A Single-Channel Noise Reduction Algorithm for Cochlear-Implant Users

A THESIS
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Ningyuan Wang

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF OR SCIENCE

Advisor: Prof. Tom Luo

December, 2015

Acknowledgements

First and foremost, I would like to thank my advisor in Electrical Engineering, Professor Tom Luo, for his time, guidance and continuous support throughout this master's thesis.

I would like to thank Professor Andrew J. Oxenham, my advisor in Department of Psychology, for his valuable comments on an earlier version of this work.

I would like to thank Heather Kreft, my lab mate, who helped me collect the experimental data from the human subjects.

Finally, I would like to thank all the research participants, for their willingness to contribute to my work. This work could not have been completed without their support from all of them.

Dedication

This thesis is dedicated to my parents and all the subjects who participated in this study.

Abstract

Despite good performance in quiet environments, there are still significant gaps in speech perception in noise between normal-hearing listeners and hearing-impaired listeners using devices like hearing aids or cochlear implants (CIs). Much effort has been invested to develop noise reduction algorithms that could fulfill these gaps, but few of them have the ability to enhance speech intelligibility without any prior knowledge of the speech signal, including both statistical properties and location information. In this study, a single-channel noise reduction algorithm, based on a noise tracking algorithm and the binary masking (BM) method, was implemented for CI users. The noise tracking algorithm was able to catch detailed spectral information of the noise with a fast noise tracker during the noise-like frames and update the estimated accumulative noise level with a slow noise tracker during speech-like frames. Next, this noise tracking algorithm was used to estimate the signal-to-noise ratio (SNR) of each temporal-spectral region, termed “time-frequency unit” in the BM method, to determine whether to eliminate or retain each unit. Finally, a sentence perception test was employed to investigate the effects of this noise reduction algorithm in various types of background noise and input SNR conditions. Results showed that the mean percent correct for CI users is improved in most conditions by the noise reduction process. Improvements in speech intelligibility were observed at all input SNR conditions for the babble and speech-shaped noise conditions; however, challenges still remain for the non-stationary restaurant noise.

Table of Contents

Chapter I: Introduction

1.1 Speech perception in cochlear-implant users

Cochlear implants (CIs) are the only reliable medical intervention that can help restore partial hearing to a totally deafened person. It converts acoustic signals into electrical pulses to stimulate residual auditory nerves to generate a sensation of hearing. The modern day CI allows many of its users to communicate in a quiet environment, as well as on the phone.

When a person becomes severely hearing impaired, very limited sensory information is received by the central nervous system (CNS), making it difficult to understand speech. However, according to previous studies, considerable information can be deleted from a speech signal with only minor deleterious effects (Shannon *et al.*, 1995). This may explain why so many CI users understand speech in quiet background so well.

Despite the good performance in quiet environment, there are still significant gaps in performance between normal-hearing people and CI users. For example, the performance of CI users for speech perception tasks with additive noise is extremely poor. At least a 15-dB loss in functional signal-to-noise ratio (SNR) can be produced in a steady-state noise background (Zeng *et al.*, 2005). Music perception is also extremely limited in CI users. Although they can access some rhythmic information, little melody and timbre information is received (McDermott, 2004).

In a study by Firszt *et al.* (2004), speech recognition was assessed using the Hearing in Noise Test (HINT) sentences (Nilsson *et al.*, 1994). Results revealed that CI recipients' performance on sentence recognition tasks was significantly poorer in noise compared with when listening at speech stimuli in quiet conditions at a soft level. In another study by Spahr and Dorman (2004), it was reported that for speech material presented at +10 dB SNR, the average speech intelligibility performance of CI recipients decreased to 70% on tasks using clean speech and to around 40% in tasks involving conversational speech. After the SNR level was lowered to +5 dB, recognition of conversational speech, on average, dropped to around 20%. Fetterman and Domico (2002) revealed a similar trend in their study; on average, CI recipients' sentence recognition scores decreased from 82% correct in quiet environment to73% at a +10 dB SNR level and to around 47% at +5 dB SNR.

Poor frequency selectivity in hearing impaired listeners has been reported as a significant factor in their inability to distinguish a speech signal in noise as compared to normal-hearing listeners (Summers and Al-Dabbagh, 1982; Baer and Moore, 1994). Recent research efforts have been focusing on state-of-the-art noise reduction solutions to improve speech intelligibility in noisy environments. Since CIs deliver electrical pulses to stimulate auditory nerves to help restore hearing sensation, multiple signal processing algorithms have been applied to convert acoustic signals into electrical stimuli (e.g. Loizou *et al.*, 2000; Zeng, 2004). As indicated previously, the majority of CI users can achieve high open-set speech recognition scores in quiet environment, regardless of the device or speech coding strategy used (e.g. Skinner *et al.*, 2002; Spahr and Dorman, 2004); however, few of them are able to overcome the problem of additive noise.

1.2 Noise reduction in cochlear-implants

Noise reduction is crucial for hearing impaired listeners to understand speech in background noise. With the help from medical devices like hearing aids or CIs, some of them may achieve nearly perfect speech recognition in quiet conditions. Unfortunately, this ability normally drops sharply with the interference of background noise (Moore *et al.*, 1985; van Tasell, 1993; Hamzavi *et al.*, 2001; Chung, 2004; Zeng, 2004).

Many of the current noise reduction algorithms can improve the output SNR, but few of them improve speech intelligibility (e.g. Li and Loizou, 2008; Kim and Loizou, 2011; Brons *et al.*, 2014). The speech distortions generated by the noise reduction processes have been considered as a main contributor to the lack of success. A traditional method used to reduce the effect of noise is to apply gain calculated from estimated SNR level to suppress the noise. However, since the power of noise cannot be accurately estimated, the speech signal can be either amplified or attenuated due to the underestimation or overestimation to the noise power, respectively. Considerable speech distortion is then introduced (Kim and Loizou, 2011), resulting in no or even negative benefits in speech intelligibility (van Tasell, 1993; Hu and Loizou, 2007; Chen *et al.*, 2012).

In general, noise reduction algorithms designed for CIs can be divided into two classes: single-microphone and multi-microphone methods. Single-microphone approaches rely mostly on statistical models of speech and noise, and therefore can only remove noise with different temporal and spectral features as speech signals (e.g., Hu & Loizou, 2002; Hu, Loizou, Li, & Kasturi, 2007; Yang & Fu, 2005). In recent years, there has been a growing tendency toward the use of noise reduction methods with multi-microphones in CI devices (e.g. van Hoesel and Clark, 1995; Wouters and Vanden Berghe, 2001; Chung, 2004; Kokkinakis and Loizou, 2008). Large improvements in SNR and, therefore considerable benefits in speech intelligibility can be obtained, but only if the target speech and the noise sources are located at different locations in space. However, if the locations of target signal source and noise source overlap, or are unknown, little or even negative benefits are expected. In that circumstance, single-channel noise reduction algorithms become more practical.

One of the first proposed single-channel noise reduction algorithms is spectral subtraction (SS). It is based on a maximum-likelihood (ML) estimator, and has been implemented in numerous applications. Its efficiency and low computational complexity have resulted in its widespread use. The general idea of SS is to estimate the spectrum of noise during gaps in the speech, and then to remove it from the noisy signal (Vary, 1985). The performance of this class of algorithms depends critically on the accuracy of the noise estimation. A conventional method to estimate the noise is to update the noise power cumulatively during speech gaps and hold it unchanged during speech frames. To determine when speech is present or absent, a voice activity detector (VAD) is required (Boll, 1979).

Despite its attractive simplicity, SS has a number of short-comings, which limit its use in applications such as hearing aids and CIs. One commonly audible artifact produced by the technique is termed “musical noise” and is generated primarily by inaccurate estimation of the noise spectrum (Goh *et al.*, 1998; Seok and Bae, 1999; Gustafsson *et al.*, 2001). Since only the average power of the noise can be estimated over time, short-term, potentially important details, such as momentary spectral peaks and valleys, are ignored by the algorithm. As a result, after the estimated average noise is removed, those residual components can produce annoying tonal sounds, which can worsen, rather than improve, speech intelligibility and perceived quality.

Over the past three decades, much effort has gone into developing methods that remove or reduce musical noise (Crozier *et al.*, 1993; Beh and Ko, 2003; Plapous *et al.*, 2006). A common technique is to set up a noise floor, that when the signal level is under a certain threshold, it will be left unprocessed (Boll, 1979). A factor to determine how much of the signal should be removed when it falls under the threshold level could be used to control the strength of this technique. Moreover, it has been noted that the influence of noise on a speech signal may not be unified, thus it is rational to apply different control factors at different frequencies (Lockwood and Boudy, 1992).

Wiener filtering is another well-studied technique in speech enhancement and is based on an optimal minimum mean-square error (MMSE) estimator of each speech spectral component (Lim and Oppenheim, 1979; Spriet *et al.*, 2004; Doclo and Moonen, 2005; Chen *et al.*, 2006). Martin (1994) developed an algorithm based on the combination of Wiener filtering and SS to overcome the limitation of VAD, so that in a speech frame, noise information could still be updated by a Wiener filter. Extended Wiener filters have been proposed to further enhance the performance of noise reduction algorithms. Multichannel Wiener filtering, for example, has been tested (Doclo *et al.*, 2007; Van Dun *et al.*, 2007; Van den Bogaert *et al.*, 2009a). In addition, the speech-distortion-weighted Wiener filter was developed to reduce speech distortion, in order to maintain speech intelligibility while cleaning speech (Spriet *et al.*, 2004; Doclo and Moonen, 2005).

The binary masking (BM) method originated from auditory scene analysis has been investigated for its ability to improve speech intelligibility for both normal-hearing and hearing-impaired listeners (e.g. Li and Loizou, 2008; Wang *et al.*, 2008; Wang *et al.*, 2009; Roman and Woodruff, 2011). The idea of the BM is to decompose the signal spectrum into a two-dimensional matrix, along time domain and frequency domain, in which each element represents a time-frequency (T-F) unit. For each unit, the corresponding local SNR is estimated. Then if the SNR is equivalent to or above a certain threshold, the gain of this unit is set to one. In contrast, a gain of zero is applied to the whole unit, if the unit is dominated by noise (Li and Loizou, 2008; Wang *et al.*, 2008).

