A WOLA-Based Real-Time Noise Reduction Algorithm to Improve Speech Perception with Cochlear Implants

Kaibao Nie[*], Tzu-Lin Ong[†], Xinzhao Liu[†], Hanqing Wen[†], Feifan Lai[*],
Wei Fan[†], Changjian Xu[†], Xindong Liu[†], and Zengjun Sun [†]

[*]University of Washington, USA. E-mail: niek@uw.edu
[†]Shanghai Listent Medical Corp, Shanghai, China

***Abstract*— Background noise poses a significant challenge to people who have a cochlear implant for restoring their hearing ability. A cochlear implant can process sound into 12 to 24 channels and it provides limited temporal and spectral information to the auditory nerve through electrical current stimulation. In this paper, a specific noise reduction algorithm was developed to accommodate the need of adaptively applying a small number of gains to the stimulation signals in cochlear implants. A sound signal was first divided into 22 channels using the WOLA (Weighted Overlap Add) spectral analysis. The spectral templates of background noise were estimated by automatically tracking energy gaps between speech segments. The gap detection algorithm utilized a mechanism like the charging and discharging of a capacitor in an envelope detector, which offers the ability for the extracted energy signal to stay at noise floors. The noise templates were updated adaptively when a segment of signal was determined to be noise. Simulations of the proposed noise reduction algorithm were performed using offline processing and it has also been implemented on the Ezairo 7150 (ON Semiconductor Corporation) DSP platform with a WOLA co-processor. Initial evaluation results showed that the signal-to-noise (SNR) ratio can be improved by up to 10 dB after noise removal.**

I. Introduction

Cochlear implants can partially restore hearing ability to people who have profound hearing loss. A cochlear implant is able to convert acoustic sound signals into electric pulses for directly activating the auditory nerve through an implantable electrode array. Patients with a cochlear implant can generally understand speech in quiet to a high level, however their speech recognition performance remains poor in background noise. Various types of noise reduction (NR) algorithms have been developed to mitigate the challenge of speech recognition in noise for cochlear implant patients. The techniques used include spectral subtraction [1,4], recursive minimum statistics [2], dual-microphone adaptive filtering [3], speech pause detection [1, 6-8], and the most recent deep learning (Deep Neural Network) [10] etc. Even though significant improvement can be potentially achieved with these techniques, the greatest challenge is to integrate the NR algorithms into existing signal processing modules in current cochlear implants, and also to implement the computationally intensive signal processing algorithms.

A cochlear implant can process sound into 12 to 24 channels, and then slowly-varying envelopes are further extracted to modulate biphasic pulse trains. Thus, it provides limited temporal and spectral information to the auditory nerve through electrical current stimulation. An effective noise reduction algorithm needs to take into consideration both spectrally and temporally smeared stimulation signals to maximize the effectiveness of noise reduction in cochlear implants. In this paper, An WOLA-based NR algorithm was proposed to reduce the computational complexity of noise estimation and to effectively determine the gain functions that can be applied to each channel in cochlear implants. The derived gain functions can be readily integrated into the existing signal processing in current cochlear implants for envelope extraction and pulse generation.

II. Noise Estimation and Gain Calculation

A. *WOLA-based noise reduction algorithm*

The WOLA filterbank processing provides an efficient way of transforming an audio signal into the frequency domain and then applying gain functions [11]. A noise reduction algorithm was then proposed to utilize the features of WOLA processing to combine stimulation feature extraction and gain application into one processing unit.

Fig. 1 shows the overall design of the WOLA-based pulse generation and noise reduction algorithm. In a 22-channel cochlear implant, an audio signal was first transformed to its spectral representation by WOLA analysis. 128 frequency bins were used in the present study and they were subsequently merged to form 22 channel stimulation signals. A noise frame detector was added for estimating the spectrum of background noise. A vector containing noise spectrum was pushed into a 4-stage buffer to store the latest spectral templates of the audio frames containing primarily background noises. The stored spectrum templates in the noise buffer were further averaged to create smoothed gain functions for mitigating potential artifacts after gain application. In addition, the averaged noise spectrum was

