Noise Suppression with Apparent Gain Threshold for Cochlear Implants: A Vocoder Simulated Verification

Nengheng Zheng*, Xiaowei Shao*, Baihan Zeng*, and Qinglin Meng†
* Shenzhen Key Lab of Advanced Communications and Information Processing, Guangdong Key Laboratory of Intelligent Information Processing, College of Information Engineering, Shenzhen University, Shenzhen, Guangdong, China
E-mail: nhzheng@szu.edu.cn
†Acoustic Lab., School of Physics and Optoelectronics, South China University of Technology, Guangzhou, Guangdong, China
E-mail: mengqinglin@scut.edu.cn

***Abstract*—Noise suppression techniques with low computational complexity and fast parameter tuning procedure are very important for the clinical implementation of cochlear implants (CI) and the postoperative rehabilitation of CI recipients. This study presents an improved noise suppression algorithm based on Mauger's work on perceptually optimized gain function for CI. The relation between the perceptually optimal Wiener parameters and a metric named apparent gain threshold (aGT) is first disclosed by CI vocoded speech perception experiments. Then a gain function based on the aGT is proposed for Wiener filtering. Instead of joint searching for the two parameters (i.e., α and β) as in Mauger's work, the proposed one requires only searching for the optimal aGT parameter and, as a result, user-dependent optimal parameter tuning could be much more efficient for CI recipients. Speech perception experiments with vocoder simulations in normal-hearing listeners are conducted to evaluate the effectiveness of the proposed method and positive results are obtained, which implies its promising implementation in clinical CI devices.**

I. Introduction

Cochlear implants (CI) are surgically implanted electronic devices aiming to recover hearing ability for patients with severe hearing loss. The progress in CI technology in the past decades has enabled CI recipients to enjoy a high level of speech understanding in quiet. However, in real-world speech communications, the target speech is usually corrupted by acoustic interference, which reduces the intelligibility of speech. The artificial electric hearing provided by CI is still far from satisfactory for noisy speech perception [1][2]. Noise suppression has been one of the key techniques for CI signal processing. Many speech enhancement algorithms have been adopted in CI to improve its noise robustness. Particularly, algorithms with low computational complexity, e.g., spectral subtraction, Wiener filtering, etc., have been widely implemented in clinical CI devices [3].

Due to the limitation of CI signal processing, e.g., coarse frequency and amplitude resolution, lack of temporal fine structure, etc., CI recipients have very different perceptual properties from normal-hearing (NH) listeners [4]. For example, most speech enhancement algorithms need to trade off between the noise removal and the speech distortion. In terms of the intelligibility of the denoised speech, NH listeners prefer to lower speech distortion, however, CI recipients are very sensitive to the noise but can tolerate high degree of speech distortion [5][6]. Furthermore, the hearing capability vary significantly across individual CI recipients. Therefore, a successful noise suppression implementation in CI usually requires a large amount of perceptual experiments to determine the user-dependent optimal parameters.

Mauger et al. [7] studied the effectiveness of two SNR-based noise suppression algorithms, i.e., the ideal binary masking (IBM) and the parametric Wiener filtering (PWF), for CI recipients' speech perception. In both algorithms, speech perception experiments were conducted to determine the perceptually optimal gain function. In comparison to IBM which requires only searching for an optimal gain threshold, PWF requires searching for two parameters (α and β). They also proposed a new metric called apparent gain threshold (aGT), and the relation between the Wiener parameters and aGT was also elaborated.

Inspired by Mauger's work, this study proposes an improved parametric Wiener filtering for CI noise suppression. Firstly, the relation between the aGT and the optimal Wiener parameters (α, β) is disclosed by a set of speech perception experiments, which tells that the searching for optimal (α, β) can be simplified to be searching for optimal aGT. Then, an aGT-dependent gain function is proposed for Wiener filtering. Speech perception experiments with vocoder simulations in normal-hearing listeners show that the proposed PWF algorithm outperforms the IBM with the same tuning complexity.