It has been well demonstrated that the BM method can improve speech intelligibility, if the statistical information of the speech and noise signals are accessible before mixing, which has been termed the “ideal binary masking” (IBM) (e.g. Anzalone *et al.*, 2006; Li and Loizou, 2008; Wang *et al.*, 2008; 2009). For example, Li and Loizou (2008) found that the intelligibility benefit from the IBM manipulation is 7 dB under speech-shaped-noise masking, 10 dB under modulated-speech-shaped-noise masking, and 15 dB under two talker-speech masking. Comparable results were also observed in Wang *et al.* (2009); speech intelligibility was improved by 11 dB and 7 dB when perceiving speech in cafeteria noise and in speech-shaped-noise (SSN), respectively. More interestingly, it has been shown that speech intelligibility can be further improved by adding background noise at moderate levels (Cao *et al.*, 2011).

However, in previous implementations of the BM method, a training session is normally required before processing, which means the prior statistical knowledge of the stimuli is required (Wang *et al.*, 2008; Kim *et al.*, 2009; Wang *et al.*, 2009; Healy *et al.*, 2013). In real applications, it is unrealistic to have all the required information available. The challenge then becomes how to implement the BM algorithm without prior knowledge.

In this study, a spectral-domain noise tracking algorithm, with a slow ML-based noise tracker and a fast Kalman-based noise tracker, which are able to update the average noise level during speech frames and to capture the detailed fluctuations of the noise signal during speech gaps, respectively, was developed. Next, this noise-tracking algorithm was used to estimate local SNR in each T-F unit, in order to remove those noise-dominated units without prior knowledge or statistical training. In chapter II, the basic theory and implementation of the noise tracking algorithm are described, along with its objective evaluations. In chapter III, a single-channel noise reduction algorithm based on the noise tracking algorithm and the BM method is introduced. The effectiveness of this noise reduction algorithm is evaluated by speech perception tests for CI users. The experiments and results are reported in chapter IV. Finally, in chapter V, conclusions are given.

Chapter II: Noise Tracking Algorithm based on Kalman filtering

2.1 Theory

The mixed noise problem is first formulated as in (2.1), where s(n) and v(n) represent the desired speech signal and additive noise respectively. v(n) is assumed to be stationary here. The x(n) is the input noisy signal, which is converted into the frequency-domain with a short-time Fourier transform (STFT) analysis. The proposed spectral-domain Kalman noise tracking algorithm (KNT) is illustrated in Figure 2.1. As with most other speech enhancement methods, a noise estimator and a gain calculator are required. In order to capture sufficient detail in the noise estimation and to keep the speech distortion as low as possible, two noise trackers constitute the noise estimator. N_s is the time-averaged slow-noise tracker, which reflects the cumulative average level of the noise spectrum, whereas N_f is a fast-noise tracker designed to acquire the fine structure of noise spectrum.

$$x(n) = s(n) + v(n) \tag{2.1}$$

In this noise tracking algorithm, when there is no speech present, the input signal, x(n), is defined as the observation process in (2.2), and the additive noise, v(n), is seen as the unobserved state process in (2.3). u(n) and w(n) represent the observation noise and the process noise, respectively.

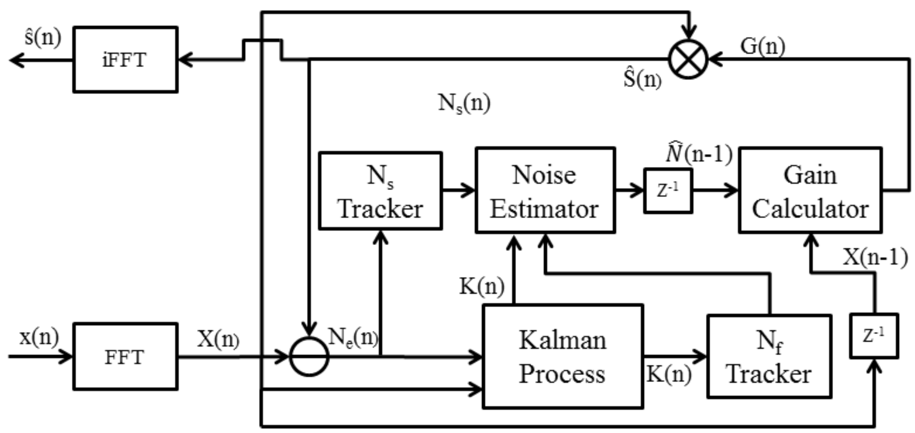

Fig. 2.1 General block diagram of KNT

The frequency-domain two-step Kalman process involves in both N_f updating and the final noise estimation ($\hat{N}$). The Kalman gain, K, varies between 0 and 1, and is used for indicating whether to weight the previous measurement more than the current observation or to weight them in a contrary way, respectively. It also determines the weights of N_s and N_f when estimating the noise ($\hat{N}$). So if the current frame is a noise-like frame, the value of K would be close to 1, and vice versa.

The last procedure is to update the gain with a continuous gain function based on the ML estimation of the noise level. The final cleaned output signal is returned to the noise estimator to update noise information for the next frame.

$$x(n) = v(n) + u(n) \quad \text{where} \quad u(n) \sim \mathcal{N}(0, R) \tag{2.2}$$

$$v(n) = v(n-1) + w(n) \quad \text{where} \quad w(n) \sim \mathcal{N}(0, Q) \tag{2.3}$$

As shown in (2.4) and (2.5), for the l^{th} time frame and k^{th} frequency bin, the current observation of noise, $N_e(k,l)$, can be obtained from input signal $X(k,l)$ and the speech signal, $\hat{S}(k,l-1)$, which is estimated from previous frame. $N_s(k,l)$ is then updated using this observation, where α is a smoothing factor.

$$N_e(k,l) = X(k,l) - \hat{S}(k,l-1) \tag{2.4}$$

$$N_s(k,l) = (1-\alpha)N_s(k,l-1) + \alpha N_e(k,l) \tag{2.5}$$

In a standard Kalman filter, Q and R correspond to the covariance of the process noise and the measurement noise, respectively. In KNT, they are computed as in (2.6) and (2.7). $Q(k,l)$ is estimated by the squared difference between $N_e(k,l)$ and

$N_e(k, l-1)$, referring to the variation of process "noise" in the system. The purpose is to track the noise level, so the variation of noise should be treated as the process noise. Similarly, $R(k,l)$ is calculated by the squared difference between $X(k,l)$ and $N_s(k,l)$, representing the observation noise. In this case, any speech signal could interfere with the noise tracking process, which means the speech signal is the observation "noise". If the current input signal, $X(k,l)$, is much stronger than the slow-noise tracker, $N_s(k,l)$, it means the speech signal is presumably contained in this frame, which triggers an enhancement in $R(k,l)$ and a reduction in $Q(k,l)$, resulting in a small Kalman gain, $K(k,l)$. Equations (2.8) - (2.12) are from standard Kalman procedures; aiming to recursively calculate $K(k,l)$ and the fast-noise tracker, $N_f(k,l)$. $N^-(k,l)$ represents the previous measurement, which is updated in each loop. β and γ are smoothing factors of Q and R, respectively.

$$Q(k,l) = (1-\beta)Q(k,l-1) + \beta\left(\frac{N_s(k,l)}{X(k,l)}\right)^2 \left(N_e(k,l) - N_e(k,l-1)\right)^2 \quad (2.6)$$

$$R(k,l) = (1-\gamma)R(k,l-1) + \gamma\left(X(k,l) - N_s(k,l)\right)^2 \quad (2.7)$$

$$P^-(k,l) = P(k,l-1) + Q(k,l) \quad (2.8)$$

$$K(k,l) = \frac{P^-(k,l)}{P^-(k,l)+R(k,l)} \quad (2.9)$$

$$N_f(k,l) = N^-(k,l) + K(k,l)\left(N_e(k,l) - N^-(k,l)\right) \quad (2.10)$$

$$P(k,l) = P^{-}(k,l)\big(1 - K(k,l)\big) \tag{2.11}$$

$$N^{-}(k,l+1) = N_f(k,l) \tag{2.12}$$

$$\widehat{N}(k,l) = N_s(k,l) + K(k,l)\big(N_f(k,l) - N_s(k,l)\big) \tag{2.13}$$

Both $N_f(k,l)$ and $N_s(k,l)$, along with $K(k,l)$, are required when estimating the current noise level, $\widehat{N}(k,l)$. In a speech-like frame, a small $K(k,l)$ is obtained, which means the noise estimator trusts $N_s(k,l)$over $N_f(k,l)$. Otherwise, in a noise-like frame, $N_f(k,l)$ is trusted so as to catch the fine structure of noise. This estimation procedure is completed in a preceding frame as in (2.13).

$$W(k,l) = \frac{X(k,l)}{\widehat{N}(k,l)} \tag{2.14}$$

$$G(k,l) = (1 - \delta)G(k,l-1) + \delta\frac{W(k,l)-1}{W(k,l)} \tag{2.15}$$

$$\widehat{S}(k,l) = X(k,l)G(k,l) \tag{2.16}$$

A continuous gain is calculated in (2.14) and (2.15). δ is a smoothing factor of gain, $G(k,l)$. The estimated speech signal in (2.16), $\widehat{S}(k,l)$, is returned to (2.4) for the next frame.

