Effect of Reflectors on Sound-Source Localization with Two Microphones*

SANDEEP A. PHATAK, RAMA RATNAM, BRUCE C. WHEELER, WILLIAM D. O'BRIEN, JR., AND**
(phatak@uiuc.edu) (Rama.Ratnam@utsa.edu) (bwheeler@uiuc.edu) (wdo@uiuc.edu)

ALBERT FENG
(afeng1@uiuc.edu)

Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

The localization performance of the source localization algorithms degrades in reverberant conditions. The performance of one such localization algorithm, the localization–extraction (LE) algorithm, was measured systematically as a function of the number of reflecting surfaces in a cubical enclosure. Localization was qualitatively measured using a localization plot and quantized using two objective parameters. A broad-band noise burst and a speech signal were used as stimuli. The degradation of the localization performance was monotonic but not uniform with an increase in the number of reflectors. The performance was found to be proportional to the bandwidth of the stimulus. The performance of the LE algorithm was benchmarked against that of a commonly used signal-subspace technique—multiple signal classification (MUSIC). The LE algorithm was less affected by reflections than the MUSIC algorithm. Degradation of the source localization under high reverberation was found to be more severe at low frequencies, which resulted in the detection of a "phantom" source at 0° for the speech signal.

0 INTRODUCTION

Reverberation severely affects human listening as well as the performance of binaural hearing aids. The performance of a binaural target-extraction algorithm for hearing aids degrades in reverberant conditions [1], [2], which can lead to poor performance of hearing aids. The binaural hearing-aid performance critically depends on the performance of the source localization algorithm in the hearing aids. Source localization is known to be adversely affected by reflections and reverberation [3]–[6]. Reflections from a reflecting surface interact with the source signal and affect the acoustic cues that are necessary for source location. In the presence of multiple reflecting surfaces, multiple reflections of sound result in a reverberant tail after the early reflections [7], making the acoustic scene more complex and the task of source localization more difficult. The human localization performance was measured quantitatively and found to degrade in the presence of one reflector [8] and six reflectors [9]. Such a systematic measurement of the effect of multiple reflecting surfaces on the performance of a localization algorithm is not available. In this study the source localization was measured as a function of the number of reflecting surfaces.

The localization–extraction (LE) algorithm [10] was primarily used to measure the source localization. A source localization algorithm can estimate the spatial distribution of coherent acoustic energy [11]. We call the graphical representation of this distribution a localization plot. The interaural cross correlation [12], the intermicrophone phase correlation [13], and a histogram of the detected direction of arrival [10], [14], [15] represent functionally similar information. Visual inspection of the localization plots and a statistical analysis of the error in the detected location of the source peak are the two commonly used techniques for analyzing localization performance [5], [10], [14], [16]. In order to compare the source localization for different conditions, it is necessary to quantify the localization performance. While there is no consistently used standard metric for quantifying localization performance, parameters such as percentage of correct responses [10], source peak amplitude [11], [15], [16], and standard deviation [4], [16] have been used in the past. Two such objective parameters (the peak amplitude and the variance, see Section 1.4) estimated from the localiza-

*Manuscript received 2005 September 20; revised 2006 February 12 and March 20.

**Now with the University of Texas, San Antonio, TX.

tion plots were used in this study to quantify and compare the localization performances.

The effect of reflectors on source localization was analyzed quantitatively using the signals recorded with two microphones in an acoustically controlled environment. The analysis was restricted to localization in azimuth only. The process of generating localization plots and calculating objective parameters was completely automated using MATLAB code. The performance of the LE algorithm was compared against that of a commonly used signal-subspace technique, multiple signal classification (MUSIC) [11]. Both localization algorithms showed monotonic degradation with increasing reverberation and the appearance of a "phantom" source for speech signals under high reverberation.

1 METHOD

1.1 Recording Setup

The microphone signals were recorded inside a plywood cube. The cube of size 1.83 m × 1.79 m × 1.83 m was constructed from approximately 1-inch-thick plywood panels. A loudspeaker and a pair of microphones were placed near the opposite corners of the cube at approximately 1.50 m from each other [Fig. 1(a)] to minimize the near-field effect of the loudspeaker. The playback of sounds and recording of the microphone signals were both controlled using MATLAB code on a laptop computer (Sony-VIAO, P4, WinXP). A power amplifier (ADCOM, GFA-535II) amplified the signals to drive the loudspeaker (ADS-L200e). The output of a pair of omnidirectional microphones (Sennheiser, MKE-2GOLD) was connected to the sound recording device (Sound Devices Inc., USBPre), which was interfaced to the laptop through a USB port. Both playback and recording were done at CD quality (16 bit, 44.1 kHz). Recorded signals were stored as WAV files and analyzed off line.

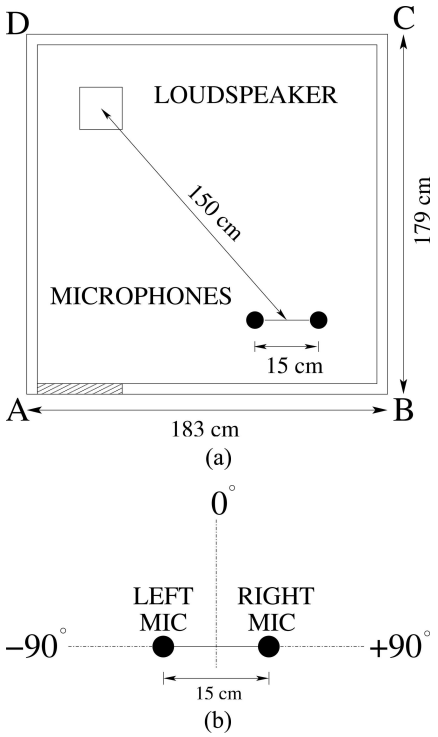


Fig. 1. (a) Top view of plywood cube used for recording signals. Shaded region near corner A shows entrance to cube. (b) Reference for determining source azimuth with respect to microphones.

The acoustic output of the loudspeaker may induce vibrations in the loudspeaker stand, which can adversely alter the loudspeaker response. Therefore a piece of acoustic foam was placed between the stand and the loudspeaker to minimize the mechanical coupling between the two. A similar technique was used to isolate the microphones mechanically from the microphone stand. The microphones were separated by 0.15 m and were adjusted to be in the horizontal plane of the loudspeaker.

The walls of the cube, including floor and ceiling, were converted into sound absorbers by fixing a 2-inch-thick layer of acoustic foam on plywood material inside the cube. Table 1 shows the absorption coefficient α of the foam as a function of frequency. The actual measurements indicated that the overall absorption coefficient of the foam was 0.85, that of the plywood was 0.19. The foam was removed from the walls, one at a time, to obtain the following six acoustically different conditions [see Fig. 1(a)]:

- No reflector ($R = 0$): All six surfaces with absorbing foam ($RT_{60} \approx 0.18$ s)
- One reflector ($R = 1$): One (CD) reflecting surface ($RT_{60} \approx 0.18$ s)
- Two reflectors ($R = 2$): Two (AB, CD) reflecting surfaces ($RT_{60} \approx 0.25$ s)
- Three reflectors ($R = 3$): Three (AB, BC, CD) reflecting surface ($RT_{60} \approx 0.28$ s)
- Four reflectors ($R = 4$): Four (AB, BC, CD, DA) reflecting surfaces ($RT_{60} \approx 0.60$ s)
- Six reflectors ($R = 6$): Six reflecting surfaces, including ceiling and floor ($RT_{60} \approx 1.0$ s).

The approximate values of the 60-dB reverberation times (RT_{60}) were obtained from the room impulse response (RIR) estimated using a maximum length sequence (MLS) signal [17]. The RIRs hit the noise floor after a 15–20-dB decay. The 60-dB decay point was therefore obtained by extrapolating from the early decay in the RIR by manually fitting a straight line to the decay curve. The RT_{60} values indicate the degree of reverberation for different acoustic conditions inside the plywood cube.

Table 1. Frequency-dependent absorption coefficient (α) of acoustic foam (melamine RPG ProFoam™) per technical specifications.

