The impact of noise power estimation on speech intelligibility in cochlear-implant speech coding strategies

Bentsen, Thomas; Mauger, Stefan J.; Kressner, Abigail Anne; May, Tobias; Dau, Torsten

Published in:
Journal of the Acoustical Society of America

Link to article, DOI:
10.1121/1.5089887

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

The impact of noise power estimation on speech intelligibility in cochlear-implant speech coding strategies (L)

Thomas Bentsen,[1] Stefan J. Mauger,[2] Abigail A. Kressner,[1,a)] Tobias May,[1] and Torsten Dau[1]

[1]*Hearing Systems Group, Department of Electrical Engineering, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark*

[2]*Cochlear Limited, Level 1, 174 Victoria Parade, East Melbourne VIC 3002, Australia*

(Received 26 August 2018; revised 17 January 2019; accepted 18 January 2019; published online 13 February 2019)

The advanced combination encoder (ACE^{TM}) is an established speech-coding strategy in cochlear-implant processing that selects a number of frequency channels based on amplitudes. However, speech intelligibility outcomes with this strategy are limited in noisy conditions. To improve speech intelligibility, either noise-dominant channels can be attenuated prior to ACE^{TM} with noise reduction or, alternatively, channels can be selected based on estimated signal-to-noise ratios. A noise power estimation stage is, therefore, required. This study investigated the impact of noise power estimation in noise-reduction and channel-selection strategies. Results imply that estimation with improved noise-tracking capabilities does not necessarily translate into increased speech intelligibility.
https://doi.org/10.1121/1.5089887

[ICB] Pages: 818–821

I. INTRODUCTION

In cochlear implant (CI) processing, a signal is decomposed into frequency channels and the signal level is used to determine the electrode stimulation intensity. In one frequently used coding strategy in devices from the manufacturer Cochlear, the advanced combination encoder (ACE^{TM}), a fixed number of channels with the largest amplitudes are selected for electrical stimulation (McDermott *et al.*, 1992; Wilson *et al.*, 1988). However, speech intelligibility outcomes with ACE^{TM} in noisy conditions with low signal-to-noise ratios (SNRs) are limited primarily because: (i) the channels with the largest amplitudes can be noise-dominated instead of speech-dominated and (ii) ACE^{TM} always selects a fixed number of channels when the signal amplitude is above a predefined threshold, irrespective of whether speech is present or absent (Hu and Loizou, 2008). In an attempt to improve the speech intelligibility in these noisy conditions, a range of different speech-coding strategies have been developed.

One group of strategies applies noise reduction prior to coding (e.g., using ACE^{TM}). Specifically, a noise power spectral density (PSD) estimate is obtained and noise-dominant channels are attenuated before the channels with the largest amplitudes are selected for stimulation. In current Cochlear-manufactured CI processors (Dawson *et al.*, 2011; Mauger *et al.*, 2012b), noise PSD estimation is based on minimum statistics (MS), where the estimate is obtained by tracking the minimum of the noisy speech PSD in a time window that typically spans over 1–3 s (Martin, 2001). Substantial speech intelligibility improvements have been demonstrated in speech-weighted noise with noise reduction based on MS-based estimators over ACE^{TM}, but the strategy failed to improve speech intelligibility in the presence of four competing talkers (Mauger *et al.*, 2012a). This may be, at least partly, because the MS-based estimator tracks changes in fluctuating noises with a delay corresponding to the duration of the time window. Since the noise PSD estimate is determined by the minimum within the time window, this can lead to an underestimation of the true noise PSD. To overcome the limitations of the MS-based estimator, other noise PSD estimators (Cohen, 2003; Cohen and Berdugo, 2002; Gerkmann and Hendriks, 2012) have been introduced and evaluated in noise-reduction strategies in CI recipients (Baumgärtel *et al.*, 2015; Hu *et al.*, 2007; Mauger *et al.*, 2012a). Specifically, Gerkmann and Hendriks (2012) proposed a noise PSD estimator based on the speech presence probability (SPP). This noise PSD estimator has been shown to track changes in the true noise PSD faster than the MS-based estimator and has been reported to be more accurate than the MS estimator in terms of the logarithmic estimation error. The present study compared these two noise PSD estimators in the context of noise reduction, and specifically investigated whether an improved accuracy (in terms of logarithmic estimation error) can translate into higher speech intelligibility.

