Speech dynamic range and its effect on cochlear implant performance

Fan-Gang Zeng[a)]
Departments of Otolaryngology, Biomedical Engineering, and Cognitive Sciences, University of California, Irvine, California 92697

Ginger Grant[b)] and John Niparko
Department of Otolaryngology, Johns Hopkins University, Baltimore, Maryland 21287

John Galvin and Robert Shannon
House Ear Institute, Los Angeles, California 90057-1922

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

(Received 4 June 2001; accepted for publication 3 October 2001)

This study examines optimal conversions of speech sounds to audible electric currents in cochlear-implant listeners. The speech dynamic range was measured for 20 consonants and 12 vowels spoken by five female and five male talkers. Even when the maximal root-mean-square (rms) level was normalized for all phoneme tokens, both broadband and narrow-band acoustic analyses showed an approximately 50-dB distribution of speech envelope levels. Phoneme recognition was also obtained in ten CLARION implant users as a function of the input dynamic range from 10 to 80 dB in 10-dB steps. Acoustic amplitudes within a specified input dynamic range were logarithmically mapped into the 10–20-dB range of electric stimulation typically found in cochlear-implant users. Consistent with acoustic data, the perceptual data showed that a 50–60-dB input dynamic range produced optimal speech recognition in these implant users. The present results indicate that speech dynamic range is much greater than the commonly assumed 30-dB range. A new amplitude mapping strategy, based on envelope distribution differences between consonants and vowels, is proposed to optimize acoustic-to-electric mapping of speech sounds. This new strategy will use a logarithmic map for low-frequency channels and a more compressive map for high-frequency channels, and may improve overall speech recognition for cochlear-implant users.
[DOI: 10.1121/1.1423926]

PACS numbers: 43.71.Es, 43.71.Ky, 43.66.Ts, 43.64.Me [CWT]

I. INTRODUCTION

A major goal in designing speech processors for cochlear implants is to optimally convert speech signals to electric currents that fit within an implant user's perceptual range. In order to make the softest speech sounds audible and the loudest still comfortable, it is important to know the dynamic range for speech sounds, the dynamic range for electric stimulation, and the appropriate conversion from speech sounds to electric currents. In clinical fitting procedures, selection of acoustic and electric dynamic ranges and conversion from acoustic amplitude to electric amplitude are part of the "mapping" process, which can play an important role in determining the outcome of cochlear-implant performance and satisfaction. Psychophysical studies have measured the dynamic range over a large electric parameter space and determined the appropriate conversion from acoustic amplitude to electric amplitude (e.g., Zeng and Shannon, 1992, 1994, 1995; Zeng, Galvin, and Zhang, 1998; Zeng and Galvin, 1999). However, less is known about how much speech information should be included in the input dynamic range for cochlear-implant speech processors (Fu and Shannon, 1999; Loizou, Dorman, and Fitzke, 2000). Here, we present new empirical data regarding the speech dynamic range and demonstrate its significance in cochlear-implant performance.

Ideally, a speech processor's input dynamic range (IDR) would be 120 dB, the typical dynamic range within which normal-hearing listeners process acoustic intensity information. This 120-dB acoustic dynamic range would then be converted into electric current values that evoke sensation between minimal to maximal loudness. However, the acoustic dynamic range must be greatly compressed to accommodate the substantially reduced range of electric stimulation for cochlear-implant listeners (about 10–20 dB; see Skinner *et al.*, 1997; Zeng and Galvin, 1999, and Table II of this study). Practically, because speech is likely the most important sound and usually has a smaller dynamic range than the 120-dB range, most implant devices have employed a much narrower 30–60-dB input dynamic range to better match the dynamic range of speech. In so doing, it is hoped that relative intensity changes from soft consonants to loud vowels are preserved perceptually for a cochlear-implant listener to understand speech.

[a)] Author to whom correspondence should be addressed. University of California, 364 Med Surge II, Irvine, CA 92697; electronic mail: fzeng@uci.edu

[b)] Present address: Cochlear Corporation, Englewood, CO.

Currently, there are more than 40 000 cochlear implant users worldwide. Nearly three-quarters of the implant users use the Nucleus device by Cochlear Corporation. In the Nucleus device, a fixed 30-dB range is used for the input dynamic range (User Manual, The Nucleus 22 Channel Cochlear Implant System, p. 4-SP). In the Med-El device, a fixed 60-dB input dynamic range is used (Stobich *et al.*, 1999). In the CLARION device, the input dynamic range can be varied between 20 and 80 dB for users of the simultaneous-analog-stimulation (SAS) strategy and between 10 and 60 dB for users of the continuous-interleaved-sampling (CIS) strategy (CLARION Device Fitting Manual, C9055003-002 Rev. C, p. 220). At present, the clinical fitting practice regarding the input acoustic dynamic range relies mostly on experience and lacks experimental validation.

It has long been assumed that speech has a roughly 30-dB dynamic range, based on classic acoustic analyses and statistical measurements on conversational speech (Dunn and White, 1940; Beranek, 1947; Fletcher, 1953). This 30-dB speech range has been used in many applications, including the Articulation Index (Kryter, 1962; ANSI, 1969, 1997). However, modern analyses using digital signal processing have shown a much wider speech dynamic range. Cox *et al.* (1988) measured distributions of short-term rms levels for conversational speech produced by 30 male and 30 female talkers. They found 40–50-dB distributions of the short-term rms speech levels in eight one-third octave bands covering a frequency range between 250 and 6300 Hz. Pavlovic (1993) noticed that speech dynamic range decreased from 50–60 to 30 dB when the constant of an exponential time window was increased from 13 to 200 ms. In addition, he found that an increase in vocal effort did not simply shift the speech dynamic range towards high values. Boothroyd *et al.* (1994) performed a similar analysis of seven phonemes produced by five female and five male talkers. They found that the overall dynamic range of these phonemes was 53 dB, and that the dynamic range was 37 dB even after adjustment in overall levels and high-frequency pre-emphasis. Stobich *et al.* (1999) calculated the distribution of envelope levels for 180 German sentences spoken by a male talker and found a dynamic range of 70 dB for these speech materials. Eddington (1999) found a 40–60-dB range in the distribution of envelope levels calculated over six frequency channels for TIMIT (Texas-Instruments-Massachusetts-Institute-of-Technology) sentences presented at a conversational level.

Perceptual studies also support the modern acoustic data that find the speech dynamic range to be greater than 30 dB. Studebaker *et al.* (1999) measured NU-6 word recognition at speech presentation levels from 64 to 99 dB SPL and speech-to-noise ratios from 28 to -4 dB. They found a slight increase in speech recognition scores (5 rau units) when the speech level was increased from 64 to 79 dB SPL. This suggests that, contrary to the commonly asserted 30-dB speech range, audibility continued to increase as the overall level increased. Moreover, if a 30-dB speech range were to be assumed, the lowest amplitudes for speech sounds would be 15 to 18 dB lower than the long-term rms level while the peak amplitudes would be 15 to 12 dB higher, according to ANSI (1969, 1997) specifications. If this were the case, word recognition should be similar between 16- and 28-dB speech-to-noise ratios. However, Studebaker *et al.* found significantly poorer speech recognition for the 16-dB than the 28-dB condition (by 5–25 rau units depending on the overall speech presentation levels). These results led Studebaker *et al.* to conclude that the effective dynamic range of speech must be at least 40–43 dB (28 dB + 15 or 12 dB).

