

Two-Microphone Spatial Filtering Improves Speech Reception for Cochlear-Implant Users in Reverberant Conditions With Multiple Noise Sources

Raymond L. Goldsworthy¹

Trends in Hearing
1–13
© The Author(s) 2014
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/2331216514555489
tia.sagepub.com


Abstract

This study evaluates a spatial-filtering algorithm as a method to improve speech reception for cochlear-implant (CI) users in reverberant environments with multiple noise sources. The algorithm was designed to filter sounds using phase differences between two microphones situated 1 cm apart in a behind-the-ear hearing-aid capsule. Speech reception thresholds (SRTs) were measured using a Coordinate Response Measure for six CI users in 27 listening conditions including each combination of reverberation level ($T_{60} = 0, 270, \text{ and } 540 \text{ ms}$), number of noise sources (1, 4, and 11), and signal-processing algorithm (omnidirectional response, dipole-directional response, and spatial-filtering algorithm). Noise sources were time-reversed speech segments randomly drawn from the Institute of Electrical and Electronics Engineers sentence recordings. Target speech and noise sources were processed using a room simulation method allowing precise control over reverberation times and sound-source locations. The spatial-filtering algorithm was found to provide improvements in SRTs on the order of 6.5 to 11.0 dB across listening conditions compared with the omnidirectional response. This result indicates that such phase-based spatial filtering can improve speech reception for CI users even in highly reverberant conditions with multiple noise sources.

Keywords

cochlear implants, speech enhancement, noise reduction, spatial filtering, reverberation

Introduction

There is a substantial history regarding the use of multiple microphones to spatially filter sounds based on their direction of arrival (Brandstein & Ward, 2001; Kokkinakis, Azimi, Hu, & Friedland, 2012). The most straightforward approach is to linearly combine microphone signals after amplitude scaling and phase shifting. This approach is linear and time-invariant and can be used to generate known directional responses such as cardioid- and dipole-response patterns. Such approaches are generally referred to as beamforming algorithms, or more specifically as delay-and-sum beamforming algorithms. Delay-and-sum beamformers can also be implemented in analog, multiport, microphone circuitry to produce directional microphones. It has been demonstrated that delay-and-sum beamforming can improve speech reception for cochlear-implant (CI) users in noisy environments (Chung & Zeng, 2009; Chung, Zeng, & Acker, 2006; Chung, Zeng, & Waltzman, 2004). The performance gain provided by delay-and-sum

beamforming is robust in reverberant environments (Chung, 2004; Desloge, Rabinowitz, & Zurek, 1997).

Arising from this class of algorithms, the closely related null-steering beamforming was devised to slowly (i.e., $>100 \text{ ms}$) adapt the beamformer-response pattern to steer spatial nulls (Frost, 1972; Griffiths & Jim, 1982; Widrow et al., 1975). The BEAMTM algorithm (Hersbach, Arora, Mauger, & Dawson, 2012; Spriet et al., 2007) implemented on Cochlear Corporation CI devices is an example of a null-steering beamformer. It has been shown that in comparison with omnidirectional processing, null-steering beamforming can improve speech reception thresholds (SRTs) in noisy environments by 7 to 16 dB in terms of the

¹Sensmetrics Corporation, Malden, MA, USA

Corresponding author:

Raymond L. Goldsworthy, Research and Development, Sensmetrics Corporation, 14 Summer Street, Suite 305, Malden, MA 02148, USA.
Email: raygold@sens.com



signal-to-noise ratio (SNR) at which listeners achieve 50% correct on speech reception measures (Spriet et al., 2007; Wouters & Vanden Berghe, 2001). Kokkinakis and Loizou (2010) demonstrated that such null-steering beamformers could be integrated—in a second-order null-steering operation—across ears yielding additional benefits. More generally, however, it has been repeatedly demonstrated that the performance of null-steering beamformers rapidly degrades with increasing reverberation and increasing number of noise sources (Greenberg & Zurek, 1992; Hamacher, Doering, Mauer, 1997; van Hoesel & Clark, 1995; Wouters & Vanden Berghe, 2001).

The spatial-filtering algorithm evaluated in this article is distinct from these aforementioned algorithms. Specifically, rather than slowly steering spatial nulls to produce a desired directional response, the algorithm uses rapid spectral-temporal analysis to determine which spectral-temporal components are dominated by target or by noise energy and then attenuate components accordingly. The general approach behind such spatial filtering was inspired by models of spatial hearing (Jeffress, 1948) leading to the pioneering signal-processing work of Kollmeier and coworkers (Kollmeier & Koch, 1994; Kollmeier, Peissig, & Hohmann, 1993) in which they demonstrated that a small, but robust, improvement in SNR could be achieved based on models of binaural interaction. Subsequent works on binaurally inspired spatial filtering have demonstrated speech reception benefits for CI users in noisy environments as large as 60 percentage points on keyword recognition and as large as 14 dB in SRT (Goldsworthy, 2005; Margo, Schweitzer, & Feinman, 1997).

A systematic study that included a delay-and-sum beamformer, two null-steering beamformers, a frequency-domain minimum-variance distortionless-response (FMV) beamformer, and the aforementioned spatial filtering of Kollmeier et al. (1993) demonstrated that the performance of the Kollmeier and FMV methods are relatively robust to reverberation compared with null-steering beamformers (Lockwood et al., 2004). They concluded that if distortionless response is required, then the FMV method is preferred, but that the Kollmeier method is also promising if maximal intelligibility is preferred.

Goldsworthy, Delhorne, Desloge, and Braidia (2014) transitioned the binaurally inspired spatial filtering of Kollmeier to an algorithm based on closely spaced microphones. They demonstrated that phase-based spatial filtering using microphones situated 1 cm apart in a behind-the-ear capsule could provide speech reception benefits between 6 and 11 dB SRT compared with an omnidirectional response when tested in a moderately reverberant environment ($T_{60} = 350$ ms). Yousefian, Kokkinakis, and Loizou (2010) and Yousefian and

Loizou (2012, 2013) developed an algorithm that used a similar spectral-temporal analysis structure using closely spaced microphones but based their attenuation function on coherence rather than directly on phase differences. In their 2013 study, they found that the benefits provided by such a coherence-based algorithm deteriorated in reverberant conditions. Specifically, they found SRT benefits of 5 to 10 dB in an anechoic condition, but benefits decreased to 4 to 7 dB and to 1 to 2 dB when tested in rooms with $T_{60} = 220$ and 465 ms, respectively.

Hersbach, Grayden, Fallon, and McDermott (2013) suggested that it was the use of coherence which caused the Yousefian and Loizou (2013) algorithm to deteriorate in reverberation; consequently, they introduced an alternate approach, using null-steering beamforming as a front end to a secondary postfilter using spectral attenuation of low SNR components. As they only evaluated this method in a sound-treated low-reverberation environment, it is unknown the extent to which that approach might be affected by higher levels of reverberation.

The spatial-filtering algorithm introduced by Goldsworthy et al. (2014) has demonstrated a degree of robustness in reverberation. The purpose of the present article is to systematically examine the effects of reverberation and the number of noise sources on the performance of that algorithm (Fennec). A study was conducted using a room simulation method (Peterson, 1986; Shinn-Cunningham, Desloge, & Kopčo, 2001) to precisely control reverberation levels and sound-source locations while maintaining all other environmental variables. Listening conditions were evaluated using combinations of reverberation level ($T_{60} = 0, 270,$ and 540 ms), number of noise sources (1, 4, and 11), and signal processing (omnidirectional response, dipole-directional response, and Fennec algorithm). The results indicate the extent to which this spatial-filtering algorithm continues to provide speech reception benefits for CI users in the presence of such environmental degradations.

