Research Article

A Method for Source-Microphone Range Estimation in Reverberant Environments Using Arrays of Unknown Geometry

Denis McCarthy and Frank Boland

Department of Electronic and Electrical Engineering, School of Engineering, Trinity College, Dublin, Ireland

Correspondence should be addressed to Denis McCarthy, demccart@tcd.ie

This paper proposes a technique for determining the distance between a sound source and the microphones in an array. The proposed "Range-Finder" algorithm is robust in the presence of reverberation and, in contrast with previously published source-localization techniques, does not require knowledge of the relative positions of the microphones. We discuss the factors affecting the accuracy of our range estimates and present the results of experiments using simulated and real data to demonstrate the efficacy of our approach.

1. INTRODUCTION

Estimating the distance between a source and a receiver has been a central problem in array signal processing since the earliest days of radar and sonar. For indoor applications, using microphone arrays, such estimates could have use in source localization or speaker tracking. In addition, they could inform decisions regarding microphone selection, allowing us to select the microphone(s) nearest the source or farthest from some likely interference. Range estimates could also have use in determining appropriate speech enhancement strategies, such as when deciding whether or not to use a dereverberation algorithm.

Typically, range is determined by measuring the time-of-flight of a transmitted or reflected soundwave and multiplying it by some known propagation speed. In [1] this is achieved by simultaneous transmission of a soundwave and a "time-stamped" radio signal. Provided that the transmitter and receiver are synchronized, the time-of-flight may be easily obtained as the difference in the times of transmission and reception. However, in a majority of cases the sources of interest will not be specifically designed transmitters and so such techniques have limited application.

Given the knowledge of the relative microphone positions, the source-microphone range may easily be obtained from estimates of the relative position of the source—an end to which a variety of solutions have been proposed.

For the sake of clarification, we note that many of the methods, presented in the literature as "source-localization" techniques, are, in fact, solutions to the related but distinct problem of delay-vector estimation, that is, obtaining the relative intersensor time-delay estimates (TDEs). Furthermore, in many cases, the source "location" is defined in terms of a bearing line only. In this paper, we use the term "source localization" to refer to the problem of estimating the position of a source, with respect to some coordinate system.

Much of the previously published work on source localization has focused on the use of TDEs (see [2] and the references therein for a review of time-delay-estimation techniques). In the two-dimensional case, source localization may be considered a practical application of Apollonius' problem of tangent circles [3]. The numerical solution to this problem, as discovered by Viète (see [4] for a description of his solution), may be easily expanded to the three-dimensional case and, given TDEs between a minimum of four microphones (three in the two-dimensional case), a source location may be found. In [5], TDEs are determined for pairs of microphones in a series of four-element, square microphone arrays. From these, source-bearing lines are calculated, with the final source location estimate being

calculated as a weighted average of the closest intersections between bearing-line pairs. In [6, 7] the authors estimate the source location via a least-squares fitting of the TDEs for an ad hoc deployment of sensors.

Relative range estimates may also be obtained from a comparison of received signal power. In [8] the authors combine TDEs and relative signal power measurements to determine the location of a source in the extreme near-field of a two-element array. In [9] the authors present a method for source localization that utilizes received signal energy only. Whilst this technique is reported as returning consistently accurate source-bearing estimates, in the presence of reverberation range estimation is shown to be subject to a significant bias.

The use of techniques employing power measurements is commonly restricted to nonreverberant acoustic environments, or to situations where the effects of reverberation are negligible. This is due to the difficulty inherent in modelling and/or mitigating against the presence of reverberation and its consequent adverse effects. Techniques that use TDEs only are preferred when reverberation is present although, as we have noted, these require knowledge of the relative microphone positions.

However, for many practical applications, microphone locations will be unknown or unreliable. Yet, the question of how to estimate the range between a sound source and a microphone, in the presence of reverberation and with the relative positions of the microphones unknown, remains largely unaddressed. We propose a solution to this problem. Our method combines relative power measurements with TDEs in such a way as to mitigate against the adverse effects of reverberation and obtain absolute source-microphone range estimates for microphones at unknown locations.

In the following section, we will briefly discuss the relevant characteristics of sound propagation in rooms. In Section 3, we derive a well-known but naïve range estimator as well as the proposed "Range-Finder" algorithm. In Section 4, we address the factors affecting range-estimate distribution. In Section 5, we present the results of a series of simulations and experiments designed to test the performance of our algorithm. We discuss the potential uses of the Range-Finder algorithm and suggest future work in Section 6.

2. SOUND PROPAGATION IN ROOMS

In a noiseless but reverberant environment, the signal received at some microphone, m_0, will consist of a direct-path component and multiple reflected components jointly referred to as reverberation. The input to the microphone may be modelled as the convolution of the source-microphone impulse response, $h_0(t)$, and the source signal, $s(t)$:

$$x_0(t) = \int_0^t s(p) h_0(t-p) dp. \tag{1}$$

In the frequency domain,

$$X_0(\omega) = S(\omega) H_0(\omega) = S(\omega) \big(H_{dp_0}(\omega) + H_{mp_0}(\omega) \big), \tag{2}$$

where H_{dp_0} is the component of H_0 due to direct-path (nonreflected) propagation and H_{mp_0} is the reverberant component due to multipath reflections. The received signal power spectrum may be calculated as follows. Note that, for clarity, we omit the dependence on ω in the sequel

$$
\begin{aligned}
|X_0|^2 &= |S|^2 |H_0|^2 \\
&= |S|^2 \big(|H_{dp_0}|^2 + |H_{mp_0}|^2 + 2\mathrm{Re}\{H_{dp_0} H_{mp_0}^*\} \big),
\end{aligned} \tag{3}
$$

where $\mathrm{Re}\{\ \}$ denotes the real component and $*$ denotes the complex conjugate.

In air, for an omnidirectional source and receiver, the power of the direct-path component of sound, received at m_0, is inversely proportional to the squared source-microphone range, that is, the squared distance between the source and the microphone,

$$|H_{dp_0}|^2 \propto \frac{1}{r_0^2}, \tag{4}$$

where $r_0 = |\vec{s} - \vec{m}_0|$ and $\vec{s}$ and $\vec{m}_0$ denote the Cartesian coordinates of the source and m_0, respectively. The direct-path component therefore decays at a rate of 6 dB per doubling of the source-microphone range. This model does not address effects due to variations of air pressure or temperature, however, in a room environment it is reasonable to assume a homogenous medium. From (4), we may derive an expression for the power of the direct-path component of the sound received at some microphone m_a:

$$|H_{dp_a}|^2 = |H_{dp_0}|^2 \left[\left(\frac{r_0}{r_a} \right)^2 \right]. \tag{5}$$

The reverberant component of an impulse response will be dependent upon factors such as the dimensions and surface absorption characteristics of the room. These vary widely from room to room and so we cannot know $|H_{mp_0}|^2$ a priori.

