Acta Acustica united with Acustica
Vol. 98 (2012) 418 – 440

DOI 10.3813/AAA.918527

Acoustic Reflection Localization from Room Impulse Responses

Sakari Tervo, Jukka Pätynen, Tapio Lokki
Aalto University School of Science, Department of Media Technology, P.O. Box 15400, 00076 Aalto, Finland.
Sakari.tervo@aalto.fi

Summary
The localization of acoustic reflections, i.e., the image-sources, is of interest when analyzing the acoustics of concert halls and auditoriums. The location is needed, for example, in room acoustic studies, auralization, inference of room geometry, or when estimating the acoustic properties of surfaces. This article studies the localization of acoustic reflections from spatial impulse responses. The contribution of this article is threefold. First, the article proposes a new method for localization that takes advantage of the time of arrival (TOA) estimation. Secondly, it is proposed that TOA and time difference of arrival (TDOA) information, present in the spatial room impulse responses, are combined in two novel ways. Thirdly, the performance of the proposed localization methods is compared to the existing state-of-the-art localization methods in the acoustic reflection localization task. Theoretical performance is investigated and experiments using real and simulated data are conducted. The TOA-based methods are found to achieve reasonably good performance in the reflection localization task. When TOA and TDOA information is combined the performance clearly improves.

PACS no. 43.60.Hg, 43.60.Jn

1. Introduction

Location of acoustic reflections, i.e, the image-sources, is a useful piece of information in room acoustic studies, auralization, room geometry inference, and in-situ measurement of acoustical properties of surfaces from room impulse responses. The locations of the reflections can be used together with the sound source location to deduce the normals and the locations of the reflective surfaces [1, 2, 3], that is, to infer the room geometry. In addition, the location of the reflection is needed for accurate time windowing of the reflection from the room impulse response when estimating, for example, the absorption coefficient of the surface from in-situ measurements [4, 5].

Localization of acoustic reflections is often performed with methods developed for sound source localization. An overall framework of the methods, considered in this article, is given in Figure 1 and in Table I. Many of the localization methods are based on time difference of arrival (TDOA), which is the difference between the arrival times of a wave front at two sensor positions. In addition, methods that are based on time of arrival (TOA), the time that the wave front takes to travel from the source to the receiver location, have been presented for source localization. TOA is available when the signal of the sound source is known, for example, in a room impulse response. Recently, methods that combine the measurements (CM) of TOA and TDOA have been introduced. In addition, some methods that use directly the received signals exist.

The objective of this article is to evaluate which of the localization methods are best applicable for the localization of acoustic reflections from spatial room impulse responses recorded with compact microphone arrays. The evaluation of the methods is done by theoretical comparison with Cramér-Rao lower bound (CRLB), Monte-Carlo simulations, and real data experiments in two enclosures.

The contribution of the article is the following. First, a new method for localization that takes advantage of TOA estimation is proposed. Second, it is proposed that TOA and TDOA information, present in the spatial room impulse responses, are combined in two novel ways. In the first combination approach, the TOA and TDOA estimation functions are combined by addition in the spatial domain. In the second one, the estimation functions are first considered as pseudo-likelihoods and then combined by multiplication in the spatial domain. Third, the performance of the methods is compared to the existing state-of-the-art localization methods and studied in the reflection localization task.

The rest of the paper is organized as follows. Research related to the localization methods applied in this article are presented in Section 2. The signal model for acoustic reflection localization is presented in Section 3. Methods for reflection localization are presented in Sections 4-6. Theoretical limits are yielded in Section 7. Simulation and real data experiments are conducted and discussed in Section 8. Finally, section 9 concludes the paper.

Received 26 August 2011,
accepted 4 February 2012.

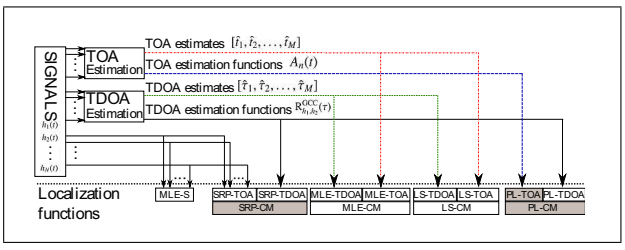

Figure 1. Overall scheme of the data flow in different localization functions. Methods introduced in this article are presented with gray background.

Table I. Contributions of the article and examples of the references for each studied method. SRP: Steered response power, MLE: maximum likelihood estimation, PL: Pseudo-likelihood, LS: least squares, TOA: Time of arrival, TDOA: Time difference of arrival, CM: Combined measurements of TOA & TDOA, S: signal model, N/A: Not applicable, * MLE-S, LS-S, SRP-TDOA, and SRP-TOA will produce the same location estimate in theory, when independent and identical errors are assumed.

	Localization Method			
Data	MLE	SRP	PL	LS
TDOA	[6]	[7]*	[8]	[9]
TOA	[6]	[10]*	Here	[11]
CM	[12]	Here	Here	[13]
S	[14]*	N/A	N/A	[15]*

2. Related work on localization methods

2.1. TDOA estimation and TDOA-based localization

Considerable research efforts have been put to the TDOA estimation problem over the last decades [16, 17, 18, 19, 20, 21, 22]. One of the most popular approaches is the generalized correlation method (GCC) [21]. Other commonly used methods are based on average difference function [18]. The accuracy of TOA and TDOA based localization is limited by the sampling frequency. In [23] parabolic fit and in [24] exponential fit are proposed for interpolating the TDOA estimate. The interpolation of the whole TDOA estimation function using exponential and parabolic fits is presented in [25].

