Hindawi Publishing Corporation
EURASIP Journal on Applied Signal Processing
Volume 2006, Article ID 93920, Pages 1–14
DOI 10.1155/ASP/2006/93920

Using Intermicrophone Correlation to Detect Speech in Spatially Separated Noise

Ashish Koul[1] and Julie E. Greenberg[2]

[1] *Broadband Video Compression Group, Broadcom Corporation, Andover, MA 01810, USA*
[2] *Massachusetts Institute of Technology, 77 Massachusetts Avenue, Room E25-518, Cambridge, MA 02139-4307, USA*

Received 29 April 2004; Revised 20 April 2005; Accepted 25 April 2005

This paper describes a system for determining intervals of "high" and "low" signal-to-noise ratios when the desired signal and interfering noise arise from distinct spatial regions. The correlation coefficient between two microphone signals serves as the decision variable in a hypothesis test. The system has three parameters: center frequency and bandwidth of the bandpass filter that prefilters the microphone signals, and threshold for the decision variable. Conditional probability density functions of the intermicrophone correlation coefficient are derived for a simple signal scenario. This theoretical analysis provides insight into optimal selection of system parameters. Results of simulations using white Gaussian noise sources are in close agreement with the theoretical results. Results of more realistic simulations using speech sources follow the same general trends and illustrate the performance achievable in practical situations. The system is suitable for use with two microphones in mild-to-moderate reverberation as a component of noise-reduction algorithms that require detecting intervals when a desired signal is weak or absent.

Copyright © 2006 A. Koul and J. E. Greenberg. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. INTRODUCTION

Conventional hearing aids do not selectively attenuate background noise, and their inability to do so is a common complaint of hearing-aid users [1–4]. Researchers have proposed a variety of speech-enhancement and noise-reduction algorithms to address this problem. Many of these algorithms require identification of intervals when the desired speech signal is weak or absent, so that particular noise characteristics can be estimated accurately [5–7]. Systems that perform this function are referred to by a number of terms, including voice activity detectors, speech detectors, pause detectors, and double-talk detectors. Speech pause detectors are not limited to use in hearing-aid algorithms. They are used in a number of applications including speech recognition [8, 9], mobile telecommunications [10, 11], echo cancellation [12], and speech coding [13].

In some cases, noise-reduction algorithms are initially developed and evaluated using information about the timing of speech pauses derived from the clean signal, which is possible in computer simulations but not in a practical device. Marzinzik and Kollmeier [11] point out that speech pause detectors "are a very sensitive and often limiting part of systems for the reduction of additive noise in speech."

Many of the previously proposed methods for speech pause detection are intended for use with single-microphone noise-reduction algorithms, where it is assumed that the desired signal is speech and the noise is not speech. In these applications, the distinction between signal and noise depends on the presence or absence of signal characteristics particular to speech, such as pitch [14, 15] or formant frequencies [16]. Other approaches rely on assumptions about the relative energy in frames of speech and noise [8, 17]. A summary of single-microphone pause detectors is found in [11].

Other methods of speech pause detection are possible when more than one microphone signal are available. Using signals from multiple microphones, information about the signal-to-noise ratio (SNR) can be discerned by comparing the signals received at different microphones. The distinction between desired signal and unwanted noise is based on the direction of arrival of the sound sources, so these approaches also operate correctly when the noise is a competing talker with characteristics similar to those of the desired speech signal.

Researchers working on a variety of applications have proposed speech pause detectors using two or more microphone signals. Examples include a three-microphone system to improve the noise estimates for a spectral subtraction

algorithm used as a front end for a speech recognition system [18]; a joint system for noise-reduction and speech coding [19]; a voice activity detector based on the coherence between two microphones to improve the performance of noise reduction algorithms for mobile telecommunications [20]. This third system requires a substantial distance between microphones, as it is only effective when the noise signal is relatively incoherent between the two microphones. A related body of work is the use of single- and double-talk detectors to control the update of adaptive filters in echo cancellers. Although there is only one microphone in this application, a second signal is obtained from the loudspeaker. A comprehensive summary of these approaches is found in [12].

In developing adaptive algorithms for microphone-array hearing aids and cochlear implants, researchers have found that it is necessary to limit the update of the adaptive filter weights to intervals when the desired signal is weak or absent. Several methods have been proposed to detect such intervals based on the correlation between microphones and the ratio of intermediate signal powers [7, 21, 22]. Greenberg and Zurek [7] propose a simple method using the intermicrophone correlation coefficient to detect intervals of low SNR that substantially improves noise-reduction performance of an adaptive microphone-array hearing aid. This method is applicable whenever two microphone signals are available and the signal and noise are distinguished by spatial, not temporal or spectral, characteristics. Despite its demonstrated effectiveness, this method was developed in an ad hoc manner. The purpose of this work is to perform a rigorous analysis of the intermicrophone correlation coefficient of multiple sound sources in anechoic and reverberant environments, to formalize the selection of parameter settings when using the intermicrophone correlation coefficient to estimate the range of SNR, and to evaluate the performance that can be obtained when optimal settings are used.

2. PROPOSED SYSTEM

Figure 1 shows the signal scenario used in this work. All sources and microphones are assumed to lie in the same plane, with the microphones in free space. Sources with angles of incidence between $-\theta_0$ and θ_0 are considered to be desired signals, while sources arriving from θ_0 to 90° and $-\theta_0$ to $-90°$ are interfering noise. Sound can arrive from any angle in a 360° range, but due to the symmetry inherent in a two-microphone broadside array, sources arriving at incident angles in the range $180° \pm \theta_0$ will also be treated as desired signals. Moreover, due to the symmetry in the definition of desired signal and noise, we restrict the following analysis to the range 0–90° without loss of generality.

Figure 2 shows the previously proposed system that uses the correlation coefficient between the two microphone signals to distinguish between intervals of high and low SNRs [7]. The microphone signals are digitized and then passed through bandpass filters with center frequency f_0 and bandwidth B. The bandpass filtered signals $x_1[n]$ and $x_2[n]$ are

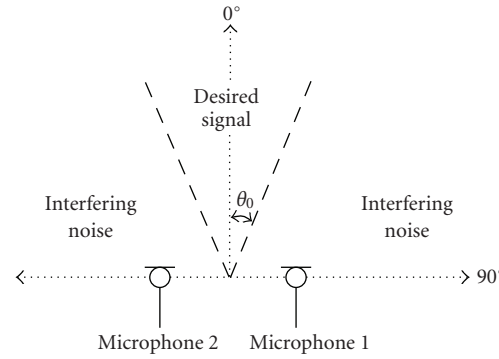

Figure 1: Signal scenario indicating the ranges of incident angles for the desired signal and interfering noise sources.

divided into N-point long segments. For each pair of segments, the corresponding intermicrophone correlation coefficient r is computed as

$$r = \frac{\sum_{n=1}^{N} x_1[n]x_2[n]}{\sqrt{\sum_{n=1}^{N} x_1^2[n] \sum_{n=1}^{N} x_2^2[n]}}. \quad (1)$$

Finally, r is compared to a fixed threshold r_0 to determine the predicted SNR range for each segment.

Because the desired signal arrives at array broadside from angles near straight-ahead, it will be highly correlated in the two microphone signals and will contribute positive values to r, provided that the source is located inside the critical distance in a reverberant environment. The interfering noise arrives from off-axis directions and should contribute negative values to r. This effect is enhanced by the bandpass filter which limits the frequency range so that signals arriving from the range of noise angles will be out of phase and produce minimum correlation values. Thus, the purpose of the bandpass filter is to enhance the ability of the intermicrophone correlation measure to distinguish between desired signal and interfering noise.

This approach is attractive for applications such as digital hearing aids, where computing resources are limited. If necessary, the correlation coefficient can be estimated efficiently using the sign of the bandpass filtered signals [7].

The proposed system has three independent parameters: the center frequency (f_0) of the bandpass filter, the bandwidth (B) of the bandpass filter, and the threshold (r_0). Another important parameter of the proposed system is the intermicrophone spacing (d). The intermicrophone spacing is not treated as a free parameter, rather it is incorporated into the analysis by normalizing two of the independent parameters (center frequency and bandwidth) as discussed in detail in Section 4.1.

In this work, the proposed system is analyzed to determine optimal settings of the three independent parameters. First, Section 3 describes a simple signal model and derives the associated probability density functions and hypothesis

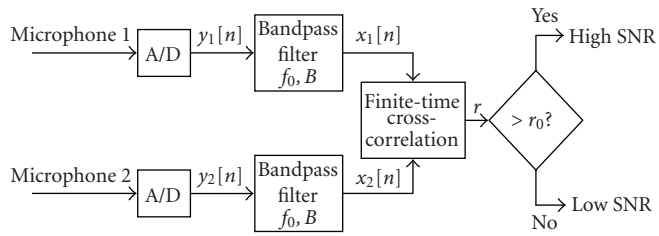

Figure 2: Block diagram of the system to estimate the intermicrophone correlation coefficient for determining range of SNR.

tests for the intermicrophone correlation. In Section 4, the analysis of Section 3 is used to examine the effects of the three parameters. In Section 4.1, theoretical results from the anechoic scenario are used to identify candidates for the optimal value of the center frequency f_0. In Section 4.2, theoretical results from the reverberant scenario are used to optimize the threshold r_0. For practical reasons described in Section 4.1, the bandwidth parameter B cannot be optimized based on the theoretical analysis; instead, it is determined from the simulations performed in Section 5.