Methods

Subjects

Six adult CI users participated in this study with relevant information summarized in Table 1. All subjects had previously participated in at least one other speech reception experiment in our laboratory. Subjects provided informed consent on their first visit to the laboratory and were paid for their participation in the study.

At the time of testing, the subjects ranged in age from 26 to 75 years (mean = 48.7 years) with five of the six subjects reporting that the cause of their hearing loss was either genetic or unknown. In these cases, the loss was typically diagnosed at birth or early childhood and

Table 1. Subject Information.

Subject	Sex	Ear tested	Age at onset of hearing loss/deafness	Etiology	Age at implantation (years)	Age at time of testing	Implant processor
S1	F	Right	Birth-progressive	Genetic	20	26	Freedom
S2	F	Right	30 s-progressive	Unknown	49	56	Harmony
S3	M	Left	50 s-progressive	Unknown	71	75	Freedom
S4	F	Right	Birth-progressive	Genetic	24	29	Freedom
S5	M	Left	12-sudden	Meningitis	13	38	Freedom
S6	F	Right	40 s-progressive	Unknown	55	68	Freedom

progressed over time. All subjects had received their CI as an adult and the age at implantation ranged from 20 to 71 years (mean = 38.7 years). All subjects had at least 4 years of CI experience. Five of the six subjects used Cochlear Freedom processors while subject S2 used the Advanced Bionics Harmony processor. Subject S5 was an N22 recipient and, consequently, used the lower stimulation rate SPEAK sound-processing strategy. All subjects were tested monaurally. Subject S4 was a bilateral implant user who used her self-selected better ear for the experiment.

Stimuli

SRTs were measured using the Coordinate Response Measure (CRM) speech materials (Bolia, Nelson, Ericson, & Simpson, 2000). The CRM speech materials consists of phrases in the form “Ready *call sign*, go to *color number* now” where *call sign* can be assigned eight different values (Arrow, Baron, Charlie, Eagle, Hopper, Laker, Ringo, or Tiger), *color* can be assigned four different values (blue, green, red, and white), and *number* can be assigned eight different values (integers between 1 and 8) yielding a total of 256 phrases. The CRM speech materials contain each of these 256 phrases each recorded for four male and four female talkers; however, for the present study, only a single male talker was used for testing. The rationale for only using one talker, rather than randomly varying the selected talker, is that, generally speaking, listeners in the real world generally know who they are talking to when they are trying to comprehend speech. There are counter examples, for instance, a listener might not know who the talker is when answering a telephone call; however, it is much more common that a listener is aware of who is speaking. Consequently, we did not want talker identification to be inherently part of the speech reception task.

The noise stimuli consisted of time-reversed speech fragments taken from recordings of IEEE sentences (Rothauser et al., 1969) made at House Research

Institute. The entire IEEE database was concatenated into a single audio file. For each unique masking noise source, a segment was randomly selected from the concatenated audio file with a duration matching the target speech stimulus. Specifically, for each trial and each unique noise position, distinct segments were selected from the database and then time reversed.

Room Simulation

A room simulation was used to provide precise control over acoustic conditions, specifically, allowing variations in reverberation while holding other aspects of the environment constant. Simulated impulse responses were generated using a modified version of the image method (e.g., see Peterson, 1986) and is identical to the method of Shinn-Cunningham et al. (2001). This room simulation method was used to generate head-related transfer functions (HRTFs) for sound-source locations within a simulated room that measured $4 \times 4 \times 4$ m. Sound-source locations and microphone positions will be described using a coordinate system having (0, 0, 0) m as the center of this room and axes parallel to the walls. The listener’s head was simulated as a sphere located in the center of this room, (0, 0, 0) m, with an 8.5 cm radius. The sound source locations were specified at the vertices of a regular icosahedron centered at (0, 0, 0) m designed such that each of the 12 sound sources were 1 m distant from the center of the simulated head. Specifically, the sound locations were the 12 combinations of

$$\left(\frac{\pm g}{\sqrt{g^2+1}}, 0, \frac{\pm 1}{\sqrt{g^2+1}} \right), \quad \left(0, \frac{\pm 1}{\sqrt{g^2+1}}, \frac{\pm g}{\sqrt{g^2+1}} \right), \quad \text{and} \\ \left(\frac{\pm 1}{\sqrt{g^2+1}}, \frac{\pm g}{\sqrt{g^2+1}}, 0 \right) \text{ m, where } g = 1 + (\sqrt{5})/2. \text{ The micro-}$$

phone locations were specified to be on the left side of the simulated head with 1 cm spacing and situated to be in an end-fire configuration pointed at the sound source located at $\left(\frac{g}{\sqrt{g^2+1}}, 0, \frac{1}{\sqrt{g^2+1}} \right)$ m. In other words, the simulated head in the center of the room was facing

the target source which was 1 m distant with a 0° azimuthal angle and an approximately 31.7° angle of elevation. The five closest sound sources were equally distributed with 63.4° angles from the axis connecting the simulated head to the target sound location (look direction). Likewise, there were five sound sources in the rear hemisphere equally distributed at angles of 126.9° from the look direction. The remaining sound source was directly behind (180°) the simulated head. The absorption coefficients of the walls of this simulated room were adjusted to control the level of room reverberation.

The absorption coefficients used were 1 (anechoic), 0.4 (mildly reverberant), and 0.2 (highly reverberant); the respective reverberation times for the reverberant energy to decay 60 dB were $T_{60}=0, 270,$ and 540 ms, respectively. HRTFs were generated at a sample rate of 22050 Hz and consisting of 16,384 samples.

Target stimuli from the CRM speech materials were always convolved with the HRTF corresponding to the simulated sound source located at $(g/\sqrt{g^2+1}, 0, 1/\sqrt{g^2+1})$ m. Depending on the listening condition, noise stimuli from the IEEE sentences were convolved with corresponding HRTFs from simulated sound-source locations which could be any of the other 11 sound-source locations (i.e., noise sources were never colocated with the target location). When using multiple noise sources, noise sources were never colocated with each other and were combined after convolution without any additional amplitude scaling. This combined noise stimulus was then added to the target stimulus at the specified SNR.

Signal Processing

In a second step, the outputs of the room simulation were processed to yield a single signal for presentation to the listener. Three signal-processing variations were considered: omnidirectional response, dipole-directional response, and the Fennec algorithm.

For the omnidirectional response, the output of the back signal was simply presented to the subject. In the free field (i.e., not mounted on the simulated sphere), the directional response of this processing would be identical from all directions. Mounted on the simulated head, it yielded no improvements in the target SNR other than those achieved naturally by head shadow.

