A METHOD FOR SOURCE-MICROPHONE RANGE ESTIMATION, USING ARRAYS OF UNKNOWN GEOMETRY, IN REVERBERANT ROOM ENVIRONMENTS

Denis L. McCarthy, Frank Boland

Department of Electronic and Electrical Engineering
Trinity College, Dublin
Dublin 2, Ireland
email: demccart@tcd.ie, fboland@tcd.ie

ABSTRACT

This paper proposes a technique for determining the distance between a sound source and the microphones in an ad-hoc distributed 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 present the results of experiments using simulated and real data to demonstrate the efficacy of our approach.

1. INTRODUCTION

Passive estimation of the distance between a source and a receiver is a central problem in array signal processing. For indoor applications, using microphone arrays, such estimates would have application 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.

Given 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.

Much of the previously published work on source localization has focused on the use of the relative intersensor time-delay-estimates (TDEs) (see [1] and 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 [2]. The numerical solution to this problem, as discovered by Viète (see [3] 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 [4], 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 [5] and [6] the authors estimate the source location via a least-squares fitting of the TDEs for an ad-hoc but known deployment of sensors.

Relative range estimates may also be obtained from a comparison of received signal power. In [7] 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 [8] 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, it is shown by the authors that range estimates are subject to a significant bias in the presence of reverberation.

The use of techniques employing power measurements is commonly restricted to anechoic 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. 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 section 2, we shall 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 present the results of a series of experiments designed to test the performance of our approach. We conclude in section 5.

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, $s(t)$.

$$x_0(t) = \int_0^t s(p) h_0(t-p) dp$$

In the frequency domain

$$X_0(\omega) = S(\omega) H_0(\omega)$$

The received signal power spectrum may be calculated as follows. Note that, for clarity, we omit the dependence on

ω in the sequel.

$$|X_0|^2 = |S|^2 |H_0|^2 \quad (1)$$

$$|X_0|^2 = |S|^2 \left(\left| H_{dp_0} \right|^2 + \left| H_{mp_0} \right|^2 + 2\,\mathrm{Re}\{ H_{dp_0} H^*_{mp_0} \} \right) \quad (2)$$

where H_{dp_0} is the component of H_0 due to direct-path (non-reflected) propagation and H_{mp_0} is the reverberant component due to multipath reflections. In room environments, we may reasonably assume a homogenous medium (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 source-microphone range, r_0, squared.

$$\left| H_{dp_0} \right|^2 \propto \frac{1}{r_0^2}$$

From this

$$\left| H_{dp_a} \right|^2 = \left| H_{dp_0} \right|^2 \left[\left(\frac{r_0}{r_a} \right)^2 \right] \quad (3)$$

The direct-path component decays at a rate of $6dB$ per doubling of distance. The reverberant component will be dependant 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 $\left| H_{mp_0} \right|^2$ *a priori*. However, an investigation of direct-to-reverberant ratio (*DRR*) in real rooms proves informative.

The *DRR* is the ratio of the received sound energy due to the direct-path component and multipath reflections. For a given bandwidth, the *DRR* for a microphone, m_0, may be defined as follows

$$DRR_0 = \frac{\int \left| H_{dp_0} \right|^2 d\omega}{\int \left| H_{mp_0} \right|^2 d\omega}$$

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 room reverberation times were determined experimentally using the transient decay method [9] and were found to be $0.6s, 0.5s$ and $1.1s$ 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 $48kHz$. 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 $5.5s$, also at a sampling rate of $48kHz$. 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 intercept of the best-fit line with the y-axis defines the spatially-averaged "*DRR*-at-1m" and we shall use this metric to describe acoustic conditions in the sequel. The slopes of the fits are $-6.12, -5.99$ and -5.915 decibels per doubling of range for the office, classroom and hall respectively. Given that $\left| H_{dp_0} \right|^2$ decays at a rate of $6dB$ per doubling of the source-microphone range, these results suggest that, in a given room, $E\left\{ \int \left| H_{mp_0} \right|^2 d\omega \right\}$ is a constant that is