3. ANALYSIS

3.1. Preliminaries

3.1.1. Assumptions

The following assumptions are made to allow a tractable analysis.

(i) There is one desired signal source and one interfering noise source in the environment.

(ii) The desired signal arrives at the microphone array from an incident angle in the range 0° to θ_0, and the interfering noise arrives from an incident angle in the range θ_0 to 90°. For both the desired signal and the interfering noise, the probability of the source arriving at any incident angle is uniformly distributed over the corresponding range of angles.

(iii) Sound sources are continuous, zero-mean, white Gaussian noise processes. Desired signal and interfering noise sources have variances σ_s^2 and σ_i^2, respectively. The signal-to-noise ratio is defined as $\text{SNR} = 10\log_{10}(W)$, where $W = \sigma_s^2/\sigma_i^2$.

(iv) Reverberation can be modelled as a spherically diffuse sound field. This is an admittedly simplified model of reverberation which is only applicable for relatively small rooms [23]. Reverberant energy is characterized by the direct-to-reverberant ratio $\text{DRR} = 10\log_{10}(\beta)$, where β is the ratio of energy in the direct wave to energy in the reverberant sound. The value of β is equal for both signal and noise sources, implying that both sources are roughly the same distance from the microphones.

(v) The filters applied to the incoming signals are *ideal* bandpass filters with center frequency f_0 and bandwidth B.

3.1.2. Signal model

While the system shown in Figure 2 processes the digitized signals, for the analysis, we consider the signals $x_1(t)$ and $x_2(t)$, continuous-time reconstructions of the bandpass filtered signals $x_1[n]$ and $x_2[n]$. For a two-microphone array in free space, these two signals can be modelled as

$$\begin{aligned} x_1(t) &= s(t) + i(t), \\ x_2(t) &= s(t-\tau_s) + i(t-\tau_i), \end{aligned} \tag{2}$$

where $s(t)$ is the desired signal after bandpass filtering, $i(t)$ is the interfering noise after bandpass filtering, and τ_s and τ_i represent the time delays between microphones for the desired signal and interfering noise, respectively. Assuming plane wave propagation, τ_s and τ_i can be expressed as

$$\tau_s = \frac{d}{c}\sin(\theta_s), \qquad \tau_i = \frac{d}{c}\sin(\theta_i), \tag{3}$$

where d is the distance separating the microphones, c is the speed of sound, and θ_s and θ_i are the incident angles of the respective sources.

The theoretical correlation coefficient ρ of the two signals is

$$\rho = \frac{E\{x_1(t)x_2(t)\}}{\sqrt{E\{x_1^2(t)\}E\{x_2^2(t)\}}}, \tag{4}$$

where $E\{\cdot\}$ denotes expected value. Under ideal conditions of stationary signals and infinite data, ρ would be the decision variable used in the system of Figure 2. However, in this application, we use the intermicrophone correlation coefficient r, defined in (1) to estimate ρ from discrete samples of the two signals over a finite time period.

3.1.3. Fisher Z-transformation

Consider the case of two random variables a and b drawn from a bivariate Gaussian distribution. We wish to obtain an estimate r of the theoretical correlation coefficient ρ using N sample pairs drawn from the joint distribution of a and b. In general, the probability distribution of the estimator r is difficult to work with directly, because its shape depends on the value of ρ.

The Fisher Z-transformation is defined as

$$z = \tanh^{-1}(r) = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right). \tag{5}$$

This yields the new random variable z which has an approximately Gaussian distribution with mean $\overline{z} = (1/2)\ln((1+\rho)/(1-\rho))$ and variance $\sigma_z^2 = 1/(N-3)$ [24]. This derived variable z has a simple distribution whose shape does not depend on the unknown value of ρ.

Due to the assumption that the signal and noise sources are Gaussian random processes, the microphone signals are jointly Gaussian random processes. Even after bandpass filtering, the input variables $x_1(t)$ and $x_2(t)$ defined in (2) are jointly Gaussian, and the Fisher Z-transformation may be applied.

3.2. Intermicrophone correlation for one source in an anechoic environment

We begin by deriving the probability density function (pdf) of r for a single source with incident angle θ. After A/D conversion and bandpass filtering, the signals $x_1[n]$ and $x_2[n]$ are rectangular bands of noise. The true intermicrophone correlation is [25]

$$\rho_\theta = \frac{\cos(kd\sin\theta)\sin\left((\pi Bd/c)\sin\theta\right)}{\left((\pi Bd/c)\sin\theta\right)}, \tag{6}$$

where k is the wavenumber,

$$k = \frac{2\pi f_0}{c}. \tag{7}$$

Using the Fisher Z-transformation, the conditional pdf of z, given a source at incident angle θ, is

$$f_{z|\theta}(z \mid \theta) = \frac{1}{\sigma_z\sqrt{2\pi}}\exp\left(-\frac{\left[z-\overline{z}(\theta)\right]^2}{2\sigma_z^2}\right) \tag{8}$$

with

$$\begin{aligned}\overline{z}(\theta) &= \frac{1}{2}\ln\left(\frac{1+\rho_\theta}{1-\rho_\theta}\right),\\ \sigma_z^2 &= \frac{1}{N-3}.\end{aligned} \tag{9}$$

Using the assumption that θ is uniformly distributed over a specific range of angles, the joint pdf for z and θ is

$$f_{z,\theta}(z,\theta) = \frac{1}{\theta_2-\theta_1}f_{z|\theta}(z \mid \theta), \tag{10}$$

where $\theta_2 = \theta_0$ and $\theta_1 = 0$ for a signal source and $\theta_2 = 90^\circ$ and $\theta_1 = \theta_0$ for a noise source. To obtain the marginal density of z, the joint density in (10) is integrated over the appropriate range of θ, that is,

$$f_z(z) = \frac{1}{(\theta_2-\theta_1)\sigma_z\sqrt{2\pi}}\int_{\theta_1}^{\theta_2}\exp\left(-\frac{\left[z-\overline{z}(\theta)\right]^2}{2\sigma_z^2}\right)d\theta. \tag{11}$$

With this expression for the pdf of z, we can use the definition of the Fisher Z-transformation to derive the pdf of the intermicrophone correlation coefficient r. Since $r = \tanh(z)$ is a monotonic transformation of the random variable z, the pdf of r can be obtained using [26]

$$f_r(r) = f_z(z)\frac{dz}{dr}. \tag{12}$$

Substituting $dz/dr = 1/(1-r^2)$ and the definition of z produces the pdf of r for a single source:

$$\begin{aligned}f_r(r) = {}& \frac{1}{(1-r^2)(\theta_2-\theta_1)\sigma_z\sqrt{2\pi}}\\ &\times\int_{\theta_1}^{\theta_2}\exp\left(-\frac{\left[\tanh^{-1}(r)-\overline{z}(\theta)\right]^2}{2\sigma_z^2}\right)d\theta.\end{aligned} \tag{13}$$

3.3. Intermicrophone correlation for two independent sources in an anechoic environment

Next, we consider the intermicrophone correlation coefficient for one signal source and one noise source in an anechoic environment, denoted by r_a. Substituting discrete-time versions of (2) into (1) yields

$$r_a = \frac{\sum_n\left(s[n]+i[n]\right)\left(s[n-\tau_s]+i[n-\tau_i]\right)}{\sqrt{\sum_n\left(s[n]+i[n]\right)^2\sum_n\left(s[n-\tau_s]+i[n-\tau_i]\right)^2}}. \tag{14}$$

The corresponding expression for the desired signal component alone is

$$r_s = \frac{\sum_n\left\{s[n]s[n-\tau_s]\right\}}{\sqrt{\sum_n s^2[n]\sum_n s^2[n-\tau_s]}}, \tag{15}$$

and for the noise component alone is

$$r_i = \frac{\sum_n\left\{i[n]i[n-\tau_i]\right\}}{\sqrt{\sum_n i^2[n]\sum_n i^2[n-\tau_i]}}. \tag{16}$$

We now make the following assumptions.

(1) The $s\times i$ cross terms in (14) are negligible when compared with the $s\times s$ and $i\times i$ terms to which they add.