The TDOA-based source localization combine the TDOA information spatially over several microphone pairs. One of the simplest methods for localization is the maximum likelihood estimation (MLE) method, which is formulated for TDOA in [6]. In MLE-TDOA, the TDOA estimation errors are given an error probability density function (PDF) and these PDFs, with TDOA estimates as their means, are then combined spatially.

When independent and identically distributed errors are assumed, MLE-TDOA can be presented as a least squares (LS) problem. With these certain assumptions the LS solution is also the best linear unbiased estimator (BLUE) [26]. The TDOA LS problem has gained a lot of attention in research [27, 9, 28, 29, 30, 6, 31, 32, 33, 34, 35], mostly because the LS solution can be given in closed form by making first some assumptions on the error or on the signal. Since the LS solutions are a special case of MLE, only MLE is considered in the rest of this paper.

A popular approach for TDOA-based localization is the steered response power (SRP), which has been studied extensively [36, 37, 38, 7, 8, 39, 25, 10, 40] and it has been followed by various modifications and optimizations [38, 39, 40]. In the SRP-TDOA method the TDOA estimation functions are combined spatially by addition. Therefore, there is no requirement for error PDF in SRP-TDOA as there is in MLE-TDOA. It has been shown that the SRP-TDOA leads to an equal localization function as the energy of the delay-and-sum beamformer when GCC is used as the TDOA estimation function [10, 41].

Quite recently, pseudo-likelihood (PL) methods for the TDOA estimation functions have been proposed in speech source localization [41, 8]. In PL-TDOA, the TDOA estimation functions are combined spatially by multiplication, as opposed to addition used in SRP-TDOA. As in SRP-TDOA, there is no need for the error PDF. That is, PL-TDOA is a non-parametric method as is SRP-TDOA. It is shown in [8], that PL-TDOA achieves better performance in the speech source localization task than SRP-TDOA.

2.2. TOA estimation and TOA-based localization

In addition to TDOA-based source localization, source localization methods based on time of arrival (TOA) estimation have been presented in the past. TOA is often also referred to as time of flight.

The most direct method for TOA estimation is a simple peak-picking algorithm [42, 43, 44]. Also, it has been proposed that statistical features, such as kurtosis, can be used to detect peaks [45]. Other methods are based on correlation or some other similarity measure and they require a priori knowledge on the signal [44, 43]. These methods are then very similar to TDOA estimation methods. Moreover, in principle also the onset detection methods used in music signal analysis could be used for TOA estimation [46]. The performance of TOA estimation is not widely studied under additive noise to the knowledge of the present authors. In addition, TOA estimation accuracy can be improved by basic Fourier-interpolation or by assuming a shape for the TOA estimation function, similarly as in the TDOA estimation.

As with TDOA, also for TOA the MLE method has been proposed [6]. In the MLE-TOA method, the error PDF is assumed for the TOA estimation error, instead of the TDOA estimation error used in the MLE-TDOA method.

The MLE method for TOA can also be formulated into a LS problem. Closed form solutions for the LS problem for TOA have been presented [11, 47, 48, 49]. Again, since the LS solutions are a special case of MLE, only MLE is considered in the rest of this paper.

Yet another method that uses the TOA-model is the energy of the delay-and-sum beamformer [10]. In the frequency domain, with continuous signals, such as speech, this method produces the same location estimate as the SRP-TDOA when cross correlation is used as the TDOA estimation function [10].

2.3. Combination of TOA and TDOA-based localization

More recently, the TOA and TDOA information have been proposed to be combined in the MLE framework [50, 13, 12, 51]. In [12] it was shown that it is advantageous to combine the TOA and TDOA information for source localization from room impulse responses. Also, LS solutions have been presented for the combination of TOA and TDOA data [13].

2.4. Maximum likelihood methods for signal and error models

In addition to MLE-TDOA and MLE-TOA, MLE solutions have been proposed for different signal and error models [15, 52, 53, 38, 54, 14]. These methods solve the maximum likelihood estimate of the location with respect to the assumed signal model and noise model which may include a model of the environment. In particular, MLE methods for reverberant conditions [54], and conditions where the microphones or source have different characteristics [15, 52, 53, 38, 14] have been presented. When independent and identically distributed errors are assumed for each receiver, and the microphone and source directivities are omnidirectional, the MLE can be presented as a LS problem [15]. In theory, the LS solution is the same as the one given by SRP-TOA or SRP-TDOA [15].

2.5. Localization of reflections and room geometry estimation

A relevant topic to the localization of reflections is the localization of the reflective surfaces, or the blind estimation of room geometry. Namely, the estimation of reflective surfaces is equivalent to localization of first order reflections. The localization of reflection, and estimation of room geometry from room impulse responses have been studied in several research articles [55, 56, 57, 3, 58, 59, 60, 61, 62, 1, 2, 63, 64, 65, 66, 67]. The approaches are based on TOA, TDOA, and direction of arrival (DOA) estimation. The TOA estimation requires that the loudspeakers and microphones are time-synchronized, and the TDOA and DOA based methods do not require synchorinization.

