The following article appeared in

Hersbach, A., Grayden, D., Fallon, J., & McDermott, H. (2013). A beamformer post-filter for cochlear implant noise reduction. Journal of the Acoustical Society of America, 133(4), 2412-2420.

and may be found at

http://link.aip.org/link/?JAS/133/2412/

A beamformer post-filter for cochlear implant noise reduction[a)]

Adam A. Hersbach[b)] and David B. Grayden[c)]
NeuroEngineering Laboratory, Department of Electrical and Electronic Engineering, University of Melbourne, Melbourne, Australia

James B. Fallon[d)] and Hugh J. McDermott[d)]
Bionics Institute, East Melbourne, Australia

(Received 20 September 2012; accepted 16 February 2013)

Cochlear implant users have limited ability to understand speech in noisy conditions. Signal processing methods to address this issue that use multiple microphones typically use beamforming to perform noise reduction. However, the effectiveness of the beamformer is diminished as the number of interfering noises increases and the acoustic environment becomes more diffuse. A multi-microphone noise reduction algorithm that aims to address this issue is presented in this study. The algorithm uses spatial filtering to estimate the signal-to-noise ratio (SNR) and attenuates time-frequency elements that have poor SNR. The algorithm was evaluated by measuring intelligibility of speech embedded in 4-talker babble where the interfering talkers were spatially separated and changed location during the test. Twelve cochlear implant users took part in the evaluation, which demonstrated a significant mean improvement of 4.6 dB (standard error 0.4, $P < 0.001$) in speech reception threshold compared to an adaptive beamformer. The results suggest that a substantial improvement in performance can be gained for cochlear implant users in noisy environments where the noise is spatially separated from the target speech. © *2013 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.4794391]

PACS number(s): 43.71.Ky, 43.72.Dv, 43.66.Ts, 43.60.Fg [MAH] Pages: 2412–2420

I. INTRODUCTION

Cochlear implant recipients often need to communicate in noisy situations in their everyday lives, where speech understanding is poorer than that usually obtained in quiet conditions. Noise-reduction signal processing can be used to enhance the intelligibility of speech signals before they are delivered to the implant. Single-microphone noise reduction techniques have been applied to cochlear implant sound processing and have demonstrated speech intelligibility improvements (Buechner *et al.*, 2010; Hu and Loizou, 2010; Dawson *et al.*, 2011). The maximum benefits are demonstrated in stationary noise with benefits of around 2.5 dB in speech reception threshold (SRT) or, equivalently, 25 percentage point improvement in word recognition. The benefit of single-channel noise reduction is reduced when the interfering noise is non-stationary, such as when the interfering noise is due to competing talkers.

Many present day hearing aids and cochlear implant sound processors utilize more than one microphone, significantly broadening the scope of noise reduction processing. Sound impinging on the microphone array is subject to a time delay between the microphone inlet ports, which results in the ability to use information about spatial location of sound sources, not possible with a single microphone. A common approach is to use two omni-directional microphones to form fixed or adaptive directional microphones, also known as beamforming (Griffiths and Jim, 1982).

The benefit of multi-microphone noise reduction becomes apparent when the target and noise are spatially separated. The direction of arrival of sound is exploited to spatially filter the signal and successfully remove the noise. Using a single, unilateral sound processor with two microphones, large improvements due to adaptive beamforming have been demonstrated for cochlear implant recipients, especially in specific acoustic scenarios involving a single interfering noise. Improvements were 7–16 dB SRT (Wouters and Vanden Berghe, 2001), 10 dB SRT (Spriet *et al.*, 2007), 8–10 dB SRT (Kompis *et al.*, 2008), and 56 percentage points in word recognition (Chung and Zeng, 2009). As the number of noise sources is increased to create a more diffuse noise field, the benefit is somewhat moderated but still significant, e.g., 2–12 dB SRT (Spriet *et al.*, 2007), 3.5 dB SRT (Chung *et al.*, 2006), 6.1 dB SRT (Gifford and Revit, 2010), and 47 percentage points word recognition (Chung and Zeng, 2009).

The Nucleus® 5 CP810 sound processor is the most recent commercially available cochlear implant sound processor available from Cochlear™ Ltd (Sydney, Australia). It has two omni-directional microphones that are used to form three different directional patterns, configurable through the commercial fitting software used to program the device. The first configuration is standard, offering a fixed directional microphone with moderate directionality. The second provides a fixed zoom pattern offering higher directionality, and the third is an adaptive beamformer called Beam® (Sydney,

[a)]Portions of this paper were presented at the International Hearing Aid Conference (IHCON), Lake Tahoe, California, 8–12 August 2012.
[b)]Author to whom correspondence should be addressed. Also at: Cochlear Ltd, East Melbourne, Australia. Electronic mail: ahersbach@cochlear.com
[c)]Also at: Centre for Neural Engineering, University of Melbourne, Australia and Bionics Institute, East Melbourne, Australia.
[d)]Also at: Departments of Otolaryngology and Medical Bionics, University of Melbourne, Melbourne, Australia.

 0001-4966/2013/133(4)/2412/9/$30.00

Australia) providing an adaptive highly directional microphone response (Spriet *et al.*, 2007).

Recently, the three directional settings available in the CP810 processor were evaluated in spatially separated noise, where four independent competing talkers were presented from locations in the rear hemi-field (Hersbach *et al.*, 2012). The location of each competing talker was randomized among seven loudspeakers arranged at 30° spacing from 90° to 270°, and at each 3 s interval during the test, one of the four competing talkers was randomly chosen and assigned a new random location. In this challenging environment, the statistically significant benefit demonstrated was 3.7 dB SRT for zoom directionality and 5.3 dB SRT for Beam, compared to the standard directionality setting. The superior performance of Beam was due to the adaptive noise cancellation stage, and represents the typical outcome when comparing fixed directional processing to adaptive beamformers, i.e., addition of the adaptive filtering stage results in improved performance.

While large benefits have been demonstrated using adaptive beamformer algorithms for cochlear implant users, there are drawbacks of this approach. It is well known that performance degrades with increased number of noise sources or the diffusiveness of the noise field, and with increased reverberation (van Hoesel and Clark, 1995; Hamacher *et al.*, 1997; Wouters and Vanden Berghe, 2001). The inability to cancel multiple noise sources simultaneously is a limitation of the adaptive filter, which steers a null in the direction of the noise in order to cancel the noise. It cannot cancel more than one direction at any time instant.

