

Phoneme recognition by cochlear implant users as a function of signal-to-noise ratio and nonlinear amplitude mapping

Qian-Jie Fu and Robert V. Shannon

*Department of Auditory Implants and Perception, House Ear Institute,
2100 West Third Street, Los Angeles, CA 90057
qfu@hei.org shannon@hei.org*

Abstract: The present study measured phoneme recognition as a function of signal-to-noise levels when different nonlinear loudness mapping functions were implemented in three cochlear implant users using a 4-channel continuous interleaved sampler (CIS) strategy. Results show that phoneme recognition scores in quiet vary only slightly when different amplitude mappings from highly compressive to weakly compressive are applied. As the level of background noise is increased, recognition scores decrease more rapidly for the strongly compressive mapping than for the weakly compressive mapping. Results indicate that, although a strongly compressive mapping between acoustic and electric amplitude produces slightly better performance in quiet, a less compressive mapping may be beneficial for implant listeners in noisy listening conditions.

©1999 Acoustical Society of America

PACS numbers: 43.71.Es, 43.71.Ky, 43.66.Ts

1. Introduction

In quiet laboratory testing conditions, many cochlear implant users with the latest implant devices can achieve high levels of open-set sentence recognition. However, performance deteriorates significantly in noisy listening conditions (Hochberg et al., 1992; Müller-Deiler et al., 1995) for even the best cochlear implant users. Several explanations for the noise susceptibility of cochlear implant listeners have been proposed recently. One of the most obvious factors is the limited spectral resolution in cochlear implants. In a recent study, Fu et al. (1998) measured phoneme recognition in five normal-hearing listeners as a function of the number of spectral channels. Results showed that as the spectral information was reduced, speech recognition deteriorated only slightly in quiet, but significantly in noise. Phoneme recognition performance of the best cochlear implant users was similar to that of normal-hearing subjects listening to a similar level of spectral reduction. A similar result was reported by Dorman and colleagues (1998a, 1998b). These results indicate that the limited spectral resolution in cochlear implant listeners is a key factor causing noise susceptibility. However, one interesting observation from the Fu et al. and Dorman et al. studies is that some implant listeners had performance comparable to normal-hearing listeners in quiet but significantly poorer performance in noise. One factor that may have contributed to this difference is the loudness mapping function between acoustic amplitude and electric current.

Instantaneous amplitudes in normal speech range over a 30 to 60 dB range (Boothroyd et al., 1994). However, implant listeners typically have dynamic ranges of only 6 to 15 dB in electric current, requiring the acoustic range to be compressed into the electric range. Fu and Shannon (1998) investigated the effect of nonlinear amplitude mapping on vowel and consonant recognition in cochlear implant users and normal-hearing listeners. They found that for both acoustic and implant listeners, the best performance was obtained

when normal loudness was preserved, which, for electrical stimulation, was obtained when a compressive power-law mapping ($p=0.22$) was applied. A traditional power-law, cross-modality model indicated that this mapping best restored the loudness growth in cochlear implant users. Performance deteriorated only slightly in both acoustic and implant listeners when the amplitude mapping function was either more compressive or more expansive. Thus, instantaneous amplitude nonlinearity has only a minor effect on phoneme recognition in quiet.

Fu and Shannon (1999) investigated the effects of nonlinear amplitude mapping on the recognition of spectrally degraded speech in noise by normal-hearing subjects. They measured vowel and consonant recognition in five normal-hearing listeners as a function of signal-to-noise ratio (S/N) with the exponent of the amplitude-mapping power function as a parameter. The results showed that nonlinear amplitude mapping produced only a mild decrement in speech recognition in quiet but could produce a large decrement in noise. Expansive nonlinear mapping provided better overall performance than linear or compressive mapping in low signal-to-noise ratio conditions.

The goal of the present study is to investigate the effect of nonlinear amplitude mapping on phoneme recognition in cochlear implant users.