In [59, 60] a technique called the spatial impulse response rendering (SIRR) is developed. The analysis part of SIRR inspects the direction of arrival of the reflection and the diffuseness of the sound field. Since the analysis is done in short time windows, the location of the reflections can be deduced using the a priori knowledge of speed of sound, the time of arrival and the estimated DOA which is calculated from sound intensity vectors.

A spherical microphone array with an integrated video camera is used in [55, 56, 57] for visually inspecting the reflections. The energy of the spherical beamformer output applied to an impulse response that is divided into short time windows is overlayed on top of a panorama video image from the center of the microphones. The location of the reflection is then inspected visually for each frame. The maximum of the beamformer output corresponds to the DOA of the reflection and the distance to the reflection is calculated from the time stamp of the current frame.

In [58] the reflections are localized using TDOA estimation with a microphone array that consists of 8 microphones. The method is demonstrated in an auditorium.

In [3] the room geometry is estimated by rotating a B-format microphone around a loudspeaker, directed towards the microphone. The estimation is based on the TOA and the DOA of the first arriving reflection in each direction. The DOA estimates is calculated from the sound intensity vectors. For each direction, a single TOA and DOA estimate is obtained. In the post-processing phase, the TOA and DOA measurements are grouped using hierarchical clustering to avoid estimating the same plane multiple times.

The tracing of reflections using highly directive loudspeaker, a compact microphone array, and localization methods is proposed in [67]. The localization is based on detecting when the reflections arrive to a microphone array based on the statistical features of the sound field and estimating the DOA based on sound intensity vectors.

A general framework localizing the reflections using a small microphone array and a loudspeaker is presented in [68]. In addition, a comparison of the performance of SRP-TDOA based methods and sound intensity vector based methods in the direction estimation task reveals that SRP-based methods should be preferred.

In [1], the reflecting plane parameters are estimated by rotating an omni-directional microphone around a loudspeaker which is directed towards a microphone. The impulse responses are transformed into an acoustic localization map from where the local maxima correspond to the plane locations. The acoustic localization map is calculated using the delay-and-sum beamformer with direct TOA mapping. As the source position is known, the plane parameters can be calculated.

In [2] the reflecting plane parameters are estimated with a common tangent algorithm in two dimensional space. The problem is first formulated into quadractic equation that describes the relation between the TOAs and plane parameters and source location. For a single reflection the solution of this quadratic equation provides the parameters of a single plane. The solution is called the common tangent algorithm (COTA). For multiple planes, the estimated TOAs are first grouped using the generalized Hough transform and then the plane for each group is solved using the COTA. The generalized Hough transform detects the TOAs that describe the same plane. The approaches in [2] are extended to three dimensions in [65]. Moreover, a closed form solution for the plane parameter estimation from the quadratic equation is presented in [66].

In [63], the COTA is applied for the estimation multiple plane parameters in two dimensional space. Whereas in [2] the grouping was done with the generalized Hough trasnform in [63] the grouping is done with an iterative search. The iteration proceeds as follows. First the parameters of the closest plane are estimated. Then the TOAs

associated with the first plane are removed and the search is performed again. This iteration is performed as many times as there are a priori known planes.

In [64] a closed form solution to the above mentioned quadratic equation that describes the relation between the TOA and plane parameters is presented for the 2-D case. In the solution, two planes are selected where the cost function is inhomogeneous. Then, the gradients of the cost function on these planes are solved analytically. The minimum of the obtained solutions corresponds to the plane parameters. Moreover, the generalized Hough transform is applied to improve the estimation of the parameters of a single plane.

Acoustic imaging for finding room geometry and other acoustic properties of enclosure is applied in [61, 62]. Acoustic imaging is based on the inverse extrapolation of the Kirchoff-Helmholtz and Rayleigh integrals. An acoustic image can be created by measuring multiple impulse responses, for example, on a line grid with B-format microphone [61, 62].

In [69] a circular microphone array is used around a loudspeaker to estimate the room geometry. A constrained room model and L1-regularized least-squares method is used to obtain the locations of walls. This method can be considered as semi-blind since it requires the knowledge of the number of walls.

The room geometry has also been estimated from continuous signals [70, 71, 69, 72, 73, 74]. The advantage of these approaches is that there is no need for controlled source signals. This paper only considers the localization of reflections from room impulse responses. Therefore, the approaches that localize the reflective surfaces or reflections from continous signals are no longer considered.

2.6. Theoretical performance

The theoretical performance of the estimation methods can be studied by using Cramér-Rao lower bound (CRLB) [26]. In CRLB, first the error PDF is assumed on the measurements. Then, the Fisher information is obtained by squaring and deriving the error PDF with respect to the parameter to be estimated. CRLB is the inverse of the Fisher information.

3. Preliminaries and Signal Model

3.1. Impulse Response and Reflection Signal Model

In a room environment, the sound $p_s(t)$, emitted from the sound source at position $\mathbf{s}$, and received at microphone n at position $\mathbf{r}_n$, is affected by the impulse response $h(t; \mathbf{r}_n, \mathbf{s})$

$$p(t; \mathbf{r}_n, \mathbf{s}) = h(t; \mathbf{r}_n, \mathbf{s}) * p_s(t) + w'(t), \tag{1}$$

where $*$ denotes convolution and, $w'(t)$ is the measurement noise, independent and normally distributed for each microphone. For simplicity, the impulse response measured at microphone n is noted with $h_n(t)$ through the rest of the paper. Furthermore, the number of microphones is denoted with N.

