

# IMPROVING CHANNEL SELECTION OF SOUND CODING ALGORITHMS IN COCHLEAR IMPLANTS

*Hussnain Ali, Feng Hong, John H. L. Hansen, and Emily Tobey*

Center for Robust Speech Systems, Cochlear Implant Laboratory  
Erik Jonsson School of Engineering and Computer Science  
The University of Texas at Dallas, Richardson TX  
hussnain.ali, feng.hong, john.hansen, etobey@utdallas.edu

## ABSTRACT

Spectral maxima sound coding algorithms, for example  $n$ -of- $m$  strategies, used in commercial cochlear implant devices rely on selecting channels with the highest energy in each frequency band. This technique works well in quiet, but is inherently problematic in noisy conditions when noise dominates the target, and noise-dominant channels are mistakenly selected for stimulation. A new channel selection criterion is proposed to address this shortcoming which adaptively assigns weights to each time-frequency unit based on the formant location of speech and instantaneous signal to noise ratio. The performance of the proposed technique is evaluated acutely with three cochlear implant users in different noise scenarios. Results indicate that the proposed technique improves speech intelligibility and perception quality, particularly at low signal-to-noise ratio. Significance of the proposed technique lies in its ability to be integrated with the existing sound coding framework employed within commercial cochlear implant processors, making it easier to adapt for resource-limited and time critical devices.

**Index Terms**— Cochlear implants, sound coding, algorithms.

## 1. INTRODUCTION

Multi-channel cochlear implant (CI) devices operate on the principle of frequency coding based on the tonotopic mapping of the cochlea, and provide stimulation to the auditory neurons via an electrode array implanted along the basilar membrane. In principle, finer frequency resolution can be achieved by increasing the number of electrodes but factors such as current spread (caused by electric stimulation) and the number of surviving auditory neurons constrain the design of CI devices. Electrode arrays used in current CI devices, such as Freedom implant by Cochlear Corporation, are able to provide up to 22 active sites for electric stimulation along the length of the cochlea. However, not all sites are used for stimulation in each stimulation cycle. This is to avoid channel interaction and

control the current spread within the cochlea. Continuous Interleaved Sampling (CIS) [1] sound coding algorithm, for example, decomposes the acoustic signal into 8 – 10 frequency bands/channels. ' $n$ -of- $m$ ' strategies, such as the Advanced Combinations Encoder (ACE) [2], decompose the signal in ' $m$ ' (22) channels and pick ' $n$ ' (8 – 12) channels in each stimulation cycle. The goal of CI sound coding strategies is to represent the meaningful sound features in a limited number of channels. ACE strategy achieves this by selecting 8 – 12 channels with the highest spectral energy, which works well in quiet; however, listening performance of CI users drop significantly in noisy environments. This limitation is inherent in the channel selection criterion of the ACE strategy which can mistakenly select channels that are dominated by noise. This is potentially the main reason that CI users are unable to tease apart meaningful features of the target speech from noise because target-dominant channels may never be activated. Therefore, an intelligent channel-selection strategy is needed which is able to classify and select channels with the highest amount of target-dominant speech, and not necessarily just energy.

A number of noise reduction algorithms for cochlear implants have been proposed over the years which are either based on signal pre-processing [3] – [5] or signal conditioning integrated with sound coding [6], [7]. While the former approach can work well in hearing aids, it is potentially susceptible to unwanted signal distortion which can easily be enhanced by CI processing (e.g., compression function emphasizes low energy sounds logarithmically) or it can be computationally intensive, thus making it unfavorable for CI processors. The later approach generally relies on spectral modification or modifying channel-selection based on the signal to noise ratio (SNR). Hu *et al.* [6], for example, used a sigmoidal-shaped function that applies attenuation to the noisy envelopes (computed by CIS strategy) inversely proportional to the estimated SNR in each channel. By discarding channels with an SNR < 0 dB (binary masking) and varying the number of active channels, [7] reported restoration of speech intelligibility to the level attained in quiet when the channel SNR was known (ideal condition). While this approach can work well in ideal conditions, one

clear disadvantage is that binary masking would completely discard channels containing speech components that are essential for speech intelligibility, but are either unfortunately dominated by noise or wrongly classified by the noise estimation algorithm. The technique proposed in this paper takes its inspiration from the former two approaches and shapes the weighting functions used in the ACE processing based on the instantaneous SNR of each time frequency (TF) unit. In addition, priority is assigned to channels containing the three speech formants, F1, F2, and F3.

