Use of S-Shaped Input-Output Functions for Noise Suppression in Cochlear Implants

Kalyan Kasturi and Philipos C. Loizou

Objectives: **The aim of this study is to assess the influence of the shape of the acoustic-to-electric mapping function on speech recognition in noise by cochlear implant listeners.**

Design: **A new acoustic-to-electric mapping function is proposed for cochlear implant users in noisy environments. The proposed s-shaped mapping function was expansive for low input levels up to a knee point level and compressive thereafter. The knee point of the mapping functions changed dynamically and was set proportional to the estimated noise floor level. The performance of the mapping function was evaluated on a sentence recognition task using IEEE sentences embedded in +5 to 10 dB SNR multitalker babble and in +5 dB SNR speech-shaped noise. Nine postlingually deafened cochlear implant users participated in the study.**

Results: **Results indicated that the same s-shaped mapping function did not yield significant improvements for all cochlear implant users. Significant benefits in speech intelligibility were observed, however, when the s-shaped mapping function was optimized to individual cochlear implant users. Significantly higher performance was achieved with the s-shaped mapping functions than the conventional log mapping function used by cochlear implant users in their daily strategy, in both multitalker (+5 and +10 dB SNR) and continuous speech-shaped (+5 dB SNR) conditions.**

Conclusions: **These results clearly indicate that the shape of the nonlinear acoustic-to-electric mapping can have a significant effect on speech intelligibility in noise when it is optimized to individual cochlear implant users. The log functions currently used in most implant processors for mapping acoustic to electric amplitudes are not the best mapping functions to use in noisy environments. This is largely because compressive functions tend to amplify low-level segments of speech along with noise, thereby decreasing the spectral contrast and effective dynamic range. In contrast, the s-shaped mapping functions, which are partly compressive and partly expansive depending on the signal level, are more suitable for noisy environments and can produce significantly higher performance than the log-mapping functions.**

(Ear & Hearing 2007;28;402–411)

Department of Electrical Engineering, University of Texas at Dallas, Richardson, Texas.

The compression of envelope amplitudes is an essential component of cochlear implant processors because it is used in the transformation of acoustic amplitudes into electrical amplitudes. This transformation is necessary because the range in acoustic amplitudes in conversational speech is considerably larger than the cochlear implant user's electrical dynamic range. In conversational speech, the acoustic amplitudes may vary within a range of 30 to 50 dB (Boothroyd, Erickson & Medwetsky, 1994; Zeng, Grant, Niparko, Galvin, Shannon, Opie & Segel, 2002). Implant listeners, however, may have a dynamic range as small as +5 dB. For that reason, the cochlear implant processors compress, using a nonlinear compression function, the acoustic amplitudes to fit the user's electrical dynamic range. The logarithmic function is commonly used for compression because it matches the loudness between acoustic and electrical stimulation (Eddington, Dobelle, Brachman, Mldadevosky & Parkin, 1978; Zeng & Shannon, 1992), that is, it restores normal loudness growth.

The effect of the shape of the compression function on speech recognition has been investigated in a number of studies (Fu & Shannon, 1998, 1999; Loizou, Poroy & Dorman, 2000; Wilson, Lawson, Zerbi & Wolford, Reference Note 1; Zeng & Galvin, 1999). Loizou et al. (2000) modified the shape of the amplitude mapping functions from strongly compressive to weakly compressive (nearly linear) by varying the exponent of a power-law function. The shape of the compression function had only a minor effect on speech recognition in quiet, with the lowest performance obtained for nearly linear mapping functions. Similar findings were also reported by Zeng and Galvin (1999) and Fu & Shannon (1998, 1999). The effect of the shape of the compression function on speech recognition in noise was investigated by Fu & Shannon (1999). While a strongly compressive mapping between acoustic and electric amplitudes produced better performance in quiet, a less compressive mapping was sometimes more beneficial in noise.

The above studies considered only logarithmic-shape functions for mapping acoustic to electrical amplitudes. These functions are compressive for the most part, and as such, tend to amplify low-level signals. Use of compressive functions for transform-

0196/0202/07/2803-0402/0 • Ear & Hearing • Copyright © 2007 by Lippincott Williams & Wilkins • Printed in the U.S.A.

ing acoustic to electrical amplitudes ought to be beneficial in quiet as it renders soft sounds audible to cochlear implant users. It is therefore not surprising that cochlear implant users perform, at least in quiet, very well with logarithmic mapping functions. The situation in noise, however, is quite different. Compressive functions amplify both noise and weak speech segments making segregation of speech from noise extremely difficult. This assertion was corroborated by Fu & Shannon (1999), who found out that performance declines dramatically in noise with strongly compressive functions.

The benefits and limitations of compression have been studied extensively in the hearing-aid literature (see review in Hohmann & Kollmeier, 1995; Hornsby & Ricketts, 2001; Levitt, 2004; Yund & Buckles, 1995), but the conclusions have been less than clear. Some (e.g., Plomp, 1988) have argued that compression in general may be detrimental to speech intelligibility due to its effect on the modulation spectrum as quantified by the speech transmission index (STI). Others (Noordhoek & Drullman, 1997; Villchur, 1989) have argued that the reduction in intelligibility in noise can not be attributed to the reductions in envelope modulations alone but may occur because the weaker elements of speech are masked. Villchur (1973) showed that the weak elements of speech can be preserved to some degree with two-band compression. He further demonstrated that speech intelligibility in noise can be improved with two-band compression despite reduction in modulation depth caused by compression (Villchur, 1973).

Compared with the use of compression in hearing aids, considerably less work has been done to study the influence of the shape of the compression function on speech recognition in noise by cochlear implant listeners. In the present study, we examine the performance of a new amplitude mapping function which may be potentially more suitable for noisy environments than the conventional log compressive function. The proposed mapping function is partly compressive and partly expansive, and has the shape of the letter S; hence, we refer to it as the "s-shaped" function.