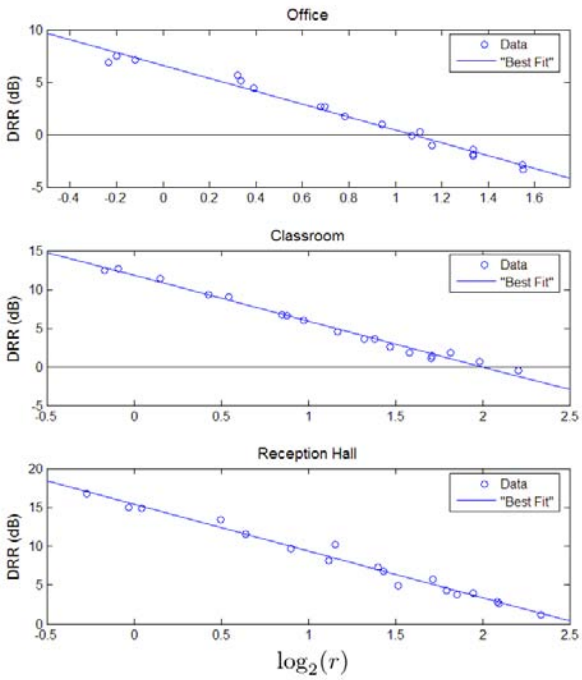

Figure 1: Direct-to-reverberant ratios versus $\log_2(r)$, where r is the source-microphone range in meters. Results shown are for an office, classroom and reception hall.

independent of r_0. We define the following

$$\begin{aligned} F_{a,b} &= \int \left| H_{mp_a} \right|^2 - \left| H_{mp_b} \right|^2 \\ &\quad + 2\,\mathrm{Re}\{ H_{dp_a} H^*_{mp_a} - H_{dp_b} H^*_{mp_b} \} d\omega \end{aligned} \quad (4)$$

Consider the cross-terms in (4). Direct path propagation applies a delay and scaling to a soundwave. 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 soundwaves. As such, $H_{dp}H^*_{mp}$ is also the sum of multiple scaled exponentials. Therefore, invoking the central limit theorem, we shall assume $\int \mathrm{Re}\{ H_{dp_a} H^*_{mp_a} \} d\omega$ and $\int \mathrm{Re}\{ H_{dp_b} H^*_{mp_b} \} d\omega$ to be zero-mean, normally-distributed random variables. Following from our previous results, we also assume $\int \left| H_{mp_a} \right|^2 d\omega$ and $\int \left| H_{mp_b} \right|^2 d\omega$ 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.

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.

$$r_a - r_0 = c\tau_a \tag{5}$$

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, $\widetilde{\tau}_a$. In noiseless, anechoic environments the direct-path sound accounts for all acoustic energy received by the microphones and so, by substituting (1) and (5) into (3) and performing algebraic manipulation, we obtain a simple estimator of r_0.

$$r_0 = \left(c\widetilde{\tau}_a\sqrt{\frac{|H_a|^2}{|H_0|^2}}\right)\left(1-\sqrt{\frac{|H_a|^2}{|H_0|^2}}\right)^{-1}$$

Unfortunately, the presence of reverberation can severely distort this estimate, making the range estimator above unsuitable for use in practical environments.

3.2 The Range-Finder Algorithm

Integrating (2) across the full bandwidth of the signal we obtain the total power of the signal received at m_0.

$$\int |X_0|^2 d\omega = \int |S|^2 \left(\left|H_{dp_0}\right|^2 + \left|H_{mp_0}\right|^2 + 2\,\mathrm{Re}\{H_{dp_0}H^*_{mp_0}\}\right)d\omega$$

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

$$\begin{aligned}\Lambda_{a,b} \;=\;& \int |S|^2 \big(\left|H_{dp_a}\right|^2 - \left|H_{dp_b}\right|^2 + \left|H_{mp_a}\right|^2 \\ & - \left|H_{mp_b}\right|^2 + 2\,\mathrm{Re}\{H_{dp_a}H^*_{mp_a} - H_{dp_b}H^*_{mp_b}\}\big)d\omega\end{aligned} \tag{6}$$

from (3) and (5)

$$\left|H_{dp_a}\right|^2 - \left|H_{dp_b}\right|^2 = \left|H_{dp_o}\right|^2 \left[\frac{r_0^2}{(r_0+c\tau_a)^2} - \frac{r_0^2}{(r_0+c\tau_b)^2}\right]$$