This paper is organized as follows. Sec. 2 describes the proposed technique which improves the channel selection in ‘n-of-m’ strategy. Sec. 3 presents evaluation of the proposed technique with 3 CI human subjects followed by discussion and conclusion in Sec. 4.

## 2. METHOD

In the clinical/standard ACE (STD\_ ACE) strategy (Fig. 1, inside the dotted block), the acoustic signal is sampled at 16 kHz, pre-emphasized, and buffered using a Blackman window into 8 ms (128 samples) analysis frames. Frame overlapping (or analysis rate) typically depends on the channel stimulation rate. For each analysis frame, 128 point FFT and magnitude squared spectrum is computed, thereby giving 64 frequency bins, each bin has a frequency resolution of 125 Hz. These bins are passed through 22 weighting filters (Fig. 2), which essentially computes the envelope of each channel. Next, 8 - 12 channels with the highest amplitudes are selected and compressed to the current levels using a loudness growth function (LGF) and the patient’s clinical MAP, which maps the acoustic amplitudes to the patient’s electrical dynamic range.

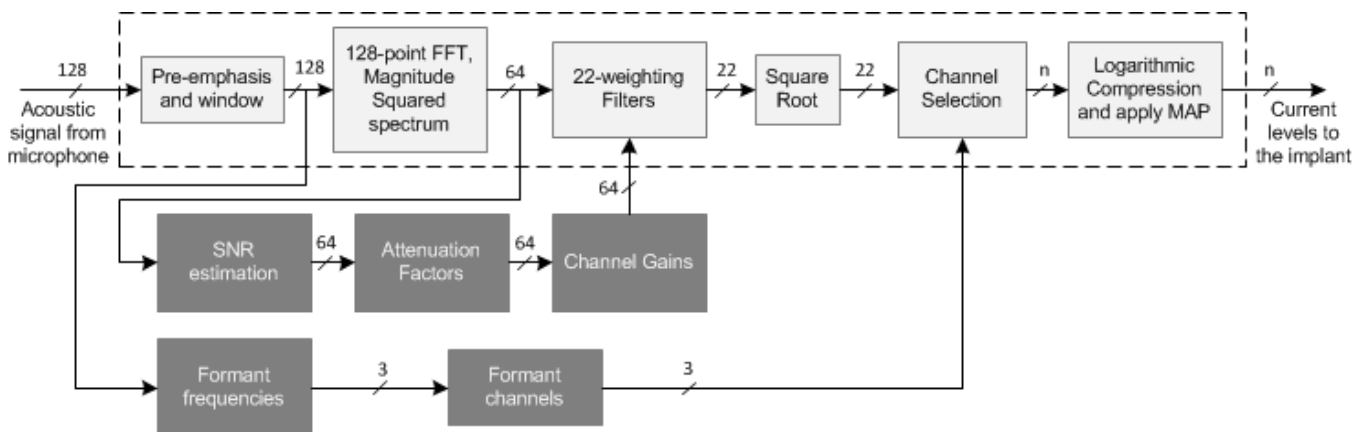
Fig. 1 shows the block diagram of the proposed technique (in conjunction with the STD\_ ACE routine). The proposed technique operates based on two principles, 1) by

assigning priority to formant bands and 2) by assigning weights to each TF unit based on the instantaneous SNR.