2. Methods

2.1 Subjects

Cochlear implant subjects were three postlingually deafened adults using the Nucleus-22 device. All had at least four years experience utilizing the SPEAK speech processing strategy and all were native speakers of American English. The Nucleus processor with the SPEAK strategy divides an input acoustic signal into 20 frequency bands, extracts the amplitude envelope from all 20 bands, and stimulates the electrodes corresponding to the 6 to 10 bands with the maximal amplitude (McDermott et al, 1992). The frequency allocation table specifies the frequency range covered by the speech processor. Two subjects (N4 and N7) used frequency allocation table 9 (150-10,823 Hz) in their clinical implant processor, and one subject (N3) used frequency allocation table 7 (120 Hz - 8,658 Hz). All implant subjects had 20 active electrodes available for use. All implant participants had extensive experience in speech recognition experiments. Table 1 contains relevant information for the three subjects, including their most recent scores on the HINT sentence test and a multitalker, 12-vowel recognition test with their 20-electrode SPEAK processor.

Table 1. Subject information on three Nucleus-22 cochlear implant listeners who participated in the present study.

Subject	Age	Gender	Cause of Deafness	Duration of use	Freq. Table	Score (HINT)	Vowel Score
N3	56	M	Trauma	7 years	7	96.2%	69.5%
N4	40	M	Trauma	5 years	9	100.0%	81.1%
N7	55	M	Unknown	5 years	9	100.0%	64.5%

2.2 Test materials and procedures

Speech recognition was assessed for medial vowels and consonants. Vowel recognition was measured in a 12-alternative identification paradigm, including 10 monophthongs (/i ɪ ε æ α ɔ u ʌ ɜ/) and 2 diphthongs (/e o/), presented in a /h/-vowel-/d/ context. The tokens for these closed-set tests were digitized natural productions from 5 males, 5 females, and 5 children, drawn from the material collected by Hillenbrand et al. (1995). Consonant recognition was measured in a 16-alternative identification paradigm for the consonants /b d g p t k l m n f s ʃ v z j θ/ presented in an /a/-consonant-/a/ context. Two repetitions of each of the 16 consonants were produced by three speakers (1 male, 2 female) for a total of 96 tokens (16 consonants * 3 talkers * 2 repeats).

Each test block included 180 tokens for vowel recognition or 96 tokens for consonant recognition. A stimulus token was randomly chosen from all 180 tokens in vowel recognition and from 96 tokens in consonant recognition and presented to the subject. Following the presentation of each token, the subject responded by pressing one of 12 buttons in the vowel test or one of 16 buttons in the consonant test, each marked with one of the possible responses. The response buttons were labeled in a /h/-vowel-/d/ context (heed, hawed, head, who'd, hid, hood, hud, had, heard, hoed, hod, hayed) for the vowel recognition task and a /a/-consonant-/a/ context followed by an example word for the consonant recognition task.

All signals were presented at comfortable audible levels through a custom implant interface system (Shannon et al., 1990). Subjects had been well familiarized with the test materials and the test procedure from prior experiments. All subjects started with a training session. Speech sounds without noise were used as training conditions. Each training session included 8 consecutive test blocks with the same mapping condition and the same speech material. Feedback was provided. Subjects started the test sessions either after 8 consecutive runs or the performance had been stabilized in three consecutive runs. In the test sessions, the order of signal-to-noise ratio conditions was randomized. The order of the five mapping conditions and the order of the vowel and consonant tests were counterbalanced across subjects. No feedback was provided in test sessions.

2.3 Signal processing

The speech signal was mixed with simplified speech spectrum-shaped noise (constant spectrum level below 800 Hz and 10-dB/octave roll-off above 800 Hz). The signal-to-noise ratio (S/N) was defined as the difference in decibels between the root-mean-square (RMS) levels of the whole speech token and the noise.

The 4-channel Continuous Interleaved Sampler (CIS) processor (Wilson et al., 1991) was implemented as follows. The signal was first pre-emphasized using a first-order Butterworth high-pass filter with a cutoff frequency of 1200 Hz and then band-pass filtered into four broad frequency bands using 8th order Butterworth filters. The five corner frequencies (-3 dB down points) of the four bands were at 300 Hz, 713 Hz, 1509 Hz, 3043 Hz, and 6000 Hz. The envelope of the signal in each band was extracted by half-wave rectification and low-pass filtering (8th order Butterworth) with a 160 Hz cutoff frequency. The amplitude histogram in each band was computed for the test materials presented at 70 dB SPL. The maximum amplitude used (A_{\max}) was set to the 99th percentile of all amplitude levels in all channels, and the minimal amplitude (A_{\min}) was set to the noise floor in the absence of sound input in all channels. The current level (E) of electric stimulation in the i^{th} band was set to the acoustic envelope value (A) raised to a power (p) (Fu and Shannon, 1998). The exponent of the power function was systematically changed from 0.05 to 0.8. This transformed amplitude was used to modulate the amplitude of a continuous, 500 pulse/sec. biphasic pulse train with a 100 μ s/phase pulse duration. The stimulus order of the 4 channels was 1-3-2-4 for electrode pairs (16,22), (11,17), (6,12), and (1,7), respectively.

