# ACOUSTIC LOCALIZATION USING REVERBERATION WITH VIRTUAL MICROPHONES

*Teemu Korhonen*

Tampere University of Technology, Department of Signal Processing
P.O. Box 553, 33101 Tampere, Finland, e-mail: teemu.korhonen@tut.fi

## ABSTRACT

This paper proposes the use of reverberant acoustic information with virtual microphones and the time difference of arrival (TDOA) concept to model the contribution of reflected signals. Performance of first order and closest wall reflection models are compared against direct path localization. Real data recordings from a stationary source are used to evaluate the performance. Spatial distributions of the source position generated with the different approaches are analyzed within a two dimensional grid. Results indicate an increase in the speaker localization accuracy using virtual setups in conjunction with the traditional direct path localization.

***Index Terms***— time difference of arrival, acoustic localization, image method, virtual arrays, reverberation

## 1. INTRODUCTION

Reverberation has usually been considered an obstacle and source of error. Especially, the time difference of arrival (TDOA) approaches deal with the ambiguity caused by the multi-path propagation [1].

A simulation method for analysis of a room impulse response was introduced by Allen and Berkley in [2]. Using a process called *image method* the analysis is accomplished with mirrored duplicates of the source.

Reversely the idea of imaging can be turned around and the microphones themselves duplicated as *virtual arrays*. This approach was demonstrated to be viable for the direction of arrival (DOA) estimation by Bergamo *et al.* [3].

In this paper, cross-correlation vectors between channel pairs are used with virtual microphones mirrored through walls. With this approach, the part of the signal correlation attributed to the wall reflection is used in the source localization as a positive contribution. A closed form parameterization for the problem using a two microphone system has been done in [4] with the use of the correlation maxima.

Here, a localization method related to the SRP-PHAT approach [5] is utilized, that is, the cross-correlation values are combined between real and virtual microphone pairs to estimate a spatial likelihood function (SLF).

The rest of the paper is ordered as follows. In Section 2, the signal model and propagation are briefly discussed. Section 3 introduces a simple localization system using virtual microphones. In Section 4, a real data scenario is presented and metrics for performance analysis are defined with results in Section 5. Section 6 discusses the results and motivates further research. Section 7 concludes the paper while highlighting the contribution.

## 2. SIGNAL MODEL AND PROPAGATION

The signal path from an acoustic source to the receiving microphones consists of direct signal propagation and multiple reflections from surfaces and scattering from objects of the environment [6]. All of the propagation paths and contributions from the equipment responses can be instanced in a unique impulse response depending on positions of the receiver and the source. This allows description of the $i$th microphone signal as a sum of convoluted source signals according to the superposition principle:

$$x_i(t) = \sum_{n=1}^{N} s_n(t) * h_{i,n}(t) + w_i(t), \tag{1}$$

where $n$th source signal $s_n$ is convolved with a source–receiver associated impulse response $h_{i,n}$ and $w_i$ represents the i.i.d. noise component of the signal.

Simplifying the nature of the sound propagation, the travel time from source $\mathbf{x}_\mathrm{S}$ to receiver $\mathbf{x}_\mathrm{R}$ can be given as function of their Euclidean distance:

$$t(\mathbf{x}_\mathrm{S}, \mathbf{x}_\mathrm{R}) = \|\mathbf{x}_\mathrm{R} - \mathbf{x}_\mathrm{S}\| c^{-1}, \tag{2}$$

where $c$ is the speed of the sound.

In a two sensor framework the time difference of arrival becomes a variable of interest. Often the sound source resides within different distance from the two receivers $\mathbf{x}_\mathrm{R1}$ and $\mathbf{x}_\mathrm{R2}$. This separation reflects directly as a difference in the propagation times of equation (2) and generates a gap of time between the sensors equal to

$$\tau(\mathbf{x}_\mathrm{S}, \mathbf{x}_\mathrm{R1}, \mathbf{x}_\mathrm{R2}) = t(\mathbf{x}_\mathrm{S}, \mathbf{x}_\mathrm{R1}) - t(\mathbf{x}_\mathrm{S}, \mathbf{x}_\mathrm{R2}). \tag{3}$$

This time delay is evident in the signals of the receiving sensors and usually can be measured using signal processing techniques, such as the generalized cross-correlation (GCC) [7].