Typically, the degree to which a room is reverberant is described with reference to a metric known as the reverberation time (RT_{60}). The RT_{60} is defined as the average time taken for the reverberant sound energy to decay by 60 dB. Although useful for conveying a general idea of how reverberant a room may be, specifying the RT_{60} gives no idea of how reverberant a recorded sound will be. Consider, for example, a recording made in a room at a distance of 1 m from a sound source. This recording will be perceived as being less reverberant than one made in the same room at 5 m from the source. This is because the direct path component decays as we get farther from the source, despite the RT_{60} being the same in each instance.

A more effective way of describing the degree of reverberation that obtains on a recording is to specify the direct-to-reverberant ratio (DRR), that is, the ratio of the received sound energy due to the direct-path component and multipath reflections. For a given bandwidth, the DRR at the
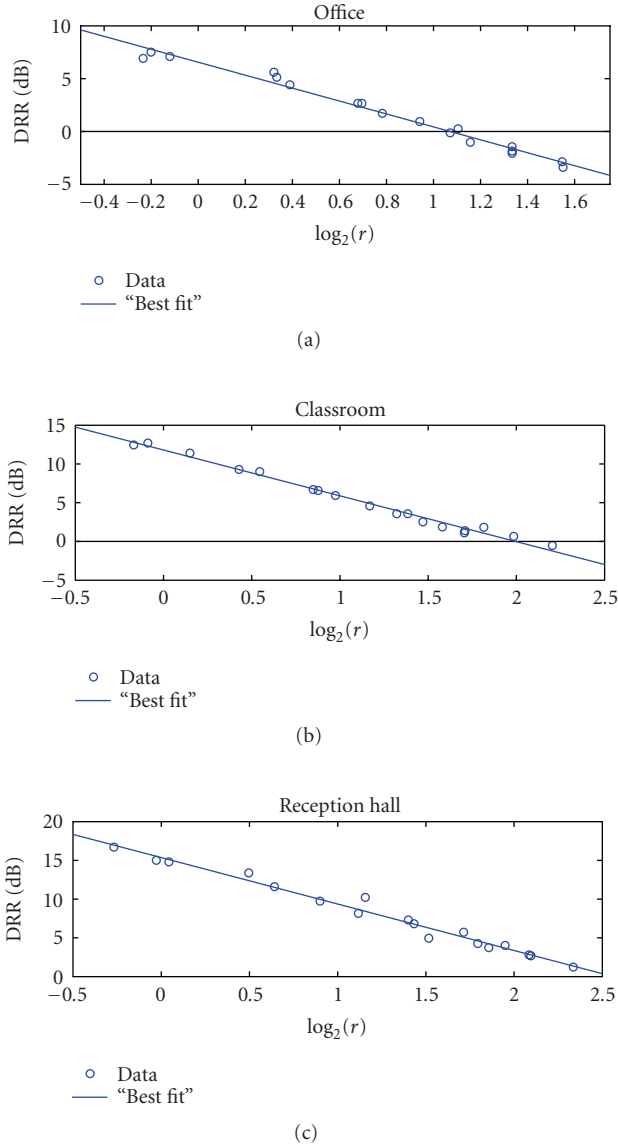
(a)

(b)

(c)

FIGURE 1: Direct-to-reverberant ratios versus $\log_2(r)$, where r is the source-microphone range. Results shown are for an office, classroom, and reception hall.

microphone, m_0, may be defined as follows:

$$\text{DRR}_0 = \frac{\int |H_{dp_0}|^2 d\omega}{\int |H_{mp_0}|^2 d\omega}. \tag{6}$$

An investigation of DRRs in real rooms proves informative. Figure 1 shows a plot of DRRs, found at a variety of locations in an office, classroom, and reception hall. The DRRs are plotted with respect to $\log_2(r)$. The reverberation times were determined experimentally using the transient decay method [10] and were found to be 0.6, 0.5, and 1.1 seconds, respectively. The DRR estimates were obtained as follows. Recordings were made at varying locations in each room and at varying distances relative to a single source—in this case a loudspeaker. The sampling rate was 48 kHz. In each instance,

the microphone was placed directly in front of the loudspeaker so as to avoid complications due to the directivity of the source. The loudspeaker produced a maximum-length-sequence (MLS) of approximate duration 5.5 seconds, also at a sampling rate of 48 kHz. These recordings were then cross-correlated with the "clean" MLS to obtain an impulse response estimate, from which a DRR estimate was calculated.

Figure 1 also shows "best-fit" linear approximations of the data. The slopes of these fits are -6.12, -5.99, and -5.915 decibels per doubling of range for the office, classroom, and hall, respectively. Given that we can expect $|H_{dp_0}|^2$ to decay at a rate of 6 dB per doubling of the source-microphone range, these results suggest that, in a given room, $E\{\int |H_{mp_0}|^2 d\omega\}$ (where $E\{ \}$ is the expectation operator) is a constant that is independent of the source-microphone range.

We define the following:

$$\begin{aligned} F_{a,b} = \int & |H_{mp_a}|^2 - |H_{mp_b}|^2 \\ & + 2\text{Re}\{H_{dp_a}H_{mp_a}^* - H_{dp_b}H_{mp_b}^*\}d\omega, \end{aligned} \tag{7}$$

where the a and b subscripts denote the impulse response components corresponding to the microphones m_a and m_b, respectively. Consider the cross-terms in (7). Direct path propagation applies a delay and scaling to a sound wave. Therefore, for any source-microphone impulse response, H_{dp} is a scaled exponential. Similarly, H_{mp} may be considered to be the sum of scaled exponentials corresponding to multiple reflected sound waves. As such, $H_{dp}H_{mp}^*$ is also the sum of multiple scaled exponentials. Therefore, invoking the central limit theorem, we will assume $\int \text{Re}\{H_{dp_a}H_{mp_a}^*\}d\omega$ and $\int \text{Re}\{H_{dp_b}H_{mp_b}^*\}d\omega$ to be zero-mean normally distributed random variables. Following from our previous results, we also assume $\int |H_{mp_a}|^2 d\omega$ and $\int |H_{mp_b}|^2 d\omega$ to be random variables distributed about the same mean. Therefore, invoking the central limit theorem once again, we may consider $F_{a,b}$ to be a zero-mean normally distributed random variable.

Note that if H_{dp} and H_{mp} are nonzero at $\omega = 0$, $\int \text{Re}\{H_{dp}H_{mp}^*\}d\omega$ will exhibit a positive bias. We may ignore this, however, as the frequency responses of real microphones will not have a nonzero component at $\omega = 0$.

As an aside, we note that a brief inspection of the results in Figure 1 reveals that although it had the greatest RT_{60}, the reception hall was not the most reverberant of the rooms in which we took measurements. This further illustrates the inadequacy inherent in characterizing the degree of reverberation in a room by specifying its RT_{60} alone. Our results do, however, suggest an alternative metric. The intercept of best-fit line with the y-axis defines the spatially averaged "DRR-at-1 m" and we will use this metric to describe acoustic conditions in the sequel.

3. RANGE ESTIMATION

In this section, we derive two range estimation algorithms: firstly a well-known but naïve range estimator that assumes an anechoic environment, and secondly the proposed algorithm, which we refer to as the Range-Finder and which is robust against the effects of reverberation.