3. Results

Figure 1A and 1B show the mean and individual scores of vowel and consonant recognition as a function of the power function exponents in the quiet condition. The dotted lines show the individual scores from three listeners, and the solid line shows the mean scores from these three subjects. The vowels were consistently recognized at about 50% correct when the exponent of the power function was increased from 0.05 to 0.4 and dropped slightly to 41.7% as the exponent of the power function further increased to 0.8. Similarly, consonant recognition changed slightly from 70% when the value of the exponent of the amplitude mapping function was 0.05, to 73% when p was 0.2, and dropped to 46% when the exponent was increased to 0.8.

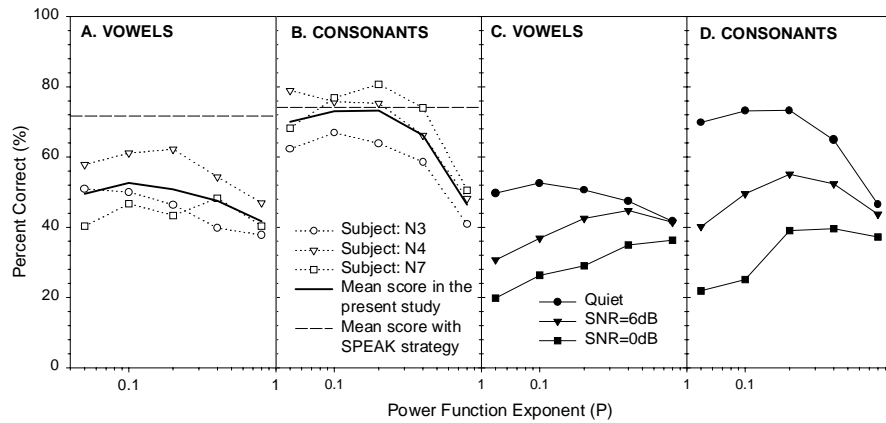


Fig. 1. Recognition scores of vowels and consonants as a function of the power function exponents in quiet and in noise. (A) Vowels (in quiet condition), (B) Consonants (in quiet condition); (C) Vowels (mean scores); (D) Consonants (mean scores).

Figures 1C and 1D show the mean vowel and consonant recognition scores as a function of the exponent of the power function in quiet and in noise. For weak compression ($p=0.8$), only a slight drop of speech performance was observed at both +6 dB SNR and 0 dB. However, a much larger reduction in performance was observed as the noise level increased for the strongly compressive conditions. When the exponent was 0.05, a 20% reduction occurred in vowel recognition and a 30% reduction in consonant recognition was observed going from quiet conditions to +6 dB SNR.

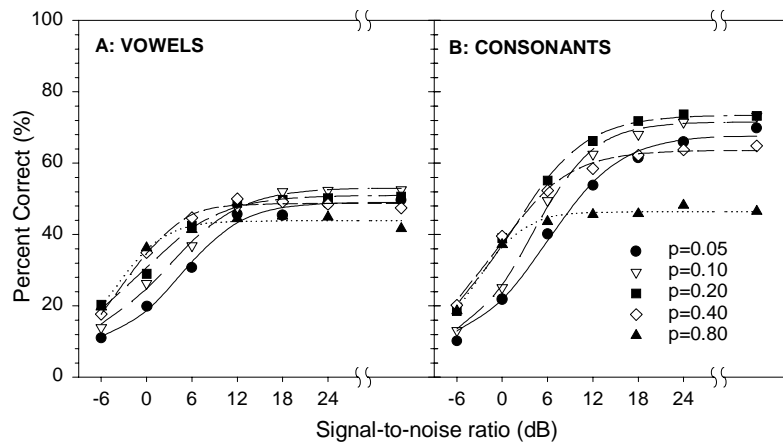


Fig. 2. Recognition scores of vowels and consonants as a function of signal-to-noise ratio. (A) Vowels; (B) Consonants. The lines represent the fitting curve based on a sigmoidal model.