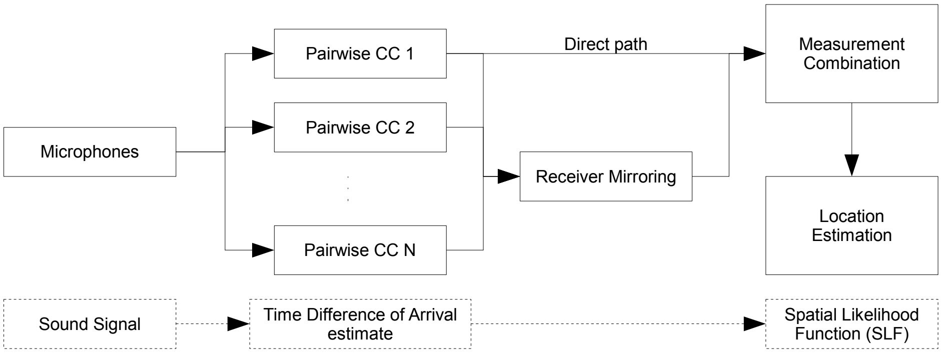

**Fig. 1**. A localization system using receiver mirroring is illustrated here with the flow of relevant information shown below (*dashed boxes*). Sound signals from a set of microphones are mapped to a time difference of arrival (TDOA) likelihood function using pairwise cross-correlation (CC). Spatial information inferred by the cross-correlations is then combined to a location estimate, which is represented as spatial likelihood function (SLF). Here, a step employing reverse imaged virtual microphones is added ("Receiver Mirroring").

## 3. SYSTEM DESCRIPTION

A simple acoustic localization system implementing microphone mirroring is described in Fig. 1. Sound signals from an array of microphones are processed in a pairwise manner using cross-correlation to measure the similarity of the signals as a function of time delay. Theoretical number of pairs from $M$ microphone signals totals to $N = \binom{M}{2}$, which can be significantly lowered by limiting the pairs to subarrays of smaller size. This partitioning of the microphones is also used in the evaluation here, where the distinct arrays of four microphones act as base sets for the microphone pairing.

While there exists multitude of weighting methods for the cross-correlation [7, 8], the resulting distribution should estimate to some degree the time difference of arrival (TDOA) for source signal(s) between the two sensors. This difference estimate relates to the impulse response of Eq. (1) since the correlations appear as spikes between different paths of propagation [1].

A time difference from the two sensor system following Eq. (3) has a non-unique mapping to the spatial coordinates. The locus of potential sources giving the exact same difference forms a half of a hyperboloid. This is expected, since sensors in the foci have a constant range difference, and therefore TDOA, to the hyperboloid locus. An example of these loci can be seen in the Fig. 2., where the values of correlation have been mapped to the spatial coordinates.

Combination of these spatial mappings from multiple microphone pair setups results in an estimate of the source position; this process is also known as multilateration. Phase transform (PHAT) [7] weighting is used here for the cross-correlation and the correlation mappings are combined using Multi-PHAT approach [9]. The analysis is performed within a 2D-grid of coordinates and the resulting distributions are compared against each other.

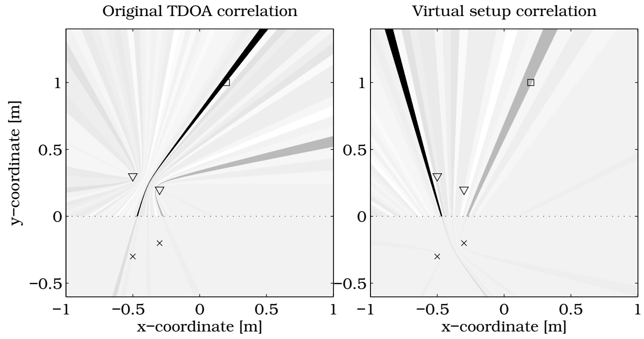

**Fig. 2**. Demonstrated here is the mapping of correlation values to hyperbolic loci from TDOA for two different array scenarios near room boundary (*dotted line*). On the left, two microhones forming an array (*triangles*) project correlation values to spatial coordinates and the source (*square*) rests on the TDOA area with the largest value of correlation. On the right, virtual microphone array (*crosses*) uses the same data and the source lies on a secondary locus. Positive contribution from the correlations reverses between the scenarios.

### 3.1. Virtual microphone setups

An integral part of the reverberant localization system is the *receiver mirroring*. Here, the microphone pairs are spatially mirrored through the walls while retaining the correlation data of the original setup. This results in a number of virtual microphones residing outside of the original room.

The contributing part of the correlation measurement changes when the array is mirrored. The hyperboloid locus associated to the direct path propagation becomes an error source while the locus with correlation from a reflected signal intersects the true source. An example is presented in Fig. 2.