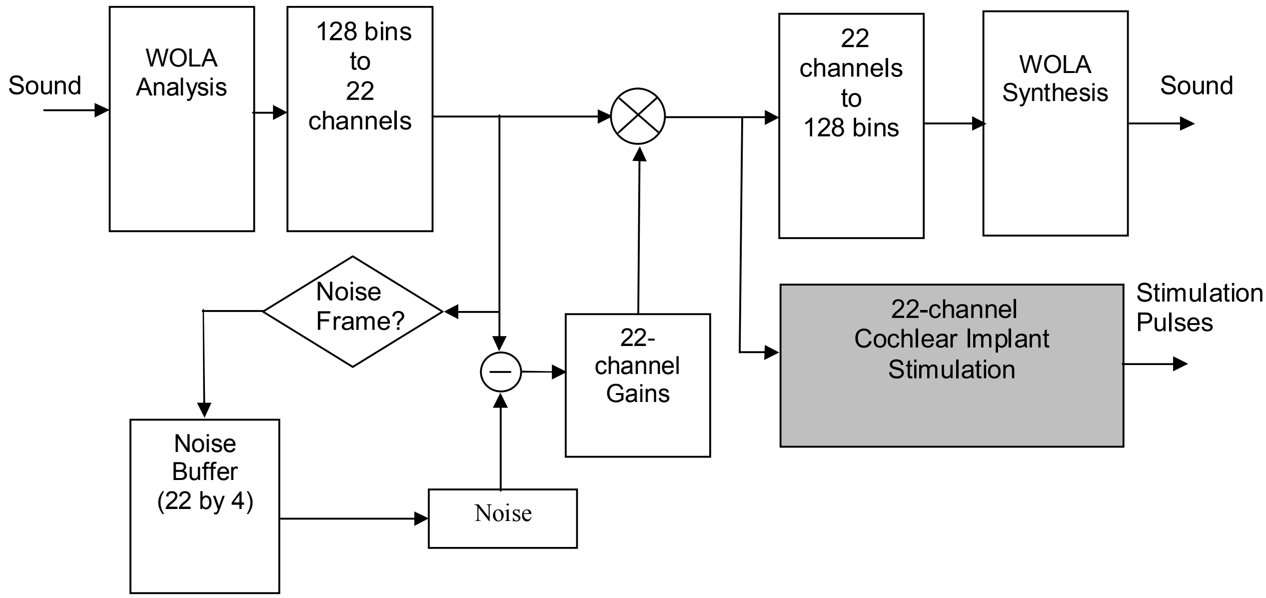

Fig. 1 Block diagram to illustrate the process of background noise estimation and gain calculation in a 22-channel cochlear implant. Sound is also re-synthesized through WOLA synthesis for evaluating the performance of the noise reduction algorithm acoustically.

filtered in the spectral domain to reduce the musical noise effect which is normally present in spectral-subtraction based noise reduction processing. Once the noise spectrum was obtained, 22 gained were calculated based on SNR estimation (RMS difference in dB) in each separate channel. For generating stimulating pulse trains, gain-modified envelopes were directly used to modulate biphasic pulse trains delivered to the cochlear electrodes. Alternatively, the 22-channel signals can be fed to the WOLA synthesis to reconstruct time-domain acoustic signals for perceptually evaluating the noise reduction algorithm.

B. Background noise estimation

One of the crucial components of this noise reduction algorithm implementation is to detect the audio frames primarily dominated by background noises. Background noise is typically nonstationary but embedded in signals in the form of energy gaps between speech segments. To track the gaps or pauses between speech segments, a dynamic energy-dip tracking method was adopted. It works in a similar way to the RC envelope/peak detector. In an envelope detector, the capacitor is charged when the input voltage exceeds the instantaneous voltage across the capacitor and otherwise it will enter a discharging cycle when the input signal reaches a peak voltage. Instead of tracking the peaks in an envelope detector, the design of the noise floor tracking should be able to follow the troughs in the energy curve.

Fig. 2 illustrates the adaptive tracking procedure of background noise. The tracking curve linearly increases with a time constant at 180 ms to catch typical speech syllabic rates (~ 2–5 Hz). When the signal's short-term RMS is below the tracking curve, the curve will follow the input until the input increases again. The charging/discharging process allows the tracking curve to stay on the noise floor when speech signals are present. A noise frame was detected at the start of the discharging cycle and was pushed into the noise spectrum buffer.