In order to enhance beamformer performance, several approaches to the design of a so-called post-filter have been explored from a theoretical perspective. Zelinski (1988), Meyer and Simmer (1997), McCowan and Bourlard (2003), and Leukimmiatis *et al.* (2006) derived optimal filters that differed in the theoretical assumptions made on the sound field. For example, Zelinski (1988) assumed a perfectly diffuse noise field, while McCowan and Bourlard (2003) assumed knowledge of the noise coherence across the microphone array. These assumptions, however, generally do not hold for many real situations and, therefore, these theoretical approaches provide limited practical solutions to improve beamformer performance in cochlear implant devices. Yousefian and Loizou (2013) proposed a dual-microphone coherence-based noise reduction algorithm for cochlear implant users and demonstrated 5–10 dB SRT improvement over a fixed directional microphone in an anechoic room with one or two competing talkers. The algorithm assumed that speech and noise were coherent across the microphone array, and was also evaluated by normal hearing listeners in various degrees of reverberation. The benefit was substantially reduced to 0–2 dB SRT when evaluated in a moderately reverberant room (T60 = 465 ms) suggesting the coherence assumption was not as strong under reverberant conditions.

As an alternative to relying on assumptions of the speech and noise coherence, Cao *et al.* (1996) used a post-filter based purely on spatial filtering. In a system with seven omni-directional microphones, they reported an objective measure of segmental signal-to-noise ratio (SNR), showing an improvement of 5.6 dB over the main beamformer stage. Wolff and Buck (2008) also proposed a spatial post-filter and described an *a posteriori* SNR estimate used in a statistically optimized maximum *a posteriori* (MAP) estimator. In a four-microphone system, they compared the MAP post-filter with a Wiener post-filter, and found substantial improvements in computer word recognition with error rate benefit of between 20 and 40 percentage points depending on the noise type and input SNR used (0–18 dB SNR). These studies provide evidence that the use of a spatial post-filter can improve beamformer output, and improve it in such a way that is superior to the theoretically optimal Wiener post-filter. In this study using a two-microphone system, we aim to demonstrate that a beamformer post-filter based on spatial filtering can improve cochlear implant performance.

This research has two main aims: (i) to propose and describe a beamformer post-filter algorithm designed to improve beamformer noise reduction performance, and (ii) to evaluate the post-filter algorithm by measuring speech intelligibility with a group of cochlear implant recipients.

II. ALGORITHM DESCRIPTION

The algorithm ("SpatialNR") is formulated as a beamformer post-filter designed to enhance intelligibility of the beamformer output. It is designed for real-time, low complexity implementation for a small separation (<2 cm) dual-microphone array typically used in hearing aids and cochlear implants. No assumption is made of the underlying speech and noise power distributions, and no *a priori* knowledge of the speech and noise signals is used in the processing; however, the target spatial location is assumed to be in front of the listener. The main issue that is addressed is the ability to cancel multiple spatially separated noise sources simultaneously, a limitation of traditional adaptive beamformers.

The algorithm estimates the SNR by analyzing the signals from front-facing and rear-facing fixed directional microphones. SNR is estimated in each frequency band and is based on the assumption that the target direction is in front. The SNR is used to control a parametric Wiener-like gain function with adjustable gain in order to attenuate frequency channels that are noisy. The result is a spatial post-filter that attenuates sound from behind and beside the listener while passing sounds from the front.

A. Algorithm kernel

The SpatialNR algorithm kernel is shown in Fig. 1. Two omni-directional microphone signals are used as inputs to a spatial pre-filter stage, where two spatial patterns are formed—a front-facing speech reference and a rear-facing noise reference. The spatial sensitivity of the speech reference and noise reference for microphones worn on the left ear of a KEMAR (Holte, Denmark) manikin (Burkhard and Sachs, 1975) are shown in Fig. 2(A). The speech reference is configured to have a front-facing polar pattern of maximum directivity index when worn on the head. Note there is increased sensitivity at around 45° on the same side of the manikin head where the microphones are worn, due to the

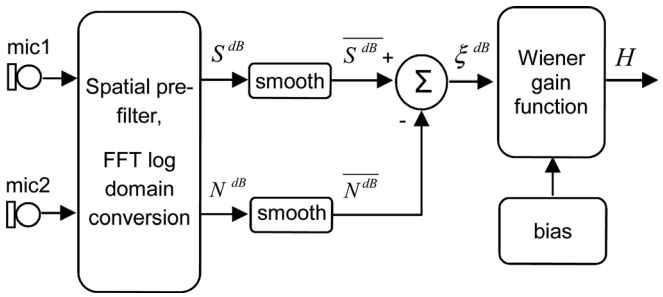

FIG. 1. SpatialNR algorithm kernel.

head shadow. The noise reference is configured to have a rear-facing polar pattern with a null in the target, or "look" direction, in this case directly in front of the listener. The spatial pre-filter signals are created with a delay-and-sum structure, by filtering and summing the microphone input signals in the time domain (e.g., Ward *et al.*, 2001). A fast Fourier transform (FFT) is used to transform these signals into the complex spectral domain, creating frequency domain representations of the speech reference, $S_k[n]$, and noise reference, $N_k[n]$, where k is the frequency index and n is the time index of overlapping FFT windows. The signals are transformed to the log domain and filtered using a first-order infinite impulse response filter with time constant of 10 ms corresponding to a value of $\beta = 0.185$ at the 489 Hz rate of successive FFT frames

$$\overline{S^{\mathrm{dB}}}_k[n] = \beta S^{\mathrm{dB}}{}_k[n] + (1-\beta)\overline{S^{\mathrm{dB}}}_k[n-1], \tag{1a}$$

$$\overline{N^{\mathrm{dB}}}_k[n] = \beta N^{\mathrm{dB}}{}_k[n] + (1-\beta)\overline{N^{\mathrm{dB}}}_k[n-1]. \tag{1b}$$

Smoothing in the log domain is used because it relates more closely to perceptual loudness. The smoothed signals are used to estimate the instantaneous SNR, $\xi_k[n]$, at each time point, n, and in each frequency band, k,

$$\xi^{\mathrm{dB}}{}_k[n] = \overline{S^{\mathrm{dB}}}_k[n] - \overline{N^{\mathrm{dB}}}_k[n]. \tag{2}$$

The SNR estimate $\xi_k[n]$ is then used as the primary means to attenuate time-frequency bins that are considered noisy. It is used to control a parametric Wiener-like gain, $H_k[n]$, with adjustable bias, α,

$$H_k[n] = \frac{\xi_k[n]}{\alpha + \xi_k[n]}. \tag{3}$$

The noise-reduced output signal, $\hat{X}_k[n]$, is created by applying the filter gain, $H_k[n]$, to the output of the beamformer stage

$$\hat{X}_k[n] = H_k[n]S_k[n]. \tag{4}$$

B. Theoretical analysis

The spatial response or directivity pattern of the SpatialNR algorithm depends on the input signal and changes over time. The spatial location and level of the noise and the target signal, the spatial pre-filtering stage, and the bias parameter, α, all have an influence on the spatial pattern. For brevity in the remainder of the text, the frequency bin index, k, and frame index, n, are dropped.