Another group of strategies selects which channels to stimulate directly based on an SNR criterion (Hu and Loizou, 2008). A frequency channel with a high instantaneous SNR conveys more reliable speech information than a frequency channel with a low instantaneous SNR, and only channels with high SNRs are therefore selected for stimulation. One approach is to select the *N*-of-*M* channels with the highest SNRs. This fixed channel-selection strategy is similar to ACE^{TM}, except that the channel-selection criterion has changed from amplitude to SNR. Alternatively, a channel is selected only if the SNR is above a local criterion (LC)

a)Electronic mail: aakress@dtu.dk

(Hu and Loizou, 2008). The number of selected channels therefore change adaptively with the SNR, such that in each processing cycle between 0 and M channels are stimulated. With this latter approach, together with *a priori* information of the clean speech and the noise signals to derive the SNR, speech intelligibility has been restored to levels obtained for speech in quiet for both speech-weighted noise and multi-talker babble (Dawson *et al.*, 2011; Hazrati and Loizou, 2013; Hu and Loizou, 2008). However, to apply these channel-selection strategies in practice, an SNR estimation algorithm is required. Given the higher accuracy of the SPP-based noise PSD estimator as compared to the MS-based estimator, the algorithm appears to be a promising candidate for this task.

The present study investigated the impact of the SPP-based estimator in a range of noise-reduction and channel-selection strategies on the speech intelligibility outcome in CI recipients. First, the SPP-based estimator was implemented in a noise-reduction strategy, and intelligibility scores were compared to those obtained with the MS-based estimator. Second, the estimated SNRs were used in both fixed and adaptive channel-selection strategies, and intelligibility scores were compared with intelligibility scores obtained with ACE™, as well as with the existing noise-reduction strategy in combination with ACE™. With this second set of comparisons, the impact of altering the channel-selection criterion was investigated. At the same time, the relative impact of altering the SNR-based channel selection from fixed to adaptive was evaluated.

II. METHODS

A. Estimation of noise power and SNR

Noisy speech was sampled at 16 kHz and buffered into $\ell = 1,\ldots,L$ frames of 8 ms duration with 1 ms step size. A short-time discrete Fourier transform with $k = 1,\ldots,K$ bins ($K = 128$) decomposed the noisy speech in the signal path. The noise PSD estimate, $\hat{\sigma}_{N,k}(\ell)$, was obtained using the MS-based and the SPP-based algorithm for each individual bin k and time frame ℓ, given the noisy speech observation, $Y_k(\ell)$ (Gerkmann and Hendriks, 2012; Martin, 2001). The noise PSD estimates were combined into $m = 1,\ldots,M$ non-overlapping auditory CI channels spaced between 245 Hz and 7279 Hz ($M = 22$), $\hat{\sigma}_{N,m}(\ell)$, and the estimated SNR, $\hat{\xi}_m(\ell)$, was computed for each CI channel:

$$\hat{\xi}_m(\ell) = \frac{|Y_m(\ell)|^2}{\hat{\sigma}_{N,m}(\ell)^2} - 1. \tag{1}$$

Finally, the estimated SNRs were then recursively smoothed across time using a time constant of 8 ms.

B. The speech coding strategies

The estimated SNRs were utilized in speech coding strategies. In the noise-reduction strategies, called "NR-MS & ACE™" and "NR-SPP & ACE™," a set of gain values were computed from a Wiener gain function optimized for CI recipients (Mauger *et al.*, 2012b). In the fixed channel-selection strategy, called "CS-SPP-FIXED," estimated SNRs were used to select the N-of-M channels with the highest SNRs. In the adaptive channel-selection strategy, called "CS-SPP-ADAPTIVE," an LC of 0 dB was first applied to the SNRs to determine which channels were speech-dominated and therefore candidates for stimulation. In order to keep the stimulation rate the same as in the CI recipients' everyday mapping, only up to N of the channels with the largest amplitudes were then stimulated in each cycle, where N is the number of maxima selected for ACE™ in each recipients' default map. To quantify the noise PSD estimation accuracy, the logarithmic estimation error was adopted (Hendriks *et al.*, 2008) across time frames ℓ and frequency channels m:

$$\begin{aligned}\mathrm{LogErr} = {} & \frac{10}{LM}\sum_{\ell=1}^{L}\sum_{m=1}^{M}\left|\min\left(0,\ \log_{10}\frac{\sigma^2_{N,m}(\ell)}{\hat{\sigma}^2_{N,m}(\ell)}\right)\right| \\ & + \frac{10}{LM}\sum_{\ell=1}^{L}\sum_{m=1}^{M}\max\left(0,\ \log_{10}\frac{\sigma^2_{N,m}(\ell)}{\hat{\sigma}^2_{N,m}(\ell)}\right).\end{aligned} \tag{2}$$

The logarithmic estimation error was computed for 10 sentences from a randomly chosen list from the Bamford-Kowal-Bench (BKB)-like corpus (Bench *et al.*, 1979) mixed with multi-talker babble from 20 talkers (Mauger *et al.*, 2012a) at 0 dB and 5 dB SNR. A linear mixed effect model was constructed to quantify the difference in logarithmic estimation error between the two noise PSD estimators.

