

RESEARCH

Open Access

Real-time spectrum estimation–based dual-channel speech-enhancement algorithm for cochlear implant

Yousheng Chen and Qin Gong*

* Correspondence: gongqin@mail.
tsinghua.edu.cn
Department of Biomedical
Engineering, Tsinghua University,
Beijing 100084, PR China

Abstract

Background: Improvement of the cochlear implant (CI) front-end signal acquisition is needed to increase speech recognition in noisy environments. To suppress the directional noise, we introduce a speech-enhancement algorithm based on microphone array beamforming and spectral estimation. The experimental results indicate that this method is robust to directional mobile noise and strongly enhances the desired speech, thereby improving the performance of CI devices in a noisy environment.

Methods: The spectrum estimation and the array beamforming methods were combined to suppress the ambient noise. The directivity coefficient was estimated in the noise-only intervals, and was updated to fit for the mobile noise.

Results: The proposed algorithm was realized in the CI speech strategy. For actual parameters, we use Maxflat filter to obtain fractional sampling points and cepstrum method to differentiate the desired speech frame and the noise frame. The broadband adjustment coefficients were added to compensate the energy loss in the low frequency band.

Discussions: The approximation of the directivity coefficient is tested and the errors are discussed. We also analyze the algorithm constraint for noise estimation and distortion in CI processing. The performance of the proposed algorithm is analyzed and further be compared with other prevalent methods.

Conclusions: The hardware platform was constructed for the experiments. The speech-enhancement results showed that our algorithm can suppresses the non-stationary noise with high SNR. Excellent performance of the proposed algorithm was obtained in the speech enhancement experiments and mobile testing. And signal distortion results indicate that this algorithm is robust with high SNR improvement and low speech distortion.

Background

The clinical cochlear implant (CI) has good speech recognition under quiet conditions, but noticeably poor recognition under noisy conditions [1]. For 50% sentence understanding [2,3], the required signal to noise ratio (SNR) is between 5 and 15 dB for CI recipients, but only -10 dB for normal listeners. The SNR in the typical daily environment is about 5–10 dB, which results in <50% sentence recognition for CI users in a normal noise environment.

Most previous studies on recognition improvement have focused on the coding strategy, design of the electrode array, and stimulation adjustment of pitch recognition, as well as on the virtual electrode technique [4,5] and optical CIs [6]. More recent efforts have focused on the microphone array technique [7,8]. This array beamforming method promises to be more effective for situations in which the desired voice and ambient noise originate from different directions, the usual work environment for CI devices.

Speech-enhancement methods include single- and multichannel techniques. Spectral estimation methods are the most widely used single-channel techniques. Typical single-channel approaches, such as the spectral subtraction [9,10], Wiener filtering [11], and subspace approach [12], are based on estimations of the power spectrum or higher-order spectrum, assume the noise to be stationary, and use the noise spectrum in the nonspeech frame to estimate the speech-frame noise spectrum. Algorithm performance sharply weakens when the noise is non-stationary, or under typical situations with music or ambient speech noise.

The microphone array technique considers the signal orientation information and focuses on directional speech enhancement. Specifically, the generalized sidelobe canceller [13] and delay beamforming [14,15] use multiple microphones to record signals for spatial filtering. For CI devices, the generalized sidelobe canceller is overly complicated and requires too many microphones, conditions that exceed the capabilities of current CI devices. Delay beamforming technologies, such as the first-order differential microphone (FDM) [16] and adaptive null-forming method (ANF) [17,18], are adopted in hearing aids. These methods need only 2 microphones, which is an appropriate set-up for the CI size constraint and real-time processing.

CI devices are similar with the hearing aids in size constraint and the requirement of front-end noise suppression. So, for CI speech enhancement, one simple solution for CI speech enhancement is to directly utilize the microphone-array-based noise-reduction methods from the present hearing aids, in which the sensor-array techniques have been more widely used. However, the difference between CI devices and hearing aids is prominent, and a direct application of these algorithms to CI speech processing is not appropriate. Firstly, the principle is very different. CI devices transfer the acoustic signal to electrical stimulation into the cochlea wirelessly, and then the electrical pulses are used to directly stimulate the acoustic nerve to yield the auditory perception. But the hearing aids only need to change the corresponding gains in different subbands for multi-frequency signal loss. In brief, the hearing aid is only an amplifier with adjustable gain in different frequency band. Secondly, the application of the microphone array technique is different. Many algorithms for speech application were borrowed from the narrowband methods in radar and antenna. Algorithms for front-end enhancement are indispensable to match the CI speech strategy. Thirdly, the solution for low frequency roll-off may be different. The hearing aids need to calibrate and preset the subband gain based on user's hearing loss. Therefore, in the hearing aid, one solution is to directly preset the subband gains in the filter banks in the processor by both taking the hearing loss and signal loss in microphone array algorithm into account. However, for CI devices with the modulated electrical pulse directly stimulate the cochlear nerves, we only need to adjust the algorithm loss. Finally, the signal distortion is different. When the enhanced signal is modulated by the CI speech strategy, the signal distortion will noticeably decreased (detailed analysis was given in the result section). Therefore,

an array for a cochlear implant is similar to a hearing aid in the speech-enhancement situation, but is different for the actual algorithm design, such as the tradeoff between speech distortion and noise suppression.

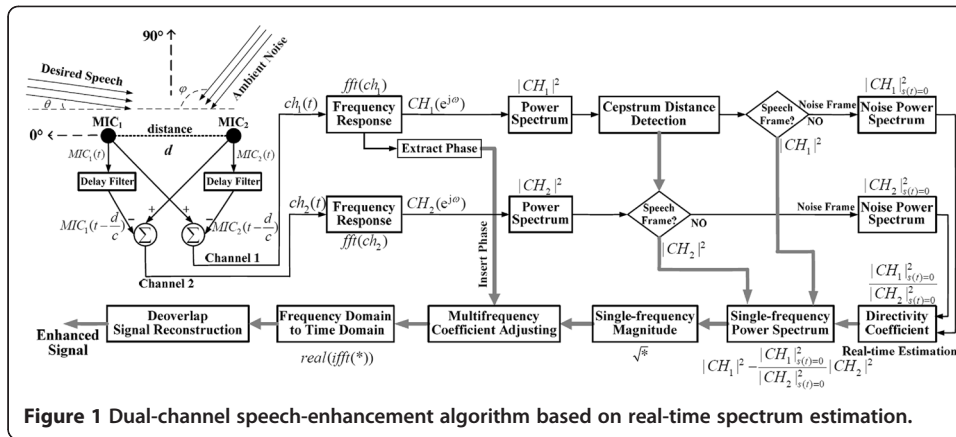
The Frost algorithm [19,20], multiple input/output inverse method (MINT) [21], minimum-variance distortionless-response technique (MVDR) [22,23] and the binaural frequency-domain minimum-variance algorithm [24] are proposed presently, with excellent performance in some specific situations. Kates [25] used a novel five-microphone end-fire array, with an MVDR included, to construct an adaptive frequency-domain noise-reduction algorithm with higher SNR improved. However, this algorithm is overly complicated, and the five-microphone array also exceeds the CI size constraint.

In daily environment, we previously proposed a low-complexity beamforming with optimal parameter to suppress the environmental stationary noise [26]. But for the music and speech noise, we need a higher SNR to weaken these ambient noises, aiming to obtain more than 10 dB SNR for the CI front-end signal acquisition. This paper focuses on directional noise suppression with one directional ambient interference for CI devices. In typical situations in which CI users want to talk with a nearby person in a conference hall or a theater, the directional voice from the lecturer or film screen must be suppressed. To weaken the directional noise in such situations, a dual-channel CI speech-enhancement algorithm was introduced that combines the single-channel power spectrum estimation and the first-order differential microphone technique of the microphone array, for beamforming and noise prediction. Our algorithm uses the dual-channel power spectrums in the noise-only intervals, including the nonstationary noise, to estimate and update the noise directivity coefficient. For noise changing in normal human walking velocity, the proposed algorithm can avoid the noise leakage and is robust to mobile noise. For spectrum estimation based speech enhancement, the speech distortion is also unavoidable in our algorithm. But when the signal is modulated in the CI speech strategy, the speech distortion will sharply decrease and the speech quality noticeably improves. The experimental results indicate that the proposed algorithm successfully achieves the desired speech reconstruction and enhancement.

Methods

For the actual daily usage of CI devices, the front speech is the desired signal that needs to be enhanced, as shown in Figure 1 (signal azimuth θ approaches 0°). The noise, including ambient music and other speech signals, originates from another direction (azimuth ϕ).