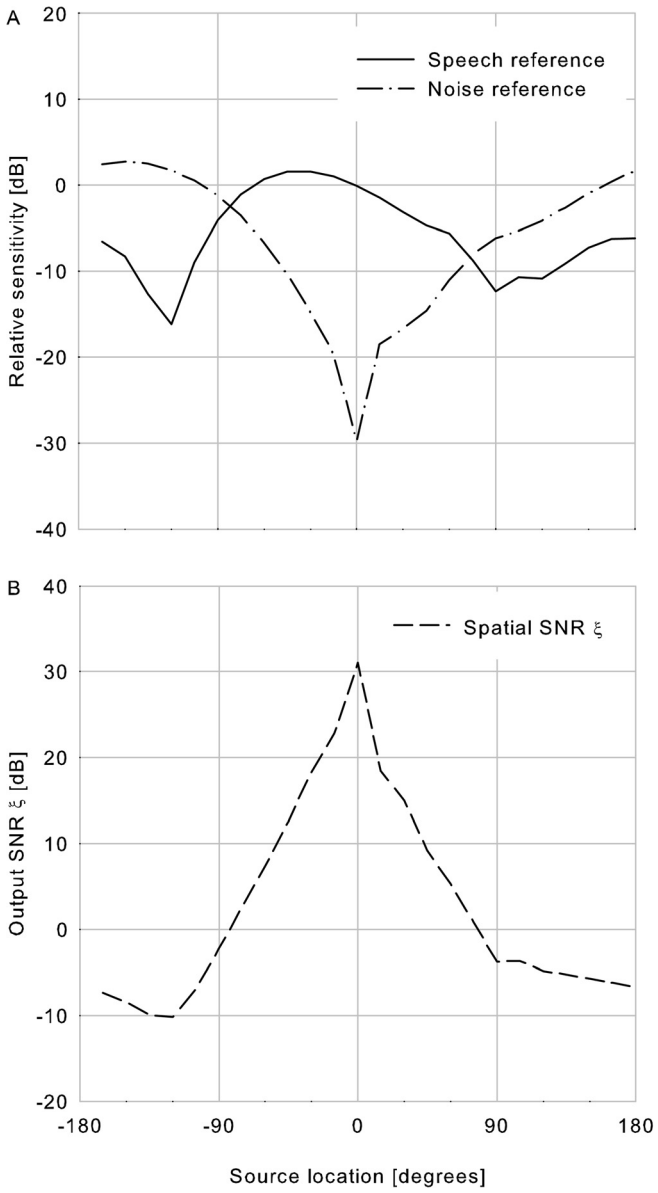

FIG. 2. (A) Single source spatial sensitivity of speech reference (zoom directionality) and noise reference signals, and (B) the resulting single source SNR ξ^{dB} according to Eq. (2). Sensitivity was recorded with a broadband speech-shaped noise stimulus presented under anechoic conditions with device worn on the left ear of a KEMAR manikin. The relative sensitivity is the RMS value across the duration of the stimulus averaged across all frequency channels.

By way of illustration, it is useful to consider a single sound source impinging on the array. The value of ξ when a single sound source is present is determined entirely by the spatial location of the sound and the spatial pre-filtering stage, specifically the sensitivity of the speech reference compared to the noise reference at that location. Figure 2(A) shows the relative sensitivity of the speech reference and noise reference, and Fig. 2(B) shows the associated ξ^{dB} that is calculated according to Eq. (2).

In contrast to the single source case, when two signals are present at two different locations, each will contribute to ξ^{dB} by an amount dependent on the level and locations of each, and the final ξ^{dB} is calculated from the SNR

contributions due to the sound at each location. This is illustrated in Fig. 3, which shows ξ^{dB} as a function of input SNR for a target signal at 0° and a single noise from different angles ranging from 30° to 180°. For example, consider the case where the noise source is at 90°. Reading from Fig. 2(B), the target source at 0° will contribute a single source spatial SNR of $\xi^{dB} = +31$ dB, and the noise source at 90° will contribute a single source spatial SNR of $\xi^{dB} = -4$ dB. The combined ξ^{dB} shown in Fig. 3 depends on the instantaneous input SNR and the spatial pre-filter response of the speech reference at the two locations. It will approach +31 dB SNR when the target signal dominates, and −4 dB SNR when the noise signal at 90° dominates. ξ^{dB} is bounded by these two extremes due to the locations of the speech and noise signals. This is evident at all noise angles, where the SNR is lower bounded and dependent on the noise location.

The bias parameter, α, is used in the calculation of noise reduction gain, H. It is used to determine the SNR threshold at which elements are considered noisy and, therefore, attenuated. The effect of the bias parameter value on the system response to a single sound source is shown in Fig. 4 for the processing conditions used in this study. The bias parameter changes the spatial response and has the perceptual effect of setting the beam width in the target direction as well as affecting the attenuation of signals outside the target area, thus providing a mechanism to alter the aggressiveness of noise reduction.

III. MATERIALS AND METHODS

The algorithms were evaluated using speech intelligibility testing in noise with a group of cochlear implant recipients. The subject group, processing conditions, hardware, and test protocol are outlined in the following sections.

