between dynamic range and speech recognition in cochlear implant users.

To accommodate the narrow dynamic range in electric hearing, all speech processors need to map or compress large variations of acoustic amplitude into an electric amplitude within the electric dynamic range. In Nucleus devices, a 30-dB acoustic range is converted into an electric current value that evokes an auditory sensation between threshold and most comfortable loudness. In Med-El devices, a 60-dB acoustic dynamic range is used. In Clarion devices, the acoustic range can be as narrow as 10 dB for users of the CIS strategy and as wide as 80 dB for users of the SAS strategy. The electric range, on the other hand, is typically determined by presenting an isolated pulse train or sinusoid on a single pair of electrodes. At present, the clinical fitting of these amplitude parameters relies mostly on experience and lacks experimental validation. We do not know whether these presently-used amplitude mappings are optimal for speech recognition in cochlear implant users, or whether different mappings are necessary for different speech strategies in individual users under different listening conditions.

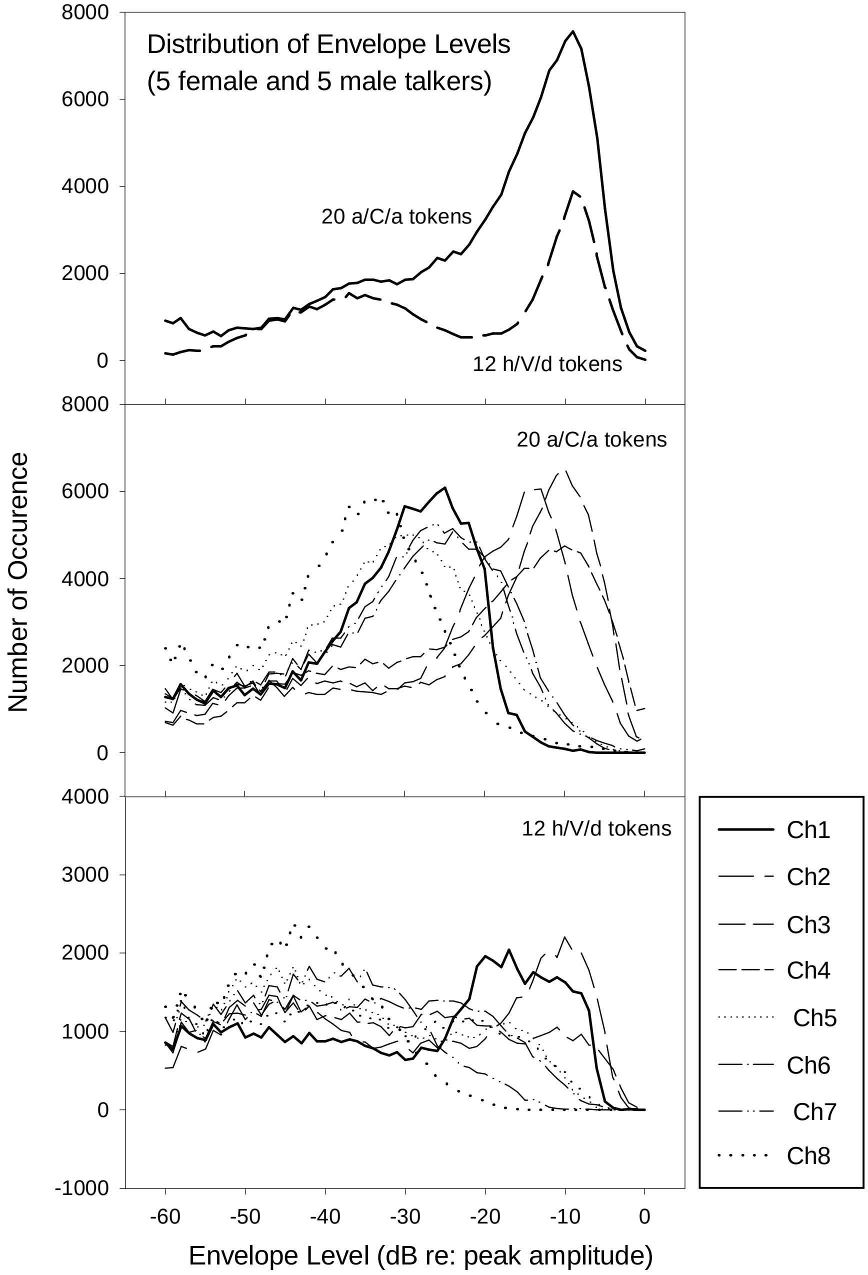
In this report, we first present results from 3 experiments. (1) We review literature and provide analysis on acoustic dynamic range of speech sounds. (2) We present data on electric dynamic range in single-electrode stimulation. (3) We develop a model of electrode interactions in multi-electrode stimulation and show how electrode interactions affect the number of independent channels and the effective electric dynamic range. In the second section, we show how acoustic and electric dynamic range manipulations and electrode interactions affect speech recognition in Clarion and Nucleus users. Finally, we identify the need to use speech sounds as an optimal and more natural way to fit cochlear implants.

Acoustic Dynamic Range

A normal-hearing person can process acoustic information that varies by 12 orders of magnitude in intensity. This large dynamic range, coupled with fine intensity resolution (200 discriminable steps), spectral, and temporal tuning, provide extremely robust speech recognition under noisy backgrounds and at speech levels from 40 to 130 dB SPL (e.g., Pollack and Pickett, 1958; Borg & Zakrisson, 1973; Studebaker et al., 1999). While the overall intensity difference due to individual talkers and the distance from the talker to the listener can be compensated for by an automatic-gain-control (AGC) circuit, the relative intensity changes from soft consonants to loud vowels must be preserved perceptually for a cochlear implant listener to understand speech. There has not been any consensus or experimental validation on how much of this acoustic range should be converted into perceptual dynamic range in a cochlear implant user.

A speech dynamic range of 30 dB is widely assumed, based on the classic acoustic analysis by Fletcher (1953) and other earlier statistical measurements on conversational speech (Dunn and White, 1940). This 30-dB dynamic range has formed the basis for many applications including the Articulation Index (ANSI, 1969; 1997) and for the fixed input dynamic range in the Nucleus device. However, modern analysis using digital signal processing has shown a much greater speech dynamic range than this classic 30-dB range. Boothroyd et al. (1994) performed one-third octave analyses of 7 phonemes produced by 5 female and 5 male talkers. They found that the overall dynamic range in these data was 53 dB, and that the dynamic range still remained at 37 dB even after adjustment in overall levels and high-frequency pre-emphasis. Eddington et al. (1999) also studied distribution of envelope levels over 6 frequency channels for the TIMIT sentences presented at a conversational level. They found that the distribution of speech envelope levels was in the range of 40-60 dB, much wider than the presumed 30-dB range. For this reason, Eddington et al. used a 60-dB acoustic range in their implementation of the MIT/GWP CIS processor. Studebaker et al. (1999) measured word recognition at high speech and noise levels and concluded from these perceptual studies that the effective dynamic range of speech must be greater than the commonly assumed value of 30 dB.

In our speech recognition tests, we typically use 20 medial consonants in a/C/a format (Shannon et al., 1999) and 12 medial vowels in h/V/d format (Hillenbrand et al., 1995). These speech tokens were produced by 5 female and 5 male talkers. The vowel overall levels were normalized based on the maximal rms level measured within a 50-ms window. Figure 1 shows distribution of envelope levels for these a/C/a and h/V/d tokens in the wide-band condition (top panel) and for the a/C/a tokens (middle panel) and the h/V/d tokens (bottom panel) in the 8-channel condition. Because of the noise floor on the bottom of the distribution, we conservatively define the speech dynamic range as the difference in the envelope levels producing between 5% and 99% accumulative occurrences. For the single-channel, wide-band condition (top panel), the acoustic dynamic range is 47 dB (from –51 to –4 dB) for consonants and 46 dB (from –50 to –4 dB) for vowels. In the 8-channel condition, the consonant dynamic range is 41, 52, 51, 50, 47, 46, 47, and 45 dB for channel 1, 2, 3, 4, 5, 6, 7, and 8, respectively. On the other hand, the vowel dynamic range is 51, 51, 53, 49, 47, 47, 42, and 36 dB for channel 1, 2, 3, 4, 5, 6, 7, and 8, respectively. Given these acoustic dynamic ranges, we shall see whether an input dynamic



cochlear implant users.