3.1. A naïve range estimator

When τ_a is the relative intersensor time-delay between m_a and m_0,

$$r_a - r_0 = c\tau_a, \tag{8}$$

where c is the speed of sound in air. Using any one of a variety of time-delay estimation techniques, we may obtain an estimate of the relative intersensor time-delay, $\tilde{\tau}_a$. In noiseless, anechoic environments the direct-path sound accounts for all acoustic energy received by the microphones and so, by substituting (3) and (8) into (5) and performing algebraic manipulation, we obtain a simple and well-known estimator of r_0:

$$r_0 = \frac{c\tilde{\tau}_a\sqrt{|H_a|^2/|H_0|^2}}{1 - \sqrt{|H_a|^2/|H_0|^2}}. \tag{9}$$

Unfortunately, in nonideal acoustic environments, the presence of interfering reverberation can severely distort this estimate, making the above range estimator unsuitable for use in practical environments. Where more than two microphones are available, the most accurate range estimate will be obtained by using only those two microphones closest to the source. These may be presumed to have the highest DRRs. The outputs of the remaining microphones will contain proportionally greater levels of reverberation and will, therefore, lead to greater distortion in the range estimates.

3.2. The Range-Finder algorithm

From (5) and (8),

$$
\begin{aligned}
&(|H_{dp_a}|^2 - |H_{dp_b}|^2) \\
&= |H_{dp_o}|^2\left[\left(\frac{r_0}{r_0 + c\tau_a}\right)^2 - \left(\frac{r_0}{r_0 + c\tau_b}\right)^2\right].
\end{aligned} \tag{10}
$$

The term in the square brackets is a function of r_0, τ_a, and τ_b which we denote as $G_{a,b}(r_0, \tau_a, \tau_b)$:

$$G_{a,b}(r_0, \tau_a, \tau_b) = \left(\frac{r_0}{r_0 + c\tau_a}\right)^2 - \left(\frac{r_0}{r_0 + c\tau_b}\right)^2. \tag{11}$$

Integrating (3) across the full bandwidth of the signal, we obtain P_0—the total received signal power at m_0:

$$P_0 = \int |S|^2\left(|H_{dp_0}|^2 + |H_{mp_0}|^2 + 2\mathrm{Re}\{H_{dp_0}H_{mp_0}^*\}\right)d\omega. \tag{12}$$

We define $\Lambda_{a,b}$ as being the difference between the total received signal power at m_a and m_b:

$$\Lambda_{a,b} = P_a - P_b. \tag{13}$$

Let us assume, for the moment, that $|S|^2$ is a constant with respect to frequency (we will return to this assumption later). Substituting (12) into (13) and performing algebraic manipulation yields

$$\Lambda_{a,b} = |S|^2\left[kG_{a,b}(r_0, \tau_a, \tau_b) + F_{a,b}\right], \tag{14}$$

where $k = \int |H_{dp_0}|^2 d\omega$. From (14), we see that the difference between the signal power received at two microphones is proportional to the sum of a scaled, deterministic function, $G_{a,b}(r_0, \tau_a, \tau_b)$, and a zero-mean and normally distributed random variable, $F_{a,b}$. We define the following vectors, noting that we have omitted the arguments of the $G_{a,b}(r_0, \tau_a, \tau_b)$ terms for clarity:

$$
\begin{aligned}
\mathbf{G} &= [G_{0,1}, G_{0,2}, \ldots, G_{1,2}, G_{1,3}, \ldots G_{M-2,M-1}]^T, \\
\mathbf{F} &= [F_{0,1}, F_{0,2}, \ldots, F_{1,2}, F_{1,3}, \ldots F_{M-2,M-1}]^T, \\
\mathbf{\Lambda} &= [\Lambda_{0,1}, \Lambda_{0,2}, \ldots, \Lambda_{1,2}, \Lambda_{1,3}, \ldots \Lambda_{M-2,M-1}]^T \\
&= |S|^2[k\mathbf{G} + \mathbf{F}].
\end{aligned} \tag{15}
$$

Once again, using any of the many well-known techniques for delay-vector estimation, we may obtain the time-delay estimates $\tilde{\tau}_a$ and $\tilde{\tau}_b$. We then define $\tilde{G}_{a,b}(r_0)$ and the corresponding vector $\tilde{\mathbf{G}}(r_0)$ from

$$\tilde{G}_{a,b}(r_0) = G_{a,b}(r_0, \tilde{\tau}_a, \tilde{\tau}_b). \tag{16}$$

Following from the Cauchy-Schwartz inequality, the optimal range estimate, $\tilde{r}_0$, is obtained by a matched-filtering of the power-difference vector, $\mathbf{\Lambda}$, with $\tilde{\mathbf{G}}(r_0)/|\tilde{\mathbf{G}}(r_0)|$:

$$\tilde{r}_0 = \arg\max_{r_0}\left[\frac{1}{|\tilde{\mathbf{G}}(r_0)|}\tilde{\mathbf{G}}(r_0)^T\mathbf{\Lambda}\right]. \tag{17}$$

Following from this estimate, we may easily obtain estimates of the remaining source-microphone ranges, $\{r_1, r_2, \ldots, r_{M-1}\}$, by inserting $\tilde{r}_0$ and the TDEs used to calculate $\tilde{\mathbf{G}}(r_0)$ into (8).

Previously, we assumed $|S(\omega)|^2$ to be a constant with respect to frequency. In many cases, including that of human speech, this is unrealistic. In reality, speech is both a lowpass and often harmonic signal. This poses particular problems. We have assumed $F_{a,b}$ to be a zero-mean, normal random variable. The analysis and experimental evidence underpinning this assumption are for broadband signals and we cannot reasonably expect it to hold for cases, such as speech, where the bulk of the energy is concentrated at low frequencies.

This problem was overcome as follows. The microphone outputs are split into individual, nonoverlapping subbands. The bandwidth of these subbands are chosen such that they are narrow enough that $|S(\omega)|^2$ is roughly constant within the subband whilst also being wide enough that there is always a direct-path speech component present. $\mathbf{\Lambda}$ is then calculated for each subband. Each $\mathbf{\Lambda}$ is normalized and, from these, an average power-difference vector, $\overline{\mathbf{\Lambda}}$, is found across all the subbands. The range estimate is found, as in (17) by a matched filtering of $\overline{\mathbf{\Lambda}}$ with $\tilde{\mathbf{G}}(r_0)/|\tilde{\mathbf{G}}(r_0)|$.

4. ESTIMATE DISTRIBUTION AND ACCURACY

Given multiple estimates for range, we might expect that, as the number of estimates increases, their mean will approach the true range. As we will see in the following section, this

is not necessarily the case. We will also show how the accuracy of a range estimate is dependant upon the actual source-microphone ranges. We restrict our analysis to the situation where we have three microphones only—the minimum number required to implement the Range-Finder. We do this both for the sake of simplicity and to allow us to employ an alternative formulation of the Range-Finder algorithm. This alternative formulation more clearly illustrates how the distribution of range estimates is related to the distribution of the ratio of normal random variables, a well-understood, albeit nontrivial, distribution that has received extensive study in the literature.