(2) The effect of time delay on the energy can be ignored such that

$$\begin{aligned}\sum_n s^2[n] &\approx \sum_n s^2[n-\tau_s],\\ \sum_n i^2[n] &\approx \sum_n i^2[n-\tau_i].\end{aligned} \tag{17}$$

(3) The SNR defined in Section 3.1.1 can be estimated from the sample data as

$$W = \frac{\sum_n s^2[n]}{\sum_n i^2[n]}. \tag{18}$$

Using the first two assumptions, (14) becomes

$$r_a = \frac{\sum_n s[n]s[n-\tau_s]+\sum_n i[n]i[n-\tau_i]}{\sum_n s^2[n]+\sum_n i^2[n]}. \tag{19}$$

Substituting (15) and (16), dividing all terms by $\sum_n i^2[n]$, and then substituting (18), we obtain

$$r_a = \frac{W r_s + r_i}{W+1} = \frac{W}{W+1} r_s + \frac{1}{W+1} r_i. \tag{20}$$

Equation (20) expresses the intermicrophone correlation as a linear combination of the correlations for signal and noise separately. The pdfs of both r_s and r_i can be obtained from (13).

For a known SNR, the pdf for r_a, a linear combination of r_s and r_i, is obtained by

$$f_{r_a|W}(r_a \mid W) = \left[\frac{W+1}{W} f_{r_s}\left(\frac{W+1}{W} r_s\right)\right] * \left[(W+1) f_{r_i}((W+1) r_i)\right], \tag{21}$$

where $*$ denotes convolution [26]. Equation (21) is the pdf of the intermicrophone correlation estimate for anechoic environments r_a conditioned on a particular value of SNR.

3.4. Reverberation

Until now, we have only considered the direct wave of the sound sources. We now consider the addition of reverberation. As described in Section 3.1.1, the reverberant sound component is modelled as a spherically diffuse sound field that is statistically independent of the direct signal and noise components. In addition, it has energy that is characterized by the direct-to-reverberant ratio β.

Analogous to (15) and (16), we define the intermicrophone correlation for the direct components r_a given by (20) and for the reverberation r_r. Applying arguments similar to those used in the previous section produces an expression for the intermicrophone correlation in the case of reverberation:

$$r = \frac{\beta r_a + r_r}{\beta+1} = \frac{\beta}{\beta+1} r_a + \frac{1}{\beta+1} r_r. \tag{22}$$

Once again, the total correlation is a linear combination of its components, and for a known direct-to-reverberant ratio, the pdf for r, a linear combination of r_a and r_r, is obtained by convolution [26]:

$$f_{r|\beta,W}(r \mid \beta, W) = \left[\frac{\beta+1}{\beta} f_{r_a|W}\left(\frac{\beta+1}{\beta} r_a \mid W\right)\right] * \left[(\beta+1) f_{r_r}((\beta+1) r_r)\right]. \tag{23}$$

Equation (23) is the pdf of the intermicrophone correlation estimate r conditioned on particular values of DRR and SNR. It requires convolution of the direct component pdf, given by (21), and the reverberant component pdf, derived below.

Under the existing assumptions, the pdf for the reverberant component is based on the intermicrophone correlation coefficient for bandlimited Gaussian white noise processes, approximated by [27]

$$\rho_r = \frac{\sin(\pi B d/c)}{\pi B d/c} \frac{\sin(kd)}{kd}. \tag{24}$$

In the following, (24) is used as the true intermicrophone correlation for reverberant sound ρ_r.

The intermicrophone correlation for reverberant sound based on sample data r_r is an estimate of ρ_r. Applying the Fisher Z-transformation,

$$z = \tanh^{-1}(r_r) = \frac{1}{2} \ln\left(\frac{1+r_r}{1-r_r}\right). \tag{25}$$

The random variable z has an approximately Gaussian distribution,

$$f_z(z) = \frac{1}{\sigma_z \sqrt{2\pi}} \exp - \left(\frac{[z-\overline{z}]^2}{2\sigma_z^2}\right) \tag{26}$$

with

$$\overline{z} = \frac{1}{2} \ln\left(\frac{1+\rho_r}{1-\rho_r}\right), \qquad \sigma_z^2 = \frac{1}{N-3}. \tag{27}$$

Applying (12) to (26) produces the pdf of intermicrophone correlation for the reverberant component,

$$f_{r_r}(r) = \frac{1}{(1-r^2)\sigma_z\sqrt{2\pi}} \times \exp\left(-\frac{[\tanh^{-1}(r_r) - \overline{z}]^2}{2\sigma_z^2}\right). \tag{28}$$

This pdf for the reverberant sound field is combined with the pdf for the direct sounds given by (21) according to (23) to obtain the pdf for the total intermicrophone correlation for signal and noise with reverberation.

3.5. Hypothesis testing

The goal of the system shown in Figure 2 is to distinguish between two situations: "low" SNR and "high" SNR, denoted by H_0 and H_1, respectively. Although the preceding analysis was performed under the assumption that the sources were white Gaussian noise processes, the system is intended to work with speech sources, detecting intervals of high and low SNRs which occur due to the natural fluctuations in speech. We define H_0 to be $10\log(W) < 0$ dB and H_1 to be $10\log(W) > 0$ dB. The choice of 0 dB as the cutoff point is motivated by the application of designing robust adaptive algorithms for microphone-array hearing aids, an application where the degrading effects of strong target signals typically occur when the SNR exceeds 0 dB [7].

The preceding analysis treated the SNR, W, as a known constant, but for the purpose of formulating a hypothesis test, it is now regarded as a random variable. Thus, it becomes necessary to know an approximate probability distribution for W. We assume that the SNR is uniformly distributed between -20 dB and $+20$ dB, so the variable $U = 10\log(W)$ is uniformly distributed between -20 and 20. Under this assumption, the two hypotheses H_0 and H_1 both have equal prior probability. In this case, the decision rule that minimizes the probability of error [28] is to select the hypothesis corresponding to the larger value of the conditional

pdf for each value of r, that is, we conclude that H_1 is true when $f_{r|H_1,\beta}(r \mid H_1,\beta) > f_{r|H_0,\beta}(r \mid H_0,\beta)$ and we conclude that H_0 is true when $f_{r|H_0,\beta}(r \mid H_0,\beta) > f_{r|H_1,\beta}(r \mid H_1,\beta)$.

To derive the conditional pdf of r under either hypotheses, the pdf given by substituting (21) and (28) into (23) is integrated over the appropriate range:

$$\begin{aligned} f_{r|H_0,\beta}(r \mid H_0,\beta) &= \int_{-20}^{0} f_{r|W,\beta}(r \mid W,\beta)\, dU, \\ f_{r|H_1,\beta}(r \mid H_1,\beta) &= \int_{0}^{20} f_{r|W,\beta}(r \mid W,\beta)\, dU. \end{aligned} \tag{29}$$

Evaluating these expressions requires substituting $W{=}10^{U/10}$.

Performance is measured by computing the probability of correct detections, that is, saying H_1 when H_1 is true,

$$P_D = \int_{r_0}^{1} f_{r|H_1,\beta}(r \mid H_1,\beta)\, dr, \tag{30}$$

and false alarms, that is, saying H_1 when H_0 is true,

$$P_F = \int_{r_0}^{1} f_{r|H_0,\beta}(r \mid H_0,\beta)\, dr, \tag{31}$$

where r_0 is the threshold defined in Section 2. We also define the probability of missed detections

$$P_M = 1 - P_D, \tag{32}$$

and the overall probability of error

$$P_E = \frac{1}{2}P_F + \frac{1}{2}P_M, \tag{33}$$

again assuming that H_0 and H_1 have equal prior probabilities.

4. ANALYTIC RESULTS

All calculations were performed in Matlab(R) on a PC with a Pentium III processor. Probability density functions were computed from (21), (23), and (28) using the Matlab(R) function *quad*. Throughout this analysis, the boundary between desired signals and interfering noise is set to $\theta_0 = 15^\circ$.

4.1. Effects of frequency and bandwidth

As described in Section 2, the three parameters to be selected are the center frequency (f_0) of the bandpass filter, the bandwidth (B) of the bandpass filter, and the threshold (r_0). Without loss of generality, we use two alternate variables in place of the center frequency and bandwidth, specifically kd in place of center frequency and fractional bandwidth in place of absolute bandwidth. Using (7), the quantity kd is related to center frequency according to

$$kd = \frac{2\pi f_0 d}{c}. \tag{34}$$

This alternate variable kd permits quantifying the center frequency parameter in a way that simultaneously incorporates both center frequency and intermicrophone distance, and we will refer to it as *relative center frequency*. The fractional bandwidth B' is defined as

$$B' = \frac{B}{f_0}. \tag{35}$$

Using (34) and (35) with (6) reveals that for a source arriving from angle θ, the true intermicrophone correlation can be expressed exclusively in terms of these two parameters, that is,

$$\rho_\theta = \frac{\cos(kd\sin\theta)\sin\left((kdB'/2)\sin\theta\right)}{\left((kdB'/2)\sin\theta\right)}. \tag{36}$$

We begin to determine the optimal value of the relative center frequency kd by examining the pdfs of the intermicrophone correlation in an anechoic environment. Figure 3 shows pdfs of r_a, computed by evaluating (21) for three values of SNR and three values of kd with fractional bandwidth $B' = 0.22$. As expected, when the microphone inputs consist of signal alone (right column of Figure 3), r_a is concentrated near +1; when the inputs consist of noise alone (left column of Figure 3), r_a takes on substantially lower values. When the microphone inputs consist of signal and noise with SNR = 0 dB (center column of Figure 3), r_a takes on intermediate values distributed according to the convolution of the two extreme cases of signal alone and noise alone. Other values of SNR produce pdfs that vary along a continuum between the cases shown in each row of Figure 3.