Figure 1. Speech dynamic ranges (or envelope level distributions) for the wide-band condition (top panel) and the 8-channel condition (for consonants see the middle panel and for vowels see the bottom panel).

Electric Dynamic Range

We have conducted extensive studies to characterize electric dynamic range with different stimuli and under various conditions including sinusoidal and pulsatile stimuli, stimulus frequency, pulse amplitude and duration, monopolar and bipolar stimulation. Our results show that the electric dynamic range is generally less than 30 dB, much narrower than the 120-dB normal acoustic range. We have developed a theoretical model to account for the narrow electric dynamic range (mostly due to the loss of cochlear compression) and to encode loudness in general (Zeng and Shannon, 1994; 1998).

Experiment 1: Dynamic range as a function of stimulus frequency

Figure 2 presents dynamic range data as a function of frequency for both sinusoidal (left-slanted, hatched areas) and pulsatile (right-slanted, hatched areas) stimuli in 8 cochlear implant users of the Ineraid device. The lower boundary of the dynamic range was electric thresholds and the upper boundary was the maximum acceptable loudness (MAL). All stimuli were 200 ms in duration. The pulsatile stimuli consisted of biphasic pulse trains of 100 us/phase. The most apical electrode was stimulated in a monopolar mode (for detailed methods see Zeng and Shannon, 1994, 1999).

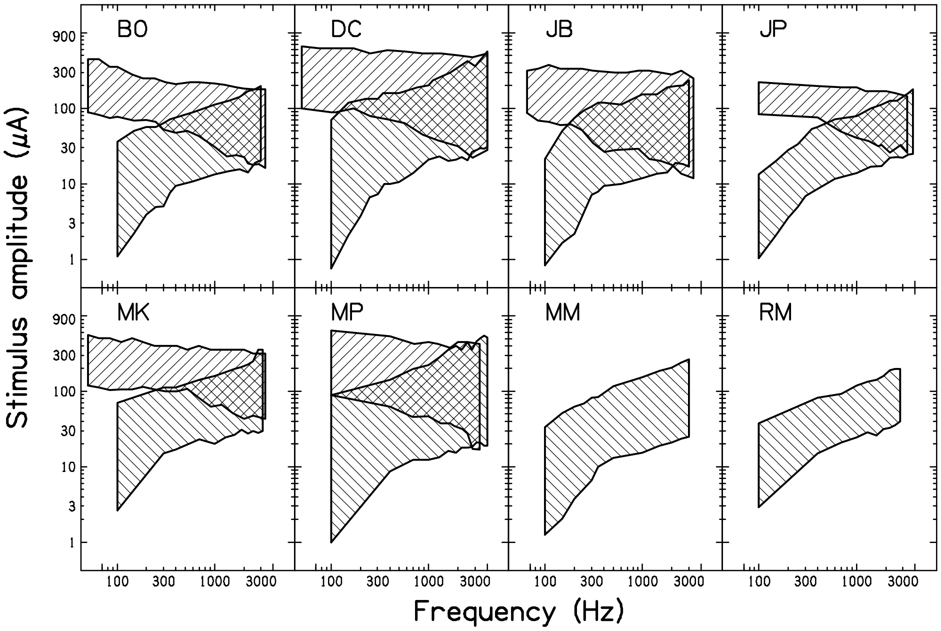


Figure 2. Electric dynamic range (Maximum acceptable loudness – threshold) as a function of frequency for sinusoid (left-slanted) and pulse (right-slanted).

For sinusoids, both thresholds and MALs increase monotonically as a function of frequency, whereas for pulses, both decrease monotonically as a function of frequency. The 100-Hz sinusoid produced the lowest threshold and the widest dynamic range (mean = 30 dB), while the 100-Hz pulse produced the highest threshold and the narrowest dynamic range (mean = 14 dB). The sinusoidal dynamic range decreases with frequency until 300-500 Hz as a result of steeper increase in thresholds than in MALs. On the other hand, the pulsatile dynamic range increases with frequency as a result of steeper decrease in thresholds than MALs. At 1000 Hz, there is no statistical difference in dynamic range between sinusoidal (19 dB) and pulsatile (18 dB) stimuli (P>0.5).

Experiment 2: Dynamic range as a function of pulse duration

Another important parameter in electric hearing is pulse duration, which has been used in combination with pulse amplitude to encode loudness (e.g., Nucleus device). Earlier researchers hypothesized that pulse amplitude can be traded for pulse duration to maintain equal loudness, as long as equal charge is maintained (Clark et al., 1987). However, more recent studies have suggested that this “equal-charge, equal-loudness” hypothesis is incorrect (Shannon, 1985; Pfingst et al., 1991; Zeng et al., 1998; McKay and McDermott, 1995, 1999).

Figure 3 shows thresholds (solid symbols) and maximum acceptable loudness (open symbols) in microamperes (y-axis) as a function of pulse duration (x-axis) in 4 Nucleus users (different symbols). Four bipolar electrode pairs, (1,3), (9,11), (20,22), and (1,22), were selected in an attempt to stimulate the basal, middle, apical, and the entire region of the cochlea, respectively. The top 4 panels show data for the 100-Hz pulse train with pulse duration varied from 10 to 4000 us/phase. The bottom 4 panels show data for the 1000-Hz pulse train with pulse duration varied from 10 to 450 us/phase (for detailed methods, see Zeng et al., 1998). Linear regression was performed for each data set on a log-log scale and is plotted as solid lines in each panel. If the “equal-charge, equal-loudness” hypothesis were true, all regression lines would have a slope of –1 (the dotted diagonal line).

The most notable result is that neither the threshold function nor the maximum loudness function abides by the “equal-charge, equal-loudness” rule. The average slope for the threshold function was –0.73 and the average slope for the maximum loudness function was –0.45, both significantly shallower than the –1 slope predicted by the “equal-charge, equal-loudness” hypothesis. Another notable result is the much lower pulse amplitude that was required to reach threshold and maximum loudness for electrodes (1,22) than that for other three electrode configurations. Presumably, the much lower threshold and maximum loudness values are due to the widely-spaced electrodes which may function like two monopolar electrodes and excite a much broader region of cochlea (i.e., more neurons, see van der Honert and Stypulkowski, 1987).

Finally, because of the steeper slope of the threshold function than the maximum loudness function, the electric dynamic range increases monotonically as a function of pulse duration. The dynamic range increased from 7-8 dB for the pulse duration of 20 us to 25-28 dB for pulse duration of 1000-2000 us. Despite of the overall lower threshold and maximum loudness levels for electrodes (1,22), their dynamic range was comparable to that produced by other three much more closely spaced bipolar electrodes.