For the dipole-directional response, the front and back signals were combined to yield a dipole-directional response that was presented to the subject. The upper branch of the signal processing in Figure 1 depicts the processing that was used to generate the dipole response. Specifically, the short-time Fourier transform (STFT)—based on a 256-point analysis window (11.6 ms) and 50% block overlap—was calculated for each of the two signals. The STFT of the back signal was subtracted from the STFT of the front signal, and frequency-dependent compensation (up to a maximum of 18 dB in order to limit noise amplification) was applied to counteract the low-frequency attenuation that resulted from the subtraction (Stadler & Rabinowitz, 1993). The inverse STFT was then taken to yield the processed output. (Note that the phase-based attenuation processing shown in Figure 1 was part of Fennec algorithm

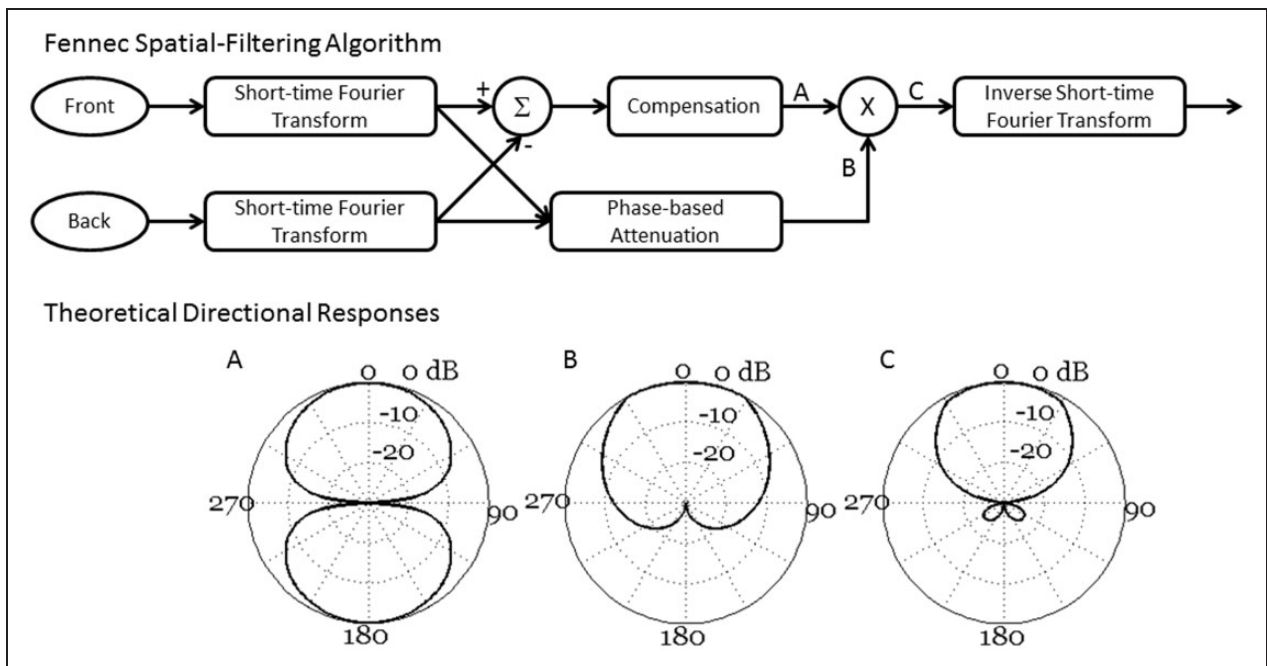


Figure 1. Fennec processing schematic and theoretical directional responses at three stages indicated by A, B, and C.

described later and is not used for the dipole-directional response.)

In the free field (i.e., when not head worn), the directional response of this processing has the pattern shown as theoretical directional response A of Figure 1 in which lateral sources from $\pm 90^\circ$ are attenuated while sources from the front and back are preserved.

For the Fennec algorithm, as shown in the schematic of Figure 1, the front and back signals were used to form a dipole-directional response as described earlier, and then an additional phase-based attenuation was generated from the two signals, which was applied to the dipole output to further reduce noise. The phase-based attenuation was calculated in the following manner. First, the time-dependent cross-spectral power density, $S_{fb}[n, k]$, for the front and back signals were calculated for each STFT using

$$S_{fb}[n, k] = \alpha \times S_{fb}[n - 1, k] + (1 - \alpha) \times \text{STFT}_f[n, k] \times \text{conj}(\text{STFT}_b[n, k]) \quad (1)$$

where $\text{STFT}_f[n, k]$ and $\text{STFT}_b[n, k]$ are the front and back signals STFTs, respectively, n is the STFT time index, k is the STFT frequency bin, $\text{conj}()$ is the complex-conjugate operator, and α is the parameter of a first-order infinite-impulse response filter used to smooth the estimate (we used $\alpha = 0.56$, which yielded a filter time constant of 10 ms).

The phase of $S_{fb}[n, k]$ was then used to estimate the direction-of-arrival that dominated the content of time/frequency cell $[n, k]$ of the STFT. Specifically, based on free-field acoustics and assuming that a single directional source from azimuth location θ ($0^\circ = \text{straight ahead}$) accounted for all energy in cell $[n, k]$, then

$$\angle(S_{fb}[n, k]) = \frac{2\pi kd}{Nc} \cos(\theta) \quad (2)$$

where $N = 256$ is the STFT block size, $d = 0.01$ m is the microphone separation, and $c = 345$ m/s is the velocity of sound. By inverting this equation, it was possible to estimate the angle of incidence, $\hat{\theta}$, for a single free-field source that would have given rise to the observed phase.

This estimated angle of incidence was then used to calculate an attenuation factor that was applied to each time/frequency cell of the dipole STFT

$$A[n, k] = \min\left(0, \frac{A_{180}(\theta - \beta)}{180 - \beta}\right) \quad (3)$$

where A_{180} is the desired attenuation in dB at 180° , and β is the angle of incidence below which the attenuation is 0 dB. For the implementation used in this study, A_{180} and β were set to 30 dB and 30° , respectively, to yield

the desired directional response shown as theoretical directional response B of Figure 1. This attenuation was calculated for each STFT time/frequency cell of the input signals and was applied to the dipole STFT prior to signal reconstruction via the inverse STFT. The theoretical, free-field Fennec directional response for a single source, combining the dipole and the phase-based attenuation, is shown as theoretical directional response C of Figure 1.

Procedures

Computer audio was delivered through an ESI U24 XL USB digital audio interface. Prior to testing, subjects were instructed to connect to the USB interface using a mains isolation cable to their CI auxiliary input. Subjects then completed 10 min of listening practice using the CRM materials in quiet. Subjects were allowed to adjust their processor sensitivity during this time but instructed to make no further changes during the testing period. Subjects used their self-selected CI processor program. Because the input to the auxiliary port has already undergone spatialization as well as spatial processing, the input is monaural and does not carry acoustic information that would allow secondary internal spatial filters to provide further compensation.

A total of 27 listening conditions were tested consisting of all combinations of reverberation level ($T_{60} = 0, 270, \text{ and } 540$ ms), number of noise sources (1, 4, and 11), and signal processing (omnidirectional response, dipole-directional response, and Fennec algorithm). The order of these 27 conditions was randomly administered. For each condition, the SRT was measured using a 1-up, 1-down decision rule that theoretically converges to 50% speech reception. For the 1 and 4 number of noise-source conditions, the location of the noise sources was randomly reassigned (never collocated with the target or with each other) for each trial. For all of the noise conditions, unique maskers were convolved with the associated HRTFs, the target speech was convolved with its associated HRTF, and then the spatialized maskers were summed together and added to the spatialized target speech at the desired SNR (as determined at the back microphone location).