The motivation behind the use of s-shaped functions is to suppress the signal falling below the noise floor (and likely dominated by noise) while retaining the signal above the noise floor (and likely dominated by speech). This can be accomplished by the use of an expansive function for signal levels falling below the noise floor and a compressive function for signal levels falling above the noise floor level. The expansive segment of the function serves as a signal attenuator, as it suppresses low-amplitude signals, while the compressive segment serves as a signal amplifier, as it amplifies low-amplitude signals. The change from expansive to compressive function occurs at the knee point of the s-shaped input-output function. This knee point is not fixed, but rather adapted from cycle to cycle to the current estimate of the noise floor. Key to the application of the proposed mapping functions is the choice of the knee point, which in the present study was set to the noise floor level estimated using an algorithm. More specifically, a noise estimation algorithm (Rangachari & Loizou, 2006) is used to track continuously the noise floor in each band and adapt the knee-point accordingly. The influence of the placement of the knee-point as well as the shape of the function (compressive versus expansive) used for low input signal levels, is evaluated in the present study on recognition of speech in noise by cochlear implant users.

METHODS

Subjects

Nine cochlear implant recipients wearing the Clarion CII (Advanced Bionics Corporation) processor participated in this Experiment. All subjects were postlingually deafened adults and had used the cochlear implant for a minimum of 3 yr. The biographical data for the nine subjects is presented in Table 1. Subjects S3 and S8 were not available for testing in all conditions.

TABLE 1. Biographical data for the nine cochlear implant users tested

Subject	Sex	Age (yr)	Years of experience with cochlear implant	IEEE sentence recognition in quiet (% correct)	Pulse rate (pps/channel)	Etiology
S1	Female	49	4	96	2840	Otosclerosis
S2	Female	36	6	96	976	Unknown
S3	Female	59	3	87	2840	Medication
S4	Female	52	3	93	2840	Unknown
S5	Female	46	4	90	2840	Unknown
S6	Male	69	4	88	2840	Unknown
S7	Female	38	4	87	2840	Hereditary
S8	Female	58	4	59	976	Unknown
S9	Female	61	3	90	2840	Medication

Sentence Material

Subjects were tested on sentence recognition. The sentence material consisted of several lists of phonetically-balanced *IEEE* sentences (IEEE Subcommittee, 1969). Twenty sentences (two lists) were used per condition. Two types of noise were used to corrupt the IEEE sentences: speech-shaped noise and multitalker babble. The (continuous) speech-shaped noise was taken from the HINT database (Nilsson, Soli & Sullivan, 1994) and the multitalker babble (10 male and 10 female talkers) was taken from the Auditec CD (St Louis, Mo).

Signal Processing

Speech material was first band-pass filtered into 16 logarithmically-spaced frequency bands using sixth-order Butterworth filters that spanned the frequency range from 350 Hz to 5500 Hz. The output of each channel was passed through a full-wave rectifier followed by a second-order Butterworth low-pass filter with a cutoff frequency of 200 Hz to obtain the envelope of each channel. The channel envelope amplitudes were finally compressed according to various mapping functions (described next) and delivered to the electrodes using the subject's daily strategy. All subjects were fitted with a high-rate CIS strategy, based on the parameters (e.g., pulse width, pulse rate) used in their daily processor. Speech material was processed off-line and delivered directly to the cochlear implant users via the Clarion Research Interface II hardware (Loizou, Stickney, Mishra & Assmann, 2003).

The s-shaped mapping of acoustic amplitudes to electrical amplitudes involved two steps. The first step estimated the noise floor level using a noise-estimation algorithm. Note that unlike voice-activity detection algorithms, noise estimation algorithms track the noise amplitude continuously, even during speech-active segments. The second step constructed the s-shaped mapping function based on the noise-floor level estimated in the first step.

Estimating the Noise Envelope Amplitude • Since the characteristics of the two types of noise (continuous speech-shaped noise and multitalker babble) used in our study differ, we used two different methods for estimating the noise envelope. For the speech shaped noise, which is stationary, the noise envelope was computed using the initial 120 msec of speech-absent segment preceding the sentences. More specifically, the noise envelopes in each channel were computed by averaging, over the initial 120-msec segment, the envelopes computed via bandpass filtering and full-wave rectification. Due to the stationary nature of the noise, the same noise envelope amplitudes were used for all subsequent segments.

A different method was used for tracking the noise envelope of multitalker babble. This method was adapted from Rangachari & Loizou (2006) and is described in detail in Appendix A. The noise estimation algorithm tracks and updates the noise envelope continuously (i.e., in every cycle) taking into account the highly nonstationary nature of multitalker babble. This algorithm tends to underestimate the amplitude of the noise envelope, as it is based on minimum tracking (Rangachari & Loizou, 2006). For that reason, we applied a bias factor (>1) to the estimated noise envelope amplitude to artificially increase the noise floor level. More specifically, if $D(\lambda, k)$ is the noise envelope amplitude computed at time λ for channel k by the noise estimation algorithm (see Eq. (15) in Appendix A), then we used the following biased estimate of the noise envelope:

$$\widehat{D}(\lambda, k) = b \cdot D(\lambda, k) \tag{1}$$

where b is the bias factor, $D(\lambda, k)$ is the noise envelope amplitude estimated using Eq. (15) and $\widehat{D}(\lambda, k)$ is the biased estimate of the noise envelope amplitude. The bias factor is used as a parameter for controlling the amount of noise suppression applied. In our experiments, we considered three different values for b: $b = 1$, $b = 2$ and $b = 2.8$. Use of $b = 1$ sets the noise floor at a low value resulting in a relatively weak suppression of the noise. The other two values of b ($b = 2$ and $b = 2.8$) set the noise floor to a relatively higher value leading to more aggressive suppression of the noise. The range (1 to 2.8) of b values was based on pilot data collected with a few cochlear implant users.