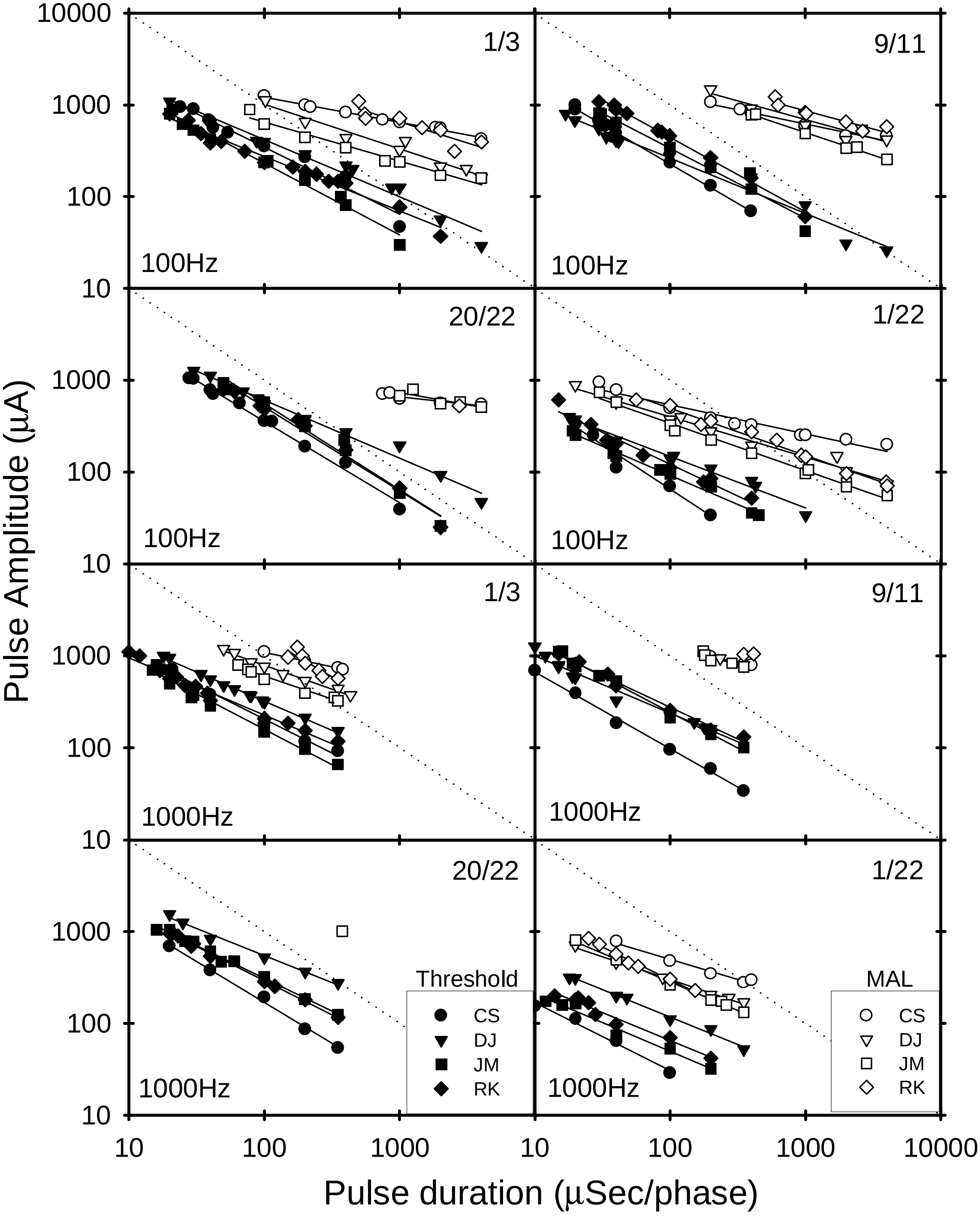


Figure 3. Trade-off between pulse duration (x-axis) and pulse amplitude (y-axis) in 4 Nucleus implant users. Thresholds are shown in solid symbols and maximum acceptable loudness in open symbols. The dotted diagonal line represents the slope for which the "equal-charge, equal-loudness" hypothesis holds.

Summary and Significance

Based on the above extensive data in thresholds and maximum loudness levels, we can draw the following conclusions. (1) For sinusoids, the dynamic range decreases monotonically as a function of frequency with the greatest value of 30 dB at 100 Hz. (2) For pulses with a 100-us/phase pulse duration, the dynamic range increases monotonically as a function of pulse frequency with the greatest value of 25 dB at 3000 Hz. (3) For pulses of 100-Hz and 1000-Hz rates, the dynamic range increases monotonically as a function of pulse phase duration with the greatest values of 20-28 dB at the maximally allowed pulse durations. We note that for conditions under which most speech processors operate, the dynamic range is more likely to be in the range of 5-20 dB, much smaller than the maximum value of 30 dB (e.g., Skinner et al., 1995; Zeng and Galvin, 1999).

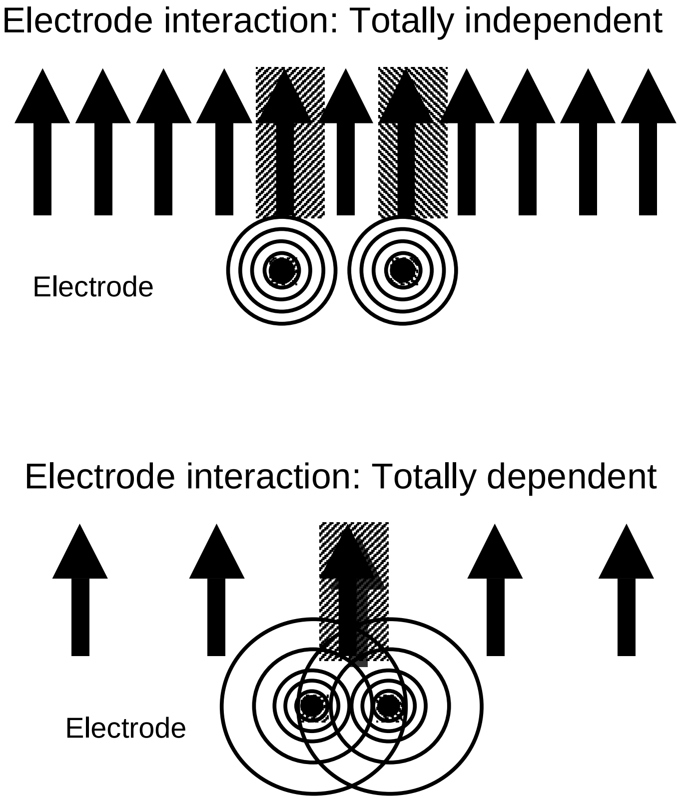
Electrode Interactions

As multi-electrode cochlear implants have become the modern standard, rarely do implant users hear individual electrode stimulation in isolation under realistic listening situations. When listening to wide-band sounds such as speech and music, multiple electrode pairs are stimulated either simultaneously (CA and SAS), in interleaved sequences (CIS and SPEAK), or both (PPS). At present, we do not know how electrodes interact to affect loudness and electric dynamic range under multiple electrode stimulation, let alone any systematic applications to the clinical fitting of speech processors. Here we present a theoretic model of electrode interactions and show that electrode interactions can greatly affect loudness and dynamic range depending on whether these electrodes stimulate independent neural channels or not. In the next section, we will present preliminary data showing that these electrode interactions need to be taken into account in order to optimally fit speech processors for cochlear implant users.

Concepts

When an array of multiple electrodes is inserted into the cochlea, the individual electrodes may or may not stimulate different populations of neurons, depending on the insertion depth, the distance between electrodes and neurons, the degree and pattern of nerve survival, and other electrical properties in the cochlea. If they stimulate different groups of neurons, they are termed as independent electrodes, or if they stimulate the same population of neurons, they are termed as dependent electrodes.

Figure 4 presents two extreme cases of electrode interactions. The top panel depicts one extreme case of totally independent electrodes as the two electrodes stimulate two independent populations of neurons. The bottom panel depicts the other extreme case of totally dependent electrodes, in which two electrodes produce directly electrical field summation and function effectively as a single neural channel.



