

Single and Multiple Microphone Noise Reduction Strategies in Cochlear Implants

Trends in Amplification
16(2) 102–116
© The Author(s) 2012
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1084713812456906
http://tia.sagepub.com


Kostas Kokkinakis, PhD¹, Behnam Azimi², Yi Hu, PhD²,
and David R. Friedland, MD, PhD³

Abstract

To restore hearing sensation, cochlear implants deliver electrical pulses to the auditory nerve by relying on sophisticated signal processing algorithms that convert acoustic inputs to electrical stimuli. Although individuals fitted with cochlear implants perform well in quiet, in the presence of background noise, the speech intelligibility of cochlear implant listeners is more susceptible to background noise than that of normal hearing listeners. Traditionally, to increase performance in noise, single-microphone noise reduction strategies have been used. More recently, a number of approaches have suggested that speech intelligibility in noise can be improved further by making use of two or more microphones, instead. Processing strategies based on multiple microphones can better exploit the spatial diversity of speech and noise because such strategies rely mostly on spatial information about the relative position of competing sound sources. In this article, we identify and elucidate the most significant theoretical aspects that underpin single- and multi-microphone noise reduction strategies for cochlear implants. More analytically, we focus on strategies of both types that have been shown to be promising for use in current-generation implant devices. We present data from past and more recent studies, and furthermore we outline the direction that future research in the area of noise reduction for cochlear implants could follow.

Keywords

cochlear implants, single-microphone noise reduction, multi-microphone noise reduction

Introduction

It is well known that background noise interferes with and also reduces the intelligibility¹ of speech and that the greater the level of background noise the greater the reduction in intelligibility. In most cases, because speech is a highly redundant signal, we are capable of understanding speech even in a moderately noisy environment. Thus, even if parts of the speech signal are masked by noise, other parts of the speech signal will convey sufficient information to render the signal sufficiently intelligible. The idea that conventional speech enhancement strategies could be used to improve important perceptual aspects of speech corrupted by additive noise was introduced almost five decades ago by Schroeder (1965) who at the time was working at AT&T Bell Labs. Since then, pioneering research intensified with contributions from several other researchers, who systematically formulated the challenging problem of noise reduction and also compared different algorithms known at the time (e.g., see Boll, 1979; Ephraim & Malah, 1984; Ephraim & VanTrees, 1995; Lim, 1978; Lim & Oppenheim, 1979). Over the years, terms such as noise removal, noise suppression, noise reduction, and speech enhancement have all been used to describe techniques that improve the intelligibility of speech in the presence of background noise.

With the rapid progression of cochlear implants (CIs),² from a speculative laboratory procedure (House & Berliner, 1991) to a widely recognized practice and a potentially life-changing use of biotechnology (Gates & Miyamoto, 2003; Wilson & Dorman, 2008a), recent research efforts have been increasingly focusing on state-of-the-art noise reduction solutions to improve speech intelligibility in noisy environments. To restore hearing sensation, CIs deliver electrical pulses to the auditory nerve by relying on sophisticated signal processing algorithms that convert acoustic inputs to electrical stimuli (e.g., see Loizou, 1998; Wilson & Dorman, 2008a, 2008b; Zeng, 2004). Nowadays, almost all CI processors perform well in quiet environments, and the majority of implant users can achieve high open-set speech recognition scores regardless of the device or speech coding strategy used (e.g., see Skinner et al., 2002; Spahr & Dorman, 2004).

¹University of Kansas, Lawrence, KS, USA

²University of Wisconsin-Milwaukee, Milwaukee, WI, USA

³Medical College of Wisconsin, Milwaukee, WI, USA

Corresponding Author:

Kostas Kokkinakis, PhD, Department of Speech-Language-Hearing Sciences and Disorders, University of Kansas, 1000 Sunnyside Avenue, Lawrence, KS 66045, USA
Email: kokkinak@ku.edu

However, in the presence of background noise, the speech recognition of CI listeners is more susceptible to background noise than that of normal-hearing (NH) listeners. This is most likely because of the limited frequency, temporal, and amplitude resolution that can be transmitted by the implant device (Qin & Oxenham, 2003).

In fact, data comparing speech reception thresholds (SRTs)³ between NH and CI listeners suggest that in steady-state noise, CI subjects' scores are on average between 10 and 15 dB higher than SRTs obtained from normal-hearing subjects. The difference is more striking in a competing-talker background. In this case, CI listeners need, on average, a signal-to-noise ratio (SNR) equal to 25 dB or higher than NH listeners to perform at the same level in noisy backgrounds (Turner, Gantz, Vidal, Behrens, & Henry, 2004). Because, for typical speech materials, a 5 dB decrease in SNR implies between 20% and 30% reduction in the overall percent correct score, from the above data, it quickly becomes evident that even a small elevation in SRT scores can substantially reduce the speech perception of CI users in background noise. The difficulty recognizing speech in noise for implant users is, in fact a combination of poor frequency and temporal resolution and channel interaction (or current spread) in the stimulating electrodes. These factors have been shown to contribute strongly to CI users' difficulty in noisy conditions, regardless of the type of noise (e.g., steady-state or modulated) present (e.g., see Fu & Nogaki, 2004). Substantial differences in speech recognition performance have also been reported for CI recipients tested in noisy versus quiet settings.

In a recent study by Firszt et al. (2004), speech recognition was assessed using the Hearing in Noise Test (HINT) sentences (Nilsson, Soli, & Sullivan, 1994). Results revealed that CI recipients' performance on sentence recognition tasks was significantly poorer in noise compared when compared with just listening at a soft conversational level in quiet. An average of 30% speech intelligibility drop was found between sentences presented at 60 dB SPL in quiet versus 60 dB SPL in the presence of noise at the SNR level of 8 dB. Still, the observed reduction in performance was only about 15% when sentences were presented at 60 and 50 dB SPL in quiet. In another study, Spahr and Dorman (2004) 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% during tasks involving conversational speech. After the SNR level was lowered to 5 dB, recognition of conversational speech, on average, dropped to around 20%. The study by Fetterman and Domico (2002) revealed a similar trend when individuals fitted with Clarion and Nucleus devices were asked to identify speech in noise. On average, CI recipients' sentence recognition scores decreased from 82% correct in quiet to 73% at a 10 dB SNR level and to around 47% at 5 dB SNR.

Noise reduction as an additional device feature has been available in hearing aids since the late 1970s (e.g., see

Bentler & Chiou, 2006). However, it was not until the mid-1990s and shortly after the first digital hearing aids appeared in the United States that some early noise reduction schemes for use in CIs were proposed (e.g., see Hochberg, Boothroyd, Weiss, & Hellman, 1992; Weiss, 1993). In general, noise reduction algorithms, considered in studies with CI listeners, fall into one of two classes: single-microphone and multi-microphone methods. Single-microphone approaches rely mostly on statistical models of speech and noise, and therefore can only differentiate between signals that have different temporal and spectral characteristics (e.g., see Hu & Loizou, 2002; Hu, Loizou, Li, & Kasturi, 2007; Loizou, Lobo, & Hu, 2005; Yang & Fu, 2005). Single-microphone noise reduction strategies have received the greatest attention, presumably because traditionally most CI devices and their sound processors have typically used only one microphone. Improving speech intelligibility in noise using just a single microphone is difficult, however, because single-channel noise reduction methods may boost the SNR level but inevitably they may also incur distortion that could degrade speech. Attaining this balance between speech distortion and noise reduction is not always feasible.

Large improvements in SNR and therefore considerable benefits in speech intelligibility can be obtained when resorting to multi-microphone noise reduction strategies, instead. In recent years, there has been a growing tendency toward the use of noise reduction methods that exploit multiple microphones in digital hearing aids and CI devices (e.g., see Chung, 2004; Chung, Zeng, & Acker, 2006; Greenberg & Zurek, 1992; Kompis & Dillier, 1994; Kokkinakis & Loizou, 2008, 2010; Levitt, 2001; Spriet et al., 2007; van Hoesel & Clark, 1995; Wouters & Van den Berghe, 2001). Because in most listening scenarios, the target speech and the noise sources are located at different locations (positions) in space, such techniques can better exploit the spatial diversity of speech and noise, in addition to their spectral and temporal differences. In applications affording use of spatially separated inputs, multi-microphone noise reduction strategies can rely on the correlation between signals received at spatially diverse sensors to identify noise and subtract it from the noisy speech.

In this review article, we identify and elucidate the most significant theoretical aspects that underpin single microphone and multiple microphone noise reduction strategies for CIs. More analytically, we focus on strategies of both types that have been shown to be promising for use in current-generation implant devices. We present data from past and more recent studies, and furthermore we outline the direction that future research in the area of noise reduction for CIs could follow.

Single-Microphone Noise Reduction

There are two types of noise reduction algorithms suitable for sound processing strategies in CI devices: one is based on preprocessing the noisy acoustic signals on the front-end