In the present study, the distribution of envelope levels was measured for two widely used speech test materials: 12 vowels in /hVd/ format (Hillenbrand *et al.*, 1995) and 20 consonants in /aCa/ format (Turner, Souza, and Forget, 1995; Shannon *et al.*, 1999). Using either a broadband analysis or a narrow-band analysis from eight frequency channels, the data showed that these speech materials have an approximately 50-dB envelope level distribution. Speech recognition in cochlear-implant listeners was then measured as a function of the input dynamic range. Consistent with acoustic data, the perceptual data also showed that an input dynamic range of 50–60 dB produced optimal performance in cochlear implant users.

II. METHODS

A. Subjects

Ten CLARION (Advanced Bionics Corporation) cochlear-implant users participated in the experiment. The implant subjects' ages ranged from 21 to 56 years (average of 42 years). Each subject had at least 1 year of experience with the implant device prior to the experiment. Seven of the implant listeners used the CIS speech processing strategy, while the other three used the SAS strategy. All subjects were postlingually deafened, except for one subject (C4). All subjects were familiar with speech tests from previous clinical evaluations. Additional subject information is listed in Table I. Five normal-hearing (NH) listeners (age range of 21–36 years) also served as a control in the experiment. Local IRB-approved informed consent was obtained and all subjects were paid for their participation.

B. CLARION speech processors

In the experiment, cochlear-implant listeners used their preferred clinical programming parameters (or map) in the speech processors. User maps were uploaded from each subject's speech processor, stored in SOFTWARE-CLINICIAN (SCLIN) for Windows environment (CLARION Device Fitting Manual), and downloaded to a laboratory S-Series speech processor to minimize equipment-related variables. Speech recognition was conducted as a function of input dynamic range (IDR). There were six possible IDR settings with the CIS processing strategy (from 10 to 60 dB in 10-dB steps) and seven possible IDR settings with the SAS processing strategy (from 20 to 80 dB in 10-dB steps). No changes other than the IDR were made to an individual's map. Volume and sensitivity controls were kept constant at their normal settings within and between test sessions.

Figure 1 illustrates the mapping relationship between input dynamic range and electric dynamic range in CLARION cochlear implants. The *x* axis (input dynamic range in dB) determines the range of acoustic signals mapped into the

TABLE I. Biographical and audiological information for cochlear implant participants.

Subject	Age	Surgery date	Device	Strategy	Implant	Etiology
C1	66	11/16/89	S-Series	CIS	Left	Otosclerosis
C2	21	8/5/98	S-Series	CIS	Right	Unknown
C3	39	7/16/97	S-Series	CIS	Left	Unknown
C4	56	11/20/96	1.2	CIS	Left	Maternal rubella
C5	55	1/17/97	S-Series	CIS	Right	Congenital
C6	56	4/25/96	1.2	CIS	Left	Meningitis
C7	35	12/5/96	1.2	CIS	Right	Ototoxicity
S1	46	1/29/98	S-Series	SAS	Left	Unknown
S2	61	5/16/97	S-Series	SAS	Right	Meniere's
S3	76	7/9/98	S-Series	SAS	Right	Unknown

electric range in μA between threshold (*T* level) and the most comfortable loudness (*M* level). Because the *x* axis is logarithmic while the *y* axis is linear, a straight line on these axes indicates a logarithmic mapping between acoustic amplitude and electric amplitude. This logarithmic transformation has been verified psychophysically to restore normal loudness growth in electric stimulation (Eddington *et al.*, 1978; Zeng and Shannon, 1992; Dorman *et al.*, 1993).

For CLARION speech processors, conversions of acoustic to electric amplitudes depend on interactions among the sensitivity setting, the input dynamic range (IDR), and the electric dynamic range. The sensitivity setting determines the peak acoustic amplitude to be mapped to the maximal electric current that evokes the most comfortable loudness (*M* level). This peak acoustic amplitude is then used as the reference level (0 dB) for the input dynamic range, which can vary from 10 to 80 dB. An IDR setting of $-X$ dB means that only *X* dB of the acoustic range below the reference peak amplitude will be mapped to an implant user's electric dynamic range (between *M* and *T* levels). Acoustic amplitudes below the IDR setting stimulate at subthreshold levels ($<T$ level).

For example, an IDR setting of -50 dB (dashed sloping line) maps the 50-dB range below the 0-dB peak reference acoustic level into the audible electric dynamic range. Presumably, any acoustic level that is outside the input dynamic range will be mapped into either a subthreshold electric level ($<T$ level) or a constant saturating level ($=M$ level). Note, however, an interchangeable relationship between the IDR and the *T*-level settings. For instance, the same acoustic-to-electric amplitude map, as determined by the -50-dB IDR and the *T*-level settings (the dashed sloping line), can also be achieved by reducing the IDR setting to -40, -30, and -20 dB, while increasing the *T* level to *T*1, *T*2, and *T*3, respectively.

C. Stimuli

Vowel stimuli consisted of 12 tokens from five male and five female talkers in /hVd/ format (Hillenbrand *et al.*, 1995). Consonant stimuli consisted of 20 tokens from five male and five female talkers in /aCa/ format (Turner, Souza, and Forget, 1995; Shannon *et al.*, 1999). The Hillenbrand vowels were 16-bit.WAV files samples at 16 kHz, and the Turner/Shannon consonants were 16-bit.WAV files sampled at 44.1 kHz. All speech tokens were normalized based on the maximal rms level from a 50-ms running window. This maximal level most likely measured the level of the steady-state portion of the vowel.

These vowel and consonant stimuli were output via a PC soundcard (Turtle Beach MultiSound Fiji board), connected to one channel of a mixer (Tucker-Davis Technologies, TDT SM1). Continuous speech-spectrum-shaped noise was generated by passing white noise (TDT WG1) through a specially designed low-pass filter with a cutoff frequency at 608 Hz and a -12-dB/octave slope (Byrne *et al.*, 1994). The noise was delivered to another channel of the mixer where it was summed with the phonemic stimuli.