II. SNR-based noise suppresion

In this study, the noise is assumed to be additive to the target speech, i.e.,

$$y(n) = x(n) + d(n) \tag{1}$$

where $x(n)$, $d(n)$ and $y(n)$ denote the target speech, the noise and the noisy speech, respectively. In spectral domain, we have

$$Y(k,l) = X(k,l) + D(k,l) \tag{2}$$

where k and l denote the frequency bin and the time frame index, respectively.

To estimate the target speech spectrum from the noisy one, a classical way, i.e., the Wiener filtering, is to find an statistically optimal gain function $G(k,l)$ such that the target spectrum can be estimated as

$$\hat{X}(k,l) = G(k,l) \cdot Y(k,l) \tag{3}$$

The parametric Wiener filtering (PWF) have been widely adopted for speech enhancement, in which a signal-to-noise ratio (SNR) dependent gain function is given as [8]

$$G(k,l) = \left(\frac{\xi(k,l)}{\xi(k,l)+\alpha}\right)^{\beta} \tag{4}$$

where $\xi(k,l)$ is the priori SNR computed as

$$\xi(k,l) = \begin{cases} \frac{|Y(k,l)|^2}{|\hat{D}(k,l)|^2} - 1, & |Y(k,l)|^2 > \left|\hat{D}(k,l)\right|^2 \\ 0, & |Y(k,l)|^2 \le \left|\hat{D}(k,l)\right|^2 \end{cases} \tag{5}$$

in which $\left|\hat{D}(k,l)\right|^2$ is the estimated power spectrum of the noise. Besides the instantaneous noise estimate, the α and β variables are also need, usually pre-determined via training, for computing the gain function.

Wang et al proposed an ideal binary masking (IBM) technique for noise suppression, in which a binary gain function was defined as [9]

$$G_{BM}(k,l) = \begin{cases} 1, & \xi(k,l) \ge T_g \\ 0, & \xi(k,l) < T_g \end{cases} \tag{6}$$

where T_g is the gain threshold and also need to be pre-determinded.

In speech enhancement, the mathematically optimal gain function usually may not result in best subjective perception. Therefore, the variables α and β in (4) and T_g in (6) need to be experimentally determined with massive subjective evaluation trials. In noise suppression for cochlear implants, the hearing capability of the CI recipients is highly subject-dependent and the perceptually optimized variables should be determined individually.

The advantage of IBM for CI noise suppression is that it is much easier to determine the perceptually optimal T_g (only one variable). However, assigning each time-frequency component to either signal or noise is surely not optimal as the addition of two signals could happen in every time-frequency component. Mauger investigated the performance of the two gain functions as given in (4) and (6) for CI noise suppression [7]. As demonstrated, the perceptually optimal gain threshold or parameters for each CI recipient are highly subject-dependent, which means that parameters tuning are necessary in clinical implementations. The work also showed that the optimal noise suppression effect can be obtained with different (α, β)s and the authors proposed a new metric, i.e., the aGT, with which the results of PWF and IBM can be compared.

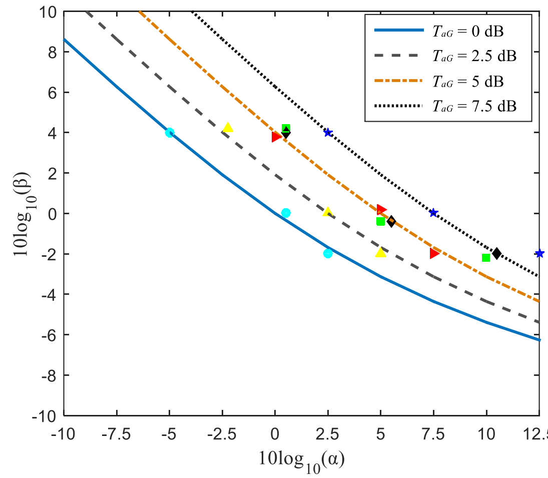

Fig. 1. Subject-dependent optimal parameters (α, β) for six subjects (each with specific marks) and the aGT curves.