Figure 1 shows an example of noise envelope estimation for a sentence corrupted in +10 dB SNR multitalker babble. Only the noise envelope amplitude of channel 3 (centered at 540 Hz) obtained with $b = 1$ in Eq. (1) is shown in this example. The true noise envelope is superimposed for comparative purposes. As can be seen, the noise estimation algorithm is capable of tracking, for the most part, sudden changes of the background noise envelope.

Construction of S-shaped Mapping Function • The estimated noise envelope amplitude $\widehat{D}(\lambda, k)$ [Eq. (1)] was subsequently used to construct the s-shaped mapping function. More precisely, the knee-point of the s-shaped function was set equal to the estimated noise envelope amplitude $\widehat{D}(\lambda, k)$. The proposed s-shaped mapping function is shown and compared with a standard log mapping function in Figure 2. It consists of the following compressive function for input levels larger than the knee-point:

$$y1 = A_1 x^{p1} + B_1 \tag{2}$$

where $p1 = -0.0001$ for logarithmic compression (Loizou et al., 2000), x is the acoustic amplitude and

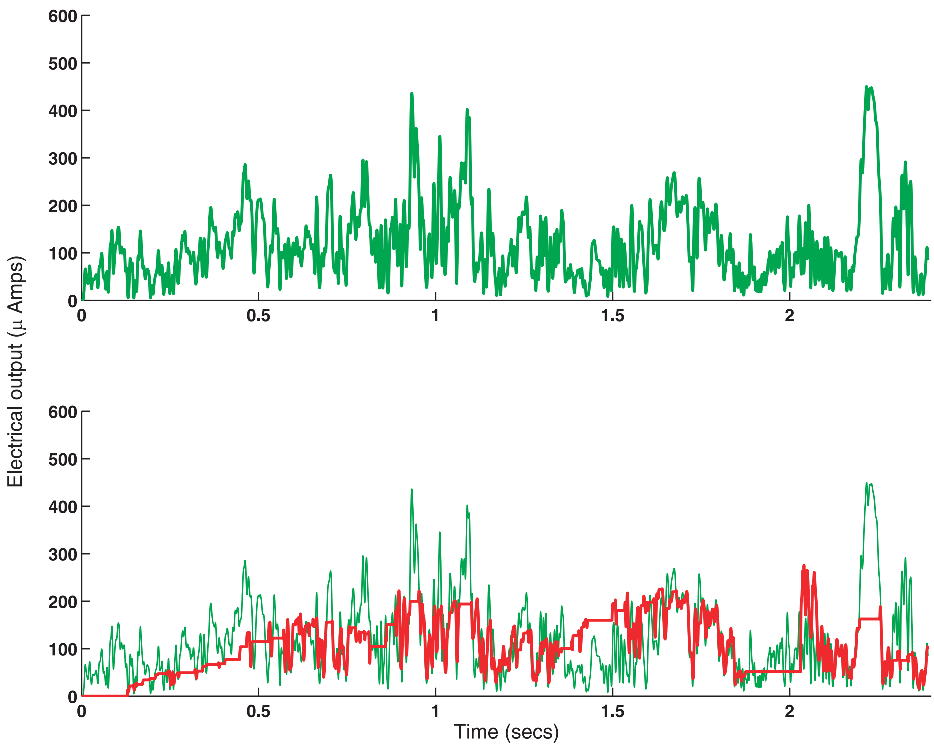

Fig. 1. Bottom panel shows the envelopes (thin lines) of noise (multitalker babble at +10 dB SNR) and the estimated noise envelopes (thick lines) obtained using a noise-estimation algorithm. For better clarity, the top panel shows the envelopes of the noise alone prior to adding it to a sentence at +10 dB SNR. The noise envelopes were extracted after bandpass filtering (channel 3, centered at 540 Hz), full-wave rectifying and low-pass filtering (200 Hz) the noise signal. The envelopes were subsequently mapped with a log function to the subject's electrical dynamic range. The estimated noise envelope (thick line, bottom panel) was computed using the algorithm given in Appendix A, with b = 1 as the bias factor.

y1 is the electrical amplitude. For input levels smaller than the knee-point, the following expansive function is used:

$$y2 = A_2 x^{p2} + B_2 \quad (3)$$

where p2 = 1.8 and the coefficients A_i, B_i are chosen to ensure that the input acoustic range [1, X_{max}] is mapped to the electrical dynamic range [THR, MCL], where X_{max} is the largest acoustic amplitude, *MCL* is the most comfortable loudness level and

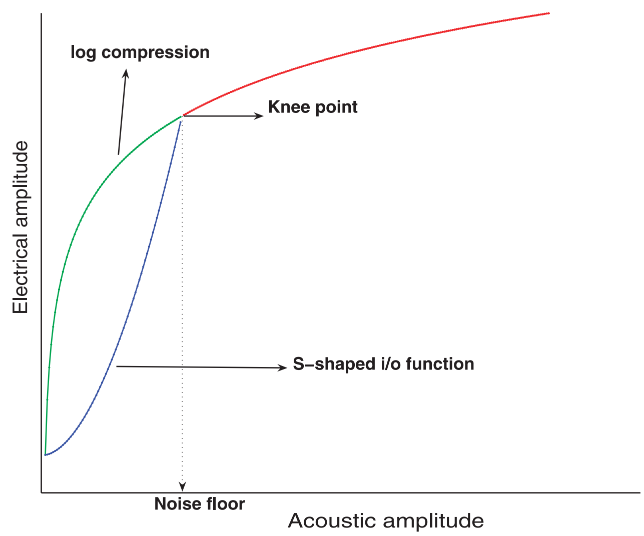

Fig. 2. The proposed s-shaped input-output function consisting of a log compressive function for input levels above the knee point and an expansive function for input levels below the knee point.