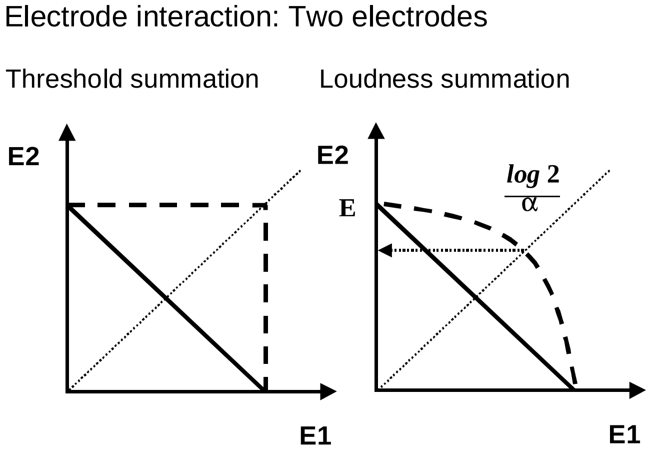
Electrode interaction: Totally independent

Electrode interaction: Totally dependent

Figure 4. Schematic representation of two totally independent electrodes (top panel) and two totally dependent electrodes (bottom panel).

A Model of Electrode Interactions and Data : 2-electrode case

Depending on whether two electrodes are dependent or not, they may produce totally different perceptual effects on electric dynamic range. To illustrate the different effects of electrode interactions on threshold and loudness, only changes in amplitude of otherwise two identical stimuli are allowed on the two electrodes. At the threshold level (left panel in Figure 5), if the electrodes are totally dependent, then perceptual threshold is reached as long as the sum of two electric currents exceed a certain level (E). We would predict a -45^{o} diagonal line for the threshold function in two-electrode simultaneous stimulation.



Electrode interaction: Two electrodes

Threshold summation

Loudness summation

Figure 5. A model of electrode interactions at the threshold level (left panel) and suprathreshold level (right panel). Solid lines represent the case of two totally dependent electrodes and dashed lines represent the case of two totally independent electrodes.

On the other hand, if the two electrodes were totally independent, then the presence of a subthreshold current on one electrode would not affect the detection of the current on the other electrode. Thus, the current should reach the threshold level on either of the two electrodes independently (dashed horizontal and vertical lines).

If the two electrodes are totally dependent, then we still have the same simple electrical field summation at the suprathreshold loudness level (the -45^{o} solid line on the right panel in Figure 5). However, if the two electrodes were totally independent, with each producing a loudness percept, then the total loudness would be doubled. To maintain the same maximum loudness, current amplitude on each electrode has to be reduced. We can derive this reduction based on our exponential loudness model in electric stimulation (Zeng and Shannon, 1992, 1994). Assuming that maximum loudness (L) is reached for current amplitude (E) on either electrode alone, that is:

$$L=k\,e^{\alpha E} \qquad (1)$$

where k and α are constants. In two-electrode simultaneous stimulation, the amplitude has to be reduced to E' (right panel in Figure 5) on each electrode to reach the same maximum loudness (L):

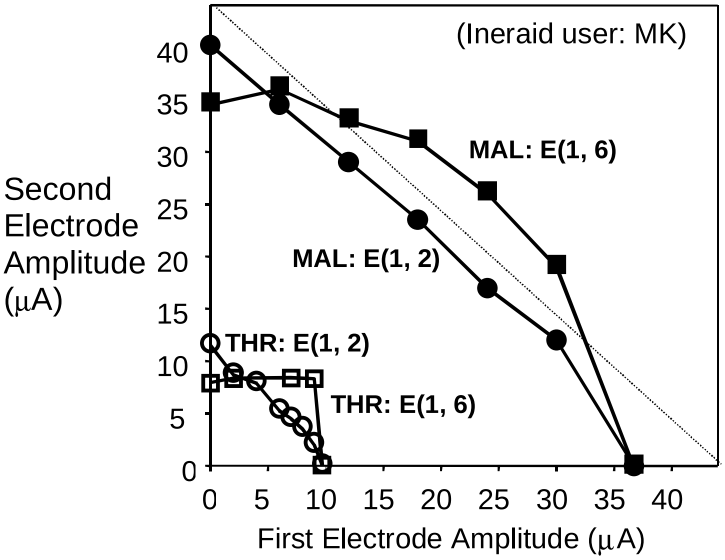
$$L=L_1+L_2= k\,e^{\alpha E'}+ k\,e^{\alpha E'} \quad =2k\,e^{\alpha E'} \qquad (2)$$

Combining Equations (1) and (2), we can derive:

$$E'=E-\log 2/\alpha \qquad (3)$$

Figure 6 shows preliminary data from one Ineraid user (Fu, Shannon, and Zeng, 1996). When the two electrodes were close to each other (1, 2), they showed a pattern of two totally dependent electrodes at both threshold (open circles) and comfortable loudness (solid circles) levels. On the other hand, when the two electrodes were far apart (1,6), they showed a pattern of totally independent electrodes.

Figure 6.



A Model of Electrode Interactions: N-electrode case

The 2-electrode model can be extended to N electrodes. Figure 7 shows how dynamic range can be differentially affected by the dependence (left panels) or independence (right panels) in electrode interactions. The top-left panel shows identical amplitude reduction (E') for both threshold (solid line) and loudness (dashed line) as a function of the number of electrodes (N) in the case of totally dependent electrodes. The top-right panel shows no amplitude reduction in threshold (solid line) and an amplitude reduction as a logarithmic function of the number of electrode (N).

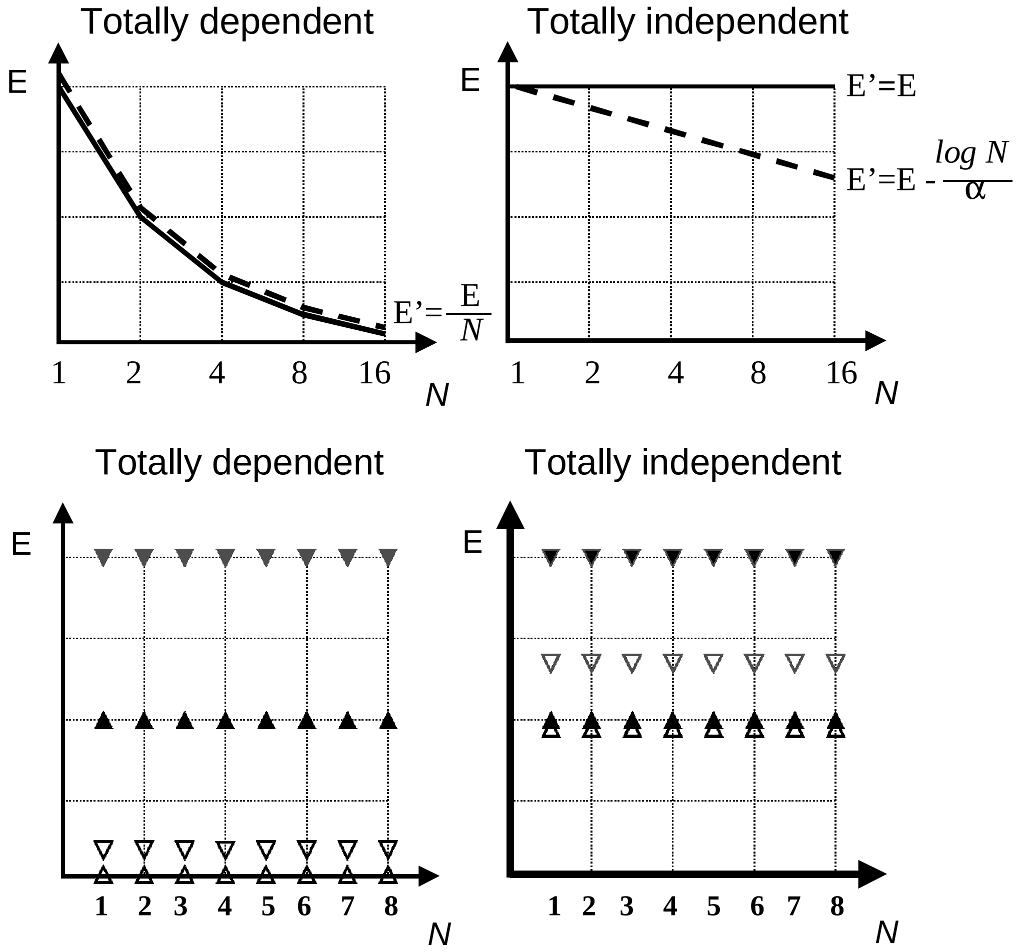


Figure 7. Amplitude reduction on each electrode as a function of the number of electrode (N) for totally dependent electrodes (top-left panel) and for totally independent electrodes (top-right panel). Dynamic range alterations for 8 totally dependent electrodes (bottom-left panel) and for 8 totally independent electrodes (bottom-right panel). Solid triangles represent thresholds and solid inverse triangles represent maximum loudness measured when each electrode was stimulated separately, whereas the open triangles represent thresholds and open inverse triangles represent maximum loudness when all 8 electrodes are stimulated simultaneously.