III. Parametric Wiener Filtering Based on Apparent Gain Threshold

Based on Mauger's work, this study proposes a parametric Wiener filtering based on aGT for CI noise suppression. Following Mauger's definition of aGT "as the SNR that attenuates the input signal by half" [7], according to (4), the relation between the aGT and (α, β) can be given by

$$\frac{1}{2} = \left(\frac{T_{aG}}{T_{aG}+\alpha}\right)^{\beta} \tag{7}$$

where T_{aG} denotes the aGT. As showed in [7], the optimal gain functions can be obtained with different (α, β)s. For a particular subject, assuming that all the optimal (α, β)s give the same T_{aG}., searching for the optimal (α, β) could be simplified to be searching for the optimal T_{aG}..

This study first conducted a set of speech perception experiments to examine the feasibility of the assumption. Six NH subjects are recruited for the CI vocoder simulated listening tests. Three optimal (α, β)s are searched for each subject, the results are given in Fig. 1. To elaborate the relation between T_{aG} and (α, β), the aGT curves for T_{aG} equals to 0 dB, 2.5 dB, 5 dB and 7.5 dB are also given in the figure. One can see that for most subjects, the subject-dependent optimal (α, β)s lie in a specific aGT curve.

Since all the (α, β)s in the same aGT curve give the same contribution to the noise suppression effect, if the optimal aGT is known for a specific subject, any points in the aGT curve could be the optimal (α, β) for the subject. For simplicity, fix the β variable to be 1, i.e., $10\log_{10}\beta=0$, form (7) we have

$$\alpha = 10^{(T_{aG}/10)} \tag{8}$$

Now the gain function in (4) can be computed as

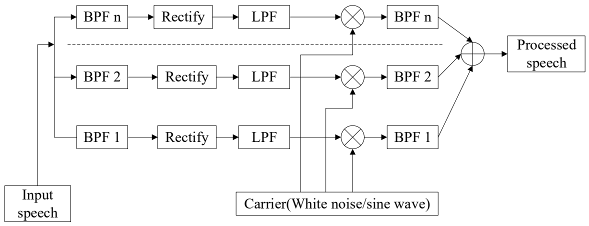

Fig. 2. Signal processing procedure for vocoder simulation

$$G(k,l) = \frac{\xi(k,l)}{\xi(k,l)+10^{(T_{aG}/10)}} \tag{9}$$

In this case, the perceptually optimal noise suppression could be achieved by tuning the T_{aG}, instead of the two parameters α and β.

IV. EXPERIMENTS

To evaluate the effectiveness of the proposed noise suppression method for CI, speech perception experiments are conducted with NH subjects using CI vocoded speech input.

A. Experiment Design

1. Signal processing procedure for vocoder simulation

Signal processing strategies for modern clinical CI devices are mostly temporal envelope-based, e.g., the continuous interleaved sampling (CIS) strategy [10]. This study adopts the CIS strategy. The vocoder simulation of CIS follows that presented in [11] and is depicted in Fig. 2. The input speech is first passed through a filter bank, which divides the fullband (80-7999 Hz) into eight subbands (channels) according to the Greenwood function [12]. In each channel, the subband output is half-wave rectified and low-pass filtered (with cutoff frequency at 250 Hz) to generate the subband envelope. The envelope is then used to modulate the tone carrier. Finally, the modulated carriers in each channel are bandpass filtered and level matched (to the input signal from the corresponding subband) and summed to produce the vocoded stimulus.

2. Participants and Speech Materials

The participants in the experiment are college students, 6 females and 6 males, all with normal hearing. Table 1 lists the details, i.e., age, gender and speech reception threshold (SRT) of the 12 participants. The SRT of a subject refers to the SNR of the speech signal that the subject can recognize at least 50% words in the speech sentence. The determination of SRT for each participant follows the same procedure as in [13].