Using Figure 3 to consider the effect of kd reveals that for any choice of the relative center frequency, for the signal alone, the pdf is heavily concentrated near $r_a = 1$, although lower values of kd produce more tightly concentrated pdfs. For the noise alone, the pattern is less evident. For $kd = \pi$, the pdf is heavily concentrated near $r_a = -1$. This is expected since noise sources originating from 90° are exactly out of phase when $kd = \pi$, and therefore have a true correlation of -1. When the value of kd deviates from this ideal situation, the noise-alone pdfs are not necessarily concentrated near $r_a = -1$.

Because the ultimate goal is to use r as a decision variable in a hypothesis test, the system will perform better when the pdfs are such that they occupy different regions of the x-axis under the two extreme conditions, with minimal overlap of the pdfs between the cases of signal alone and noise alone. Therefore, at first glance, it might appear that selecting the relative center frequency of $kd = \pi$ is the optimal choice for this parameter. However, careful examination of Figure 3 reveals that the noise-alone pdf for $kd = \pi$ spans a very large range, with a tail in the positive r_a direction reaching values close to $r = +1$. Since overlap of the signal-alone and noise-alone pdfs will adversely affect the performance of the hypothesis test, this long tail is an undesirable feature. Examining the noise-alone pdf for $kd = 4\pi/3$, which is less concentrated about $r_a = -1$ but has less overlap with the corresponding signal-alone pdf, indicates that this parameter setting should not be eliminated as a candidate.

This suggests using the moments of the pdfs about the corresponding extreme values as appropriate metrics to select the relative center frequency parameter kd. The moment

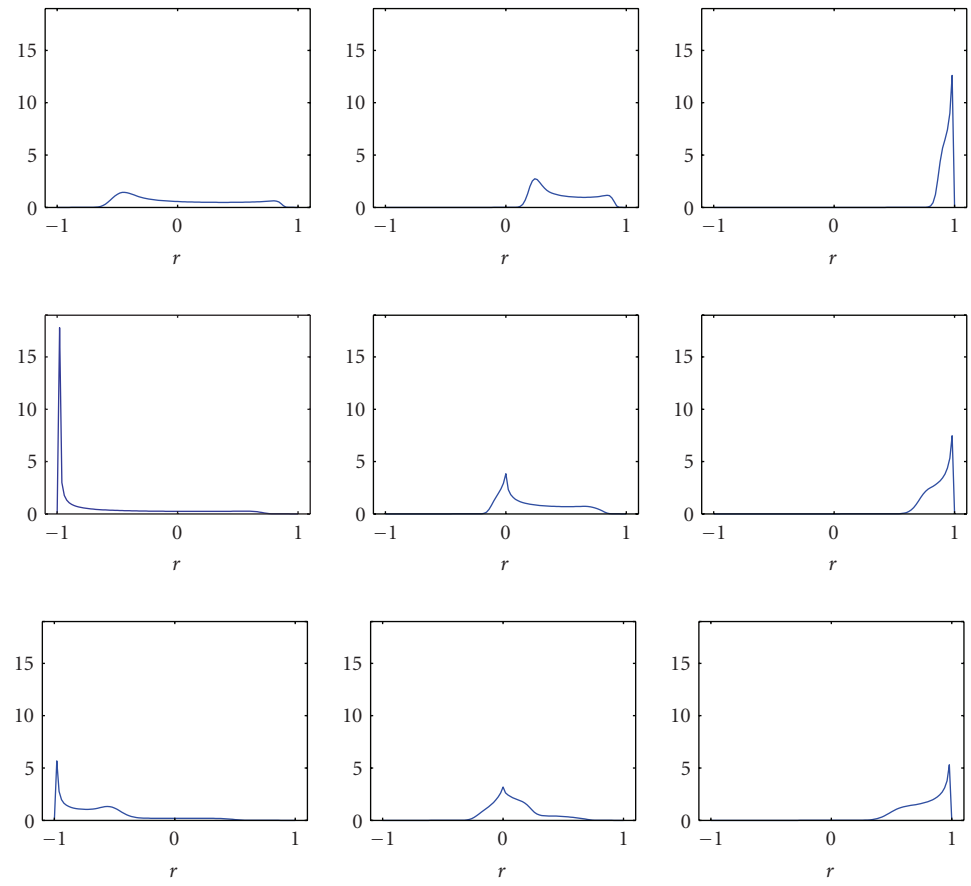

FIGURE 3: Probability density functions of the estimated intermicrophone correlation coefficient for two sources in an anechoic environment, $f_{r_a|W}(r_a \mid W)$, computed from (21), for three SNRs ($-\infty$, 0, and $+\infty$ dB) and for three values of relative center frequency ($kd = 2\pi/3, \pi, 4\pi/3$), with fractional bandwidth $B' = 0.22$ and $\theta_0 = 15^\circ$. The first row represents $kd = 2/3\pi$, the second row represents $kd = \pi$, and the third row represents $kd = 4/3\pi$. The first column represents noise alone, the second column represents SNR = 0 dB, and the third column represents signal alone.

of the signal-alone pdf about +1 and the moment of the noise-alone pdf about −1 will quantify how concentrated each pdf is about the desired extreme value, while penalizing long tails deviating from that value. Low values of the moment are desirable, indicating more concentrated pdfs.

Figure 4 shows the second moments of the signal- and noise-alone pdfs as a function of kd for several values of fractional bandwidth. The lines in Figure 4(a) are monotonic, indicating that reducing kd always causes the signal-alone pdf to be more concentrated about +1. Figure 4(b) shows that the moment of the noise-alone pdf has a local minimum for $kd \approx 1.3\pi$, with a slight variation due to bandwidth. The moments of the noise-alone pdf are an order of magnitude larger than those of the signal-alone pdfs, so in terms of optimizing the overall performance, relatively greater weight should be given to the noise-alone pdfs.

Based on Figure 4, the rest of this work considers two choices of relative center frequency $kd = \pi$ and $kd = (4/3)\pi$. The value of $kd = (4/3)\pi$ is chosen because it is near the minimum of the noise-alone pdf for the lower values of fractional bandwidth. The value $kd = \pi$ is selected since for this value, the moment for the noise-alone pdf is still within the relatively broad region about its minimum, while being considerable lower for the signal-alone pdf.

Figure 4 also shows that for the idealized scenario of white Gaussian noise sources, increasing the bandwidth parameter B' slightly increases the moments. This will have a small but detrimental effect on the performance. However, in a practical system, where the desired signal is speech, a relatively wide bandwidth is required to capture enough energy from the speech signal to minimize adverse affects due to relative energy fluctuations in different frequency regions. The current theoretical analysis is necessarily based on idealized signals, while the final system will operate on speech sources. Therefore, the selection of the bandwidth parameter will be evaluated via simulations in Section 5.

4.2. Effects of reverberation and threshold selection

Figure 5 shows the pdfs of the intermicrophone correlation r for signal and noise computed by evaluating (23) for three values of SNR and three levels of reverberation. Because the

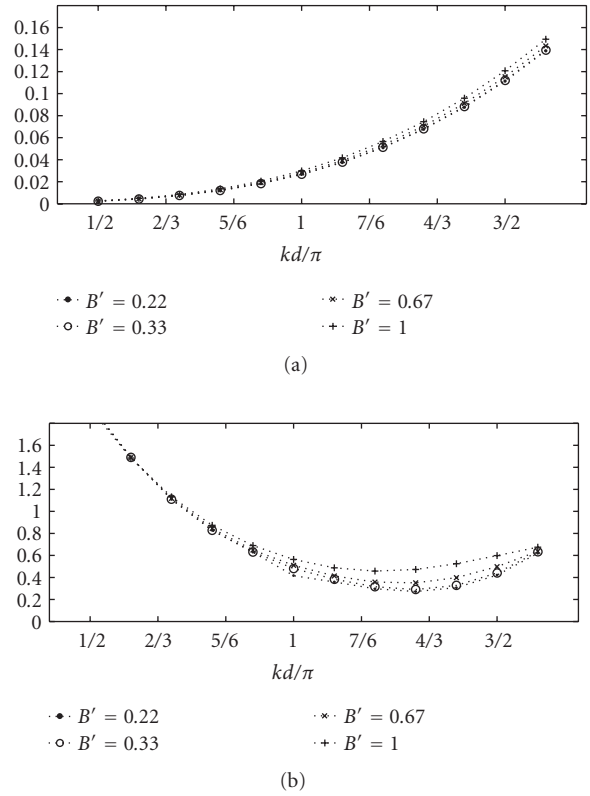

FIGURE 4: Second moments of pdfs as a function of relative center frequency kd, with $\theta_0 = 15^\circ$. The multiple curves are for different values of fractional bandwidth B'. (a) Moment of signal-alone pdf about +1. (b) Moment of noise-alone pdf about −1.

system is dependent on the directional information contained in the direct wave of the signals, it is not expected to perform well in strong reverberation. Accordingly, we restrict the level of reverberation to $\beta \geq 1$, corresponding to DRRs greater than 0 dB. Comparing the top row of Figure 5 (anechoic) to the middle and bottom rows reveals that the effect of reverberation is to shift the center-of-mass of the pdfs away from the extreme values of ±1 and towards more moderate values of r. This increases the overlap between the signal-alone and noise-alone pdfs, thereby increasing the probability of error of the hypothesis test.