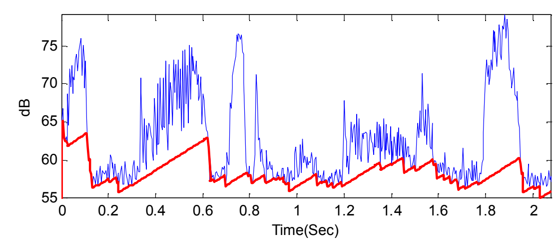

Fig. 2 Tracking of the background noise floor with the charging & discharging method. The blue curve represents the short-term RMS levels of the sound signal.

C. Channel gain calculation

For each channel, the signal-to-noise (SNR) was calculated for determining the applied gain function to each individual envelope signal. The SNR to gain mapping function is presented in Fig. 3. A threshold of 10 dB was selected to either apply 0 dB gain to speech dominated frames or an attention gain to noise dominated frames.

As shown in Fig. 3, the gain was set to 0 dB for estimated SNRs above 10 dB. Otherwise, the channel gain would be decreased at a slope of -1, -2, or -3 dB/ per SNR dB at different noisy environments. In extremely noisy listening conditions, the slope of attenuation can be set at -3 to aggressively reduce background noise. The threshold of 10 dB was chosen based on previous human perceptual experiments, which is also consistent with the psychometric function for speech perception found in cochlear implant patients [9,12]. It has been shown that signals with SNR above 10 dB contribute most for cochlear implant patients to understand speech.

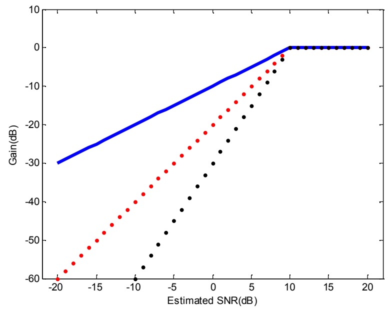

Fig. 3 Channel gain calculation based on estimated SNR. The slope factor can be set to -1(solid blue), -2(dashed red) or -3 (dashed, black) for various degrees of noisy environments.

III. RESULTS

The proposed algorithm was first evaluated with a simulation tool in Matlab. As an example, a HINT (Hearing in Noise Test) sentence 'Big dogs can be dangerous' was mixed with a speech-shaped noise at 10 dB SNR. The speech plus noise signal was then processed by the noise reduction algorithm in off-line signal processing.

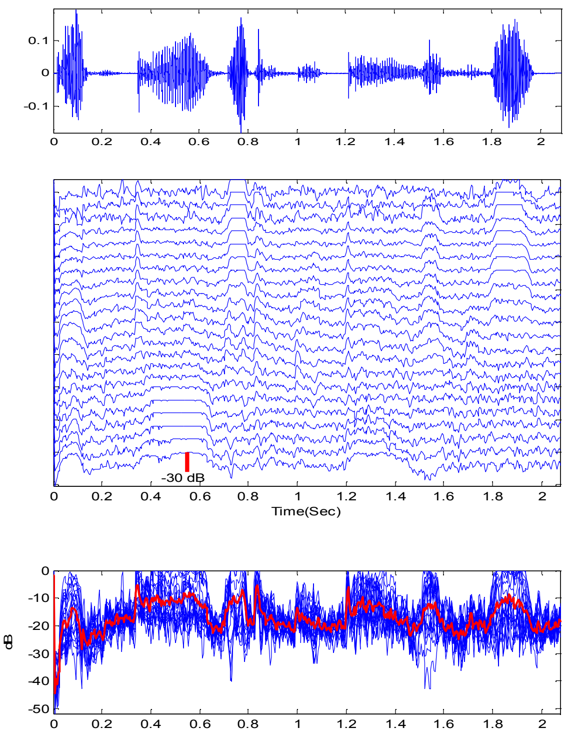

Fig. 4 Clean speech (top), 22 individual channel gains (middle) and the average gain curve (bottom, red) for a speech sound 'Big dogs can be dangerous' mixed with noise. The SNR was set to 10 dB with a speech-shaped noise from the HINT test battery. The gain attenuation slope was at -2.