Frequency (Hz)	100	160	200	250	315	400	500	625
α	0.14	0.13	0.16	0.30	0.42	0.66	1.08	1.26
Frequency (Hz)	800	1000	1250	1600	2000	2500	3150	4000
α	1.14	1.11	1.00	0.96	0.99	0.99	1.03	1.01

Due to the small size of the plywood cube, it was not possible to change the source azimuth by moving the loudspeaker. Instead, three source azimuth positions were obtained by rotating the microphone pair. The microphone pair was initially oriented parallel to side AB [Fig. 1(a)], so that the loudspeaker was at −45° with respect to the frame of reference, shown in Fig. 1(b). The microphone pair was rotated to achieve orientations of 0° and +45°.

1.2 Signals

A speech signal ("The North Wind and the Sun," US-English female talker from narrative recordings of International Phonetic Association) and a broad-band Gaussian white-noise signal (cutoff at 20 kHz) were used as input sounds at the source. The signals and their spectra are shown in Fig. 2. The energy in the speech signal is concentrated at frequencies below 11 kHz. White Gaussian noise of 0.3-ms duration was generated using the *randn* function in MATLAB. Both signals were stored in 16-bit 44.1-kHz WAV format.

In addition to the two signals, an MLS signal was also played for calibration. The sound pressure level (SPL) for the MLS signal was measured using a sound level meter (B&K 2260 Investigator) and was 78.5 dB SPL with uniform frequency weights. Using the measured SPL for the MLS signal, the playback level of white noise and speech was determined to be 65.5 and 65.0 dB SPL, respectively, in the absence of reflectors. The ambient noise level was 38.2 dB SPL in the absence of reflectors and 43.2 dB SPL in the presence of six reflectors. The increase in the ambient noise was due to noise leakage from outside the plywood cube after removal of the acoustic foam. Thus the average signal-to-noise ratio (SNR) was about 25 dB. Furthermore the spectral analysis of the noise showed that 76.5% of the background noise energy was at frequencies below 200 Hz. Thus the background noise was negligible with respect to the signal at frequencies above 200 Hz.

1.3 Localization Algorithms

1.3.1 Localization–Extraction (LE) Algorithm

The LE algorithm [10] is based on a delay-line model of the localization mechanism in the human midbrain, proposed by Jeffress [18]. LE localizes sources in the frontal half of the horizontal plane and extracts the selected sources. We used only the localization part of the algorithm.

The LE is a frequency-based algorithm, that is, the source is localized independently in each frequency bin. In each frequency bin a dual delay line adds phase delays to the signal in one channel and advances the phase of the signal in the other channel, in uniform steps, and subtracts the pairs of phase-modified outputs. When the phase difference between the two channels, introduced by the delay lines, compensates for the intermicrophone phase delay, the difference between the pair of phase-modified outputs is minimum. Thus each subtraction unit corresponds to a particular intermicrophone phase difference (IPD), which in turn is mapped to intermicrophone time difference (ITD). This mapping is not one to one for wavelengths smaller than twice the intermicrophone distance, and so multiple ITDs can result in the same IPD [10]. The ITD, however, has a one-to-one mapping with the azimuth in

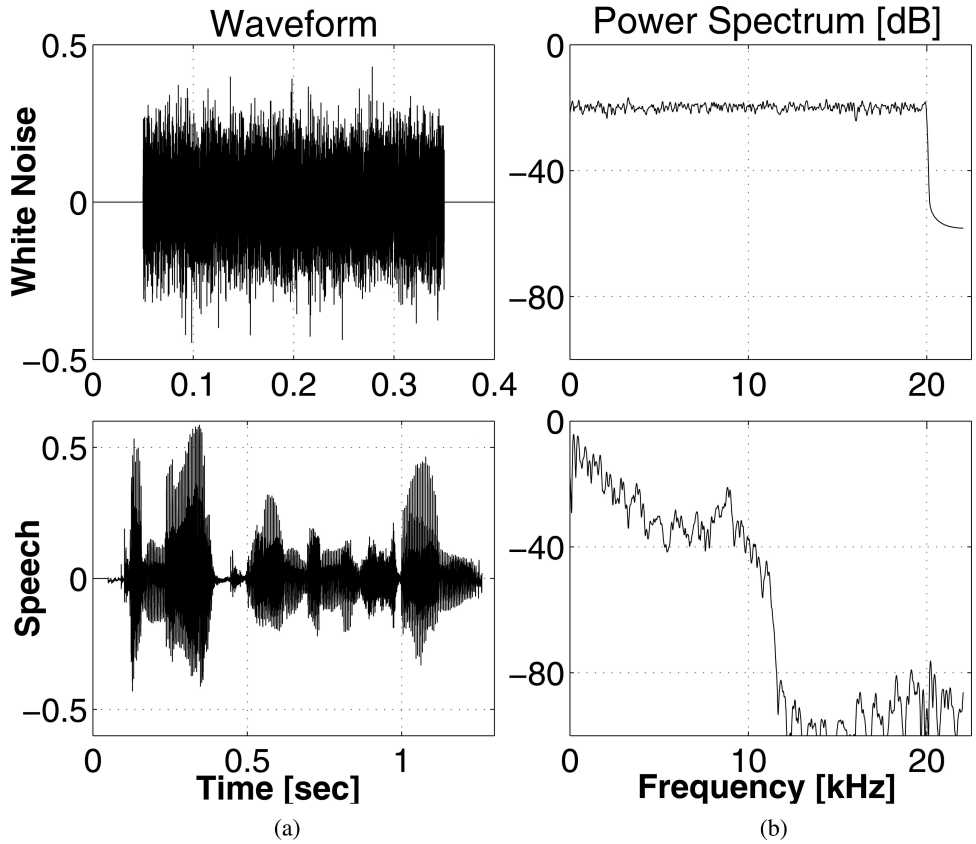


Fig. 2. Signals used. (a) White noise and speech. (b) Corresponding power spectral densities.

the frontal horizontal plane. Thus a source at a particular azimuth will give rise to "ambiguous" sources at all those azimuths that yield the same IPD as that yielded by the true azimuth. The number and location of the ambiguous sources depend on the frequency, giving rise to the ambiguity contours in the coincidence plot [Fig. 3(a)], which is a plot of detected source azimuth in each frequency bin. When the coincidence plot is summed over frequency, we get an ensemble of detected source locations, called the localization plot [Fig. 3(b)]. The multiple contours, visible in the coincidence plots, are due to the phase ambiguity. The contour corresponding to the correct source location is almost a vertical line because the ITD for the actual source location is constant at all frequencies. The ambiguity contours are curved as the phase-ambiguous ITDs change with frequency. There are more phase ambiguities at higher frequencies than at lower frequencies.

It should be noted that although the LE algorithm is biologically inspired, it deals only with IPD and does not take into account the intermicrophone intensity difference (IID), which is an important localization cue for mammals. Interaural intensity cues, primarily caused by the head shadow, are particularly helpful at higher frequencies, where IPD produces location ambiguities. IID information may be useful in applications such as the binaural hearing aid, but is not relevant to localization using the free-field microphones.

In the MATLAB implementation of the LE algorithm that was used in this study, the azimuth in the frontal plane (−90° to +90°) was divided into 361 ITD bins with equal time-delay increments. When the intermicrophone distance d is very small compared to the source–microphone distance, the intermicrophone time difference τ is related to the azimuth θ by the trigonometric relationship

$$\tau = \frac{d}{c} \cdot \sin\theta \tag{1}$$

where $c = 343$ m/s is the speed of sound. The azimuth corresponding to each bin varies nonlinearly from −90° to +90°, with the best angular resolution available in the medial plane, that is, at 0°, where the ITD is also zero.

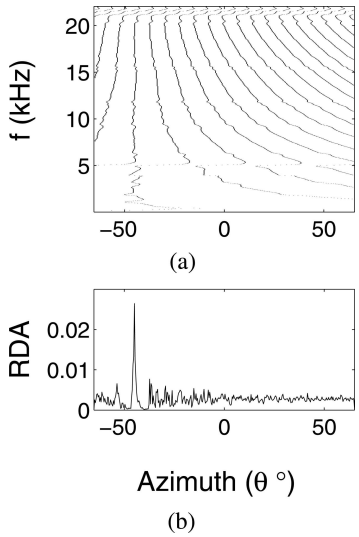


Fig. 3. (a) Typical coincidence plot for white-noise signal, no-reflector condition. (b) Corresponding localization plot. Ordinate-Relative detection amplitude (RDA).