The speech materials are sentences taken from the Mandarin Hearing in Noise Test (MHINT) database [14]. The database contains 40 training sentences and 240 test sentences, each consisting of 10 Mandarin words. The training sentences are for measuring the subject-dependent SRTs. In the experiment, subject-dependent SNR is adopted to generate the noisy signal and the SNR is set to be subject-dependent SRT minus 1 dB. Noise signals used in this experiment are the speech-spectrum shaped noise (SSN) and the 11-talker babble noise (BBN).

Table I: Description for the 12 subjects in this study

Subject	Gender	Age (yr)	SRT for SSN (dB)	SRT for BBN (dB)
1	male	22	0.5	-3.0
2	male	25	3.5	-2.5
3	female	23	3.0	-2.0
4	female	23	1.0	-3.0
5	male	26	3.0	-4.5
6	male	22	3.0	-1.5
7	female	23	-1	-3.5
8	female	23	1.5	-3
9	male	24	0	-3.5
10	female	24	-1.5	-3
11	male	23	2.5	-4.5
12	female	23	-2.5	-2

3. Test conditions

This study evaluated the performance of two noise suppression methods, i.e., the proposed parametric Wiener filtering with apparent gain threshold (PWF) and the ideal binary masking with gain threshold (IBM). Noise estimate is also essential in noise suppression. Two noise estimate methods, i.e., the improved minima controlled recursive averaging (Imcra) method proposed by Cohen [13] and the ideal noise estimate (that is, assuming the noise spectrum is known, Ideal), are adopted to compare the performance of PWF and IBM at different noise estimate confidence. Therefore, there are in total four noise suppression strategies, i.e., I. Ideal + IBM, II. Ideal + PWF, III. Imcra + IBM and IV. Imcra + PWF, to be compared.

For each type of noise, there are six test conditions, i.e., noisy speech without noise reduction (noted as Un), denoised speech with gain threshold (T_g for IBM) or apparent gain threshold (T_{aG} for PWF) of -5dB, 0dB, 5dB, 7.5dB and 10dB. Each condition has 5 test sentences. Therefore, each subject is presented 240 different sentences (4 strategies, 2 noises, 6 test conditions and 5 sentences at each condition).

The experiments are carried out in a soundproof room. The denoised speech is first passed through the vocoder simulation system to generate the vocoded signal and then is presented to the subject via a Roland Quad-Capture UA-55 audio interface and a Sennheiser HD 650 headset. The sounds are presented at a comfortable level (approximately 70 dB). The subject is instructed to repeat as much of the sentence as possible. Each sentence could be presented up to three times upon the response of the subject. No feedback about the correctness of the responses is given during the experiment.

B. Results

Figure 3 shows the mean word recognition rates of the 12 subjects at the 6 test conditions for the 4 different strategies. The target speech signals are corrupted by SSN (Fig. 3 (a)) or BBN (Fig. 3 (b)). Without noise suppression, the recognition