The bottom two panels in Figure 7 show how dynamic range is altered in the case of 8 totally dependent electrodes (bottom-left panel) and of 8 totally independent electrodes (bottom-right panel). If the eight electrodes are totally dependent, then only 1/8th of the original current amplitude is required to reach both threshold and maximum loudness when all 8 electrodes are simultaneously stimulated. In this case, the eight electrodes function effectively as a single-electrode implant. However, if the eight electrodes are totally independent, then the threshold stays at the same value (for clarity of demonstration, they were shifted downward slightly) while the maximum loudness was reduced by the amount of $\log 8/\alpha$. The present model presents two extreme forms of electrode interactions. In reality, electrode interactions are likely to be somewhere between. We shall examine such a case in Exp. 3 in the next section.

Summary and Significance

Electrode interactions can be in the form of direct electrical field summation if the electrodes are totally dependent, or in the form of loudness summation if the electrodes are totally independent. Electrode interactions are different between threshold and suprathreshold measures. The present model suggests that a significant decrease in thresholds indicates greater dependence among electrodes. Electrode interactions significantly affect speech processor fittings in that dependent electrodes reduce the number of effective perceptual channels while independent electrodes reduce the electric dynamic range. These changes in the number of effective channels and in electric dynamic range should be taken into account in the speech processor fitting.

Dynamic range and Speech Recognition

In this section, we present experimental results that show how acoustic dynamic range, electric dynamic range, and electrode interaction affect cochlear implant users' performance in speech recognition. We recruited Clarion users to study the effects of acoustic input dynamic range on speech recognition because the Clarion device allows a wide range of selection of input dynamic ranges. We recruited both Nucleus and Clarion users to study the effects of electric dynamic range on speech recognition. Finally, we studied the effect of electrode interaction on speech recognition in Clarion users, particularly those using the SAS strategy. Speech materials included 20 consonants (a/C/a) and 12 vowels (h/V/d), produced by 5 female and 5 male talkers. Speech was delivered through a speaker (TANNOY Reveal CE) producing a 65 dB SPL overall level in the head area when the subject was seated. The general methods of conducting speech recognition experiments can be found elsewhere (Zeng and Galvin, 1999).

Experiment 1: Effects of Acoustic Dynamic Range

Objective. To study whether there is an optimal setting of the acoustic dynamic range in speech processors for cochlear implant users.

Methods : Seven Clarion users (4 CIS and 3 SAS) participated in this experiment. The input dynamic range in speech processors was systematically varied from 10 to 70 dB in CIS users and 20 to 80 dB in SAS users (Figure 8).

Other parameters including the sensitivity control, T levels and M levels were left unchanged in the original processor setting.

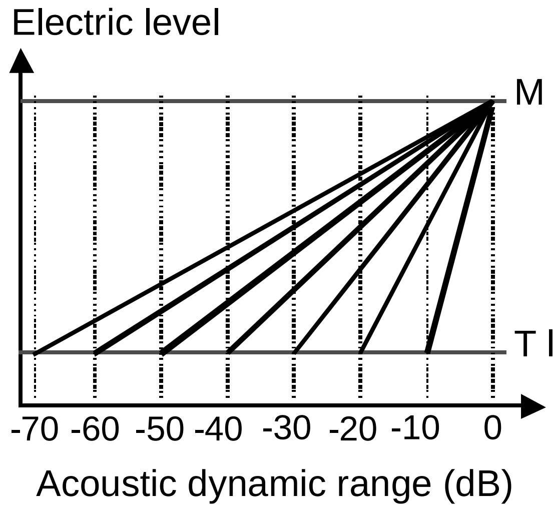


Figure 8. Selection of input acoustic dynamic range in Clarion devices.

Results. Figure 9 shows individual recognition scores (y-axis) as a function of the acoustic dynamic range (x-axis). Despite great individual variability, there appears to be an optimal acoustic dynamic range of 50 dB for both consonant and vowel recognition. Speech recognition either stays unchanged or decreases slightly for dynamic range wider than 50 dB, but it decreases sharply for dynamic ranges narrower than 30 dB. The largest decrease was observed for subject, JM, whose consonant scores dropped from 80% with the 50-dB acoustic dynamic range to 55% with the 30-dB range and to 10% with the 10-dB range.

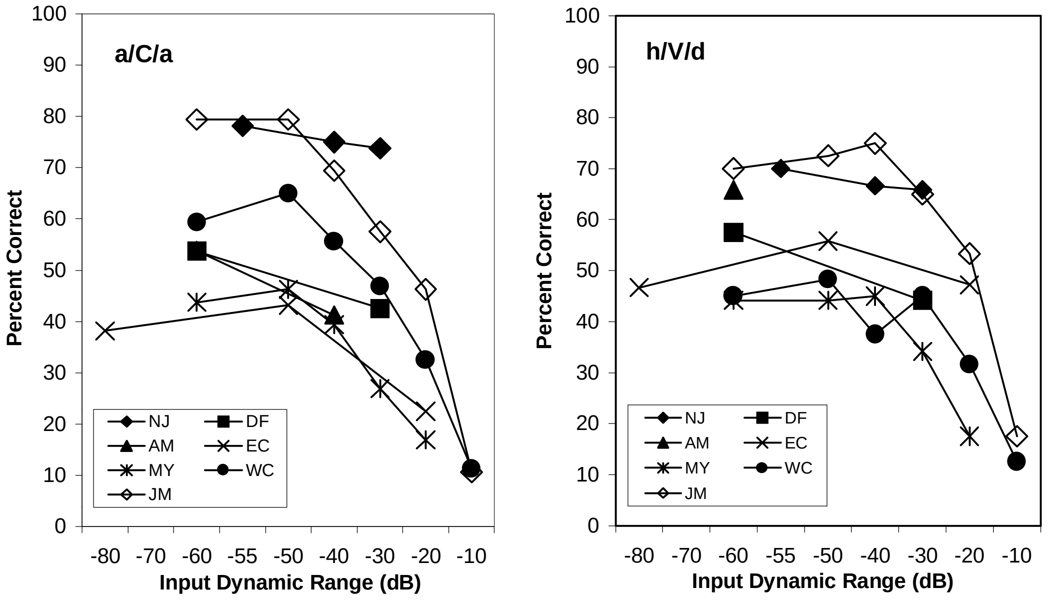


Figure 9. Effects of input acoustic dynamic range on recognition of consonants (left panel) and vowels (right panel).

Significance. Our data show that, at least for the multiple talker materials tested here, the optimal acoustic dynamic range is about 50 dB. This value, not too surprisingly, is consistent with the dynamic range of speech materials used in the experiment (see the "Acoustic Dynamic Range" section before). We note, however, that this optimal value is observed for speech recognition in quiet. It remains to be seen whether it is still optimal under noise conditions.

Experiment 2: Effects of Electric Dynamic Range

Objective. To study how changes in electric dynamic range affect speech recognition in cochlear implant users, and whether these effects, if any, are dependent on speech processing strategies.