In the previous section, candidate values of kd were determined based on the pdfs for the anechoic case. Figure 5 illustrates that the signal-alone and noise-alone pdfs are affected equally by the simple model of reverberation used in this work, indicating that the analysis of the effect of kd in the anechoic case also applies to reverberation.

The next step is to determine the optimal range for the threshold r_0. Because the effect of reverberation is to bring the signal-alone and noise-alone pdfs closer together, we must include reverberation as we consider the threshold selection. Furthermore, until now we have based our analysis on the conceptually simple signal- and noise-alone pdfs shown in the right and left columns of Figures 3 and 5. However, in this application, we are not attempting to distinguish between signal-alone from noise-alone cases; we wish to select a threshold that will minimize the probability of error when classifying combinations of signal and noise at various SNRs. Therefore, to select the threshold, we consider the signal scenario described in conjunction with the hypothesis tests in Section 3.5.

Figure 6 shows the conditional pdfs for the hypothesis test as given by (29) for three levels of reverberation. Given equal prior probabilities for the two hypotheses, the optimum choice of the threshold r_0 is the value at which the pdfs corresponding to H_0 and H_1 intersect. However, as seen in Figure 6, the value of r at which this intersection occurs is not constant; it varies with the level of reverberation. A practical system must use one threshold to operate robustly across all levels of reverberation. The threshold cannot be selected to account for the level of reverberation, which is an unknown environmental variable.

Figure 7 shows the probability of error given by (33) as a function of the threshold r_0 for two values of kd. For $kd = \pi$, any choice of threshold in the range 0–0.2 minimizes the probability of error, regardless of the level of reverberation. For $kd = (4/3)\pi$, the minimum probability of error varies somewhat with threshold, but using $r_0 = 0$ provides near-optimal performance for all levels of reverberation.

5. SIMULATIONS

This section presents the results of computer simulations of the SNR-detection system shown in Figure 2. These simulations were performed in Matlab(R). The sound sources were sampled at 10 kHz. The bandpass filters were 81-point FIR filters designed using the Parks-McClellan method. The filtered signals were broken into frames of 100 samples (10 ms), which is appropriate for tracking power fluctuations in speech. For each frame, the sample correlation coefficient is computed according to (1). This value is compared to the threshold. If it exceeds the threshold, then the system declares H_1 (high SNR), otherwise it declares H_0 (low SNR).

The desired signal and interference sources were first convolved with their respective source-to-microphone impulse responses and then added together. These impulse responses were generated numerically using the image method [29, 30]. The simulated room was $5.2 \times 3.4 \times 2.8$ m. The microphones were centered at the coordinates (2.7, 1.4, 1.6) m along the array axis which was a line through the coordinates (2.7495, 1.3505, 1.600) m. Three intermicrophone distances of $d = 7, 14$, and 28 cm were used. All sources in the room were located on a circle around the array center in the horizontal plane at height of 1.7 m. The forward direction ($\theta = 0$) is defined to be directly broadside of the array in the direction of positive coordinates, and increasing the incident angle refers to clockwise progression of source angle when viewed from above. The radius of source locations and coefficient of absorption for the walls vary with the specified level of reverberation. For the anechoic environment, the radius was 1.0 m and the absorption coefficient of all surfaces was 1.0. For DRR = 3 dB ($\beta = 2$), the radius was 1.07 m and the absorption coefficient was 0.6. For DRR = 0 dB ($\beta = 1$), the

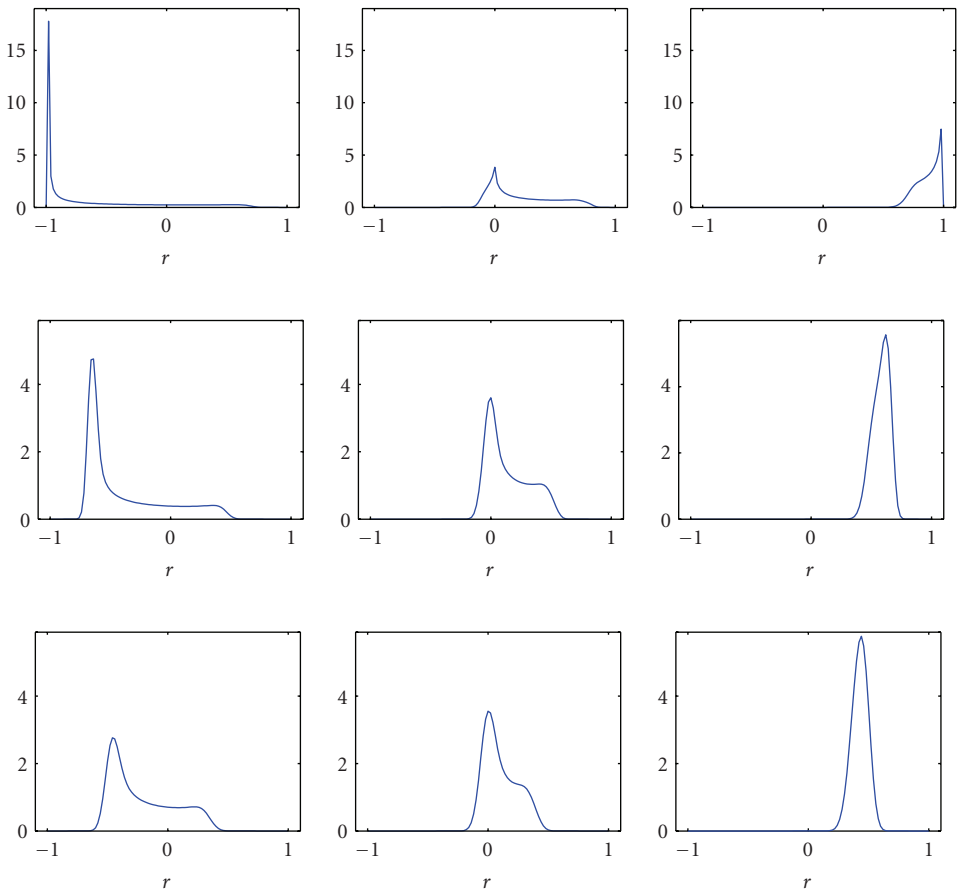

Figure 5: Probability density functions of the estimated intermicrophone correlation coefficient for two sources in varying levels of reverberation $f_{r|\beta,W}(r \mid \beta, W)$ computed from (23), for three SNRs ($-\infty$, 0, and $+\infty$ dB) and three levels of reverberation (DRR = 0, 3, and $+\infty$ dB represents by the three rows), with relative center frequency of $kd = \pi$, fractional bandwidth $B' = 0.22$, and $\theta_0 = 15°$. The first column represents noise alone, the second column represents $SNR = 0$ dB, and the third column represents signal alone.

radius was 1.62 m and the absorption coefficient was again 0.6.

The desired signal source angle varied between 0° and 12° and the interfering noise source angle varied between 18° and 90°, both in 4° increments. For each of the resulting 76 combinations of signal and noise source angles, the system generated predictions of high and low SNRs for each 10-millisecond frame. These results were then compared to the true SNRs for each frame to determine the detection and false alarm rates.

5.1. Simulations with white Gaussian noise

Simulations were performed using desired signal and interfering noise sources consisting of 28000-sample long segments of white Gaussian noise. The variance of the interfering noise source was constant at a value of one. The desired signal source consisted of a series of 2000-sample intervals each with a constant variance; the variance increased in steps of 3 dB between intervals such that the SNR ranged from −19.5 dB to 19.5 dB. This input is structured so that the SNR is less than 0 dB for the first 14000 samples, and the SNR is greater than 0 dB for the last 14000 samples. Thus, the first half of the signal was used to determine the false alarm rate P_F, and the second half was used to determine the detection rate P_D. The values of P_D and P_F were averaged over all combinations of source angles for desired signals and interfering noise.