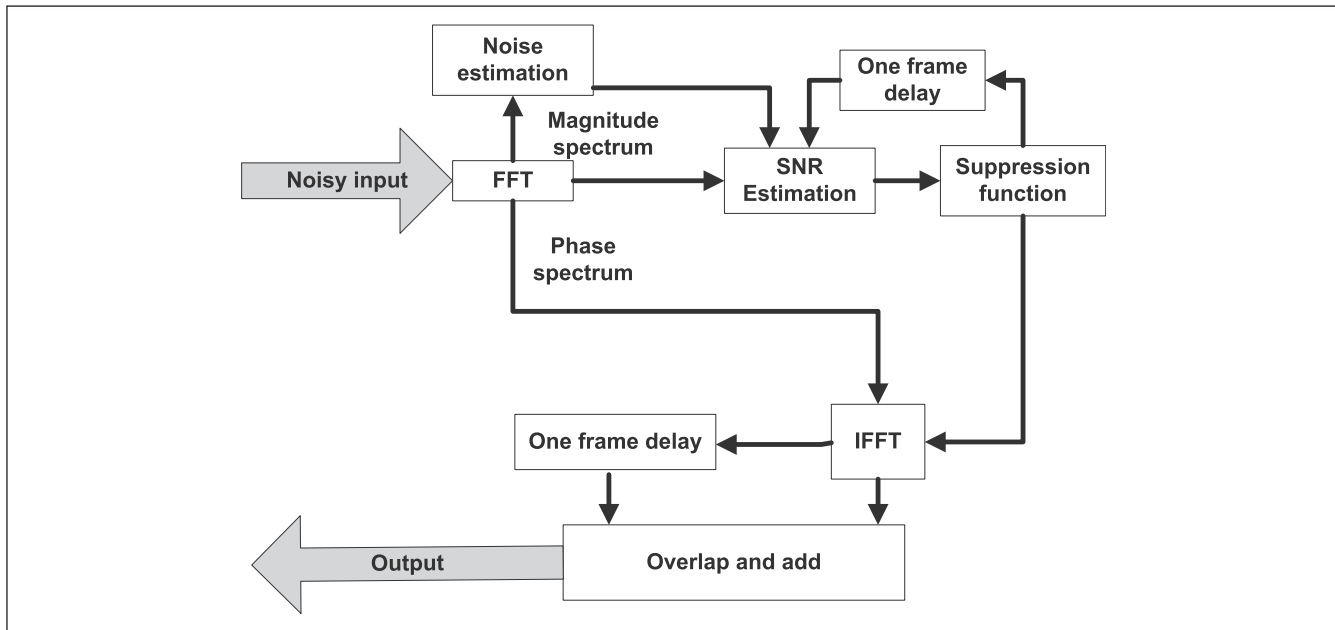


Figure 1. Block diagram of single-microphone noise reduction methods based on preprocessing the noisy acoustic input signals

side located before the radio-frequency (RF) link, which transmits the audio stream to the internally implanted receiver. This approach is similar to speech enhancement currently used in most modern communication devices (e.g., cellphones). The other type of single-microphone noise reduction strategies is based on applying some form of attenuation directly on the noisy electrical envelopes.

Noise Reduction on Noisy Acoustic Inputs

Hochberg et al. (1992) and Weiss (1993) were the first to assess the effect of using a digital single-channel noise suppression algorithm (INTEL) to acoustically process noisy speech inputs in CIs. The enhanced acoustic stimuli obtained with the INTEL strategy were presented to normal-hearing individuals and also to users of the Nucleus 22 implant device. In both populations, INTEL was shown to reduce noise consistently over the entire range of the different input signal-to-noise ratios tested. In addition, results with consonant-vowel-consonant (CVC) words presented in speech-shaped random noise indicated that this noise reduction strategy could improve the phoneme recognition threshold of the implant group significantly. Much later, Yang and Fu (2005) found significant improvements to sentence recognition in stationary speech-shaped noise at different SNRs in a group of seven CI users that were fitted with different CI devices. Figure 1 describes the general block diagram for noise reduction algorithms implemented in the frequency domain.

At the same time, Loizou et al. (2005) investigated the potential benefits of first preprocessing the noisy acoustic input with a custom subspace-based noise reduction

algorithm. The subspace algorithm was originally developed for suppressing white input noise by Ephraim and Van Trees (1995) but was later extended to handle colored noise (e.g., speech-shaped noise) by Hu and Loizou (2002). The underlying principle of the subspace algorithm is based on the projection of the noisy speech vector (i.e., consisting of a segment of speech) onto two subspaces: the “noise” subspace and the “signal” subspace (Ephraim & Van Trees, 1995). The noise subspace contains only signal components because of the noise, and the signal subspace contains primarily the clean signal. Therefore, an initial estimate of the clean signal can be estimated by removing the components of the signal in the noise subspace and retaining only the components of the signal in the signal subspace. By doing this, noise can be, in most cases, suppressed to a large extent.

Loizou et al. (2005) summarized the mathematical treatment for the subspace-based noise reduction algorithm as follows. Let \mathbf{y} be the noisy speech vector, and let $\bar{\mathbf{x}} = \mathbf{H} \times \mathbf{y}$ be the estimated clean speech signal vector, where \mathbf{H} is the transformation matrix for the estimation of clean speech signal. In essence, the noise reduction problem is formulated as that of finding a transformation matrix \mathbf{H} , which would yield the estimated clean signal when applied to the noisy vector. After applying such a transformation to the noisy signal, the error between the estimated signal vector $\bar{\mathbf{x}}$ and the true clean signal vector \mathbf{x} is expressed as: $\boldsymbol{\varepsilon} = \bar{\mathbf{x}} - \mathbf{x} = (\mathbf{H} - \mathbf{I}) \times \mathbf{x} + (\mathbf{H} \times \mathbf{n})$, where \mathbf{n} is the noise vector. Because the estimated transformation matrix will not be perfect, it will introduce some speech distortion, which is quantified by the first error term $(\mathbf{H} - \mathbf{I}) \times \mathbf{x}$. In addition, the second error term $(\mathbf{H} \times \mathbf{n})$ will quantify the amount of noise distortion introduced by the transformation matrix. Because the speech and noise

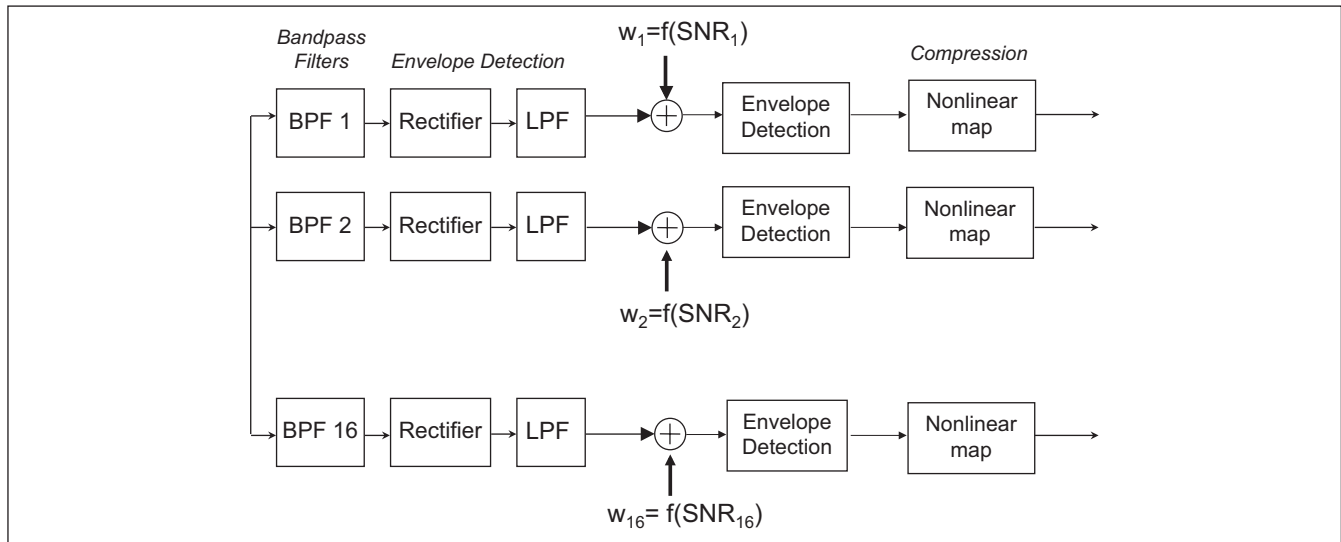


Figure 2. Block diagram of single-microphone noise reduction methods based on attenuating the noisy electrical envelopes through envelope-weighting

distortion (as defined earlier) are decoupled, one can estimate the optimal transformation matrix \mathbf{H} that would minimize the speech distortion subject to the noise distortion falling below a preset threshold. The solution to this constrained minimization problem for the transformation matrix is presented more analytically in Hu and Loizou (2002).

The authors in Loizou et al. (2005) tested this subspace-based noise reduction algorithm using HINT sentences (Nilsson et al., 1994) corrupted in speech-shaped noise at a 5 dB SNR level. Speech intelligibility scores were obtained with 14 Clarion CI recipients and results were compared against the users' standard sound processing strategy, either the continuous interleaved stimulation (CIS) strategy or the simultaneous analog stimulation (SAS) strategy. These findings indicated that when switching to the subspace-based single-microphone noise reduction strategy almost all subjects benefited from the noise reduction strategy and managed to raise their speech perception scores significantly. These findings are presented more analytically in "Noise Reduction on Noisy Electrical Envelopes."

Noise Reduction on Noisy Electrical Envelopes

The practice of preprocessing noisy acoustic inputs with a suitable noise reduction strategy has yielded modest but statistically significant results. However, such approaches to noise reduction have three main drawbacks: (1) preprocessing algorithms often introduce unwanted acoustic distortion in the signal, (2) some algorithms (e.g., subspace algorithms) are computationally complex (and consequently power hungry) and fail to integrate well with existing CI strategies, and (3) it is not straightforward to always fine tune (or optimize) the operation of a particular algorithm to individual

users. Ideally, noise reduction algorithms should be easy to implement and be integrated into existing coding strategies. From a computational standpoint, the simplest way to overcome these issues is to directly apply attenuation to the electrical envelopes according to the intensity contrast between the speech signal and noise signal. By transferring the energy in frequency channels directly to electric outputs we can reduce the limitations normally associated with acoustic waveform reconstruction.

Envelope-Weighting. Such a strategy that can efficiently suppress noise by attenuating electrical envelopes, while avoiding the intermediate acoustic waveform reconstruction stage is described in Figure 2. This practical algorithm can directly tackle noisy electrical envelopes and can therefore be easily integrated in existing strategies used in commercially available implant devices (Hu et al., 2007). The proposed algorithm is based on envelope-weighting of each spectral channel and therefore fits into the general category of algorithms that perform noise suppression by spectral modification (e.g., spectral subtraction, Wiener filtering). In this algorithm, the enhanced signal envelopes are obtained by applying a weight (taking values in the range 0-1) to the noisy envelopes of each channel. The weights are chosen to be inversely proportional to the estimated SNR of each channel. Envelope amplitudes in channels with a high SNR are multiplied by a weight close to one (i.e., left unaltered), whereas envelope amplitudes in spectral channels with a low SNR level are multiplied by a weight close to zero (i.e., heavily attenuated).