Methods. Four Nucleus users and three Clarion users participated in this experiment. For Nucleus users, the electric dynamic range was reduced to 25% of the original by increasing the T level by 75% ("75% T" condition) and to a binary value by a combination of increasing the T level by 75% and decreasing the C level by 24% (for details, see Zeng and Galvin, 1999). For Clarion users, the electric dynamic range was halved by either decreasing the M level by 50% ("50% M") or by increasing the T level by 50% ("50% T").

Results. Figure 10 shows effects of reducing electric dynamic range on vowel and consonant recognition in 4 Nucleus users of the SPEAK strategy.

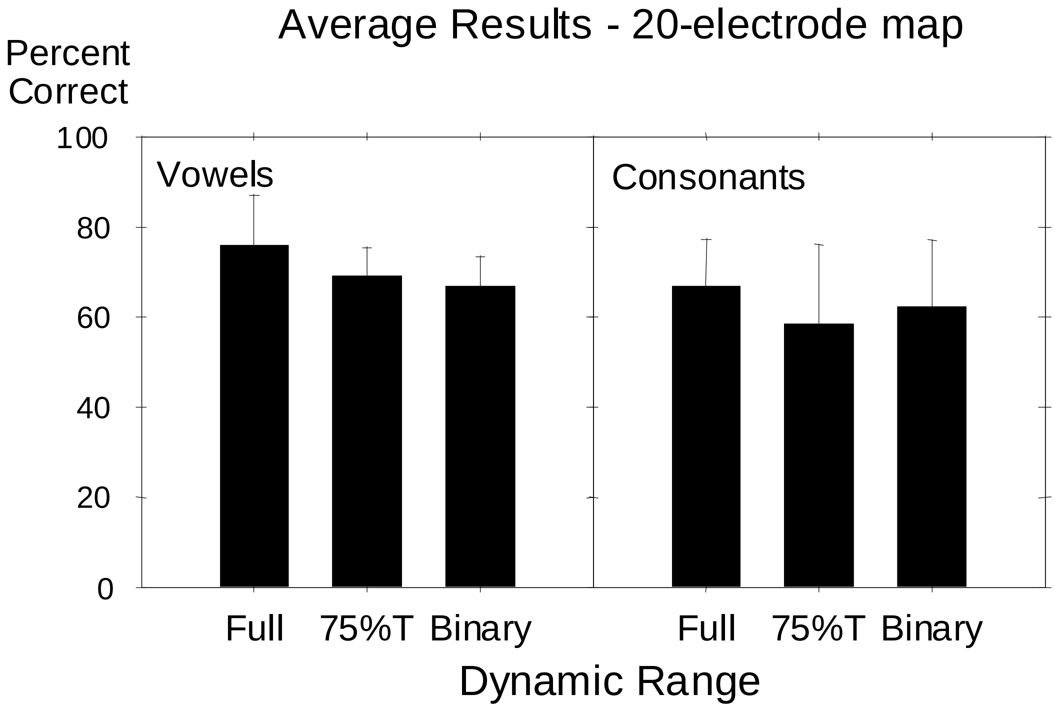


Figure 10. Effects of reducing electric dynamic range on recognition of vowels (left panel) and consonants (right panel) in Nucleus users.

The surprising results are that speech recognition was not significantly affected even when the electric dynamic range was reduced to a binary representation of the acoustic amplitude. We interpret this finding as a result of robust representation of spectral cues for speech recognition employed by the SPEAK strategy.

On the other hand, when the electric dynamic range was halved, a significant decrease in speech recognition was observed for all Clarion users, except for subject, DF, whose consonant scores were unchanged (Figure 11). The decrease in speech recognition due to reduction in electric dynamic range probably reflects the dependence on temporal cues for Clarion CIS and SAS users. We intend to collect data in additional Clarion users to further establish the relationship between electric dynamic range and speech recognition.

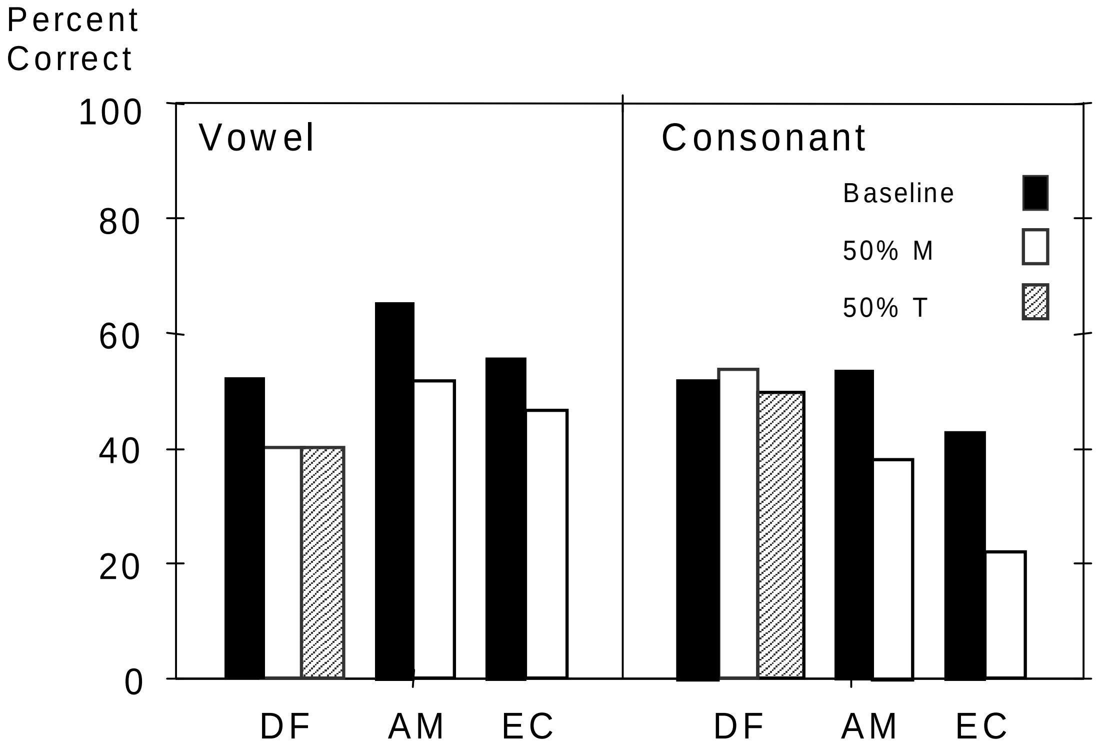


Figure 11. Effects of reducing electric dynamic range on recognition of vowels (left panel) and consonants (right panel) in Clarion users.

Significance. Our preliminary data suggest that the effects of electric dynamic range on speech recognition depend on speech processing strategies. Because changes in electric dynamic range affect mostly the representation of temporal cues, they adversely affect speech recognition for the CIS and SAS users more than for the SPEAK users.

Experiment 3: Effects of electrode interactions

Objective. As we showed earlier, due to direct electric field summation, loudness summation, or both, thresholds and loudness levels could be drastically different between single-electrode and multi-electrode stimulations. There are no established standards on how to take these electrode interactions into account in the clinical fitting of speech processors. For example, the existing manufacturers' guidelines generally recommend lowering the overall volume or setting the T levels to some minimal values to overcome the electrode interaction effect. These guidelines cannot guarantee an optimal fitting in which the softest sounds are audible while the loud sounds are still comfortable. Our long-term objective is to develop such an optimal fitting procedure for multi-electrode cochlear implants. While more experiments are needed to form a solid basis for the optimal fitting, we report some preliminary data here to show that it is possible to fit speech processors with speech sounds and that such a fitting procedure can produce improved speech recognition.