4.1. An alternative formulation of the Range-Finder

The range estimate, $\tilde{r}_o$, is that which maximizes the expression in (17). For two vectors with given norms, the dot product of the vectors is a maximum when they are proportional. Therefore, we may write $\tilde{\mathbf{G}}(\tilde{r}_0) \propto \mathbf{\Lambda}$. For the three-microphone case, this implies

$$[\tilde{G}_{0,1}(\tilde{r}_0), \tilde{G}_{0,2}(\tilde{r}_0)] \propto [\Lambda_{0,1}, \Lambda_{0,2}]. \tag{18}$$

Using an equivalent expression, we define $Q_{0,1,2}$:

$$\frac{\tilde{G}_{0,1}(\tilde{r}_0)}{\tilde{G}_{0,2}(\tilde{r}_0)} = \frac{\Lambda_{0,1}}{\Lambda_{0,2}} = Q_{0,1,2}, \tag{19}$$

and from this, we obtain an alternative formulation for the Range-Finder:

$$\tilde{r}_0 = \arg\min_{r_0} \left[\left| Q_{0,1,2} - \frac{\tilde{G}_{0,1}(r_0)}{\tilde{G}_{0,2}(r_0)} \right| \right]. \tag{20}$$

For 3 microphones there are, of course, 5 further permutations of Q ($Q_{0,2,1}$, $Q_{1,2,0}$, etc.). However, all may be shown to yield identical range estimates and so we will consider only $Q_{0,1,2}$. Furthermore, to simplify our analysis, we will assume that $0 \leq \tau_1 \leq \tau_2$. We note that this relationship is for simplicity only and is not an absolute requirement. Rather, it is merely a result of the arbitrary way in which we assign labels to the microphones. Once again, omitting the arguments of the $G_{a,b}(r_0, \tau_a, \tau_b)$ terms for clarity:

$$Q_{0,1,2} = \frac{G_{0,1} + (F_{0,1})/k}{G_{0,2} + (F_{0,2})/k}. \tag{21}$$

From (21), we see that $Q_{0,1,2}$ is the ratio of normally distributed and correlated random variables, with unknown variances and means of $G_{0,1}$ and $G_{0,2}$, respectively. Such a ratio is itself a Cauchy distributed random variable.

4.2. Cauchy distribution

In [11] it is shown that, following a translation and a change of scale, $Q_{0,1,2}$ has the same distribution as the ratio of two uncorrelated normal random variables of unity variance, $N(\alpha, 1)/N(\beta, 1)$. The real constants α and β may be calculated as follows:

$$\alpha = \pm\frac{G_{0,1}/\sigma_{0,1} - \rho G_{0,2}/\sigma_{0,2}}{\sqrt{1 - \rho^2}}, \qquad \beta = \frac{G_{0,2}}{\sigma_{0,2}}, \tag{22}$$

where $\sigma_{a,b}$ is the standard deviation of $(F_{a,b})/k$, ρ is the correlation between $\Lambda_{0,1}$ and $\Lambda_{0,2}$ (which may be shown to be 0.5), and the sign of α is chosen to be the same as that of β. For the sake of simplicity and to avoid unwieldy equations, the following discussion will be with reference to the simplified standard form $N(\alpha, 1)/N(\beta, 1)$. From [12], the probability density function (PDF), $p(t)$, of $N(\alpha, 1)/N(\beta, 1)$ may be given as shown below:

$$p(t)$$
$$= \frac{\exp\{-0.5(\alpha^2 + \beta^2)\}}{\pi(1 + t^2)} \left[1 + q\exp\{0.5q^2\} \int_0^q \exp\{-0.5x^2\}dx\right],$$
$$q = \frac{\beta + \alpha t}{\sqrt{1 + t^2}}. \tag{23}$$

Figure 2 shows the PDFs for varying values of α and β. A very wide variety of distribution shapes are possible and the ones shown are chosen for specific illustrative purposes. For a more complete selection of graphs please see [12]. Shown also is α/β (dashed line). In Figure 2, the distributions are not symmetric about α/β. In addition and contrary to what we might expect, the "mean" of $N(\alpha, 1)/N(\beta, 1)$ is not α/β. In fact, strictly speaking, the mean and variance of $N(\alpha, 1)/N(\beta, 1)$ do not exist. This is because $N(\alpha, 1)/N(\beta, 1)$ is undefined when the denominator equals zero.

In practice, we may calculate a pseudomean and pseudovariance by considering only those estimates that fall within certain bounds. A natural bound would be that value of $Q_{0,1,2}$ corresponding to a range estimate of zero meters (negative range estimates cannot be correct). In setting such bounds, however, we should be mindful that the consequent truncation of the PDF may introduce a bias into the pseudomean.

In general, when defined within sufficiently wide bounds, the pseudomean tends towards α/β for $|\alpha|, |\beta| \gg 1$, as occurs when $G_{0,b} \gg \sigma_{0,b}$. Furthermore, under these conditions, $Q_{0,1,2}$ tends to have quite a narrow distribution (see Figure 2(c)). Unfortunately, the converse is also the case. In general, without knowing $\sigma_{0,1}$ or $\sigma_{0,2}$, we cannot calculate/estimate the distribution of $Q_{0,1,2}$ and, hence, cannot quantify the bias that any given bounds may introduce. We can, however, identify certain situations in which such a bias is likely to be very large. Consider the case where $r_0 \gg c\tau_b$, that is, when the array is remote from the source. From inspection of (11), we see that under these conditions, $G_{0,b} \to 0$. As a result, $Q_{0,1,2}$ is widely distributed, causing our range estimates to exhibit a large variance and, depending upon the bounds used, the mean of the range estimates to be subject to a potentially large bias.