2.2 Implementation

A sampling rate of 22050 Hz is applied in this implementation. The window length is 25 ms with 50% overlap. The signal is an IEEE sentence, "The lease ran out in sixteen weeks," spoken by a male speaker. The noise used in this illustration is SSN at +5-dB SNR. For comparison and evaluation, a simple SS algorithm with a perfect VAD was implemented. As shown in the top panel of Figure 2.2, the slow-noise tracker and the fast-noise tracker are presented in cyan and green curves, respectively. The estimated noise by KNT is shown in the blue curve. In a speech-like frame, with a strong speech signal, the final estimated noise level is much closer to that of the slow-noise tracker, while in a noise-like frame, its curve almost coincides with that of the fast-noise tracker. This specific behavior of the estimated noise curve is determined by the inverse of a posteriori SNR estimation as in (2.6); the ratio of the noisy signal power to the estimated noise power. When a high value of a posteriori SNR estimation is found, it means the current input signal is stronger than the estimated noise. In that case, the slow-noise tracker should be trusted over the fast-noise tracker, and vice versa. In the bottom panel of Figure 2.2, it is illustrated that the estimated noise by KNT is able to follow the detailed fluctuations of the noise in noise-like frames, whereas the cyan curve of the SS algorithm reflects only the average spectrum of the noise.

2.3 Illustration and evaluation

In Figure 2.3, the mean-square error (MSE) of noise estimation for both KNT and SS is shown. The MSE for each frequency bin is calculated across all time frames and converted into dB. The MSEs of SS and KNT are represented as blue curve and green curve, respectively. A critical point seems to be reached at about 3500 Hz, whereby the MSE of KNT is lower than that of SS below 3500 Hz, but is higher than that of SS above 3500 Hz. One possible explanation for this pattern is the energy distribution of speech, which has most energy concentrated at lower (< 3500 Hz) frequencies. A weak speech signal means inaccurate speech-like frame identification, resulting in inaccurate noise estimation. Given that most useful information for speech perception is provided by the low-frequency region (Baer *et al.*, 2002), this is unlikely to be an important issue in practice.

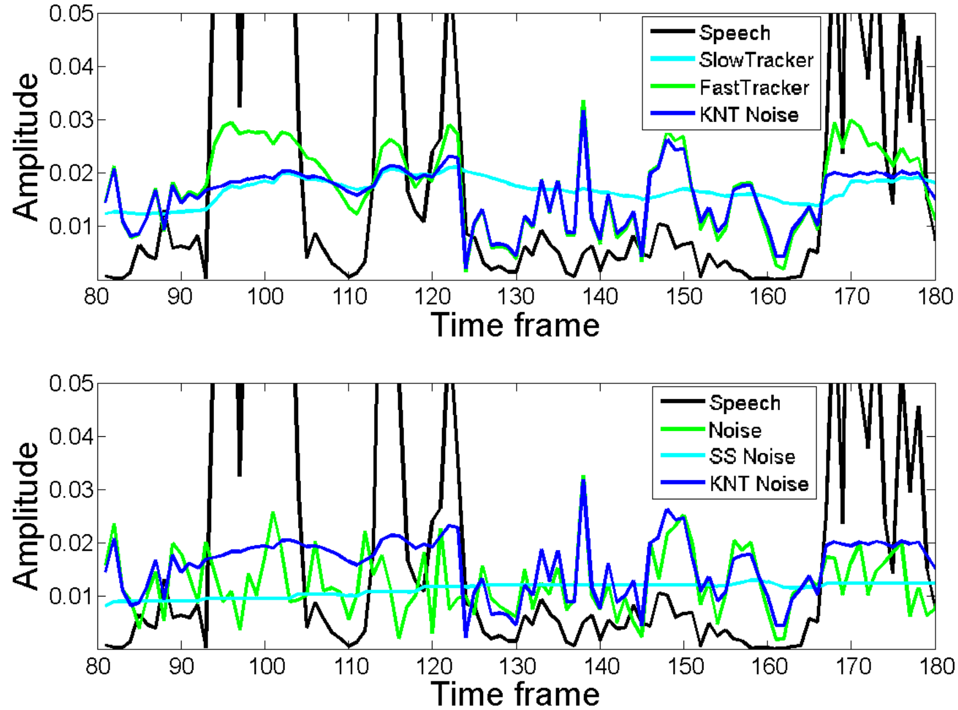

Fig. 2.2 Noise estimation by KNT (the 160th frequency bin, frame 80 to 180 of a sentence, "The lease ran out in sixteen weeks," spoken by a male speaker). The top panel presents the slow-noise tracker (N_s, cyan), fast-noise tracker (N_f, green) and also the estimated noise by KNT ($\hat{N}$, blue). In the bottom panel, the estimated noise ($\hat{N}$, blue) by KNT is compared with the estimated noise by SS with a perfect VAD (cyan). The original input noise is displayed in green.

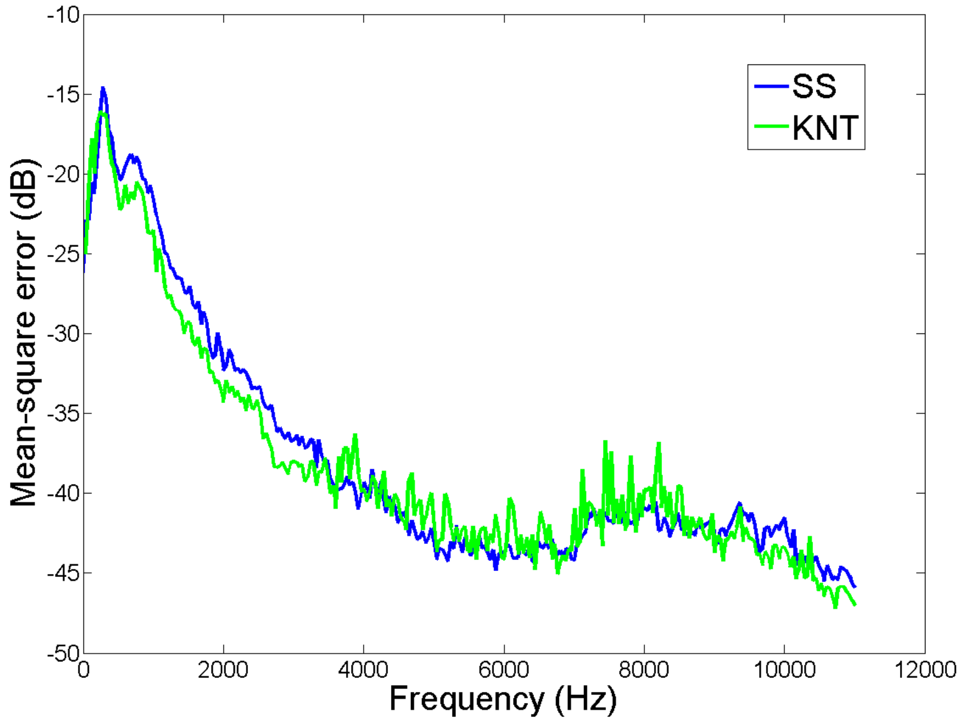

Fig. 2.3 MSE of noise estimation for KNT and SS as a function of frequency (Hz).

Table 2.1 Output SNRs in different noise and input SNR conditions.

Noise	Input SNR (dB)	KNT(dB)	MMSE(dB)	WF(dB)	SS(dB)
SSN	0	8.40	7.14	11.7	6.30
	5	13.4	11.5	14.4	10.8
	10	18.0	15.5	17.7	15.3
Babble	0	5.40	3.84	6.46	4.32
	5	11.2	8.99	11.3	9.41
	10	16.6	13.4	15.4	14.3
Restaurant	0	3.72	2.05	3.17	3.28
	5	10.0	7.55	9.42	8.62
	10	15.6	12.5	14.6	13.6

In the comparison experiment, the test material contained 80 randomly selected IEEE sentences, spoken by one male speaker and one female speaker. The KNT algorithm was tested under three noise conditions: SSN, babble noise (20 talkers), and restaurant noise. The results from standard SS, Wiener filter (WF) (Plapous *et al.*, 2006) and a minimum-mean square error (MMSE) algorithm (Ephraim and Malah, 1984; 1985) were evaluated for comparison.

Table 2.2 Output PESQ scores in different noise and input SNR conditions.

Noise	Input SNR (dB)	KNT	MMSE	WF	SS
SSN	0	1.47	1.61	1.19	1.38
	5	2.07	2.06	1.79	1.79
	10	2.43	2.45	2.37	2.16
Babble	0	1.55	1.55	0.89	1.58
	5	1.96	1.92	1.61	1.93
	10	2.30	2.32	2.14	2.25
Restaurant	0	1.56	1.60	1.02	1.59
	5	1.99	1.98	1.54	1.93
	10	2.26	2.34	2.09	2.24

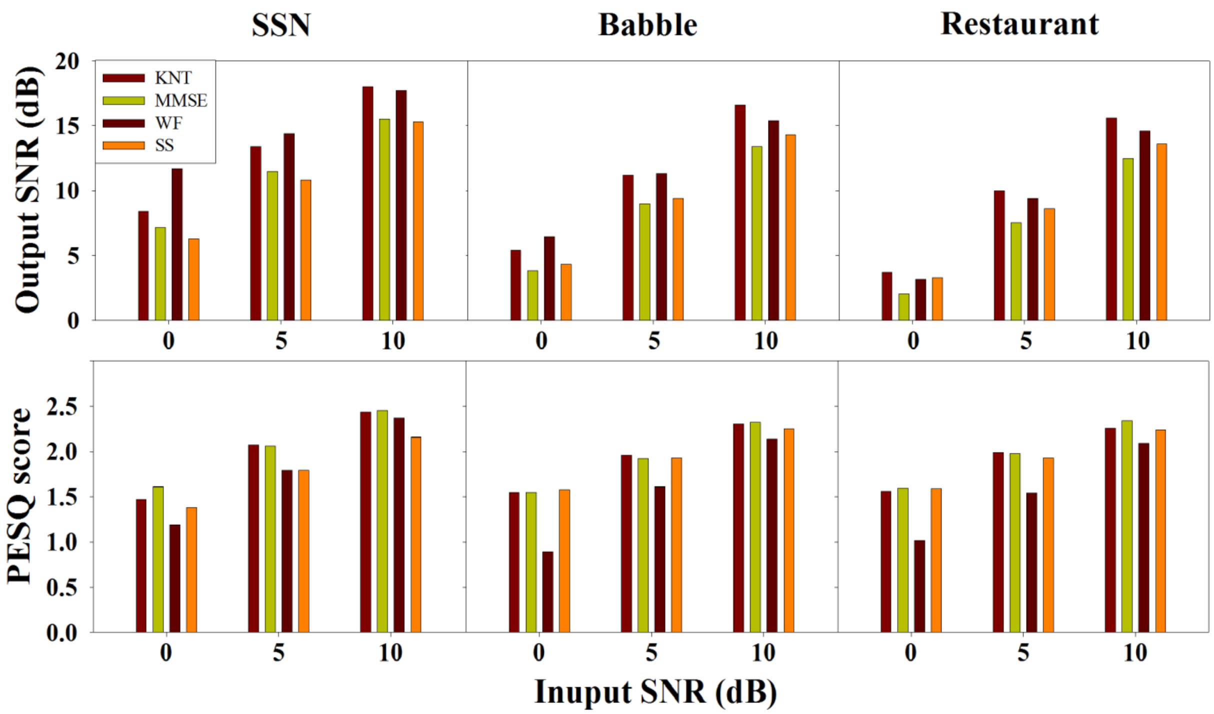

Fig. 2.4 Output SNRs and PESQ scores in different noise conditions as a function of input SNR. The average output SNRs are presented in top panels. Average PESQ scores are shown in bottom panels. From left to right, each column represents a different noise condition (SSN, babble noise and restaurant noise).