For a given trial, a speech phrase was randomly drawn from 256 possible combinations of call sign, color, and number. Speech phrases were drawn with replacement, as the CRM materials do not contain contextual information. The subject was provided a graphical user interface on a personal computer that contained response buttons for the four colors and eight numbers. Reception of the call sign was not scored, as it was determined in pilot testing that call sign identification was significantly lower than either color or number identification. A trial was scored as correct only when both the

color and number were correctly identified. Subjects were given visual feedback indicating when they correctly identified the color or number as the response buttons would flash green for 400ms (as opposed to red for incorrect responses).

The initial SNR was 0dB which was thereafter decreased by 2 dB after each correct response and increased by 2 dB after each incorrect response. This procedure continued for 20 reversals, and the SRT was calculated as the average of the last eight reversals.

Results

The results consisted of 486 SRT measures including all combinations of six subjects, three repetitions, three reverberation levels, three numbers of noise sources, and three signal-processing algorithms. Measured SRTs were analyzed using a three-way repeated-measures analysis of variance with reverberation, number of noise sources, and signal processing as factors. Figure 2 plots the SRTs for each subject averaged across repetitions for each condition. Figure 3 plots the SRTs averaged across subject and repetitions for each condition.

Reverberation was significant, $F(2, 10) = 58.6$, $p < .001$. Illustrating the effect of reverberation, SRTs averaged across subjects on the 1 noise-source, omnidirectional response, conditions increased from -2.5 to 0.0

to 6.0 dB as T_{60} increased from 0 to 270 to 540 ms, respectively. The interaction between reverberation and the number of noise sources was significant, $F(4, 20) = 5.7$, $p = .003$, with reverberation generally affecting performance more in the single noise-source condition. The interaction between reverberation and signal processing was not significant, $F(4, 20) = .55$, $p = .70$, indicating that while reverberation decreased speech reception for all conditions, the average benefit provided by the algorithms was not significantly affected by reverberation. Figure 4 illustrates this by plotting the average SRT benefits for the dipole-directional and Fennec algorithms compared with the omnidirectional response for each reverberation level. The SRT benefits were calculated as the SRT difference between algorithms averaged across subjects, repetitions, and number of noise sources. This average benefit for the Fennec algorithm compared with the omnidirectional response decreased from 8.5 to 7.7 to 7.5 dB SNR, which was not a significant decrease ($p > .1$).

The number of noise sources was significant, $F(2, 10) = 19.3$, $p < .001$, as was the interaction between the number of noise sources and signal processing, $F(4, 20) = 10.8$, $p < .001$. Illustrating the effect of the number of noise sources, SRTs averaged across subjects on the anechoic, omnidirectional response, conditions increased from -2.5 to 0.8 to 1.7 dB for the 1, 4, and

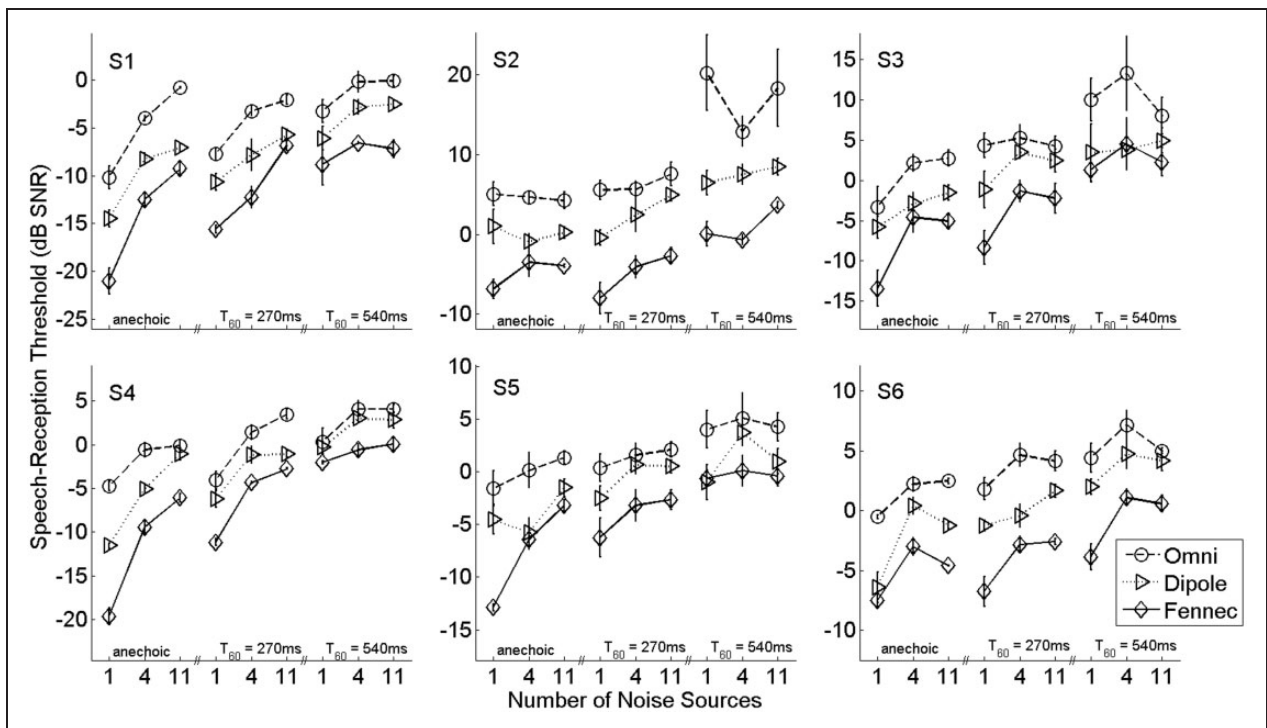


Figure 2. Speech reception thresholds for each subject and each condition averaged across repetitions. Error bars represent 1 standard error of the mean.