All of the simulations with white noise used an intermicrophone spacing of $d = 14$ cm together with two sets of system parameters. In the first set, $kd = \pi$ and $r_0 = 0.1$. With $d = 14$ cm, this results in a center frequency of $f_0 = 1238$ Hz. In the second parameter set, $kd = (4/3)\pi$ and $r_0 = 0$, resulting in a value of $f_0 = 1650$ Hz. For both parameter sets, the fractional bandwidth B' varied between 0.1 and 1.5, corresponding to actual bandwidths of 124 Hz to 1856 Hz for the first parameter set and 165 Hz to 2475 Hz for the second set.

Figure 8 shows the results of these simulations, displaying the detection, error, and false alarm rates as functions of fractional bandwidth for the two values of kd and three levels of reverberation. This figure also includes the probabilities

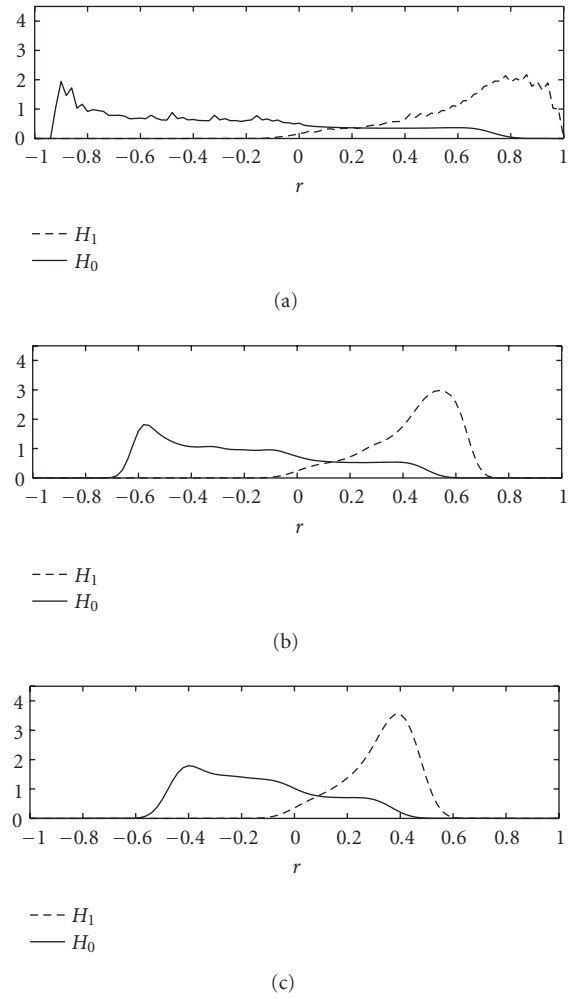

Figure 6: Conditional probability density functions of the estimated intermicrophone correlation coefficient for the two hypotheses $f_{r|H_0,\beta}(r \mid H_0, \beta)$ and $f_{r|H_1,\beta}(r \mid H_1, \beta)$, computed as in (29) with relative center frequency of $kd = \pi$, fractional bandwidth $B' = 0.22$, and $\theta_0 = 15^\circ$ for three levels of reverberation (a) DRR = $+\infty$, (b) DRR = 3 dB, (c) DRR = 0 dB.

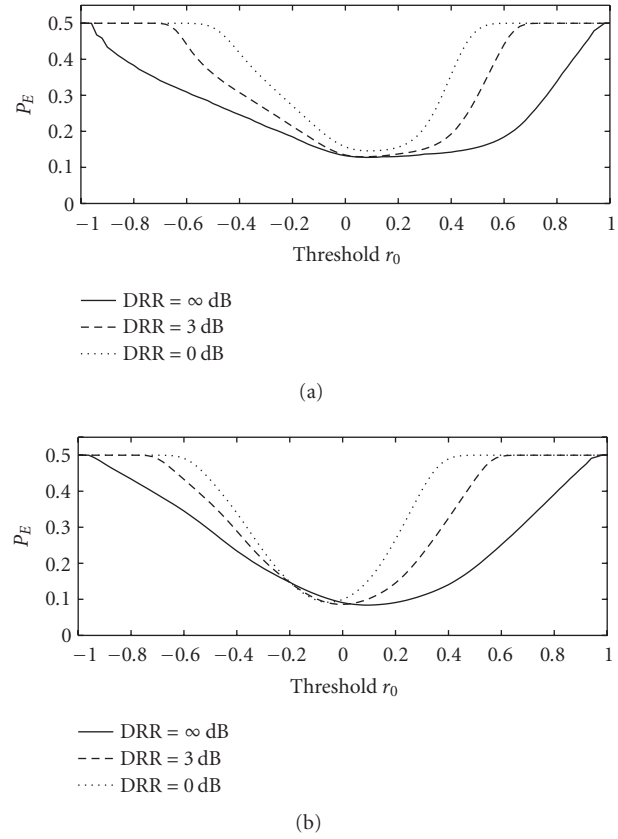

Figure 7: Probability of error P_E as a function of threshold r_0 for two values of relative center frequency (kd = (a) π, (b) $4\pi/3$) and three levels of reverberation (DRR = 0, 3, and $+\infty$ dB), with fractional bandwidth $B' = 0.22$ and $\theta_0 = 15^\circ$.

of detection, false alarm, and error as predicted by the analysis in Section 4. The agreement between the analytic and simulation results is quite good, especially for the anechoic condition. Minor but systematic deviations are apparent in the false alarm and error rates for the reverberant conditions, which is not surprising considering the oversimplified model of reverberation as a spherically diffuse sound field that was used in the analysis, but not in the simulations.

Overall, the best performance is obtained with low-to-moderate values of the fractional bandwidth. As predicted by Figure 4, large values of the fractional bandwidth increase the overlap between the pdfs, thereby increasing the error rate. However, the noise simulation results indicate that performance is relatively constant for a relatively wide range of fractional bandwidths. While both values of kd perform comparably, there is a slight benefit in using $kd = (4/3)\pi$.

5.2. Simulations with speech

More realistic simulations were performed using speech as the desired signal and babble as the noise signal. The speech source was 7-second long, formed by concatenating two sentences [31] spoken by a single male talker. The noise source consisted of 12-talker SPIN babble [32] trimmed to the same length as the speech material and normalized to have the same total power. The "true" SNR was calculated for each 10-millisecond frame by taking the ratio of the total power in the speech segment to the total power in the babble segment. The "true" SNRs were compared to the system outputs to determine the detection and false alarm rates, which were averaged over all combinations of signal and noise angles.

The speech simulations investigated three intermicrophone spacings $d = 7, 14$, and 28 cm, all with $kd = (4/3)\pi$ and $r_0 = 0$.[1] This resulted in center frequencies of $f_0 = 3300, 1650$, and 825 Hz for $d = 7, 14$, and 28 cm, respectively. The fractional bandwidth varied between 0.1 and 1.5. For $d = 7$ cm,

[1] Speech simulations were also performed with $kd = \pi$ and $r_0 = 0.1$. However, since the effect of kd on performance was comparable for both speech and noise simulations, those results are not presented here.

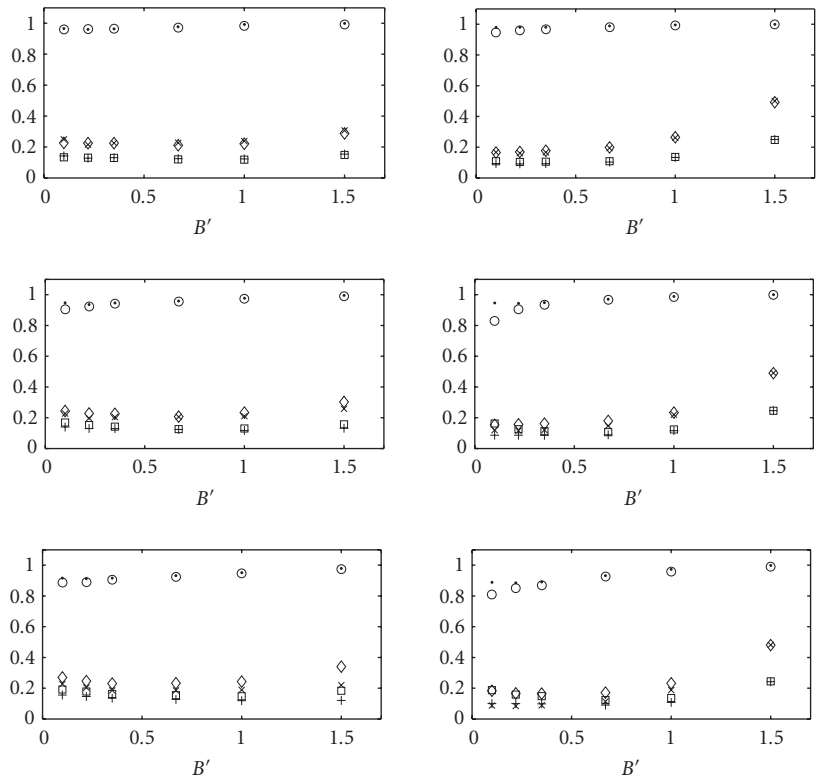

Figure 8: System performance as a function of fractional bandwidth B' for three levels of reverberation (DRR = 0, 3, and $+\infty$ dB) and two values of relative center frequency ($kd = \pi, 4\pi/3$). The plots show detection rates (circle), false alarm rates (diamond), and error rates (square) from the simulations with white noise along with the theoretical probabilities of detection (dot), false alarm (x), and error (+) predicted by the analysis in Section 4. The first row represents DRR = ∞, the second row represents DRR = 3 dB, and the third row represents DRR = 0 dB. The first column represents $kd = \pi$ and the second column represents $kd = 4/3\pi$.

the larger fractional bandwidths ($B' = 1.0$ and 1.5) were not simulated because they corresponded to frequency ranges that exceeded the signals' 5 kHz bandwidth.