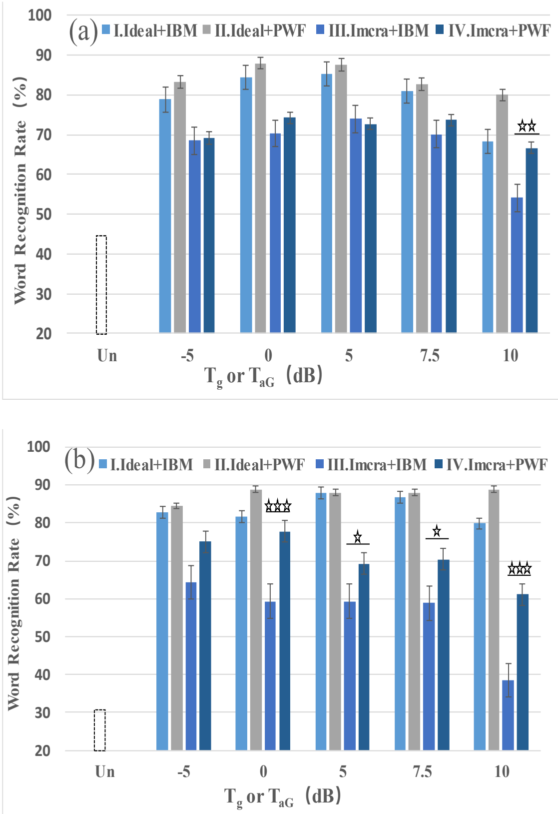

Fig. 3. Group mean recognition rates for the four noise suppression strategies at different GT or aGT. (a) Speech-spectrum shaped noise; (b) Babble noise. Error bars are the standard deviation of group means. Statistical significance between IBM and PWF, with the same noise estimate method, is shown by bars with asterisks denoting the level of the significance (***, **, * representing p<0.001, p<0.01 and p<0.05, respectively). No significance if there is no bar.

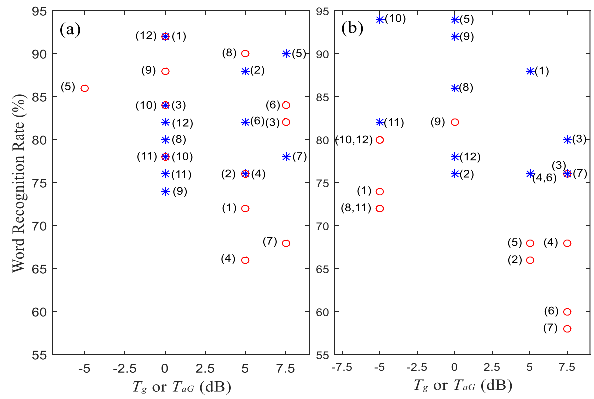

Fig. 4. Word recognition rates for two noise suppression strategies at subject-dependent optimal thresholds. (a) Speech-spectrum shaped noise; (b) Babble noise. The Mark (8, 11)○ means that both subject 8 and subject 11 obtain best recognition rate about 72% at $T_g = -5$ dB; ∗ (8) means that subject 8 obtains best recognition rate about 86% at $T_{aG} = 0$ dB.

rates are only 43.7% for SSN and 30.2% for BBN. For both IBM and PWF, the performances with ideal noise estimate are much better than that with Imcra, which is reasonable since noise estimate is critical for SNR-based noise suppression.

With known noise spectrum, the difference between IBM and PWF is insignificant for both noise types. On the other hand, when the noise estimate is not perfect, significant differences between IBM and PWF are observed. For example, strategy III and IV perform comparably for SSN at most threshold conditions, except that strategy IV significant outperforms strategy III at threshold of 10 dB. For BBN, strategy IV shows significant superiority over strategy III at all thresholds except at threshold of -5 dB.

To further investigate the effectiveness of the proposed method for each subject with subject-dependent aGT, Fig. 4 gives the best word recognition rate achieved by each subject with PWF or IBM for both noise types. As shown, for SSN, 6 subjects perform better with PWF and the remaining 6 subjects perform better with IBM; for BBN, however, most subjects except subject 12 obtain better recognition rate with PWF than with IBM. In SSN, the average recognition rates of the 12 subjects are 80.5% and 81.7% for IBM and PWF, respective; in BBN, the average rates are 71.3% and 78% for IBM and PWF, respectively.

V. Conclusions

In this study, we propose an improved parametric Wiener filtering for CI noise suppression. By adopting the apparent gain threshold, searching for the perceptually optimal parameters (i.e., α and β) can be simplified to be searching for the perceptually optimal aGT (i.e., T_{aG}). This could be advantageous in clinical CI implementation where tuning for subject-dependent perceptually optimal parameters (which is usually very time consuming) could be crucial for postoperative rehabilitation. Perceptual experiments with vocoder simulation in normal-hearing listeners demonstrate the potential advantages of the proposed algorithm in clinical CI implementation.