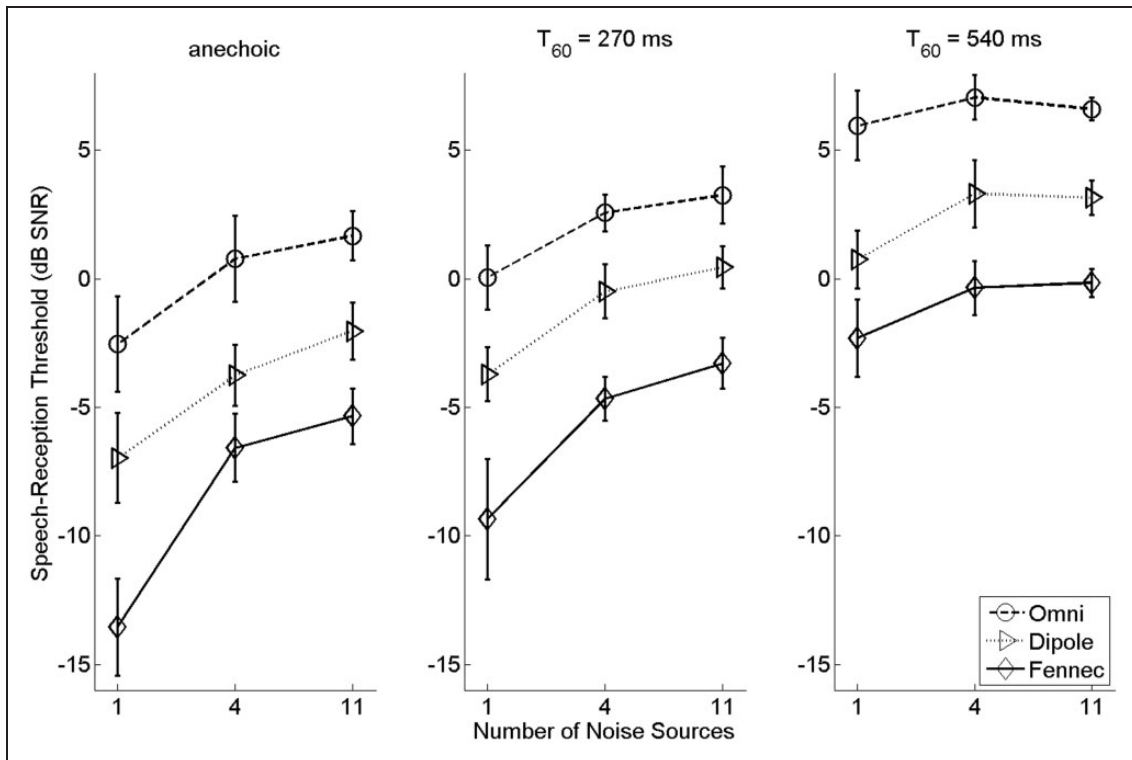


Figure 3. Speech reception thresholds for each condition averaged across subject and repetitions. Error bars represent 1 standard error of the mean.

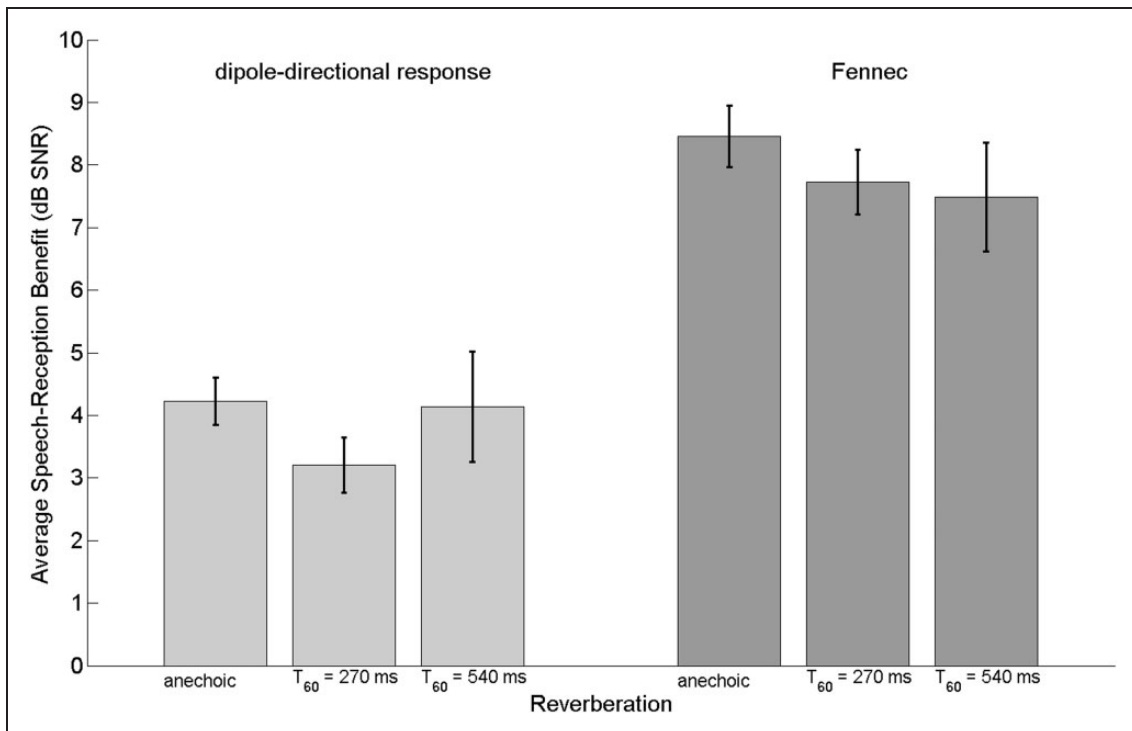


Figure 4. The average speech reception benefit relative to the omnidirectional response for the dipole-directional response and for the Fennec algorithm averaged across subjects, repetitions, and number of noise sources for each reverberation level tested. Error bars represent 1 standard error of the mean.

11 noise-sources conditions, respectively. Figure 5 illustrates this effect with average SRT benefits for the dipole-directional response and Fennec algorithms compared with the omnidirectional response plotted for each number of noise sources. The significant interaction between the number of noise sources and signal processing can be observed in the larger decrease in SRT benefits provided by the Fennec algorithm for the 11 noise-sources condition. Specifically, SRT benefits decreased from 9.6 to 7.3 to 6.8 dB SNR.

Signal processing was significant, $F(2, 10) = 69.8$, $p < .001$, as illustrated by Figure 3 which shows a clear SRT benefit of the dipole-directional response over the omnidirectional response, and a clear SRT benefit of the Fennec algorithm over the dipole-directional response for all conditions tested.

Paired-sample t tests were calculated comparing spatial processing based on the SRT scores averaged across subjects for each of the nine acoustic conditions (i.e., three reverberation levels and three number of noise sources). This post-hoc analysis indicated that SRT performance was significantly ($p < .01$) better with the Fennec algorithm compared with the dipole-directional response for all conditions except for the $T_{60} = 540$ ms and 4 noise-source condition, which only resulted in

$p = .02$. SRT performance was significantly ($p < .01$) better with the dipole-directional response compared with the omnidirectional response for all conditions except for the $T_{60} = 540$ ms and 11 noise-source condition, which only resulted in $p = .03$. In general, the observed spatial processing SRT benefits were robust across conditions but with a small decrease in the most reverberant conditions.

Comparison of Results to Cardioid Response Based on Physical SNR Benefit

The conditions for the preceding experiment were selected to measure the benefit derived from the Fennec algorithm compared with no spatial processing (omnidirectional response) and compared with linear spatial processing (dipole-directional response). The dipole-directional response was selected as the linear spatial processing condition, as the Fennec algorithm contains a dipole-directional response as part of its operation. Thus, the comparison between Fennec and the dipole-directional response indicates the additional benefit derived from the adaptive processing. In practice, the use of cardioid microphones is more common in present-day CIs and hearing aids; consequently, additional