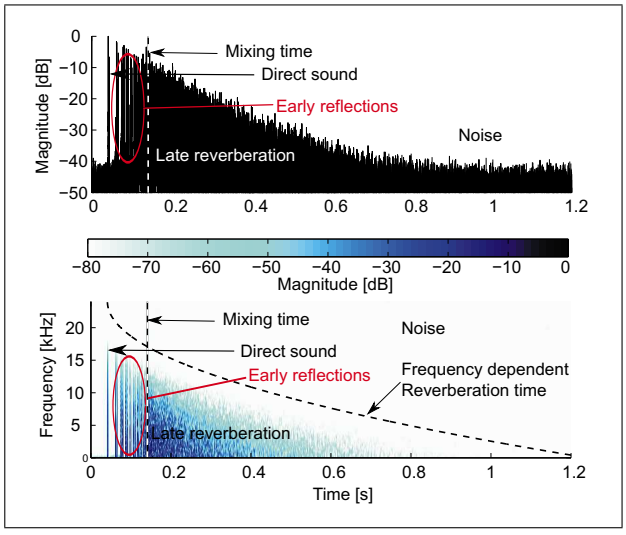

Figure 2. Impulse response in the time-domain (upper) and in the time-frequency-domain (lower). The early reflections studied here appear before the late reverberation starts.

A room impulse response can be roughly divided into three parts, the direct sound, the early reflections, and the late reverberation. Figure 2 illustrates these three categories in principle. The early reflections should be identifiable until the so-called mixing time[1].

In this paper, the impulse response is considered as

$$\begin{aligned} h_n(t) &= \sum_{k=1}^{K} h_n^k(t) + w(t) \\ &= \sum_{k=1}^{K} \int_0^{\pi} H_n^k(\omega) \mathrm{e}^{\mathrm{j}\omega t} \, \mathrm{d}\omega + w(t), \end{aligned} \tag{2}$$

where ω is the angular frequency, $h_n^k(t)$ are the impulse-like reflection signals, $k = 1, \ldots, K$ indicated the index of each reflection and $w(t)$ is measurement noise, independent, and identically distributed with normal distribution for each microphone.

Each of the impulse-like reflection signals $h_n^k(t)$ consists of the impulse response of the loudspeaker $s(t)$ and attenuation factor a_n^k via

$$h_n^k(t) = a_n^k s\big(t - t_n^k(\mathbf{x})\big) + w_n(t), \tag{3}$$

where $t_n^k(\mathbf{x})$ is the TOA related to the distance of the path of a reflection k at location $\mathbf{x}_k$,

$$t_n^k(\mathbf{x}_k) \triangleq t(\mathbf{r}_n; \mathbf{x}_k) = c^{-1} \|\mathbf{r}_n - \mathbf{x}_k\|, \tag{4}$$

and c is the speed of sound. The location of the reflection $\mathbf{x}$ refers to the image source location [78], see Figure 3 for further explanation. In the frequency domain the signal model is given as

$$H_n^k(\omega) = A_n^k(\omega)\, S(\omega) \mathrm{e}^{-\mathrm{j}\omega t_1(\mathbf{x})} + W_n(\omega) \tag{5}$$

[1] One definition of the mixing time is given in [75] and several ad-hoc methods for its estimation in [43, 76, 77].

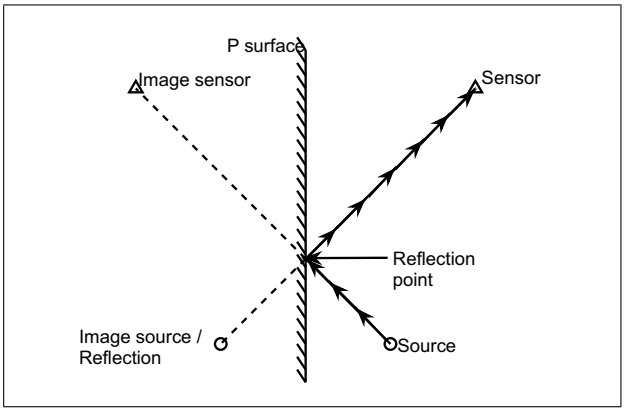

Figure 3. Concept of the image source and the reflection.

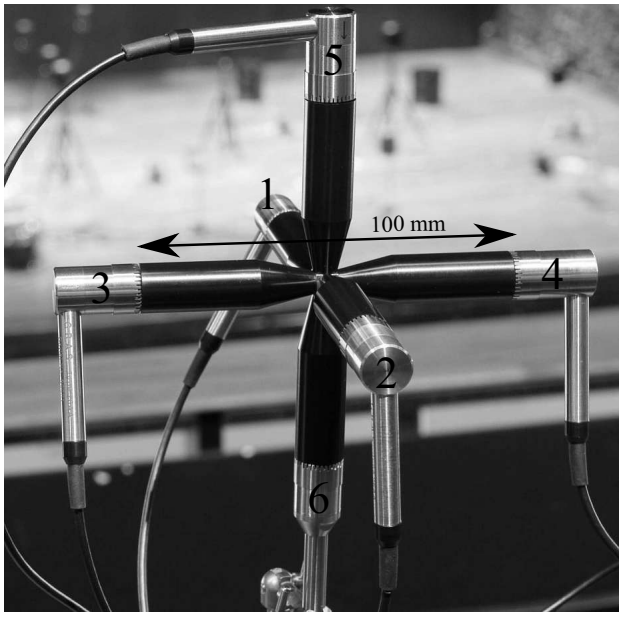

Figure 4. G.R.A.S. sound intensity vector type 50VI microphone array used in this paper. The distance between a microphone pair on each axis is 100 mm.