Figure 1 shows the flow chart of the proposed dual-channel speech-enhancement algorithm. Signals recorded by two omnidirectional microphones and two delaying signals by the delay filters are summed to yield dual-channel outputs. Firstly, the signal frequency response in each channel is extracted to obtain the power spectrum and the cepstrum distance. The cepstrum distance differentiates the desired-speech and noise-only power spectrums. Then the noise-only power spectrums are adopted to estimate the noise directivity coefficient. The narrowband signal magnitude is estimated by these power spectrums, including the desired-speech and noise-only segments, and the directivity coefficient. Furthermore, the narrowband signal magnitude, also named as single-frequency magnitude, is adjusted to yield the multifrequency magnitude for the desired speech (broadband signal with the compensation for low-frequency loss). And the



phase information in channel 1 is extracted and inserted for signal reconstruction to obtain the enhanced speech signal. Our proposed algorithm is theoretically depicted as below:

Two omnidirectional microphones are spaced a distance d apart. The desired speech comes from a direction θ ($\theta \rightarrow 0^\circ$) and the ambient noise from ϕ , both being recorded by MIC_1 and MIC_2 . If we denote the desired speech and ambient noise obtained by MIC_1 as $s(t)$ and $n(t)$, then the recorded signal by MIC_1 can be written as

$$MIC_1(t) = s(t) + n(t) \quad (1)$$

The spatial difference between the microphones results in a time delay for the recorded signal by MIC_2 , shown as

$$MIC_2(t) = s(t - d/c \cdot \cos\theta) + n(t - d/c \cdot \cos\phi) \quad (2)$$

where c is the speed of the sound. The recorded signals $MIC_1(t)$ and $MIC_2(t)$ are both delayed by the fractional delay filter for a fixed time d/c . Based on the FDM method, $ch_1(t)$ and $ch_2(t)$ can be defined as the sum of the original recorded signals and the corresponding delayed signals, described by Eqs. (3) and (4):

$$ch_1(t) = s(t) + n(t) - s\left(t - \frac{d}{c} - \frac{d}{c} \cos\theta\right) - n\left(t - \frac{d}{c} - \frac{d}{c} \cos\phi\right) \quad (3)$$

$$ch_2(t) = s\left(t - \frac{d}{c} \cos\theta\right) + n\left(t - \frac{d}{c} \cos\phi\right) - s\left(t - \frac{d}{c}\right) - n\left(t - \frac{d}{c}\right) \quad (4)$$

The frequency responses for time-domains $ch_1(t)$ and $ch_2(t)$ can be expressed as

$$CH_1(e^{j\omega}) = \left(1 - e^{-j\omega d/c \cdot (1 + \cos\theta)}\right)S(e^{j\omega}) + \left(1 - e^{-j\omega d/c \cdot (1 + \cos\phi)}\right)N(e^{j\omega}) \quad (5)$$

$$CH_2(e^{j\omega}) = \left(e^{-j\omega d/c \cdot \cos\theta} - e^{-j\omega d/c}\right)S(e^{j\omega}) + \left(e^{-j\omega d/c \cdot \cos\phi} - e^{-j\omega d/c}\right)N(e^{j\omega}) \quad (6)$$

where $\omega = 2\pi f$, f corresponds to a narrowband signal frequency. And the corresponding power spectrum for $ch_1(t)$ is:

$$\begin{aligned} |CH_1(e^{j\omega})|^2 = & 2(1 - \cos(\omega d/c \cdot (1 + \cos\theta))) |S(e^{j\omega})|^2 + A_1 A_2^* (S(e^{j\omega}) N^*(e^{j\omega})) \\ & + A_1^* A_2 (S^*(e^{j\omega}) N(e^{j\omega})) \\ & + 2(1 - \cos(\omega d/c \cdot (1 + \cos\phi))) |N(e^{j\omega})|^2 \end{aligned} \quad (7)$$

where $A_1 = 1 - e^{-j\omega_c^d(1 + \cos\theta)}$ and $A_2 = 1 - e^{-j\omega_c^d(1 + \cos\phi)}$.

In Eq. (7), each framed data set (of about 23 ms in duration) is used to calculate the corresponding statistical average of the power spectrum. For daily CI application, the desired signal $s(t)$ and ambient noise $n(t)$ are uncorrelated, given in Eq. (8).

$$E(S(e^{j\omega}) N^*(e^{j\omega})) = E(S^*(e^{j\omega}) N(e^{j\omega})) = 0 \quad (8)$$

Substituting Eq. (8) into Eq. (7) gives the statistical power spectrum set in channel 1:

$$\begin{aligned} E(|CH_1(e^{j\omega})|^2) = & 2(1 - \cos(\omega d/c(1 + \cos\theta))) E(|S(e^{j\omega})|^2) \\ & + 2(1 - \cos(\omega d/c(1 + \cos\phi))) E(|N(e^{j\omega})|^2) \end{aligned} \quad (9)$$

For CI devices, the desired speech generates from the front, and the signal direction θ approaches 0° . Thus, $\cos\theta \approx 1$, which simplifies Eq. (9) and gives

$$\begin{aligned} E(|CH_1(e^{j\omega})|^2) = & 2(1 - \cos(2\omega d/c)) E(|S(e^{j\omega})|^2) \\ & + 2(1 - \cos(\omega d/c \cdot (1 + \cos\phi))) E(|N(e^{j\omega})|^2) \end{aligned} \quad (10)$$

Similarly, the simplified statistical power spectrum for each framed data in channel 2 is written as:

$$E(|CH_2(e^{j\omega})|^2) = 2(1 - \cos(\omega d/c(1 - \cos\phi))) E(|N(e^{j\omega})|^2) \quad (11)$$

Seen from Eq. (10), the power spectrum for each framed data set in channel 1 includes the power spectra of desired speech and the ambient noise. Equation (11) only contains the power spectrum of the ambient noise in channel 2. In addition, the power spectra of the noise in channels 1 and 2 are different, which are functions of the noise azimuth.

To estimate the directivity of the ambient noise, the power spectra of the 2 channels are calculated at the noise frame of desired speech, where $s(t) = 0$ and $S(e^{j\omega}) = 0$:

$$E(|CH_1(e^{j\omega})|^2) \Big|_{s(t)=0} = 2(1 - \cos(\omega d/c \cdot (1 + \cos\phi))) E(|N(e^{j\omega})|^2) \Big|_{s(t)=0} \quad (12)$$

$$E(|CH_2(e^{j\omega})|^2) \Big|_{s(t)=0} = 2(1 - \cos(\omega d/c \cdot (1 - \cos\phi))) E(|N(e^{j\omega})|^2) \Big|_{s(t)=0} \quad (13)$$

For each framed data set, the statistical average of the power spectrum can be used to estimate the power spectrum for desired speech, and yield the magnitude estimation in Eq. (14).

$$|\hat{S}(e^{j\omega})| = \frac{\left| E(|CH_1(e^{j\omega})|^2) - \frac{E(|CH_1(e^{j\omega})|^2)|_{s(t)=0}}{E(|CH_2(e^{j\omega})|^2)|_{s(t)=0}} E(|CH_2(e^{j\omega})|^2) \right|^{0.5}}{2 \sin(2\pi fd/c)} \quad (14)$$

Eq. (14) indicates the algorithm for magnitude estimation must obtain the power spectrum of the framed data set of each channel, as well as the power spectrum of each channel at the noise-only frame.

$E(|CH_1(e^{j\omega})|^2)|_{s(t)=0}/E(|CH_2(e^{j\omega})|^2)|_{s(t)=0})$, defined as the directivity coefficient, is a function of the noise azimuth ϕ . The noise power spectrums in channels 1 and 2 are different, which are also the functions of the noise azimuth. The directivity coefficient is used to estimate the gain of noise power spectrums between channels 1 and 2. When the inter-channel noise gain is estimated accurately and the two channels' noise power spectrums are then balanced to approximately the same, the ambient noise can be attenuated and the desired speech signal can be extracted. The noise directivity coefficient is further analyzed and simplified by Eq. (15):

$$\begin{aligned} \frac{E(|CH_1(e^{j\omega})|^2)|_{s(t)=0}}{E(|CH_2(e^{j\omega})|^2)|_{s(t)=0}} &= \frac{\sin^2(0.5\omega d/c \cdot (1 + \cos\phi))}{\sin^2(0.5\omega d/c \cdot (1 - \cos\phi))} \approx \frac{(0.5\omega d/c \cdot (1 + \cos\phi))^2}{(0.5\omega d/c \cdot (1 - \cos\phi))^2} \\ &= \cot^4(\phi/2) \end{aligned} \quad (15)$$

For an actual CI size constraint of $d \approx 0.01$ m, the directivity coefficient approaches to $\cot^4(\phi/2)$. This result indicates that the estimation of the directivity coefficient is robust, because it only depends on the noise direction ϕ . This simplified form of noise directivity coefficient also implies that, for noise direction with slowly varying, the adjusted gain for noise reduction can be accurately obtained with excellent algorithm stability.

Realization

Fractional delay

In Figure 1, the recorded signals $MIC_1(t)$ and $MIC_2(t)$ are sampled at the 44.1 kHz sampling rate by the AD converter as $MIC_1(n)$ and $MIC_2(n)$. These digital signals are then delayed by the fractional delay filter with an algorithm offset of d/c . In our hardware platform, the system design specifies the intermicrophone distance d to be at or near 1 cm, corresponding to 1.297 sampling points.

To obtain this accurate fractional delaying, we use the maximal flat (Maxflat) criteria [27-30] to design a fourth-order finite impulse response (FIR) filter for the required system delaying. For CI devices, the required speech energy is primarily in the low-frequency band, peaking near 1 kHz; this Maxflat FIR filter matches the desired speech characteristic well. And the fourth-order Maxflat FIR filter is given by

$$h(n) = [-0.0400, 0.6995, 0.4433, -0.1220, 0.0192] \quad (16)$$

For $f_s = 44.1$ kHz, the digital signals $MIC_1(n)$ and $MIC_2(n)$ are delayed, with d/c offsets of $MIC_1(n-1.297)$ and $MIC_2(n-1.297)$, respectively.

For the time-domain delay of 1.297 sampling points, the ideal system frequency response is $e^{-j\omega \times 1.297}$, which is a linear-phase all-pass filter. The proposed fourth-order

Maxflat FIR filter approaches the ideal filter at the range of 0–6000 Hz. The errors of magnitude and phase responses are plotted in Figure 2.

The proposed fourth-order Maxflat FIR filter agrees well with the ideal filter at the signal range of 0–6000 Hz, a range that includes most of the subbands of the CI filter bank. The maximal error for the magnitude response and phase response are less than 0.3% and 0.4%, respectively. Additionally, this filter can obtain the required delaying signal easily, with low computation complexity.

This Maxflat digital filter is used to cover frequencies between 0 Hz and 6000 Hz as the CIS strategy does. The range of the CIS actually depends on the corner and center frequencies of the filters, therefore, the frequency range will change based on different channel quantities. But both the present CI filter banks (8, 16, 24 channels etc.) are primarily within these range. So, the required signal delaying is accurate.