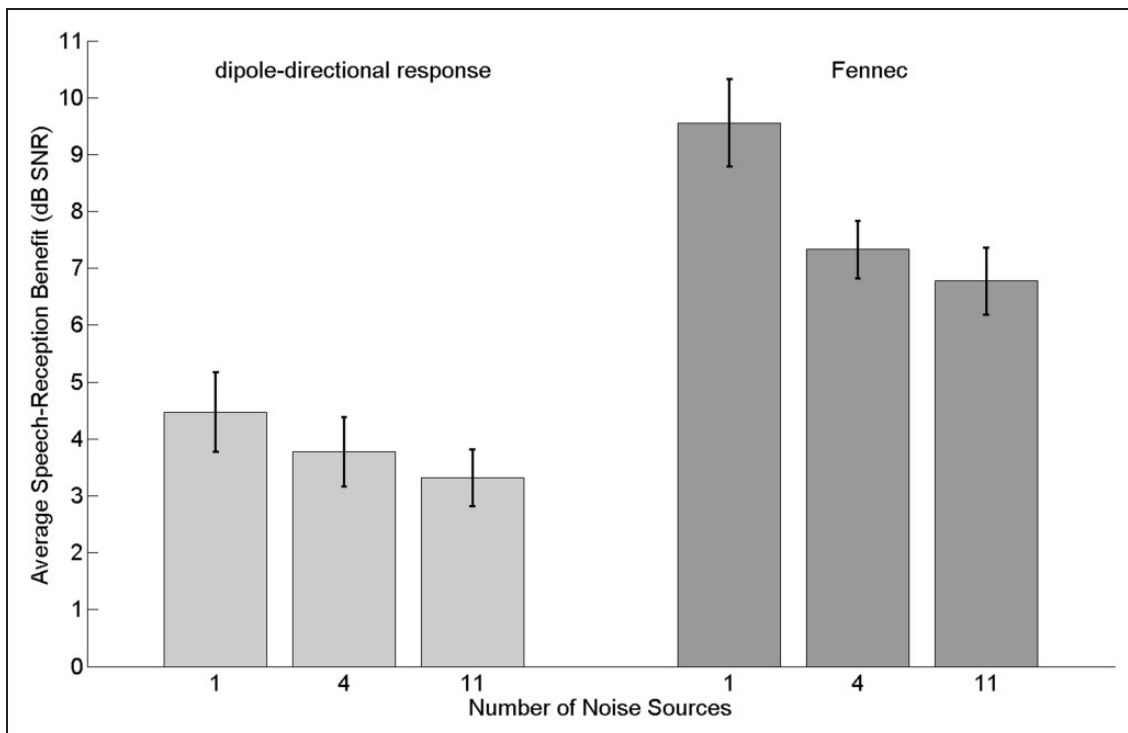


Figure 5. The average speech reception benefit relative to the omnidirectional response for the dipole-directional response and for the Fennec algorithm averaged across subjects, repetitions, and reverberation levels for each number of noise sources tested. Error bars represent 1 standard error of the mean.

analysis was calculated to compare the results of the perceptual experiment with physical SNR benefits derived from a cardioid-directional response.

Analysis was performed by processing speech material through the room simulation as described in the Methods section and then calculating the output SNR benefit relative to an omnidirectional response for a cardioid-directional response, a dipole-directional response, and the Fennec algorithm. Analysis was calculated for three reverberation levels (anechoic, $T_{60} = 270$ ms, and $T_{60} = 540$ ms) combined with 11 different options for the number of environmental noise sources (e.g., the number of simultaneous noise sources ranged between 1 and 11). For each processing condition, a sentence was randomly selected and processed through the room simulation with target speech positioned straight ahead and noise sources generated based on the reverberation time and number of noise sources for the particular condition. As in the experimental methods, the noise sources were time-reversed speech segments randomly drawn from the IEEE database. The input SNR into the spatial processing was set to 0 dB SNR as measured at the back microphone in the room simulation.

The physical SNR was calculated as the average power of the target speech compared with the average power of the noise sources as measured at the output of

the spatial processing. For the cardioid-directional and dipole-directional responses, as those systems are linear, the output SNR was calculated by processing the target speech and noise sources separately. For the Fennec algorithm, as it is nonlinear, the effective output SNR was calculated by processing the combined target speech plus noise through the Fennec system and then freezing the algorithm structure; the target speech and noise were then processed separately through the algorithm with frozen spatial processing weights. For each condition, 1,000 iterations of this procedure were calculated, and the output SNR was taken as the average, in decibels, across iterations.

Figure 6 summarizes the physical SNR benefit provided by the three spatial processing algorithms relative to an omnidirectional response. The primary purpose of this analysis was to compare the SNR benefit achieved by the Fennec algorithm compared with the cardioid-directional response. The measured SNR benefit for the cardioid-directional response always fell between the benefit achieved by the Fennec algorithm and by the dipole-directional response, with performance being closer to the dipole-directional response as both reverberation and number of noise sources increased.

This analysis provides a comparison between the Fennec algorithm with a cardioid-directional response

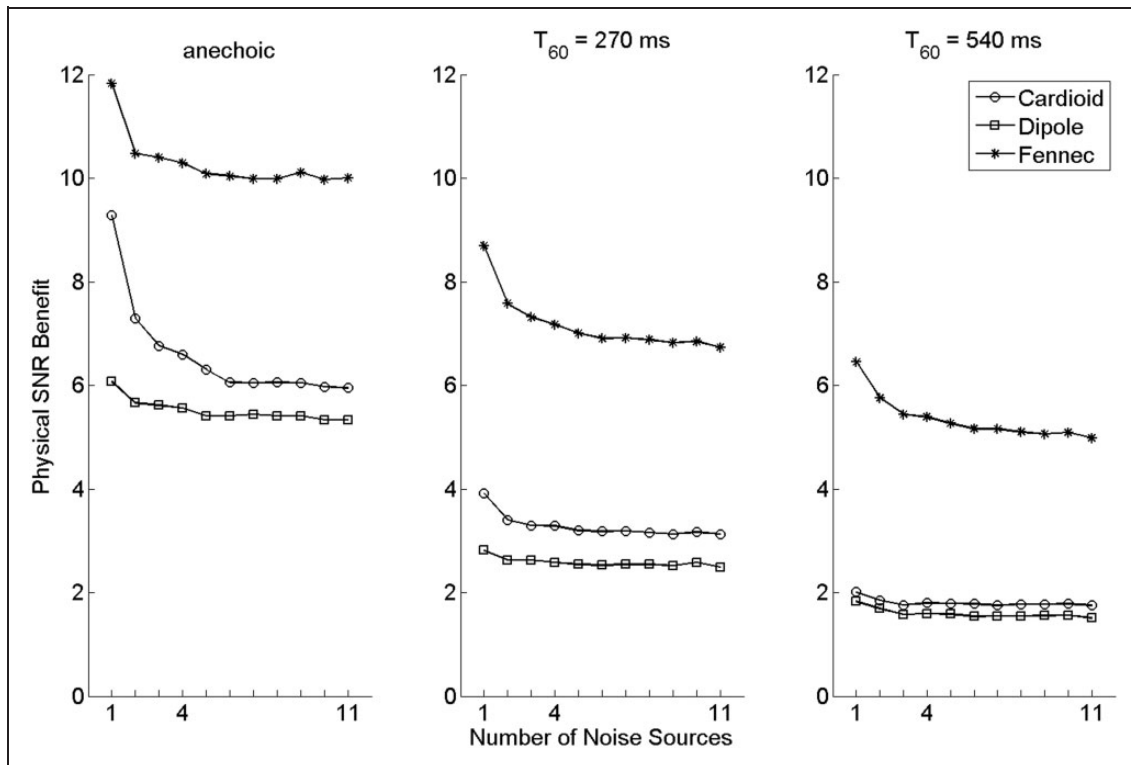


Figure 6. The average physical signal-to-noise ratio benefit relative to an omnidirectional response for the cardioid-directional response, the dipole-directional response, and the Fennec algorithm.

for the conditions considered in the reported perceptual experiment with CI users. However, it should be kept in mind that the measured physical SNR benefit will not perfectly predict the observed SRT benefit, as the perceptually measured SRT will also depend on inherent characteristics of the noise (e.g., degree of modulations) as well as the precise SRT value that the subject converges (i.e., algorithm performance is probably SNR dependent). Nevertheless, the analysis demonstrates that the Fennec algorithm provides improved performance over the cardioid-directional response as measured by the physical output SNR for the conditions tested.