where the attenuation, source, noise, and received signal have spectral densities $G_{a,a}(\omega)$, $G_{s,s}(\omega)$, $G_{w_n,w_n}(\omega)$, and $G_{h_n,h_n}(\omega)$, respectively. For clarity the index k is omitted from hereon.

The attenuation factor $A_n(\omega)$ is dependent at least on the properties of the surface and air absorption, the distance from the source to the microphones, and the directivity of the source and of the microphones. Although in real situations the phase of the reflection depends on the frequency, especially if the surface is uneven, the analysis of the room impulse responses assumes that with early reflections the phase is frequency independent, i.e. $A(\omega) \in \mathbb{R}$. Thus, ideal specular reflections are assumed. In addition, the analysis does not differentiate between diffraction and reflections, but all the events are considered to be reflections.

The analysis in this article is done for spatial impulse responses measured with a compact microphone array. An example of the microphone array setup is shown in Figure 4. The array has $N = 6$ microphones and the spacing between the two microphones on an axis is 100 mm. Other microphone arrays can also be used with the presented methods, however, all the results are presented here are for this setup.

The processing of the spatial impulse response measured with a compact microphone array is usually done in short time windows [55, 56, 59, 60, 67, 68]. Based on previous knowledge [79, 59, 60, 67], a good window size for the analysis of early reflections is from 1 ms to 4 ms. Here, the time indexes, i.e. the starting and ending points of the analysis time window are denoted with $t_{\text{start}} \leq t \leq t_{\text{end}}$.

It is further assumed that there is only one specular reflection present per analysis window. This is generally true in the early part of the impulse response for the first order reflections, especially if a short analysis window is used, large spaces, such as auditoriums or concert halls are under investigation, and the length of the reflection signal $s(t)$ is less than the length of the analysis window. Theoretical limitations for the analysis window size can be derived trivially from the echo density, which is given e.g. in [80, p. 92].

Since the microphone array is compact, it is assumed that the attenuation factor is equal for all microphones and a single reflection $A_n(\omega) = A(\omega)$. In addition, all the other spectral densities are also assumed to be independent of the microphone for the same reason. If the intra-sensor distances would be large this assumption could not be made.

With the assumption of only one specular reflection per analysis window, the relation between the measurement noise and the signal energy and attenuation factor define the signal-to-noise ratio (SNR). A recognizable feature of the room impulse response, also shown in Figure 2, is the fact that the SNR decreases as the time increases. Thus, the reflections that arrive later in time have a lower SNR. The SNR of each reflection in the impulse response measurement is defined here as

$$\text{SNR} = \frac{\text{E}\left\{h^2(t)\right\}_{t_{\text{start}}}^{t_{\text{end}}}}{\text{E}\left\{w^2(t)\right\}}, \tag{6}$$

where $\text{E}\{\cdot\}$ is the expected value.

4. Time difference of arrival estimation

TDOA (Time difference of arrival) is the time difference between two TOAs and it is calculated with spherical wave propagation model as

$$\begin{aligned}\tau_{i,j}(\mathbf{x}) &\triangleq \tau(\mathbf{r}_i, \mathbf{r}_j; \mathbf{x}) = t(\mathbf{r}_i; \mathbf{x}) - t(\mathbf{r}_j; \mathbf{x}) \\ &= c^{-1}\left(\|\mathbf{r}_i - \mathbf{x}\| - \|\mathbf{r}_j - \mathbf{x}\|\right).\end{aligned} \tag{7}$$

In 3-D, a TDOA is presented by a paraboloid.

In the TDOA estimation, the task is to estimate the time delay $\tau_{i,j} = t_i - t_j$ between two received signals $h_i(t)$ and $h_j(t)$. The maximum argument of the estimation function $R_{h_i,h_j}(\tau)$ is the TDOA estimate, i.e.,

$$\hat{\tau}_{i,j} = \arg\max_{\tau}\left\{R_{h_i,h_j}(\tau)\right\}. \tag{8}$$

Here GCC [21] and average squared difference function (ASDF) [18] are used for TDOA estimation in the reflection localization task. Due to the measurement noise $w(t)$, a TDOA is corrupted by an additive noise component,

$$\hat{\tau}_{i,j} = \tau_{i,j} + \varepsilon_{\mathrm{TDOA}}, \tag{9}$$

where $\hat{\cdot}$ denotes estimate. The error components are assumed to be normally distributed with zero mean, $\varepsilon_{\mathrm{TDOA}} \sim \mathcal{N}(0, \sigma^2_{\mathrm{TDOA}})$, where σ^2_{TDOA} is the variance of the TDOA error component $\varepsilon_{\mathrm{TDOA}}$.