4.3. The effect of array geometry

The actual source-microphone ranges determine the values of r_0, τ_1 and τ_2. We have seen how these parameters can affect the distribution of $Q_{0,1,2}$ and bias its pseudomean away from $G_{0,1}/G_{0,2}$. In this respect, therefore, the accuracy with which we may estimate range is determined by the array geometry. Array geometry also determines the extent to which a
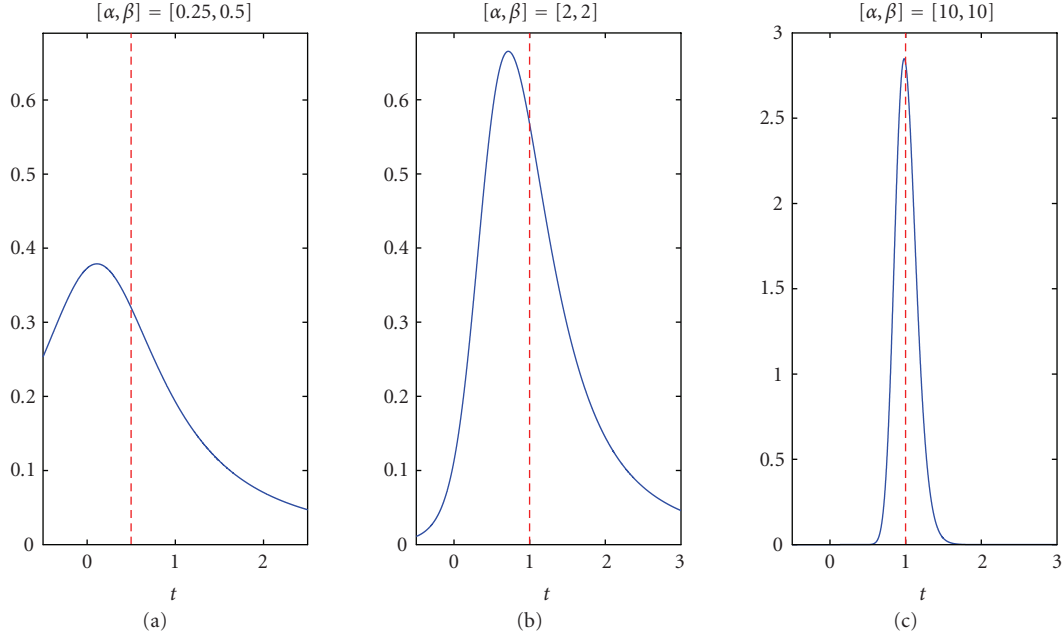
FIGURE 2: Portions of the PDFs of $N(\alpha, 1)/N(\beta, 1)$, also shown is α/β (dashed line).

bias/error in $Q_{0,1,2}$ translates into an error in the corresponding range estimate. To investigate this second effect of array geometry, we examine how a fixed bias, ξ, translates into an error in the range estimate.

Consider an estimate, $\tilde{r}_0$, of the true range, r_0, and let us assume that this estimate contains some error, ϵ_0:

$$\frac{\widetilde{G}_{0,1}(\tilde{r}_0)}{\widetilde{G}_{0,2}(\tilde{r}_0)} = Q_{0,1,2} = \frac{G_{0,1}}{G_{0,2}} + \xi. \tag{24}$$

As an illustrative example, we plot $G_{0,1}/G_{0,2}$ against r_0 for $[c\tau_1, c\tau_2] = [1\,\mathrm{m}, 5\,\mathrm{m}]$ in Figure 3. Outside of a small region around $r_0 = 0$, as r_0 increases the slope of the graph reduces and ϵ_0 becomes larger.

Figure 4, showing $|(d/dr_0)(G_{0,1}/G_{0,2})|$ with respect to $c\tau_1/r_0$ and $c\tau_2/r_0$, provides a more complete description of how array geometry affects estimate accuracy. Note that the region where $c\tau_2/r_0 < 1$ is not shown as in this region $|(d/dr_0)(G_{0,1}/G_{0,2})| \to \infty$, obscuring the remaining detail in the graph. However, it is the region where $(r_0 + c\tau_1)/r_0 \approx (r_0 + c\tau_2)/r_0$ that is of particular interest. Here, $|(d/dr_0)(G_{0,1}/G_{0,2})|$ approaches zero leading to a very large ϵ_0. In the extreme case, where $\tau_1 = \tau_2$, no range estimate may be found as $G_{0,1}/G_{0,2}$ will be unity for all values of r_0. Similarly, no range estimate may be found if τ_1 or τ_2 equals zero, as $G_{0,1}/G_{0,2}$ will be zero or undefined, respectively, for all values of r_0.

The analysis in this section has been limited to the three microphone case. However, the results of our analysis have implications for implementations of the Range-Finder using any number of microphones. To obtain accurate range estimates, we require access to a minimum of three microphones for which no two are equidistant (or approximately equidistant) from the sound source. Furthermore, we will
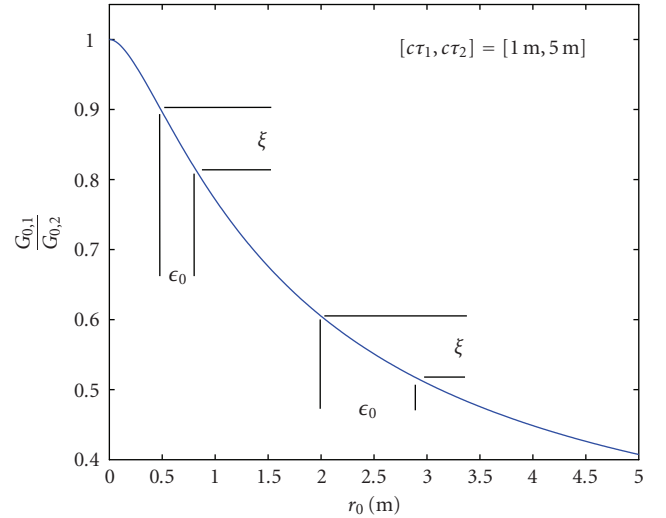

FIGURE 3: $G_{0,1}/G_{0,2}$ versus r_0 for $[c\tau_1, c\tau_2] = [1\,\mathrm{m}, 5\,\mathrm{m}]$. Range estimate error increases with r_0.

not achieve accurate range estimation when $r_0 \gg c\tau_1, c\tau_2$. Under such conditions we may expect $Q_{0,1,2}$ to exhibit a wide distribution and significant bias. This bias/error will then translate into a large error in the range estimate due to $(r_0 + c\tau_1)/r_0 \approx (r_0 + c\tau_2)/r_0$.

We should not, therefore, apply the Range-Finder algorithm in what might be considered the classical microphone array scenario, that of closely spaced microphones and a distant, "farfield" source. Rather, successful implementation would require microphones to positioned in such a way that they are unlikely to be equidistant from the source and, ideally, we will have access to at least 3 microphones for
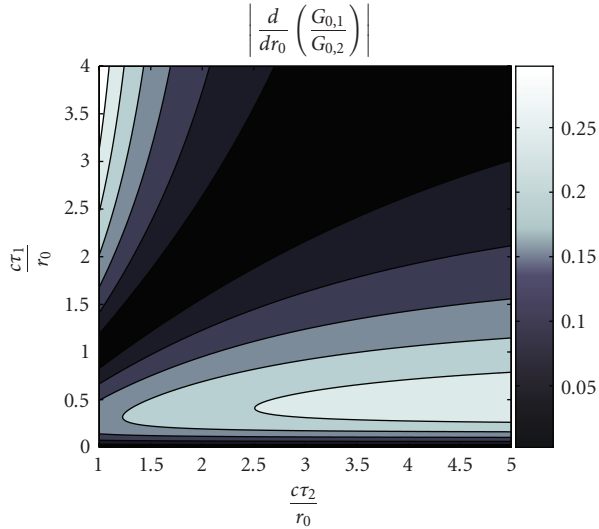
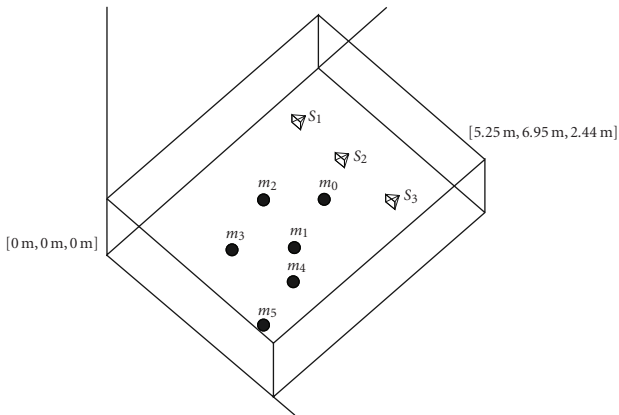FIGURE 4: $|(d/dr_0)(G_{0,1}/G_{0,2})|$ with respect to $c\tau_1/r_0$ and $c\tau_2/r_0$.