The summed speech and noise stimuli were amplified (Crown D-75) and presented to the listener via a Tannoy Reveal speaker mounted on a double-walled, sound-treated booth (IAC). Each subject was positioned in the center of the sound-treated room, facing the speaker (about 1 meter away, at 0° azimuth and at ear level). A calibration vowel /a/ was generated to have the same rms level as the average vowel level in both tests and to produce a conversational level of 65 dBA. The noise was attenuated via a programmable attenu-

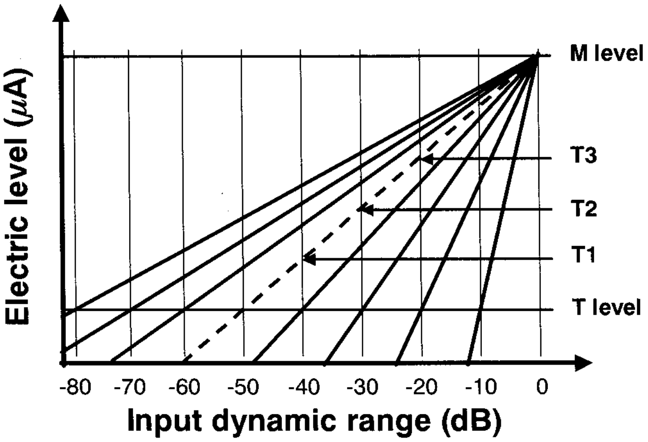

FIG. 1. Conversion from input dynamic range (*x* axis) to electric dynamic range (*y* axis) in CLARION devices. The reference acoustic level (0 dB) represents the peak amplitude, as determined by the sensitivity control, to be mapped to the most comfortable loudness level in electric stimulation (*M* level). Input dynamic range setting determine the range below the 0-dB reference acoustic level to be mapped into audible electric range between *M* and *T* levels. *T* level is electric threshold, representing 50% detection. Note the interchangeable relationship between IDR and T level settings in determining the acoustic-to-electric amplitude map (see the text in Sec. II B for detailed information).

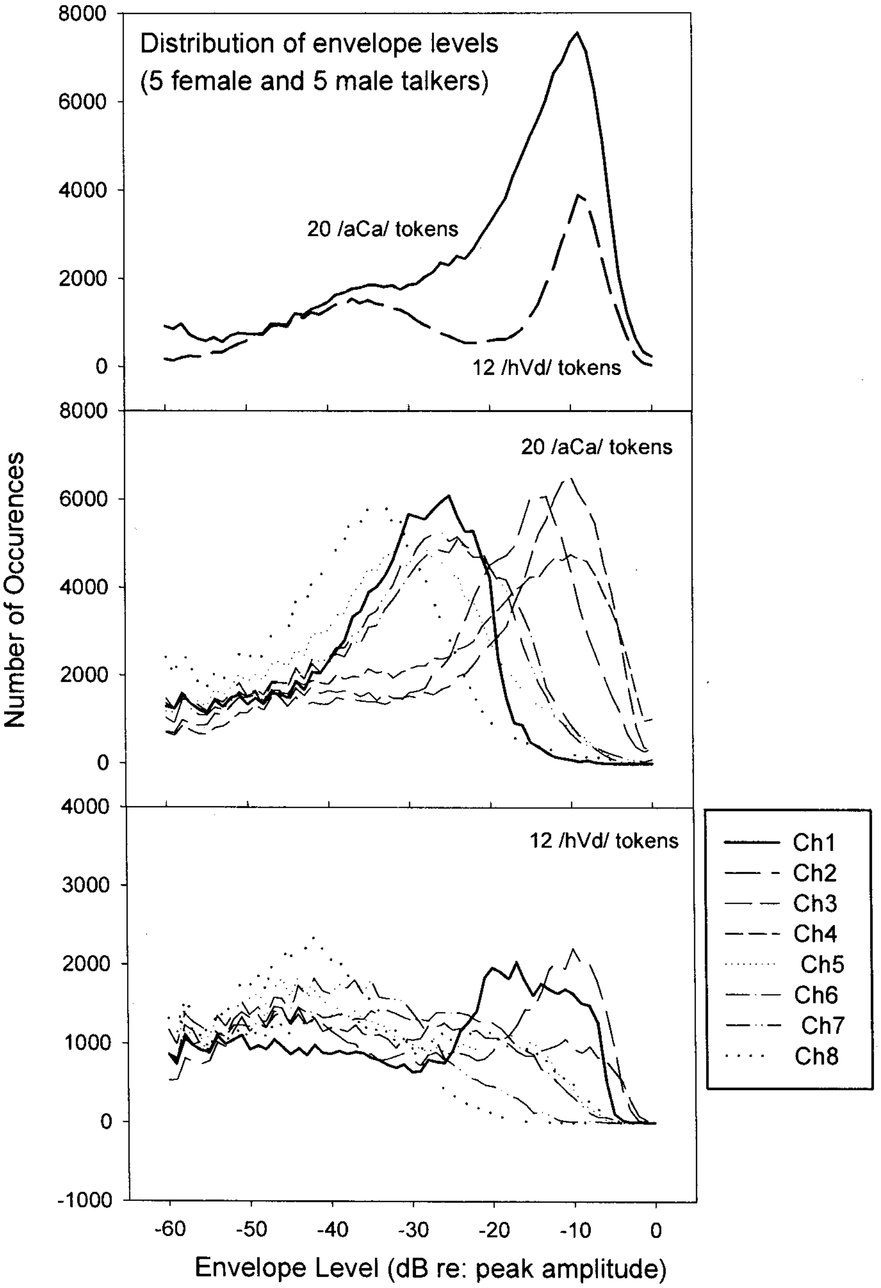

FIG. 2. Speech dynamic ranges in terms of envelope level distributions for the broadband condition (top panel) and the eight-narrow-band conditions (for consonants see the middle panel and for vowels see the bottom panel).

ator (TDT PA4) to achieve a +5-dB speech-to-noise ratio (i.e., the noise had a level of 60 dBA).

D. Procedures

Distribution of speech envelope levels was calculated for both broadband (250–6800 Hz) and narrow-band analysis. In the broadband analysis, the envelope of the acoustic signal was extracted by full-wave rectification and low-pass filtering (an Elliptical IIR filter with 160-Hz cutoff frequency and −6 dB per octave slope). The 160-Hz low-pass filter had an equivalent time window of roughly 6 ms, about half of the shortest integration value (13 ms) used in the Pavlovic study. A histogram recorded the number of occurrences for envelope amplitude (*re*: peak amplitude). Because coarticulation cues are always present in the adjacent sounds, the histogram also included contribution from the initial and final phonemes (/h/ and /d/ for vowels and /a/ for consonants). Because of the noise floor on the bottom of the distribution, the speech dynamic range was conservatively defined as the difference in the envelope levels producing between 5% and 99% accumulative occurrences. In the narrow-band analysis, the broadband signal was divided into eight narrow bands (fourth-order Elliptical IIR filters with cutoff frequencies at 250, 500, 875, 1150, 1450, 2000, 2600, 3800, and 6800 Hz, respectively). These filters approximately correspond to the frequency analysis filters used in the CLARION cochlear implant. The band-specific envelope was extracted and its amplitude histogram was constructed in the same way as for the broadband analysis.