4.1. Generalized correlation method

The most commonly used TDOA estimation method is the generalized correlation method [21]. The generalized cross correlation (GCC) function between two received impulse responses h_i and h_j is calculated as [21]

$$\mathrm{R}^{\mathrm{GCC}}_{h_1,h_2}(\tau) = \mathcal{F}^{-1}\left\{\mathcal{W}(\omega)G_{h_1,h_2}(\omega)\right\}, \tag{10}$$

where $\mathcal{W}(\omega)$, $\mathcal{F}^{-1}$, and $\hat{G}_{h_1,h_2}(\omega)$ are the weighting function, inverse Fourier transform, respectively, and cross spectral density between h_i and h_j.

4.1.1. Maximum likelihood weighting

The maximum likelihood weighting for the GCC function is given as [21]

$$\mathcal{W}^{\mathrm{ML}}_{h_1,h_2}(\omega) = \frac{1}{\left|G_{h_1,h_2}(\omega)\right|}\frac{C_{h_1,h_2}(\omega)}{\left[1 - C_{h_1,h_2}(\omega)\right]} \tag{11}$$

where

$$C_{h_1,h_2}(\omega) = \frac{\left|G_{h_1,h_2}(\omega)\right|^2}{G_{h_1h_1}(\omega)G_{h_2h_2}(\omega)} \tag{12}$$

is the magnitude squared coherence function. For the derivation of ML weighting function see [21]. Since the noises are assumed to be uncorrelated the true spectral densities can be written as [21]

$$G_{h_1,h_2}(\omega) = G_{a,a}(\omega)G_{s,s}(\omega)\mathrm{e}^{-\mathrm{j}\omega\tau_{i,j}}, \tag{13}$$

$$G_{h_1,h_1}(\omega) = G_{a,a}(\omega)G_{s,s}(\omega) + G_{w_1,w_1}(\omega), \tag{14}$$

$$G_{h_2,h_2}(\omega) = G_{a,a}(\omega)G_{s,s}(\omega) + G_{w_2,w_2}(\omega). \tag{15}$$

Then, by using these equivalences in equation (11), one has [21]

$$\mathcal{W}^{\mathrm{ML}}_{h_1,h_2}(\omega) = \Big[G_{a,a}(\omega)G_{s,s}(\omega)\Big]\Big[G_{w_1,w_1}(\omega)G_{w_2,w_2}(\omega) + G_{a,a}(\omega)G_{s,s}(\omega)(G_{w_1,w_1}(\omega) + G_{w_2,w_2}(\omega))\Big]^{-1}. \tag{16}$$

In practical situation, since the signal is an impulse response, it is easy to estimate the noise auto power spectral density $G_{w_1,w_1}(\omega)$ from the beginning of the impulse response. Then, the auto spectral density of the source signal and the attenuation is given by the inverse of equation (14), e.g., $G_{a,a}(\omega)G_{s,s}(\omega) = G_{h_1,h_1}(\omega) - G_{w_1,w_1}(\omega)$.

If the noise can not be estimated, the first version of the ML weighting in equation (11) can be used, but the coherence should then be estimated using for example Welch's approach [81, 82]. Coherence estimation can be problematic for non-stationary signals [17]. In addition, since it includes additional computational load it is not used in this article.

4.1.2. Other weighting functions

Two weighting functions of GCC that do not require the a priori knowledge of the auto power spectral densities are considered. These weighting functions are called as the direct cross correlation (CC) weighting [21]

$$\mathcal{W}^{\mathrm{CC}}(\omega) = 1 \tag{17}$$

and phase transform (PHAT) [21]

$$\mathcal{W}^{\mathrm{PHAT}}_{h_i,h_j}(\omega) = 1/\|G_{h_i,h_j}(\omega)\|. \tag{18}$$

4.2. Average square difference function

Similar to the generalized correlation method are the difference function based methods [18]. In these approaches, two signals are subtracted from each other, while the other signal is delayed by the TDOA. Here, the absolute squared difference function is also tested [20, 18],

$$\mathrm{R}^{\mathrm{ASDF}}_{h_i,h_j}(\tau) = \int_{-T/2}^{T/2}\left[h_i(t) - h_j(t-\tau)\right]^2 \mathrm{d}t, \tag{19}$$

where T is the length of the integration window. With ASDF, instead of the maximum, the minimum argument of the estimation function is the TDOA estimate

$$\hat{\tau}_{i,j} = \arg\min_{\tau}\left\{\mathrm{R}^{\mathrm{ASDF}}_{h_i,h_j}(\tau)\right\}. \tag{20}$$

5. Time of arrival estimation

TOA is given by equation (4). In TOA estimation, the time delay t_n of a signal is estimated. In a short time window the maximum argument of the TOA estimation function $D_n(t)$ is the TOA estimate

$$\hat{t}_n = t_{\mathrm{start}} + \arg\max_{t}\left\{D_n(t)\right\}, \tag{21}$$

where t is limited by the starting point, and the ending point of the time window, i.e., $t_{\mathrm{start}} \le t \le t_{\mathrm{end}}$.

A TOA is corrupted by an additive noise component due to the measurement noise $w(t)$,

$$\hat{t}_n = t_n + \varepsilon_{\mathrm{TOA}}, \tag{22}$$

Also for TOA, the error components are assumed to be normally distributed with zero mean, $\varepsilon_{\mathrm{TOA}} \sim \mathcal{N}(0, \sigma^2_{\mathrm{TOA}})$, where σ^2_{TOA} is the variance of the TOA error component $\varepsilon_{\mathrm{TOA}}$.