Acknowledgment

This work is jointly support by National Natural Science Foundation of China (Project 61771320, 11704129) and Shenzhen Science and Technology Foundation (Project JCYJ20170302145906843). Qinglin Meng is the corresponding author.

References

[1] K. Nie, G. Stickney, and F.G. Zeng, “Encoding frequency modulation to improve cochlear implant performance in noise,” *IEEE Trans. Bio.Med. Eng.*, vol. 52, pp. 64–73, 2005.

[2] K. Kokkinakis and P.C. Loizou, “The impact of reverberant self-masking and overlap-masking effects on speech intelligibility by cochlear implant listeners,” *J. Acoust. Soc. Am.*, vol. 130, pp. 1099–1102, 2011.

[3] 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 & Hearing*, vol. 32, pp. 382-390, 2011.

[4] M.K. Qin, and A.J. Oxenham, "Effects of simulated cochlear implant processing on speech reception in fluctuating maskers," *J. Acoust. Soc. Am*., vol. 114, pp. 446-454, 2003.

[5] R. Koning, N. Madhu, J. Wouters, "Ideal time-frequency masking algorithms lead to different speech intelligibility and quality in normal-hearing and cochlear implant listeners," *IEEE Trans. Biomedical Engineering*, vol. 62, pp. 331-341, 2015.

[6] R. Huber, T. Bisitz, T. Gerkmann, J. Kiessling, H. Meister, B. Kollmeier, "Comparison of single-microphone noise reduction schemes: Can hearing impaired listeners tell the difference?" *International Journal of Audiology*, *early online*, pp. 1-7, 2017.

[7] S.J. Mauger, P.W. Dawson, and A.A. Hersbach, "Perceptually optimized gain function for cochlear implant signal-to-noise ration based noise reduction," *Acoustical Society of American*, vol. 131(1), pp. 327-336, 2012.

[8] J.S. Lim, and A.V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proc. IEEE*, vol. 67, pp. 1586-1604, 1979.

[9] D.L. Wang, U. Kjems, M.S. Pedersen, J.B. Boldt, and T. Lunner, "Speech intelligibility in background noise with ideal binary time-frequency masking," *J. Acoust. Soc. Am*. vol. 118, pp. 2336-2347, 2009.

[10] B. S. Wilson, C. C. Finley, D. T. Lawson, R. D. Wolford, D.K. Eddington, and W. M. Rabinowitz, "Better speech recognition with cochlear implants," *Nature*, vol. 352, pp. 236–238,1991.

[11] N.A. Whitmal, S.F. Poissant, R.L. Freyman, and K.S. Helfer, "Speech intelligibility in cochlear implant simulations: Effects of carrier type, interfering noise, and subject experience," *J. Acoust. Soc. Am*, vol. 122(4), pp. 2376-2388, 2007.

[12] D.D. Greenwood, "A cochlear frequency-position function for several species-29 years later," *J. Acoust. Soc. Am*, vol. 87(6), pp. 2592-2605, 1990.

[13] Q.L. Meng, N.H. Zheng, and X. Li, "Mandarin speech-in-noise and tone recognition using vocoder simulation of the temporal limits encoder for cochlear implants," *J. Acoust. Soc. Am*. vol. 139(1), pp. 301-310, Jan. 2016.

[14] L. L. Wong, S. D. Soli, S. Liu, N. Han, and M. W. Huang, "Development of the mandarin hearing in noise test (MHINT)." *Ear& Hearing*, vol. 28, no. 2 Suppl, pp. 70S–74S, 2007.

[15] I. Cohen, "Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging," I*EEE Trans. Speech Audio Process*. vol.11, pp. 75-466, 2003.