C. Study design

The subjects participated in two sessions, and in each session four different strategies were tested. In Session 1, the strategies ACE™, NR-MS & ACE™, CS-SPP-FIXED, and CS-SPP-ADAPTIVE were tested in speech-weighted noise to compare the channel-selection strategies with existing speech-coding strategies, as well as to assess the impact of altering the SNR-based channel selection from fixed to adaptive. In Session 2, ACE™, NR-MS & ACE™, NR-SPP & ACE™ and the best performing SNR-based channel-selection strategy of the two in Session 1 were tested in the multi-talker babble condition. In particular, Session 2 investigated if an improved accuracy of the noise PSD estimator translates into higher speech intelligibility in the context of noise reduction.

D. Hardware and procedure

The strategies were implemented with Simulink in a real-time system developed by Cochlear Limited. BKB-like sentences from a female speaker were mixed with noise, and the corrupted sentences were presented at 0 deg azimuth 1.2 m in front of the recipients at 65 dB sound pressure level via a loudspeaker in a sound isolated booth. Twelve CI recipients participated in the study. The subjects were native speakers of Australian English, and the age spanned from 37 to 85 yr with a median age of approximately 69 yr. The CI usage time ranged from 1 to 13 yr with a median of 8 yr. All but one subject were stimulated with $N = 8$ maxima out of

$M = 22$ electrodes while the remaining subject was stimulated with $N = 12$ maxima. The subjects were tested with an adaptive speech reception threshold (SRT) task (Dawson *et al.*, 2013). Each strategy was evaluated with two runs, and the test order was counterbalanced within the session and randomized across subjects. A linear mixed effect model was constructed for the SRTs from each session.

III. RESULTS

A. Evaluation of the noise-reduction strategies

Prior to the evaluation, the least square mean of the logarithmic estimation error was computed for the MS-based (2.9 dB) and for the SPP-based noise PSD estimator (1.8 dB). The improvement of the SPP-based relative to the MS-based estimator (1.1 dB; $p < 0.0001$) is consistent with the literature for similar conditions (Gerkmann and Hendriks, 2012). Figure 1 shows measured SRTs in speech-weighted noise in Session 1 [Fig. 1(a)] and in multi-talker babble in Session 2 [Fig. 1(b)]. No statistically significant difference between the NR-MS & ACE^{TM} and the NR-SPP & ACE^{TM} strategies was observed in multi-talker babble. The results therefore suggest that speech intelligibility does not improve significantly with the more accurate SPP-based estimator relative to the MS-based estimator. Finally, in speech-weighted noise the existing noise-reduction strategy (NR-MS & ACE^{TM}) improved the SRT compared to ACE^{TM} alone by about 1.6 dB ($p < 0.01$) [Fig. 1(a)], which is consistent with previously reported findings (Dawson *et al.*, 2011; Mauger *et al.*, 2012a; Mauger *et al.*, 2012b).

B. Evaluation of the channel-selection strategies

The fixed and the adaptive channel-selection strategies were first compared and are shown in Fig. 1(a). The CS-SPP-ADAPTIVE strategy was found to decrease the mean SRT scores by 1.63 dB as compared to the CS-SPP-FIXED strategy ($p < 0.01$). The adaptively changing channel selection therefore improved the speech intelligibility relative to the fixed channel selection in the CI recipients. However, when comparing the CS-SPP-ADAPTIVE strategy with the ACE^{TM} strategy, there was no significant difference in mean SRT in speech-weighted noise. Moreover, the SRT increased by 1.53 dB ($p < 0.0001$), i.e., speech intelligibility was worse, with the CS-SPP-ADAPTIVE strategy in the presence of multi-talker babble [see Fig. 1(b)]. Therefore, neither of the two SNR-based channel-selection strategies improved speech intelligibility relative to ACE^{TM}.