The underlying assumption is that channels with low SNR are heavily masked by noise and therefore contribute little, if any, information about the speech signal. As such, these low-SNR channels are heavily attenuated (or annihilated) leaving only the high-SNR channels, which are likely to contribute more useful information to the listener. With

that in mind, we resort to weighting functions that apply heavy attenuation in channels with low SNR and at the same time little or no attenuation in channels with high SNR levels (e.g., see Hu et al., 2007). An appropriate choice⁴ is the sigmoidal-shaped function $g(i, \ell) = \exp(-b/\text{SNR}(i, \ell))$, where $b = 2$, $g(i, \ell)$ denotes the weighting function ($0 < g(i, \ell) < 1$), and $\text{SNR}(i, \ell)$ denotes the estimated instantaneous SNR in the i th channel and at stimulation cycle ℓ . Following the weighting function computation, the enhanced temporal envelope can be subsequently obtained by using $s(i, \ell) = g(i, \ell) \times y(i, \ell)$, where $s(i, \ell)$ represents the enhanced signal and $y(i, \ell)$ denotes the noisy envelope of the i th channel at stimulation cycle ℓ . As shown in Hu et al. (2007), the proposed sigmoidal-shaped function is capable of providing significant benefits to CI users' speech perception in noise, and much of the success of this algorithm can be attributed to the improved temporal envelope contrast. The sigmoidal function preserves the envelope peaks and also deepens the envelope valleys thereby increasing the effective envelope dynamic range within each channel.

Envelope-Selection. Although the noise suppression methods discussed earlier have shown promising results, there still exists a substantial performance gap between CI listeners' speech recognition in noisy listening conditions and in quiet. This can be attributed mainly to the fact that it is not possible to accurately estimate the SNR at each frequency bin or each spectral channel. To circumvent this problem, Hu and Loizou (2008) proposed a noise reduction strategy that can operate under the assumption that the true SNR values in each spectral channel are known *a priori*. Because the true SNR values are assumed to be known, each frequency channel (or equivalent envelope) is selected only if its corresponding SNR is larger than or equal to 0 dB. In a similar vein, a channel with SNR level which is smaller than 0 dB is discarded.

The main idea is that spectral channels with low SNR values (e.g., $\text{SNR} < 0$ dB) contain mainly masker-dominated envelopes and therefore contribute little, if any, information about the speech signal. However, channels with high SNR levels (e.g., $\text{SNR} \geq 0$ dB) contain target-dominated envelopes and can be retained as they contain reliable information about the target input. The SNR envelope-selection criterion can be implemented simply by multiplying the noisy signal by a binary time-frequency (T - F) mask or equivalently a binary gain function to the electrical envelopes of the noisy stimuli. This approach was recently tested in CI listeners and the authors showed that the SNR channel-selection criterion is capable of restoring the speech intelligibility in noise for CI listeners to the level attained in quiet even at extremely low input SNR levels (e.g., -10 dB). For this reason, this strategy is referred to as the optimal ACE (opACE; see Hu & Loizou, 2008). It is also worth mentioning that different forms of such binary time-frequency (T - F) masks and other similar local channel-selection criteria have been previously applied with much success to noise reduction for normal-hearing individuals

(Brungart, Chang, Simpson, & Wang, 2006; Li & Loizou, 2008) and reverberation suppression for CI listeners (Kokkinakis, Hazrati, & Loizou, 2011).

This envelope-selection strategy is promising; nonetheless, its implementation poses a considerable challenge in real-world applications. This is because of the fact that the SNR values for each spectral channel need to be estimated from the mixture envelopes (corrupted speech), which is a formidable task. In fact, Hu et al. (2007) showed that most conventional noise estimation algorithms perform poorly in estimating the SNR. These algorithms are derived by minimizing a certain optimization criterion (e.g., mean-square error), and they are expected to perform well in all noisy environments, which is clearly an ambitious goal. In other words, conventional noise estimation algorithms are not optimized for a particular listening situation, and thus do not take into account the differences in temporal and spectral characteristics of real-world maskers.

In the study by Hu and Loizou (2010), an algorithm was designed to select channels for stimulation based on estimated SNRs in each spectral channel. Rather than relying on knowledge of the local SNR, the authors proposed to take advantage of the distinctive temporal and spectral characteristics of different real-world maskers, which can be learned by resorting to machine-learning techniques. The proposed noise reduction algorithm uses Gaussian mixture models (GMMs) to learn how to use the distinctive temporal and spectral characteristics of the different maskers in real world. The proposed algorithm consists of two steps: (1) a training stage and (2) a speech intelligibility enhancement stage. In the training stage, with access to the temporal envelopes of the speech signals (typically from a large corpus) and the masker signals, the SNRs are computed for each spectral channel, and the binary status of the channels being speech dominated (with a binary gain of 1) or being masker dominated (with a binary gain of 0) is determined.

In the next step, the features extracted from the noisy mixture temporal envelopes, and the corresponding binary gains are used to train for each spectral channel two GMMs representing two corresponding feature classes: target-dominated and masker-dominated. Note that in this stage features similar to amplitude modulation spectrograms (AMS) can be used (e.g., see Kollmeier & Koch, 1994; Tchorz & Kollmeier, 2003). In the enhancement stage, a Bayesian classifier is used to classify the spectral channel into two classes using the extracted features: target-dominated and masker-dominated. The appropriate electrode is then selected for stimulation only when the spectral channel is target-dominated. Figure 3 depicts the block diagram of noise reduction methods based on envelope-selection as described in Hu and Loizou (2010).

Clinical Verification. The target speech stimuli used for testing were sentences from the IEEE database (IEEE, 1969). Each sentence is composed of approximately 7 to 12 words and in total there are 72 lists of 10 sentences each produced

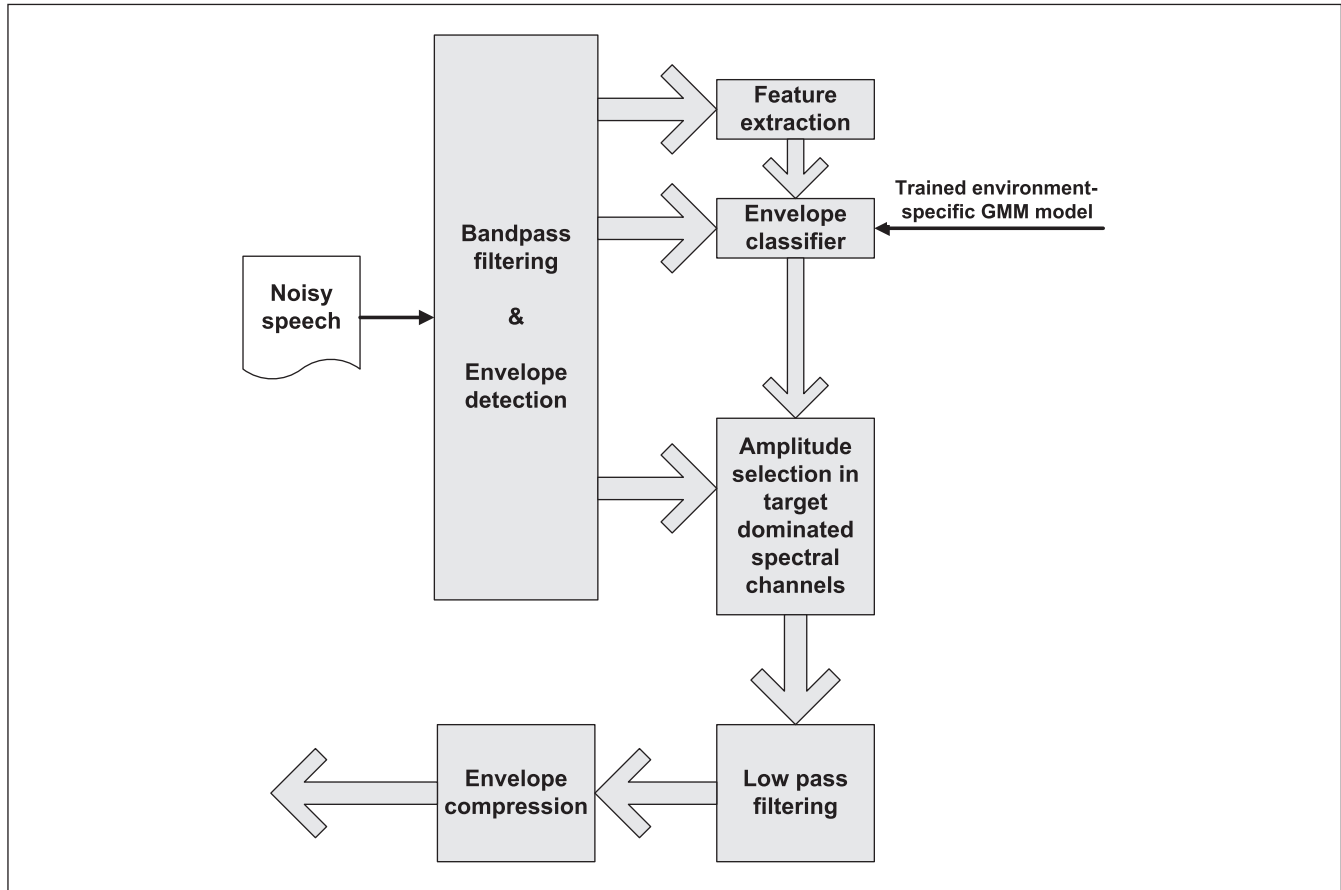


Figure 3. Block diagram of single-microphone noise reduction methods based on attenuating the noisy electrical envelopes through envelope-selection

by a single talker. The root-mean-square amplitude of all sentences was equalized to the same root-mean-square value approximately 65 dBA.

Figure 4 shows subjects' performance on identification of words in sentences corrupted by 5 dB speech-shaped noise (preprocessed by the subspace algorithm or left unprocessed). The mean score obtained using sentences preprocessed by the subspace algorithm was equal to around 44%, while the standard error of mean was 6.2%. On the contrary, the mean score obtained using unprocessed sentences was 19% and the standard error was 6.6% (Loizou et al., 2005). The sentence scores obtained with the subspace algorithm were significantly higher than the scores obtained with the unprocessed sentences [$F(1,13) = 33.1, p < 0.0005$]. As it can be seen from Figure 4, most subjects benefited from the noise reduction algorithm. Subject SS4's score, for instance, improved from 0% correct to 40% correct. Similarly, subjects SS1 and SS2 scores improved from roughly 0% to 50% correct. The above results indicate that the subspace algorithm can provide significant benefits to CI users in sentence recognition in noise.