Directivity estimation and noise-frame identification

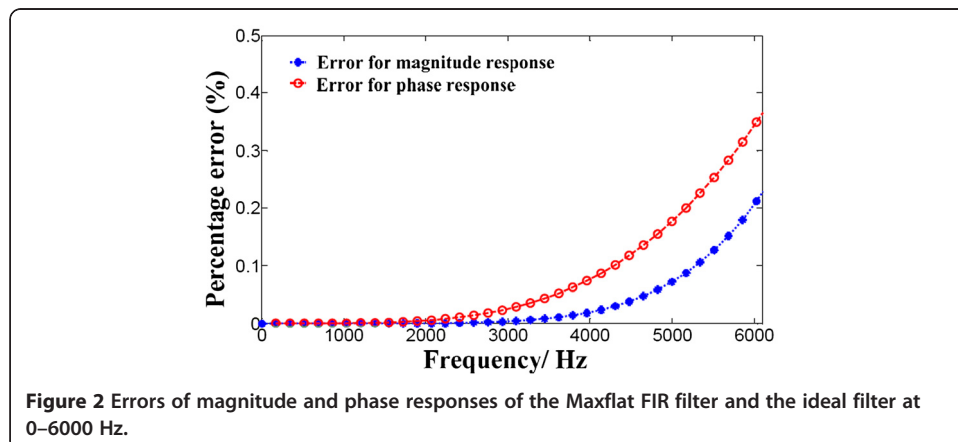
The directivity coefficient (Figure 1 and Eq. (15)) is obtained from the power spectrums of the 2 channels at the noise-only frames, which correspond to the time-domain signals of $ch_1(t)|_{s(t)=0}$ and $ch_2(t)|_{s(t)=0}$, respectively. We used the cepstrum [31-33] method to differentiate the desired speech frame and the noise frame.

The anterior several frames of data are considered to be noise. The cepstrum vector is denoted as C , which is expressed by a series of vector coefficients c_i . Then, the spectrum density function is given by Eq. (17)

$$\log S(\omega) = \sum_{n=-\infty}^{\infty} c_n e^{-j\omega n} \text{ and } c_0 = \int_{-\pi}^{\pi} \log S(\omega) \frac{d\omega}{2\pi} \quad (17)$$

The average cepstrum coefficients of the anterior several frames are used to estimate and obtain the cepstrum vector C . The cepstrum vector is updated by the current cepstrum vector C_i and the previous cepstrum vector C_{i-1} . The update equation is $C_i = \beta C_i + (1 - \beta) C_{i-1}$, where β is the update weight. In our algorithm, β is 0.85 and the corresponding cepstrum distance is given by

$$d_{cep}|_{(i+1)} = A \sqrt{(c_0|_{(i+1)} - c_0|_i)^2 + 2 \sum_{n=1}^p (c_n|_{(i+1)} - c_n|_i)^2} \quad (18)$$



Equation (18) with a predefined threshold value is used to differentiate the noise frames for both channels, $ch_1(t)|_{s(t)=0}$ and $ch_2(t)|_{s(t)=0}$. Then, the corresponding power spectra are used to obtain the real-time noise directivity coefficient.

Broadband signal adjustment

The speech is a broadband signal with multifrequency information. For the first order differential microphone, we previously proposed the normalized beamforming method for the gain [34] adjustment. But in this paper, we can compensate the gains directly in the dual-channel algorithm. Equation (14) provides the magnitude estimation for the desired speech changes as a function of signal frequency f . The coefficient is a function of the signal frequency denoted as Eq. (19).

$$\lambda(f) = 1/(2 \sin(2\pi fd/c)) \quad (19)$$

Therefore, the magnitude estimation for the desired speech is modulated by the $\lambda(f)$ to change its multifrequency gain. This coefficient function is monotone, decreasing between 100 and 6000 Hz. We design a digital filter to approximate $\lambda(f)$. The codomain of $\lambda(f)$ is between 0.6 and 27 when the frequency is between 100 and 6000 Hz. Because the maximal magnitude response of a filter is always 1, the desired coefficient function is actually written as Eq. (20).

$$\lambda'(f) = \frac{1}{30}\lambda(f) = \frac{1}{60 \sin(2\pi fd/c)} \quad (20)$$

We design a 1st order Butterworth filter, $Butter(f)$, with band-pass cutoff frequencies of $0.00045 f_s$ and $0.0045 f_s$, to approach the coefficient function $\lambda'(f)$, as shown in Figure 3.

The proposed filter for multi-frequency adjusting is highly consistent with the required coefficient function between 100 and 6000 Hz. Because $\lambda(f) = 30\lambda'(f)$, the filtered signal needs an additional 30 times gain (or 29.54 dB) for signal energy rebalancing.

For a CI speech-processing strategy based on a filter bank, such as the continuous interleaved sampling strategy (CIS [35]) or advanced combined encoder strategy (ACE [36]), the adjusting coefficients for the multi-frequency signal can be transferred directly to the corresponding subband filters. According to the characteristics of the cochlea [37], the speech signal can be divided into 16 bands (Table 1) and the corresponding band

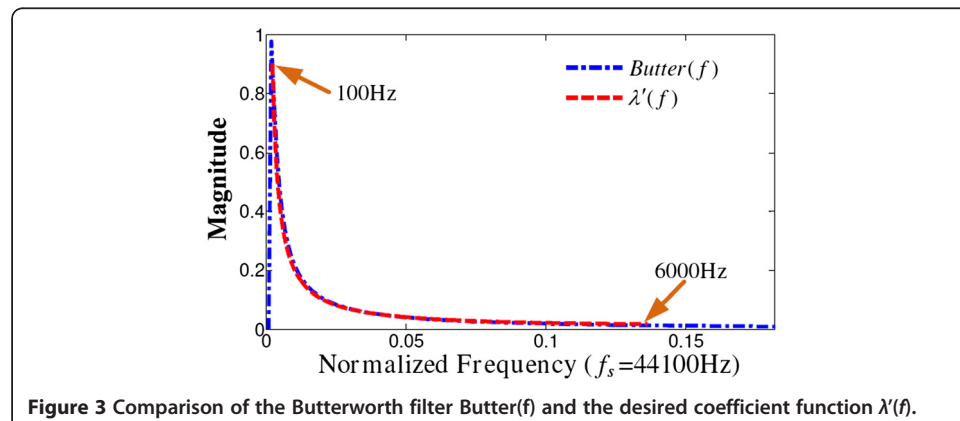
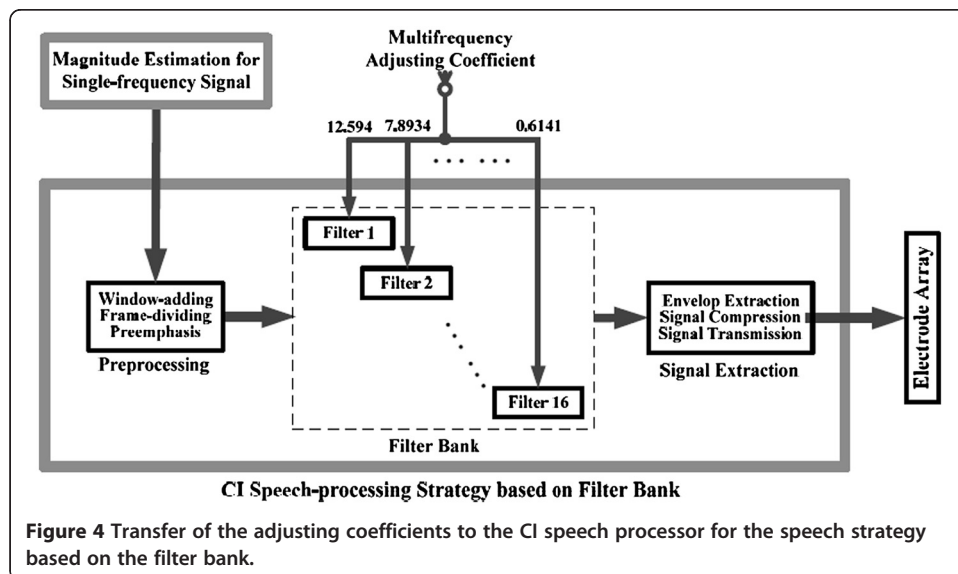


Table 1 Parameters of each sub-band in CI filter bank and the corresponding adjusting coefficients

	Band edge (Hz)	Center frequency (Hz)	Adjusting coefficient
Channel 1	[156, 276]	216	12.5294
Channel 2	[276, 410]	343	7.8934
Channel 3	[410, 560]	485	5.5861
Channel 4	[560, 730]	645	4.2047
Channel 5	[730, 922]	826	3.2883
Channel 6	[922, 1138]	1030	2.6428
Channel 7	[1138, 1380]	1259	2.1685
Channel 8	[1380, 1653]	1517	1.8071
Channel 9	[1653, 1960]	1807	1.5255
Channel 10	[1960, 2305]	2133	1.2135
Channel 11	[2305, 2694]	2500	1.1217
Channel 12	[2694, 3131]	2913	0.9752
Channel 13	[3131, 3623]	3377	0.8557
Channel 14	[3623, 4176]	3840	0.7578
Channel 15	[4176, 4798]	4487	0.6781
Channel 16	[4798, 5498]	5148	0.6141

center frequencies are also listed in Table 1. Eq. (19) is used to calculate the desired adjusting coefficient for each band of the filter bank, shown in this Table and Figure 4.

Figure 4 describes the transmission of multi-frequency adjusting coefficients to the CI processor. This method of directly transmitting the parameters to the CI filter bank requires very little additional calculation, which is suitable to the situation of a filter bank-based strategy. For the situation in which the CI processor uses the speech-processing strategy without a filter bank, the proposed Butterworth filter (Figure 3) should be adopted for the coefficient adjusting.



Signal reconstruction

In our CI speech-enhancement platform, the sampling rate is 44.1 kHz and the Hamming window is used for framing, with a window length of 1024 sampling points. Each frame is about 23 ms in duration, with 50% overlap.