Methods. Two Clarion users, one CIS user (DF) and one SAS user (AM), participated in this experiment. Figure 12 shows how dynamic range changed when using speech sounds to fit the processor in subject, AM. We first performed the standard clinical fitting procedure to establish the T and M levels on each individual electrode (open symbols in Figure 12). Then we set the T level equal to M level in the clinical fitting system and, by doing so, we fixed the current-level contour across the electrodes. We changed the overall volume control to move this current-level contour up and down while using live voice to ask the subject to report at what level the speech was first audible ("Speech-adjusted T level") and was maximally comfortable ("Speech-adjusted M level").

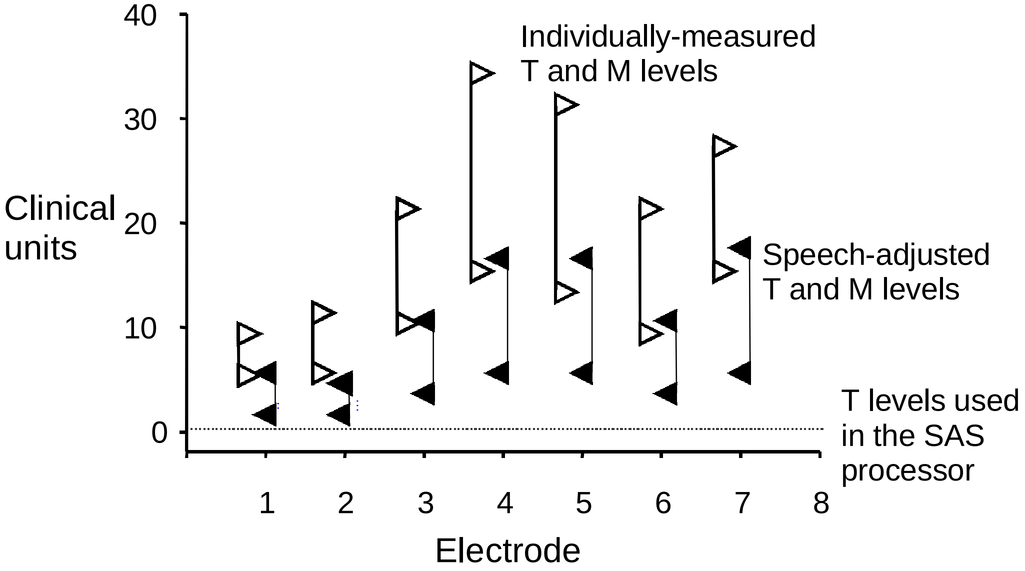


Figure 12. Changes in electric dynamic ranges due to electrode interactions. Open symbols represent original T and M levels measured for individual electrodes in isolation. Solid symbols represent modified T and M levels measured using live speech as calibration signal. The dashed line represents a suggested default setting of the T levels in the SAS processor.

Figure 12 show that the T and M levels were greatly reduced with the new speech-adjusted procedure (solid symbols). These results suggest that speech-adjusted M-levels are needed to ensure that stimulation not be overly loud when all eight electrodes are stimulating. However, when fitting a patient with the SAS speech processor, clinicians are recommended to use a low current value (e.g., 3, the dotted line in Figure 12), rather than the individually measured T levels. In this case, weak speech sounds may be too soft, if not inaudible, as they are mapped to below speech-adjusted threshold levels. If our speech-adjusted T and M levels represent the realistic listening situations for implant users, then the recommended standard fitting procedure for SAS processors would provide a sub-optimal mapping between the acoustic and electric dynamic ranges.

Results. Figure 13 shows vowel and consonant recognition with the standard clinical procedure (hatched bars) and the speech-adjusted fitting procedure (solid bars) for two Clarion users with two input dynamic range settings. Improvement in speech recognition scores ranged from 2 (vowels, AM60) to 20 (consonants, DF30) percentage points when the new speech-adjusted procedure was used instead of the standard fitting procedure.

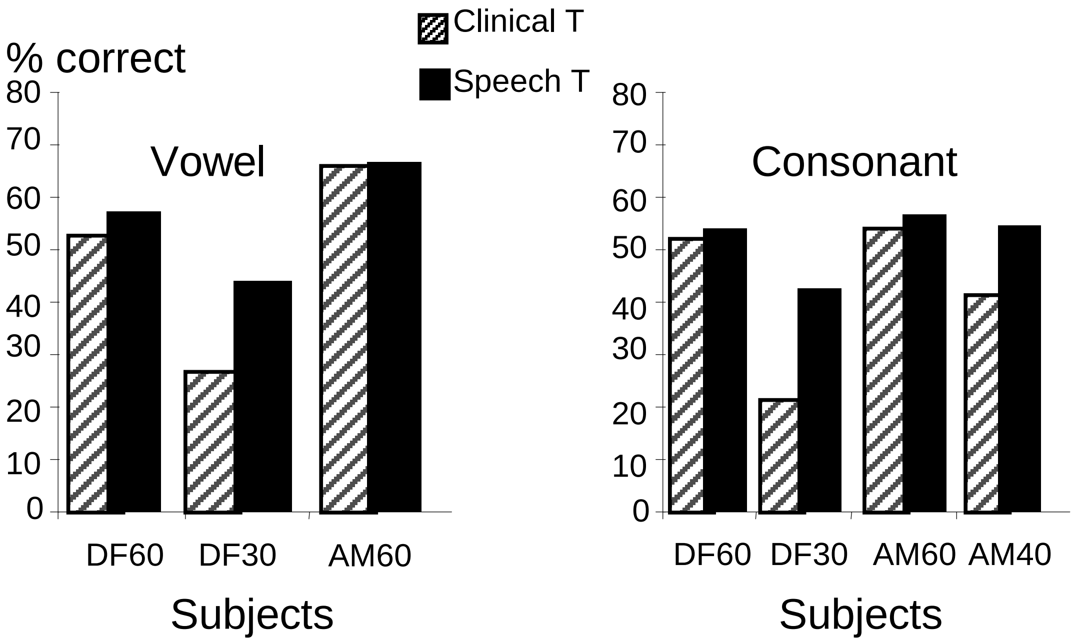


Figure 13. Vowel (left panel) and consonant (right panel) recognition with clinically adjusted T and M levels (hatched bars) and with speech-adjusted T and M levels (filled bars).

Significance. The present preliminary data suggest that we should consider electrode interactions in fitting speech processors and also develop new fitting procedures that are appropriate for speech and other environmental sounds under realistic listening conditions. Fitting speech processors with speech sounds can be extremely important for children who use a cochlear implant, as speech is the most frequently and naturally present sound under realistic listening situations.

Future Directions: Fitting Implants under Realistic Listening Conditions

We have explored the effects of acoustic dynamic range, electric dynamic range, and electrode interactions on speech recognition in cochlear implant listeners. In the future, we intend to optimize acoustic dynamic range not only for speech sounds but music and other environmental sounds. We also intend to optimize electric dynamic range in terms of reducing unwanted electrode interactions and taking the loudness perception of dynamic stimuli into account (Zeng and Shannon, 1995; Zhang and Zeng, 1997). Based on our preliminary data, we believe that we should fit cochlear implants with wide-band, dynamic signals that mimic realistic listening conditions.

Plans for the Next Quarter:

Hardware – Nucleus Interface. We have developed a Motorola 56309 interface to the Nucleus 22 and 24 implant devices. The interface works presently in SEMA mode and programs are being developed to allow access to the newer and faster Embedded Mode. Hardware and software development will continue on this interface.

Hardware – Clarion Research Interface. We are presently developing a research interface for the new generation of Clarion cochlear implants under contract from Advanced Bionics Corp. We will report on the progress on this interface in future quarters as it is relevant to this contract.

Hardware – Portable Processors. We will design a custom board to take signal from the Motorola 56309 board to the rf transmission stages of both Nucleus and Clarion devices. The new board will be designed to be relatively compact and power-efficient so that it can be packaged for portable use.