The input SNR varied from 0 dB to 10 dB in 5-dB steps. Estimated output SNR and objective PESQ score (Beerends *et al.*, 2002; Rix *et al.*, 2002) were selected as two criteria for the algorithm performance. The output SNR was estimated by first applying the gain in (2.15) in each frame on both clean speech and pure noise signals, then calculated the ratio of their power. As shown in the top panel of Figure 2.4, the output SNRs were enhanced by all the four algorithms in all noise conditions. The improvements made by KNT were generally larger than those produced by SS and MMSE. The average output SNR in each noise condition (noise type and input SNR) was treated as a sample, and a one-way ANOVA test was run, indicating there was a significant main effect on noise reduction method [$F(3,24)=22.683$, $p<0.001$]. A pairwise comparison test showed significantly better performance by KNT than by SS ($p<0.001$) and MMSE ($p<0.001$). However, there was no significant difference between the results of KNT and WF ($p=0.676$). Significant difference was also observed in PESQ score [$F(3,24)=15.393$, $p<0.001$]. In the pairwise comparison test, the PESQ score of WF was significantly lower than that of KNT ($p=0.001$), MMSE ($p<0.001$) and SS ($p=0.03$). The same results are shown in Table 2.1 and Table 2.2.

Chapter III: Speech Enhancement Algorithm with Binary Masking Method

3.1 Motivation and theory

The unnatural speech perception generated by the discrete binary gain applied to a speech signal in the BM method is considered to be a disadvantage of this technique. However, CI users may not be affected from such distortions triggered by the BM method. Acoustic signals are converted into electrical pulses that are used to stimulate auditory nerves for CI users, so the speech signal they receive may be far from natural sound. Furthermore, it has been reported that speech intelligibility can be enhanced with the IBM method for both normal-hearing listeners and hearing-impaired listeners (e.g. Li and Loizou, 2008; Roman and Woodruff, 2011). Therefore, the BM method could be a potentially effective way to improve speech intelligibility in noise for CI users, with an acceptable loss in sound quality.

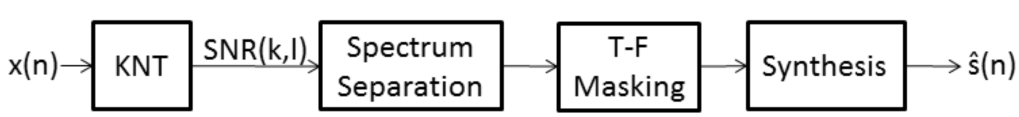

Fig. 3.1 A schemitic overview of KNT-BM method

The BM method takes values of 0 and 1, and compares the individual SNR of each T–F unit against a threshold. Ideally, if the information of the speech and noise is accessible before mixing, the perfect local SNR could be calculated and used to determine the binary gain of each T-F unit, which is IBM (Wang *et al.*, 2008; 2009). However, in the real application, perfect SNR is normally not available due to the lack of information of the input signal. A method to estimate the local SNR values is then considered a key component. In this chapter, a single-channel speech enhancement algorithm based on the KNT and BM method is implemented. Fig. 3.1 gives an overview of four phases of the KNT-BM method. The SNR is estimated by the KNT algorithm as described below.

3.2 Implementation

The first part of the KNT-BM method is to estimate local SNR of each frequency bin. Then a spectral analysis phase is performed to map the input signal into the time-frequency domain, dividing the input signal into T-F units. This is followed by a classification step, during which the target speech regions of the input signal are distinguished from competing noise regions. Finally, the algorithm removes the noise regions and retains the target speech regions.

$$\mathrm{SNR}(k,l) = \left(\frac{X(k,l)}{\tilde{N}(k,l)} - 1\right)^2 \tag{3.1}$$

The SNR of each frequency bin is computed as in (3.1), which is equivalent to $(\mathrm{W}(\mathrm{k},\mathrm{l})-1)^2$ as described in (2.14). Next, the bin array of that temporal frame is divided into 16 equally spaced channels (Princen and Bradley, 1986). Each T-F unit is defined as a group of bins in one temporal window within an individual spectral channel. The number of bins used in our case is 17 per unit. As in (3.2), the gain of the k^{th} bin, $\mathrm{G}(\mathrm{k},\mathrm{l})$, is set to 1, if $\mathrm{SNR}(\mathrm{k},\mathrm{l})$ is larger than the determined threshold, T; otherwise, it is set to 0. After the spectral analysis, all the isolated survived bins are eliminated to reduce the effects of estimation error. For example, if $\mathrm{G}(\mathrm{k},\mathrm{l})$ is 1, but $\mathrm{G}(\mathrm{k}-1,\mathrm{l})$ and $\mathrm{G}(\mathrm{k}+1,\mathrm{l})$ are both 0, then $\mathrm{G}(\mathrm{k},\mathrm{l})$ is set to 0. Finally, if this T-F unit still contains survived bins, the gains of all the bins within this T-F unit will be set to 1, to retain the speech information and avoid potential distortion. The cleaned speech is then synthesized based on the masking results as in (3.3). The sampling rate and window length are the same as used in Chapter 2.

$$\mathrm{G}(\mathrm{k},\mathrm{l})=\begin{cases}0, & \mathrm{SNR}(\mathrm{k},\mathrm{l})<\mathrm{T}\\ 1, & \mathrm{SNR}(\mathrm{k},\mathrm{l})\geq \mathrm{T}\end{cases} \tag{3.2}$$

$$\hat{\mathrm{S}}(\mathrm{k},\mathrm{l})=\mathrm{X}(\mathrm{k},\mathrm{l})\mathrm{G}(\mathrm{k},\mathrm{l}) \tag{3.3}$$

It should be noted that one common spectral anylysis method of BM algorithms is based on a bank of gammatone filters (e.g. Li and Loizou, 2008; Wang *et al.*, 2008), as it is assumed to be similar to what human cochlea does when processing sounds (Moore *et al.*, 1990). However, in this method, the signal spectrum is divided into a bank of equally spaced channels, because the local SNRs are estimated from the KNT algorithm, in which each freuency bin is weighted equally.

3.3 Illustration

The effects of the KNT-BM method by removing the SSN from a sample sentence, "Mend the coat before you go out," spoken by a male speaker are shown in Figure 3.2. Spectrograms of original signal, noisy signal and cleaned signal are shown from top to bottom panels, respectively. The input SNR is 10 dB. The bottom panel shows most of the noise has been removed by the KNT-BM algorithm. However, since KNT may not always be able to distinguish soft speech from noise, some distortion may occur to the soft speech components, which may undermine the overall speech quality. To further investigate the effects on speech intelligibility, a speech perception experiment was conducted for CI users in next Chapter.

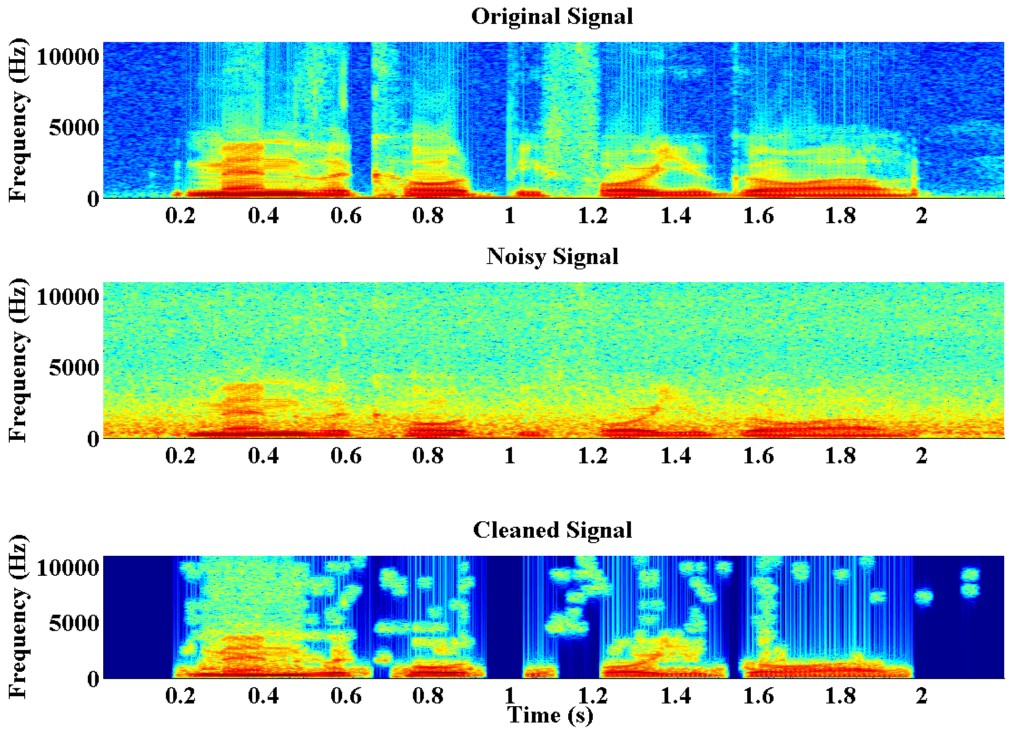

Fig. 3.2 Spectrum of original signal, noisy signal and cleaned signal by KNT-BM (from top to bottom respectively) at an input SNR=10 dB.