The speech magnitude is estimated by Eq. (14). The signal phase of the original data (channel 1) is used directly for signal construction, because the human cochlea is relatively insensitive to phase information. The gain of the single-frequency signal is adjusted by the proposed Butterworth filter or by directly transmitting the adjusting coefficients in the filter-bank based CI processor. This broadband signal, expressed in the form of frequency response, is processed by the subsequent processing of the inverse Fourier transform and deoverlapping to reconstruct the enhanced speech signal.

Results

Hardware platform

A dual-channel CI front-end hardware platform was constructed (Figure 5). The dual microphones were linearly spaced 1 cm apart. The recorded signals, which were obtained with a real-time acquisition process controlled by a software interface, were transmitted to the computer.

The experiments were carried out in a chamber, or to be exact, an actual office measured by 8 m × 8 m × 4 m with the room reverberation time $T_{60} = 450$ ms. Two microphone modules, placed at the center (O), recorded the signal. P₁-P₁₂, which represent 12 testing points at 15° intervals, were marked 1.5 m from the microphones arranged in a semicircle. P₁ indicates the forward direction for playing the desired speech by Speaker 1; the other locations (P₂-P₁₂) indicate the directions for playing the ambient noise by Speaker 2. The recorded signals by this hardware system, after amplification, filtering, and analogue to digital conversion, were transmitted to the computer via a USB interface for further analysis.

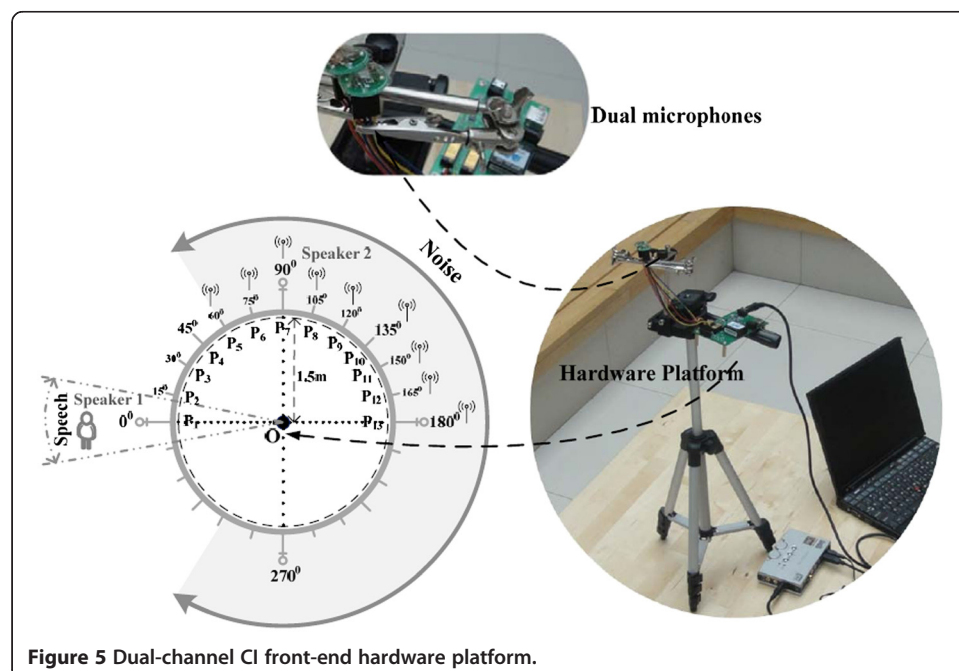


Figure 5 Dual-channel CI front-end hardware platform.

Speech-enhancement result

For this test, a speech signal (i.e., the English sentence “I heard you called me” spoken by a native speaker (American English) as material) was played by Speaker 1 at P_1 . Speaker 2 played ambient noise, including 2 types of noise signals, at P_7 . One noise signal was the theme song “My Heart Will Go On” from the movie Titanic; the other was the speech signal in an interviewing scene, with part of the speech content being “My background and work experience are tailor-made for this position. I studied marketing as an undergrad here in Taiwan”. The desired speech and the ambient noise were played at the same power. For this situation, the SNR was approximately 0 dB, which is an extremely poor noise environment. The enhancement results were compared with those of the single-channel method of spectral subtraction (Figure 6).

Figure 6 (a) is the original waveform of the desired speech, played by Speaker 1 in P_1 . Speaker 2 located at P_7 plays the noise. Figure 6 (b-1) and (b-2) are the plots recorded by the hardware platform (located at O) corresponding to the situations of music noise and speech noise, respectively.

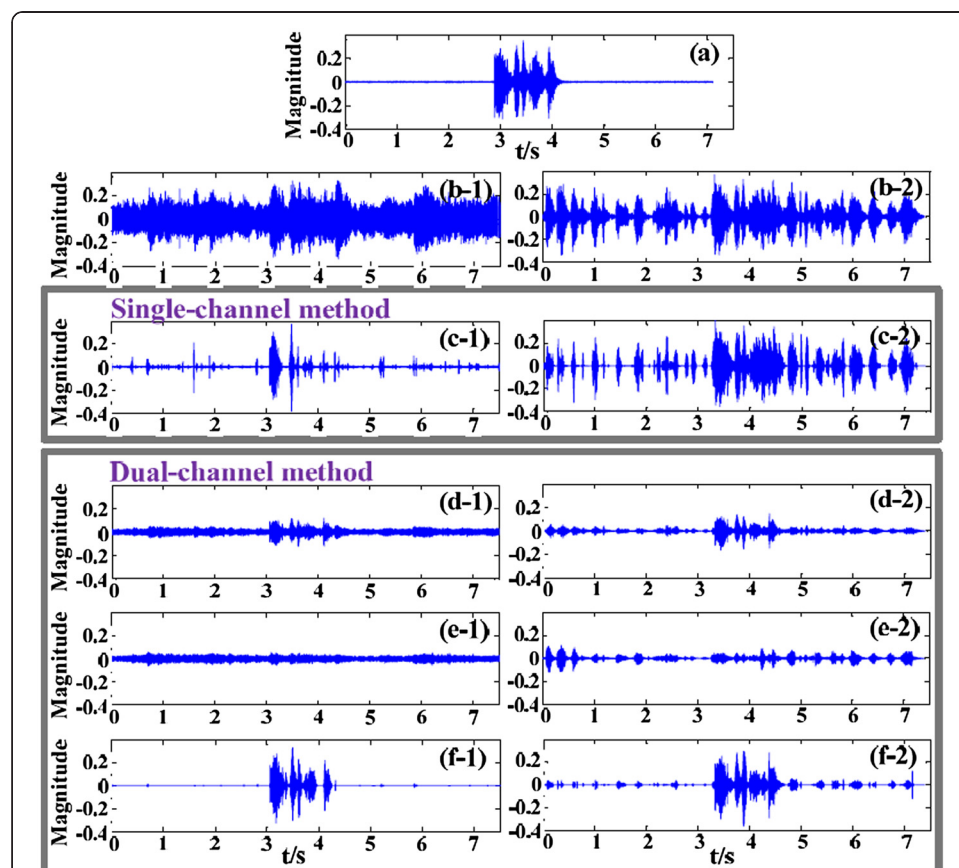


Figure 6 Speech-enhancement results: comparison of the proposed algorithm and the single-channel method. (a) Original signal played as the desired speech. (b) Signal recorded by the omnidirectional microphone at O (**b-1**: includes the desired speech and the music noise; **b-2**: includes the desired speech and the ambient speech noise). The corresponding speech-enhancement results for the single-channel method (**c-1** and **c-2**) and our proposed algorithm (**f-1** and **f-2**) are plotted, and **d-1**, **e-1** and **d-2**, **e-2** are the corresponding signal outputs of channel 1 and channel 2 for music and speech noises respectively.

In this paper, the improved SNR is defined by Eq. (21) [38,39]:

$$\Delta SNR = \sum_{j=1}^J w_j (SNR_{j,out} - SNR_{j,in}) \quad (21)$$

where J is the frequency band quantity, w_j is the corresponding weight for different band given in [39], and the input SNR and output SNR is given by Eqs. (22) and (23).

$$SNR_{in} = 10 \lg \left(\frac{\sum_n |s(n)|^2}{\sum_n |n(n)|^2} \right) \quad (22)$$

$$SNR_{out} = 10 \lg \left(\frac{\sum_n |\hat{s}(n)|^2}{\sum_n |\hat{n}(n)|^2} \right) \quad (23)$$

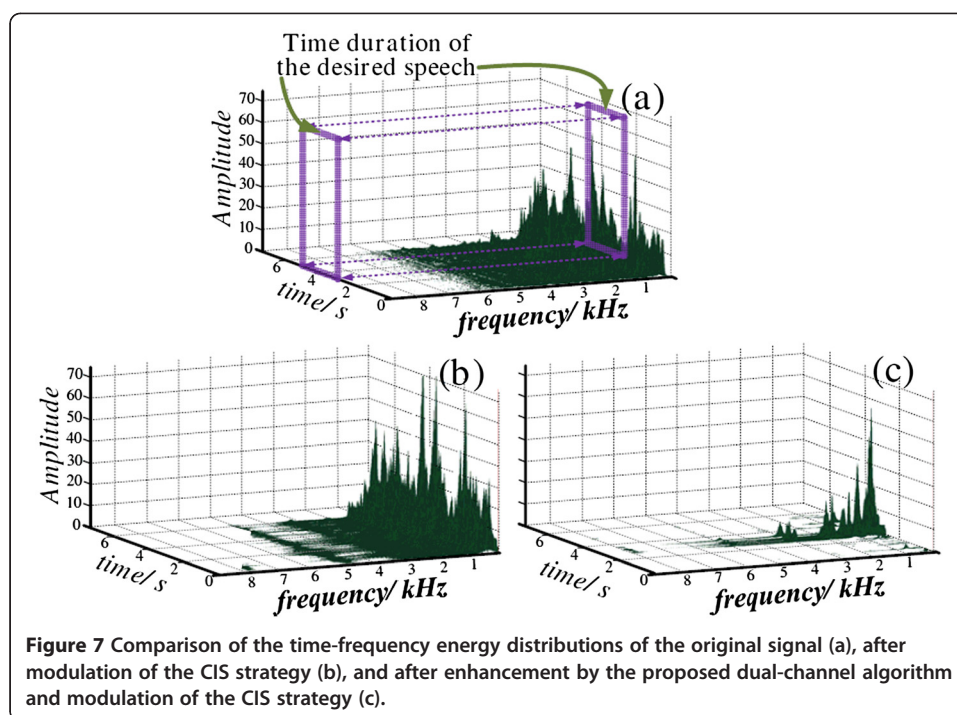
where the output SNR uses the estimation of $\hat{s}(n)$ and $|\hat{n}(n)|$ to obtain.