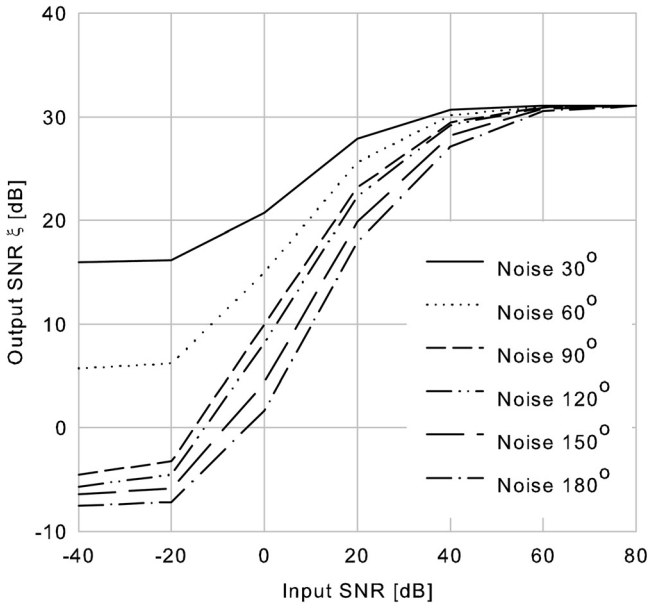

FIG. 3. SNR ξ^{dB} as a function of instantaneous input SNR for target speech from 0° embedded in speech-shaped noise from six different locations ranging from 30° to 180°. ξ^{dB} is upper-bounded by the single source SNR contribution of the target speech, and lower bounded by the single source SNR contribution of the noise.

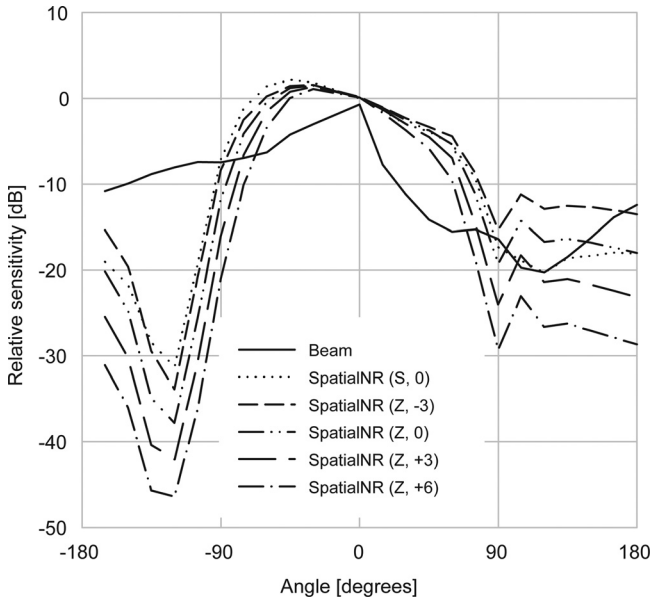

FIG. 4. Single source spatial sensitivity of Beam and SpatialNR processing including variants (shown in brackets) evaluated clinically in this study. *S* and *Z* indicate the signals to which the SpatialNR gains were applied (standard and zoom, respectively), and the following value is the bias parameter in dB. A broadband speech shaped noise stimulus was presented from each location under anechoic conditions with the device worn on the left ear of a KEMAR manikin. The system transfer function was measured from input to output, and thus represents the whole system response for each processing condition. The bias parameter has the effect of altering the beam width in the target directions as well as changing the attenuation of signals outside the target area.

A. Subjects

A group of 12 adult cochlear implant recipients participated in the study. Their mean age was 65.6 yrs [standard deviation (SD) 13.2 yrs] and mean time of implant use was 3.2 yrs (SD 1.6 yrs). All subjects were implanted with a Cochlear Nucleus® (Sydney, Australia) implant system and were current users of the CP810 or Freedom® sound processor and therefore had previous experience listening to the Beam algorithm.

In the evaluation, users' existing map parameters were transferred from their own device to the real-time evaluation system. All subjects used the Advanced Combination Encoder program (McDermott *et al.*, 1992; Vandali *et al.*, 2000) with sound processing features Autosensitivity (Patrick *et al.*, 2006) and Adaptive Dynamic Range Optimization (ADRO®) enabled (Blamey, 2005), the combination of which has provided beneficial outcomes for CI recipients when listening in noise (Wolfe *et al.*, 2011). The C-sound pressure level (SPL) parameter was set to 75 dB for all subjects. Stimulation rate and number of maxima are shown for each recipient in Table I. All subjects were tested monaurally using the real-time system to activate the cochlear implant in their best performing ear, while any direct acoustic transmission was attenuated using ear-plugs when necessary.

B. Processing conditions

Several variations of the SpatialNR algorithm were evaluated in this study. In particular, bias values, α, were evaluated at 2, 1, 0.5, and 0.25 corresponding to biasing ξ^{dB} by −3 dB,

TABLE I. Demographic details and cochlear implant map configuration for the 12 subjects who participated in the study. R: right ear, L: left ear, HA: hearing aid.

Subject	Age (years)	Implant use (years)	Stimulation rate (Hz)	Number of maxima	Test ear	Non-test ear device
S1	53	3	900	8	R	HA
S2	73	2	900	12	L	HA
S3	72	1	1200	8	L	Freedom
S4	45	5	1800	8	L	HA
S5	42	3	900	8	L	—
S6	82	2	900	8	R	CP810
S7	65	3	900	8	L	HA
S8	67	3	900	8	R	HA
S9	68	2	500	8	L	HA
S10	79	4	250	8	R	HA
S11	80	3	900	8	R	HA
S12	61	7	900	8	R	—

0 dB, +3 dB, and +6 dB SNR, respectively. In addition to varying the bias, a second algorithmic variable was tested where the noise reduction gains were applied to the standard directional signal as opposed to the zoom signal. This variation was tested using $\alpha = 1$ to form a total of five SpatialNR conditions. They are abbreviated as follows: SpatialNR (S,0), SpatialNR (Z,−3), SpatialNR (Z,0), SpatialNR (Z,+3), and SpatialNR (Z,+6), where S represents standard, Z represents zoom, and the bias value is in units of dB.

An additional condition was also evaluated to gauge benchmark performance. Beam, an adaptive generalized sidelobe canceller with energy-based voice activity detection, is commercially available in the CP810 speech processor. It was used as a baseline for comparison because it has demonstrated superior noise reduction performance to other directional microphone options available in the device (Hersbach *et al.*, 2012).

The Beam algorithm and the SpatialNR algorithm used a common spatial pre-filter that formed the speech reference (zoom directionality) and noise reference signals in the time domain (Fig. 5). The Beam algorithm used these inputs to an adaptive finite impulse response filter stage updated using a least mean squares algorithm to minimize the energy in the output signal.