Figure 5 shows subjects' performance on identification of words in sentences embedded in 5 dB SNR multitalker

babble and processed by the proposed algorithm that applied signal-to-noise-ratio-based weighting to electric envelopes. The mean score obtained using sentences preprocessed by the proposed SNR-weighting algorithm was approximately 39%, whereas the standard error of mean was equal to 4.7%. The mean score obtained using unprocessed (noisy) sentences was equal to almost 29% and the standard error was 5.2%. The difference in scores was significant ($F(1,7) = 24.97, p = 0.002$). As it can be seen from Figure 5, nearly all subjects (except S5) benefited from the proposed noise reduction algorithm.

Figure 6 shows subjects' performance on identification of words in sentences corrupted by babble noise at 5 dB SNR level using environment-optimized noise reduction based on GMMs. Performance was measured in terms of percent of words identified correctly (all words were scored). *Post hoc* tests (Scheffe, corrected for multiple comparisons) were run to assess the statistical significance between different conditions. For all the noise conditions, the opACE strategy was significantly better than the GMM-based noise reduction algorithm. The highest performance was obtained with the opACE strategy. For all the noise conditions, performance

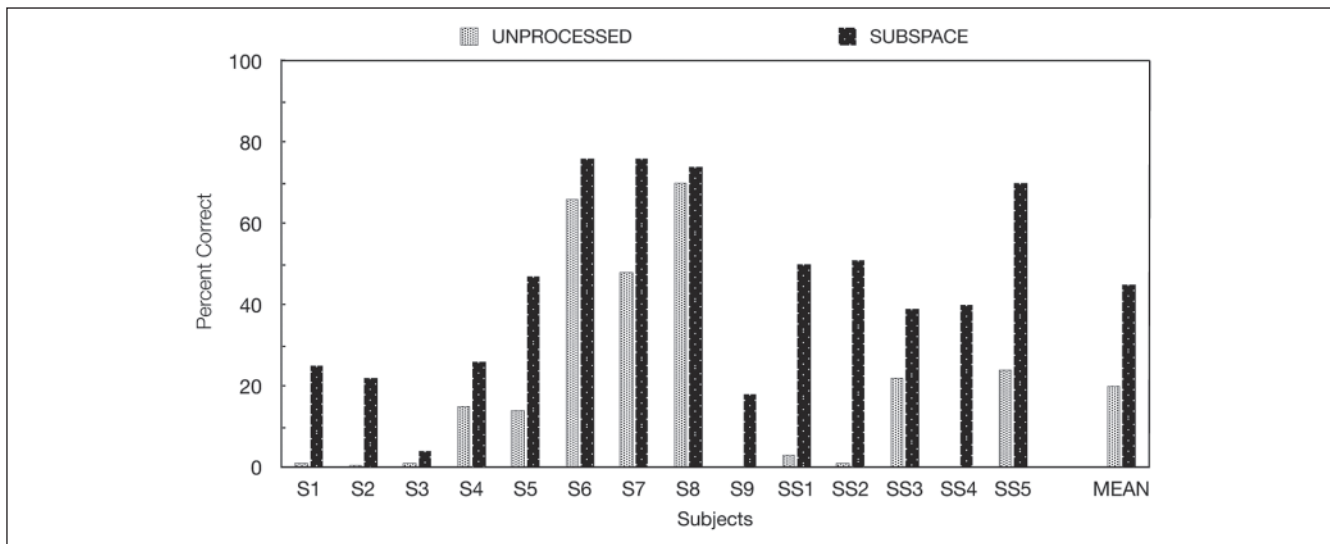


Figure 4. Subjects’ performance on identification of words in sentences embedded in 5 dB SNR speech-shaped noise and preprocessed by the subspace algorithm or left unprocessed. Subjects S1-S9 were Clarion CII patients and subjects SS1-SS5 were Clarion S-Series patients.

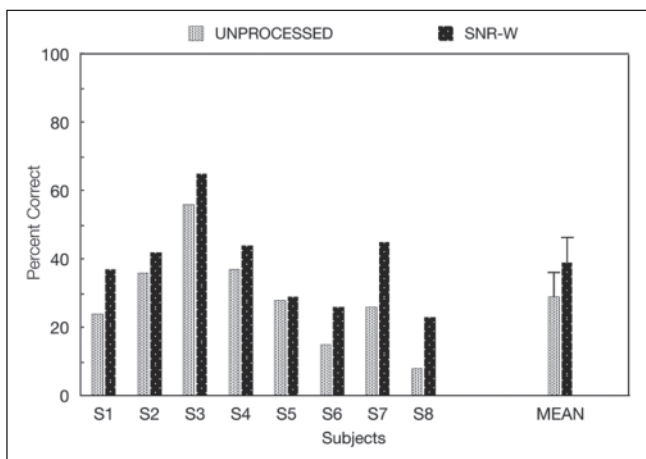


Figure 5. Subjects’ performance on identification of words in sentences embedded in 5 dB SNR multitalker babble (CIS+N) and processed by the proposed algorithm (SNR-W). Error bars indicate standard errors of the mean.

with the unprocessed sentences (UN) was significantly lower than both the GMM-based noise reduction algorithm (GMM) and the opACE strategy.

Multi-Microphone Noise Reduction

Nowadays, an increasing number of profoundly deaf individuals are being fitted bilaterally with bilateral CI (BCI) systems. Numerous studies conducted with BCI listeners suggest that bilateral cochlear implantation can improve speech intelligibility in noise by up to 3 to 5 dB (e.g., see Tyler et al., 2002; Litovsky et al., 2004; Loizou et al., 2009). Most bilat-

eral CIs are fitted with either two microphones in each ear or one microphone in each of the two (one per ear) behind-the-ear (BTE) processors. The Nucleus Freedom processor, for instance, employs a rear omnidirectional microphone, which is equally sensitive to sounds from all directions, as well as an additional directional microphone pointing forward.⁵ In a directional microphone, sound waves originating from directly behind the listener are mechanically delayed through the use of a small screen (diaphragm), whereas sounds originating from the front are allowed to reach the port of the microphone unimpeded. Resorting to directional microphones provides an effective yet simple form of spatial processing.

In fact, a number of recent studies have shown that the overall benefit simply by adding a directional microphone in the prosthetic device can be about 3 dB within low reverberation settings when compared to processing with just an omnidirectional microphone (Chung et al., 2006; Wouters & Van den Berghe, 2001). The same effect of a considerably improved spatial directivity can be replicated with two omnidirectional microphones.⁶ In this case, in lieu of a mechanical diaphragm, a simple filter can be used to delay the signal entering the rear microphone. In both paradigms, the desired effect is that sounds from the front are left unimpeded, whereas sounds originating from directly behind the microphone are completely canceled. This form of spatial processing is particularly useful in typical noisy backgrounds where the target speaker is always assumed to be in the front of the listener and noise sources are spatially distributed in one or both hemifields.

The benefit in noisy backgrounds because of better spatial directivity can be increased even further by resorting to adaptive beamformers. Adaptive beamformers can somewhat efficiently optimize the microphone inputs by steering nulls

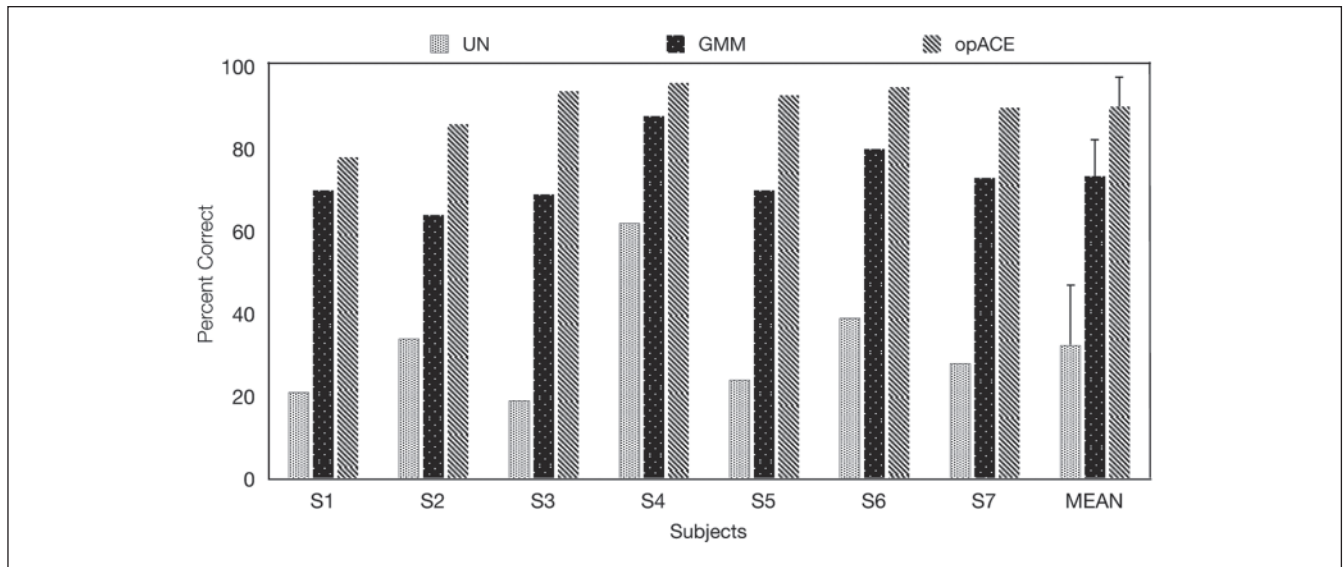


Figure 6. Mean percent correct scores for babble noise at 5 dB SNR level. The error bars denote ± 1 standard error of the mean. UN represents the condition for the unprocessed noisy speech, GMM represents the proposed GMM-based noise reduction algorithm, and opACE represents the opACE processing strategy.

toward sources of interference. Often this is done by adjusting the beamformer filter weights to minimize the output power subject to certain constraints for the desired direction, namely the “look” direction. Adaptive beamformers are extensions of differential microphone arrays, where the suppression of sources of interference is carried out by adaptive filtering of the signals at the microphones. A computationally attractive realization of an adaptive beamformer is the generalized sidelobe canceller (GSC) structure (Griffiths & Jim, 1982).