Since the problem in TOA estimation is similar to TDOA estimation, the TDOA estimation methods introduced above can also be applied for TOA estimation, given that the source signal $s(t)$ is known a priori. Next the GCC estimation methods are presented for the TOA estimation and a simple peak picking method is reviewed. The ASDF for TOA estimation is the same for TDOA estimation, with the exception that the other received signal is replaced by $s(t)$.

5.1. Auto-correlation method

This method requires a priori information on the sound source used. First a reference $s(t)$ is measured for the applied sound source in free-field conditions: in an anechoic chamber, or windowed from an in-situ impulse response. The reference represents the waveform of the emitted impulse response from the source. The reference is then correlated with the impulse response

$$D^{\mathrm{AC}}_{s,h_1}(t) = \int_{-T/2}^{T/2} s(\xi) h_1(\xi - t)\,\mathrm{d}\xi, \tag{23}$$

where AC denotes auto-correlation.

Defrance *et al.* use similar auto-correlation approach for detecting reflections from a single impulse response [42, 43]. In addition, similar auto-correlation method has been used to detect the TOA of a reflection as a preliminary task before absorption coefficient calculations [83].

5.1.1. Maximum likelihood weighting

The auto-correlation function can be given in the frequency domain as the generalized correlation function

$$\mathrm{D}^{\mathrm{AC}}_{s,h_1}(\tau) = \mathcal{F}^{-1}\left\{\mathcal{W}_{s,h_1}(\omega) G_{s,h_1}(\omega)\right\}. \tag{24}$$

By definition, the maximum likelihood weighting also for this method is given by equation (11). Since the other signal is the true signal without noise the spectral densities can be written as

$$G_{s,h_1}(\omega) = A_1(\omega) G_{s,s}(\omega) \mathrm{e}^{-\mathrm{j}\omega t_1}, \tag{25}$$

$$G_{h_1,h_1}(\omega) = G_{a,a}(\omega) G_{s,s}(\omega) + G_{w_1,w_1}(\omega). \tag{26}$$

Then, the ML weighting for the auto-correlation method is given as

$$\begin{aligned} W^{\mathrm{ML-AC}}_{s,h_1}(\omega) &= \frac{1}{\left|G_{s_1,h_1}(\omega)\right|} \frac{C_{s,h_1}(\omega)}{\left[1 - C_{s,h_1}(\omega)\right]} \\ &= \frac{A_1(\omega)}{G_{w_1,w_1}(\omega)} \end{aligned} \tag{27}$$

where $C_{s,h_1}(\omega)$ is the magnitude squared coherence between s and h_1.

The problem with the AC method is that, a real loudspeaker emits different impulse responses to different directions. Thus, as a priori knowledge, the AC method requires the impulse response of the loudspeaker to each direction. This can be artificially done using the sparse impulse response technique [67], where a directional loudspeaker is spanned to all different directions. Then each emitted impulse response is ideally the same.

5.2. Maximum absolute pressure

Peak detection is a straightforward method for TOA estimation. It is assumed that the arriving sound wave introduces an impulse, a local maximum or minimum, that can be detected. The maximum argument is then the estimated TOA

$$\hat{t}_n = \arg\max_t \left\{|h_n(t)|\right\}. \tag{28}$$

Windowing or filtering may be applied to the impulse response prior to maximization.

5.3. Other methods

The statistical features of impulse response differ when there is a reflection present in the analysis window [76, 79, 84, 45]. One way of measuring the statistical difference is the kurtosis [45]. Another option is to detect the peak from a local absolute pressure ratio between the current absolute pressure and its surroundings [79]. These statistical approaches are not used in this article.

6. Localization functions

This section presents localization methods from earlier research that are applied in this article for the localization of reflections. Also three novel localization methods for reflections are proposed.

For each method, the maximum argument of the localization function $P(\mathbf{x})$ is the location estimate, i.e.

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} \left\{P(\mathbf{x})\right\}. \tag{29}$$

For notational convenience, a TDOA is denoted by $\tau_m(\mathbf{x})$, where $m = \{i, j\} = 1 \ldots M$ is a tuple, and M is the number of microphone pairs. Morever, the TDOA estimates are denoted with $\hat{\tau}_m$, and the TDOA estimation function $R_{h_i,h_j}(\tau)$ with $R_m(\tau)$. In this article, the number of microphones is $N = 6$, and the number of microphone pairs is $M = 15$. Then, the microphone pairs m from 1 to 15 are $\{\{1,2\}, \{1,3\}, \{1,4\}, \{1,5\}, \{1,6\}, \{2,3\}, \{2,4\}, \{2,5\}, \{2,6\}, \{3,4\}, \{3,5\}, \{3,6\}, \{4,5\}, \{4,6\}, \{5,6\}\}$.