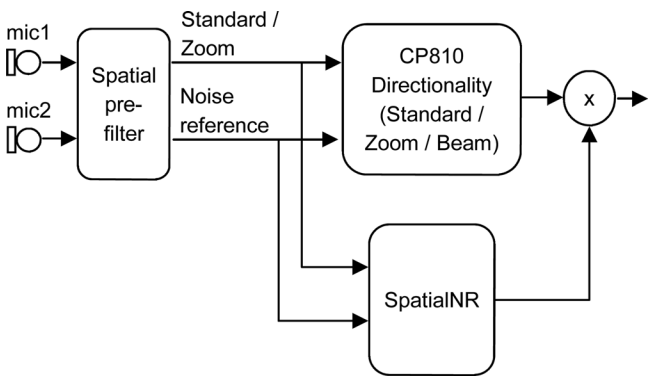

FIG. 5. Signal path integration used to evaluate the SpatialNR algorithm. The noise reduction gains were calculated from the front- and rear-facing directional patterns and the noise reduction gains were applied as a post-filter to the main beamformer stage.

C. Hardware

All signal processing was performed on a real-time computer-based speech processor, previously used for similar investigations (Dawson *et al.*, 2011). In this study, the system performed the same processing as the CP810 commercial device, but had the experimental SpatialNR algorithm integrated for the purpose of evaluation (Fig. 5). The system used a sampling rate of 15.7 kHz, and 128 point FFT frames were acquired every 16 samples at a rate of 489 Hz. Software control was used to switch between the Beam and SpatialNR algorithms and to control the SpatialNR parameters under investigation. During evaluation, the acoustic signal was captured using a modified CP810 sound processor, which contained only the microphone components that was connected to the real-time processor.

D. Test environment

The noise condition used in this study was developed by Hersbach *et al.* (2012) and was designed to improve on traditional tests, aiming to represent real-life environments where talkers are spatially separated and involved in conversation. To emulate this, the interfering maskers had different locations that were assigned randomly for each sentence presentation. The algorithm under evaluation was thus forced to adapt to the varying environment. The test is suitable for the evaluation of a wide variety of noise reduction algorithms, including multi-microphone and binaural algorithms.

Similar to the Hersbach *et al.* (2012) study, the target speech was always presented from the front direction, while the noise was presented using the remaining seven loudspeakers arranged in the rear hemi-field at discrete locations evenly spaced at 30° intervals from 90° to 270° azimuth. One of the noise conditions, 4-talker babble used by Hersbach *et al.* (2012), was selected for testing in this study to enable an indirect comparison to the standard and zoom directionality settings. To generate the noise environment for each sentence presentation, the allocations of the four independent talkers comprising the 4-talker babble were assigned randomly among the seven loudspeaker positions. The test was administered in a sound-treated booth measuring $4 \times 5 \times 2$ m^3. The target speech level was set to 65 dB SPL and the noise level was adjusted according to the subject's response to each sentence. Therefore, both the noise level and noise locations were altered after each sentence. Contrary to the Hersbach *et al.* (2012) study, where up to one mid-sentence change to one interferer was used, the noise locations in this study remained fixed for the duration of each sentence. Following the change after each subject response, the noise was presented in the new configuration for 3 s before the next sentence was presented, allowing both the subject and the algorithm under evaluation some time to adapt to the changed noise.

E. Test protocol

The study used a repeated-measure, single-subject design, in which each user served as his/her own control. The Australian Sentence Test in Noise (AuSTIN) was used

to assess speech intelligibility (Dawson *et al.*, 2013). The test is an adaptive SRT task and was used to obtain the SNR for 50% morphemes correct in each of the processing conditions. Each subject participated in two test sessions. A single SRT was collected for each of the six processing conditions in each test session. The test order was randomized for each subject, and counterbalanced by reversing the test order in the second session. In total, two SRTs were collected for each processing condition, and each SRT was calculated after presenting 20 short sentences. The sentence material was developed and recorded for Australian use and is based on the Bamford-Kowal-Bench sentences (Bench *et al.*, 1979). According to the AuSTIN, the number of morphemes repeated correctly was recorded and the level of the noise was increased if more than 50% were correct, otherwise the noise level was reduced. The noise level was adjusted by 4 dB for the first four presentations and by 2 dB thereafter. The SRT was calculated from the mean SNR of the final 16 sentences, thus excluding the first four sentences.

The group results were analyzed using 2-way repeated measures analysis of variance (ANOVA) with processing condition and test session as factors. Test session was used as a factor in order to assess the impact of training or acclimatization to the SRT test or to the sound processing involved in each of the conditions. *Post hoc* pairwise comparisons were performed using the Student-Newman-Keuls method.

IV. RESULTS

SRTs for each subject are shown in Fig. 6 for the six processing conditions evaluated, while the SRT benefit relative to the Beam baseline is shown in Fig. 7 for each of the five SpatialNR processing conditions summarized for the group of 12 subjects. The benefit relative to Beam was computed by subtracting the SpatialNR SRT from the Beam SRT for each individual, and is shown in Fig. 7 with the standard error of the mean as error bars.

Analysis of the group results using a two-way ANOVA with processing condition and test session as factors revealed a significant main effect of processing condition ($F^{5,55} = 41.69$, $P < 0.001$), no significant main effect of test session ($P = 0.324$), and no significant interaction term (processing condition × test session, $P = 0.984$). This indicates that training or acclimatization was unlikely to have affected the SRTs recorded at the two test sessions.

Post hoc pairwise analysis showed that all SpatialNR conditions were significantly better than Beam ($P < 0.001$) regardless of the bias parameter value or whether the SpatialNR filter weights were applied to the standard or zoom signal. SpatialNR (S,0) showed 2.7 dB [standard error (SE) 0.4] benefit over Beam ($P < 0.001$). Changing the signal to which the filter weights were applied from standard to zoom provided a 1.5 dB (SE 0.4, $P < 0.001$) improvement such that SpatialNR (Z,0) was 4.2 dB (SE 0.2) better than Beam ($P < 0.001$). The effect of changing the bias parameter from −3 dB to +3 dB resulted in a performance benefit of 1.2 dB (SE 0.5, $P < 0.05$) while all other comparisons of bias parameter settings failed to reach a significant difference. A bias value of +3 dB ($\alpha = 0.5$) showed the largest improvement over the Beam condition of 4.6 dB (SE 0.4, $P < 0.001$).