TABLE 1: The coordinates of the microphone and source locations for the simulated room. Coordinates are in meters.

(m)	m_0	m_1	m_2	m_3	m_4	m_5	S_1	S_2	S_3
x	3	3	2	2	4	4	1	2.5	4
y	4	3	3	2	2	1	5.5	5.5	5.5
z	2	1	2	1	2	1	1	1	1

path component. Statistical, random reverberant tails were used for the remaining reflections. Two "source signals"— a maximum-length sequence (MLS) of 5.5 seconds in duration and concatenated voice samples of approximately 13 seconds total duration, both bandlimited to avoid aliasing— were convolved with each impulse response to obtain the simulated "recordings." The TDEs were calculated geometrically, using the source and microphone coordinates and a known speed of sound.

The recordings were split into segments of 8192 samples and windowed using a Hamming window. The segment overlap was 50%. In the case of the speech recordings, the signals were separated into eight nonoverlapping subbands with bandwidth 10/16 kHz and $\overline{\Lambda}$ was determined as described in Section 3. For each segment, the Range-Finder algorithm (original formulation (17)) was then used to estimate the distance between the sources and each of the microphones. Negative range estimates and estimates greater than 5 m were ignored—having been determined that wider boundaries did not increase the accuracy of the range estimates.

To investigate the effect of reverberation, the DRR at 1 m of the simulated room was varied by applying an appropriate scaling to the direct-path components of the simulated impulse responses. Range estimates were then obtained as previously described. The results for each source are shown in Figures 6 and 7. The mean of the range estimates, $\pm$ one standard deviation, is shown with respect to the DRR at 1 m. The results shown relate to the estimates of r_0 only. Estimates of the remaining ranges (r_1 to r_5) are omitted because, as is apparent from (8), these will exhibit an identical bias and distribution to those corresponding to r_0. Note that m_0 is the closest microphone to each source. The estimates of r_0 will, therefore, exhibit the greatest percentage error.

The means of the results obtained using the voice recordings are slightly more accurate than those found using the MLS recordings, albeit with a significantly greater variance. Each set of graphs shows that the range estimates are subject to a negative bias that reduces as the reverberation levels decrease. In Section 4.2, we discussed the factors that may explain the presence of a bias in the range estimates. While it is not necessarily the case that any such bias should be negative, from inspection of the PDFs in Figure 2 we see that the density below the mean tends to be greater than that above. Therefore, we may speculate that, for a finite number of estimates, any bias present would tend to be negative, although the precise nature of such a bias is ultimately determined by the reverberation levels present and the array geometry and estimate bounds used.

In Figure 8, the performance of the Range-Finder algorithm is compared to that of the naïve range estimator
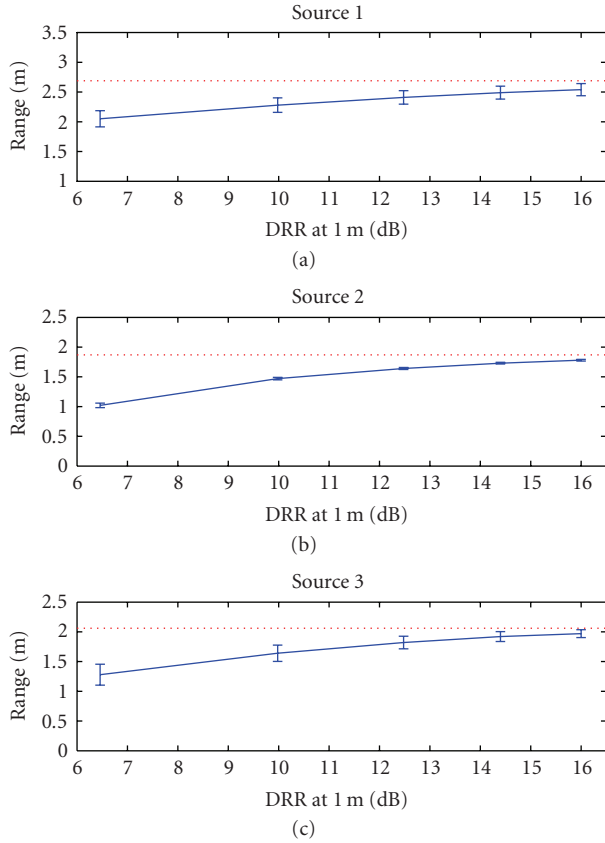

FIGURE 5: A diagram of the simulated room and setup. For precise coordinates of the microphones and loudspeakers, see Table 1.

which $r_0 \ll c\tau_1 \ll c\tau_2$. We will discuss this further and consider the potential applications of the Range-Finder algorithm in Section 6.

5. SIMULATIONS AND EXPERIMENTS

5.1. Simulations

A series of simulations were performed to examine the performance of the Range-Finder algorithm and compare it to that of the naïve range estimator under varying reverberant conditions. Our simulated environment, Figure 5, was a simple rectangular room of dimensions [5.25 m, 6.95 m, 2.44 m] and uniform surface absorption coefficient of 0.3. In this room, we simulated three omnidirectional sources and six omnidirectional microphones (see Table 1 for coordinates). The sampling frequency used was 10 kHz. The source-microphone impulse responses were generated using an acoustic modeling software package [13]. A ray tracing algorithm was used to determine first 20 milliseconds of the impulse response after and including the arrival of the direct-

Figure 6: Mean range estimates ± standard deviation for source producing an MLS.

Figure 7: Mean range estimates ± standard deviation for a voice source.

derived in Section 3. The estimates made using the naïve range estimator were found using the two microphones closest to the source so as to achieve the best possible results. The results shown are for Source 2 but are illustrative of the results obtained for the other sources. In both the voice and MLS cases, the Range-Finder algorithm outperforms the naïve range estimator.