The signals were windowed using a 5-ms Hamming window with 1.25-ms overlap. The windowed signals were filtered in 1024 equally spaced frequency bins via a 2048-point fast Fourier transform. A coincidence plot and a localization plot were obtained for each time frame.

1.3.2 Multiple-Signal-Classification (MUSIC) Algorithm

MUSIC [11] is a covariance-based localization algorithm that estimates the direction of arrival (DOA) of wave fronts arriving from multiple sources using the eigenstructure of the covariance matrix of the input signal. MUSIC models the data as a superposition of point sources and uncorrelated noise. If there are M sensors and D sources, then the covariance matrix has D nonzero eigenvalues, and the eigenvectors corresponding to them constitute the signal subspace. The remaining N eigenvectors that correspond to the zero eigenvalues constitute the noise subspace, where $N = MD$. If $E(\theta)$ is the $M \times N$ matrix of the eigenvectors in the noise subspace and if $a(\theta)$ is the steering vector that characterizes the amplitude and the phase of a wave front arriving from azimuth θ (that is, the DOA), then the energy is given by

$$P(\theta) = \frac{1}{a^*(\theta) \cdot E(\theta) \cdot E^*(\theta) \cdot a(\theta)} \tag{2}$$

where * represents the complex conjugate. The contour $P(\theta)$ is, by definition, the localization plot.

A frequency-based MUSIC localization algorithm that localizes the sources independently in each frequency bin was used for this study. For $P(\theta)$ to resolve D sources, the condition $D < M$ should be met, and because the value of M, that is, the number of sensors, for this experiment is 2, the maximum value that D can have is 1. In other words, with two microphones the frequency-based MUSIC localization algorithm cannot localize more than one source in each frequency bin. However, it can detect different sources in different frequency bins, depending on the distribution of the spectral strength of the sources.

A MATLAB implementation [19] of the two-dimensional MUSIC algorithm was used for this study. The number of ITD bins was set to 361 in order to be consistent with the implementation of the LE algorithm.

1.4 Localization Plot and Objective Parameters

Localization plots have been used commonly to qualitatively report the performance of localization algorithms. A localization plot is an ensemble of detected source localizations as a function of azimuth and is obtained by summing the coincidence plot over frequency. Therefore the localization plots obtained by integrating over different frequency ranges and different lengths of the signal cannot be compared directly. However, a localization plot, when normalized to have unit area under the curve, becomes an empirical probability distribution of detecting a source as a function of azimuth and can be compared across different conditions. The ordinate to the normalized localization

plot is labeled relative detection amplitude (RDA) [Fig. 3(b)]. In this study the localization performances of both algorithms were quantized using a set of two objective parameters that were extracted from the localization plot. The two objective parameters are as follows.

1) *Relative detection amplitude (RDA) of source peak* P_S RDA is the amplitude of the source peak in the normalized localization plots and is equivalent to the parameters normalized peak height [15] and percent correct score [10], used in the past. A higher RDA value of the source peak indicates a greater likelihood of localizing the source.

2) *Normalized variance of source peak* V_S The normalized variance of the source peak located in the S^{th} ITD bin is defined as

$$V_S = \frac{1}{P_S} \cdot \frac{1}{(\Delta\tau)^2} \cdot V'_S \tag{3}$$

where

$$V'_S = \frac{1}{2N+1} \sum_{n=S-N}^{S+N} P_n \cdot [(n-S) \cdot \Delta\tau]^2 \tag{4}$$

Here $\Delta\tau$ is the ITD bin width, that is, the time delay between consecutive bins, and P_n is the relative detection amplitude in the nth ITD bin. V'_S is an estimate of the centralized second moment of the source peak within $\pm N$ bins about the detected source peak. The value of V'_S depends on the amplitude of the source peak. Because the purpose of this parameter is to capture the shape, in particular, the width of the source peak, the effect of the height of the source peak is removed by normalizing the values of P_n, that is, dividing by the source amplitude P_S. Note that the factor P_S can be taken out of the summation. Finally, to make the quantity V'_S dimensionless and independent of the ITD bin width, it is normalized by dividing by $(2N + 1)$ to get the normalized variance V_S. The value of N is decided using the localization plot, so that $(S \pm N)$ covers the source peak but avoids the adjoining peaks. The value of N depends on the ITD bin width, and $N = 15$ was used in this experiment. The normalized variance is a measure of the width or the spread of the source peak. The spread of the source peak indicates the sharpness of the source as perceived by the localizer. Thus the value of the normalized variance of the source peak indicates the sharpness of the localized source peak.

Calculating the objective parameters requires the azimuth of the detected source peak. Though it was easy to visually locate the source peak in the localization plots, a peak detection scheme was necessary to automate the process. Peak detection became particularly difficult in reverberant conditions when the source peak in the localization plot was not sharp. Also, it was necessary to distinguish the source peak from the peaks that correspond to phase-ambiguous locations. A simple peak detection algorithm was used to detect the source peak location and was found to be working effectively even in reverberant conditions. The peak detection algorithm selected the highest peak within ±30 ITD bins of the approximate source ITD (that is, the ITD bin corresponding to one of the three source locations −45°, 0°, and 45°) to be the detected source peak.

2 RESULTS

2.1 Effect of Signal Spectrum

Fig. 3(a) shows a typical coincidence plot for a white-noise signal, and Fig. 3(b) shows the corresponding normalized localization plot. The plots were obtained using the LE algorithm in the absence of reflectors, with the source (the loudspeaker) located at −45° azimuth and at a distance of 1.50 m from the center of the microphone pair. The source was localized at the correct azimuth for frequencies up to 20 kHz for white noise [Fig. 3(a)] and up to 11 kHz for speech [Fig. 4(a)], which matches the upper cutoff of the bandwidths of the two signals (Fig. 2). Above these frequencies there was very weak spectral energy in the signals, and the source detection of the LE algorithm was unreliable. When the coincidence plots were summed over the entire frequency range, the localization plot for the white-noise burst [Fig. 3(b)] showed a sharp peak at the correct source azimuth, whereas the localization plot for speech [Fig. 4(b)] showed a relatively smaller peak at the correct source location. If the coincidence plot for speech was summed over the frequency range of 0–10 kHz, then the normalized localization plot showed a stronger source peak [Fig. 4(c)]. Hence the coincidence plot was summed from 0 to 10 kHz for further analysis with the speech signal. Note that the localization plots were normalized and hence could be compared directly, even if the coincidence plots were summed over different frequency ranges.

Because the localization decisions in the LE algorithm were based only on the ITDs and not on the IIDs, the localization was not affected by the intensity of the signal. This can be observed clearly in Fig. 5 The relative detection amplitude of the source peak P_S was high when the

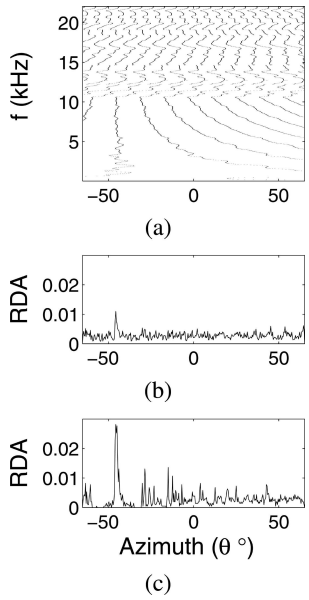


Fig. 4. (a) Typical coincidence plot in a time frame for speech signal, no-reflector condition. (b), (c) Corresponding localization plots when coincidence plot was summed; (b) over entire frequency range; (c) from 0 to 10 kHz.

spectrum was spread over a wide frequency range, independent of the intensity of the signal.

2.2 Effect of Reflectors

Fig. 6 shows the time-averaged localization plots generated using LE for a source at −45°. It can be observed that for both noise and speech, the source peak decreased when the number of reflectors was increased, and it almost diminished for two or more reflectors. As evident from the objective parameters, the degradation of the source peak was of two types: 1) the peak broadened and 2) it decreased in amplitude.