To evaluate the benefits of resorting to a multiple microphone noise reduction for CI listeners, Spriet et al. (2007) investigated the performance of the BEAM strategy in the Nucleus Freedom speech processor with five CI users. The performance with the BEAM strategy was evaluated at two noise levels and with two types of noise, speech-shaped noise and multitalker babble. On average, the algorithm tested lowered the SRT by approximately 5 to 8 dB, as opposed to just using a single directional microphone to increase the direction-dependent gain of the target source. Such spatial processing strategies, as those found in the latest CI sound processors, including traditional directional microphones and adaptive beamformers, have also been implemented in several of the latest hearing aid devices. The observed benefit of these strategies in hearing aids ranges from about 4 dB to almost 12 dB improvements in SNR. For instance, an increased SNR due to spatial processing has been shown to correspond to an improvement on speech recognition tasks in noise of about 40% to almost 70% (e.g., see Greenberg & Zurek, 1992; Kompis & Dillier, 1994; Levitt, 2001; Ricketts, 2001, 2005). At the same time, although significant, the observed improvement in speech intelligibility will depend largely on the spatial location or configuration of noise and target sources, as well as the amount of

reverberant energy in the room. For instance, it has been shown that the performance of noise reduction strategies based on adaptive beamforming drops significantly even in moderately reverberant surroundings (e.g., see Greenberg & Zurek 1992; Kokkinakis & Loizou, 2007).

One popular solution to circumvent this problem, especially in reverberant listening situations, is to increase the number of microphones that are made available in the device. In the following sections, we describe noise reduction strategies based on two and four microphones. Although, CIs do not operate on synchronized processors, and therefore there is no communication between the two sides so that auditory information can be processed in a coordinated mode, in our paradigm we assume that the CI devices in the left and right sides of the listener are synchronized such that auditory streams from both sides can be captured synchronously and processed together. Multi-microphone strategies can be classified into two different types, based on the total number of microphones available: (1) two-microphone bilateral, whereby we assume access to a total of two directional microphones (left and right) with each microphone placed on opposite sides of the head (Kokkinakis & Loizou, 2008) and (2) four-microphone bilateral, whereby we assume access to a total of four microphones, namely two-omnidirectional (left and right) and two-directional microphones (left and right) with each set of directional and omnidirectional microphones placed on opposite sides of the head (Kokkinakis & Loizou, 2010).

Noise Reduction With Two Microphones

Two-Microphone Beamforming Strategy. The 2M2-BEAM noise reduction strategy utilizes the two-microphone BEAM running in an independent electrical stimulation mode. In

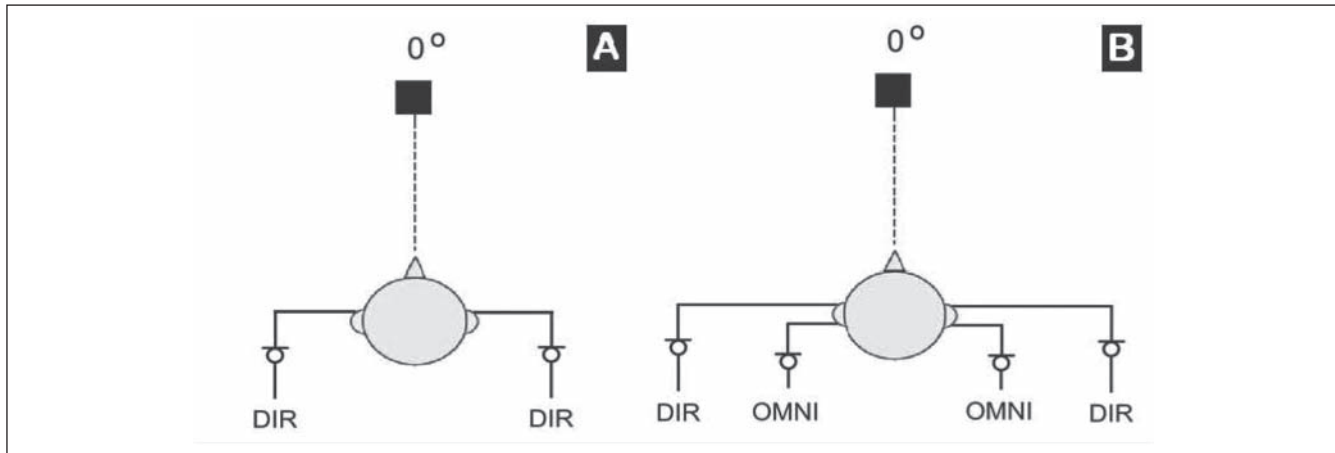


Figure 7. (a) Two-microphone (one per side) bilateral and (b) four-microphone (two per side) bilateral configuration

this stimulation mode, the left and right processors are unsynchronized. The 2M2-BEAM employs two (one per ear) BTE units furnished with one-directional and one-omnidirectional microphone each and is therefore capable of delivering the signals bilaterally to the CI user. In this paradigm, each BEAM combines a directional microphone with an extra omnidirectional microphone placed closed together in an end-fire array configuration to form the target and noise references. The inter-microphone distance is usually fixed at 8 mm (Patrick, Busby, & Gibson, 2006).

In the BEAM strategy,⁷ the first stage utilizes spatial preprocessing through an adaptive two-microphone system that combines the front directional microphone and a rear omnidirectional microphone to separate speech from noise. The output from the rear omnidirectional microphone is filtered through a fixed finite impulse response (FIR) filter. The output of the FIR filter is then subtracted from an electronically delayed version of the output from the front directional microphone to create the noise reference (Spriet et al., 2007; Wouters & Van den Berghe, 2001). The filtered signal from the omnidirectional microphone is then added to the delayed signal from the directional microphone to create the speech reference. This spatial preprocessing increases sensitivity to sounds arriving from the front while suppressing sounds that arrive from the sides. The two signals with the speech and noise reference are then fed to an adaptive filter, which is updated with the normalized least-mean-squares (NLMS) algorithm in such a way as to minimize the power of the output error (Greenberg & Zurek, 1992). The 2M2-BEAM strategy is currently implemented in commercially available bilateral CI processors, such as the Nucleus Freedom and the new Nucleus 5 sound processor (CP810).

Two-Microphone Spatial Separation Strategy. The two-microphone spatial enhancement via source separation (2M-SESS) strategy can be classified as a two-microphone binaural strategy, because it relies on having access to a total of two directional microphones (left and right) with each microphone placed on opposite sides of the head. This configuration is

illustrated in Figure 7(a). Here, we assume that the CI devices in the left and right sides can be synchronized such that auditory streams from both sides can be captured synchronously and processed together. As outlined in Figure 8, the 2M-SESS strategy operates by estimating a total of four adaptive linear filters that can undo the mixing effect by which two composite signals are created when the target and noise sources propagate inside a natural acoustic environment. The 2M-SESS can spatially separate and further suppress noise and hence the CI user can maximize speech intelligibility by focusing only on the extracted target source. Although in principle one can allow the CI user to select which enhanced signal output to listen to, the 2M-SESS strategy is implemented so that it can reject the interferer and deliver only the recovered speech waveform diotically to the listener.

To adaptively estimate the unmixing (or separating) filters, the 2M-SESS strategy employs the frequency-domain natural gradient algorithm (FD-NGA) described in Kokkinakis and Nandi (2006). The 2M-SESS strategy runs in an adaptive offline mode and relies on a multipass processing scheme. Therefore, the filter estimation is performed iteratively over a block of data and the estimates obtained in the last iteration are then used to process the same data blocks. That is, the same blocks of data are reused. This is a strategy typically employed in adaptive algorithms to obtain sufficiently good estimates of the filters. This scheme is not amenable to real-time implementation, and thus it can be used only when large amounts of training data are available.

Noise Reduction With More Than Two Microphones

Four-Microphone Spatial Separation Strategy. The four-microphone spatial enhancement via source separation (4M-SESS) processing strategy is the two-stage multi-microphone extension of the conventional BEAM noise reduction approach. It is based on co-ordinated or cross-side stimulation by collectively using information available on both the left and

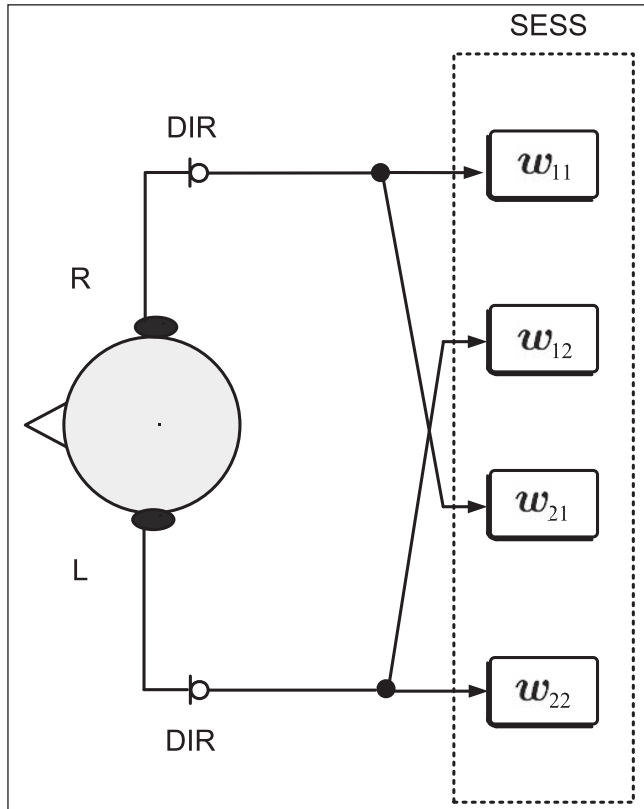


Figure 8. Block diagram of the 2M-SESS noise reduction processing strategy

right sides. The 4M-SESS can be classified as a four-microphone binaural strategy, because it relies on having access to four microphones, namely two omnidirectional (left and right) and two directional microphones (left and right) with each set of directional and omnidirectional microphones placed on opposite sides of the head. This particular arrangement is depicted in Figure 7(b).