IV. DISCUSSION AND CONCLUSION

The current study confirmed findings in Gerkmann and Hendriks (2012) that the SPP-based estimator is more accurate in tracking the true noise PSD than the MS-based estimator in the multi-talker babble condition in terms of the logarithmic estimation error. Nevertheless, the results from the listening experiment showed that the improved accuracy in noise PSD estimation does not translate into an increase in measured speech intelligibility. Two points may help explain this observation. First, the SPP-based noise PSD estimate changed more rapidly over time, and the gain values therefore also varied more quickly over time. The CI recipients are accustomed to a more slowly changing noise-reduction strategy (NR-MS & ACE^{TM}), since this noise-reduction strategy is integrated in the participants' everyday sound processors and has most likely been used on a daily basis for many years. A lack of familiarity with the SPP-based noise-reduction strategy may thus have affected the results. Second, the logarithmic estimation error does not indicate for which time frames and frequency channels a noise PSD estimator is tracking the true noise PSD with high accuracy, i.e., whether the accuracy is high when speech is present or absent. The results therefore suggest that the logarithmic estimation error is not a good predictor of the speech intelligibility outcome.

Neither of the channel-selection strategies improved the speech intelligibility relative to the well-established ACE^{TM} strategy. There may be three possible explanations for this.

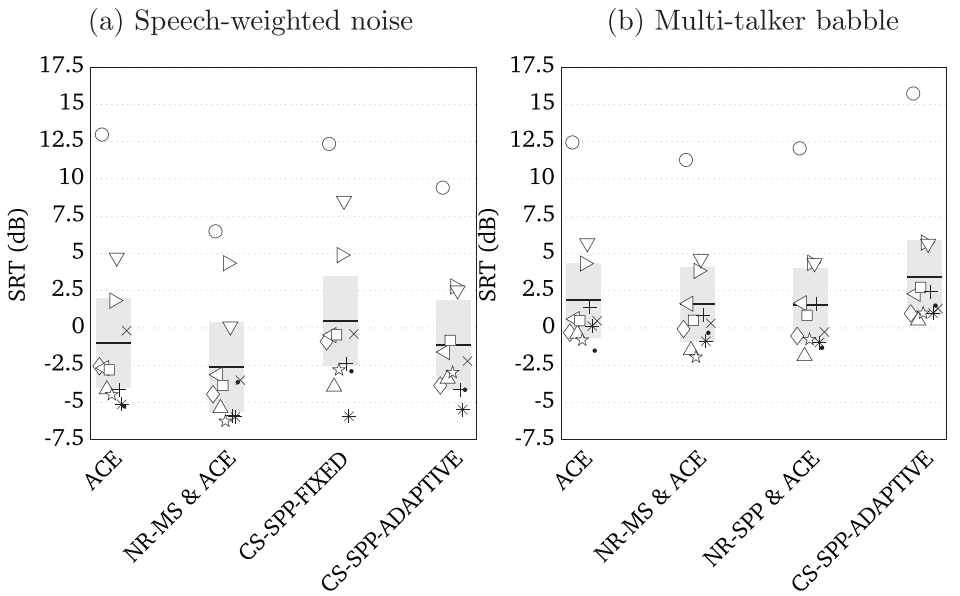

FIG. 1. Measured SRTs for the speech coding strategies in speech-weighted noise (a) and in multi-talker babble (b). Individual SRTs for each of the 12 CI recipients are shown with different symbols. Horizontal black bars illustrate the least square means, and the gray shaded boxes show the 95% confidence limits of the least square means predictions. To assess any difference between strategies, the differences of the least-squares means were computed following the Tukey multiple comparison testing.

First and foremost, even though the SPP-based noise PSD estimator has decreased the logarithmic estimation error, it does not appear to be accurate enough for SNR-based channel selection, since performance with these strategies was not close to that obtained with SNR-based channel selection based on *a priori* SNRs (Hu and Loizou, 2008). Second, a lack of training with the channel-selection strategies by the CI recipients may have influenced the performance. Finally, an experimental constraint was that only up to N channels were stimulated in the adaptively changing channel-selection strategy, where $N = 8$ for most of the participants. In comparison, up to 16 (out of the 16) channels were available for stimulation in Hu and Loizou (2008) when the SNR was high. However, this limited subset of N-of-M channels seems sufficient for ACE™, and therefore, it is unlikely to be the primary explanation for the lack of any speech intelligibility improvement.

The impact of altering the SNR-based channel selection from fixed to adaptive was also investigated. Results indicated that the adaptively-changing channel selection resulted in a higher speech intelligibility than the fixed channel selection in speech-weighted noise. Specifically, fewer than N channels were stimulated in the CI recipients when the instantaneous SNR was low in the speech gaps, and therefore, the CI recipients were exposed to less noise-induced stimulation. Reducing stimulation in speech gaps has previously been shown to be important for improving speech intelligibility in noise, because CI recipients can tolerate significantly lower levels of noise in the speech gaps than in the speech segments (Qazi *et al.*, 2013).