For the music noise, use of the single-channel method (panel c-1) based on spectral subtraction weakened most of the music noise, but much of the transient impulse at the nonstationary part of the music noise remained. Panels d-1 and e-1 plot the signal outputs in channel 1 and channel 2 in our dual-channel system. Comparison of (d-1) and (e-1), the magnitude attenuation or enhancement were different for the desired speech and music noise. And this characteristic can remarkably be seen in (d-2) and (e-2) for the ambient speech noise. The waveforms of the two channels were similar in time domain, but the gains in channel 2 were discrepant, with respectively about 0.3 and 2.8 gains for the desired speech and ambient speech compared with channel 1. These gains changed when the noise moved. The previous directivity coefficient in our algorithm was used to estimate the noise gain between channels 1 and 2 in the noise-only intervals. For accurate noise gain estimation, two channels' noise power can be adjusted to nearly the same but noticeably discrepant for the power of the desired speech. Then the desired speech can be extracted from the subtraction of the adjusted signals in these two channels. Our proposed method suppressed the overall music noise, including the instantaneous noise (panel f-1). This method also suppressed the ambient speech noise, with nearly 20 dB SNR improvement (panel f-2). The single-channel method did not adequately suppress the ambient speech noise (panel c-2). The comparison indicates that the proposed dual-channel speech-enhancement algorithm successfully suppresses the nonstationary noise, which adds to its practical value.

The enhanced speech signal was further processed in CI speech processor. The CI speech strategy extracts and encodes the signal and then wirelessly transmits it to the electrode array. We adopted the widely used CIS strategy. The sine simulation model was used to implement the CI signal-processing. The energy distribution after CIS processing is shown in Figure 7.

For this test, the original signal, recorded by the platform, contained desired speech and ambient music noise. The signal duration was 7.5 sec and the desired speech was located in the time axis approximately between 3 sec and 5 sec. Figure 7 (a) describes the time-frequency energy distribution of original signal, which was dispersed in the frequency range between 0 and 6000 Hz.

Figure 7(b) describes the energy distribution of the signal after the modulation of the CIS strategy. This speech strategy divided the original signal by the filter bank and then



modulated the subband signal, by using the center frequency of each band to characterize its corresponding information for further speech synthesis. The plot indicates that the CIS modulation only changed the frequency-domain energy distribution, but maintained the time-domain energy distribution. As a result, the signal energy concentrated in the corresponding center frequency of each subband, and the ambient noise in the time domain was not suppressed.

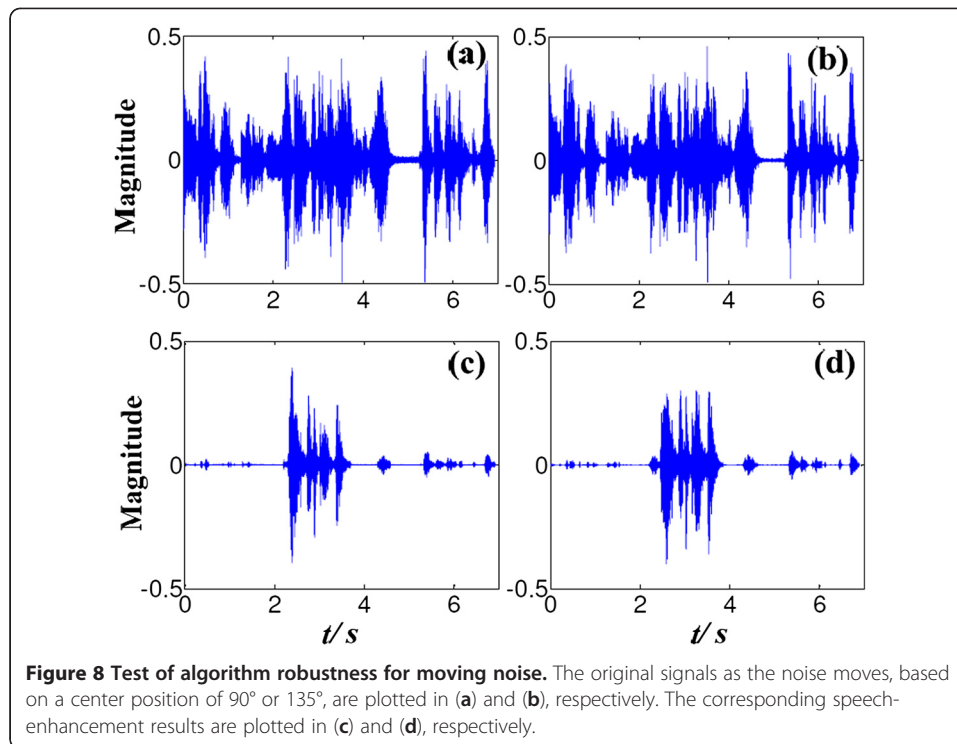
Figure 7(c) describes the energy distribution of the signal after enhancement by the proposed dual-channel algorithm and modulation of the CIS strategy. The energy distribution changed in the time domain, primarily between 3 sec and 5 sec, and the ambient noise was sharply weakened.

Together, the plots in Figure 7 indicate that the speech enhancement can achieve the following 2 purposes. First, the desired speech remained while the ambient noise was sharply suppressed, which improved the CI speech recognition. Second, the global signal energy was lowered, and the CI battery life was prolonged, because information from the signal and the energy both are transmitted wirelessly to the inner part of the CI device.

Algorithm robustness and signal distortion analysis

The test aims to the analysis of algorithm robustness when the ambient noise was moving. The desired speech was played by Speaker 1 (P_1 in Figure 5). Speaker 2 (located at P_7 , about 90° azimuth) played the ambient noise. During the testing, Speaker 2 moved back and forth at a speed of 1 m/s, corresponding to normal human walking velocity. The moving range was about 30° , from 75° to 105° . Another test, with the speaker located at P_{10} and moving from 120° to 150° at the same speed, was performed. The speech-enhancement results are shown in Figure 8.

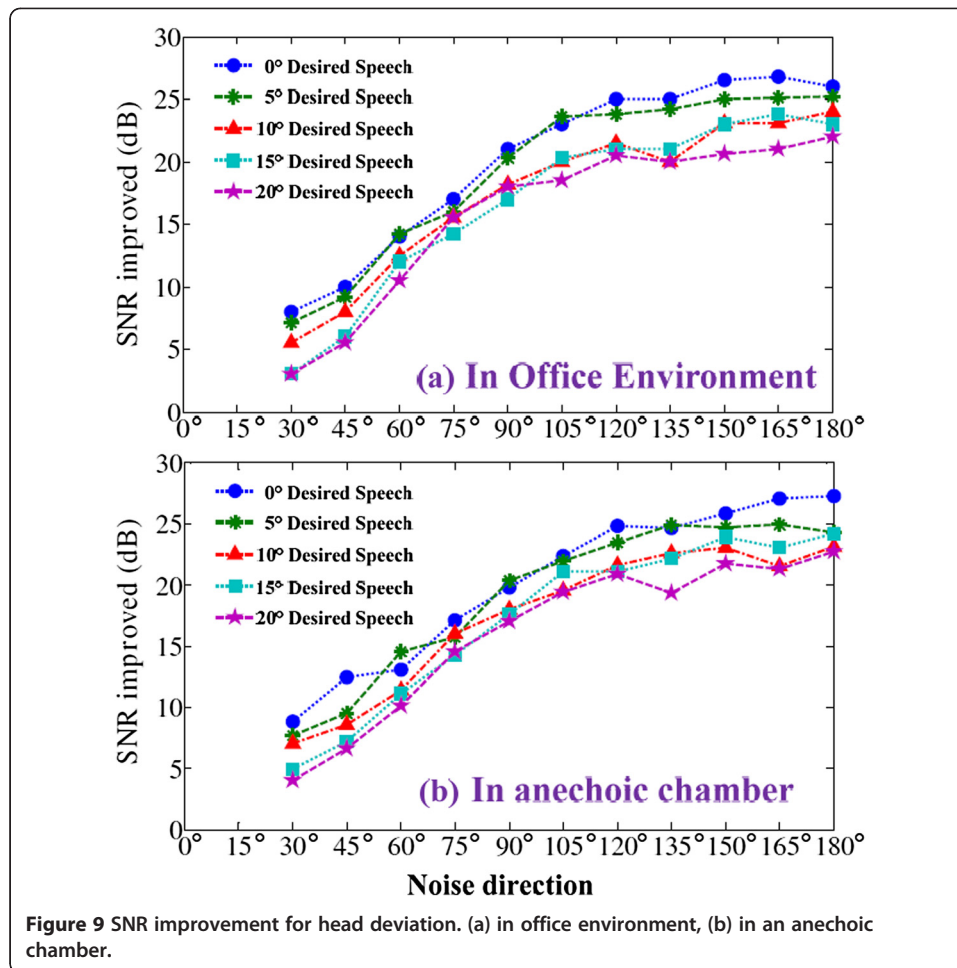
For the tests in these 2 situations (moving noises from 90° and 135°), the original signals recorded by the omnidirectional microphone were plotted in (a) and (b),



respectively. The original signal consists of the desired speech and the ambient speech noise. The noise suppression results are plotted in (c) and (d), respectively. A comparison of these plots reveals that the proposed algorithm also effectively weakens moving noise, with an SNR improvement of about 15 dB. The conventional noise-reduction methods need to reconvergent in algorithm for coefficients updating, and will always result in noise leakage and noticeable SNR decrease. The mentioned MVDR method, one of the most widely used adaptive beamformer, can choose and adjust the filter coefficients to minimize the output power with the constraint that the desired signal is not to be distorted. For moving noise, the MVDR method will also partly result in noise leakage, which attenuates the algorithm performance. The proposed algorithm calculates the noise directivity coefficient for moving noise, and also remains excellent performance with a few attenuation of SNR. As shown in result, the proposed algorithm is advantageous to avoid the noise leakage and is robust to mobile noise.