Closed-set vowel and consonant recognition was measured separately using custom software (Robert, 1999). During each test, listeners heard five presentations of each phoneme spoken by each of the ten talkers; presentation of each

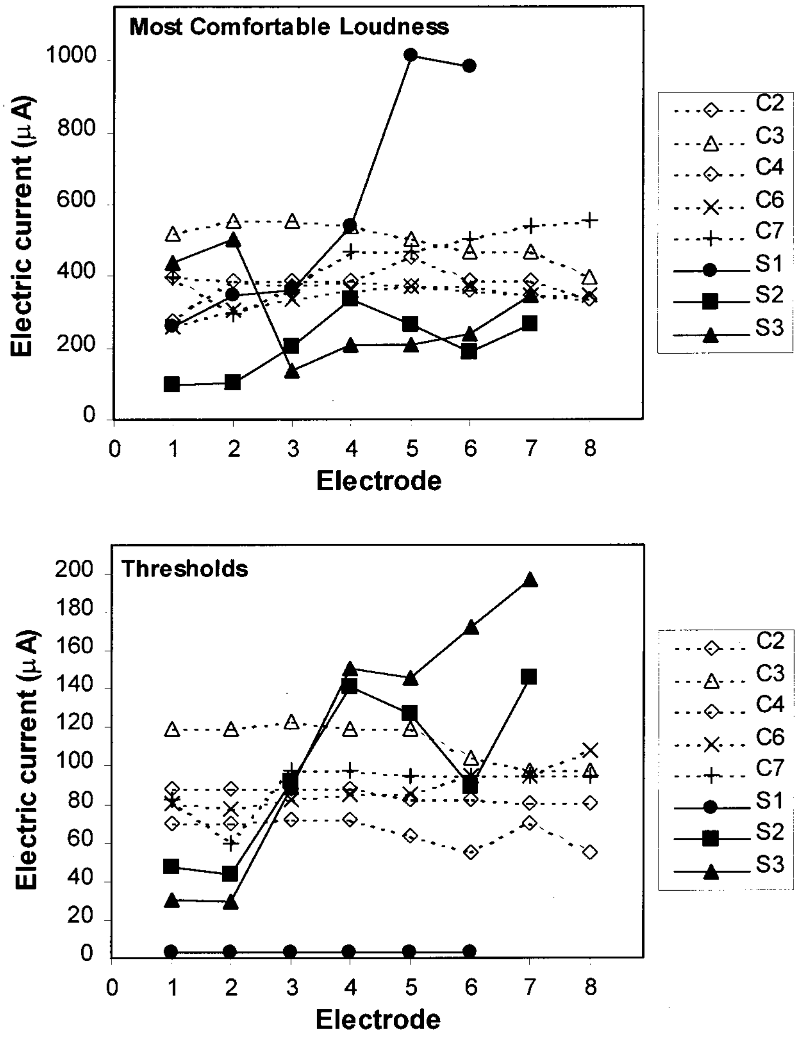

FIG. 3. Top panel displays the most comfortable loudness (*M* level) as a function of electrodes (*x* axis). Bottom panel displays threshold (*T* level) as a function of electrodes (*x* axis). Note both the greater intersubject- and intrasubject variability for the SAS users (filled symbols connected by solid lines) than for the CIS users (open symbols connected by dashed lines).

phoneme was randomized within the test. Listeners' responses to the stimuli were stored in a confusion matrix. No trial-by-trial feedback was given regarding the correctness of the response. Normal-hearing listeners listened to the original phoneme sounds in quiet and in noise (+5-dB S/N) with speech presented at 65 dBA. Implant listeners were tested first in quiet and then in noise. The test order of different input dynamic ranges was pseudorandomized. Each listener was given 15 min to acclimate to the experimental processor and was allowed to preview all stimuli before formal test sessions. Due to time limitation, some conditions were not tested.

III. RESULTS

A. Speech dynamic range

Figure 2 shows distribution of envelope levels for the /aCa/ and /hVd/ tokens in the broadband analysis (top panel) and for the /aCa/ tokens (middle panel) and the /hVd/ tokens (bottom panel) in the eight-channel analysis. First, note the dominant envelope level distribution at high levels for the broadband analysis. A small "bump" in the distribution at low levels (more obvious in the vowel envelope) most likely reflects the contribution from the lower-amplitude consonants. This is clearly illustrated in the narrow-band analysis, which shows a strong distribution at low levels for the high-frequency channels (dotted lines in the middle and bottom panels).

The broadband analysis (top panel) shows the acoustic dynamic range to be 47 dB (from −51 to −4 dB) for consonants and 46 dB (from −50 to −4 dB) for vowels. For the eight-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,

TABLE II. Electric dynamic range in dB for five CIS users (C2, C3, C4, C6, and C7) and three SAS users (S1, S2, and S3). Note that S1 had six usable electrodes while S2 and S3 and seven usable electrodes.

Electrode	C2	C3	C4	C6	C7	S1	S2	S3
1	13.1	12.8	11.9	10.1	13.6		6.1	23.3
2	12.8	13.4	14.8	11.9	13.9	38.7	7.5	24.8
3	12.8	13.1	14.2	12.2	11.3	41.3	6.9	3.7
4	12.8	13.1	14.5	12.5	13.6	41.6	7.5	2.9
5	13.1	12.5	17.1	12.8	13.9	45.1	6.4	3.2
6	12.8	13.1	16.8	11.9	14.5	50.6	6.6	2.9
7	12.8	13.6	14.8	11.3	15.1	50.3	5.2	4.9
8	12.5	12.2	15.7	10.2	15.4			
Average	12.8	13.0	15.0	11.6	13.9	44.6	6.6	9.4
Std dev.	0.2	0.5	1.6	1.0	1.2	5.0	0.8	10.0