Chapter IV: Speech Perception Tests for Cochlear-Implant Users

4.1 Method

4.1.1 Subjects

Twelve post-lingually deafened CI users participated in this study and were compensated for their time. Information regarding the individual CI users is provided in Table 4.1.

4.1.2 Materials and procedures

The test stimuli were IEEE sentences corrupted by babble noise (20 talkers), SSN or restaurant noise. Three input SNR conditions (0, +5 and +10 dB) were tested; resulting in nine test conditions altogether. Each condition contained two test blocks (10 sentences/block); one with a male talker and one with a female talker. In addition to the 18 blocks with the KNT-BM processing, 18 reference (unprocessed) blocks were also tested. The test order of the 36 test blocks was randomized for each participant. A Matlab-based testing program was designed (see the picture of interface in Fig. 4.1). Each sentence was played only once and subjects were instructed to type all the words they heard.

Table 4.1 CI patients information

Subject code	Gender	Age (Yrs)	CI use (Yrs)	Etiology	Duration HL prior to implant (Yrs)	Speech processing strategy
D02	F	65.5	13.7	Unknown	1	HiRes-P with Fidelity120
D10	F	61.2	12.7	Unknown	8	HiRes-S with Fidelity120
D24	M	65.1	7.6	Progressive	27	HiRes-S with Fidelity120
D27	F	63.6	6.1	Otoscerlosis	13	HiRes-S with Fidelity120
D28	F	66.4	12.4	Familial Progressive SNHL	7	HiRes-S with Fidelity120
D35	F	55.6	2.6	High Fever	Unknown	HiRes-S with Fidelity120
D39	M	68.2	6.6	High Fever	Unknown	HiRes Optima-S
N13	M	77.4	25.0	Hereditary	4	SPEAK
N14	M	71.4	21.8	Progressive SNHL	1	SPEAK
N32	M	47.8	18.0	Maternal Rubella	<1	SPEAK
P10	F	18.5	13.7	Congenital Auditory neuropathy	5.5	ACE
P13	F	26.0	3.0	Sudden SNHL	1	ACE

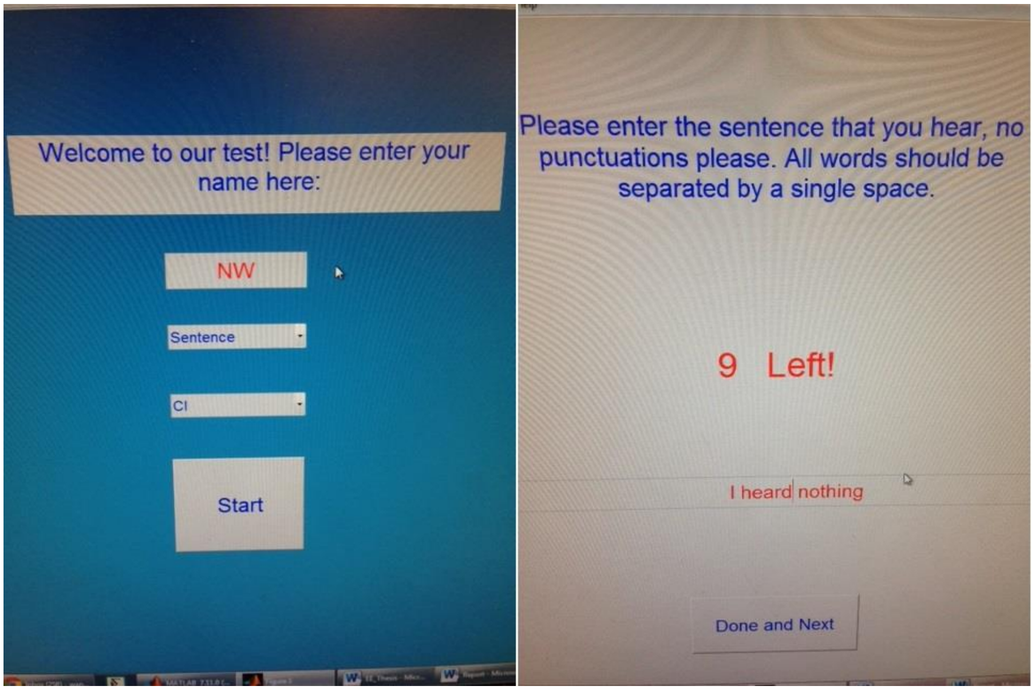

Fig. 4.1 Interface of the test program

The stimuli were generated digitally and played out from a LynxStudio L22 24-bit soundcard at a sampling rate of 22.5 kHz via a loudspeaker (subject 1-meter at 0°azimuth) to the subjects seated in a double-walled sound-attenuating chamber. The sentences were presented at 70 dB SPL. All the subjects used their daily processors and coding strategies during the experiment.

4.2 Results

4.2.1 Babble noise

For the babble noise condition, individual results are shown in Fig. 4.2. A two-way (noise reduction process × input SNR level) within-subjects ANOVA was run. Significant main effects were observed in both the noise reduction process [$F(1,11)=20.31$, $P=0.001$] and input SNR level [$F(2,22) = 27.804$, $p<0.001$], indicating the KNT-BM algorithm could improve speech intelligibility of CI users in the presence of babble noise at various SNR conditions. A significant interaction between them [$F(2,22) = 3.984$, $p = 0.033$] suggests the effects of KNT-BM varies with the input SNR level. In general, it works better in the more favorable SNR conditions (5 and 10 dB) as indicated from the results; positive improvements were observed for most subjects. However, in 0-dB condition, only two subjects showed considerable improvement.

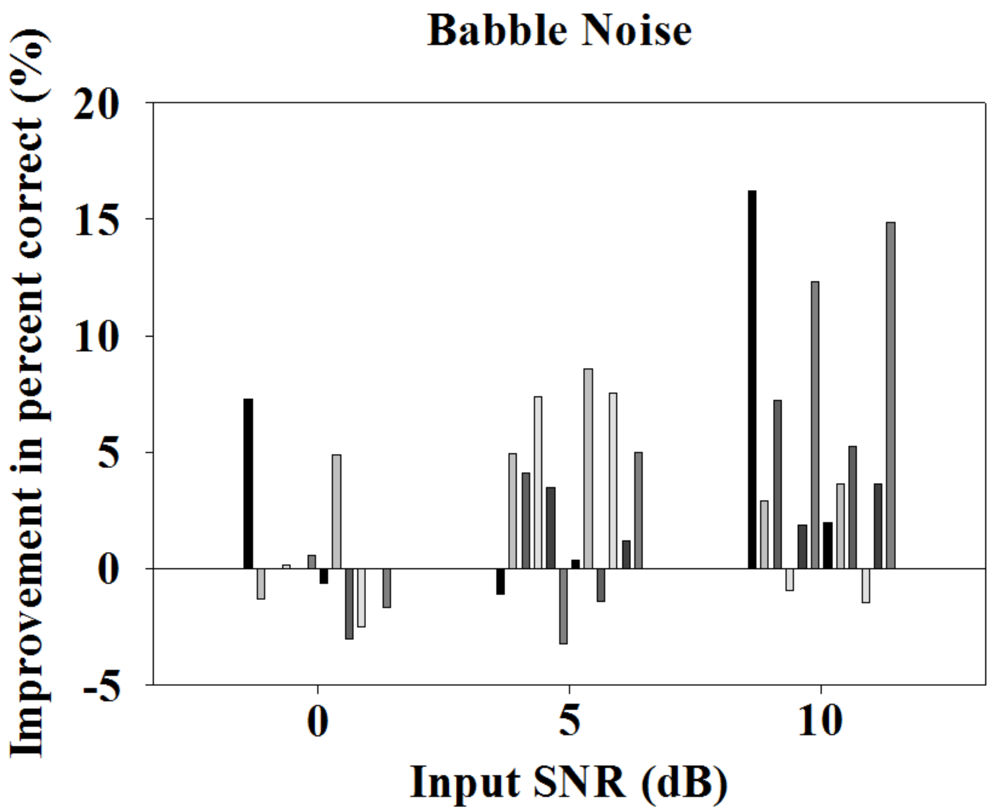

Fig. 4.2 Individual results from babble noise conditions. Y-axis represents the improvement in percent correct with the KNT-BM method compared to the unprocessed reference conditions. X-axis indicates the input SNR level. Each bar stands for an individual subject.

4.2.2 Speech-shaped noise

The individual results for the SSN are shown in Fig. 4.3. The same statistical analysis as used in the babble noise condition was applied to the results from the SSN conditions. A two-way within-subjects ANOVA showed significant main effects for both the noise reduction process [F(1,11) = 6.537, p = 0.027] and input SNR level [F(2,22) = 24.323, p < 0.001]; indicating this algorithm could improve speech intelligibility of CI users when listening in SSN. No significant interaction was found between the two factors [F(2,22) = 2.871, p = 0.078], which is consistent with the observation that there are positive effects for most subjects in all three SNR conditions (see Fig. 4.3).