5.2. Experiments

A series of recordings were made to test the Range-Finder under real conditions. The room used was the office, which was chosen for being a highly reverberant environment that would best highlight the superior performance of the Range-Finder over the naïve range estimator. Six microphones were positioned at distances of between 0.8 m and 3 m from a loudspeaker, at intervals of roughly 0.5 m. The loudspeaker and microphones were arranged so as to be approximately colinear, so as to avoid errors due to the directionality of the source. Voice and MLS signals were produced by the loudspeaker. The microphone outputs were recorded before being bandlimited and downsampled to a sampling rate of 10 kHz. These recordings were then split into segments of 8192 samples and windowed using a Hamming window. The segment overlap was 50%. The TDEs were found using a PHAT-GCC [14] and range estimates were obtained for each segment.
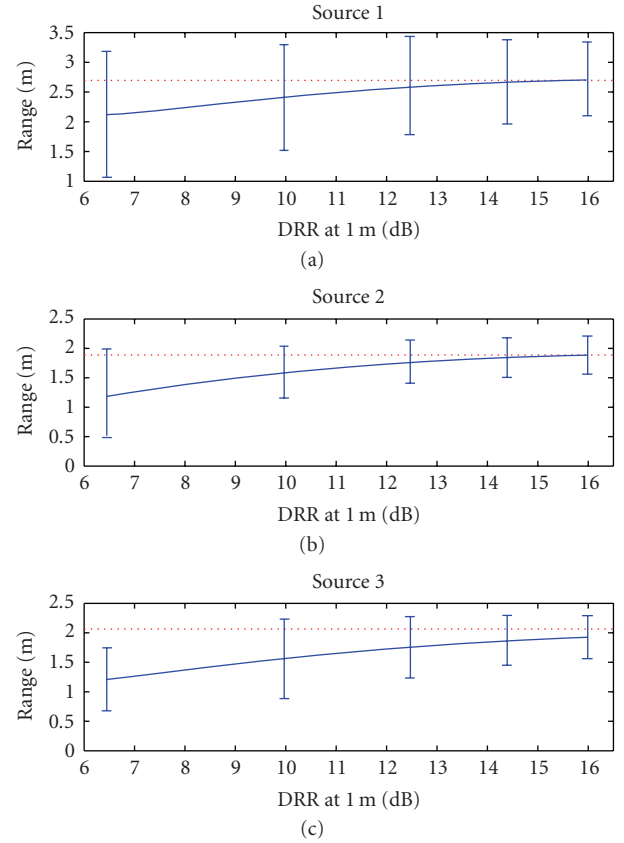
This procedure was repeated for each of three setups in which the loudspeaker and microphones were arranged colinearly along the length and each diagonal of the office, respectively.

The results are shown in Figure 9 and, as with the simulations, clearly show the superior performance of the Range-Finder method. As before, the variances of the results found using voice recordings are greater than those found using MLS recordings, however, there is no noticeable trend with respect to the bias in the mean of the estimates.

6. DISCUSSION

We have proposed a method for estimating source-microphone ranges that is robust against the effects of reverberation. We have discussed the factors affecting the distribution and accuracy of the range estimates obtained by our method and have presented simulated and real experimental results demonstrating its efficacy.

In contrast with source-localization techniques, our method requires no information regarding microphone locations in order to return a range estimate. However, our analysis in Section 4 revealed that the accuracy of the range estimates so obtained is, nonetheless, affected by the relative positioning of the microphones and the sound source. In particular, it was found that we can expect the range estimates to be inaccurate if $r_0 \gg c\tau_1, c\tau_2, \ldots, c\tau_{M-1}$. Rather, successful
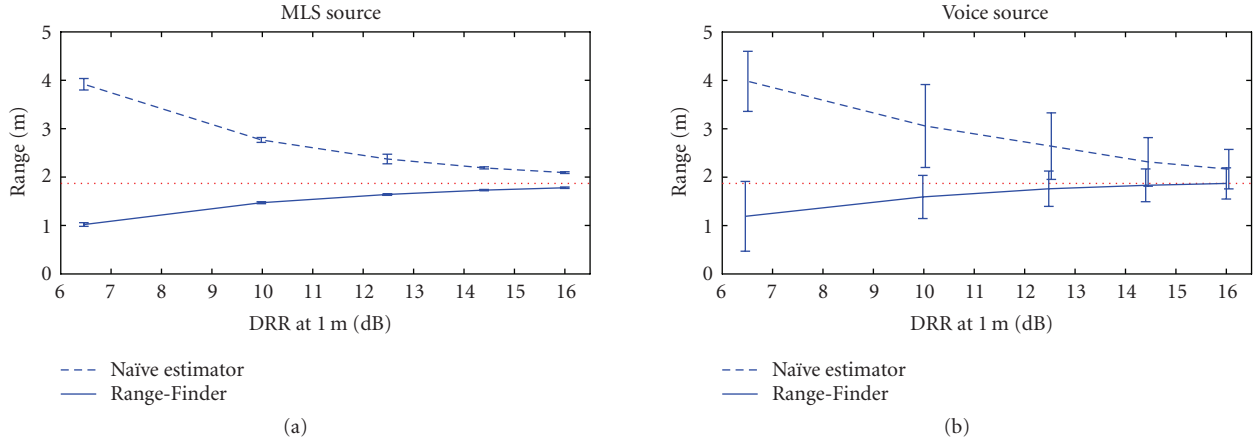
FIGURE 8: A comparison of mean range estimates ($\pm$ one standard deviation) for the naïve range estimator and the Range-Finder algorithm.
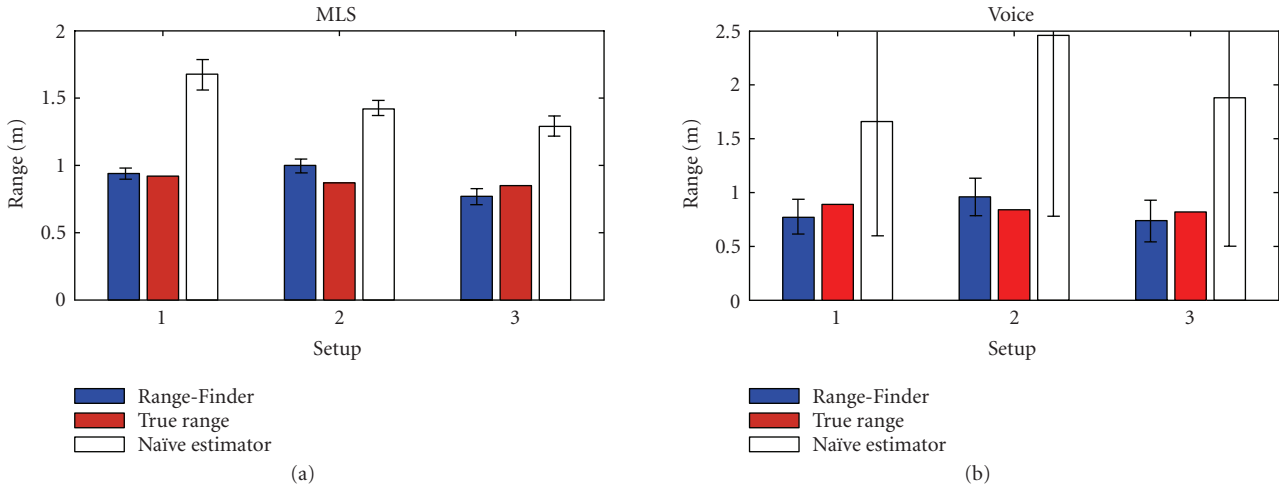

FIGURE 9: Mean range estimates $\pm$ standard deviation from real-room recordings.

implementation of the Range-Finder requires that the microphones be positioned such that there is a sufficient "spread" in the distances from the source to each microphone.