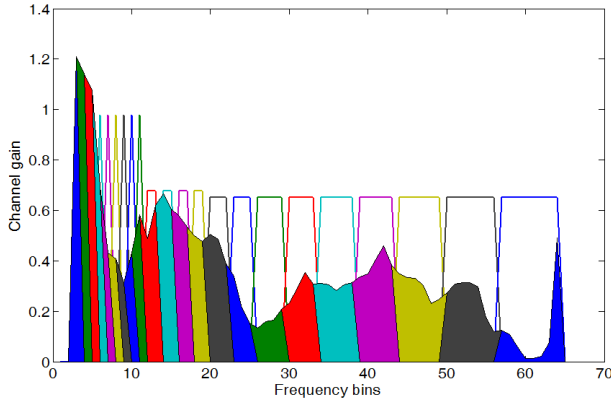
### 2.1 Assigning priority to the formant bands

The frequencies of the first three speech formant (F1 – F3) peaks as well as their trajectory over time provide valuable cues to listeners for vowels, glides and stop-consonant perception [8]. This is the reason that feature-extraction strategies [9] – [13] have been popular in earlier generation CI processors of 1980s and 1990s. F0/F2 [9], [10], and F0/F1/F2 [11] strategies extract formant locations (F1 and F2) and stimulate the corresponding electrodes at a rate of F0 pulses/sec (pps) for voiced segments and an average rate of 100 pps for unvoiced segments. The MULTIPLEAK (MPEAK) strategy [12], [13] stimulates four electrodes at a time and always activates electrode numbers 4 and 7 for F1 and F2 respectively, and then selects the remaining two based on the spectral content of the speech signal.

Spectral maxima based sound strategies were later adapted to encode the entire spectrum of the speech signal of whom ACE is the prime example. The shortcoming in spectral maxima algorithms, as noted earlier, comes from the fact that channel selection is based on the largest filter amplitudes which are not necessarily the spectral peaks, and hence may not encode the major formant frequencies. Several maxima may come from a single spectral peak [14]. This can be problematic in noise, which tends to reduce the dynamic range of the spectrum as well as the spectral contrast (peak-to-valley ratio on LPC spectrum). Thus, preference will be given to noise dominant channels irrespective of the presence/absence of spectral peaks. In fairness, the F1 spectral peak is preserved to a certain degree in noise which works to the advantage of ACE. Although, the location of peaks of higher formants may not be affected as much in noise, spectral smearing and reduced spectral contrast would give preference to the noise dominant



**Fig.1.** Signal flow in the standard ACE strategy (shown inside the dotted block). Processing blocks for the proposed technique are shown in the darker tone. Numbers on connecting arrows represent the frame size in number of samples at each step.



**Fig. 2.** Weighting filters used in the ACE strategy plotted in solid line. Spectrum shaping by soft-masking technique shapes the spectrum based on SNR of each TF unit as shown by the color-filled plots.

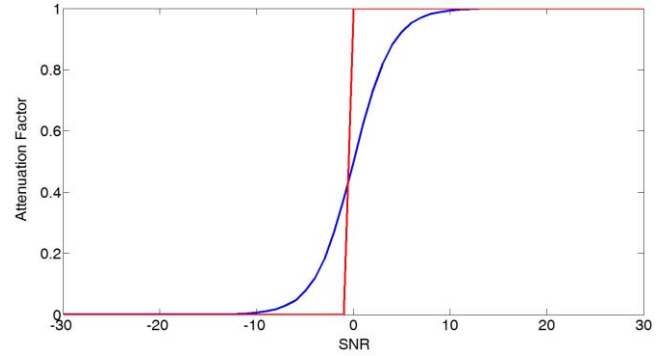
channels.

The proposed technique continuously computes the first three formant peaks (F1 – F3) in each analysis cycle and assigns priority to the channels corresponding to the formant frequencies during the channel selection process. Formant frequencies are computed by solving for the roots of the linear prediction coefficients (LPC). Formant continuity constraints are imposed to avoid unwanted distortion.