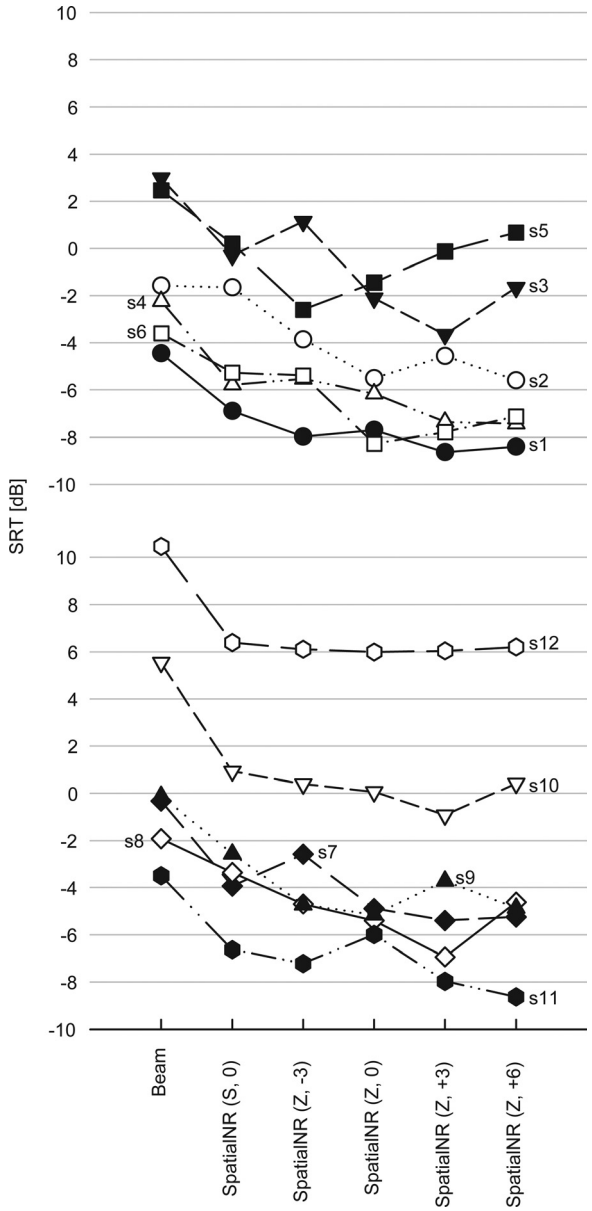

FIG. 6. Individual SRT results in spatially separated 4-talker babble from the rear hemi-field for Beam and each of the five SpatialNR conditions. Lower scores indicate improved performance.

Results at the individual level show there was a range of different performers within the group, indicated by the range of SRTs recorded for the baseline Beam condition. The range covered more than 15 dB, from −4.4 dB to +10.5 dB. A Pearson product moment correlation was performed to analyze the relationship between the baseline Beam SRT and the improvement due to each of the SpatialNR conditions, however, no statistically significant relationship was found [$P = 0.065$ SpatialNR (S,0), $P = 0.137$ SpatialNR (Z,−3), $P = 0.067$ SpatialNR (Z,0), $P = 0.452$ SpatialNR (Z,+3), $P = 0.895$ SpatialNR (Z,+6)].

In order to analyze the relationship between bias value and test SNR, a Pearson product moment correlation was performed between the Beam SRT and the bias value which

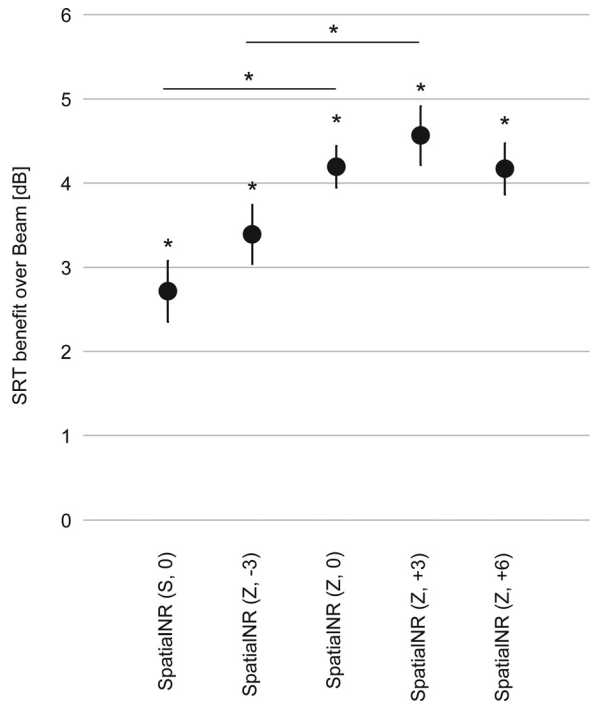

FIG. 7. SRT benefit relative to Beam in spatially separated 4-talker babble from the rear hemi-field. The mean of the difference scores (Beam-Processing Condition) are shown and error bars indicate the standard error of the mean. Stars indicate a statistically significant difference for the group between mean SRTs ($P < 0.001$).

resulted in the largest improvement for each subject, however, no significant correlation was found ($P = 0.189$).

V. DISCUSSION

The SpatialNR (Z,+3) algorithm showed a statistically significant improvement in SRTs of 4.6 dB over the Beam algorithm. Given that Beam has previously been shown to provide a 5.0 dB improvement in SRTs compared to the standard directionality setting in a different group of subjects (Hersbach *et al*., 2012), the overall improvement of SpatialNR over the standard setting is likely to have increased speech intelligibility from 0% to 100% in the noise environment used for evaluation, and so could provide substantial improvements for cochlear implant users in commonly encountered noisy situations.

The main reason for superior performance of SpatialNR over Beam is the ability to remove multiple interfering noise sources at different locations simultaneously. Beam must adapt a filter to minimize the total energy in the output and while it is possible to obtain very large attenuation of a single masker, the effectiveness is reduced when multiple noise sources are presented simultaneously from different locations (van Hoesel and Clark, 1995; Hamacher *et al*., 1997; Wouters and Vanden Berghe, 2001). In contrast, the SpatialNR algorithm can cancel many sources simultaneously from many locations because all sources contribute to a common SNR estimate, which is used to remove the noise.

Two variable aspects to the SpatialNR algorithm were investigated: changing the signal to which the filter weights were applied (standard/zoom) and the bias parameter, α. For all variants, the performance was significantly better than Beam, and SpatialNR (Z,+3) demonstrated the best improvement of 4.6 dB SRT.