For actual CI devices, the dual microphones may not remain exactly collinear with the forward direction. Additionally, the head offset for CI users results in an orientation deviation of the desired speech. For daily face-to-face conversation, the microphone bias is primarily $<20^\circ$. In this test, the orientation offset was 20° , and the ambient noise moved from 30° to 180° (the mirror-reversed orientation is between 180° and 330°), which covers the most probable range of head deviation and the noise direction. The desired speech (5° intervals) and the ambient noise (15° intervals) were played. The speech-enhancement results are plotted in Figure 9(a), in which the original input SNR was 0 dB.

Figure 9 describes the SNR improvement (in dB) for all situations in which the noise comes from 30° to 180° and the desired speech is played at the azimuth of 0° , 5° , 10° , 15° , or 20° . In the office environment with $T_{60} = 450$ ms, for a fixed



speech direction, the improved SNR (Figure 9 (a)) was higher when the noise azimuth approached 180° (backward), and was lower when the noise approached the desired speech (forward). For the situation of speech deviation to the 0° azimuth, greater offset resulted in less SNR improvement. The plot also indicates that, for a speech deviation range of 0° to 20° and noise range of 60° to 180°, the SNR improvement was >10 dB. For a noise azimuth of 180° to 300°, the expected analogous and mirror-reversed result was obtained. For comparison, experiments were also carried out in an anechoic chamber ($T_{60} = 100$ ms), and the SNRs improvement for head deviation are plotted (Figure 9 (b)). A set of similar SNR results are obtained, with only 1 to 3 dB globally SNR increased in situation of anechoic environment. The room reverberation actually influences the algorithm performance, but in an acceptable constraint.

The prevalent algorithms can be seen as the beamformers to extract the desired signal from a certain direction while minimizing the output power of the ambient noise from another direction. These methods are advantageous in low speech distortion or high noise-reduction performance. Therefore, a compromise between signal distortion and noise suppression is needed. For spectrum estimation and subtraction based method, the distortion of the desired speech is unavoidable in our algorithm. To delineate the speech

distortion, the speech distortion index is used [40-42], with the vector expression given in Eq. (24).

$$v_{sd}(\mathbf{h}) = \frac{(\mathbf{h}_1 - \mathbf{h})^T \mathbf{R}_{xx} (\mathbf{h}_1 - \mathbf{h})}{\mathbf{h}_1^T \mathbf{R}_{xx} \mathbf{h}_1} \quad (24)$$

And the actual expression for calculation is given in Eq. (25).

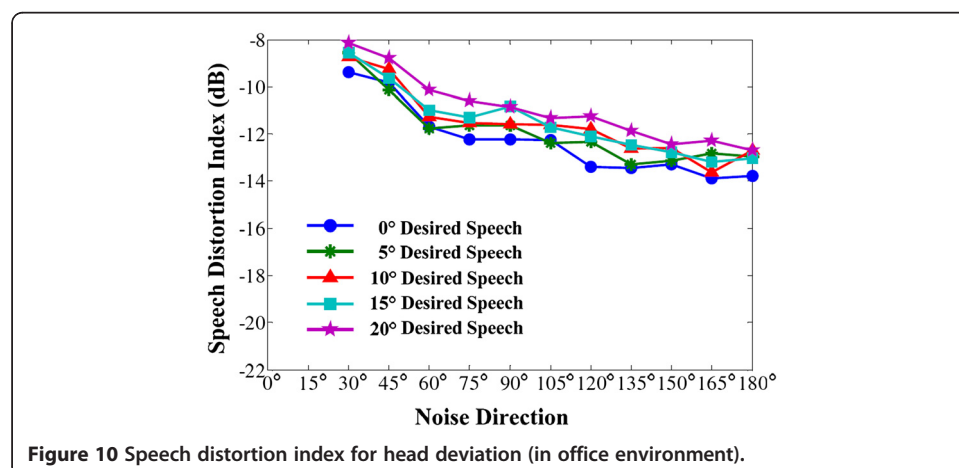
$$v_{sd}(\mathbf{h}) = \frac{\sum_{i=1}^L \frac{\lambda_i}{(1 + \lambda)^2} b_{i1}^2}{\sum_{i=1}^L \lambda_i b_{i1}^2} \quad (25)$$

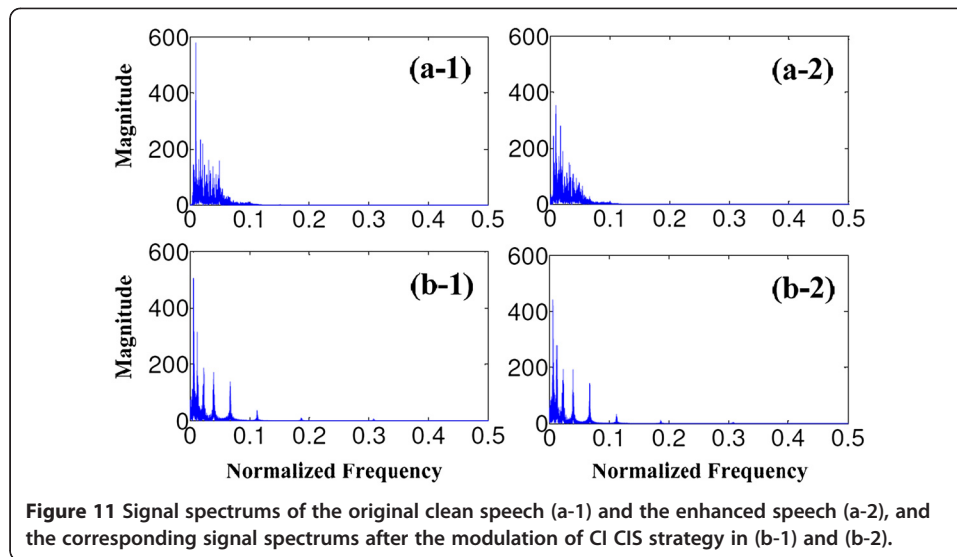
where λ_i is the element of diagonal matrix given in [42]. Actually, the speech distortion index presents the attenuation between speech power and the original clean speech.

And the distortion results are presented in Figure 10.

Figure 10 depicts the speech distortion index (dB) for head deviation, corresponding the office-environment experiments also for all situations in which the noise comes from 30° to 180° and the desired speech is played at the azimuth from 0° to 20°. As expect, the speech distortion is noticeably large, ranging from -8 to -14 dB. It implies that our algorithm obtains high SNR but with a little bit large of speech distortion. Compared with other conventional methods in decreasing speech distortion, the proposed algorithm is not advantageous, or of great disadvantage compared with the time-domain beamformers (ANF etc.). But after the modulation of the CIS strategy in CI processor, with the signal envelope and signal information of the enhanced speech extracted and transferred to the CI electrode array, the signal distortion will be attenuated. For clear comparison, the signal magnitude spectrums are used to analyze the speech distortion, shown in Figure 11.

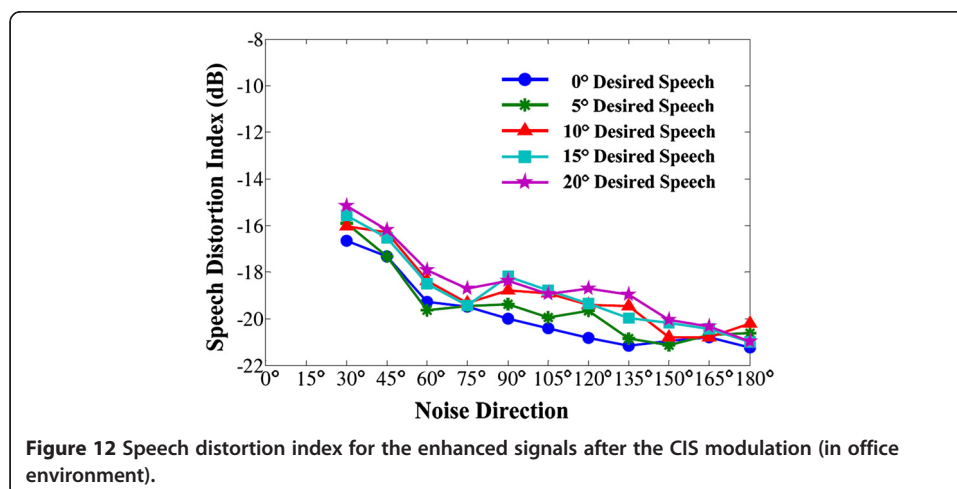
In Figure 11, panels a-1 and a-2 describe the signal spectrums of the original clean speech and the enhanced speech, and panels b-1 and b-2 are the corresponding spectrums after CIS modulation. Comparing the result in panel a-1 and a-2 (speech-enhancement result before the processing of CIS strategy) the spectrum difference is noticeable. But after the CIS processing (panels b-1 and b-2), with the signal energy





concentrated in the central frequency of each sub-band in the CI filter bank, signal spectrums change to be approximately the same. These results imply that the final modulated signal to the CI electrodes will become low distorted. To further quantify the speech distortion for the enhanced signal after the CIS processing, another graph of the speech distortion index is plotted (Figure 12).