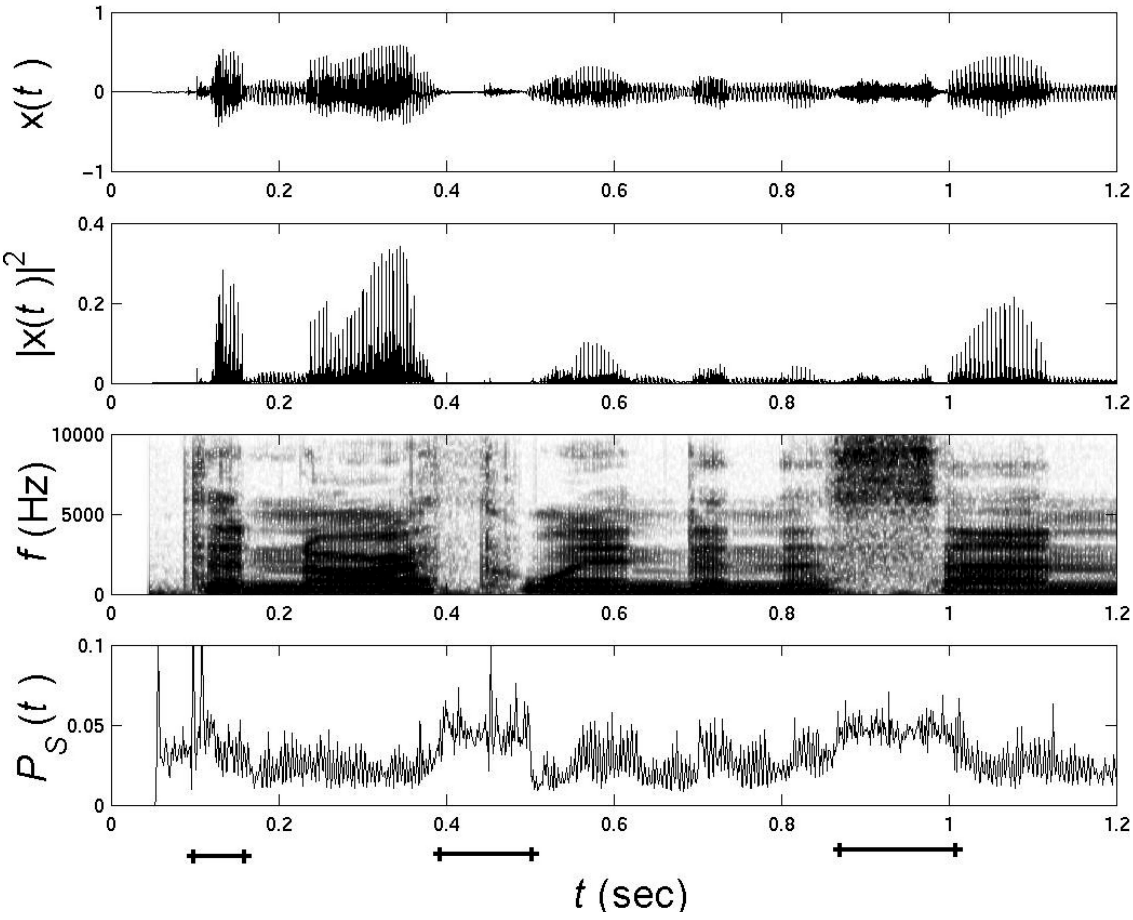


Fig. 5. Speech signal $x(t)$, square of its amplitude $|x(t)|^2$, spectrogram f (Hz), and corresponding RDA values $P_S(t)$. For calculating RDA, coincidence plot was summed over 0–10 kHz. Solid lines at bottom show the approximate intervals when $P_S(t)$ is relatively high.

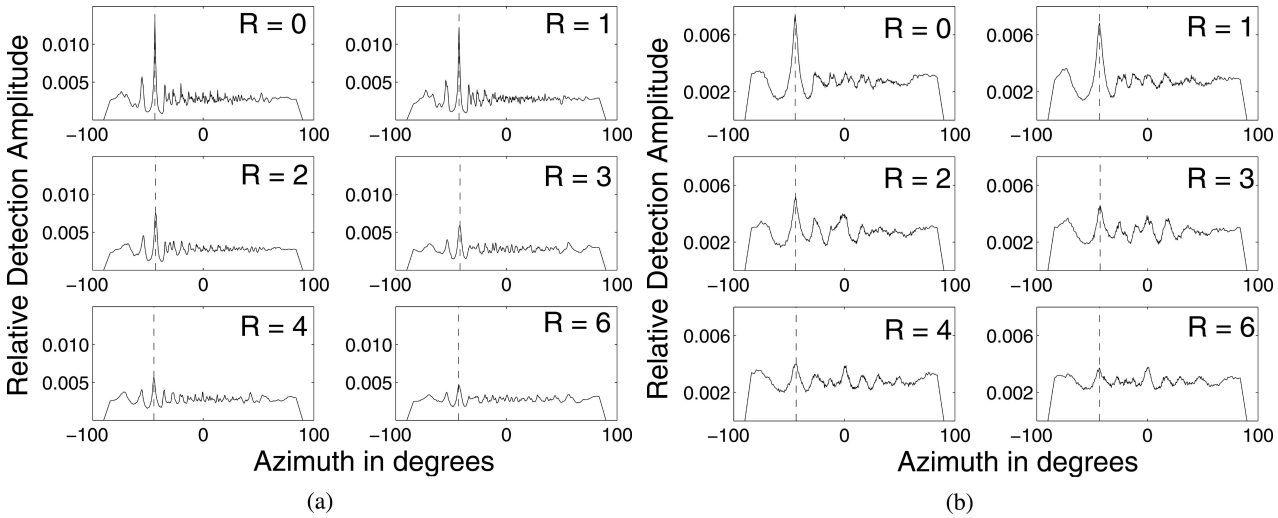


Fig. 6. Time-averaged localization plots for six acoustic conditions. (a) White noise. (b) Speech. Source azimuth ≈ −45°. R denotes the number of reflecting surfaces in each condition. Dashed vertical lines denote locations of detected source peaks. Note the difference in ordinate limits of the localization plots for two signals.

Figs. 7 and 8 show the means and the standard deviations of the objective parameters for the three source azimuths (–45°, 0°, +45°) as a function of the number of reflectors. The dotted lines with * markers show the objective parameter values at the intended source azimuth when the source was off. This was the baseline performance of the LE localizer. In the absence of any source, the probability of detecting the source should be uniform over all azimuth, and the localization plot should ideally be flat with the amplitude equal to 1/(# azimuth bins) = $1/361 = 2.77 \times 10^{-3}$. So theoretically, the baseline values of the relative detection amplitude and the normalized variance should be 2.77×10^{-3} and 80, respectively. The mean values of the objective parameters for the LE algorithm in the absence of a source were very close to the theoretically expected values. This was a sanity check for the localizer.

For white noise (Fig. 7) the mean value of the relative detection amplitude decreased from about 0.014 in the absence of reflectors to about 0.007 for six reflectors, a change of 50%. In the presence of six reflectors, the mean RDA value was close to the baseline performance. Comparatively, the mean value of the normalized variance of the source was farther from the baseline performance. The degradation of the objective parameters was greatest between one and four reflectors for white noise.

For the speech signal (Fig. 8) the change in the mean of the normalized variance was within one standard deviation for all conditions, but the mean RDA value decreased significantly from one to two reflectors. The normalized variance of the source peak for speech was always greater than that for white noise, indicating that the source forms a sharper peak with white noise. For the no-reflector and one-reflector conditions, the relative detection amplitude of the source peak was greater for white noise than for speech.