## 2.2 Assigning attenuation factor based on the SNR

Both [6] and [7] apply binary and soft masking techniques respectively to channels with low SNR. Given that a channel can comprise of as many as 8 frequency bins, for example, for higher frequencies (more if the number of stimulation sites is less than 20), channel classification would be compromised for bin-widths of 2 or more. The proposed technique estimates the SNR for each TF unit,  $X(i,j)$ , where  $X$  is magnitude squared spectrum of the  $i^{\text{th}}$  analysis frame and  $j^{\text{th}}$  frequency bin. This yields a total of 64 SNR values for each analysis frame (stimulation cycle). Based on the computed SNR, an attenuation factor is generated. In the present study, we analyzed both binary and soft masking techniques. In binary masking, a weight of 0 was assigned to the  $SNR(i,j) < 0$  dB, for the rest (i.e.,  $SNR(i,j) \geq 0$  dB), a binary value of 1 was assigned. In the soft masking approach, a sigmoidal-shaped function was considered which plateaus for SNRs  $> 15$  dB and floors to 0 for SNRs  $< -15$  dB. Both weighting functions are shown in Fig. 3. The 64 weighting values generated for each TF unit are then used to shape the gain of weighting functions. This is illustrated in Fig. 2 for the soft masking technique.

In order to evaluate the effectiveness of the proposed technique, tests were first performed with the SNR of each TF unit known *a priori*. The results from this experiment would validate if the proposed technique is



**Fig. 3.** Binary weighting (red) and soft-masking (blue) attenuation functions.

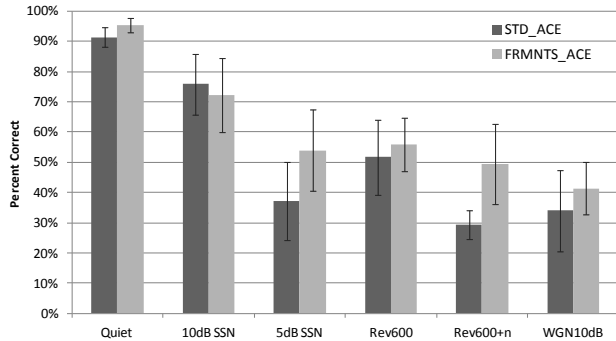
effective. In the second phase, the instantaneous SNR of each TF unit was estimated using improved minimum controlled recursive average (IMCRA) algorithm [15]. While any SNR estimation algorithm could be used, IMCRA utilizes the spectrum components of each TF unit,  $X(i,j)$ , which are already computed by ACE. Furthermore, the advantages of the IMCRA method are particularly notable in adverse environments involving nonstationary noise, weak speech components, and low SNR conditions.

## 3. PROCEDURE, EVALUATION, AND RESULTS

A total of 3 CI users participated in this acute study. All participants were native speakers of American English and fitted with Nucleus 24 multichannel device manufactured by Cochlear Corp. All participants used ACE as their speech processing strategy. IEEE sentences [16] were used as the speech stimuli for testing. 10 sentences for each test condition were used. The algorithms were implemented offline in MATLAB and stimuli were presented via UT Dallas's PDA-based research platform [17].

Two sets of experiments were conducted. In the first set, the effectiveness of assigning priority to formant channels (F1, F2, and F3) was evaluated both in terms of speech intelligibility and perception quality. All words were marked for correctness. A total of 6 conditions were tested in experiment 1, namely speech in quiet, speech in 10 dB SNR speech shaped noise (SSN), speech in 5 dB SNR SSN, speech in 10 dB SNR white Gaussian noise (WGN), speech in reverberation with reverberation time  $T_{60} = 600$  ms, and finally, speech in reverberation ( $T_{60} = 600$  ms) and 10dB noise. Fig. 4 shows the mean intelligibility scores for STD\_ ACE and formant-based ACE (FRMNTS\_ ACE) technique. While there is very little to no significant improvement in intelligibility at high SNRs, there was an improvement of 17% for speech at 5 dB SNR SSN, and 20% when noise was added to the reverberant signal (Rev600+n). Modest improvement was also observed for speech in WGN at 10 dB SNR.