THR is the threshold level of electric hearing expressed in microamperes. The coefficients A_1, B_1, A_2 and B_2 are determined according to the following equations:

$$A_1 = (MCL - THR)/(X_{max}^{p1} - 1) \quad (4)$$

$$B_1 = (THR - A_1) \quad (5)$$

$$A_2 = (Y_{knee} - THR)/(K^{p2} - 1) \quad (6)$$

$$B_2 = (THR - A_2) \quad (7)$$

$$Y_{knee} = A_1 K^{p1} + B_1 \quad (8)$$

where K is the knee-point [set in our case to the noise floor level, that is, $K = \widehat{D}(\lambda, k)$ estimated as per Eq. (1)] and Y_{knee} is the electrical amplitude corresponding to the knee point (assuming log compression). It is interesting to note that the log-mapping function is a special case of the s-shaped function when the knee-point is set at $b = 0$.

Figure 3 shows examples of envelope amplitudes produced using the s-shaped mapping function ($b = 2$ and $b = 2.8$) and the log-mapping function for a segment of a sentence corrupted in +10 dB SNR multitalker babble. Only the envelope amplitudes of channel 3 (centered at 540 Hz) are shown in this example. It is clear that noise primarily affected the envelope valleys, with little effect on the envelope peaks. The s-shaped mapping functions suppressed the noise in the valleys while preserving the peaks of speech. Stronger noise suppression was achieved with $b = 2.8$ than with $b = 2$. Also, compared with the envelopes obtained with the log mapping function,

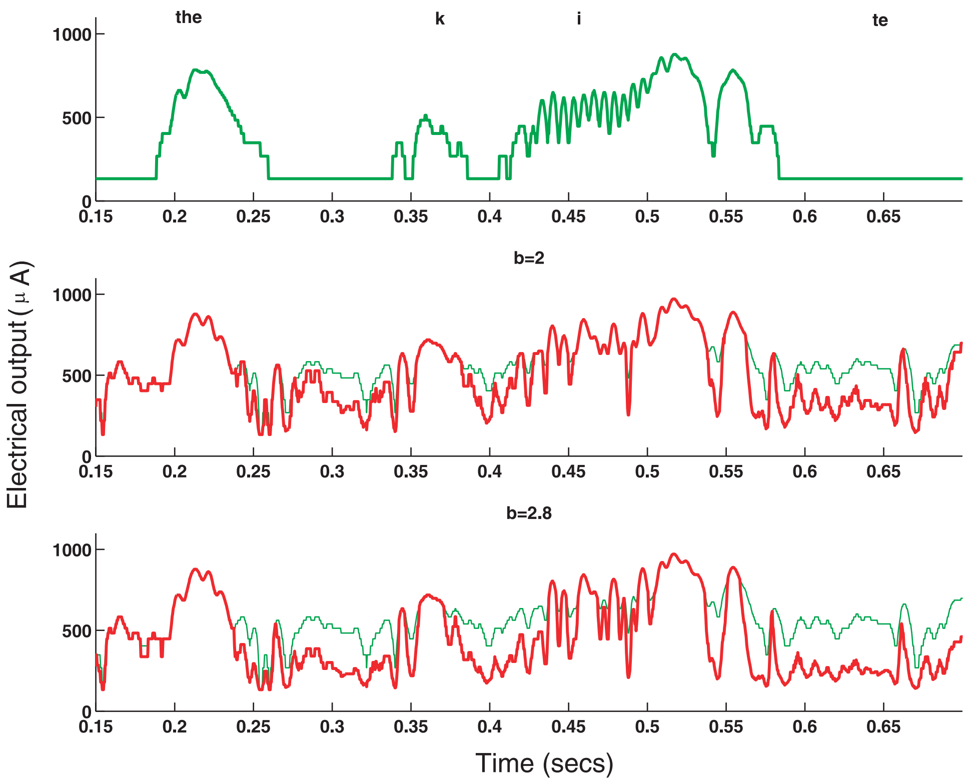

Fig. 3. Top panel shows the envelopes computed for channel 3 (centered at 540 Hz) for the segment "the kite" excerpted from the IEEE sentence "The kite flew wildly in the high wind." The log function was used for mapping the acoustic to electrical amplitudes. Bottom two panels show the envelopes extracted using the s-shaped input-output function (thick line) with b = 2 (middle panel) and b = 2.8 (bottom panel) for the same IEEE sentence corrupted in +10 dB multitalker babble. The envelope amplitudes obtained using the log mapping function (thin line) are overlaid for comparative purposes.

a better temporal envelope contrast was achieved with the s-shaped mapping functions.

Procedure

The cochlear implant subjects were tested with the Clarion research interface-II hardware using similar experimental setup as in Loizou et al. (2003). Sentences were processed off-line with the various s-shaped input-output functions and presented to the subjects directly at a comfortable level through the Clarion research interface-II. The comfortable level was adjusted separately for each condition by the implant users. The subjects were instructed to write down the words they heard. A practice session with ten sentences presented in quiet was used in the beginning. The practice session lasted for 5 to 10 minutes. After the practice session, the subjects were tested on the various conditions incorporating the different input-output functions. Speech-shaped noise was added to the IEEE sentences at +5 dB SNR, and multitalker babble was added to the sentences at +5 and +10 dB SNR. Speech-shaped noise was added only at +5 dB SNR, as we observed ceiling effects at +10 dB with most of our subjects. Three different values of the bias factor were considered: $b = 1$, $b = 2$ and $b = 2.8$, yielding three different knee-point conditions for the s-shaped input-output functions. The log mapping function used in the subject's daily processor was also tested as the control condition. The order of the various test conditions was randomized among subjects to avoid order effects. Twenty sentences were used for each test condition.