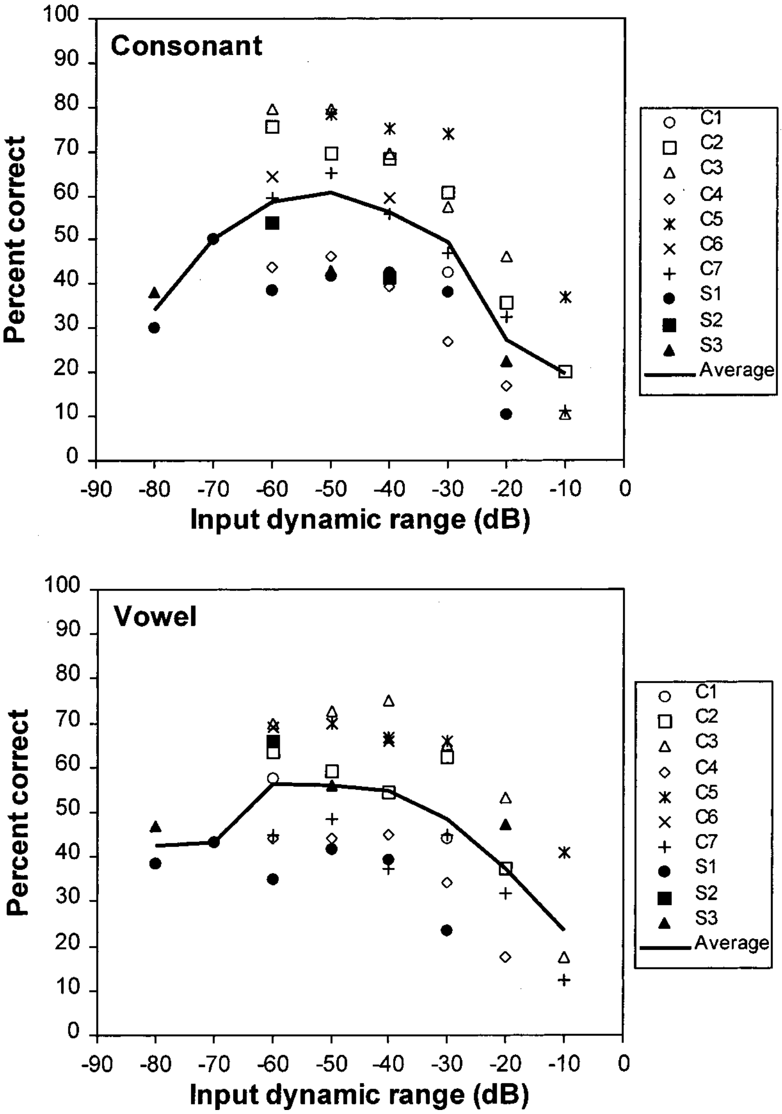

FIG. 4. Consonant (top) and vowel (bottom) recognition scores (*y* axis) in quiet as a function of the input dynamic range (*x* axis). Individual data are represented by symbols (unfilled symbols for CIS users and filled symbols for SAS users). The solid line represents the average data.

2, 3, 4, 5, 6, 7, and 8, respectively. Given these acoustic dynamic ranges, we shall see whether an input dynamic range setting of roughly 50 dB would produce optimal speech recognition in cochlear-implant users.

B. Electric dynamic range

Figure 3 shows the most comfortable loudness (*M* levels, top panel) and threshold (*T* levels, bottom panel) as a function of electrode position in 5 of the 7 CIS users and all 3 SAS users. These *M* and *T* levels are presented in microamps. Note the greater intersubject variability in both *M* and *T* levels for the SAS users compared to the CIS users. Also note the greater intrasubject variability across electrodes for the SAS users than the CIS users. The high *M* levels for subject S1 may actually be much lower as they approach the saturation portion of the current source in the CLARION S-series devices (CLARION Device Fitting Manual, p. 20).

Table II shows the calculated electric dynamic ranges, defined as the dB difference between *M* and *T* levels. Table II confirms the visual impression in Fig. 3 that the SAS users have both greater intersubject and intrasubject variability in dynamic range than the CIS users. The electric dynamic range averaged across electrodes was 12.8, 13.0, 15.0, 11.6, and 15.0 dB for the CIS users, C2, C3, C4, C6, and C7, respectively. On the other hand, the averaged electric dynamic range was 44.6, 6.6, and 9.4 dB for the SAS users, S1, S2, and S3, respectively. The unusually large dynamic range for S1 may actually be much lower and could be calculated by accessing the subject's internal device. Similarly, the variability in dynamic range across each subject's electrodes is much smaller (standard deviation ranges from 0.2 to 1.6 dB) for the CIS users than the SAS users (standard deviation ranges from 0.8 to 10.0 dB). The reasons for the differences between the CIS and SAS user results are not clear, but may be related to the difference in electrode configurations and stimulating waveforms between the two strategies.

C. Phoneme recognition in quiet

Figure 4 shows both the group average (line) and individual data (symbol). The top panel shows consonant recognition (*y* axis) as a function of input dynamic range (*x* axis), while the bottom panel shows vowel recognition (*y* axis) as a function of input dynamic range (*x* axis). For the five

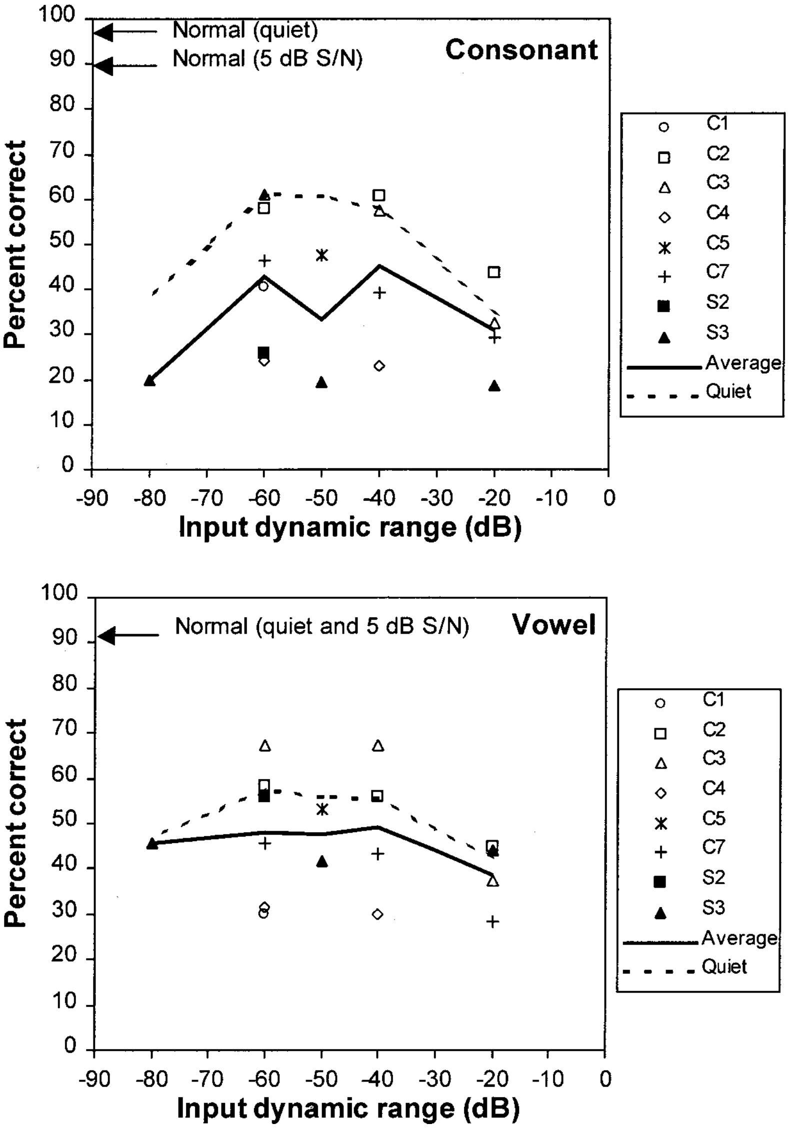

FIG. 5. Consonant (top) and vowel (bottom) recognition scores (*y* axis) in noise as a function of the input dynamic range (*x* axis). Individual data are represented by symbols (unfilled symbols for CIS users and filled symbols for SAS users). The solid line represents the average data. For comparison, the correspondent average data in quiet (see the text) are also included (dashed line). Arrow lines represent the average control data from five normal-hearing listeners.

normal-hearing listeners, the average score for consonant recognition was 97% and the score for vowel recognition was 93%. For the implant listeners, the best average score was about 40 percentage points lower than the normal-hearing controls; even the best individual score was still about 15 percentage points lower than the controls.