Fig. 4 shows the estimated 22 channel gains during the entire speech signal. The top panel contains the clean speech prior to adding the noise. Flat horizontal line segments in the middle panel indicate 0 dB gain where the speech signal is relatively strong (>10 dB SNR). The average of all 22 gains over time was plotted on the bottom panel of Fig. 4 highlighted with a bold curve in red. The patterns of adaptive gain adjustments are generally consistent with the modulation envelope of speech syllables plotted on the top panel of Fig. 4. As a result, background noise will be significantly attenuated when noise is dominant and perceptually important speech segments at higher SNR still can be maintained.

Fig. 5 compares the speech spectrum before and after gain application. The top panel gives the WOLA analysis results with 128 bins for the HINT sentence used. Stored noise spectrum templates are presented in the middle panel. The bottom panel is the speech spectrum after applying all the 22 channel gains but prior to WOLA synthesis. In the first 0.1 sec, the gain curves were being initialized and therefore it represents a transitional period for the algorithm to start working. After applying the derived gains, the majority of the noise components across the entire frequency spectrum were successfully removed, creating a relatively clean speech sound.

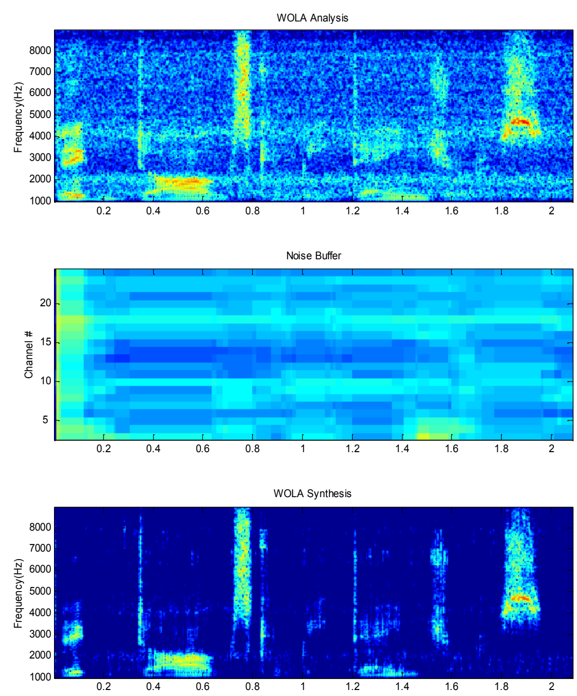

Fig. 5 128-bin WOLA analysis result of the noisy speech (top), 22 channels of estimated noise spectrum (middle), and the cleaned spectrum for WOLA synthesis.

The time-domain waveforms before and after noise reduction are given in Fig. 6 to visually examine the effects of noise reduction.

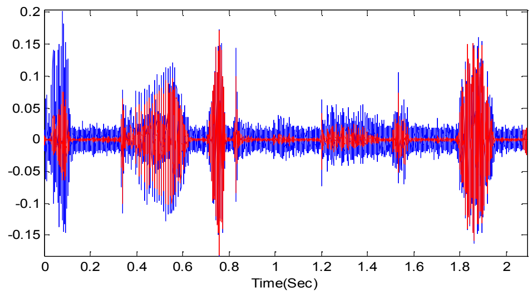

Fig. 6 Comparison of the time waveforms of the noisy speech (blue) and synthesized speech (red) by WOLA synthesis (SNR=10 dB).

To assess the noise reduction algorithm running in real time, the HINT sentence was also played through a speaker and internal signals before and after noise removal were transmitted out through an SPI (Serial Peripheral Interface) interface port for analysis. Signal-to-noise ratio was calculated by dividing the recorded signals into speech and noise blocks. Table 1 presents the SNR values for the input and output plus the improvement in SNR after noise attenuation. The slope was set to -2 in the actual implementation on the DSP platform. At 2.9 dB SNR, the amount of improvement can reach 11.4 dB since the attenuation gain was quite aggressive and the gain function operated at a sharp slope near 0 dB. As the SNR was increased to 7.8 and 10.7 dB, SNR improvement was at roughly 10.7 and 9 dB respectively.