Figure 9 shows the results of these simulations, displaying the detection, error, and false alarm rates as a function of fractional bandwidth for three values of d and three levels of reverberation. Comparing the columns in Figure 9 confirms that the overall performance is relatively unaffected by microphone spacing when comparing systems based on the normalized parameters kd and B'. The exception is the smaller microphone spacing ($d = 7$ cm), where small fractional bandwidths produce relatively more detections and false alarms, leading to comparable overall error rates.

Comparing the middle column of Figure 9 to the right-hand column of Figure 8 reveals that for the same parameter settings, the use of speech signals leads to substantial reductions in system performance, as evidenced by higher error and false alarm rates and lower detection rates. The discrepancies between Figures 8 and 9 are explained by the observation that in the case of the speech signals, the SNRs are not uniformly distributed in the range -20 dB to 20 dB, as was assumed in the analysis. This assumption was true for the noise simulation. In the case of speech, values of the short-time SNR tend to be concentrated at less extreme values, where the system does not perform as well. In fact, the majority of errors made by the system occur when the SNR is close to 0 dB, and therefore in transition between the two hypotheses. This is illustrated in Figure 10, which shows the true short-term SNRs for a 3-second speech segment and the values of intermicrophone correlation computed according to (1), along with the locations of misses and false alarms.

Another major difference between Figures 8 and 9 is the more pronounced effect of bandwidth on speech when compared with noise sources. For the noise signals, the energy was uniformly distributed across the bandwidth, but this is not the case for speech signals. As discussed in Section 4, selection of the bandwidth represents a tradeoff between the theoretical considerations, which dictate smaller bandwidths, and practical considerations, which require that the system captures sufficient energy from the nonstationary speech signal to minimize adverse affects of the relative energy fluctuations in different frequency regions. The simulation results in Figure 9 suggest that for speech signals, fractional bandwidths in the range 0.67 to 1.0 yield the best performance.

6. SUMMARY AND CONCLUSIONS

This paper describes a system for determining intervals of "high" and "low" signal-to-noise ratios when the signal and

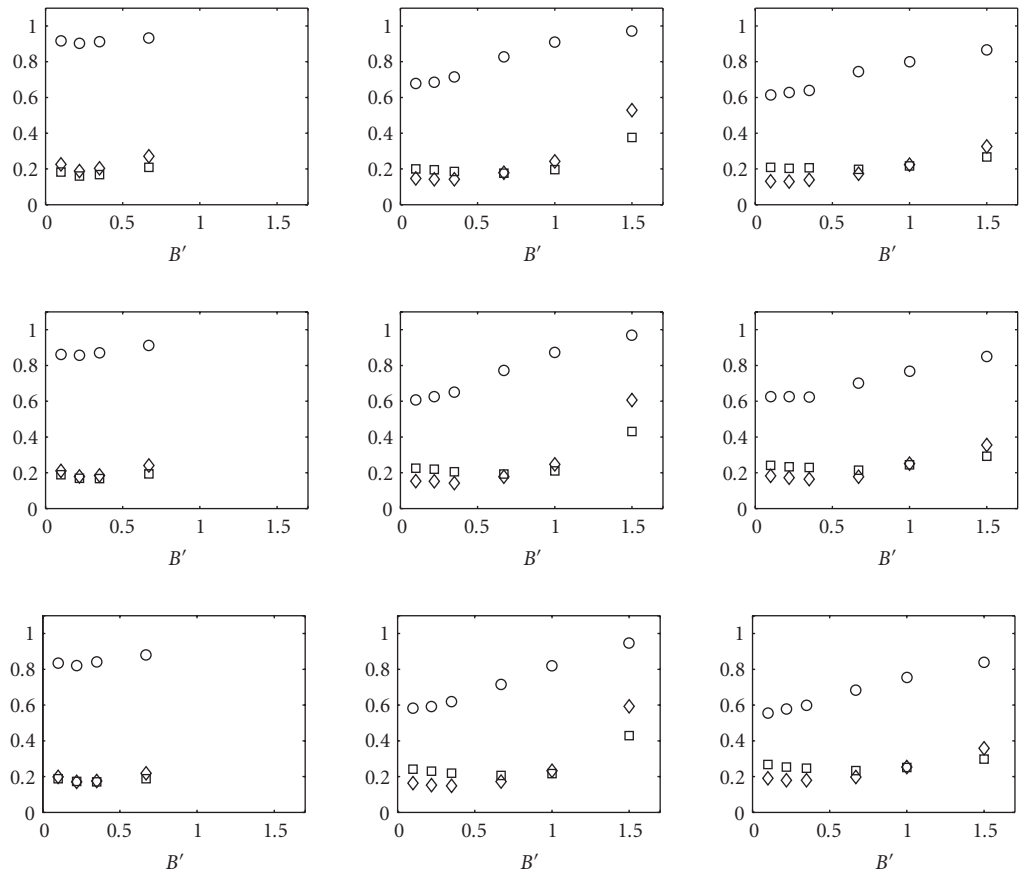

Figure 9: System performance as a function of fractional bandwidth B' for three levels of reverberation (DRR = 0, 3, and $+\infty$ dB) and three intermicrophone spacings ($d = 7, 14, 28$ cm), with relative center frequency ($kd = 4\pi/3$). The plots show detection rates (circle), false alarm rates (diamond), and error rates (square) from the simulations with speech. The first row represents DRR = ∞, the second row represents DRR = 3 dB, and the third represents DRR = 0 dB. The first column represents $d = 7$ cm, the second column represents $d = 14$ cm, and the third column represents $d = 28$ cm.

noise arise from distinct spatial regions. It uses the correlation coefficient between two microphone signals as the decision variable in a hypothesis test. The system has three parameters: the center frequency of the bandpass filter, the bandwidth of the bandpass filter, and the threshold for the decision variable. We performed a theoretical analysis based on a signal scenario that includes two spatially separated sound sources and a simple model of reverberation. By deriving conditional probability density functions of the intermicrophone correlation coefficient under both hypotheses, we gained insight into optimal selection of the system parameters. Results of simulations using white Gaussian noise for the sound sources were in close agreement with the theoretical results. More realistic simulations using speech sources followed the same general trends and illustrated the performance that can be obtained in practical situations with the parameters determined by the analysis, specifically, $kd = (4/3)\pi$, $B' = 0.67 - 1.0$, and $r_0 = 0$.

The contributions of this work are twofold. First, it provides an example of how speech detection systems can be analyzed and optimized. Rigorous comparison of the many speech detection systems proposed in the literature is often hampered by the differing conditions under which they are evaluated. If theoretical analyses similar to the one performed here were available, they would greatly facilitate the comparison of different speech detection systems. Second, for the particular speech detection system considered here, the analysis provides simple and widely applicable guidelines for the selection of parameters.

The system considered in this work is only applicable in situations when two microphone signals are available. It is further limited in that it is only expected to work in mild-to-moderate reverberation. The current study was restricted to a signal model consisting of a broadside array configuration, microphones in free space, a single interfering noise source, and simple models of reverberation. Future work should (1) consider endfire array configurations; (2) investigate the effect of mounting the microphones near the head for the hearing-aid application; (3) assess the performance of the system in the presence of multiple interferers; (4) quantify the degradation in performance with increasing levels of reverberation; and (5) evaluate the system with recorded (rather than simulated) sound signals. A study addressing these issues will more completely establish the potential of