Average phoneme identification performance can be described by a nonmonotonic function, with best performance occurring at IDRs of −40 to −60 dB and decreased performance at lower and higher IDRs. Individual data showed a similar trend, but the range of performance varied greatly. Individual performance variability ranged between 30 and 45 percentage points for all except the −70- and −80-dB IDR conditions (data were collected in two of the three SAS users but could not be obtained in the CIS users at these IDRs).

A one-way analysis of variance (ANOVA) confirmed that the input dynamic range is a significant factor affecting speech recognition in CLARION cochlear-implant users [consonants: $F(7,36)=6.19$, $p<0.01$; vowels: $F(7,33)=2.79$, $p<0.05$]. A paired t-test indicated no significant difference in consonant recognition between the −50- and the −60-dB IDR conditions ($p>0.05$), but significantly poorer performance for the remaining narrower IDR conditions ($p<0.01$). For vowel recognition, there was no significant difference between −40-, −50-, and −60-dB IDR conditions ($p>0.05$); however, all produced better performance than the −10-, −20-, and −30-dB IDR conditions. No statistical test was conducted between the medium and the −70/−80-dB IDR conditions because of the small number of subjects. The present data suggest that the input dynamic range should be set to 50 dB or greater in order to achieve optimal speech performance in quiet.

D. Phoneme recognition in noise

Figure 5 shows consonant and vowel recognition as a function of input dynamic range for the 5-dB speech-to-noise ratio condition. For comparison, the dashed line shows the averaged data for the previous quiet condition. Because not all subjects were tested for every condition, the averaged data for the quiet condition is shown for only those conditions where corresponding, noise data were available. A paired t-test revealed that noise significantly lowered both consonant ($p<0.001$) and vowel recognition scores (p

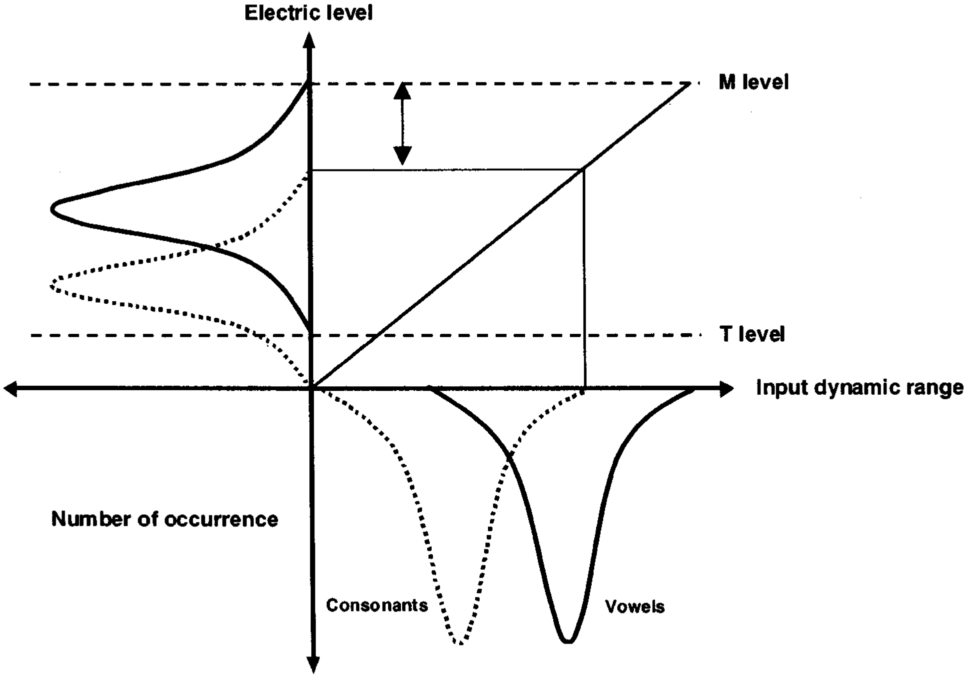

FIG. 6. Effects of envelope level distribution. I. Logarithmic mapping for both consonants and vowels. The right-bottom quadrant shows idealized acoustic envelope level distribution for consonants (dotted line) and vowels (solid line). The right-top quadrant shows the logarithmic acoustic-to-electric conversion. The left-top quadrant shows electric envelope level distribution. Note that a portion of low electric envelope levels is mapped below threshold (*T* level) and also that a portion of electric dynamic range is unused (indicated by the line with double arrowhead).

<0.01). The presence of noise seemed to "flatten" the consonant and vowel recognition functions observed in quiet. This trend is particularly apparent in vowel recognition performance. Average consonant recognition in noise was more degraded by wide IDR settings (a decrease of 20 percentage points for IDRs between -50 and -80 dB) than by narrow IDR settings (merely a decrease of 4 percentage points for the -20-dB IDR setting). CLARION fitting procedures recommend reducing the IDR to aid in noise suppression and to a limited extent, the present results support that recommendation. More detailed assessment of the effect of IDR settings on phoneme recognition in noise is difficult due to the large individual variability as well as the limited data collection in the present study.