Discussion

The results demonstrate that phase-based spatial filtering provides SRT benefits for CI users in reverberant conditions with multiple noise sources. Reverberation did not significantly affect performance for the levels considered ($T_{60} = 0, 270, \text{ and } 540 \text{ ms}$) with average SRT benefits for the Fennec algorithm relative to an omnidirectional response only decreasing from 8.5 to 7.5 dB SNR. The number of noise sources did significantly affect performance for the conditions tested with average SRT benefits for the Fennec algorithm relative to an omnidirectional response decreasing from 9.6 to 6.8 dB SNR for the 1 and 11 noise-sources conditions, respectively. Demonstrating the efficacy of the Fennec algorithm, the observed SRT benefit for the $T_{60} = 540 \text{ ms}$, 11 noise-sources, condition was 6.8 and 3.3 dB SNR relative to the omni- and dipole-directional responses, respectively. Thus, the Fennec algorithm provided substantial SRT benefits even in a highly reverberant condition with multiple noise sources.

That the Fennec algorithm is robust to reverberation is a consequence of using phase-based attenuation with closely spaced microphones. In a reverberant environment, both the target speech and interfering noise sources will generate reflections leading to acoustic combinations of target direct path, target reflections, noise-source direct paths, and noise-source reflections. Of these components, only the target direct path is likely to occur in the straight-ahead direction; while it is possible for components to reflect coincidentally in the straight-ahead direction, it is much less likely. Consequently, when target speech is dominating a spectral-temporal component, the resulting phase difference between microphones will correctly indicate the straight-ahead direction. When target speech is not dominating a spectral-temporal component, the phase difference will be a combination of the noise-source direct paths and reflections, which will combine to indicate a direction other than the straight-ahead direction and consequently will be attenuated.

There is an advantage of using phase-based analysis when the microphones are collinear with the target

sound source; in such a case, the expected phase difference of sounds originating from the target position produce extreme value of the phase difference. For example, for the configuration considered in this article, assuming the speed of sound to be 340 m/s, then the 1-cm spacing between microphones would produce a timing delay of $29.4 \mu\text{s}$ for a straight-ahead sound source. At the other extreme, sounds incident collinear with the microphones arriving from behind the listener would produce a relative delay of $-29.4 \mu\text{s}$. Any combination of noise sources and reflections would produce phase differences that fall between these two values; that is, it is not possible for multiple noise sources to combine in a manner that would average out to produce an acoustic delay at the extreme value of $29.4 \mu\text{s}$. This point is relevant, as it is only true when the target sound is collinear with the two microphones. If the system was redesigned using a binaural configuration of microphones, then a target source straight ahead of the listener would ideally produce a microphone timing difference of 0. It is possible in that configuration for independent sounds from the left and from the right of the listener to combine in such a manner as to produce an average timing delay that would appear to be arriving from the straight-ahead position. This is one advantage of using the closely spaced microphones placed in a behind-the-ear capsule where the microphones are collinear with the straight-ahead position.

Another advantage of this configuration when focusing on the straight-ahead position is that the axis of symmetry for the spatial filtering is collinear with the microphones. That is, to produce the three-dimensional directional response of the system, the directional responses in Figure 1 would be rotated about the axis of symmetry connecting the two microphones. For the present configuration, this axis of symmetry would be the axis connecting 0° and 180° ; consequently, the resulting three-dimensional directional response associated with the response C of Figure 1 would be rotated about that axis producing a focused area centered in three-dimensional space around the target source. This is not true for the binaural configuration, as the axis of symmetry for that configuration would be the line collinear with those microphones which would be the axis connecting 90° and 270° ; consequently, the resulting three-dimensional directional response generated from response C of Figure 1 would produce a torus in its projected shape and sounds incident from directly above or behind the listener would also generate acoustic characteristics similar to the straight-ahead position. The preceding argument is rationale for using the microphone configuration suggested in the present article when focusing on sounds arriving from straight ahead of the listener. For more complex situations in which the target sound source is not always in the straight-ahead direction, then a

combination of microphones would be advantageous. The most obvious for hearing solutions would be the use of two closely spaced microphones over both ears with communications between all four microphones.

In contrast to the Fennec algorithm, null-steering beamforming is relatively sensitive to reverberation, as the process of null-steering beamforming requires an initial estimate be made of noise that is free of contamination from the target speech signal. Typically, this noise-only reference signal is generated by a linear combination of microphones (i.e., Welker, Greenberg, Desloge, & Zurek, 1997). For example, for a binaural microphone configuration and a target sound source that is straight ahead of the listener, the noise-only reference might be formed simply by subtracting the two microphones under the assumption that the target portion of the microphone signals are precisely the same at each microphone. However, this assumption of target equality rapidly degrades in reverberation with reflections leaking into the noise-only reference rapidly decreasing the performance of the overall system. It might be possible to improve upon such null-steering beamforming by using the Fennec algorithm as a more sophisticated method of producing the noise-only reference that is free from target reflections.

Considering the effect of the number of noise sources, that the speech reception benefits provided by the Fennec algorithm decrease with increasing noise sources can be explained by processing issues as the combined noise stimuli become less time variant. For a single, time-reversed speech noise source, the noise will contain the spectral-temporal variations inherent in speech, which, when competing with the target speech, will produce a wide range of spectral-temporal variations in the instantaneous SNR. That is, the individual components of the STFT analysis will contain greater variability. But for the 4 and 11 noise-sources conditions, the combined noise spectrum approaches a more of a multiple talker babble condition. In such a condition, the range of variations in the instantaneous SNR is reduced, and the consequent performance of the phase-based attenuation deteriorates.

Although performance in multiple noise sources decreases compared with a single noise source, the Fennec algorithm does continue to provide a high degree of speech reception benefits. For example, on the $T_{60}=540$ ms and 11 noise-sources condition, the Fennec algorithm provides a 3.3-dB SNR benefit over the dipole-directional response and a 6.8-dB SNR benefit over the omnidirectional response. While this benefit is smaller than the corresponding benefit for the anechoic, 1 noise-source condition, which is 6.6 dB SNR over the dipole-directional response and 11.0 dB SNR over the omnidirectional response, it remains a substantial improvement.

Beyond the evaluation of the spatial-filtering algorithm, the results indicate an important trend in CI perception. Specifically, the results provide further evidence that CI users can obtain masking release in fluctuating noise. The average SRTs for the CI users tested in the anechoic condition with the omnidirectional response increased from -2.5 to 0.8 to 1.7 dB for the 1, 4, and 11 noise-sources conditions, respectively. The SNR for these conditions were balanced based on the root mean square levels of the target speech and the combined noise sources. Thus, that performance was best in the 1 noise-source condition indicates that there is something inherently easier about the 1 noise-source (time-reversed speech) condition compared with the mixed 4 or 11 noise-source conditions. Goldsworthy, Delhorne, Braida, and Reed (2013) demonstrated that, on average, CI users obtain a small but significant masking release (3–5 dB SNR) in temporally gated noise with a 10-Hz gating frequency compared with performance in stationary noise. The masking release observed in the present study between time-reversed speech and 4 and 11 multiple talker babble using time-reversed speech was of similar size (3.3–4.2 dB SNR).