For both signals the degradation of the objective parameters was relatively less when the source was at 0°. This effect was more prominent for the speech signal. A closer analysis of the localization plot revealed the appearance of

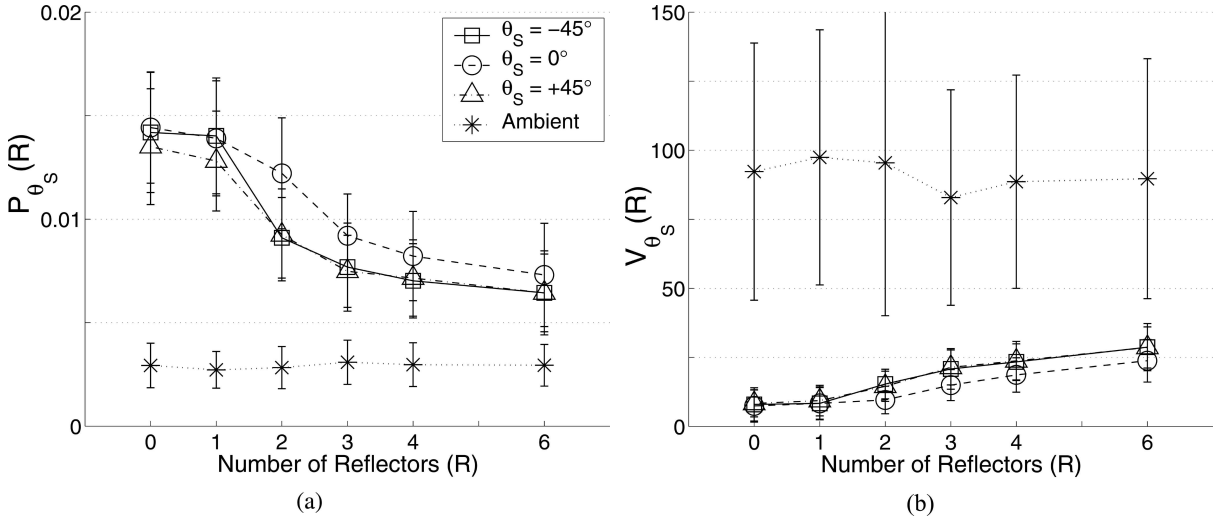


Fig. 7. Means and standard deviations for objective parameters. (a) RDA. (b) Normalized variance of source peak, obtained using LE for white noise and for three source azimuths. * indicates baseline performance in ambient condition (absence of source).

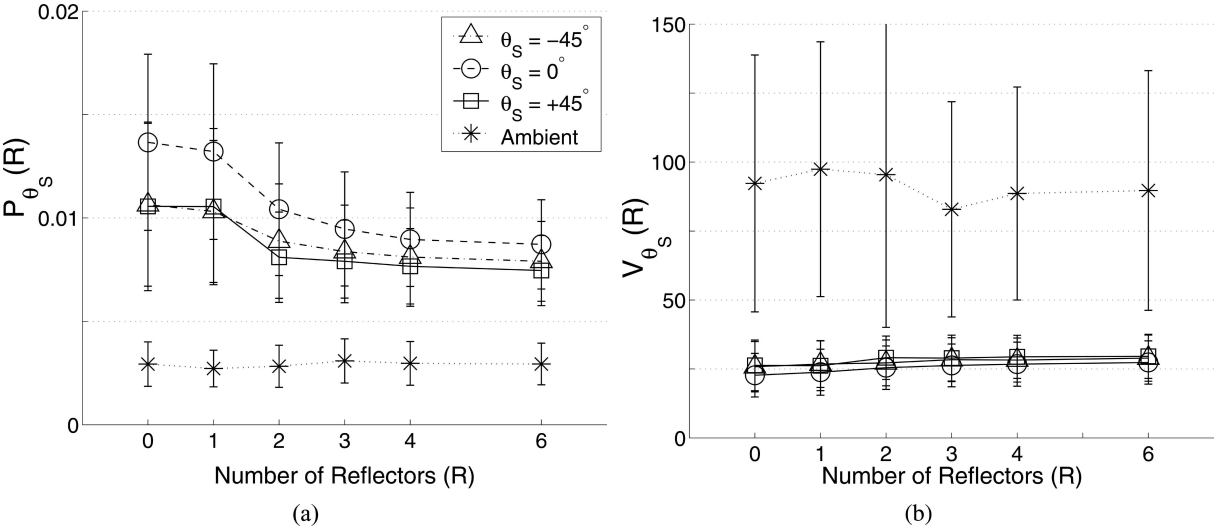


Fig. 8. Means and standard deviations for objective parameters. (a) RDA. (b) Normalized variance of source peak, obtained using LE for speech.

a weak peak centered at 0°. This peak was particularly significant for the speech signal in the presence of four or more reflectors. The position of this "phantom" source was always at 0°, irrespective of the source azimuth.

2.3 Comparison with MUSIC

The contours in the coincidence plots for MUSIC (not shown) were almost identical to those for the LE algorithm. Hence the coincidence plots were summed from 0 to 10 kHz for speech and over the entire frequency range for white noise. As described in Section 1.3, the coincidence plots for LE have the detected source azimuths, whereas the coincidence plots for MUSIC have the amplitudes of cross correlation between the signal from one microphone and the time-delayed signal from the other, as a function of the time delay. The time delays are mapped to the DOA according to Eq. (1). However, coincidence plots and localization plots for both algorithms represent the same information as far as localization is concerned.

MUSIC showed worse source localization (Fig. 9) than LE (Fig. 6). The phantom source peak at 0° for the speech signal is stronger and wider than that for LE, which masks the localization in the adjacent ITD bins. The appearance of the phantom source peak at the same location for the speech signal is independent of the source azimuth and is observed for both localization algorithms, although with different magnitudes. This suggests that the phantom source is a characteristic of the speech signal corrupted by multiple reflections and reverberation, and not of the localization algorithm.

Implementation of the MUSIC algorithm used for this study generated a single coincidence plot averaged over the entire duration of the signal, which was then summed over frequency to obtain a localization plot. Thus there was only one objective parameter value per condition for each of the two signals, and therefore there are no error bars in the graphs of the objective parameters (Figs. 10 and 11). Also, the implementation of MUSIC was designed to localize only when significant source energy was present. So the baseline performance of MUSIC in the absence of

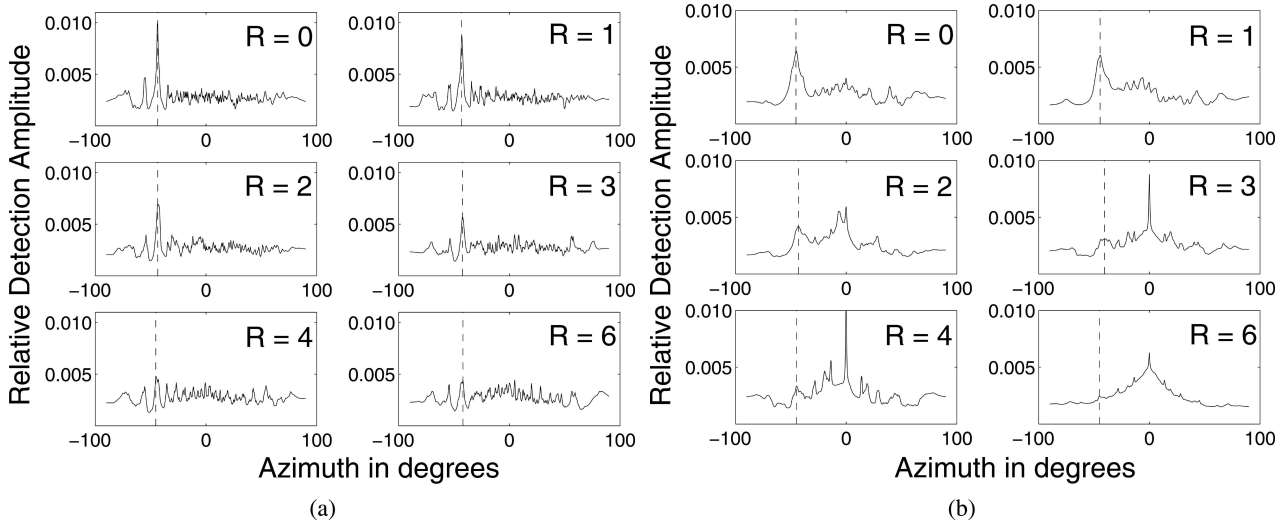


Fig. 9. Localization plots, obtained using MUSIC, for six acoustic conditions. (a) White noise. (b) Speech. Source azimuth ≈ −45°. Dashed vertical lines denote locations of detected source peaks.