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

$$G_{a,b}(r_o,\tau_a,\tau_b) = \left(\frac{r_0}{r_0+c\tau_a}\right)^2 - \left(\frac{r_0}{r_0+c\tau_b}\right)^2 \tag{7}$$

Let us assume, for the moment, that $|S|^2$ is a constant with respect to frequency (we shall return to this assumption later). Substituting (4) and (7) into (6) yields

$$\Lambda_{a,b} = |S|^2 \left[kG_{a,b}(r_o,\tau_a,\tau_b) + F_{a,b}\right] \tag{8}$$

where $k = \int \left|H_{dp_o}\right|^2 d\omega$. From (8), we see that the difference between the signal power received at two microphones is proportional to the sum of a scaled, deterministic function, $G_{a,b}(r_o,\tau_a,\tau_b)$, 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 $G_{a,b}(r_o,\tau_a,\tau_b)$ terms for clarity.

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

(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

Table 1: The coordinates of the microphone and source locations for the simulated room

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

$$\widetilde{G}_{a,b}(r_o) = G_{a,b}(r_o,\widetilde{\tau}_a,\widetilde{\tau}_b)$$

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

$$\widetilde{r}_0 = \arg\max_{r_0} \left[\frac{1}{\left|\widetilde{\mathbf{G}}(r_0)\right|}\widetilde{\mathbf{G}}(r_0)^T\mathbf{\Lambda}\right] \tag{9}$$

The range of the remaining microphones may easily be estimated by substituting $\widetilde{r}_0$ into equation (5).

Previously, we assumed $|S(\omega)|^2$ to be a constant with respect to frequency. For many signals, including speech, this is unrealistic. In reality, speech is 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 them to hold for cases, such as speech, where the signal has a non-uniform spectral density.

This problem was overcome as follows. The microphone outputs are split into non-overlapping 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 signal component present. $\mathbf{\Lambda}$ is then calculated for each subband. Each $\mathbf{\Lambda}$ is normalized and, from these, an average power-difference-vector, $\overline{\mathbf{\Lambda}}$, found across all the subbands. The range estimate is found, as in (9) by a matched-filtering of $\overline{\mathbf{\Lambda}}$ with $\widetilde{\mathbf{G}}(r_0)/\left|\widetilde{\mathbf{G}}(r_0)\right|$.

4. SIMULATIONS AND EXPERIMENTS

4.1 Simulations

A series of simulations were performed to examine the performance of the Range-Finder algorithm. Our simulated environment was a simple rectangular room of dimensions $[5.25m, 6.95m, 2.44m]$ 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 $10kHz$.

The source-microphone impulse responses were generated using an acoustic modeling software package [12]. A raytracing algorithm was used to determine first $20ms$ of the impulse response after and including the arrival of the direct-path component. Statistical, random reverberant tails were

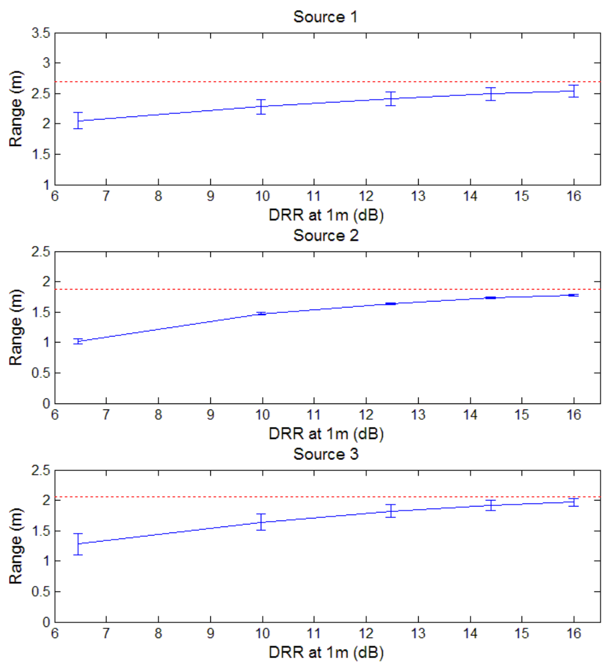

Figure 2: Mean range estimates ± standard deviation for source producing an MLS.