IV. DISCUSSION

The present results have both theoretical and practical significance. Theoretically, the acoustic analysis results showed that multitalker phonemes have an approximately 50-dB distribution of envelope levels, which is much wider than the commonly assumed 30-dB speech dynamic range. In the broadband (250–6800 Hz) analysis, the distribution of consonant and vowel envelope levels, particularly the vowel envelope levels, showed a bimodal pattern (top panel in Fig. 2). This bimodal distribution disappeared in the narrow-band analysis (middle and bottom panels in Fig. 2), approximating a normal distribution with different means for different frequency bands. The high-frequency channels had a shifted

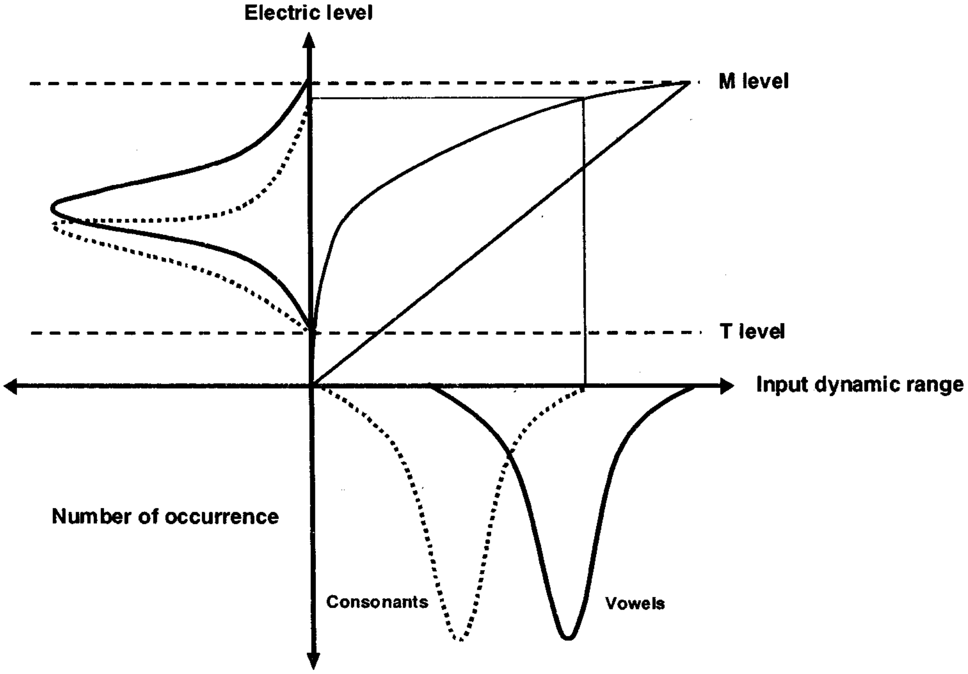

FIG. 7. Effects of envelope level distribution. II. Logarithmic mapping for vowels and more compressive mapping for consonants. The right-bottom quadrant shows idealized acoustic envelope level distribution for consonants (dotted line) and vowels (solid line). The right-top quadrant shows the logarithmic acoustic-to-electric conversion (straight line) for vowels and the more compressive conversion (curved line) for consonants. The left-top quadrant shows electric envelope level distribution. Note the improved use of electric dynamic range for consonants.

distribution towards lower envelope levels than the low-frequency channels. Presumably, the high-frequency channels carry mostly consonant information such as fricatives and stops, while the low-frequency channels carry mostly vowel information. In practical clinical fittings, this difference in envelope level distribution may significantly affect how consonants and vowels should be mapped into an implant user's audible electric range.

Acoustic-to-electric amplitude mapping has been studied extensively in users of auditory brainstem implants (Shannon, Zeng, and Wygonski, 1992), the Med-El/CIS-Link Ineraid devices (Boex *et al.*, 1995; Wilson *et al.*, 1999; Loizou, Poroy, and Dorman, 2000), and the Nucleus devices using either four-channel CIS-type processing (Fu and Shannon, 1998) or the SPEAK strategy (Zeng and Galvin, 1999). A general trend noted in these studies was that a more compressive map would produce better consonant recognition than a less compressive map, while the degree of compression appears to have a small opposite effect, if at all, on vowel recognition (Boex *et al.*, 1995; Zeng and Galvin, 1999). The present acoustic analysis can account for these observations.

Figure 6 shows a case where the acoustic envelope amplitude of both consonants (dotted line) and vowels (solid line) is mapped into the electric level using the same logarithmic function (assuming input dynamic range is in dB and electric level is in microamps). The two horizontal dashed lines represent the electric threshold (T level) and the most comfortable loudness (M level). Because the consonant envelope distribution was about 20 dB lower than the vowel envelope distribution (Fig. 2 middle and bottom panels), the consonants are likely to be mapped into a less-than-optimal electric range. First, some low envelope levels may be mapped into electric levels below threshold (the lower horizontal dotted line, top-left quadrant). Second, some of the upper portion of the electric dynamic range (indicated by the line with the double arrowhead) may not be utilized because few amplitude envelope levels ($<1\%$) are present. Third, most envelope levels are likely mapped into the lower portion of the electric dynamic range where intensity discrimination and modulation detection are both poor (Nelson *et al.*, 1996; Zeng *et al.*, 1998; Fu, 2000).

On the other hand, if a more compressive map is used for consonants, then all three undesirable effects can be alleviated. Figure 7 shows the same map as in Fig. 6 for vowels but a more compressive map for consonants (the curved line on the right-top quadrant). The compression will raise previously inaudible low envelope levels above threshold, reduce the unused portion of the electric dynamic range, and map more of the envelope into the upper electric dynamic range where intensity discrimination and modulation are optimal. One negative trade-off for the more compressive mapping is the possibility that some low-level noise may become audible. Another negative trade-off is the slightly distorted envelope level distribution (see the mapped consonant envelope distribution in electric domain, left-top quadrant). However, previous results on the effects of changing consonant-to-vowel ratios (Freyman, Nerbonne, and Cote, 1991) and amplitude compression (Souza and Turner, 1996; Van Tasell and Trine, 1996) indicated that this distortion should produce little, if any, decrease in consonant recognition.

Theoretically, under laboratory conditions in which the envelope level distribution for test materials is known, one can optimally set each channel's mapping function based on the mean and standard deviation of the envelope level distribution of that channel. Under realistic listening situations in which speech materials cannot be controlled and real-time processing is required, more compressive mapping for high-frequency channels relative to low-frequency channels will help map the consonant envelope levels into the full electric dynamic range. In other words, cochlear-implant users may achieve better overall speech recognition with a logarithmic map for low-frequency channels and a more compressive map for high-frequency channels. Such implementation is not feasible with the present clinical fitting systems. A future study using a research interface to the cochlear implant is required to implement the different mapping functions for different frequency channels and to evaluate its predicted improvement in speech recognition.

V. CONCLUSIONS

The present study measured the speech dynamic range using 20 consonants and 12 vowels spoken by five female and five male talkers. The present study also measured speech recognition in CLARION implant users as a function of input acoustic dynamic range. The acoustic and perceptual data support the following conclusions:

(1) The speech dynamic range is about 50 dB, much wider than the commonly assumed 30-dB dynamic range.
(2) An input dynamic range of 50–60 dB is required to support optimal speech recognition in cochlear implants.
(3) Current cochlear-implant users may benefit from a new amplitude mapping strategy that uses a logarithmic map for low-frequency channels and a more compressive map for high-frequency channels.

ACKNOWLEDGMENTS

The authors thank Lendra Friesen, Mark Robert, and John Wygonski for help in patient recruiting, speech test, and engineering support, respectively. We also thank Ginger Stickney, Rachel Cruz, Chris Turner, and two anonymous reviewers for comments on the manuscript. This study was supported by National Institutes of Health (Nos. RO1-DC-02267 and N01-DC-92100).