The use of multiple virtual microphone setups lessens the

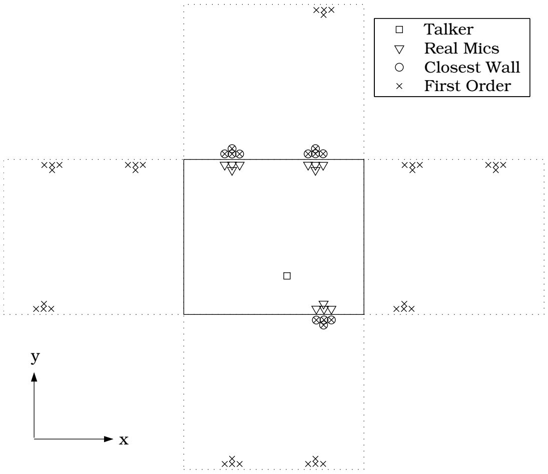

**Fig. 3**. Original microphone locations and their virtual counterparts for the real data scenario are presented here. Actual room (*solid lines*) is mirrored with each of its walls and resulting images are shown (*dotted lines*). The set of virtual microphones used by the closest wall approach is included in the set used by the complete first order reflection method.

adverse effect of the main locus from a single setup. For virtual setups, hyperboloids from the direct-path propagation are not assumed to correlate with any source inside the enclosure and their intersections are mostly coincidental, evening out their effect.

Two scenarios using different number of virtual microphones are investigated. The first method uses only the virtual setups mirrored from the nearest wall (*circles* in Fig. 3). The second uses all of the reflections (*crosses*). The closest wall reflections double the amount of microphones and the complete wall reflections multiply the number of virtual microphones by the number of walls. However, the increase in computational load is minimal, since the pairwise cross-correlation need not to be recalculated for the virtual setups.

## 4. PERFORMANCE ANALYSIS

The different methods were compared using real data recordings of a monologue in two different locations. The source position is assumed semistationary and is annotated accordingly.

### 4.1. Recording setup

Recording environment is a meeting room with dimensions of $4.53 \times 3.96 \times 2.59$ m. The room contains some furniture, three small diffusors on the walls and a projection canvas. Three microphone arrays are set on the walls. The reverberation time for the room is $0.25$ s, which has been measured using the maximum-length sequence (MLS) technique [2]. The room configuration has been detailed in [9].

The microphone arrays consist of four microphones each. Shape of an array is an upside down T-shape parallel to a wall with the fourth microphone set out of the plane. The total number of microphones in the room is $M = 12$ and using an array grouping descriped in Section 3, the number of real microphone pairs totals to $N = 18$.

The sound source is a seated person uttering a pre-defined sentence. The source is facing the center of the room and annotated to the xyz-coordinates of $\langle 2.60, 0.99, 1.14 \rangle$ for the first position and $\langle 2.30, 3.00, 1.13 \rangle$ for the second.

### 4.2. Performance metrics

The comparison of the methods is done within an uniformly spread, two dimensional grid $G_t$ using a spatial resolution of 20 mm. Data is processed in Hanning windowed frames $t \in [1 \ldots T]$ of length 23.2 ms. For every method of interest, the performance is evaluated using grid weights

$$w_{t,\mathbf{x}} = \prod_{p \in P} R^{\mathtt{PHAT}}(\tau_{p,\mathbf{x}}, t), \tag{4}$$

with $R^{\mathtt{PHAT}}(\tau_{p,\mathbf{x}}, t)$ being the PHAT-weighted cross-correlation normalized between [0, 1]. The TDOA value $\tau_{p,\mathbf{x}}$ of Eq. (3) is indexed by a grid point $\mathbf{x} \in G_t$ and a pair $p$ from the set of microphone pairs $P$, which contains both true and virtual pairs used by the method of choice.

A weighted distance error based (WDE) metric for grid points $\mathbf{x}$ is used here as

$$e_{\mathtt{WDE}} = \frac{1}{T} \sum_{t=1}^{T} \sum_{\mathbf{x} \in G_t} w_{t,\mathbf{x}} \|\mathbf{x} - \mathbf{x}_{\mathrm{S}}\|, \tag{5}$$

where the grid weights $w_{t,\mathbf{x}}$ sum to unity for any single frame $t$ and the annotated true location is the constant $\mathbf{x}_{\mathrm{S}}$.

The performance metric (5) measures the average source distance of distribution mass produced by the method of choice. Since the majority of the mass should be centered tightly around the true position, another metric (WSDE) using square distance $\|\cdot\|^2$ is also used in the evaluation.