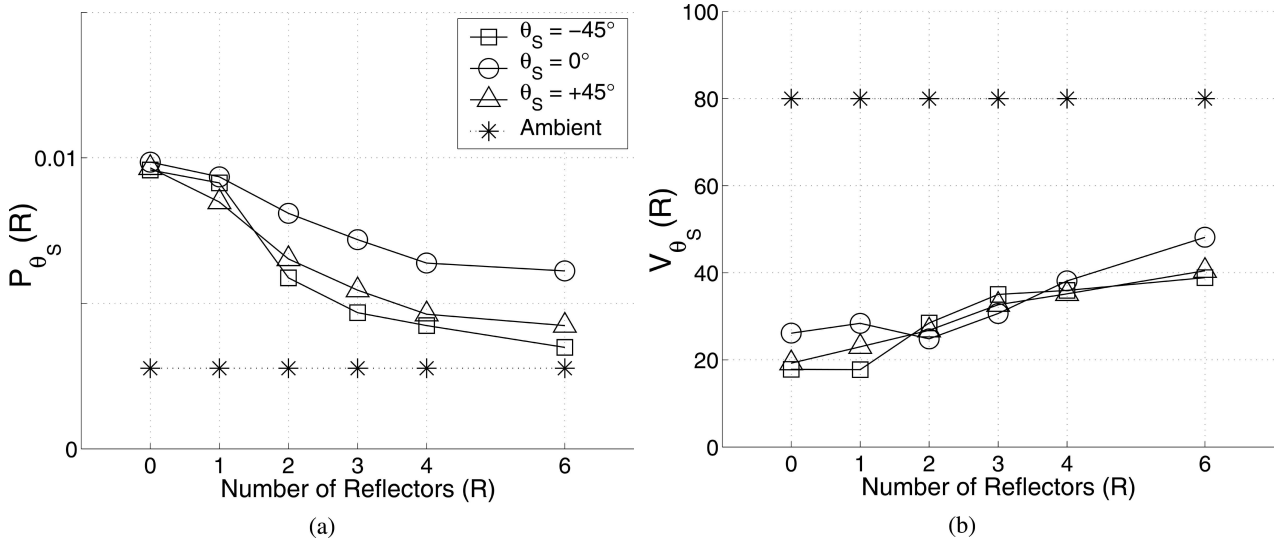


Fig. 10. Objective parameters. (a) RDA. (b) Normalized variance of source peak obtained using MUSIC for white noise and for three source azimuths. * indicates theoretically expected value of baseline performance in ambient condition.

a source could not be measured, and the theoretically expected values in the absence of a source are plotted for reference (dotted line, * marker) on the objective parameter graphs.

Unlike LE, there was a significant change in the normalized variance for both signals. However, similar to LE, most of the degradation of the objective parameters was between one and four reflectors for white noise and between one and two reflectors for speech. For speech, the RDA value dropped below the baseline performance; however, in the absence of error bars it was not possible to know the significance of this effect. For white noise the values of the objective parameters for MUSIC were always closer to the baseline performance than those for LE. But for speech the localization parameters showed very high performance when the source was at 0°. The high performance in this case was not due to the actual source peak, but to the phantom source peak that is generated at 0° by default.

3 DISCUSSION

The effect of the number of reflectors on source localization was measured using the localization plots and quantified using two objective parameters. There was a monotonic degradation of the localization performance with an increasing number of reflectors. The degradation effect is of two types: the probability of correctly detecting the source azimuth (given by RDA) decreased, and the source appeared to "broaden" in azimuth, as suggested by the increase in normalized variance. Both effects indicate that qualitatively, early reflections and reverberation reduce the coherence in the source signal, making it appear more diffused. Moreover, by quantifying the shape of the source peak in terms of the objective parameters, it is now possible to calibrate the localization performance in a reverberant room.

The performance of the frequency-based localization algorithm depends on the bandwidth of the source signal. Restricting the frequency range of summation for the coincidence plot to match the bandwidth of the speech signal helped improve the localization performance for the speech signal. However, the probability of localizing the correct source azimuth was greater and the source image was sharper for white noise than for the speech signal, even after adjusting the frequency range of summation of the coincidence plot for speech. The white noise, being a wide-band stationary signal, spans the entire frequency range of summation, resulting in consistently sharp source peaks in the localization plot. Speech, however, is a quasi-stationary signal, and its spectrum changes with time (Fig. 5). When the spectrum was wide enough in frequency to cover the range of summation, which was 0–10 kHz, the localization plot showed a strong source peak with a high relative detection amplitude. However, when the speech signal spectrum was narrow, the source peak had low amplitude, even if the speech signal was strong. Monitoring the speech spectrum over time and changing the frequency range of summation dynamically to match the time-varying speech spectrum may improve the localization performance for speech. Though human localization is much more complex because of nonlinear effects such as the precedence effect, human localization shows similar patterns. For example, humans have higher localization accuracy for wide-band noise or spectrally dense complex tones than for low-frequency, slow-onset tones, both in an absorbing room as well as in the presence of reflectors [8], [9]. The source peak amplitude in a localization plot is very high during the speech onset (~0.1 s in Fig. 5), which is characterized by a transient with broad spectrum. It is known that the speech onsets also play an important role in human localization [20]. The most likely reason for these similarities is that the mammalian auditory system integrates the localization information over frequency, and so does a frequency-based localization algorithm. It should be noted, however, that human localization performance is much superior to the localization algorithms in reverberant conditions due to complex operations in the human auditory system, such as echo suppression and the ability to latch on to a source after detecting its onset.

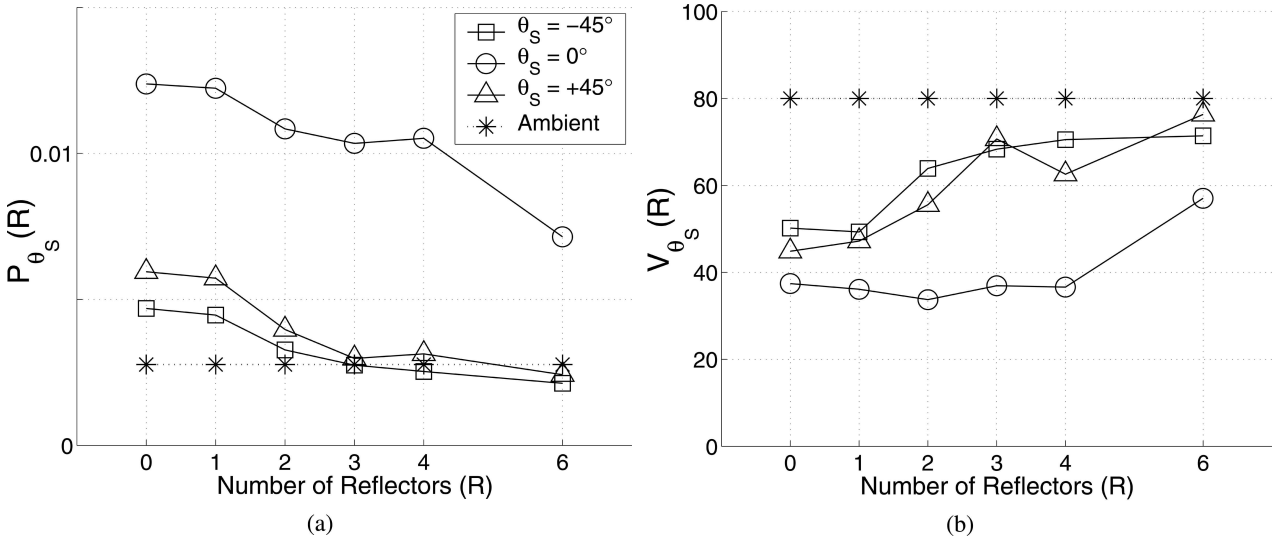


Fig. 11. Objective parameters. (a) RDA. (b) Normalized variance of source peak obtained using MUSIC for speech.

Using an auditory filter bank instead of a Fourier transform could be an alternative for improving the localization of speech signals. This is because in an auditory filter bank covering 0–20 kHz, with the center frequencies of the filters given by the Greenwood function [21], about 82% of the filters span the frequency range of 0–10 kHz while the remaining filters span frequencies from 10 to 20 kHz. Thus the frequencies up to 10 kHz, where the speech energy is high, contribute much more to the localization plot than the frequencies above 10 kHz. Using such a filter bank may improve the localization performance for speech, but it will adversely affect the localization of signals such as high-frequency tones, broad-band chirps, and white noise. It may also boost the phantom source phenomenon.

The reduction in localization performance, as quantized by the objective parameters, was not uniform with the increase in the number of reflectors. The reduction was greatest from one to four reflectors for white noise, but it was concentrated between one and two reflectors for speech. The relative detection amplitude of the source peak dropped very close to the baseline performance in the presence of six reflectors. One of the important differences between one reflector and two reflectors was that the two-reflector condition had a pair of reflectors facing each other, which creates persistent standing waves with a frequency that depends on the distance between the reflectors. There are two and three such pairs under four-reflector and six-reflector conditions, respectively, which gives rise to more complex normal modes or eigentones of standing waves [22]. Higher eigentone density results in a longer reverberation decay. This can be observed in the relatively high jumps in the approximate reverberation times (RT_{60}) from one to two reflectors, from three to four reflectors, and from four to six reflectors. A longer decay causes more temporal smearing of the localization cues, resulting in poorer localization performance.