ANSI (**1969**). ANSI S3.5-1969, "American National Standards Methods for the Calculation of the Articulation Index" (American National Standards Institute, New York).

ANSI (**1997**). ANSI S3.5-1997, "American National Standards Methods for the Calculation of the Speech Intelligibility Index" (American National Standards Institute, New York).

Beranek, L. L. (**1947**). "The design of speech communication systems," Proc. Inst. Radio Eng. **35**, 880–890.

Boex, C. S., Pelizzone, M., Piloux, V., and Montandon, P. (**1995**). "Use of loudness scaling measurements to determine compressive mapping in speech processing for cochlear implants," Abstracts of 1995 Conference on Implantable Auditory Prostheses, p. 57.

Boothroyd, A., Erickson, F. N., and Medwetsky, L. (**1994**). ''The hearing aid input: A phonemic approach to assessing the spectral distribution of speech,'' Ear Hear. **6**, 432–442.

Byrne, D., Dillon, H., Tran, K. (**1994**). ''An international comparison of long-term average speech spectra,'' J. Acoust. Soc. Am. **96**, 2108–2120.

Cox, R. M., Matesich, J. S., and Moore, J. N. (**1988**). ''Distributions of short-term rms levels in conversational speech,'' J. Acoust. Soc. Am. **84**, 1100–1104.

Dorman, M. F., Smith, L., Parkin, J. L. (**1993**). ''Loudness balance between acoustic and electric stimulation by a patient with a multichannel cochlear implant,'' Ear Hear. **14**(4), 290–292.

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

Eddington, D. K., Dobelle, W. H., Brackmann, D. E., Mladejovsky, D. E., and Parkin, J. L. (**1978**). ''Auditory prostheses research with multiple channel intracochlear stimulation in man,'' Ann. Otol. Rhinol. Laryngol. **87** (Suppl 53), 5–39.

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* (Van Nostrand, New York), pp. 69–88.

Freyman, R. L., Nerbonne, G. P., and Cote, H. A. (**1991**). ''Effect of consonant–vowel ratio modification on amplitude envelope cues for consonant recognition,'' J. Speech Hear. Res. **34**, 415–426.

Fu, Q.-J. (**2000**). ''Auditory temporal resolution and speech performance in cochlear implant users,'' J. Acoust. Soc. Am. **108**, 2600.

Fu, Q.-J., and Shannon, R. V. (**1998**). ''Effects of amplitude nonlinearity on phoneme recognition by cochlear implant users and normal-hearing listeners,'' J. Acoust. Soc. Am. **104**, 2570–2577.

Fu, Q.-J., and Shannon, R. V. (**1999**). ''Effect of acoustic dynamic range on phoneme recognition in quiet and noise by cochlear implant users,'' J. Acoust. Soc. Am. **106**, L65–70.

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

Kryter, K. D. (**1962**). ''Methods for the calculation and use of the articulation index,'' J. Acoust. Soc. Am. **34**, 1689–1697.

Loizou, P. C., Dorman, M., and Fitzke, J. (**2000**). ''The effect of reduced dynamic range on speech understanding: Implications for patients with cochlear implants,'' Ear Hear. **21**, 25–31.

Loizou, P. C., Poroy, O., and Dorman, M. (**2000**). ''The effect of parametric variations of cochlear implant processors on speech understanding,'' J. Acoust. Soc. Am. **108**, 790–802.

Nelson, D. A., Schmitz, J. L., Donaldson, G. S., Viemeister, N. F., and Javel, E. (**1996**). ''Intensity discrimination as a function of stimulus level with electric stimulation,'' J. Acoust. Soc. Am. **100**, 2393–2414.

Pavlovic, C. V. (**1993**). ''Problems in the prediction of speech recognition performance of normal-hearing and hearing-impaired individuals,'' in *Acoustical Factors Affecting Hearing Aid Performance*, edited by G. A. Studebaker and I. Hochberg, 2nd ed. (Allyn and Bacon, Boston), pp. 221–234.

Robert, M. E. (**1999**). ''CONDOR: Documentation for version 1.3©,'' House Ear Institute, Los Angeles, CA.

Shannon, R. V., Zeng, F.-G., and Wygonski, J. (**1992**). ''Speech recognition using only temporal cues,'' in *The Auditory Processing of Speech: From Sounds to Words*, edited by M. E. H. Schouten (Mouton de Gruyter, Berlin), pp. 263–274.

Shannon, R. V., Jensvold, A., Padilla, M., Robert, M. E., and Wang, X. (**1999**). ''Consonant recordings for speech testing,'' J. Acoust. Soc. Am. **106**, L71–74.

Skinner, M. W., Holden, L. K., Holden, T. A., Demorest, M. E., and Fourakis, M. S. (**1997**). ''Speech recognition at simulated soft, conversational and raised-to-loud vocal efforts by adults with cochlear implants,'' J. Acoust. Soc. Am. **101**, 3766–3782.

Souza, P. E., and Turner, C. W. (**1996**). ''Effect of single-channel compression on temporal speech information,'' J. Speech Hear. Res. **39**, 901–911.

Stobich, B., Zierhofer, C. M., Hochmair, E. S. (**1999**). ''Influence of automatic gain control parameter settings on speech understanding of cochlear implant users employing the continuous interleaved sampling strategy,'' Ear Hear. **20**(2), 104–116.

Studebaker, G. A., Sherbecoe, R. L., McDaniel, D. M., Gwaltney, C. A. (**1999**). ''Monosyllabic word recognition at higher-than-normal speech and noise levels,'' J. Acoust. Soc. Am. **105**, 2431–2444.

Turner, C. W., Souza, P. E., and Forget, L. N. (**1995**). ''Use of temporal envelope cues in speech recognition by normal and hearing-impaired listeners,'' J. Acoust. Soc. Am. **97**, 2568–2576.

Van Tasell, D. J., and Trine, T. D. (**1996**). ''Effects of single-band syllabic amplitude compression on temporal speech information in nonsense syllables and in sentences,'' J. Speech Hear. Res. **39**, 912–922.

Wilson, B. S., Lawson, D. T., Zerbi, M., and Wolford, R. D. (**1999**). ''Speech processors for auditory prostheses,'' NIH Project NO1-DC-8-2105, Third Quarterly Progress Report.

Zeng, F.-G., and Galvin, J. J. (**1999**). ''Amplitude compression and phoneme recognition in cochlear implant listeners,'' Ear Hear. **20**, 60–74.

Zeng, F.-G., Galvin, J. J., and Zhang, C. Y. (**1998**). ''Encoding loudness by electric stimulation of the auditory nerve,'' NeuroReport **9**, 1845–1848.

Zeng, F.-G., and Shannon, R. V. (**1992**). ''Loudness balance between acoustically and electrically stimulated ears,'' Hear. Res. **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,'' Ann. Otol. Rhinol. Laryngol. **104** (Suppl 166), 235–238.