Conclusion

A two-microphone spatial-filtering algorithm dubbed Fennec was evaluated in acoustic conditions in which reverberation and number of noise sources were systematically increased. The Fennec algorithm provided speech reception benefits for all six CI users tested including for the most challenging acoustic condition tested which was highly reverberant ($T_{60}=540$ ms) and had 11 simultaneous, spatially distributed, noise sources. The results indicate that the Fennec algorithm is fairly robust to both reverberation and number of noise sources and is thus a promising solution for improving speech reception for CI users in challenging listening conditions.

Declaration of Conflicting Interests

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

Funding

The author declared receipt of the following financial support for the research, authorship, and/or publication of this article: The research was funded by: NIH NIDCD R43 DC010524-02.

References

- Bolia, R. S., Nelson, W. T., Ericson, M. A., & Simpson, B. D. (2000). A speech corpus for multitalker communications research. *Journal of the Acoustical Society of America*, *107*, 1065–1066.

- Brandstein, M., & Ward, D. (2001). *Microphone arrays: Signal processing techniques and applications*. New York, NY: Springer.
- Chung, K. (2004). Challenges and recent developments in hearing aids: Part I. Speech understanding in noise, microphone technologies and noise reduction algorithms. *Trends in Amplification*, 8, 83–124.
- Chung, K., & Zeng, F.-G. (2009). Using hearing aid adaptive directional microphones to enhance cochlear implant performance. *Hearing Research*, 250, 27–37.
- 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.
- Chung, K., Zeng, F.-G., & Waltzman, S. (2004). Using hearing aid directional microphones and noise reduction algorithms to enhance cochlear implant performance. *Acoustics Research Letters Online*, 5, 56–61.
- Desloge, J. G., Rabinowitz, W. M., & Zurek, P. M. (1997). Microphone-array hearing aids with binaural output — Part I: Fixed-processing systems. *IEEE Transactions on Speech and Audio Processing*, 5, 529–542.
- Frost, O. L. (1972). An algorithm for linearly constrained adaptive array processing. *Proceedings of the IEEE*, 60, 926–935.
- Goldsworthy, R. L. (2005). *Noise reduction algorithms and performance metrics for improving speech reception in noise by cochlear implant users* (Doctoral thesis). Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA.
- Goldsworthy, R. L., Delhorne, L. A., Braidia, L. D., & Reed, C. M. (2013). Psychoacoustic and phoneme identification measures in cochlear-implant and normal hearing listeners. *Trends in Amplification*, 17, 27–44.
- Goldsworthy, R. L., Delhorne, L. A., Desloge, J. G., & Braidia, L. D. (2014). Two-microphone spatial filtering provides speech reception benefits for cochlear implant users in difficult acoustic environments. *The Journal of the Acoustical Society of America*, 136, 867–876.
- Greenberg, J. 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.
- Hamacher, V., Doering, W. H., Mauer, G., Fleischmann, H., & Hennecke, J. (1997). Evaluation of noise reduction systems for cochlear implant users in different acoustic environments. *American Journal of Otolaryngology*, 18, S46–S49. (Retrieved from http://journals.lww.com/otology-neurotology/Abstract/1997/11001/Evaluation_of_Noise_Reduction_Systems_for_Cochlear.21.aspx).
- Hersbach, A. A., Arora, K., Mauer, S. J., & 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 and Hearing*, 33, e13–e23.
- Hersbach, A. A., Grayden, D. B., Fallon, J. B., & McDermott, H. J. (2013). A beamformer post-filter for cochlear implant noise reduction. *Journal of the Acoustical Society of America*, 133, 2412–2420.
- Jeffress, L. A. (1948). A place theory of sound localization. *Journal of Comparative and Physiological Psychology*, 41, 35–39.
- Kokkinakis, K., Azimi, B., Hu, Y., & Friedland, D. R. (2012). Single and multiple microphone noise reduction strategies in cochlear implants. *Trends in Amplification*, 16, 102–116.
- 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.
- 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.
- Kollmeier, B., Peissig, J., & Hohmann, V. (1993). Real-time multiband dynamic compression and noise reduction for binaural hearing aids. *Journal of Rehabilitation Research and Development*, 10, 82–94.
- Lockwood, M. E., Jones, D. L., Bilger, R. C., Lansing, C. R., O'Brien, W. D., Wheeler, B. C., . . . , Feng, A. S. (2004). Performance of time- and frequency-domain binaural beamformers based on recorded signals from real rooms. *Journal of the Acoustical Society of America*, 115, 379–391.
- Margo, V., Schweitzer, C., & Feinman, G. (1997). Comparisons of Spectra 22 performance in noise with and without an additional noise reduction preprocessor. *Seminars in Hearing*, 18, 405–415.
- Peterson, P. M. (1986). Simulating the response of multiple microphones to a single acoustic source in a reverberant room. *Journal of the Acoustical Society of America*, 80, 1527–1529.
- Shinn-Cunningham, B. G., Desloge, J. G., & Kopčo, N. (2001, October). *Empirical and modeled acoustic transfer functions in a simple room: Effects of distance and direction*. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, W2001-4, New Paltz, NY.
- 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.
- Stadler, R. W., & Rabinowitz, W. M. (1993). On the potential of fixed arrays for hearing aids. *Journal of the Acoustical Society of America*, 80, 1332–1342.
- Rothausler, E. H., Chapman, W. D., Guttman, N., Hecker, M. H. L., Nordby, K. S., Silbiger, H. R., . . . , Weinstock, M. (1969). *IEEE recommended practice for speech quality measurements*, 17, 225–246.
- van Hoesel, R. J., & 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.
- Welker, D. P., Greenberg, J. E., Desloge, J. G., & Zurek, P. M. (1997). Microphone-array hearing aids with binaural output — Part II: A two-microphone adaptive system. *IEEE Transactions on Speech and Audio Processing*, 5, 543–551.
- Widrow, B., Glover, J. R., McCool, J. M., Kaunitz, J., Williams, C. S., Hearn, R. H., . . . , Goodlin, R. C. (1975).

- Adaptive noise cancelling: Principles and applications. *Proceedings of the IEEE*, 63, 1692–1716.
- Wouters, J., & Vanden 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.
- Yousefian, N., Kokkinakis, K., & Loizou, P. C. (2010, August). *A coherence-based algorithm for noise reduction in dual-microphone application*. European Signal Processing Conference (EUSIPCO'10), Alborg, Denmark, pp. 1904–1908.
- Yousefian, N., & Loizou, P. C. (2012). A dual-microphone speech enhancement algorithm based on the coherence function. *IEEE Transactions on Audio, Speech, and Language Processing*, 20, 599–609.
- Yousefian, N., & Loizou, P. C. (2013). A dual-microphone algorithm that can cope with competing-talker scenarios. *IEEE Transactions on Audio, Speech, and Language Processing*, 21, 145–155.