Table 1 Calculated SNR of the input and output of NR module implemented on Ezairo 7150.

Input SNR (dB)	Output SNR(dB)	SNR improvement(dB)
2.9	14.3	11.4
7.8	18.5	10.7
10.7	19.7	9.0

IV. Conclusions and Discussions

Cochlear implant users are more susceptible to background noise due to the limited spectral and temporal resolution in electric hearing. In this study, a feasible real-time noise reduction algorithm was developed and implemented on the Ezairo 7150 DSP system in conjunction with a WOLA co-processor designed specifically for hearing devices such as hearing aids and cochlear implants. The general algorithm can also be implemented on other platforms with some modifications.

The WOLA-based noise removal method utilizes a noise floor tracking technique to dynamically update the estimated noise spectrum. The 'charging' and 'discharging' process when tracking energy troughs can increase the chance of catching spectral and temporal prosperities of the background noise. The actual performance of the proposed noise reduction algorithm needs to be further tested with cochlear implant patients.

Acknowledgment

This study was partially supported by the University of Washington VM Bloedel Hearing Center and Shanghai Listent Medical Tech Co Ltd.

References

[1] L.P. Yang and Q.J. Fu, "Spectral subtraction-based speech enhancement for cochlear implant patients in background noise (L)," *The Journal of the Acoustical Society of America*, vol. 117(3), pp. 1001-1004, 2005
[2] P.W. Dawson, S.J. Mauger, and A.A. Hersbach, "Clinical Evaluation of Signal-to-Noise Ratio–Based Noise Reduction in Nucleus® Cochlear Implant Recipients," *Ear and Hearing*, vol. 32, pp. 382-90, 2011
[3] J.B. Maj, J. Wouters, and M. Moonen, "Noise reduction results of an adaptive filtering technique for dual-microphone behind-the-ear hearing aids," *Ear Hearing*, vol. 25(3), pp. 215-29, 2004
[4] K. Kokkinakis, B. Azimi, Y. Hu, and D.R. Friedland, "Single and multiple microphone noise reduction strategies in cochlear implants," *Trends in Amplification*, vol. 16(2), pp.102-16, 2012
[5] S.J. Mauger, P.W. Dawson, and A.A. Hersbach, Perceptually optimized gain function for cochlear implant signal-to-noise ratio based noise reduction," *The Journal of the Acoustical Society of America*, vol. 131, pp. 327-36, 2012
[6] M. Marzinzik and B. Kollmeier, "Speech pause detection for noise spectrum estimation by tracking power envelope dynamics," *IEEE Trans. Speech and Audio Processing*, vol. 10, pp. 341 - 351, 2002
[7] R. Martin. "Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics," *IEEE Transactions on Speech and Audio Processing*," vol. 9(5), pp. 504 - 512, 2001
[8] I. Cohen. "Noise Spectrum Estimation in Adverse Environments: Improved Minima Controlled Recursive Averaging," *IEEE Transactions on Speech and Audio Processing*, vol. 11(5), pp. 466-475, 2018
[9] K. Kasturi and P.C. Loizou, "Use of S-Shaped Input-Output Functions for Noise Suppression in Cochlear Implants," *Ear and Hearing*, vol. 28, pp. 402-411, 2007
[10] Y.H. Lai, Y. Tsao, and X. Lu etc., "Deep Learning-Based Noise Reduction Approach to Improve Speech Intelligibility for Cochlear Implant Recipients," *Ear and Hearing*. doi: 10.1097/AUD.0000000000000537[Epub ahead of print], 2018
[11] R. Brennan, T. Schneider, and W. Zhang, "Ultra-Low-Power Noise Reduction Strategies Using a Configurable Weighted Overlap-Add Coprocessor," Dspfactory Ltd., 2012
[12] Y. Hu, P.C. Loizou, N. Li, and K. Kasturi, "Use of a sigmoidal-shaped function for noise attenuation in cochlear implants," *J. Acoust. Soc. Am.*, vol. 122(4):EL, pp. 128–134, 2007