## 5. RESULTS

The results for different methods for both of the metrics are documented in Table 1. The performance of a zero information, uniform weight distribution is also included for comparative purposes. The methods of the interest are:

- Baseline design employing direct path propagations
- Information from closest wall reflections
- Complete first order wall reflections
- Baseline including the closest wall reflections
- Baseline including the complete first order reflections

The uniform distribution sets the bar for all of the other methods. Since the weights are equal, no target information is inferred and metric value of Eq. (5) depends solely on the reference position and room dimensions. The baseline design concentrates the distribution near the true source position and therefore has a better WDE and WSDE than the uniform case.

Interestingly enough, both reflection models infer some usable information alone, even surpassing the baseline performance in the case of the full first order reflections. The methods employing baseline design and the additional reverberant information give the best results.

## 6. DISCUSSION

In scope of this paper some items of interest were left unaddressed and thus deserve further research. The number of virtual arrays was limited to the first order reflections from the walls. Ceiling and floor reflections and scattering from major objects were omitted. Similarly, the reflection and propagation models were overly simplified.

The use of the second and higher order reflections brings new challenges as the number of virtual TDOA setups grows exponentially, and the propagation times extend with the extra distance of the image rooms. A window mismatch can rise with short enough windows of calculation.

In addition, the effective contributions from the virtual arrays need more investigation. The results of Section 5 indicated positive contribution from the method using complete first order reflections. However, further analysis shows that the virtual setups mirrored from the wall farthest from the source actually impair the combined performance. This issue could be addressed, e.g. in a form of a distance dependent weighting.

The author would like to point out that a realistic localization scheme should use another parameterization with lower computational costs, instead of the grid method used here for the analysis. A good numerical approach is the particle filtering method with comprehensive tutorial presented in [10].

**Table 1**. Results for different methods are shown here for both talker positions. The performance metrics are weighted (square) distance errors. Results for the uniform distribution are included for purely comparative purposes. Last two methods employing the baseline design with different amounts of wall reflections give the best results.

| | **Position 1** | | **Position 2** | |
|---|---|---|---|---|
| **Method** | **WDE** | **WSDE** | **WDE** | **WSDE** |
| *Uniform* | 1.86 | 4.11 | 1.86 | 4.09 |
| Baseline | 1.75 | 3.85 | 1.74 | 3.76 |
| Closest Wall (CW) | 1.79 | 3.94 | 1.80 | 3.94 |
| First Order (FO) | 1.71 | 3.71 | 1.67 | 3.51 |
| Baseline + CW | 1.62 | 3.52 | 1.62 | 3.48 |
| Baseline + FO | 1.53 | 3.31 | 1.51 | 3.20 |

These points will be addressed with additional work supplemented with more extensive real data evaluations and parameter simulations.

## 7. CONLUSION

This paper proposes the use of image method with virtual microphone setups for cross-correlation based localization. The method does not significantly increase the computational complexity while still harnessing the destructive effects of reverberation induced cross-correlation peaks into constructive information. The contribution of positive information was considered from two reflection schemes: the complete first order wall reflections and a limitation to the closest wall reflections. Both approaches were shown to increase the likelihood mass near the source position, therefore resulting in improved localization accuracy.

## 8. REFERENCES

[1] J. Chen, J. Benesty, and Y. Huang, "Performance of GCC- and AMDF-based time-delay estimation in practical reverberant environments," *EURASIP Journal on Applied Signal Processing*, vol. 2005, no. 1, pp. 25–36, 2005.

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

[3] P. Bergamo *et al.*, "Collaborative sensor networking towards real-time acoustical beamforming in free-space and limited reverberance," *IEEE Transactions on Mobile Computing*, vol. 3, no. 3, pp. 211–224, Jul.-Sep. 2004.

[4] W. Yan, W. Qun, B. Danping, and J. Jin, "Acoustic localization in multi-path aware environments," ICCCAS 2007, 11-13 Jul. 2007, pp. 667–670.

[5] J. H. DiBiase, *A High-Accuracy, Low-Latency Technique for Talker Localization in Reverberant Environments Using Microphone Arrays*, Ph.D. thesis, Brown University, Providence, Rhode Island, 2000.

[6] L. Beranek, *Acoustics*, American Institute of Physics, New York, NY, USA, 1986.

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

[8] J. Chen, J. Benesty, and Y. Huang, "Time delay estimation in room acoustic environments: An overview," *EURASIP Journal on Applied Signal Processing*, vol. 2006, 2006.

[9] P. Pertilä, T. Korhonen, and A. Visa, "Measurement combination for acoustic source localization in a room environment," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2008, 2008.

[10] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, Feb. 2002.