As outlined schematically in Figure 9, in the 4M-SESS binaural processing strategy, the speech reference on the right side is formed by adding the input to the left omnidirectional microphone to the delayed version of the right directional microphone signal and the noise reference on the right is estimated by subtracting the left omnidirectional microphone signal from a delayed version of the right directional microphone signal. In a similar manner, to create the speech reference signal on the left side, the signals from the left directional microphone and right omnidirectional microphone are summed together. The noise reference on the left side is formed by subtracting the right omnidirectional microphone signal from a delayed version of the left directional microphone signal. Now by assuming that the noise source is placed on the right of the listener, this procedure leads to a signal with an amplified noise level on the right side but also yields an output with a substantially reduced noise level in the left ear. After processing the microphone signals containing two speech and two noise reference signals binaurally with one BEAM processor per ear, the two

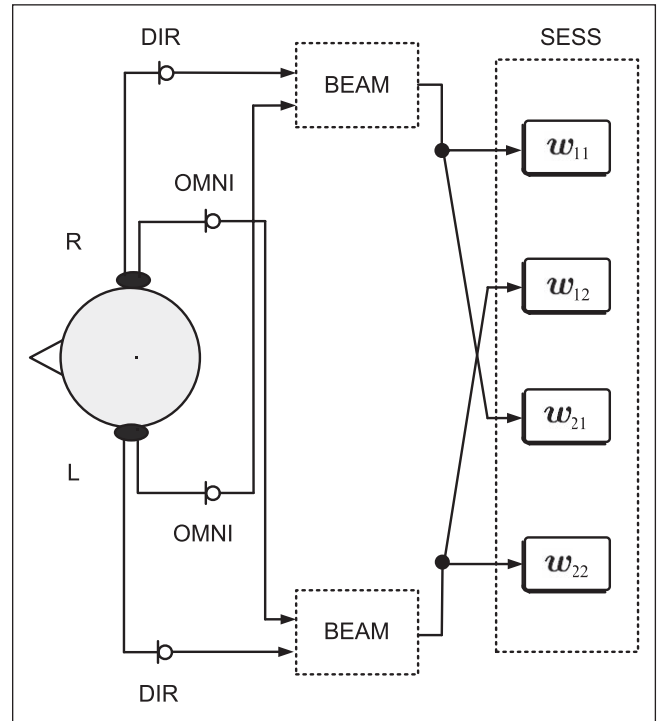


Figure 9. Block diagram of the 4M-SESS noise reduction processing strategy

microphone outputs from the two BEAM processors containing the generated speech and noise reference signals are enhanced further by employing a BSS strategy.

The 4M-SESS strategy is the four-microphone online extension of the two-microphone algorithm in Kokkinakis and Loizou (2008) and unlike the 2M-SESS strategy is amenable to real-time implementation. The 4M-SESS strategy operates by estimating a total of four FIR filters that can undo the mixing effect by which two composite signals are generated when the target and noise sources propagate inside an acoustic environment. The filters are computed after only a single pass with no additional training. The 4M-SESS strategy operates on the premise that the target and noise source signatures are spatially separated and thus their individual form can be retrieved by minimizing the statistical dependence between them.

In statistical signal processing theory, this configuration is referred to as a fully determined system, where the number of independent sound sources is equal to the number of microphones available for processing. Initializing the filters used in the 4M-SESS strategy with those obtained with each of the two BEAMs results in a substantial reduction in the total number of filter coefficients required for adequate interference rejection and substantially speeds up the convergence of the algorithm. Similarly to the 2M-SESS strategy, the implementation of the 4M-SESS strategy requires access to a single processor driving two CIs, such that signals from the left and right sides are captured synchronously and processed together.

Clinical Verification. The target speech stimuli used for testing were sentences from the IEEE database (IEEE, 1969). Each sentence is composed of approximately 7 to 12 words and in total there are 72 lists of 10 sentences each produced by a single talker. The root-mean-square amplitude of all sentences was equalized to the same root-mean-square value approximately 65 dBA. Every sentence in the IEEE speech corpus that was produced by a male talker was designated as the target speech. The simulated target location was always placed directly in front of the listener at a 0° azimuth. Subjects were tested in conditions with either one or three interferers. In the single interferer conditions a single speech-shaped noise source was presented from the right side of the listener (+90°). In the condition where multiple interferers were present, three interfering noise sources were placed on the right side only (30°, 60°, 90°).

The noisy stimuli in the case of a single interferer were processed with the following stimulation strategies: (1) unilateral presentation using the unprocessed input to the directional microphone on the side ipsilateral to the noise source, (2) bilateral stimulation using the unprocessed inputs from the two-directional microphones, (3) bilateral stimulation plus noise reduction using the 2M2-BEAM strategy, and (4) diotic stimulation plus noise reduction using the 4M-SESS processing strategy. The noisy stimuli generated when three interferers were present were processed with the following processing strategies: (1) bilateral stimulation using the unprocessed inputs from the two directional microphones, (2) bilateral stimulation plus noise reduction using the 2M2-BEAM strategy, and (3) diotic stimulation plus noise reduction using the 4M-SESS processing strategy.

Speech intelligibility scores obtained with the 2M2-BEAM and 4M-SESS multi-microphone noise reduction strategies are shown in Figure 10 (one interferer) and Figure 11 (three interferers). As shown in Figure 10, when the single noise source arrives from the side ipsilateral to the unilateral CI, speech understanding in noise was quite low for all subjects when presented with the input to the fixed directional microphone. Although one would expect that processing with a directional microphone alone would result in a more substantial noise suppression this is not the case here. In contrast, improvements in speech understanding were obtained when listeners were relying on using both CI devices. When compared to the unprocessed unilateral condition, the mean improvement observed was 14 percentage points due to processing when the SNR was fixed at 5 dB in the single interferer scenario.

As expected, the bilateral benefit was smaller when multiple noise sources were present (see Figure 11). In contrast, the 4M-SESS strategy yielded a considerable benefit in all conditions tested and for all five subjects. As evidenced from the scores plotted in Figures 10 and 11, the observed benefit over the subjects' daily strategy ranged from 30 percentage points when multiple interferers were present to around 50 percentage points for the case where only a single interferer emanated from a single location in

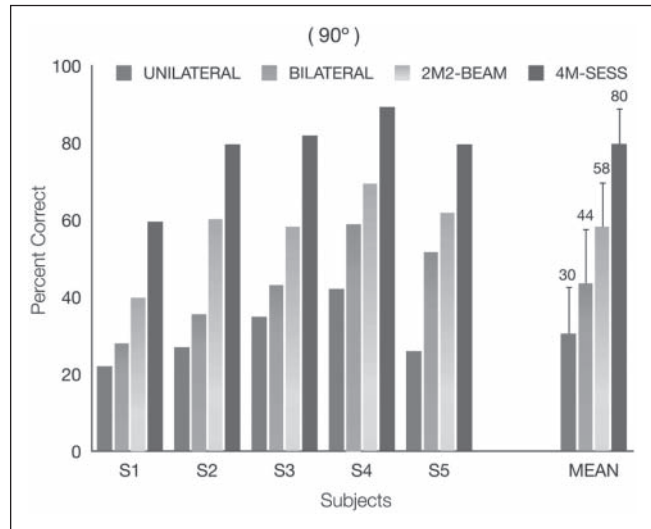


Figure 10. Percent correct scores by five Nucleus 24 users using both CI devices tested on IEEE sentences originating from the front of the listener (0° azimuth) and embedded in a single speech-shaped noise source placed on the right side of the listener (+90°). Error bars indicate standard deviations.

space. This improvement in performance with the proposed 4M-SESS strategy was maintained even in the challenging condition where three noise sources were present.

In addition, the overall benefit after processing with the 4M-SESS processing strategy was significantly higher than the benefit received when processing with the binaural 2M2-BEAM noise reduction strategy. In both the single and multinoise source scenarios the 4M-SESS strategy employing two contralateral microphone signals led to a substantially increased performance, especially when the speech and the noise source were spatially separated. The observed improvement in performance with the 4M-SESS strategy can be attributed to having a more reliable reference signal. In essence, combining spatial information from all four microphones forms a better representation of the reference signal, leading to a better target segregation than that made available with only a monaural input or two (independent) binaural inputs.

Future Directions

CI devices provide a high level of functional hearing to deaf patients, and the sophistication of the implant hardware continues to improve rapidly. In fact, the most significant advances in CI performance over the past 20 years have been made because of improvements in basic CI technology, such as faster speech processors and more electrodes (e.g., see Zeng, Rebscher, Harrison, Sun, & Feng, 2008). Below, we propose some potential avenues for further investigation into next-generation sound processors:

Multicore processors. A cochlear implant is a type of battery-powered embedded system, and power consumption

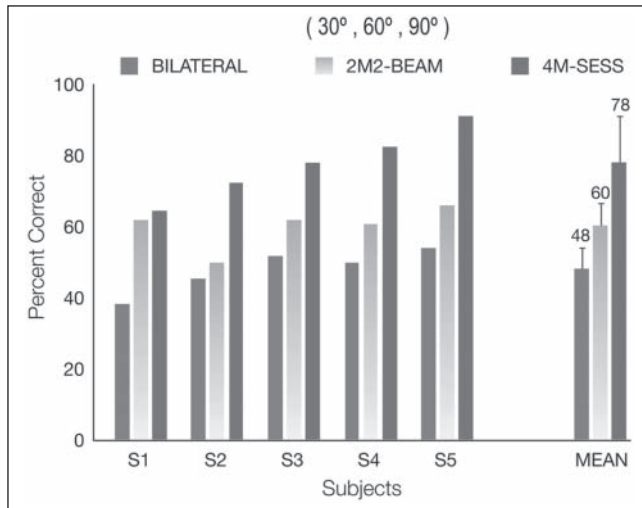


Figure 11. Percent correct scores by five Nucleus 24 users using both CI devices tested on IEEE sentences originating from the front of the listener (0° azimuth) and embedded in three speech-shaped noise sources distributed on the right side of the listener (30° , 60° , 90°). Error bars indicate standard deviations.

is among the most critical factors to consider during product design.⁸ For this reason, compared to other modern electronic devices (e.g., smartphones) that run between 1 and 2 GHz clock speeds, many cochlear implant sound processors need to maintain a low-power profile operation and are therefore clocked on speeds below 100 MHz. This amounts to a difference in computational power of a factor that ranges between 10 and 20. For instance, in the last few years, cochlear implants users have lived in world of limited computational power and devices that provide a fairly limited budget of at most 120 to 180 million-instructions-per-second (MIPS) (e.g., see Patrick et al., 2006; Zeng et al., 2008). Such slow clock speeds have made it difficult for developers to integrate single-microphone and multi-microphone noise suppression strategies for real-time signal processing in current sound processors. We believe that the next phase of improvements in cochlear implant performance can come from incorporating high-performance⁹ multicore processors to process signals and routines in parallel with one another. By having multiple cores to work with, software designers can dedicate groups of processors assigned to specific tasks. For instance, a number of cores can be dedicated to performing highly computationally intensive audio processing, whereas other cores could handle wireless interfaces, external memory, and user interface functions. A fairly powerful 40-core asynchronous processor running on clock speeds in excess of 700 MHz while delivering approximately 25,000 MIPS that could potentially be used in next-generation auditory prostheses was recently made commercially available (Intella-sys, 2009).