For the speech quality tests, the same sentences processed with STD\_ ACE and FRMNTS\_ ACE strategies



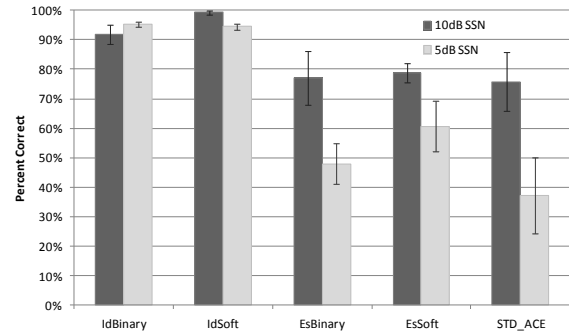
**Fig. 4.** Mean speech intelligibility scores of 3 CI users in Exp. 1. Error bars represent standard error of the mean.

was streamed back to back and CI users were asked to rate the quality of the second sentence as compared to the first in terms of being pleasant, clear and free of sound distortions on a scale of -3 to 3, with 0 being ‘about the same’, -3 being much worse, and +3 indicating much better. On average, the user response was between +1 and 0, indicating slightly better to no difference in all test conditions. The subjects reported that words “popped out” more in the FRMNTS\_ACE strategy.

In the second experiment, four techniques were assessed separately, namely ideal-binary (IdBinary), ideal-soft (IdSoft), estimated-binary (EsBinary), and estimated-soft (EsSoft). Ideal conditions represent when the SNR was known *a priori*, while the estimated conditions represent when the SNR was estimated using IMCRA. Binary conditions correspond to the output of the attenuation factor to binary values, while the soft conditions represent output from the sigmoidal attenuation function, as described earlier. Speech intelligibility and quality measures were assessed for each technique with speech in 10 dB SNR SSN, and speech in 5 dB SNR SSN. The results from STD\_ACE were used as baseline scores for comparison.

The mean intelligibility scores for experiment 2 are presented in Fig. 5. The results show that both IdBinary and IdSoft techniques are able to restore speech intelligibility to the level equivalent to speech in quiet. This establishes the effectiveness of the proposed technique in masking noise if cues to the SNR are available. Fig. 5 also presents intelligibility scores when the noise is estimated. There was no significant improvement at 10 dB SNR level. However, at 5 dB SNR, improvement is seen with both EsBinary and EsSoft, but later results in significant gains in intelligibility (>20 percent). The results indicate that the proposed technique can potentially improve speech intelligibility in low-SNR conditions.

Quality tests for Experiment 2 indicate an average score of “Much Better” as compared to the STD\_ACE (unprocessed) for both ideal conditions. For the estimated SNR at 10 dB and 5 dB SNR levels, the average score was (+2) corresponding to “Better” as compared to the



**Fig. 5.** Speech intelligibility scores for Experiment 2. Comparison of the proposed technique using IdBinary, IdSoftm, EsBinary, and EsSoft approaches with STD\_ACE strategy. Error bars represent standard error of the mean.

STD\_ACE. The quality scores between binary and soft masking approaches indicate slight preference for the soft masking approach, with an average score of +0.5.

#### 4. CONCLUSION

This study considered two approaches for improving the channel selection process of spectral maxima sound coding algorithms for cochlear implant systems. The first approach assigns priority to the channels containing the formant frequencies, while the second approach adaptively assigns weights to each time-frequency (TF) unit based on the estimated SNR in each stimulation cycle. Two types of maskers, binary and soft, were used to assign attenuation factor to each TF unit based on SNR. Both approaches were evaluated independently and synergistically on 3 CI users. The intelligibility scores indicate significant improvement at low SNR levels, with speech at 5 dB SNR and speech masked by reverberation and noise. The quality scores revealed very little preference to the formants-based approach over the standard ACE, and soft-masking over the binary masking. However, the noise masking approach was greatly preferred over standard ACE. The intelligibility scores for the two approaches revealed some benefit of the formants-based technique over the standard ACE, and soft masking over binary masking at low SNR levels.

The proposed technique is inherently limited by the accuracy of formant locations and SNR estimation. Speech intelligibility reached to the level obtained in quiet when SNR was known *a priori*. The results from this study indicate the potential of the proposed technique to improve speech perception of CI users in adverse listening environments.