The appearance of a phantom source at 0° azimuth for speech in the presence of two or more reflecting surfaces was the most interesting and intriguing result of this study. The separation of the curves for $\theta_S = 0°$ from those for the other two values of θ_S is much greater for the speech signal than for the white noise, because the phantom peak phenomenon was dominant for speech. No phantom source was observed for white noise, suggesting that this phenomenon must be dominant at lower frequencies, where the speech energy is concentrated, while leaving the higher frequency signals relatively less corrupted.

This argument was found to be true when the coincidence plots for different conditions were compared (Fig. 12). In the presence of the two reflectors, the source is localized at the correct location above 7 kHz, indicated by the thick vertical strip at −45°, but the coincidence plot is smeared out over the azimuth below 7 kHz. There is also a faint vertical line at 0° from 2 to 7 kHz, which hints at the formation of the phantom source. The phantom source, though seen in the coincidence plots, is not seen in the localization plots for white noise because the coincidence plot is summed over the entire frequency range of 0–22.05 kHz for white noise and the phantom source phenomenon is limited to a relatively small range of frequencies. However, the phantom peak is noticeable in the localization plots for speech because the range of summation of the coincidence plots for speech is limited to 0–10 kHz. The absorption coefficient for plywood is small at low frequencies [23], and the reverberation in the plywood cube has longer decays at lower frequencies, resulting in more temporal smearing at lower frequencies. Temporal smearing alters the localization cues in speech signals, resulting in the appearance of the phantom source that causes a reduction in the localization performance of the actual source at the low frequencies. However, the exact nature of the link between the alteration of localization cues in reverberation and the formation of a phantom source at 0° is not clear from this experiment. The localization plot, and hence the objective parameters, cannot distinguish the phantom source from the real source at 0° and therefore underestimate the degradation of the localization performance when the real source is at 0°. In other words, the objective parameters overestimate the localization performance in reverberation when the source is at 0°, resulting in the $P_{\theta_S}(R)$ curves for $\theta_S = 0°$ to be higher than the curves for other values of θ_S for speech.

The performance of the LE algorithm was compared to that of the MUSIC algorithm. MUSIC localizes the source using a second-order metric (the covariance matrix), whereas the LE algorithm uses a first-order metric (subtraction of the delayed signals). In the LE algorithm the difference of the delay-pair outputs in each frequency bin is a smooth function of the azimuth. However, a localization decision is made in each frequency bin, thereby converting the smooth function into a unit impulse function with the nonzero value at the detected source azimuth. Thus the localization plot for LE is the sum of the unit impulse functions and has sharper peaks than the localization plot for MUSIC, which is the sum of smooth $P(\theta, f)$ contours. Though MUSIC belongs to the class of conventional covariance-based localization algorithms that are known to fail in reverberant environments [3], it is a widely known and utilized algorithm and was therefore used for benchmarking the performance of the LE algorithm. In general the localization performance of LE was better than that for MUSIC in reverberation, both qualitatively (localization plot) and quantitatively (objective parameters). The phantom source peak was stronger and wider for MUSIC than for LE. Thus the LE algorithm

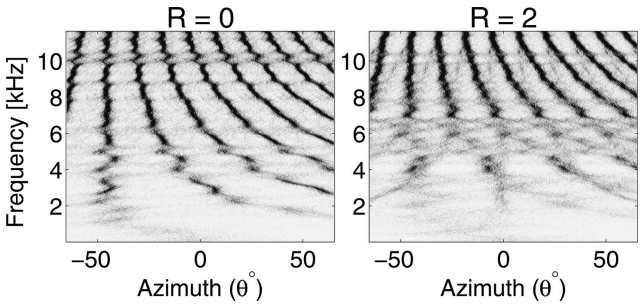


Fig. 12. Average of coincidence plots generated by LE for white-noise source at −45°. Left panel: No-reflector condition ($R = 0$). Right panel: Two-reflector condition ($R = 2$).

could be a better alternative to conventional localization algorithms in reverberant environments.

4 CONCLUSIONS

1) The degradation of the source localization performance is not uniform with an increase in the number of reflectors. The degradation is relatively high when a pair of reflecting surfaces facing each other is added to the environment.

2) The localization is better for white noise than for speech because of the higher bandwidth of white noise. Adjusting the frequency range for the summation of the coincidence plot to match the bandwidth of speech improves the localization performance for the speech signal.

3) The localization of the speech signal is affected more than that of the white noise in reverberant conditions because the temporal smearing of the signal due to reverberation is greater at lower frequencies. Alteration of the localization cues due to reflections and reverberation gives rise to a phantom source at 0° in the localization plot.

4) The localization performance of the LE algorithm is less affected by reflections and reverberation than that of the MUSIC algorithm.

5 ACKNOWLEDGMENT

This study was supported by the National Institutes of Health under grant PHS 1 R21 DC04840-01. The authors are grateful to all members of the Intelligent Hearing Aid Group at the Beckman Institute, UIUC, for their inputs in this research. They wish to thank the Integrated Systems Laboratory, UIUC, for permitting the use of the plywood cube and the recoding equipment.

REFERENCES

[1] M. Goueygou, M. E. Lockwood, M. E. Elledge, R. C. Bilger, D. L. Jones, C. R. Lansing, C. Li, W. D. O'Brien, Jr., and B. C. Wheeler, "The Effect of Reverberation on a New Binaural Noise Cancellation Algorithm for Hearing Aid," *J. Acoust. Soc. Am.,* vol. 106, pp. 2278–2279 (1981).

[2] M. E. Lockwood, D. L. Jones, R. C. Bilger, C. R. Lansing, W. D. O'Brien, Jr., B. C. Wheeler, and A. S. Feng, "Performance of Time- and Frequency-Domain Binaural Beamformers Based on Recorded Signals from Real Rooms," *J. Acoust. Soc. Am.,* vol. 115, pp. 379–391 (2004).

[3] M. Tanaka and Y. Kaneda, "Performance of Sound Direction Estimation Methods under Reverberant Conditions," *J. Acoust. Soc. Jpn.,* vol. 14, pp. 291–292 (1993).

[4] B. Champagne, S. Bédard, and A. Stéphenne, "Performance of Time-Delay Estimation in the Presence of Room Reverberation," *IEEE Trans. Speech Audio Process.,* vol. 4, pp. 148–152 (1996).

[5] T. Gustafsson, B. D. Rao, and M. Trivedi, "Source Localization in Reverberant Environments: Modeling and Statistical Analysis," *IEEE Trans. Speech Audio Process.,* vol. 11, pp. 791–803 (2003).

[6] H. F. Silverman, Y. Yu, J. M. Sachar, and W. R. Patterson III, "Performance of Real-Time Source Location Estimators for a Large-Aperture Microphone Array," *IEEE Trans. Speech Audio Process.,* vol. 13, pp. 593–606 (2005).

[7] B. Blesser, "An Interdisciplinary Synthesis of Reverberation Viewpoints," *J. Audio Eng. Soc.,* vol. 49, pp. 867–903 (2001 Oct.).

[8] B. Rakerd and W. M. Hartmann, "Localization of Sound in Rooms, II: The Effects of a Single Reflecting Surface," *J. Acoust. Soc. Am.,* vol. 78, pp. 524–533 (1985).

[9] W. M. Hartmann, "Localization of Sounds in Rooms," *J. Acoust. Soc. Am.,* vol. 74, pp. 1380–1391 (1983).

[10] C. Liu, B. C. Wheeler, W. D. O'Brien, Jr., R. C. Bilger, C. R. Lansing, and A. S. Feng, "Localization of Multiple Sound Sources with Two Microphones," *J. Acoust. Soc. Am.,* vol. 108, pp. 1888–1905 (2000).

[11] R. Schmidt, "Multiple Emitter Location and Signal Parameter Estimation," *IEEE Trans. Antennas Propogat.,* vol. AP 34, pp. 276–280 (1986).

[12] J. Blauert and W. Cobben, "Some Considerations of Binaural Cross Correlation Analysis," *Acustica,* vol. 39, pp. 96–104 (1978).