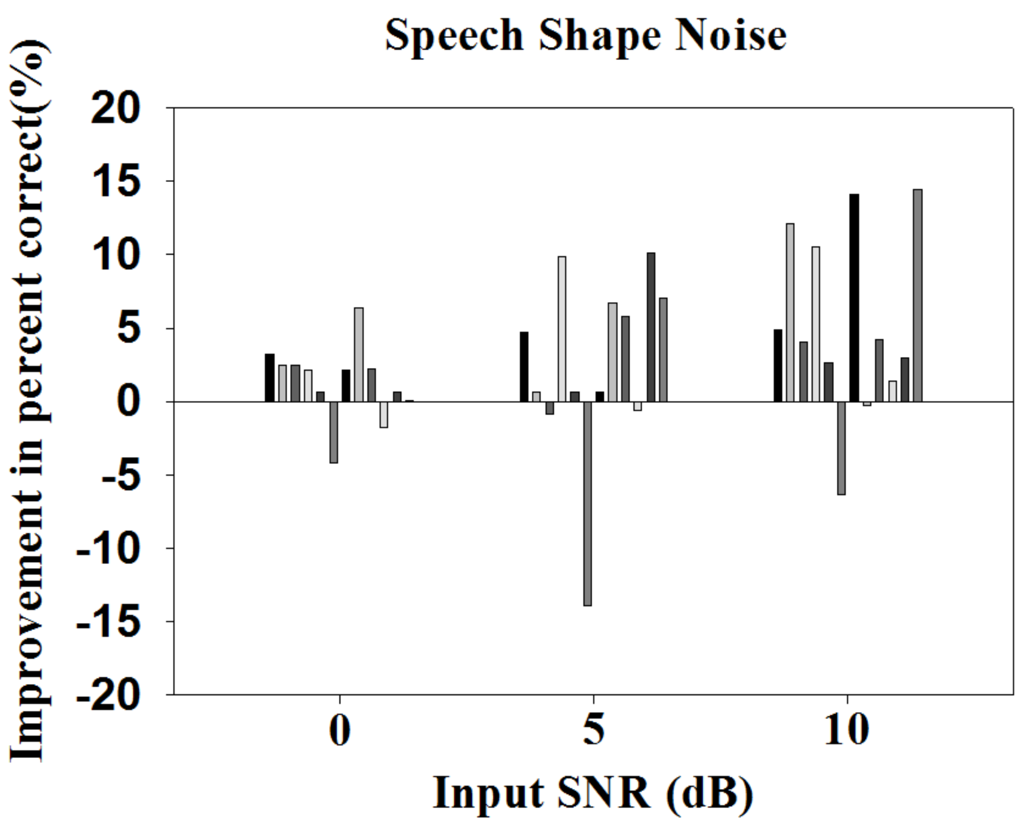

Fig. 4.3 Individual results from SSN conditions. Y-axis represents the improvement in percent correct with the KNT-BM method compared to the unprocessed reference

conditions. X-axis indicates the input SNR level. Each bar stands for an individual subject.

4.2.3 Restaurant noise

The individual results for the restaurant noise are shown in Fig. 4.4. Unlike the previous two noise conditions, no significant main effect was observed for the noise reduction process [$F(1,11) = 2.837$, $p = 0.12$], indicating the KNT-BM algorithm does not improve speech intelligibility of CI users when listening in restaurant noise. A significant main effect was still observed for the input SNR level [$F(2,22) = 27.373$, $p < 0.001$], as well as the interaction between the two factors [$F(1,11) = 9.698$, $p = 0.001$]. It seems most subjects were able to benefit from the noise reduction process in 0-dB and 5-dB SNR conditions; unfortunately, a strong negative effect occurred in the 10-dB SNR condition, which means the KNT algorithm may not be able to distinguish restaurant noise, which is the most non-stationary among the three types of noise, from the speech signal.

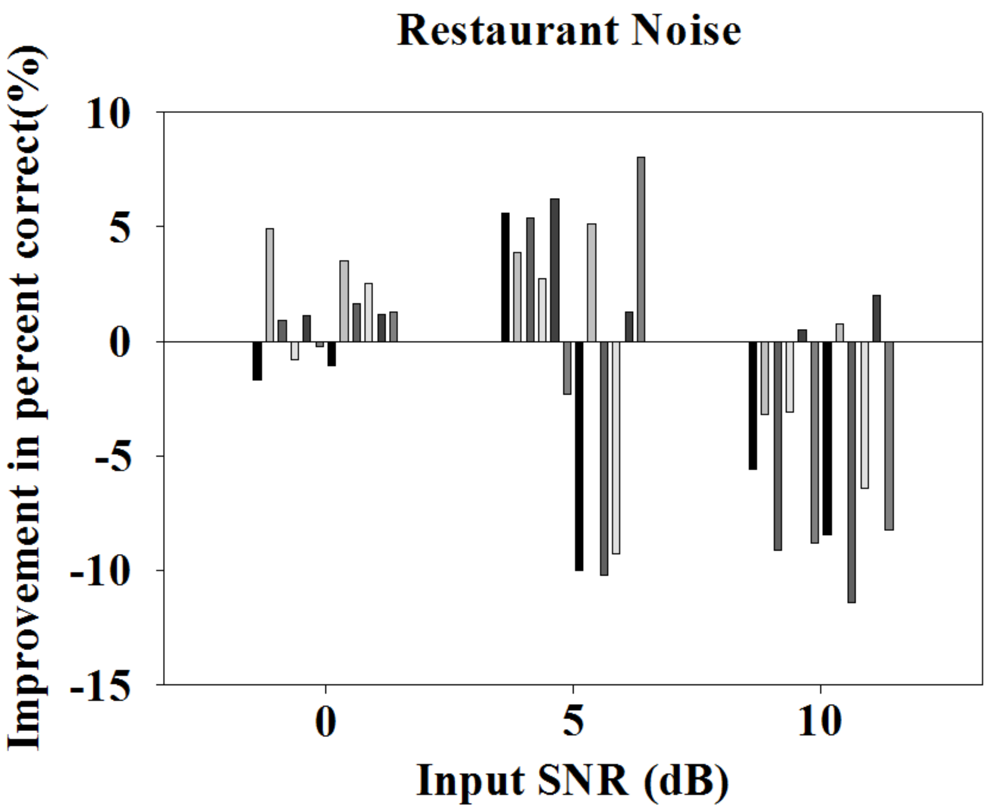

Fig. 4.4 Individual results from restaurant noise condition. Y-axis represents the improvement in percent correct with the KNT-BM method compared to the unprocessed reference conditions. X-axis indicates the input SNR level. Each bar stands for an individual subject.

4.2.4 Average results

The average results are shown in Fig. 4.5. The CI users showed improvement in all conditions, except the 10-dB restaurant noise condition. One explanation to consider is that it is almost impossible to track restaurant noise, because of its non-stationary features, with the KNT algorithm. When the input SNR is low, some noise is removed, although much is not. Nevertheless, the speech intelligibility is enhanced or retained, indicating the masking effects produced by the noise contribute more than the speech distortion created by the noise reduction process to the deficits in speech intelligibility in low SNR conditions. However, when the input SNR is high, e.g. 10 dB, some of the speech-dominated regions may be eliminated instead, which results in a degradation in speech intelligibility of CI users. In all other conditions, higher input SNRs corresponds to better perception.

Table. 4.2 Average results from all three noise conditions.

Noise	Input SNR	Unprocessed (%)	Processed (%)
Babble	0 dB	4.03	4.35
	5 dB	12.06	15.13
	10 dB	22.06	27.68
Speech-shaped	0 dB	7.94	9.31
	5 dB	18.58	21.14
	10 dB	30.03	35.42
Restaurant	0 dB	8.21	9.32
	5 dB	14.29	14.81
	10 dB	31.16	26.07

Another interesting finding is that the performance of subjects in the 0-dB SNR condition with restaurant noise is better than that measured in 0-dB SNR condition with babble noise or SSN. Masking release produced by noise gap may explain this mismatch with our expectation (Oxenham and Simonson, 2009; Leger *et al.*, 2012), that restaurant noise is the most harmful to speech intelligibility in all SNR conditions, including the 0-dB SNR condition. The detailed average results have been exhibited in Table 4.2.

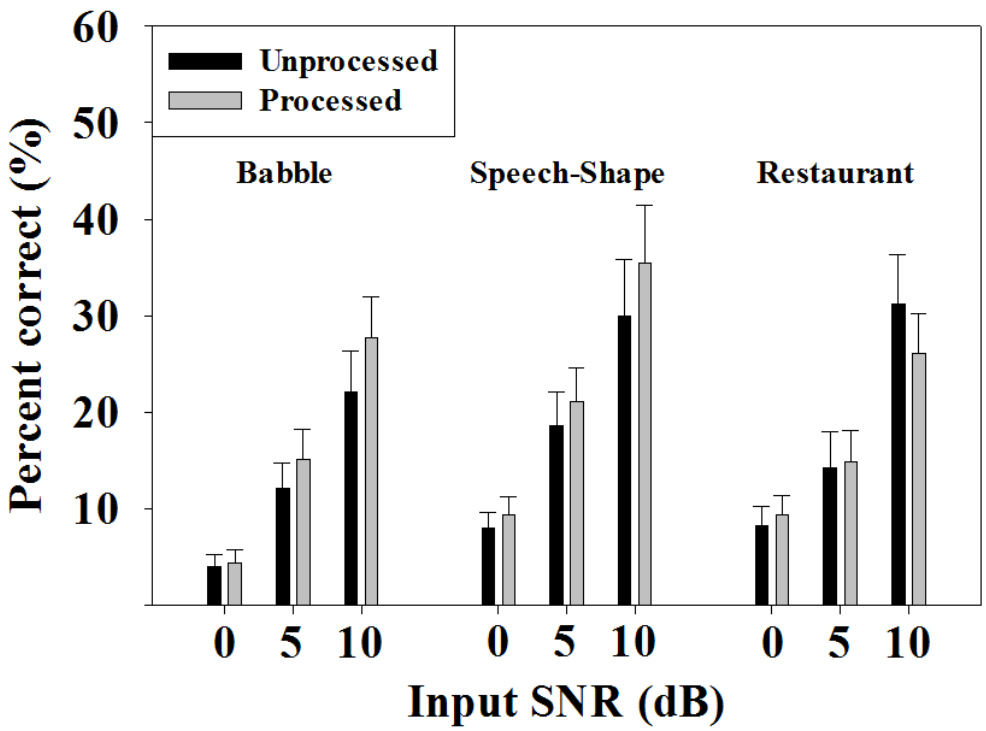

Fig. 4.5 Average results from all three noise conditions. Y-axis represents the mean percent correct in speech perception tests from all CI users. X-axis indicates the input SNR level. The black bars and gray bars stand for unprocessed and processed conditions, respectively.