Figures 2A and 2B show the mean scores of vowel and consonant recognition, respectively, as a function of S/N ratio with different amplitude mappings. Both vowel and consonant scores gradually increased as signal-to-noise (S/N) ratio increased for all mapping conditions. The phoneme recognition threshold (PRT) was defined as the S/N level that produced 50% of the performance level in quiet. The lines represent the best fit of a simple sigmoidal model (Fu et al., 1998; Fu and Shannon, 1999) with three parameters: PRT, the slope of the function at PRT (β), and the performance level in quiet.

Figure 3 shows the PRT and slope as a function of the power function exponents. The PRT for both vowels and consonants improved significantly as the mapping function changed

from a strong compression ($p=0.05$) to a weak compression ($p=0.8$). However, the slopes of the vowel and consonant functions at PRT were relatively constant.

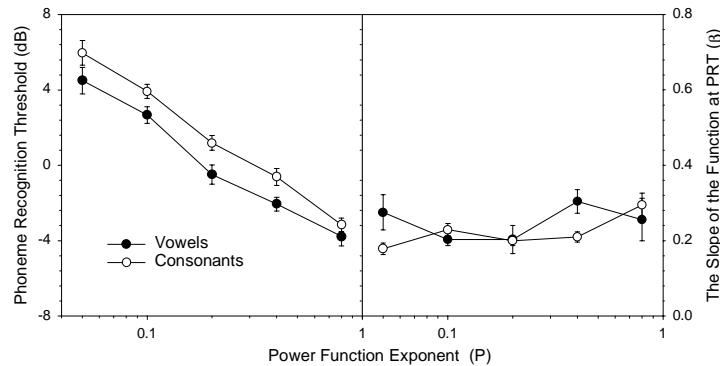


Fig. 3. Phoneme recognition threshold and the slope of vowel and consonant recognition as a function of the power function exponent. Error bars represent \pm one standard deviation.

4. Discussion and conclusions

The present results demonstrate that nonlinear acoustic-to-electric amplitude mappings have only a minor effect on phoneme recognition in quiet, and are consistent with the previous finding in normal-hearing listeners (Fu and Shannon, 1998). However, as the S/N level decreases, the effect of nonlinear amplitude mapping becomes dramatic and asymmetric: performance with weakly compressive mappings declines mildly in noise, but performance declines dramatically in noise with a strongly compressive amplitude mapping.

Two factors have possibly contributed to the differential effect of nonlinear amplitude mapping on speech performance in quiet and noise. One is the loudness-related effect (altered loudness growth function), and the other is the effective S/N ratio after compression. From an information transmission viewpoint, one would expect to obtain the highest vowel and consonant recognition scores when the normal loudness growth was preserved in electric stimulation, as demonstrated in the previous study (Fu and Shannon, 1998). The curve of performance in quiet could be used to roughly estimate the loudness-related effect. Vowel and consonant scores dropped only slightly when the amplitude mapping function was either more compressive or less compressive than the optimal mapping function, indicating that only a moderate degradation was caused by the loudness-related factor. Because the noise was introduced in addition to the loudness-related distortion, different amplitude mapping functions also result in a different effective S/N ratio at the output of the power function compressor. In other words, the effective S/N level will be dependent on the power function exponents. A linearly shifted PRT and a relatively constant slope as a function of power function exponents suggest that the differential effect of nonlinear amplitude mapping in noise was due to the change of the effective S/N ratio caused by compression.