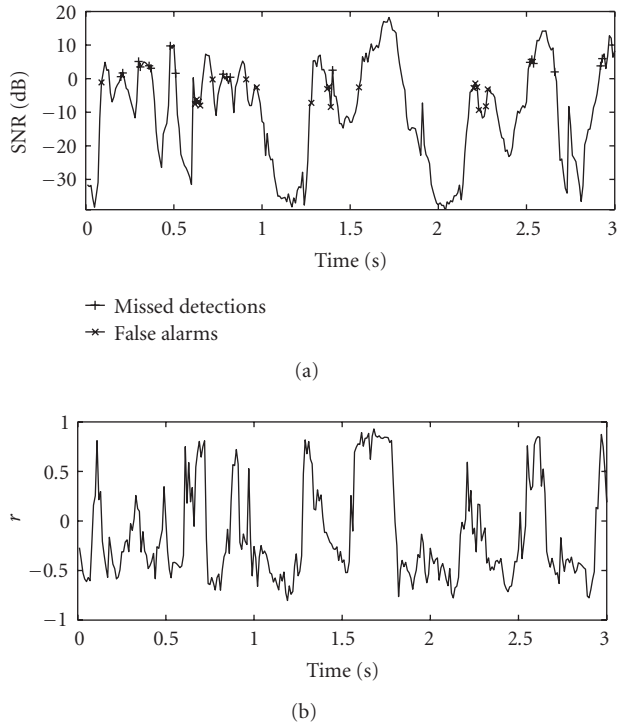

Figure 10: Simulation results for a desired speech source at 8° and interfering babble at 86° azimuth, combined to produce a long-term SNR of 0 dB. The sources were in an anechoic environment with 14 cm microphone spacing. (a) Short-time SNR as a function of time for a 3-second segment of speech. (b) Estimated intermicrophone correlation coefficient r for the same speech and babble segment as in (a), computed for $kd = 4/3\pi$ and $B' = 0.22$. Using a threshold of $r_0 = 0$, the symbols in (a) indicate frames, where there were missed detections ("+") and false alarms ("x").

the proposed system for use in speech-enhancement and noise-reduction algorithms that require identification of intervals when the desired signal is weak or absent.

ACKNOWLEDGMENTS

The authors are grateful to Pat Zurek, who suggested the use of the Fisher Z-transformation and outlined portions of the derivation presented in Section 3, and to three anonymous reviewers, who provided valuable feedback on an earlier version of this paper. This work was supported by the National Institute of Deafness and Other Communicative Disorders under Grant 1-R01-DC00117.

REFERENCES

[1] R. Plomp, "Auditory handicap of hearing impairment and the limited benefit of hearing aids," *Journal of the Acoustical Society of America*, vol. 63, no. 2, pp. 533–549, 1978.

[2] T. C. Smedley and R. L. Schow, "Frustrations with hearing aid use: candid reports from the elderly," *The Hearing Journal*, vol. 43, no. 6, pp. 21–27, 1990.

[3] S. Kochkin, "MarkeTrak V: consumer satisfaction revisited," *The Hearing Journal*, vol. 53, no. 1, pp. 38–55, 2000.

[4] S. Kochkin, "MarkeTrak V: 'why my hearing aids are in the drawer': the consumers' perspective," *The Hearing Journal*, vol. 53, no. 2, pp. 34–42, 2000.

[5] D. Van Compernolle, "Hearing aids using binaural processing principles," *Acta Oto-Laryngologica: Supplement*, vol. 469, pp. 76–84, 1990.

[6] M. Kompis and N. Dillier, "Noise reduction for hearing aids: Combining directional microphones with an adaptive beamformer," *Journal of the Acoustical Society of America*, vol. 96, no. 3, pp. 1910–1913, 1994.

[7] J. E. Greenberg and P. M. Zurek, "Evaluation of an adaptive beamforming method for hearing aids," *Journal of the Acoustical Society of America*, vol. 91, no. 3, pp. 1662–1676, 1992.

[8] D. Van Compernolle, W. Ma, F. Xie, and M. Van Diest, "Speech recognition in noisy environments with the aid of microphone arrays," *Speech Communication*, vol. 9, no. 5-6, pp. 433–442, 1990.

[9] H. Kobatake, K. Tawa, and A. Ishida, "Speech/nonspeech discrimination for speech recognition system under real life noise environments," in *Proc IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '89)*, vol. 1, pp. 365–368, Glasgow, Scotland, UK, May 1989.

[10] D. K. Freeman, G. Cosier, C. B. Southcott, and I. Boyd, "The voice activity detector for the Pan-European digital cellular mobile telephone service," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '89)*, vol. 1, pp. 369–372, Glasgow, Scotland, UK, May 1989.

[11] M. Marzinzik and B. Kollmeier, "Speech pause detection for noise spectrum estimation by tracking power envelope dynamics," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 2, pp. 109–118, 2002.

[12] C. Breining, P. Dreiscitel, E. Hansler, et al., "Acoustic echo control. An application of very-high-order adaptive filters," *IEEE Signal Processing Magazine*, vol. 16, no. 4, pp. 42–69, 1999.

[13] J. Stegmann and G. Schroder, "Robust voice-activity detection based on the wavelet transform," in *Proceedings of IEEE Workshop on Speech Coding For Telecommunications Proceeding*, pp. 99–100, Pocono Manor, Pa, USA, September 1997.

[14] R. Tucker, "Voice activity detection using a periodicity measure," *IEE Proceedings. I: Communications, Speech, and Vision*, vol. 139, no. 4, pp. 377–380, 1992.

[15] J. Pencak and D. Nelson, "The NP speech activity detection algorithm," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '95)*, vol. 1, pp. 381–384, Detroit, Mich, USA, May 1995.

[16] J. D. Hoyt and H. Wechsler, "Detection of human speech in structured noise," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '94)*, vol. 2, pp. 237–240, Adelaide, Australia, April 1994.

[17] J. T. Sims, "A speech-to-noise ratio measurement algorithm," *Journal of the Acoustical Society of America*, vol. 78, no. 5, pp. 1671–1674, 1985.

[18] M. Akagi and T. Kago, "Noise reduction using a small-scale microphone array in multi noise source environment," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '02)*, vol. 1, pp. 909–912, Orlando, Fla, USA, May 2002.

[19] M. W. Hoffman, Z. Li, and D. Khataniar, "GSC-based spatial voice activity detection for enhanced speech coding in the presence of competing speech," *IEEE Transactions Speech Audio Processing*, vol. 9, no. 2, pp. 175–178, 2001.

[20] R. Le Bouquin-Jeannès and G. Faucon, "Study of a voice activity detector and its influence on a noise reduction system," *Speech Communication*, vol. 16, no. 3, pp. 245–254, 1995.

[21] M. Kompis, N. Dillier, J. Francois, J. Tinembart, and R. Hausler, "New target-signal-detection schemes for multimicrophone noise-reduction systems for hearing aids," in *Proceedings of 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS '97)*, vol. 5, pp. 1990–1993, Chicago, Ill, USA, October–November 1997.

[22] R. J. M. van Hoesel and G. M. Clark, "Evaluation of a portable two-microphone adaptive beamforming speech processor with cochlear implant patients," *Journal of the Acoustical Society of America*, vol. 97, no. 4, pp. 2498–2503, 1995.

[23] P. Janecek, "A model for the sound energy distribution in work spaces based on the combination of direct and diffuse sound fields," *Acustica*, vol. 74, pp. 149–156, 1991.

[24] M. G. Bulmer, *Principles of Statistics*, Dover, New York, NY, USA, 1979.

[25] W. M. Hartmann, *Signals, Sound, and Sensation*, Springer, New York, NY, USA, 1998.

[26] H. P. Hsu, *Probability, Random Variables, and Random Processes*, McGraw-Hill, New York, NY, USA, 1997.

[27] H. Nélisse and J. Nicolas, "Characterization of a diffuse field in a reverberant room," *Journal of the Acoustical Society of America*, vol. 101, no. 6, pp. 3517–3524, 1997.

[28] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*, John Wiley & Sons, New York, NY, USA, 1968.

[29] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.

[30] P. M. Peterson, "Simulating the response of multiple microphones to a single acoustic source in a reverberant room," *Journal of the Acoustical Society of America*, vol. 80, no. 5, pp. 1527–1529, 1986.

[31] IEEE, "IEEE recommended practice for speech quality measurements," Tech. Rep. IEEE 297, Institute of Electrical and Electronics Engineers, Washington, DC, USA, 1969.

[32] D. N. Kalikow, K. N. Stevens, and L. L. Elliot, "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," *Journal of the Acoustical Society of America*, vol. 61, no. 5, pp. 1337–1351, 1977.

Ashish Koul received the B.S. and M.Eng. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology in 2001 and 2003, respectively. While at MIT, he served as a Research Assistant in the Sensory Communications Group within the Research Laboratory of Electronics, where he was involved in applications of digital signal processing in hearing-aid design. Currently, he is employed as an Engineer working on research and development in the Broadband Video Compression Group at the Broadcom Corporation in Andover, Mass.

Julie E. Greenberg is a Principal Research Scientist in the Research Laboratory of Electronics at the Massachusetts Institute of Technology (MIT). She also serves as the Director of Education and Academic Affairs for the Harvard-MIT Division of Health Sciences and Technology (HST). She received a B.S.E. degree in computer engineering from the University of Michigan, Ann Arbor (1985), an S.M. in electrical engineering from MIT (1989), and a Ph.D. degree in medical engineering from HST (1994). Her research interests include signal processing for hearing aids and cochlear implants, as well as the use of technology in bioengineering education. She is a Member of IEEE, ASEE, and BMES.