Figure 12 describes a graph of the speech distortion index for the enhancement of the desired speech after the CIS processing. The speech distortion indexes are sharply small, mainly between -18 and 21 dB. A smaller value of the speech distortion index means the desired signal is less distorted. Compared with figure 10, it prominently depicts a set of low speech distortion for the application of CI devices, which is in an acceptable range of signal distortion. For CI front-end signal acquisition, our algorithm is advantageous for a large amount of SNR improvement, but disadvantageous in a little bit of greater speech distortion when comparing with other low-distortion algorithms. However, for signal modulation and transmission by the CI CIS strategy, the speech distortion is sharply decreased.



These results, including the test of moving noise, head deviation (Figure 9) and algorithm evaluation of speech distortion (Figures 10, 11 and 12), indicate that the proposed algorithm is robust and flexible for CI speech-enhancement, with great SNR improvement and low speech distortion.

Discussion

Approximation of directivity coefficient

The approximation of the directivity coefficient in Eq. (15), corresponding to the noise azimuth ϕ , is important to algorithm performance. The hardware platform was also constructed for the experiments to evaluate the approximation. Firstly, the test was carried out in an anechoic chamber ($T_{60} = 100$ ms). The loudspeaker played music as the ambient noise at 90° , 180° and 270° orientations respectively. The target speech located at 0° orientation, with equal-power signal play from the loudspeaker (SNR = 0). The calculated orientations for the noise azimuth ϕ in Eq. (15) are 81° , 192° and 280° . The orientation error is about 10° for this situation. And the corresponding results are 77° , 171° and 285° respectively for the test in office environment ($T_{60} = 450$ ms), with about 15° errors. These errors is acceptable for the CI application, therefore, the directivity coefficient can be applied in the estimation of desired signal power spectrum in Eq. (14).

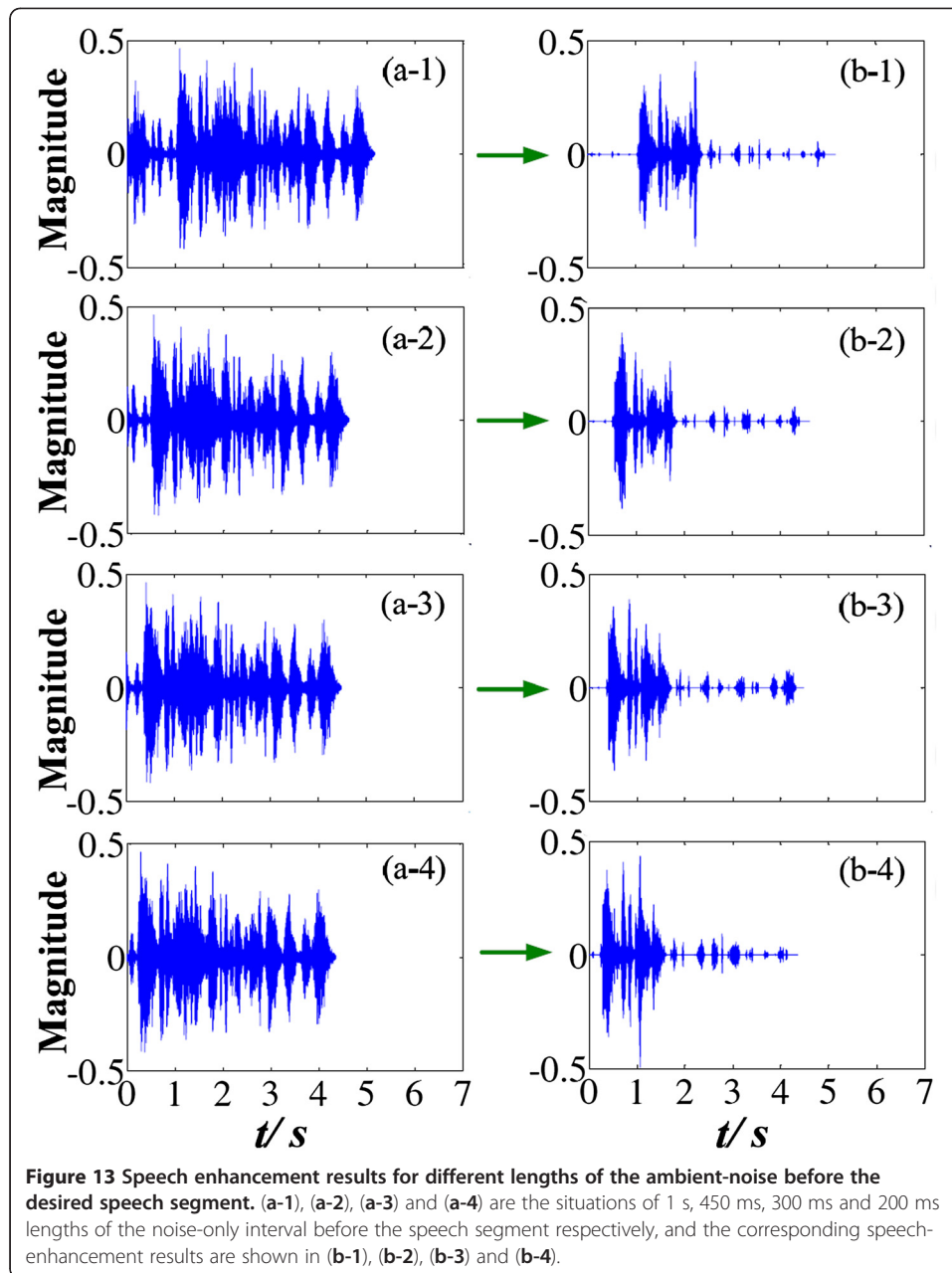
Algorithm constraint for noise estimation

As detailed analysis of the directivity coefficient in the aforementioned sections (seen in Figure 1, Eqs. (14) and (15)), the cepstrum method) was used for noise estimation. To simplify the test, as well as to be convenient for algorithm analysis and performance evaluation, the previous experiments use two long noise-only periods (about 3 seconds each) before and after the desired speech segment (about 1 second). However, the length of the ambient noise interval does not need as long as 3 seconds. For situations of shorter ambient noise, the results for speech enhancement are listed in Figure 13.

There are noise-only segments before and after the desired speech segment. In our algorithm, the anterior noise segment (before the desired speech) is used to estimate the directivity coefficient, and the length of which will influence the algorithm performance. For situations of the ambient noise with length no shorter than 450 ms (panels a-1 and a-2), the enhanced signals (panels b-1 and b-2) still remain great SNR improvement. For shorter length of ambient noise (panels a-3 and a-4), the SNR decreases noticeably (panels b-3 and b-4), as well as more speech distortion. To remain algorithm performance and low speech distortion in CIS modulation for CI devices, the minimal length of the noise-only period before the speech segment is about 0.5 second. This delaying time is acceptable for CI users in daily conversation. Specially, if we do not need a great SNR improvement, the length of the noise-only segment for pre-estimation can decrease to be about or less than 200 ms.

Distortion in CI processing

The previous results indicate that the proposed algorithm introduced a bit larger distortion to the desired speech but with noticeable low distortion after the CI processing. This phenomenon may result from the speech processing strategy of CI devices. The CIS strategy is a speech processing strategy to extract the signal information from the



time-domain envelope and then be transferred to the electrode array. This CIS speech strategy primarily includes the process of window-adding, frame-dividing, pre-emphasis, sub-band dividing, envelop extraction and signal compression. As the CIs only have several channels, correspond to a few specific stimulating rates, the extracted envelope may lose lots of signal information. In addition, in each band of the CI channel, only one frequency (correspond to the center frequency of the CI filter bank) is applied to modulate the desired signal. That is, a set of sinusoidal signals (only 16 or 24 for CI devices) are modulated by the corresponding envelope of the band-pass signals in the CI filter bank. Therefore, the single-frequency modulation processing seems to be a smoothness process to reduce the distortion. For example, taking the distortion in the band of [1653, 1960] (channel 9) into account. If the 1700 Hz signal is strengthened

but the 1900 Hz signal is weakened, the distortion will be introduced in this band. But after the CI modulation, both frequencies in this band, the center frequency 1807 Hz sinusoidal signal is applied to modulate the envelope based on the whole band energy. Therefore, the difference between the same bands will be smoothed, and the speech distortion after CI processing may come from the difference between different bands. Consequently, the CI speech strategy can reduce speech distortion and more aggressive algorithm can be applied in the CI application.

Algorithm performance

The proposed algorithm is an aggressive noise-suppression method with high SNR improved but a bit large distortion. But the distortion can be reduced in the CI processing and we can obtain excellent performance. The prevalent Frost algorithm based methods, such as linearly constrained minimum-variance and MINT algorithms, can suppress the noise with less signal distortion. These methods use iterative-adaptive technique to update the filter coefficients by gradient estimation and are advantageous in minimizing the ambient noise with low or no distortion for desired signal. Whereas, when the moving noise (or in the situation that the noise changes its azimuth) is present, these methods will weaken their noise-reduction performance. To obtain the optimal filter coefficients, if the algorithm does not reconvergent at the beginning or too slow to update the new coefficients, the noise will leakage and the desired performance will be attenuated. Other methods, MVDR and the binaural frequency-domain minimum-variance algorithm etc., present effective ways for noise suppression. Though these algorithms can converge more quickly, they will also cause noise leakage and the algorithm performance will be weakened. The approximation of directivity coefficient in our algorithm is tested, and about 10° error can be found. The directivity coefficient estimation is accurate enough to separate the desired speech and the ambient noise. For the trade-off between SNR improvement and speech distortion, the prevalent optimal-filter methods minimize the speech distortion while guarantying a certainly level of SNR improvement or maximize the SNR improvement while guarantying a certainly level of speech distortion. So it is hard for these algorithms to obtain both high SNR and low signal distortion. However, in the application of cochlear implant, the CI processing helps to obtain excellent speech enhancement while guarantying low speech distortion in the proposed algorithm.