[13] E. Jan and J. Flanagan, "Sound Source Localization in Reverberant Environments Using an Outlier Estimation Algorithm," in *Proc. 4th Int. Conf. on Spoken Language (ICSLP 96),* vol. 3 (1996), pp. 1321–1324.

[14] S. Rickard and F. Dietrich, "DOA Estimation of Many w-Disjoint Orthogonal Sources from Two Mixtures Using Duet," in *Proc. 10th IEEE Signal Processing Workshop (SSA P2000),* pp. 311–314.

[15] D. Banks, "Localization and Separation of Simultaneous Voices with Two Microphones," *IEE Proc. I,* vol. 140, pp. 229–234 (1993).

[16] R. Roy and T. Kailath, "ESPRIT-Estimation of Signal Parameters via Rotational Invariance Techniques," *IEEE Trans. Acoust., Speech, Signal Process.,* vol. 37, pp. 984–995 (1989).

[17] M. R. Schroeder, "Integrated-Impulse Method for Measuring Sound Decay without Using Impulses," *J. Acoust. Soc. Am.,* vol. 66, pp. 497–500 (1979).

[18] L. A. Jeffress, "A Place Theory of Sound Localization," *J. Comp. Physiol. Psychol.,* vol. 41, pp. 35–39 (1948).

[19] S. Mohan, M. L. Kramer, B. C. Wheeler, and D. L. Jones, "Localization of Nonstationary Sources Using a Coherence Test," in *Proc. 2003 IEEE Statistical Signal Processing Workshop* (2003), pp. 453–456.

[20] D. R. Perrott, "Role of Signal Onset in Sound Localization," *J. Acoust. Soc. Jpn.,* vol. 45, pp. 436–445 (1969).

[21] D. D. Greenwood, "A Cochlear Frequency-Position Function for Several Species 29 Years Later," *J. Acoust. Soc. Am.,* vol. 87, pp. 2592–2605 (1990).

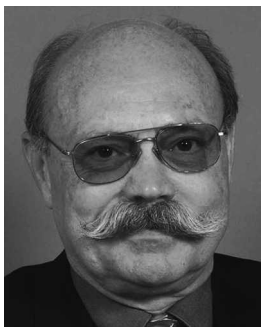
[22] H. Kuttruff, *Room Acoustics,* 2nd ed. (Van Nostrand, London, 1991).

[23] C. W. Kosten, "International Comparison Measurements in the Reverberation Room," *Acustica,* vol. 10, pp. 400–411 (1960).

THE AUTHORS

S. A. Phatak

B. C. Wheeler



W. D. O'Brien, Jr.

A. Feng

Sandeep Phatak was born in Pune, India, in 1979. He received a B.Eng. degree in electronics and telecommunications engineering from the University of Pune in 2000 and an M.S. degree in electrical engineering from the University of Illinois at Urbana-Champaign (UIUC) in 2003. He is currently pursuing a Ph.D. degree in the field of human speech recognition at UIUC. His research interests are acoustics, psychoacoustics, speech perception, and speech recognition.

●

Rama Ratnam was born in Rangoon, Burma, in 1963. He obtained a B.Tech. degree in chemical engineering from the Indian Institute of Technology, New Delhi, in 1985. He moved to the United States in 1991 to pursue doctoral research in sensory neurobiology at the University of Illinois at Urbana-Champaign (UIUC) and received a Ph.D. degree in biophysics and computational biology in 1998. His thesis is on the detection of auditory signals in spatially separated noise by neurons in the frog auditory midbrain.

Starting in 1985, he worked in the Indian industry as a process design engineer. He then assumed the position of staff scientist with the Chemical Engineering Division of the National Chemical Laboratory, Pune, India, where he worked on process control and process plant dynamics. After receiving his Ph.D. degree and until 2001 he was a postdoctoral research associate at UIUC, where he carried out research in electrosensory processing in weakly electric fish. From 2001 to 2004 he was a research scientist with the Intelligent Hearing Aid Project at the Beckman Institute, UIUC, working on the acoustical analysis of reverberation and on sound perception in reveberant spaces. Since 2004 he has been assistant professor of Systems and Computational Neuroscience in the Department of Biology at the University of Texas at San Antonio. His interests are in information processing in the brain, analysis, and characterization of natural signals and natural environments, and development of biomedical devices.

●

Bruce Wheeler received a B.S. degree from the Massachusetts Institute of Technology, Cambridge, MA, and M.S. and Ph.D. degrees from Cornell University, Ithaca, NY, all in electrical engineering.

He has been with the University of Illinois at Urbana-Champaign since 1980 and is currently interim head of the Department of Bioengineering. He is also a professor of Electrical and Computer Engineering at its Beckman Institute. He has served as an associate head of the Electrical and Computer Engineering Department and as chair of the Neuroscience Program.

Dr. Wheeler's research interests lie in the application of electrical engineering methodologies, including signal processing, to biological problems.

●

William D. O'Brien, Jr., was born in Chicago, IL, in 1942. He received B.S., M.S., and Ph.D. degrees in 1966, 1968, and 1970 from the University of Illinois, Urbana-Champaign.

From 1971 to 1975 he worked with the Bureau of Radiological Health (currently the Center for Devices and Radiological Health) of the U.S. Food and Drug Administration. Since 1975 he has been at the University of Illinois, where he is the Donald Biggar Willet Professor of Engineering. He is also professor of Electrical and Computer Engineering and of Bioengineering, College of Engineering; professor of Bioengineering, College of Medicine; professor of Nutritional Sciences, College of Agricultural, Consumer and Environmental Sciences; professor of Speech and Hearing Science, College of Applied Life Studies; research professor in the Beckman Institute for Advanced Science and Technology; and research professor in the Coordinated Science Laboratory. He is the director of the Bioacoustics Research Laboratory. His research interests involve the many areas of acoustic- and ultrasound-tissue interaction, including biological effects and quantitative acoustic imaging, for which he has published 309 papers.

Dr. O'Brien is a fellow of the Institute of Electrical and Electronics Engineers, the Acoustical Society of America, and the American Institute of Ultrasound in Medicine, and a founding fellow of the American Institute of Medical and Biological Engineering. He was recipient of the IEEE Centennial Medal (1984), the AIUM Presidential Recognition Awards (1985 and 1992), the AIUM/WFUMB Pioneer Award (1988), the IEEE Outstanding Student Branch Counselor Award for Region 4 (1989), the AIUM Joseph H. Holmes Basic Science Pioneer Award (1993), and the IEEE Ultrasonics, Ferroelectrics, and Frequency Control Society Distinguished Lecturer (1997–1998). He received the IEEE Ultrasonics, Ferroelectrics, and Frequency Control Society's Achievement Award for 1998 and its Distinguished Service Award for 2003, and the IEEE Millennium Medal in 2000. He has served as cochair of the 1981, 2001, and 2003 IEEE Ultrasonic Symposia, and as general chair of the 1988 IEEE Ultrasonics Symposium. He has served as president (1982–1983) of the IEEE Sonics and Ultrasonics Group (currently the IEEE Ultrasonics, Ferroelectrics, and Frequency Control Society), editor-in-chief (1984–2001) of the *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control,* president (1988–1991) of the American Institute of Ultrasound in Medi-

cine, treasurer (1991–1994) of the World Federation for Ultrasound in Medicine and Biology, and on the Board of Directors (1988–1993) of the American Registry of Diagnostic Medical Sonographers.

●

Albert S. Feng received B.S. and M.S. degrees in electrical engineering from the University of Miami, Coral Gables, in 1968 and 1970, respectively, and a Ph.D. degree in electrical engineering and neurobiology and behavior from Cornell University, Ithaca, NY, in 1975.

Prior to joining the University of Illinois in 1977, he has been with the University of California at San Diego and Washington University, St. Louis, MO. At present he is a professor of Molecular and Integrative Physiology, Bioengineering, Neuroscience, and Biophysics and Computational Biology at the Beckman Institute, University of Illinois, as well as Richard and Margaret Romano Professional Scholar and director, Sensory Neuroscience Training Program at Illinois. He is the author or coauthor of numerous scientific and technical publications as well as several patents.

Dr. Feng is a fellow of the American Association for the Advancement of Science and the Acoustical Society of America and a member of the International Society for Neuroethology.