Alternative manufacturing technologies. Investigate the application of alternative design and manufacturing

principles for next-generation sound processors by resorting to application-specific integrated circuits (ASICs). To our knowledge, current sound processors used in HA and CI devices today are typically built using older-generation ASIC technologies, which consume significantly more power when compared to more efficient state-of-the-art ASIC chips (Zeng et al., 2008). Recent advances in bipolar metal oxide semiconductor (BiMOS) fabrication technologies and silicon-germanium-on-insulator (SGOI) chips are making the design and implementation of faster and more efficient internal ASICs more feasible than ever before. In fact, experimental prototypes of such next-generation low-power and small-scale sound processors for cochlear implant devices have been recently developed (e.g., see Sit et al., 2007). These devices have been shown to significantly extend battery life while still benefiting a considerable boost in computational power. We believe that technological advances in this front could enable other computationally powerful and power-efficient fully neural stimulating interfaces to be developed, thus bringing them a step closer to universal accessibility and widespread clinical use.

Mobile and cloud computing. Recent advances in mobile and cloud computing, if taken advantage, may enable future speech processors with access to almost unlimited computational resources by establishing a wireless-based connection to the Internet. Over the next years, we envision such low-cost, low-power speech processors for use in cochlear implant devices, with enhanced mobile connectivity and Internet functionality. Today, a very large number of commercially available hearing aid devices offer Bluetooth wireless connectivity. Bluetooth-enabled hearing aids currently use wireless technology either for communication between the two sides or for connecting to other Bluetooth-enabled devices (e.g., see Burrows, 2010). The strong adoption of Bluetooth technology by today's hearing aid manufacturers strongly suggests that a similar trend could be followed in next-generation cochlear implants soon. The advantages of such functionality are numerous. In principle, Bluetooth wireless connectivity could be used to relay information from the sound processor to a smartphone¹⁰ or any other wireless platform. By establishing a Bluetooth connection with an Internet capable handheld device, the sound processor on the implant device could then relay (or stream) data for real-time processing on the smartphone itself. Alternatively, the sound processor could use the smartphone device to wirelessly submit data to a dedicated Internet-based computing platform for further processing. In theory, next-generation sound processors could rely entirely on cloud computing systems to compensate for the lack of advanced hardware and storage capabilities.

Wireless binaural links. Bilateral electrical stimulation is far superior to monaural stimulation, and bilateral cochlear implants (BCIs) offer considerable advantages over monaural fittings. This is due to the ability of BCIs to reinstate, at

least to some extent, the interaural amplitude and timing difference cues that allow people with normal hearing to lateralize sounds in the horizontal plane and to attend selectively to a primary auditory input (e.g., speech), among multiple other sound sources at different locations (e.g., see Kokkinakis & Loizou, 2010; Litovsky et al., 2004; Litovsky, Parkinson, & Arcaroli, 2009). Nowadays, individuals fitted with BCIs can achieve high open-set speech recognition scores of 80% or higher from having two ears stimulated instead of one. In addition, as discussed previously (see “Multi-Microphone Noise Reduction”), cochlear implant systems that come with sound processors equipped with multiple microphones, have been shown to provide reasonable improvements in speech intelligibility even in challenging noisy listening settings. Still, in all bilateral paradigms available today, there are two independently operating devices on the left and right ears. To experience the true benefits of binaural hearing, the two sound processors (left and right) need to collaborate with one another to ensure that binaural cues are presented in a consistent manner to the implant user. To fully realize binaural hearing functionality, we could potentially rely on a wired link between the two devices. Such a scheme, however, may be unacceptable to some users from an aesthetic point of view. This makes the use of a wireless link the most viable alternative. In the presence of an ear-to-ear wireless binaural link, a simple practical scheme would be to transmit all the microphone signals from one ear to the other in order to obtain a better estimate of the desired target signal. Instead, a more intelligent scheme that would only select specific auditory input streams from the left and right microphones could be implemented. In both cases, optimizing which microphone signals to transmit would be critical; especially since wireless transmission of speech data can be power intensive. The potential of such schemes in restoring spatial cues necessary for binaural perception needs to be assessed. The investigation of such binaural strategies that rely on reduced transmission bandwidths for use in next-generation bilateral devices is currently underway in our laboratories.

Author's Note

The experimental data presented in this study were collected and analyzed by the first and third author.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by Grants R03 DC 008882 (Kostas Kokkinakis) and R03 DC 008887 (Yi Hu) awarded from the National Institute of Deafness and other Communication Disorders (NIDCD) of the National Institutes of Health (NIH).

Notes

1. Intelligibility is an objective measure, which measures the percentage of spoken words that can be correctly identified by a listener. In contrast, speech quality is highly subjective in nature and is indicative of the extent to which the listener is comfortable with the speech signal. These two measures may not always correlate well (e.g., see Loizou, 2007; Kim, Lu, Hi, & Loizou, 2009).
2. According to the Food and Drug Administration (FDA), as of May 2010, approximately 200,000 people worldwide have received implants. In the United States, roughly 45,000 adults and around 25,500 children have received cochlear implants. There are currently three cochlear implant systems in use in the United States: (1) the Harmony processor and the HiRes 90K (repackaged Clarion CII device) implant manufactured by the Advanced Bionics Corporation in Sylmar, CA; (2) the Nucleus 5 system with the CP810 sound processor and CI 512 implant manufactured by Cochlear Limited in Sydney, Australia; and (3) the OPUS 2XS sound processor and the CONCERTO implant system manufactured by MED-EL Corporation in Innsbruck, Austria. These devices vary in size, water resistivity, and reliability. They also differ in the input dynamic range (IDR), stimulation rate (temporal resolution), number of electrodes (spectral resolution), type of processor, and sound processing strategies they use.
3. The speech reception threshold (SRT) is defined as the signal-to-noise ratio (SNR) necessary to achieve 50% intelligibility. In general, high SRT values correspond to poor speech intelligibility.
4. Note that this specific function is fairly suitable because it has a sigmoidal shape similar to the human listener's psychometric function of intelligibility versus SNR.
5. Microphones operate by sensing the pressure difference on either side of a thin sheet known as a diaphragm. Ultimately, there are really only two fundamental microphone types—omnidirectional and directional. Omnidirectional microphones are designed to be equally sensitive to sounds arriving from all directions, without essentially aiming to favor one direction over another, while directional microphones are sensitive to sounds emanating from only one direction and reject sounds coming from different azimuths outside the desired pickup area (polar pattern) (e.g., see Chung, 2004; Chung et al., 2006).
6. Unlike the Nucleus Freedom sound processor which uses an omnidirectional microphone plus a hardware dual-port directional microphone, the new Nucleus 5 sound processor (CP810) uses two-omnidirectional microphones which can be combined to produce several directional responses. The CP810 dual omnidirectional microphones are more flexible to control and can more accurately tune the directional patterns to create an optimal speech and noise reference.
7. For a more thorough explanation of the BEAM strategy, the reader is referred to Spriet et al. (2007).
8. The power consumption of the most recent Nucleus 5 sound processor (CP810) by Cochlear Americas ranges from 20 mW to around 100 mW (Cochlear Limited, 2009).

9. In the past, devices from Cochlear have used low-performance multicore processors.
10. According to recent survey data published by Pew Research Center's Internet and American Life Project on smartphone adoption and usage, among all cellphone owners tested, 35% of all American adults own a smartphone as of 2011. More interestingly, the same study documents evidence of a strong trend towards increased smartphone ownership among those respondents who are aged between 18 to 24 and 25 to 34 with the percentages of ownership being 50% and 58%, respectively (Smith, 2011). Based on this data, it is very likely that a large number of future cochlear implantees will also own a smartphone.