The choice of signal to which the SpatialNR gains were applied significantly affected speech intelligibility, with the choice of zoom being 1.5 dB better than standard. This benefit is due to the spatial preprocessing of the zoom signal, which has a directivity pattern that attenuates signals from the rear more heavily than the standard configuration. The difference is smaller than measured in a direct comparison between standard and zoom directionality without the SpatialNR post-filter, which resulted in a 3.9 dB difference in SRT in the same noise environment (Hersbach *et al*., 2012). This indicates that there is some overlap in the noise removed by zoom directionality and SpatialNR, and indicates some redundancy in the information used by the two algorithms.

The bias parameter value had some impact on performance and, although not all parameter values that were evaluated resulted in statistically significant performance differences compared with one another, the maximum benefit over the Beam condition was found to be 4.6 dB with bias value of +3 dB ($\alpha = 0.5$). The bias value has the effect of changing the aggressiveness of noise reduction. As the bias value (in dB) is increased, the amount of noise reduction is increased, but more speech is also removed as a consequence leading to speech distortion. Adjusting the bias value provides a means of counterbalancing the relative amount of noise reduction and speech distortion in the output signal. A bias value of −3 dB ($\alpha = 2$) was significantly worse than a bias value of +3 dB, an expected outcome given the degree of noise attenuation is reduced, allowing more noise to pass through the system. Although there was no significant difference between other values of the bias parameter, increasing the bias value beyond +6 dB is likely to lead to decreased performance due to excessive speech distortion, although this hypothesis cannot be verified with the present data. In other noise environments where there are interfering maskers closer to the target signal, a bias parameter value greater than +6 dB may be suitable to achieve maximum noise reduction while maintaining an acceptable degree of speech distortion.

The large improvements demonstrated are expected to have significant benefit for cochlear implant users since the noise environment under which the evaluation took place can be considered a reasonably difficult condition for understanding speech, and may simulate some difficult real-world environments. During adaptive SRT testing, the SpatialNR algorithm was evaluated across a wide range of SNRs spanning more than 15 dB due to the different cochlear implant users within the group. No significant relationship between the baseline Beam SRT and SpatialNR SRTs was found which indicates the test SNR did not have a significant impact on the performance benefit obtained with SpatialNR. In addition, no significant relationship was found between the baseline Beam SRT and the bias value which provided maximum benefit for each subject, indicating that the best bias parameter was not dependent on the test SNR. It appears the algorithm is robust to changes in the input SNR supporting its suitability for a wide range of listening environments.

However, the locations of interfering talkers were restricted to the rear hemi-field and, since both Beam and SpatialNR algorithms are based on spatial filtering, it is expected that the benefit would be reduced when maskers are located closer to the target and would be least effective when the target and masker are co-located. In this instance, single-channel noise reduction is a more viable alternative, and has been shown to provide speech intelligibility benefit, at least in situations where the masker is less modulated than the target signal (Dawson *et al*., 2011; Hersbach *et al*., 2012).

The sound-treated room used for presentation of the noisy environment had low reverberation, and it is well known that adaptive beamformer performance declines as reverberation increases. This is directly linked to the spatial response of the noise reference, which ideally has a deep null in the target direction but under the effects of reverberation the null depth is reduced. Further evaluation of the SpatialNR algorithm in the presence of reverberation will assist in assessing the likely effect on day-to-day use of the algorithm in real-world environments.

Although substantial noise reduction performance was demonstrated, such aggressive and successful noise reduction may have some detrimental outcomes in daily situations. This may become apparent when the target signal is not directly in front of the listener, or when sound from behind is not considered entirely as interfering noise. Common situations occur when a cochlear implant user is interested in signals both from in front and behind. SpatialNR is unlikely to be suitable for this kind of scenario and, in order to become effective in a commercial device, manual or automatic switching of the algorithm is likely to be required to accommodate the wishes of the user.

No effect of training or acclimatization was measured across the two test sessions for any of the processing conditions evaluated in this study. This indicates that training is unlikely to have played a significant role in the SRT data gathered in this study. However, listening experience is known to play a key role affecting cochlear implant performance when new sound processing strategies are evaluated, and further exposure to the test processing conditions could lead to increased benefit for individual listeners.

The SpatialNR algorithm has been developed specifically for use in cochlear implant sound processing, but may be transferrable to acoustic applications such as hearing aids. Two similar algorithms based on spatial filtering were evaluated in acoustic applications and were found to give superior performance compared to the main beamformer (Cao *et al*., 1996), and compared to beamformer with Wiener post-filter (Wolff and Buck, 2008). This indicates the SpatialNR algorithm may be used in acoustic applications with a likelihood of success. However, the bias parameter used to tune the aggressiveness of noise reduction, which changes the relative amounts of speech distortion and noise reduction, would probably need to be tuned differently for acoustic applications. For example, Brungart *et al*. (2006) showed that normal hearing listeners perform best with a negative gain threshold, while cochlear implant listeners perform best with a positive gain threshold (Mauger *et al*., 2012). This indicates that acoustic applications may need a less aggressive (lower) bias value, maintaining speech quality at the expense of noise reduction performance. Another issue that relates to acoustic processing, not investigated in the present study, is reconstruction of the time-domain audio signal through the use of an inverse FFT. Cochlear implant processing requires no such transformation because the frequency channel energy is used to determine the level of electric stimulation delivered by the cochlear implant electrodes. Reconstruction of a time-domain signal could introduce unwanted artifacts if the noise reduction gains change too quickly over time or across neighboring frequencies.

VI. CONCLUSION

A multiple-microphone noise reduction algorithm (SpatialNR) was evaluated in a complex noisy environment where the noise sources changed location during the test. Speech intelligibility tests with cochlear implant recipients revealed a significant benefit of SpatialNR processing compared to a commercially available adaptive beamformer. This was likely due to the design of SpatialNR, which allowed simultaneous removal of multiple noise sources that is not possible with the adaptive beamformer. The bias parameter of the SpatialNR algorithm provided a mechanism for trading off speech distortion and noise reduction, and was used to control the aggressiveness of noise reduction. The benefit obtained with the experimental algorithm was large, comparable to the benefit observed when a second microphone was introduced into Nucleus cochlear implant systems.

ACKNOWLEDGMENTS

Many thanks to the cochlear implant recipients who volunteered their time to participate in this study. Thanks to Pam Dawson for statistical advice and Jennie Gorrie for data collection. This research was carried out at Cochlear Ltd, Research and Applications Laboratory, Melbourne, and was financially supported in part by the HEARing CRC established and supported under the Australian Government's Cooperative Research Centre's program. The Bionics Institute acknowledges the support it receives from the Victorian Government through its Operational Infrastructure Support Program.