Conclusions

The proposed speech-enhancement algorithm based on a dual-channel microphone array and spectral estimation technique aims to suppress the directional noise and improve the speech recognition of CI devices. A hardware platform was constructed and the experiments were carried out in an office to evaluate the algorithm performance in a real working environment for CI users. The experimental results indicated the excellent algorithm performance for speech enhancement. For stationary and moving noises, in orientations from lateral to rear, the improvements in SNR were 20 and 15 dB, respectively. For the situation of $\pm 20^\circ$ speech deviation and a broad range of noise azimuths from 60° to 300° , SNR improvement of >10 dB was maintained. Also, the speech distortion was very low when evaluating the modulated signal in CIS processing. The proposed algorithm is robust to mobile noise and signal orientation deviation and is applicable to the improvement of the front-end signal acquisition and speech recognition for CI devices.

Abbreviations

CI: Cochlear implant; SNR: Signal to noise ratio; FDM: First-order differential microphone; ANF: Adaptive null-forming method; MINT: Multiple input/output inverse method; MVDR: Minimum-variance distortionless-response technique; Maxflat: Maximal flat; FIR: Finite impulse response; CIS: Continuous interleaved sampling strategy; ACE: Advanced combined encoder strategy.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

YC initiated and conceived the algorithm, designed the hardware system and experiments and analyzed the data. QG is the corresponding author. This study originated from QG's idea and problems were solved under her directions. QG drafted this paper's manuscript, including content and overall arrangement. QG was also responsible for revising this manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors are grateful to the support of National Natural Science Foundation of China under the grant No. 61271133, Beijing Natural Science Foundation under the grant No. 3082012, Basic Development Research Key Program of Shenzhen under the grant No. JC201105180808A.

Received: 18 June 2012 Accepted: 27 July 2012

Published: 24 September 2012

References

1. Chung K, Zeng FG: **Using hearing aid adaptive directional microphones to enhance cochlear implant performance.** *Hear Res* 2009, **250**:27–37.
2. Nelson PB, Jin SB, Carney AE: **Understanding speech in modulated interference: cochlear implant users and normal-hearing listeners.** *J Acoust Soc Am* 2003, **113**(2):961–968.
3. Zeng FG, Nie K, Stickney GS, Kong YY, Vongphoe M, Bhargava A, Wei C, Cao K: **Speech recognition with amplitude and frequency modulations.** *Proc Natl Acad Sci U S A* 2005, **102**(7):2293–2298.
4. Donaldson GS, Kreft HA, Litvak L: **Place-pitch discrimination of single- versus dual-electrode stimuli by cochlear implant users (L).** *J Acoust Soc Am* 2005, **118**(2):623–626.
5. Kwon BJ, van den Honert C: **Dual-electrode pitch discrimination with sequential interleaved stimulation by cochlear implant users.** *J Acoust Soc Am* 2006, **120**(1):1–6.
6. Izzo AD, Richter CP, Jansen ED, et al: **Laser stimulation of the auditory nerve.** *Lasers Surg Med* 2006, **38**(8):745–753.
7. Chung K, Zeng FG, Acker KN: **Effects of directional microphone and adaptive multichannel noise reduction algorithm on cochlear implant performance.** *J Acoust Soc Am* 2006, **120**:2216–2227.
8. Spriet A, Van Deun L, Eftaxiadis K, Laneau J, Moonen M, Van Dijk B, Van Wieringen A, Wouters J: **Speech understanding in background noise with the two-microphone adaptive beamforming BETM in the Nucleus Freedom cochlear implant system.** *Ear Hear* 2007, **28**(1):62–72.
9. Boll SF: **Suppression of acoustic noise in speech using spectral subtraction.** *IEEE Trans Acoust Speech Signal Process* 1979, **27**(2):113–120.
10. Kamath S, Loizou P: **A multi-band spectral subtraction method for enhancing speech corrupted by colored noise.** In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, Orlando, Florida, 2002*:675–678.
11. Scalart P, Filho JV: **Speech enhancement based on a priori signal to noise estimation.** In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, volume 2, Atlanta, GA, Volume 2, 1996*:629–632.
12. Ephraim Y, Van Trees H: **A signal subspace approach for speech enhancement.** *IEEE Trans Speech Audio Process* 1995, **3**(4):251–266.
13. Bitzer J, Simmer KU, Kammeyer KD: **Theoretical noise reduction limits of the generalized sidelobe canceller (GSC) for speech enhancement.** In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, Phoenix, AZ, Volume 5, 1999*:2965–2968.
14. Flanagan JL: **Computer-steered microphone arrays for sound transduction in large rooms.** *J Acoust Soc Am* 1985, **78**:1508–1518.
15. Marciano JS, Vu TB: **Reduced complexity beam space broadband frequency invariant beamforming.** *Electron Lett* 2000, **36**:682–683.
16. Elko GW, Pong ATN: **A simple adaptive first-order differential microphone.** In *Proceedings of IEEE workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, USA, 1995*:169–172.
17. Fischer S, Simmer KU: **Beamforming microphone arrays for speech acquisition in noisy environments.** *Speech Commun* 1996, **20**(3):215–227.
18. Luo FL, Yang J, Pavlovic C, et al: **Adaptive null-forming scheme in digital hearing aids.** *IEEE Trans Acoust Speech Signal Process* 2002, **50**(7):1583–1590.
19. Frost OL: **An algorithm for linearly constrained adaptive array processing.** *Proc IEEE* 1972, **60**:926–935.
20. Van Veen BD, Buckley KM: **Beamforming: a versatile approach to spatial filtering.** *IEEE Trans Acoust Speech Signal Process* 1988, **5**:4–24.
21. Miyoshi M, Kaneda Y: **Inverse filtering of room acoustics.** *IEEE Trans Acoust Speech Signal Process* 1988, **2**(36):145–152.
22. Capon J: **High resolution frequency-wavenumber spectrum analysis.** *Proc IEEE* 1969, **57**:1408–1418.

23. Johnson DH: The application of spectral estimation methods to bearing estimation problems. *Proc IEEE* 1982, **70**:1018–1028.
24. Lockwood ME, Jones DL, et al: Performance of time- and frequency-domain binaural beamformers based on recorded signals from real rooms. *J Acoust Soc Am* 2004, **115**(1):379–391.
25. Kates JM, Weiss MR: A comparison of hearing-aid array-processing techniques. *J Acoust Soc Am* 1996, **99**:3138–3148.
26. Qin G, Yousheng C: Parameter section methods of delay and beamforming for cochlear implant speech enhancement. *Acoust Phys* 2011, **57**(4):542–550.
27. Thiran JP: Recursive digital filters with maximally flat group delay. *IEEE Trans Circuits Theory* 1971, **18**(6):659–664.
28. Cooklev T, Nishihara A: Maximally flat FIR filters. In *Proc. IEEE Int. Symp. Circuits Syst. Chicago, USA*; 1993:96–99.
29. Valimaki V, Laakso TI: Principles of fractional delay filters. In *IEEE ICASSP'00*; 2000:3870–3873.
30. Martin E, Arild L: Maximally flat FIR and IIR fractional delay filters with expanded bandwidth. In *European Signal Processing Conference. Poznan, Poland*; 2007:1038–1042.
31. Oppenheim AV, Schfer R: *Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall; 1975.
32. Furui S: Cepstral analysis technique for automatic speaker verification [J]. *IEEE Trans Acoust Speech Signal Process* 1981, **29**(2):254–272.
33. Viikki O, Bye D, Laurila K: A recursive feature vector normalization approach for robust speech recognition in noise. In *Proceedings of ICASSP' 98*. Seattle, WA, USA: IEEE Acoustics, Speech and Signal Processing Society; 1998:733–736.
34. Yousheng C, Qin G: A normalized beamforming algorithm for broadband speech using a continuous interleaved sampling strategy. *IEEE Trans Audio Speech Language Process* 2011, **20**(3):868–874.
35. Wilson BS, Lawson DT, Zerbi M, et al: Design and evaluation of a Continuous Interleaved Sampling(CIS) processing strategy for multichannel cochlear implants. *J Rehabil Res Dev* 1993, **30**(1):110.
36. Psarros CE, Plant KL, Lee K, et al: Conversion from the SPEAK to the ACE strategy in children using the Nucleus 24 cochlear implant system: speech perception and speech production outcomes. *Ear Hear* 2002, **23**(18):18.
37. Greenwood DD: A cochlear frequency-position function for several species-29 years later. *J Acoust Soc Am* 1990, **87**(6):2593–2605.
38. Greenberg JE, Peterson PM: Intelligibility-weighted measures of speech-to-interference and speech system performance. *J Acoust Soc Am* 1993, **94**(5):3009–3010.
39. Acoustical Society of American: *American national standard methods for calculation of the speech intelligibility index*; 1997. ANSI S3.5.
40. Chen JD, Benesty J, Huang YT: New insights into the noise reduction wiener filter. *IEEE Trans Audio Speech Language Process* 2006, **14**(4):1218–1234.
41. Benesty J, Chen JD, Huang YT: Binaural noise reduction in the time domain with a stereo setup. *IEEE Trans Audio Speech Language Process* 2011, **19**(8):2260–2272.
42. Chen JD, Benesty J, Huang Y: On the optimal linear filtering techniques for noise reduction. *Speech Commun* 2007, **49**:305–316.

doi:10.1186/1475-925X-11-74

Cite this article as: Chen and Gong: Real-time spectrum estimation-based dual-channel speech-enhancement algorithm for cochlear implant. *BioMedical Engineering OnLine* 2012 **11**:74.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