4.3 Future improvements

In the KNT-BM algorithm implemented above, the threshold, T, is fixed through the process. Results from both objective evaluation and speech perception test indicate better performance in high-SNR conditions relative to low-SNR conditions. Poor accuracy in noise estimation in such low-SNR conditions is considered to be the reason. A crude estimation in noise level may arouse severe speech distortion, which may explain why there is little improvement in performance of CI users in the speech perception test when the input SNR is low.

$$T(k,l+1) = (1-\tau)T(k,l) + \tau SNR(k,l) \tag{4.1}$$

One way to overcome this issue is to apply SNR-sensitive threshold instead the fixed, as in (4.1). In each frame, the threshold is updated by the SNR value estimated from the prior frame. The advantage is obvious. When the estimated SNR is low, the threshold decreases, so that more T-F units will be retained to limit speech distortion. When the estimated SNR is high, the algorithm becomes more aggressive with the higher threshold, resulting in the cleaner and more intelligible speech signal.

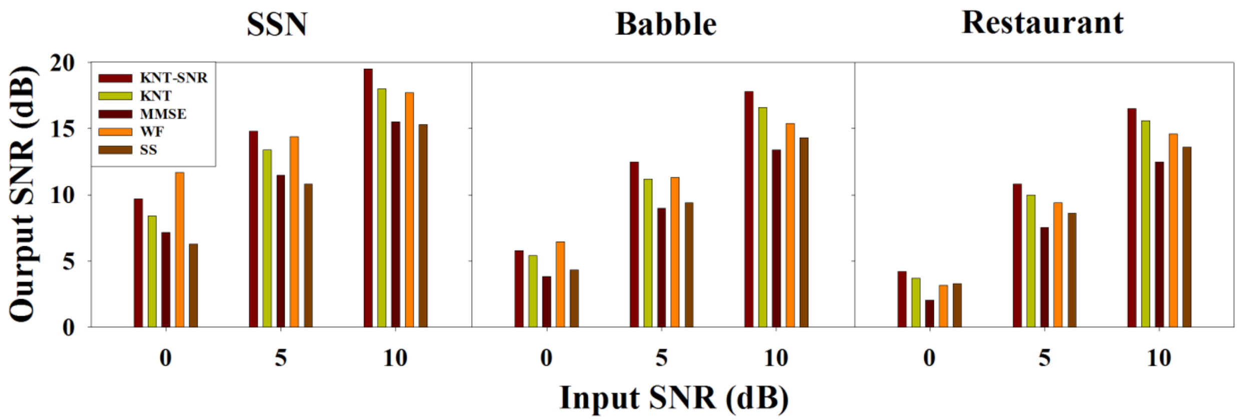

Fig. 4.6 Output SNR in different noise conditions as a function of input SNR. From left to right, each bar represents the result from one algorithm (KNT-SNR, KNT, MMSE, WF, SS), and each panel represents a noise condition (SSN, babble noise and restaurant noise).

To evaluate the performance of SNR-sensitive KNT algorithm (KNT-SNR) and investigate its potential ability, a simple output SNR comparison test is made among it and other four noise estimation algorithms. The results are shown in Fig. 4.6. The output SNRs of KNT-SNR are consistently higher than that of KNT throughout all noise conditions, which is further confirmed by a paired-sample t-test [$t(8)=8.23$, $p<0.001$]. Based on this finding, the future work should be focused on developing SNR and environment (noise type) sensitive algorithms to further improve the performance of such noise reduction algorithms.

Chapter V: Conclusions

Despite good performance in quiet environment, there are still significant gaps in auditory perception between normal-hearing listeners and CI users. For example, the performance of CI users in a speech perception task with additive noise is poorer than normal-hearing listeners. Very few of the current noise reduction algorithms are able to enhance speech intelligibility without prior knowledge of the noisy speech. In this study, a single-channel noise reduction algorithm, based on a noise tracking algorithm and the BM method, is implemented based on Matlab and tested in CI users.

In chapter II, a Kalman filtering process is used to track the detailed fluctuations of noise. This KNT algorithm is able to catch detailed spectral information of the noise with a fast noise tracker when no speech signal is present. When speech is introduced, the overall estimation of the noise level is updated by a slow noise tracker instead. Since the slow noise tracker is able to update the general level of the noise in real time, no independent VAD is required in KNT. In objective measurements, the output SNRs of the cleaned speech are significantly higher after being processed by KNT compared to the conventional SS and MMSE algorithms. In addition, its PESQ scores are significantly higher than that of WF in all noise conditions. The results suggest better performance can be achieved with the KNT rather than with the traditional algorithms, generally speaking.

In chapter III, a single-channel speech enhancement algorithm based on the KNT and BM method is implemented. Since acoustic signals are converted into electrical pulses delivered to auditory nerves by CIs, CI users may not suffer from speech distortions triggered by the BM method as much as normal-hearing listeners, because they are used to listening to unnatural speech all the time. Furthermore, it has been reported speech intelligibility can be enhanced with the IBM method for both normal-hearing listeners and hearing-impaired listeners (e.g. Li and Loizou, 2008; Roman and Woodruff, 2011). Therefore, the KNT-BM method could be a potentially effective way to improve speech intelligibility in noise for CI users with an acceptable loss in sound quality.

The first part of the KNT-BM method is to estimate the local SNR of each frequency bin. Then a spectral analysis phase is performed to map the input signal into the time-frequency domain, dividing the input signal into T-F units. This is followed by a classification step, during which the target speech regions of the input signal are distinguished from competing noise regions. Finally, the algorithm removes the noise regions and retains the target speech regions. Figure 3.2 showed an example where the majority of noise was removed by the KNT-BM algorithm in the 10-dB SNR SSN condition.

In Chapter IV, a speech perception test was employed to investigate the effects of KNT-BM method on improving speech intelligibility of CI users in various types of background noise. The mean percent correct of CI users was improved in all conditions with noise reduction process, but the 10-dB restaurant noise condition. One explanation is considered as it is almost impossible to track restaurant noise with KNT algorithm, because of its non-stationary features. When the input SNR is low, the speech intelligibility is enhanced or retained, indicating the masking effects produced by noise contribute more than speech distortion created by noise reduction process to the deficits in speech intelligibility in low-SNR conditions. However, when the input SNR is high, e.g. 10 dB, some of the speech-dominated regions may be eliminated instead, which results in a degradation in speech intelligibility of CI users. In all the other conditions, higher input SNR means higher perception correct.

In summary, this KNT-BM algorithm improves CI subjects' speech perception in background noise. It provides more benefit in babble noise or SSN conditions, as improvements in speech intelligibility are observed in all input SNR conditions, although larger improvements are found in higher input SNR conditions. However, challenges still remain for restaurant noise, which is the most non-stationary noise tested. Future work is required to restore speech intelligibility for hearing-impaired listeners when perceiving speech in various types of noise with single-channel noise reduction algorithms, in which SNR-sensitive threshold should be applied. Although multi-channel methods, such as beamforming (Peterson *et al.*, 1987; Greenberg and Zurek, 1992; van Hoesel and Clark, 1995) or multi-channel wiener filter (Doclo and Moonen, 2005; Doclo *et al.*, 2007; Van den Bogaert *et al.*, 2007; Van den Bogaert *et al.*, 2009b) have been well studied, there is still considerable room for single-channel algorithms to be improved with the help from other techniques, e.g. deep learning or cognition-based models.

References

Anzalone, M. C., Calandruccio, L., Doherty, K. A., and Carney, L. H. (**2006**). "Determination of the potential benefit of time-frequency gain manipulation," Ear Hear. **27**, 480-492.

Baer, T., and Moore, B. C. (**1994**). "Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech," J. Acoust. Soc. Am. **95**, 2277-2280.

Baer, T., Moore, B. C., and Kluk, K. (**2002**). "Effects of low pass filtering on the intelligibility of speech in noise for people with and without dead regions at high frequencies," J. Acoust. Soc. Am. **112**, 1133-1144.

Beerends, J. G., Hekstra, A. R., Rix, A. W., and Hollier, M. P. (**2002**). "Perceptual evaluation of speech quality (PESQ) - The new ITU standard for end-to-end speech quality assessment - Part II - Psychoacoust model," J. Audio Eng. Soc. **50**, 765-778.

Beh, J., and Ko, H. (**2003**). "SS using spectral harmonics for robust speech recognition in car environments," Lect. Notes. Comput. Sc. **2660**, 1109-1116.

Boll, S. F. (**1979**). "Suppression of Acoustic Noise in Speech Using SS," IEEE T. Acoust. Speech **27**, 113-120.

Brons, I., Houben, R., and Dreschler, W. A. (**2014**). "Effects of noise reduction on speech intelligibility, perceived listening effort, and personal preference in hearing-impaired listeners," Trends Hear. **18**.

Cao, S., Li, L., and Wu, X. (**2011**). "Improvement of intelligibility of ideal binary-masked noisy speech by adding background noise," J. Acoust. Soc. Am. **129**, 2227-2236.

Chen, J., Baer, T., and Moore, B. C. (**2012**). "Effect of enhancement of spectral changes on speech intelligibility and clarity preferences for the hearing impaired," J. Acoust. Soc. Am. **131**, 2987-2998.

Chen, J. D., Benesty, J., Huang, Y., and Doclo, S. (**2006**). "New insights into the noise reduction Wiener filter," IEEE T. Acoust. Speech **14**, 1218-1234.

Chung, K. (**2004**). "Challenges and recent developments in hearing aids. Part I. Speech understanding in noise, microphone technologies and noise reduction algorithms," Trends Amplif. **8**, 83-124.