References

- Bentler, R. A., & Chiou, L.-K. (2006). Digital noise reduction: An overview. *Trends in Amplification*, *10*, 67-82.
- Boll, S. F. (1979). Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on Acoustics, Speech and Signal Processing*, *2*, 113-120.
- Brungart, D., Chang, P., Simpson, B., & Wang, D. (2006). Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. *Journal of the Acoustical Society of America*, *120*, 4007-4018.
- Burrows, D. L. (2010). Bluetooth technology in hearing aids: A clinician's perspective. *Perspectives on Audiology*, *6*, 4-8.
- Chung, K. (2004). Challenges and recent developments in hearing aids: Microphone technologies and noise reduction algorithms. Part I. *Trends in Amplification*, *8*, 83-124.
- Chung, K., Zeng F.-G., & Acker, K. N. (2006). Effects of directional microphone and adaptive multichannel noise reduction algorithm on cochlear implant performance. *Journal of the Acoustical Society of America*, *120*, 2216-2227.
- Cochlear Limited (2009). *Nucleus 5 (CP810) sound processor technical specifications*. Centennial, CO, USA.
- Ephraim, Y., & Malah, D. (1984). Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech and Signal Processing*, *32*, 1109-1121.
- Ephraim, Y., & VanTrees, H. L. (1995). A signal subspace approach for speech enhancement. *IEEE Transaction on Speech and Audio Processing*, *3*, 251-266.
- Fetterman, B. L., & Domico, E. H. (2002). Speech recognition in background noise of cochlear implant patients. *Otolaryngology Head and Neck Surgery*, *126*, 257-263.
- Firszt, J., Holden, L., Skinner, M., Tobey, E., Peterson, A., Gaggl, W., . . . Wackym, P. A. (2004). Recognition of speech presented at soft to loud levels by adult cochlear implant recipients of three cochlear implant systems. *Ear and Hearing*, *25*, 375-387.
- Fu, Q., & Nogaki, G. (2004). Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing. *Journal of the Association for Research in Otolaryngology*, *6*, 19-27.
- Gates, G. A., & Miyamoto, R. T. (2003). Cochlear implants. *New England Journal of Medicine*, *349*, 421-423.
- Greenberg, Z. E., & Zurek, P.M. (1992). Evaluation of an adaptive beamforming method for hearing aids. *Journal of the Acoustical Society of America*, *91*, 1662-1676.
- Griffiths, L. J., & Jim, C. W. (1982). An alternative approach to linearly constrained adaptive beamforming. *IEEE Transactions on Antennas and Propagation*, *30*, 27-34.
- Hochberg, I., Boothroyd, A., Weiss, M., & Hellman, S. (1992). Effects of noise and noise suppression on speech perception for cochlear implant users. *Ear and Hearing*, *13*, 263-271.
- House W. F., & Berliner K. I. (1991). Cochlear implants: From idea to clinical practice. In H. Cooper (Ed.), *Cochlear implants: A practical guide* (pp. 9-33). London, UK: Whurr.
- Hu, Y., & Loizou, P. C. (2002). A subspace approach for enhancing speech corrupted with colored noise. *IEEE Signal Processing Letters*, *9*, 204-206.
- Hu, Y., Loizou, P. C., Li, N., & Kasturi, K. (2007). Use of a sigmoidal-shaped function for noise attenuation in cochlear implants. *Journal of the Acoustical Society of America*, *122*, 128-134.
- Hu, Y., & Loizou, P. C. (2008). A new sound coding strategy for suppressing noise in cochlear implants. *Journal of the Acoustical Society of America*, *124*, 498-509.
- Hu, Y., & Loizou, P. C. (2010). Environment-specific noise suppression for improved speech intelligibility by cochlear implant users. *Journal of the Acoustical Society of America*, *127*, 3689-3695.
- IEEE (1969). IEEE recommended practice speech quality measurements. *IEEE Transactions on Audio and Electroacoustics*, *17*, 225-246.
- Intelliasys (2009). *SEK 40C18 Data Sheet*. Cupertino, CA, USA.
- Kim, G., Lu, Y., Hi, Y., & Loizou, P. C. (2009). An algorithm that improves speech intelligibility in noise for normal-hearing listeners. *Journal of the Acoustical Society of America*, *126*, 1486-1494.
- Kokkinakis, K., & Nandi, A. K. (2006). Multichannel blind deconvolution for source separation in convolutive mixtures of speech. *IEEE Transactions on Audio, Speech, and Language Processing*, *14*, 200-213.
- Kokkinakis, K., & Loizou, P. C. (2007). Signal separation by integrating adaptive beamforming with blind deconvolution. *Lecture Notes Computer Science*, *4666*, 495-503.
- Kokkinakis, K., & Loizou, P. C. (2008). Using blind source separation techniques to improve speech recognition in bilateral cochlear implant patients. *Journal of the Acoustical Society of America*, *123*, 2379-2390.
- Kokkinakis, K., & Loizou, P. C. (2010). Multi-microphone adaptive noise reduction strategies for coordinated stimulation in bilateral cochlear implant devices. *Journal of the Acoustical Society of America*, *127*, 3136-3144.
- Kokkinakis, K., Hazrati, O., & Loizou, P. C. (2011). A channel-selection criterion for suppressing reverberation in cochlear implants. *Journal of the Acoustical Society of America*, *129*, 3221-3232.
- Kollmeier, B., & Koch, R. (1994). Speech enhancement based on physiological and psychoacoustical models of modulation perception and binaural interaction. *Journal of the Acoustical Society of America*, *95*, 1593-1602.

- Kompis, M., & Dillier, N. (1994). Noise reduction for hearing aids: Combining directional microphones with an adaptive beamformer. *Journal of the Acoustical Society of America*, 96, 1134-1143.
- Levitt, H. (2001). Noise reduction in hearing aids: A review. *Journal of Rehabilitation Research and Development*, 38, 111-121.
- Li, N., & Loizou, P. C. (2008). Factors influencing intelligibility of ideal binary-masked speech: Implications for noise reduction. *Journal of the Acoustical Society of America*, 123, 1673-1682.
- Lim, J. S. (1978). Evaluation of a correlation subtraction method for enhancing speech degraded by additive white noise. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 26, 471-472.
- Lim, J. S., & Oppenheim, A. V. (1979). Enhancement and bandwidth compression of noisy speech. *Proceedings of the IEEE*, 67, 1586-1604.
- Litovsky, R., Parkinson, A., Arcaroli, J., Peters, R., Lake, J., Johnstone, P., & Yu, G. (2004). Bilateral cochlear implants in adults and children. *Archives of Otolaryngology—Head & Neck Surgery*, 130, 648-655.
- Litovsky, R., Parkinson, A., & Arcaroli, J. (2009). Spatial hearing and speech intelligibility in bilateral cochlear implant users. *Ear and Hearing*, 30, 419-431.
- Loizou, P. C. (1998). Mimicking the human ear. *IEEE Signal Processing Magazine*, 15, 101-130.
- Loizou, P. C., Lobo, A., & Hu, Y. (2005). Subspace algorithms for noise reduction in cochlear implants. *Journal of the Acoustical Society of America*, 118, 2791-2793.
- Loizou, P. C. (2007). *Speech enhancement: Theory and practice*. Boca Raton, FL: CRC Press.
- Loizou, P. C., Hu, Y., Litovsky, P., Yu, G., Peters, R., Lake, J., & Roland, P. (2009). Speech recognition by bilateral cochlear implant users in a cocktail party setting. *Journal of the Acoustical Society of America*, 125, 372-383.
- Nilsson, M., Soli, S. D., & Sullivan, J. (1994). Development of hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America*, 95, 1085-1099.
- Patrick, J. F., Busby, P. A., & Gibson, P. J. (2006). The development of the Nucleus Freedom cochlear implant system. *Trends in Amplification*, 10, 175-200.
- Qin, M. K., & Oxenham, A. J. (2003). Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. *Journal of the Acoustical Society of America*, 114, 446-454.
- Ricketts, T. A. (2001). Directional hearing aids. *Trends in Amplification*, 5, 139-176.
- Ricketts, T. A. (2005). Directional hearing aids: Then and now. *Journal of Rehabilitation Research and Development*, 38, 133-144.
- Schroeder, M. R. (1965). Apparatus for suppressing noise and distortion in communication signals, U. S. Patent No. 3,180,936, Filed, December 1960, Issued, April 1965, United States Patent and Trademark Office (USPTO).
- Sit, J.-J., Simonson, A. M., Oxenham, A. J., Faltys, M. A., & Sarpeshkar, R. (2007). A low-power asynchronous interleaved sampling algorithm for cochlear implants that encodes envelope and phase information. *IEEE Transactions on Biomedical Engineering*, 54, 138-149.
- Skinner, M., Holden, L. K., Whitford, L. A., Plant, K. L., Psarros, C., & Holden, T. A. (2002). Speech recognition with the Nucleus 24 SPEAK, ACE, and CIS speech coding strategies in newly implanted adults. *Ear and Hearing*, 23, 207-223.
- Smith, K. (2011). *Smartphone adoption and usage*. Retrieved from <http://www.pewinternet.org/Reports/2011/Smartphones.aspx>
- Spahr, A., & Dorman, M. (2004). Performance of subjects fit with the Advanced Bionics CII and Nucleus 3G cochlear implant devices. *Archives of Otolaryngology—Head & Neck Surgery*, 130, 624-628.
- Spriet, A., Van Deun, L., Eftaxiadis, K., Laneau, J., Moonen, M., Van Dijk, B., . . . Wouters, J. (2007). Speech understanding in background noise with the two-microphone adaptive beamformer BEAM in the Nucleus Freedom cochlear implant system. *Ear and Hearing*, 28, 62-72.
- Tyler R. S., Gantz B. J., Rubinstein J. T., Wilson B. S., Parkinson A. J., Wolaver A., Preece, J. P., . . . Lowder M. W. (2002). Three-month results with bilateral cochlear implants. *Ear and Hearing*, 23, 80-89.
- van Hoesel, R. J. M., & Clark, G. M. (1995). Evaluation of a portable two-microphone adaptive beamforming speech processor with cochlear implant patients. *Journal of the Acoustical Society of America*, 97, 2498-2503.
- Tchorz, J., & Kollmeier, B. (2003). SNR estimation based on amplitude modulation analysis with applications to noise suppression. *IEEE Transactions on Speech and Audio Processing*, 11, 184-192.
- Turner, C. W., Gantz, B. J., Vidal, C., Behrens, A., & Henry, B. A. (2004). Speech recognition in noise for cochlear implant listeners: Benefits of residual acoustic hearing. *Journal of the Acoustical Society of America*, 115, 1729-1735.
- Weiss, M. (1993). Effects of noise and noise reduction processing on the operation of the Nucleus 22 cochlear implant processor. *Journal of Rehabilitation Research and Development*, 30, 117-128.
- Wilson, B. S., & Dorman, M. F. (2008a). Cochlear implants: A remarkable past and a brilliant future. *Hearing Research*, 242, 3-21.
- Wilson, B. S., & Dorman, M. F. (2008b). Cochlear implants: Current designs and future possibilities. *Journal of Rehabilitation Research and Development*, 45, 695-730.
- Wouters, J., & Van den Berghe, J. (2001). Speech recognition in noise for cochlear implantees with a two microphone monaural adaptive noise reduction system. *Ear and Hearing*, 22, 420-430.
- Yang, L.-P., & Fu, Q.-J. (2005). Spectral subtraction-based speech enhancement for cochlear implant patients in background noise. *Journal of the Acoustical Society of America*, 117, 1001-1004.
- Zeng, F. G. (2004). Trends in cochlear implants. *Trends in Amplification*, 8, 1-34.
- Zeng, F. G., Rebscher, S., Harrison, W., Sun, X., & Feng, H. (2008). Cochlear implants: System design, integration, and evaluation. *IEEE Reviews in Biomedical Engineering*, 1, 115-142.