This then precludes the application of the Range-Finder method to the classical scenario of closely spaced microphones and a farfield source. Nonetheless, there are several scenarios in which this requirement is likely to be met and, hence, to which we may successfully apply the Range-Finder method. Consider, for example, the case in which it is required to capture the contributions of a large and distributed group of talkers using a finite number of remote microphones. Under such conditions, it may be found that the classical approach of concentrating the microphones in a closely spaced array causes many of the participants to be a significant distance from all available microphones. As the DRR of recorded sound reduces with increasing distance (see Figure 1) this could cause the contributions from some talkers to be degraded unacceptably. We may, then, prefer to distribute the microphones throughout or around the group of participants such that every potential talker is sufficiently close and has unobstructed access to at least one microphone. Given the wide distribution of the microphones, it

is also likely that, when the sound source is any given talker, we will have access to at least three microphones for which $r_0 \ll c\tau_1 \ll c\tau_2$. We may, therefore, expect accurate range estimates.

We also note that it is often most advantageous to be able to estimate source-microphone ranges in scenarios in which these are not equal for all microphones (so that we may determine which microphones are closest/farthest away, etc.). In addition, when microphones are widely separated, determining their relative locations is likely to be cumbersome and prone to error. Where microphones are frequently moved, say in response to changes in the distribution of participant talkers, it may not be practical to measure microphone locations at all. The Range-Finder algorithm is, therefore, most effective in precisely those scenarios in which it may be required to estimate source-microphone ranges in the absence of reliable microphone-location information.

Our analysis in Section 4 identified scenarios in which the Range-Finder is likely to be inaccurate. Conversely, however, it is possible to specify situations in which the Range-Finder will perform well where many source-localization techniques fail completely. Consider, for example, a situation

in which the microphones and sound source are colinear. For such a setup, the intersensor time delays will be identical for all r_0 (assuming that the source is not in the interior of the array). As a result, no TDE-based localization technique can return a unique estimate of r_0. Where the source and microphones are nearly colinear, we can expect significant error in our range estimates due to errors in the TDEs.

It is apparent, therefore, that the relative positions of the microphones and sound source have a significant bearing upon the accuracy or otherwise of source localization algorithms as well as that of the Range-Finder method. For this reason, any experimental comparisons made between their relative performances would yield scenario-specific results that could not be considered valid in general.

So far, we have assumed an omnidirectional source. In doing so, we have ignored a very pressing practical problem. In reality, sources of interest are likely to be directional and the received sound intensity will depend not only upon the microphone's distance from the source but also its relative azimuth and elevation. If the azimuth-elevation-dependant gain were known for each microphone, it could easily be included in our formulation of the Range-Finder. However, we are unlikely to have such information, or, indeed, to know the orientation of the source relative to the microphones. A further complicating factor is that source directionality is frequency-dependant, with sources typically becoming increasingly directional with frequency.

We should, however, be careful not to overstate the difficulties that directionality presents. Some studies would suggest that directivity would not be a significant factor at frequencies below 4 kHz and within an azimuth of $\pm 30°$ relative to the direction in which a talker is facing [15]. If we could assume that the microphones were within some angular boundaries relative to the source, then we may apply the Range-Finder with confidence. Yet, in the absence of comprehensive data regarding azimuth-elevation-dependant gain for the source of interest, it is hard to see how we might specify and justify the required angular boundaries. We therefore require such data and are limited in application when it is not available.

We note that not all microphones need to be within the specified boundaries; only a minimum of 3 need be and the remaining ranges may be found from the TDEs. Future work will focus on determining the directionality of typical sources and on methods for automatically determining which, if any, of the microphones we should use in the presence of a directional source.

We also note that, when the source and microphones are colinear, the directionality of the source does not pose a problem. However, as previously mentioned, given such a setup, TDE-based source localization techniques will fail. This, therefore, suggests a role for the Range-Finder as an auxiliary source localization algorithm.

ACKNOWLEDGMENTS

REFERENCES

[1] L. Girod and D. Estrin, "Robust range estimation using acoustic and multimodal sensing," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '01)*, vol. 3, pp. 1312–1320, Maui, Hawaii, USA, October-November 2001.

[2] J. Chen, J. Benesty, and Y. Huang, "Time delay estimation in room acoustic environments: an overview," *EURASIP Journal on Advances in Signal Processing*, vol. 2006, Article ID 26503, 19 pages, 2006.

[3] D. Gisch and J. M. Ribando, "Apollonius' problems: a study of their solutions and connections," *American Journal of Undergraduate Research*, vol. 3, no. 1, pp. 15–26, 2004.

[4] E. W. Weisstein, ""Apollonius' Problem" from MathWorld-A wolfram web resource," http://mathworld.wolfram.com/ApolloniusProblem.html.

[5] M. S. Brandstein, J. E. Adcock, and H. F. Silverman, "A closed-form method for finding source locations from microphone-array time-delay estimates," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '95)*, vol. 5, pp. 3019–3022, Detroit, Mich, USA, May 1995.

[6] K. Yao, R. E. Hudson, C. W. Reed, D. Chen, and F. Lorenzelli, "Blind beamforming on a randomly distributed sensor array system," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1555–1567, 1998.

[7] Y. Huang, J. Benesty, G. W. Elko, and R. M. Mersereau, "Real-time passive source localization: a practical linear-correction least-squares approach," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 8, pp. 943–956, 2001.

[8] H. Teutsch and G. W. Elko, "An adaptive close-talking microphone array," in *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (ASSP '01)*, pp. 163–166, New Paltz, NY, USA, October 2001.

[9] S. T. Birchfield and R. Gangishetty, "Acoustic localization by interaural level difference," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '05)*, vol. 4, pp. 1109–1112, Philadelphia, Pa, USA, March 2005.

[10] K. S. Sum and J. Pan, "On the steady-state and the transient decay methods for the estimation of reverberation time," *Journal of the Acoustical Society of America*, vol. 112, no. 6, pp. 2583–2588, 2002.

[11] G. Marsaglia, "Ratios of normal variables," *Journal of Statistical Software*, vol. 16, no. 4, pp. 1–10, 2006.

[12] G. Marsaglia, "Ratios of normal variables and ratios of sums of variables," *Journal of the American Statistical Association*, vol. 60, no. 309, pp. 193–204, 1965.

[13] EASE, "Enhanced acoustic simulator for engineers," version 4.0, http://www.renkus-heinz.com/ease/.

[14] C. H. Knapp and G. C. Carter, "Generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, pp. 320–327, 1976.

[15] J. Huopaniemi, K. Kettunen, and J. Rahkonen, "Measurement and modeling techniques for directional sound radiation from the mouth," in *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (ASSP '99)*, pp. 183–186, New Paltz, NY, USA, October 1999.