Overall, the results of the study indicate that a noise power estimation with improved noise-tracking capabilities, and therefore a higher accuracy, does not necessarily translate to increased speech intelligibility when the noise PSD estimation is utilized for noise reduction, nor for when it is utilized for SNR-based channel selection. However, results indicate that, for SNR-based channel selection with CI recipients, the application of an LC is important to reduce detrimental noise-induced stimulation in the speech gaps.

ACKNOWLEDGMENTS

This study was carried out in collaboration with Cochlear Limited, and clinical testing was conducted at Cochlear Melbourne, Australia. We thank Dr. Kerrie Plant and Evelyn Do for audiological and clinical assistance. In addition, this work was supported by the Oticon Centre of Excellence for Hearing and Speech Sciences and by the Danish Council for Independent Research with Grant No. DFF-5054-00072.

Baumgärtel, R. M., Hu, H., Krawczyk-Becker, M., Marquardt, D., Herzke, T., Coleman, G., Adiloğlu, K., Bomke, K., Plotz, K., and Gerkmann, T. (**2015**). "Comparing binaural pre-processing strategies II: Speech intelligibility of bilateral cochlear implant users," Trends Hear. **19**, 2331216515617917.

Bench, J., Kowal, Å., and Bamford, J. (**1979**). "The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children," Br. J. Audiol. **13**(3), 108–112.

Cohen, I. (**2003**). "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging," IEEE Trans. Speech Audio Process. **11**(5), 466–475.

Cohen, I., and Berdugo, B. (**2002**). "Noise estimation by minima controlled recursive averaging for robust speech enhancement," IEEE Signal Process. Lett. **9**(1), 12–15.

Dawson, P. W., Hersbach, A. A., and Swanson, B. A. (**2013**). "An adaptive Australian sentence test in noise (AuSTIN)," Ear Hear. **34**(5), 592–600.

Dawson, P. W., Mauger, S. J., and Hersbach, A. A. (**2011**). "Clinical evaluation of signal-to-noise ratio–based noise reduction in nucleus® cochlear implant recipients," Ear Hear. **32**(3), 382–390.

Gerkmann, T., and Hendriks, R. C. (**2012**). "Unbiased MMSE-based noise power estimation with low complexity and low tracking delay," IEEE Trans. Audio, Speech, Lang. Process. **20**(4), 1383–1393.

Hazrati, O., and Loizou, P. C. (**2013**). "Comparison of two channel selection criteria for noise suppression in cochlear implants," J. Acoust. Soc. Am. **133**(3), 1615–1624.

Hendriks, R. C., Jensen, J., and Heusdens, R. (**2008**). "Noise tracking using dft domain subspace decompositions," IEEE Trans. Audio, Speech, Lang. Process. **16**(3), 541–553.

Hu, Y., and Loizou, P. C. (**2008**). "A new sound coding strategy for suppressing noise in cochlear implants," J. Acoust. Soc. Am. **124**(1), 498–509.

Hu, Y., Loizou, P. C., Li, N., and Kasturi, K. (**2007**). "Use of a sigmoidal-shaped function for noise attenuation in cochlear implants," J. Acoust. Soc. Am. **122**(4), EL128–EL134.

Martin, R. (**2001**). "Noise power spectral density estimation based on optimal smoothing and minimum statistics," IEEE Trans. Speech Audio Process. **9**(5), 504–512.

Mauger, S. J., Arora, K., and Dawson, P. W. (**2012a**). "Cochlear implant optimized noise reduction," J. Neural Eng. **9**(6), 065007.

Mauger, S. J., Dawson, P. W., and Hersbach, A. A. (**2012b**). "Perceptually optimized gain function for cochlear implant signal-to-noise ratio based noise reduction," J. Acoust. Soc. Am. **131**(1), 327–336.

McDermott, H. J., McKay, C. M., and Vandali, A. E. (**1992**). "A new portable sound processor for the University of Melbourne/nucleus limited multielectrode cochlear implant," J. Acoust. Soc. Am. **91**(6), 3367–3371.

Qazi, O. U., van Dijk, B., Moonen, M., and Wouters, J. (**2013**). "Understanding the effect of noise on electrical stimulation sequences in cochlear implants and its impact on speech intelligibility," Hear. Res. **299**, 79–87.

Wilson, B., Finley, C., Weber, B., White, M., Farmer, J., Wolford, R., Merzenich, M., Lawson, D., Kenan, P., and Schindler, R. (**1988**). "Comparative studies of speech processing strategies for cochlear implants," Laryngoscope **98**(10), 1069–1077.