RESULTS

Performance was assessed in terms of percent of words identified correctly (all words were scored). The individual subjects' scores on sentence recognition are shown in Figure 4 for the s-shaped mapping functions constructed using three different values of the bias factor [Eq. (1)]. Two-way ANOVA (repeated measures) with SNR and location of knee-point (bias factor) as within subject factors indicated a nonsignificant effect of the location of knee-point [$F(3, 18) = 2.027$, $p = 0.146$], a significant effect of SNR [$F(1, 6) = 127.4$, $p < 0.0005$] and a nonsignificant interaction [$F(3, 18) = 2.206$, $p = 0.123$] between knee point and SNR. The scores obtained with the log-mapping function were also included in the ANOVA, as the log-mapping function can be considered as a special case of the s-shaped mapping function with the bias factor in Eq. (1) set to $b = 0$.

The above statistical analysis (ANOVA) indicated a nonsignificant effect of the location of knee-point of the s-shaped input-output function on speech intelligibility. This suggests that if we were to fit all cochlear implant users with the same s-shaped map-

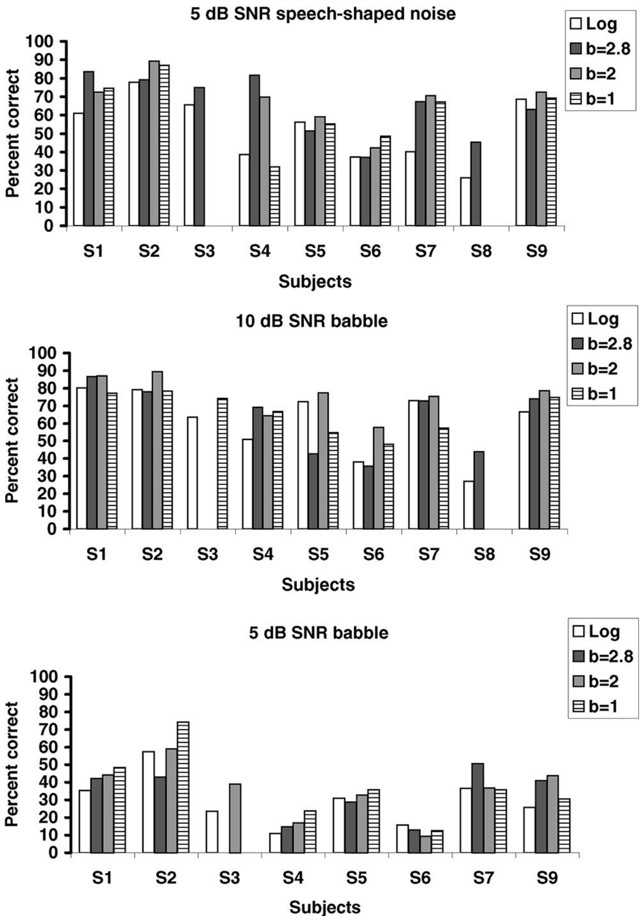

Fig. 4. Individual subject's scores (percent words identified correctly) for recognition of IEEE sentences embedded in +5 dB speech-shaped noise and +5 to 10 dB multitalker babble for two types of mapping functions, log and s-shaped. Three different values of the bias factor (b = 1, b = 2 and b = 2.8) were considered in the s-shaped mapping function. Subjects S3 and S8 were not tested in all conditions.

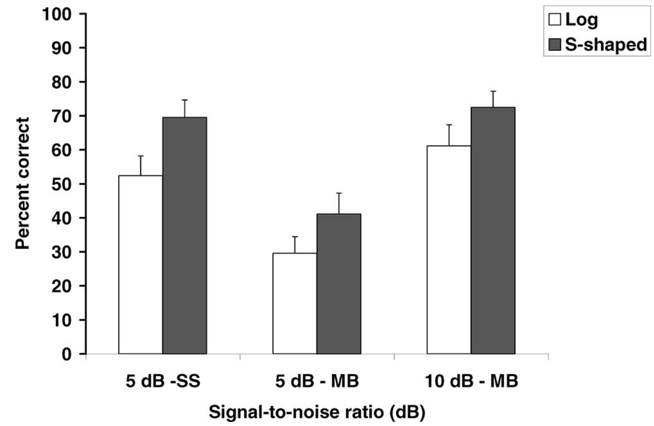

Fig. 5. Mean scores (percent words identified correctly) for recognition of IEEE sentences embedded in +5 dB speech-shaped (SS) noise and +5 to 10 dB multitalker babble (MB) for two types of mapping functions, log and s-shaped. These scores were computed by averaging the scores obtained with the "best" knee-point values (Fig. 4) by each cochlear implant user. Error bars indicate standard errors of the mean. The error bars reflect the variability in performance after optimizing the s-shaped function to individual cochlear implant users, and as such, might underestimate the true variability in performance with the various knee-point values.

ping function constructed using the same bias factor (i.e., same knee-point value), then we would not observe any benefit in speech intelligibility. While that is highly desirable in clinical applications (for ease of fitting), the above analysis does not take into account the inherent variability among cochlear implant users owing to differences in etiology, patterns of neuron survival, and so forth. The underlying peripheral physiological differences in cochlear implant users are partly responsible for the observed differences in performance. Indeed, from Figure 4 we observe that the individual subjects' performance varied across the three values of the knee point tested. The "best" knee-point value, in terms of intelligibility improvement, differed across subjects. The majority (6 of 9) of the cochlear implant users performed the best with $b = 2$ at +10 dB SNR (babble), but there was much variability among subjects in the "best" knee-point value for the +5 dB SNR babble and steady-state noise conditions.

The data in Figure 4 suggest that it might be possible to optimize the s-shaped function (in terms of the location of the knee-point) to individual cochlear implant users for better performance. While such a tuning can not be done, at this point, automatically via an algorithm or by using a psychophysical procedure, it can be done manually by trying out different knee-point values and choosing the value that works the best for each user. We refer to that value as the "best" knee-point value for a particular cochlear implant user. To determine whether such a tuning would yield statistically significant benefits to intelligibility, we subjected the scores obtained with the "best" knee point value (bias factor) for each cochlear implant user into a different ANOVA. We summarize in Figure 5 the mean scores obtained in the various SNR conditions for the two types of maskers, after selecting the best knee-point value of each subject. These scores are compared against those obtained with the log-mapping functions used in the subject's daily processors. The mean scores obtained with the optimized s-shaped mapping function were found to be significantly higher than the scores obtained with the log-mapping functions in +5 dB SNR speech shaped noise [$F(1, 8) = 15.85$, $p = 0.004$], in +5 dB SNR multitalker babble [$F(1, 7) = 21.89$, $p = 0.002$] and in +10 dB SNR multitalker babble [$F(1, 8) = 30.51$, $p = 0.001$].