#### ACKNOWLEDGEMENT

This work was supported by the National Institutes of Health under contract N01-DC-6-0002 from the National Institute on Deafness and Other Communication Disorders and R01 DC010494-01A.

## 12. REFERENCES

- [1] B. Wilson, C. Finley, D. T. Lawson, R. D. Wolford, D. K. Eddington, W. M. Rabinowitz, "Better speech recognition with cochlear implants," *Nature*, vol. 352, pp. 236-238, July 1991.
- [2] A. E. Vandali, L. A. Whitford, K. L. Plant, and G. M. Clark, "Speech perception as a function of electrical stimulation rate using the Nucleus 24 cochlear implant system," *Ear and Hearing*, vol. 21 (6), pp. 608–624, 2000.
- [3] J. Wouters, and J. B. Vanden., "Speech recognition in noise for cochlear implantees with a two microphone monaural adaptive noise reduction system," *Ear and Hearing*, vol. 22, issue 5, pp. 420 – 430, 2001.
- [4] P. C. Loizou, "Speech processing in vocoder-centric cochlear implants," *Adv. Oto-Rhino-Laryngol.* vol. 64, pp. 109 – 143.
- [5] R. V. Hoesel, and G. Clark, "Evaluation of a portable two-microphone adaptive beamforming speech processor with cochlear implant patients," *J. Acoust. Soc. Am.* vol. 97, issue 4, pp. 2498–2503.
- [6] Y. Hu, P. C. Loizou, N. Li, K Kasturi, "Use of a sigmoidal-shaped function for noise attenuation in cochlear implants," *J. Acoust. Soc. Am.* vol. 122, issue 4, 2007.
- [7] Y. Hu, P. C. Loizou, "A new sound coding strategy for suppressing noise in cochlear implants," *J. Acoust. Soc. Am.* vol. 124, issue 1, pp. 498–509, 2008.
- [8] P. C. Loizou, "Speech Enhancement – Theory and Practice," CRC Press, 2012.
- [9] G. Clark, "The University of Melbourne-Nucleus multielectrode cochlear implant," *Advances in Oto-Rhino-Laryngology*, vol. 38, pp. 1 – 189, 1987.
- [10] P. Seligman, J. Patrick, Y. Tong, G. Clark, R. Dowell, and P. Crosby, "A signal processor for a multiple-electrode hearing prosthesis," *Acta Otolaryngologica*, pp. 135 – 139. (Suppl. 411), 1984.
- [11] P. Blamey, R. Dowell, and G. Clark, "Speech recognition for 40 patients receiving multichannel cochlear implants," *Archives of Otolaryngology, Head and Neck Surgery*, vol. 112, pp. 1054 – 1059, 1986.
- [12] J. Patrick, P. Seligman, D. Money, and J. Kuzma, "Engineering," in Cochlear prosthesis (G. Clark, Y. Tong, and J. Patrick, eds.), Edinburgh: Churchill Livingstone, 1990, pp. 99 - 124.
- [13] J. Patrick and G. Clark, "The Nucleus 22-channel cochlear implant system," *Ear and Hearing*, vol. 12, pp. 3 – 9, 1991.
- [14] P. C. Loizou, "Mimicking the human ear," *IEEE Signal Processing Magazine*, vol. 15, no. 5, pp. 101 – 130, Sep. 1998.
- [15] I. Cohen, "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging," *IEEE Transactions on Speech and Audio Processing*, vol. 11, issue 5, pp. 466-475, 2003.
- [16] IEEE Subcommittee, "IEEE Recommended Practice for Speech Quality Measurements," *IEEE Trans. Audio and Electroacoustics*, vol. AU-17, no. 3, pp. 225-246, 1969.
- [17] H. Ali, A. P. Lobo, P. C. Loizou, "Design and Evaluation of a Personal Digital Assistant-based Research Platform for Cochlear Implants," *IEEE Transactions on Biomedical Engineering*, vol.60, no.11, pp.3060,3073, Nov. 2013.