Experiments – Psychophysics. We will continue to collect forward masking patterns an evaluate the psychophysical excitation patterns of electrical stimuli that vary in current level, pulse phase duration, and electrode stimulation mode. We will relate these measures to measures of loudness with the same parameters.

Experiments – Speech Processor Design. (1) Continue data collection on “holes in hearing” experiment with normal-hearing listeners, (2) complete data collection on “noise and number of channels” study with Nucleus-24 listeners and normal-hearing listeners, and (3) collect data on the effects of stimulation rate on speech recognition with Clarion and Nucleus-24 listeners.

Publications and Presentations in this Quarter:

Friesen, L.M., Shannon, R.V., and Slattery, W.H. (1999). Effects of frequency allocation on phoneme recognition with the Nucleus 22 cochlear implant, Amer. J. Otol., 20(6), 729-734.

Fu, Q.-J. and Shannon, R.V. (1999). Effect of acoustic dynamic range on phoneme recognition in cochlear implant listeners, Journal of the Acoustical Society of America (Acoustic Research Letters Online), 106(6), L65-L70.

Shannon, R.V., Jensvold, A., Padilla, M., Robert, M., and Wang, X. (1999). Consonant recordings for speech testing, Journal of the Acoustical Society of America (ARLO), 106(6), L71-L74.

Shannon, R.V. and Zeng, F.-G. (1999). NIH Neural Prostheses Workshop, Bethesda, MD, Oct 12, 1999.

Shannon, R.V. (1999). Critical cues for auditory pattern recognition of speech: Implications for cochlear implant speech processor design, Department of Cognitive Sciences, UC Irvine, October 6. (Invited talk)

Shannon, R.V. (1999). The relative importance of amplitude, temporal, and spectral cues in speech recognition, USC Biomedical Engineering Dept., Oct 26. (Invited talk)

Shannon, R.V. (1999). Critical cues for auditory pattern recognition of speech: Implications for cochlear implant speech processor design, Kresge Hearing Research Institute, University of Michigan, Ann Arbor, 3 Nov. (Invited talk)

Shannon, R.V. (1999). The relative importance of amplitude, temporal, and spectral cues in speech recognition, Association of Students in Biomedical Engineering, USC, Dec 1. (Invited talk)

Shannon, R.V. (1999). Critical cues for speech recognition, LA Chapter meeting of the Acoustical Society of America, Dec 14, Los Angeles. (Invited talk)

Wei, C., Cao, K., Wang, Z, and Zeng, F.-G. (1999). Rate discrimination and tone recognition in Mandarin-speaking cochlear-implant listeners. Chinese Journal of Otorhinolaryngology 34(2), 82-88.

Zeng, F.-G., and Shannon, R.V. (1999). Psychophysical laws revealed by electric hearing. NeuroReport 10(9), 1931-1935.

Zeng, F.-G. (1999). Auditory Information Processing: What have we learned from cochlear implants, Neuroscience and Cognitive Science Seminar, University of Maryland, College Park, Oct 8, 1999 (Invited talk).

Zeng, F.-G. (1999). Advances in Cochlear Implants, Aural Rehabilitation and Assistant Listening Devices Workshop, The NWL Foundation for the Hearing Impaired, Taipei, Oct. 21, 1999 (Invited talks).

Zeng, F.-G. (1999). Psychophysics and Cochlear Implants, National Normal University, Kao-Hsiung, Oct. 25, 1999 (Invited talks).

References

Boothroyd, A., Erickson, F.N., and Medwetsky, L. (1994). “The hearing aid input: a phonemic approach to assessing the spectral distribution of speech,” Ear & Hearing, 15, 432-442.

Borg, E., Zakrisson, J.E. (1973). Stapedius reflex and speech features. J Acoust Soc Am. 54(2), 525-527.

Dunn, H.K., and White, S.D. (1940). Statistical measurements on conversational speech. J. Acoust Soc Am. 11, 278-288.

Eddington, D. (1999). Speech Processors for Auditory Prostheses, NIDCD Contract N01-DC-6-2100, Final Progress Report.

Fletcher, H. (1953). Speech and Hearing in Communication, New York: Krieger.

Fu, Q.J., Shannon, R.V. (1999). Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing. J Acoust Soc Am. 105(3), 1889-1900.

Fu, Q.-J., Shannon, R.V., Zeng, F.-G., and Chatterjee, M. (1996). Electrode interactions measured by loudness summation in cochlear implant listeners, Journal of the Acoustical Society of America, 100(4), 2631.

Hillenbrand, J., Getty, L., Clark, M., and Wheeler, K. (1995). Acoustic characteristics of American English vowels, J. Acoust. Soc. Am., 97, 3099-3111.

Kessler, D.K., Osberger, M.J., Boyle, P. (1997). CLARION patient performance: an update on the adult and children's clinical trials. Scand Audiol Suppl. 47, 45-49.

McKay, C.M., McDermott, H.J., and Clark, G.M. (1995). “Loudness summation for two channels of stimulation in cochlear implants: effects of spatial and temporal separation,” Annals Otol Rhinol Laryngol, 104 (9), Suppl 166, pp. 230-233.

McKay and McDermott (1999). The perceptual effects of current pulse duration in electrical stimulation of the auditory nerve, J. Acoust Soc Am., 106(2), 998-1009.

Pfingst, B.E., De Haan, D.R., and Holloway, L.A. (1991). Stimulus features affecting psychophysical detection thresholds for electrical stimulation of the cochlea. I: Phase duration and stimulus duration, J. Acoust. Soc. Am., 90, 1857-1866.

Pollack, I., Pichett, J.M. (1958). Masking of speech by noise at high sound levels. J Acoust Soc Am. 30, 127-130.

Shannon, R.V. (1985). Loudness summation as a measure of channel interaction in a multichannel cochlear implant. In Cochlear Implants, R.A. Schindler and M.M. Merzenich, Eds., Raven Press, New York.

Shannon, R.V., Jensvold, A., Padilla, M., Robert, M., and Wang, X. (1999). Consonant recordings for speech testing, Journal of the Acoustical Society of America (ARLO), 106(6), L71-L74.

Skinner, M.W., Clark, G.M., Whitford, L.A., et 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, M.W., Holden, L.K., Holden, T.A., Demorest, M,E. (1995). Comparison of procedures for obtaining thresholds and maximum acceptable loudness levels with the nucleus cochlear implant system. J Speech Hear Res. 38(3), 677-689.

Studebaker, GA, Sherbecoe, RL, McDaniel, DM, Gwaltney, CA. (1999). Monosyllabic word recognition at higher-than-normal speech and noise levels. J Acoust Soc Am. 105(4), 2431-2444.

Zeng, F.-G., and Shannon, R.V. (1992). Loudness balance between electric and acoustic stimulation. Hearing Research 60, 231-235.
Zeng, F.-G., and Shannon, R.V. (1994). Loudness-coding mechanisms inferred from electric stimulation of the human auditory system. Science 264, 564-566.
Zeng, F.-G., and Shannon, R.V. (1995). Loudness of simple and complex stimuli in electric hearing. Annals of Otology, Rhinology and Laryngology 104 (suppl. 166), 235-238.
Zeng, F.-G., J.J. Galvin, and C.Y. Zhang (1998). Encoding loudness by electric stimulation of the auditory nerve. NeuroReport 9, 1845-1848.
Zeng, F.-G., and Galvin, J.J. (1999). Amplitude compression and phoneme recognition in cochlear implant listeners. Ear and Hearing 20, 60-74.
Zeng, F.-G., and Shannon, R.V. (1999). Psychophysical laws revealed by electric hearing. NeuroReport 10, 1931-1935.
Zhang, C.Y., and Zeng, F.-G. (1997). Loudness of dynamic stimuli in acoustic and electric hearing. The Journal of Acoustical Society of America 102, 2925-2934.