These results clearly indicate that the shape of the nonlinear acoustic-to-electric mapping can be tuned to some degree for individual subjects to obtain maximum benefit on speech intelligibility in noise. After such tuning, higher performance can be achieved with s-shaped mapping functions than log mapping functions.

DISCUSSION

The results demonstrated that the shape of the nonlinear acoustic-to-electric mapping function can yield improved sentence recognition in noise when the knee values are optimized for individual listeners. The log functions currently used in most implant processors for mapping acoustic to electric amplitudes are not the best mapping functions for noisy environments. This is largely because compressive functions tend to amplify low-level segments of speech along with noise, thereby decreasing the spectral contrast and effective dynamic range. In contrast, the s-shaped mapping functions, which are partly compressive and partly expansive depending on the signal level, are more suitable for noisy environments and can produce significantly better performance than the log-mapping functions. Results from this study demonstrated that the choice of the knee-point in the s-shaped function is critical and that the optimal knee value (bias factor), in terms of highest speech recognition score, varies across subjects.

One factor has possibly contributed to the benefit of s-shaped mapping functions: improved spectral contrast. As can be seen in Figure 3, s-shaped processing preserves the envelope peaks and deepens the valleys, which are otherwise filled for the most part with noise. The s-shaped function utilizes *selectively* the benefits of the compression and expansion functions. When the signal level is substantially above the noise floor, i.e., during periods that speech is present (e.g., voiced segments), compression is at work preserving the peaks and strong elements of speech. When the speech signal is weak or absent, as during stop closures, the expansion is at work suppressing the background noise rather than boosting it. This selective application of compression and expansion lends itself to improved spectral contrast, as illustrated in Figure 3. This figure shows the envelopes obtained with s-shaped and log functions for the words "the kite" excerpted from the IEEE sentence "The kite flew wildly in the high wind." embedded in +10 dB SNR babble. Note that the noisy signal is severely attenuated (due to expansion) during the /k/ and /t/ stop closures thereby enhancing the peak-to-trough contrast. In contrast, the signal during the closures is amplified when log compression is used and the spectral contrast is subsequently reduced. More precisely, in this example, the peak-to-trough ratio improved from roughly 4 dB with the log compression function to 11 dB with the s-shaped function (see bottom panel in Figure 3, segments $t = 0.23$ to 0.33 sec and $t = 0.53$ to 0.6 sec). It is also worth noting here that the relatively weak stop consonant /k/ (segment $t = 0.35$ to 0.4) was not attenuated with the s-shaped function.

We speculate that the s-shaped functions improve spectral contrast without compromising loudness. While the use of linear mapping functions generally improves spectral contrast, it is done at the expense of reducing significantly the loudness of the acoustic stimuli thereby rendering most speech segments inaudible or not sufficiently loud. Several studies (e.g., Fu & Shannon, 1998) have confirmed that any dramatic deviation from a power-law (log type) compressive function will deteriorate performance. The s-shaped functions maintain the log mapping function for input levels above the estimated noise floor (knee-point). As such, they preserve the loudness of signals falling above the noise floor while reducing the loudness of signals falling below the knee-point and possibly dominated by noise.

S-shaped functions suppress the signal falling below the knee-point assumed to contain primarily noise. It is reasonable to ask why suppress and not annihilate (e.g., zero out) any signal falling below the knee point. One would expect that that would provide better suppression of the noise. We do not believe that is true for two main reasons. First, eliminating the signal falling below the knee-point (noise floor) would be a reasonable, and perhaps a better, approach provided that we are somehow capable of estimating the noise level (knee point) very accurately. That is not the case in practice, as the noise level constantly changes, and the best we can do is to estimate, at least conservatively, the noise floor level. Any errors in overestimating the noise floor level would wipe out segments of speech containing useful information thereby degrading intelligibility. Second, the frequent switching from signal-on to signal-off across and within each channel would produce undesirable distortion effects. Hence, suppressing rather than zeroing out the signal falling below the knee point seems to be a safer approach. There is also evidence from intelligibility studies (Drullman, 1995) suggesting that while removing noise from the envelope peaks has no effect on intelligibility, removing the speech signal from the noisy (envelope) valleys can produce a decrement in intelligibility. Hence, weak speech segments falling below the noise level do contribute to intelligibility and cochlear implant users are somehow able to utilize this information.

It is interesting to note that the input-output function of the normal cochlea is level-dependent much like the proposed s-shaped input-output function. It is known, for instance, that the compressive action of outer hair cells does not operate at very low (and very high) input levels (see review in Cooper, 2004; Ruggero, Rich, Recio, Narayan & Robles, 1997). More specifically, the basilar membrane's input-output function is linear at low intensity levels up to breakpoint level (knee-point), and becomes compressive thereafter. It is also linear at high input levels. This has prompted hearing aid companies to adopt similar input-output functions with linear mapping for low and extremely high input levels, and compressive mapping for intermediate (conversational level) levels (Dillon, 2001). Unlike the input-output functions used in hearing aids, which are based on fixed knee-point levels (around 50 dB SPL), the proposed input-output function changes its knee point adaptively, at least in constantly changing noise environments (e.g., multitalker babble).