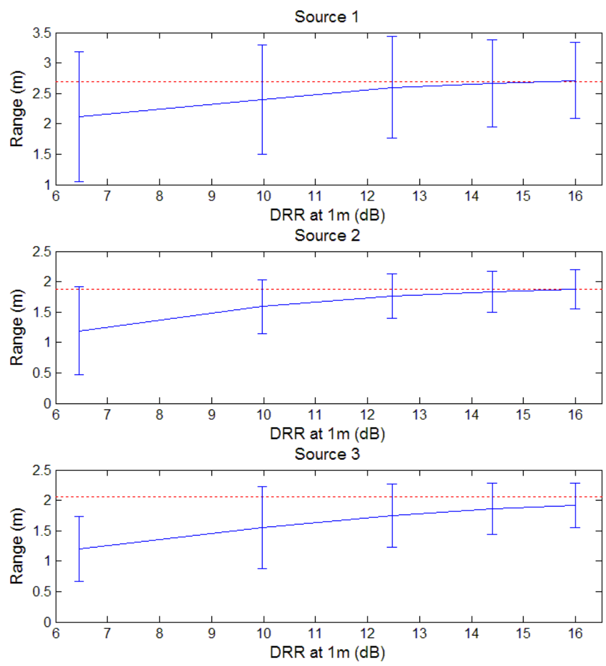

Figure 3: Mean range estimates ± standard deviation for a voice source.

used for the remaining reflections. The direct path components of the impulse responses were scaled to obtain varying direct-to-reverberant ratios. Two "source signals" - a maximum-length-sequence (MLS) of approximate duration $5.5s$ and concatenated voice samples of approximately $13s$ total duration, both bandlimited to avoid aliasing - were convolved with each impulse response to obtain the simulated "recordings".

The recordings were split into segments of 8196 samples and windowed using a Hamming window. The segment overlap was 50%. The TDEs were calculated geometrically, using the source and microphone coordinates and a known speed of sound. For the speech recordings, the signals were separated into eight non-overlapping subbands with bandwidth $\frac{10}{16}kHz$. $\overline{\Lambda}$ was determined as described in section 3. The Range-Finder algorithm was then used to obtain range estimates. Negative range estimates were ignored.

The results for each source are shown in figures (2) and (3). The mean range estimates, ± one standard deviation are shown with respect to the spatially averaged "*DRR*-at-$1m$".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.

In figure (4), the performance of the Range-Finder algorithm is compared to that of the naïve range estimator 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.

4.2 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. For each of three different setups, six microphones were placed at varying distances in front of a loudspeaker. The microphones and loudspeaker were approximately collinear so as to avoid errors due to the directionality of the source. Once again, voice and MLS signals were recorded before being bandlimited and downsampled to a sampling rate of $10kHz$. The recordings were split into segments of 8196 samples and windowed using a Hamming window. The segment overlap was 50%. The TDEs were found using a PHAT-GCC [13]. The results obtained are shown in figure (5) 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.

5. CONCLUSION

We have proposed a method for estimating source-microphone range that is robust against the effects of reverberation and requires no information regarding microphone locations. We have presented the results of a series of experiments, demonstrating its efficacy in simulated and real

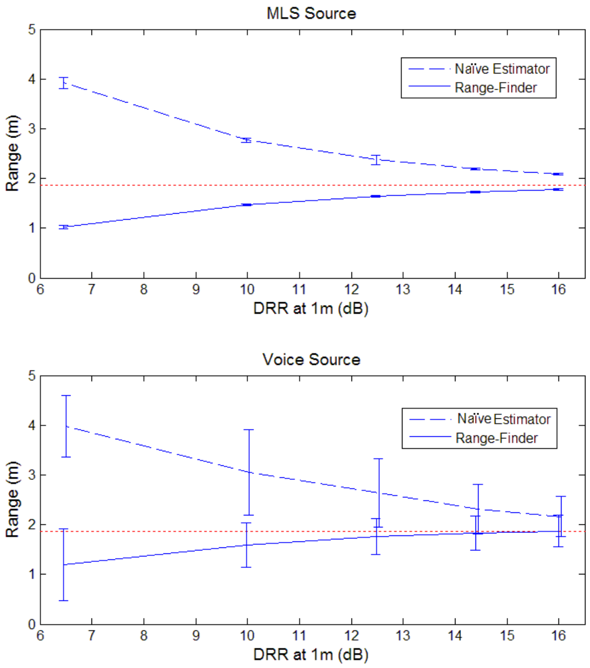

Figure 4: A comparison of mean range estimates (± one standard deviation) for the Naïve range estimator and the Range-Finder algorithm.