One observation worth mentioning is the difference between the performance in the present study and the performance with the SPEAK strategy. Although the consonant scores for the 4-channel CIS strategy were comparable to those for the SPEAK strategy, the vowel scores for the SPEAK strategy (dashed line in Figure 1) were about 20% higher than the highest scores in Fig. 1 for the 4-channel CIS strategy. Two factors may contribute to this significant difference in vowel recognition. One is the processing strategy itself. The 20-electrode SPEAK strategy provides significantly better spectral resolution than the 4-channel CIS strategy (Fu and Shannon, 1998). The other possible factor is long-term learning because the present results are based on an "acute" study, where no adjustment time for the new speech processor was provided. It is quite possible that the difference will become smaller after significant exposure to the new speech processor with the 4-channel CIS strategy.

The present results also show an interesting similarity between vowel and consonant recognition. In quiet, neither vowel nor consonant recognition was strongly affected by amplitude compression, although consonant recognition did deteriorate more than vowel recognition for most linear mapping ($p=0.8$). Further analysis showed that performance on manner cues (Miller and Nicely, 1955) suffered most in this condition. The amplitude mapping exponent had a similar effect on the PRT for vowels and consonants (Figure 3).

The data in the present study showed that the PRT was highly dependent on amplitude mapping. This suggests that at least part of the large variability in performance across cochlear implant users may be due to nonoptimal amplitude mapping. Implant listeners who have an amplitude mapping function that is too compressive are at a disadvantage in noise compared with implant listeners with less compressive mappings. One implication of the results is that less compressive mappings may be better overall for mixed quiet and noisy conditions. Amplitude mappings that are more linear than the optimal loudness mapping may be slightly poorer in quiet conditions, but would still allow reasonably good speech recognition in noise. In contrast, strongly compressive mappings would allow a similar level of speech recognition in quiet, but would be considerably worse in noise.

Acknowledgments

We wish to thank Prof. James Hillenbrand for allowing us to use the multitalker vowel test materials. The research was supported by grants R03-DC-03861, R01-DC-01526, and contract N01-DC-92100 from NIDCD.

References and links:

- Boothroyd, A., Erickson, F.N., and Medwetsky, L. (1994). "The hearing aid input: a phonemic approach to assessing the spectral distribution of speech," *Ear and Hearing* **15**, 432-442.
- Dorman, W.F., Loizou, P.C., and Fitzke, J. (1998a). "The identification of speech in noise by cochlear implant patients and normal-hearing listeners using 6-channel signal processors," *Ear and Hearing* **19**, 481-484.
- Dorman, W.F., Loizou, P.C., Fitzke, J., and Tu, Z. (1998b). "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.* **104**, 3583-3585.
- Fu, Q.-J., and Shannon, R.V. (1998). "Effects of amplitude nonlinearity on speech recognition by cochlear implant users and normal-hearing listeners," *J. Acoust. Soc. Am.* **104**, 2571-2577.
- Fu, Q.-J., Shannon, R.V., and 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.
- Fu, Q.-J. and Shannon, R.V. (1999). "Recognition of spectrally degraded speech in noise with nonlinear amplitude mapping," *Proceedings of 1999 IEEE International Conference of Acoustics, Speech, and Signal Processing*, Vol. 1, 369-372.
- 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.
- Hochberg, I., Boothroyd, A., Weiss, M., Hellman, S. (1992). "Effects of noise and noise suppression on speech perception by cochlear implant users," *Ear and Hearing* **13**, 263-271.
- McDermott, H.J., McKay, C.M., and Vandali, A.E. (1992). A new portable sound processor for the University of Melbourne/Nucleus Limited Multichannel cochlear implant, *J. Acoust. Soc. Amer.*, **91**, 3367-3371.
- Miller, G. and Nicely, P. (1955). "An analysis of perceptual confusions among some English consonants." *J. Acoust. Soc. Am.* **27**, 338-352.
- Müller-Deiler, J., Schmidt, B.J., and Rudert, H. (1995). "Effects of noise on speech discrimination in cochlear implant patients," *Ann. Otol. Rhinol. Laryngol.* **166**, 303-306.
- Shannon, R.V., Adams, D.D., Ferrel, R.L., Palumbo, R.L., and Grantgenett, M. (1990). "A computer interface for psychophysical and speech research with the Nucleus cochlear implant," *J. Acoust. Soc. Am.* **87**, 905-907.
- Wilson. B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., & Rabinowitz, W. M. (1991). New levels of speech recognition with cochlear implants, *Nature*, **352**, 236-238.