The construction of the proposed acoustic-to-electric mapping functions depends on the choice of the knee-point, which is critically important. If the knee-point is overestimated, then weak consonants (e.g., /f/, /s/) might be heavily attenuated; possibly diminishing speech intelligibility. But, as shown in the examples in Figure 1 and 3, that was rarely the case. On the other hand, if the knee-point is underestimated, then too little noise will be suppressed yielding a reduced benefit in intelligibility. Clearly, the effect of underestimating the knee-point is less severe as the resulting mapping function resembles that of the log function used in the user's daily processor. In our study, the knee-point was adjusted according to the estimate of the noise floor (see Fig. 2), hence, much of the success of the s-shaped mapping function can be attributed to the relatively accurate estimates of the knee-point (noise floor). We expect larger improvements in performance with more accurate estimates of the noise floor level, and therefore better estimates of the knee point. Further research is therefore needed to improve the accuracy and speed (e.g., adaptation time) of noise tracking algorithms.

Comparing the performance of cochlear implant users on the two types of maskers used (steady-state speech-shaped noise and multitalker babble), we noted that implant listeners performed better with the steady-state masker than the rapidly fluctuating babble masker. This was true despite the fact that the temporal characteristics of the multitalker babble masker are similar to those of steady-state noise in that no temporal dips are available in the multitalker babble waveform that would allow the listeners to "glimpse" the target speech. Previous studies (Qin and Oxenham, 2003; Stickney, Zeng, Litovksy & Assmann, 2004) have shown that unlike normal-hearing listeners, cochlear implant users perform better in steady-state noise than in the presence of a single competing talker. In stark contrast to normal-hearing listeners, implant listeners do not obtain a release of masking in the presence of a competing talker. The findings of the present study extend the findings of Stickney et al. (2004) and support the view that cochlear implant users can segregate competing sounds more easily if the masker is steady-state noise than competing talkers.

CONCLUSIONS

The present study assessed the performance of a new nonlinear acoustic-to-electrical mapping function on speech recognition in noise by cochlear implant users. More specifically, the performance of an s-shaped mapping function, which was expansive for low input levels up to a knee point and compressive thereafter, was compared with the performance of the log-mapping function currently used in the subject's daily processors. Results indicated that the same s-shaped mapping function did not yield significant improvements for all cochlear implant users. Significant benefits in speech intelligibility were observed, however, when the s-shaped mapping function was tuned to individual cochlear implant users. Following the tuning, the s-shaped mapping function, produced significantly higher scores than the log-mapping function, in all noise conditions (+5 dB SNR speech-shaped noise and +5 to 10 dB SNR multitalker babble).

APPENDIX A

This Appendix describes the noise estimation algorithm used in our studies. This algorithm was adapted from Rangachari & Loizou (2006).

The smoothed envelope of noisy speech is first computed using the following first-order recursive equation:

$$P(\lambda, k) = \eta P(\lambda - 1, k) + (1 - \eta) Y(\lambda, k) \qquad (9)$$

where $P(\lambda, k)$ is the smoothed envelope, λ is the time index, k is the channel index, $Y(\lambda, k)$ is the envelope of the noisy speech (extracted through bandpass filtering, rectification and low-pass filtering) and η is a smoothing constant. Next, a nonlinear rule is used to track the minimum of the noisy speech envelope by continuously averaging past envelope values:

If $P_{\min}(\lambda-1, k) < P(\lambda, k)$ then

$$P_{\min}(\lambda, k) = \gamma P_{\min}(\lambda - 1, k) + \frac{1-\gamma}{1-\beta}(P(\lambda, k) - 0.96P(\lambda - 1, k)) \quad (10)$$

else

$$P_{\min}(\lambda, k) = P(\lambda, k)$$

end

where $P_{\min}(\lambda, \mathrm{k})$ is the local minimum of the noisy speech envelope and β and γ are constants determined experimentally. The look-ahead factor β controls the adaptation time of the local minimum.

After the minimum tracking, the next step is to determine speech presence in each channel. Let the ratio of the smoothed noisy envelope and its local minimum be defined as:

$$S_r(\lambda, k) = P(\lambda, k)/P_{\min}(\lambda, k) \quad (11)$$

If the above ratio is found to be greater than a preset threshold, it is taken as a channel containing speech otherwise it is taken as a channel containing either noise or insignificant amounts of speech energy. This is based on the principle that the envelope of the noisy speech will be nearly equal to its local minimum when speech is absent. Hence the smaller the ratio is in Eq. (11), the higher the probability that it will be a noise-only region and vice versa. The speech-presence decision made for each channel at time index λ can be summarized as follows:

If $S_r(\lambda, k) > \delta$

$$I(\lambda, k) = 1 \quad \text{speech present} \quad (12)$$

else

$$I(\lambda, k) = 0 \quad \text{speech absent}$$

end

where δ is a threshold determined experimentally. From the above rule, a so-called speech-presence probability, $p(\lambda, k)$, is updated using the following first-order recursion:

$$p(\lambda, k) = \alpha_p p(\lambda - 1, k) + (1 - \alpha_p)I(\lambda, k) \quad (13)$$

where α_p is a smoothing constant. Note that the above recursion implicitly exploits the correlation for speech presence in adjacent frames. Using the above speech-presence probability estimate, we compute the time-frequency dependent smoothing factor as follows:

$$\alpha_s(\lambda, k) \triangleq \alpha_d + (1 - \alpha_d)p(\lambda, k) \quad (14)$$

where α_d is a constant. Note that $\alpha_s(\lambda, k)$ takes values in the range of $\alpha_d \leq \alpha_s(\lambda, k) \leq 1$.

Finally, after computing the frequency-dependent smoothing factor $\alpha_s(\lambda, k)$ using Eq. (14), we update the noise envelope amplitude as follows:

$$D(\lambda, k) = \alpha_s(\lambda, k)D(\lambda - 1, k) + (1 - \alpha_s(\lambda, k))Y(\lambda, k) \quad (15)$$

where $D(\lambda, k)$ is the estimate of the noise envelope of the kth channel at time index λ.