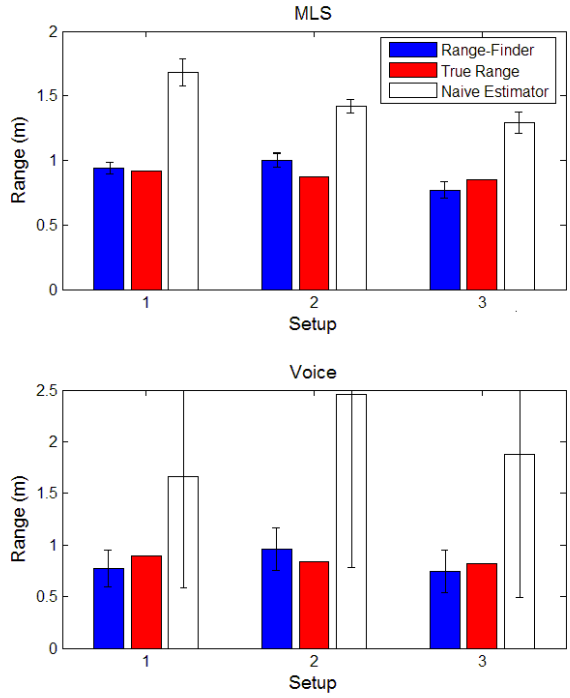

Figure 5: Mean estimates of the range of the closest microphone ± one standard deviation, from real-room recordings

environments. Future work will seek to determine and compensate for the effect of non-ideal sources and sensors.

REFERENCES

[1] J. Chen, J.Benesty, and Y. Huang, “Time delay estimation in room acoustic environments: An overview,” *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. 1-19

[2] D. Gisch, J.M. Ribando,"Apollonius’ problems: a study of their solutions and connections", *American Journal of Undergraduate Research*, vol 3, 2004, pp 15-26

[3] E.W. Weisstein, "Apollonius’ Problem." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/ApolloniusProblem.html

[4] M.S. Brandstein, J.E. Adcock, H.F. Silverman, "A closed-form method for finding source locations from microphone-array time-decay estimates," *Proc ICASSP 1995*, 9-12 May, vol.5, no.pp.3019-3022,

[5] Yiteng Huang, J. Benesty, G.W. Elko, "An efficient linear-correction least-squares approach to source localization," Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on, pp.67-70, 2001

[6] K. Yao, R.E. Hudson,C.W. Reed, D. Chen, F. Lorenzelli, "Blind beamforming on a randomly distributed sensor array system," *Selected Areas in Communications, IEEE Journal on* , vol.16, pp.1555-1567, Oct. 1998

[7] H. Teutsch, G.W. Elko, "An adaptive close-talking microphone array," *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the* , New Paltz, New York, pp.21-24

[8] S.T. Birchfield, R. Gangishetty, "Acoustic localization by interaural level difference," *Proc ICASSP 2005*, Philadelphia, 18-23 March, pp.1109-1112

[9] K.S. Sum, J. Pan, "On the steady state and transient decay methods for estimation of reverberation time", *Journal of the Acoustical Society of America*, vol 112, issue 6, pp. 2583-2588

[10] S.M. Griebel, M.S. Brandstein, "Microphone array source localization using realizable delay vectors ," *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the* , New Paltz, New York, pp.71-74

[11] J.C. Chen, R.E. Hudson, Kung Yao, "Maximum-likelihood source localization and unknown sensor location estimation for wideband signals in the near-field," *Signal Processing, IEEE Transactions on* , vol.50, no.8pp.1843-1854, Aug 2002

[12] EASE; Enhanced Acoustic Simulator for Engineers, version 4.0, http://www.renkus-heinz.com/ease/

[13] C. Knapp, G. Carter, "The generalized correlation method for estimation of time delay," *Acoustics, Speech, and Signal Processing, IEEE Transactions on* , vol.24, no.4, pp. 320- 327, Aug 1976