Crozier, P. M., Cheetham, B. M. G., Holt, C., and Munday, E. (**1993**). "Speech Enhancement Employing SS and Linear Predictive Analysis," Electron Lett. **29**, 1094-1095.

Doclo, S., and Moonen, M. (**2005**). "On the output SNR of the speech-distortion weighted multichannel Wiener filter," IEEE Signal Proc. Lett. **12**, 809-811.

Doclo, S., Spriet, A., Wouters, J., and Moonen, M. (**2007**). "Frequency-domain criterion for the speech distortion weighted multichannel Wiener filter for robust noise reduction," Speech Commun. **49**, 636-656.

Ephraim, Y., and Malah, D. (**1984**). "Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator," IEEE T. Acoust. Speech **32**, 1109-1121.

Ephraim, Y., and Malah, D. (**1985**). "Speech Enhancement Using a Minimum Mean-Square Error Log-Spectral Amplitude Estimator," IEEE T. Acoust. Speech **33**, 443-445.

Fetterman, B. L., and Domico, E. H. (**2002**). "Speech recognition in background noise of cochlear implant patients," J. Assoc. Res. Otolaryngol. **126**, 257-263.

Firszt, J. B., Holden, L. K., Skinner, M. W., Tobey, E. A., Peterson, A., Gaggl, W., Runge-Samuelson, C. L., and Wackym, P. A. (**2004**). "Recognition of speech presented at soft to loud levels by adult cochlear implant recipients of three cochlear implant systems," Ear Hear. **25**, 375-387.

Goh, Z., Tan, K. C., and Tan, B. T. G. (**1998**). "Postprocessing method for suppressing musical noise generated by SS," IEEE T. Speech Audio P**6**, 287-292.

Greenberg, J. E., and Zurek, P. M. (**1992**). "Evaluation of an adaptive beamforming method for hearing aids," J. Acoust. Soc. Am. **91**, 1662-1676.

Gustafsson, H., Nordholm, S. E., and Claesson, I. (**2001**). "SS using reduced delay convolution and adaptive averaging," IEEE T. Speech Audio P**9**, 799-807.

Hamzavi, J., Franz, P., Baumgartner, W. D., and Gstoettner, W. (**2001**). "Hearing performance in noise of cochlear implant patients versus severely-profoundly hearing-impaired patients with hearing aids," Audiology **40**, 26-31.

Healy, E. W., Yoho, S. E., Wang, Y., and Wang, D. (**2013**). "An algorithm to improve speech recognition in noise for hearing-impaired listeners," J. Acoust. Soc. Am. **134**, 3029-3038.

Hu, Y., and Loizou, P. C. (**2007**). "A comparative intelligibility study of single-microphone noise reduction algorithms," J. Acoust. Soc. Am. **122**, 1777.

Kim, G., and Loizou, P. C. (**2011**). "Gain-induced speech distortions and the absence of intelligibility benefit with existing noise-reduction algorithms," J. Acoust. Soc. Am. **130**, 1581-1596.

Kim, G., Lu, Y., Hu, Y., and Loizou, P. C. (**2009**). "An algorithm that improves speech intelligibility in noise for normal-hearing listeners," J. Acoust. Soc. Am. **126**, 1486-1494.

Kokkinakis, K., and Loizou, P. C. (**2008**). "Using blind source separation techniques to improve speech recognition in bilateral cochlear implant patients," J. Acoust. Soc. Am. **123**, 2379-2390.

Leger, A. C., Moore, B. C., and Lorenzi, C. (**2012**). "Temporal and spectral masking release in low- and mid-frequency regions for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **131**, 1502-1514.

Li, N., and Loizou, P. C. (**2008**). "Factors influencing intelligibility of ideal binary-masked speech: implications for noise reduction," J. Acoust. Soc. Am. **123**, 1673-1682.

Lim, J. S., and Oppenheim, A. V. (**1979**). "Enhancement and Bandwidth Compression of Noisy Speech," P IEEE **67**, 1586-1604.

Lockwood, P., and Boudy, J. (**1992**). "Experiments with a Nonlinear Spectral Subtractor (Nss), Hidden Markov-Models and the Projection, for Robust Speech Recognition in Cars," Speech Commun. **11**, 215-228.

Loizou, P. C., Poroy, O., and Dorman, M. (**2000**). "The effect of parametric variations of cochlear implant processors on speech understanding," J. Acoust. Soc. Am. **108**, 790-802.

Martin, R. (**1994**). "SS based on minimum statistics," Proc. Eur. Signal Process., 1182-1185.

McDermott, H. J. (**2004**). "Music perception with cochlear implants: a review," Trends Amplif. **8**, 49-82.

Moore, B. C., Peters, R. W., and Glasberg, B. R. (**1990**). "Auditory filter shapes at low center frequencies," J. Acoust. Soc. Am. **88**, 132-140.

Moore, B. C. J., Laurence, R. F., and Wright, D. (**1985**). "Improvements in speech intelligibility in quiet and in noise produced by two-channel compression hearing aids," Brit. J. Audiol. **19**, 175-187.

Nilsson, M., Soli, S., and Sullivan, J. A. (**1994**). "Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am. **95**, 1085-1099.

Oxenham, A. J., and Simonson, A. M. (**2009**). "Masking release for low- and high-pass filtered speech in the presence of noise and single-talker interference," J. Acoust. Soc. Am. **125**, 457-468.

Peterson, P. M., Durlach, N. I., Rabinowitz, W. M., and Zurek, P. M. (**1987**). "Multimicrophone adaptive beamforming for interference reduction in hearing aids," J. Rehabil. Res. Dev. **24**, 103-110.

Plapous, C., Marro, C., and Scalart, P. (**2006**). "Improved signal-to-noise ratio estimation for speech enhancement," IEEE T. Audio Speech **14**, 2098-2108.

Princen, J. P., and Bradley, A. B. (**1986**). "Analysis Synthesis Filter Bank Design Based on Time Domain Aliasing Cancellation," IEEE T. Acoust Speech **34**, 1153-1161.

Rix, A. W., Hollier, M. P., Hekstra, A. P., and Beerends, J. G. (**2002**). "Perceptual evaluation of speech quality (PESQ) - The new ITU standard for end-to-end speech quality assessment - Part I - Time-delay compensation," J. Audio Eng. Soc. **50**, 755-764.

Roman, N., and Woodruff, J. (**2011**). "Intelligibility of reverberant noisy speech with ideal binary masking," J. Acoust. Soc. Am. **130**, 2153-2161.

Seok, J. W., and Bae, K. S. (**1999**). "Reduction of musical noise in SS method using subframe phase randomisation," Electron Lett. **35**, 123-125.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303-304.

Skinner, M. W., Holden, L. K., Whitford, L. A., Plant, K. L., Psarros, C., and Holden, T. A. (**2002**). "Speech recognition with the nucleus 24 SPEAK, ACE, and CIS speech coding strategies in newly implanted adults," Ear Hear. **23**, 207-223.

Spahr, A. J., and Dorman, M. F. (**2004**). "Performance of subjects fit with the Advanced Bionics CII and Nucleus 3G cochlear implant devices," Arch. Otolaryngol. Head Neck Surg. **130**, 624-628.

Spriet, A., Moonen, M., and Wouters, J. (**2004**). "Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction," Signal Process. **84**, 2367-2387.

Summers, I. R., and Al-Dabbagh, A. D. (**1982**). "Simulated loss of frequency selectivity and its effects on speech perception," Acoust. Lett. **5**, 129-132.

Van den Bogaert, T., Doclo, S., Wouters, J., and Moonen, M. (**2009a**). "Speech enhancement with multichannel Wiener filter techniques in multimicrophone binaural hearing aids," J. Acoust. Soc. Am. **125**, 360-371.

Van den Bogaert, T., Doclo, S., Wouters, J., and Moonen, M. (**2009b**). "Speech enhancement with multichannel Wiener filter techniques in multimicrophone binaural hearing aids," J. Acoust. Soc. Am. **125**, 360-371.

Van den Bogaert, T., Wouters, J., Doclo, S., and Moonen, M. (**2007**). "Binaural cue preservation for hearing aids using an interaural transfer function multichannel Wiener filter," Int. Conf. Acoust. Speech, 565-568.

Van Dun, B., Wouters, J., and Moonen, M. (**2007**). "Multi-channel wiener filtering based auditory steady-state response detection," Int. Conf. Acoust. Speech, 929-932.

van Hoesel, R. J., and Clark, G. M. (**1995**). "Evaluation of a portable two-microphone adaptive beamforming speech processor with cochlear implant patients," J. Acoust. Soc. Am. **97**, 2498-2503.

van Tasell, D. J. (**1993**). "Hearing loss, Speech, and Hearing Aids," J. Speech Hear. Res. **36**, 228-244.

Vary, P. (**1985**). "Noise Suppression by Spectral Magnitude Estimation - Mechanism and Theoretical Limits," Signal Process. **8**, 387-400.

Wang, D., Kjems, U., Pedersen, M. S., Boldt, J. B., and Lunner, T. (**2008**). "Speech perception of noise with binary gains," J. Acoust. Soc. Am. **124**, 2303-2307.

Wang, D., Kjems, U., Pedersen, M. S., Boldt, J. B., and Lunner, T. (**2009**). "Speech intelligibility in background noise with ideal binary time-frequency masking," J. Acoust. Soc. Am. **125**, 2336-2347.

Wouters, J., and Vanden Berghe, J. (**2001**). "Speech recognition in noise for cochlear implantees with a two-microphone monaural adaptive noise reduction system," Ear Hear. **22**, 420-430.

Zeng, F. G. (**2004**). "Trends in cochlear implants," Trends Amplif. **8**, 1-34.

Zeng, F. G., Nie, K., Stickney, G. S., Kong, Y. Y., Vongphoe, M., Bhargave, A., Wei, C., and Cao, K. (**2005**). "Speech recognition with amplitude and frequency modulations," Proc. Natl. Acad. Sci. U.S.A. **102**, 2293-2298.