The overall algorithm can be summarized as follows. After classifying the individual channels into speech present/absent using Eq. (12), we update the speech presence probability using Eq. (13) and then use this probability to update the time-frequency dependent smoothing factor in Eq. (14). The noise envelope estimate is finally updated according to Eq. (15) using a time-frequency dependent smoothing factor.

The following smoothing constants and parameters were used in our implementation of Equations (9)–(15): $\eta = 0.5$, $\alpha_p = 0.5$, $\alpha_d = 0.8$, $\gamma = 0.998$, $\beta = 0.5$, $\delta = 12$.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Gail Donaldson, Dr. Fan-Gang Zeng, and two anonymous reviewers for providing valuable feedback to this manuscript. This research was supported by grant R01 DC007527 from the National Institute of Deafness and other Communication Disorders, NIH.

Address for correspondence: Dr. Philipos C. Loizou, Department of Electrical Engineering, University of Texas at Dallas, PO Box 830688, EC 33, Richardson, TX 75083-0688. E-mail: loizou@utdallas.edu.

Received June 5, 2006; accepted January 21, 2007.

REFERENCES

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

Cooper, N. (2004). Compression in the peripheral auditory system, in *Compression: From cochlea to cochlear implants*, (S Bacon, R Fay, & A Popper, eds.), New York, Springer Verlag, 18–61.

Dillon, H. (2001). *Hearing Aids* NY, Thieme.

Drullman, R. (1995). Speech intelligibility in noise: relative contribution of speech elements above and below the speech level. *Journal of the Acoustical Society of America*, *98*, 1796–1798.

Eddington, D., Dobelle, W., Brachman, D., Mladevosky, M., Parkin, J. (1978). Auditory prosthesis research using multiple intracochlear stimulation in man. *Annals of Otology, Rhinology and Laryngology*, *87*, 1–39.

Fu, Q.-J., Shannon, R. (1998). Effect of amplitude nonlinearity on phoneme recognition by cochlear implant users and normal-hearing listeners. *Journal of the Acoustical Society of America*, *104*, 2570–2577.

Fu, Q.-J., Shannon, R. (1999). Phoneme recognition by cochlear implant users as a function of signal-to-noise ratio and nonlinear amplitude mapping. *Journal of the Acoustical Society of America*, *106*, L18–L23.

Hohmann, V., Kollmeier, B. (1995). The effect of multichannel dynamic compression on speech intelligibility. *Journal of the Acoustical Society of America*, *97*, 1191–1195.

Hornsby, B., Ricketts, T. (2001). The effects of compression ratio, signal-to-noise ratio, and level on speech recognition in normal-hearing listeners. *Journal of the Acoustical Society of America*, *109*, 2964–2973.

IEEE Subcommittee (1969). IEEE Recommended Practice for Speech Quality Measurements, *IEEE Trans.Audio and Electroacoustics AU-17*, 225–246.

Levitt, H. (2004). Compression Amplification, in *Compression: From cochlea to cochlear implants.* (S Bacon, R Fay, and A Popper , eds.), New York, Springer Verlag, 153–183.

Loizou, P., Poroy, O., Dorman, M. (2000). The effect of parametric variations of cochlear implant processors on speech understanding. *Journal of the Acoustical Society of America*, *108*, 790–802.

Loizou, P., Stickney, G., Mishra, L., Assmann, P. (2003). Comparison of speech processing strategies used in the Clarion implant processor. *Ear and Hearing*, *24*, 12–19.

Nilsson, M., Soli, S., Sullivan, J. (1994). Development of Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America*, *95*, 1085–1099.

Noordhoek, I., Drullman, R. (1997). Effect of reducing temporal intensity modulations on sentence intelligibility. *Journal of the Acoustical Society of America*, *101*, 498–502.

Plomp, R. (1988). The negative effect of amplitude compression in multichannel hearing aids in the light of the modulation-transfer function. *Journal of the Acoustical Society of America*, *83*, 2322–2327.

Qin, M., Oxenham, A. (2003). Effects of simulated cochlear implant processing on speech reception in fluctuating maskers. *Journal of the Acoustical Society of America*, *114*, 446–454.

Rangachari, S., Loizou, P. (2006). A noise estimation algorithm for highly non-stationary environments. *Speech Communication*, *28*, 220–231.

Ruggero, M., Rich, N., Recio, A., Narayan, S., Robles, L. (1997). Basilar-membrane responses to tones at the base of the chinchilla cochlea. *Journal of the Acoustical Society of America*, *101*, 2151–2163.

Stickney, G., Zeng, F., Litovsky, R., Assmann, P. (2004). Cochlear implant speech recognition with speech maskers. *Journal of the Acoustical Society of America*, *116*, 1081–1091.

Villchur, E. (1973). Signal processing to improve speech intelligibility in perceptive deafness. *Journal of the Acoustical Society of America*, *53*, 1646–1657.

Villchur, E. (1989). Comments on "The negative effect of amplitude compression in multichannel hearing aids in the light of the modulation-transfer function" [J. Acoust. Soc. Am. 83, 2322–2327 (1988)]. *Journal of the Acoustical Society of America*, *86*, 425–427.

Yund, E., Buckles, K. (1995). Enhanced speech perception at low signal-to-noise ratios with multichannel compression hearing aids. *Journal of the Acoustical Society of America*, *97*, 1224–1240.

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

Zeng, F.-G., Grant, G., Niparko, J., Galvin, J., Shannon, R., Opie, J., Segel, P. (2002). Speech dynamic range and its effect on cochlear implant performance. *Journal of the Acoustical Society of America*, *111*, 377–386.

Zeng, F.-G., Shannon, R. (1992). Loudness balance between acoustic and electric stimulation. *Hearing Research*, *60*, 231–235.

REFERENCE NOTE

1. Wilson B., Lawson D., Zerbi M., Wolford R. (1999). Speech Processors for Auditory Prostheses, NIH Project N01-DC-8-2105, Third Quarterly Progress Report.