Bench, J., Kowal, A., and Bamford, J. (**1979**). "The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children," Br. J. Audiol. **13**, 108–112.

Blamey, P. J. (**2005**). "Adaptive dynamic range optimization (ADRO): A digital amplification strategy for hearing aids and cochlear implants," Trends Amplif. **9**, 77–98.

Brungart, D. S., Chang, P. S., Simpson, B. D., and Wang, D. (**2006**). "Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation," J. Acoust. Soc. Am. **120**, 4007–4018.

Buechner, A., Brendel, M., Saalfeld, H., Litvak, L., Frohne-Buechner, C., and Lenarz, T. (**2010**). "Results of a pilot study with a signal enhancement algorithm for HiRes 120 cochlear implant users," Otol. Neurotol. **31**, 1386–1390.

Burkhard, M. D., and Sachs, R. M. (**1975**). "Anthropometric manikin for acoustic research," J. Acoust. Soc. Am. **58**, 214–222.

Cao, Y. C., Sridharan, S., and Moody, M. (**1996**). "Speech enhancement using microphone array with multi-stage processing," IEICE Trans. Fundamentals **E79A**, 386.

Chung, K., and Zeng, F. G. (**2009**). "Using hearing aid adaptive directional microphones to enhance cochlear implant performance," Hear. Res. **250**, 27–37.

Chung, K., Zeng, F. G., and Acker, K. N. (**2006**). "Effects of directional microphone and adaptive multichannel noise reduction algorithm on cochlear implant performance," J. Acoust. Soc. Am. **120**, 2216–2227.

Dawson, P. W., Hersbach, A. A., and Swanson, B. A. (**2013**). "An adaptive Australian Sentence Test In Noise (AuSTIN)," Ear Hear, in press.

Dawson, P. W., Mauger, S. J., and Hersbach, A. A. (**2011**). "Clinical evaluation of signal-to-noise ratio-based noise reduction in Nucleus(R) cochlear implant recipients," Ear Hear. **32**, 382–390.

Gifford, R. H., and Revit, L. J. (**2010**). "Speech perception for adult cochlear implant recipients in a realistic background noise: effectiveness of preprocessing strategies and external options for improving speech recognition in noise," J. Am. Acad. Audiol. **21**, 441–451.

Griffiths, L., and Jim, C. (**1982**). "An alternative approach to linearly constrained adaptive beamforming," IEEE Trans. Antennas Propag. **30**, 27–34.

Hamacher, V., Doering, W. H., Mauer, G., Fleischmann, H., and Hennecke, J. (**1997**). "Evaluation of noise reduction systems for cochlear implant users in different acoustic environment," Am. J. Otol. **18**, S46–S49.

Hersbach, A. A., Arora, K., Mauger, S. J., and Dawson, P. W. (**2012**). "Combining directional microphone and single-channel noise reduction algorithms: A clinical evaluation in difficult listening conditions with cochlear implant users," Ear Hear. **33**, e13–e23.

Hu, Y., and Loizou, P. C. (**2010**). "Environment-specific noise suppression for improved speech intelligibility by cochlear implant users," J. Acoust. Soc. Am. **127**, 3689–3695.

Kompis, M., Bertram, M., Senn, P., Muller, J., Pelizzone, M., and Hausler, R. (**2008**). "A two-microphone noise reduction system for cochlear implant users with nearby microphones - Part II: Performance evaluation," EURASIP J. Advances Signal Process. **2008**, 451273.

Leukimmiatis, S., Dimitriadis, D., and Maragos, P. (**2006**). *An Optimum Microphone Array Post-Filter for Speech Applications* (International Speech Communication Association, Bonn, Germany), pp. 2142–2145.

Mauger, S. J., Dawson, P. W., and Hersbach, A. A. (**2012**). "Perceptually optimized gain function for cochlear implant signal-to-noise ratio based noise reduction," J. Acoust. Soc. Am. **131**, 327–336.

McCowan, I. A., and Bourlard, H. (**2003**). "Microphone array post-filter based on noise field coherence," IEEE Trans. Speech Audio Process. **11**, 709–716.

McDermott, H. J., McKay, C. M., and Vandali, A. E. (**1992**). "A new portable sound processor for the University of Melbourne/Nucleus Limited multielectrode cochlear implant," J. Acoust. Soc. Am. **91**, 3367–3371.

Meyer, J., and Simmer, K. U. (**1997**). "Multi-channel speech enhancement in a car environment using Wiener filtering and spectral subtraction," in *International Conference on Acoustics, Speech, and Signal Processing*, April, Munich, Germany, pp. 1167–1170.

Patrick, J. F., Busby, P. A., and Gibson, P. J. (**2006**). "The development of the Nucleus Freedom Cochlear implant system," Trends Amplif. **10**, 175–200.

Spriet, A., Van Deun, L., Eftaxiadis, K., Laneau, J., Moonen, M., van Dijk, B., van Wieringen, A., and Wouters, J. (**2007**). "Speech understanding in background noise with the two-microphone adaptive beamformer BEAM in the Nucleus Freedom Cochlear implant system," Ear Hear. **28**, 62–72.

Vandali, A. E., Whitford, L. A., Plant, K. L., and Clark, G. M. (**2000**). "Speech perception as a function of electrical stimulation rate: Using the Nucleus 24 cochlear implant system," Ear Hear. **21**, 608–624.

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

Ward, D. B., Kennedy, R. A., and Williamson, R. C. (**2001**). "Constant directivity beamforming," in *Microphone Arrays: Signal Processing Techniques and Applications*, edited by M. Brandstein and D. Ward (Springer Verlag, Berlin).

Wolfe, J., Schafer, E. C., John, A., and Hudson, M. (**2011**). "The effect of front-end processing on cochlear implant performance of children," Otol. Neurotol. **32**, 533–538.

Wolff, T., and Buck, M. (**2008**). "Spatial maximum *a posteriori* post-filtering for arbitrary beamforming," in *Hands-Free Speech Communication and Microphone Arrays*, 2008, pp. 53–56.

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

Yousefian, N., and Loizou, P. C. (**2013**). "A dual-microphone algorithm that can cope with competing-talker scenarios," IEEE Trans. Audio Speech Lang. Process. **21**, 145–155.

Zelinski, R. (**1988**). "A microphone array with adaptive post-filtering for noise reduction in reverberant rooms," in *International Conference on Acoustics, Speech, and Signal Processing*, 1988, Vol. 2575, pp. 2578–2581.