6.1. Maximum likelihood estimation for time of arrival and time difference of arrival based localization

The MLE function for TDOA is given as the joint probability density function [6]

$$\begin{aligned} P_{\text{MLE-TDOA}}(\mathbf{x}) &= \prod_{m=1}^{M} p\left(\hat{\tau}_m; \tau_m(\mathbf{x})\right) \\ &= \frac{\exp\left(-\frac{1}{2}\left[\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}(\mathbf{x})\right] \Sigma^{-1} \left[\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}(\mathbf{x})\right]^{\mathrm{T}}\right)}{(2\pi)^{(M)/2} \sqrt{\det(\Sigma_{\mathrm{TDOA}})}}, \end{aligned} \tag{30}$$

where $p(\hat{\tau}_m; \tau_m(\mathbf{x}))$ is the normal error probability density function for a TDOA estimate,

$$\hat{\boldsymbol{\tau}} = \left[\hat{\tau}_1, \hat{\tau}_2, \ldots, \hat{\tau}_M\right], \tag{31}$$

$$\boldsymbol{\tau}(\mathbf{x}) = \left[\tau_1(\mathbf{x}), \tau_2(\mathbf{x}), \ldots, \tau_M(\mathbf{x})\right], \tag{32}$$

$$\Sigma_{\mathrm{TDOA}} = \mathbf{I}\sigma^2_{\mathrm{TDOA}}, \tag{33}$$

and $\mathbf{I}$ is the identity matrix, σ_{TDOA} is the error standard deviation and $\tau_m(\mathbf{x})$ is given by equation (7).

The MLE function for TOAs, assuming normally distributed errors is given as [6]

$$P_{\text{MLE-TOA}}(\mathbf{x}) = \prod_{n=1}^{N} p(\hat{t}_n; t_n(\mathbf{x})) \tag{34}$$

$$= \frac{\exp\left(-\frac{1}{2}\left[\hat{\boldsymbol{t}} - \boldsymbol{t}(\mathbf{x})\right] \Sigma^{-1} \left[\hat{\boldsymbol{t}} - t(\mathbf{x})\right]^{\mathrm{T}}\right)}{(2\pi)^{(N)/2} \sqrt{\det(\Sigma_{\mathrm{TOA}})}}, \tag{35}$$

where $p(\hat{t}_n; t_n(\mathbf{x}))$ is the normal error probability density function for a TDOA estimate,

$$\hat{\boldsymbol{t}} = \left[\hat{t}_1, \hat{t}_2, \ldots, \hat{t}_M\right], \tag{36}$$

$$\boldsymbol{t}(\mathbf{x}) = \left[t_1(\mathbf{x}), t_2(\mathbf{x}), \ldots, t_M(\mathbf{x})\right], \tag{37}$$

$$\Sigma_{\text{TOA}} = \mathbf{I}\sigma^2_{\text{TOA}}, \tag{38}$$

with σ_{TOA} as the error standard deviation and $\tau_m(\mathbf{x})$ is given by equation (4).

For combining the TOA and TDOA information with MLE an assumption is made, that the TDOA and TOA have independent errors. Then, the MLE function for combined measurements (CM) is given as the multiplication of MLE-TOA and MLE-TDOA functions [12],

$$P_{\text{MLE-CM}}(\mathbf{x}) = P_{\text{MLE-TOA}}(\mathbf{x}, \sigma_{\text{TOA}}) \cdot P_{\text{MLE-TDOA}}(\mathbf{x}, \sigma_{\text{TDOA}}). \tag{39}$$

If different error variances σ^2_{TOA} and σ^2_{TDOA} are assumed for TOA and TDOA, respectively, the MLE-TOA and MLE-TDOA functions have different weightings. In [12], it is found that $\sigma^2_{\text{TOA}} = \sigma^2_{\text{TDOA}}$ is a reasonable choice.

6.2. Maximum likelihood estimation for the signal model

Earlier, the maximum likelihood estimation was formulated with respect to TOA and TDOA estimates. It is also possible to formulate the MLE directly with respect to the signal and noise models [54]

$$\begin{aligned} P_{\text{MLE-S}}(\mathbf{x}) &= \prod_{\omega} p\big(\boldsymbol{H}(\omega); \mathbf{x}\big) \\ &= \prod_{\omega} \Bigg\{ \Big[\exp\big(-1/2 \big[\boldsymbol{H}(\omega) - \boldsymbol{D}(\omega, \mathbf{x}) A(\omega) S(\omega)\big] \\ &\qquad \boldsymbol{Q}^{-1}(\omega) \big[\boldsymbol{H}(\omega) - \boldsymbol{D}(\omega, \mathbf{x}) A(\omega) S(\omega)\big] \big) \Big] \\ &\qquad \cdot \Big[(2\pi)^{N/2} \sqrt{\det(\boldsymbol{Q}(\omega))} \Big]^{-1} \Bigg\}, \end{aligned} \tag{40}$$

where

$$\boldsymbol{H}(\omega) = \left[H_1(\omega), H_2(\omega), \ldots, H_N(\omega)\right]^{\mathrm{T}}, \tag{41}$$

$$\boldsymbol{D}(\omega, \mathbf{x}) = \left[\mathrm{e}^{-\mathrm{j}\omega t_1(\mathbf{X})}, \mathrm{e}^{-\mathrm{j}\omega t_2(\mathbf{X})}, \ldots \mathrm{e}^{-\mathrm{j}\omega t_N(\mathbf{X})}\right]^{\mathrm{T}}, \tag{42}$$

$$\boldsymbol{Q}(\omega) = \mathbf{I}\sigma^2_F. \tag{43}$$

where $\sigma^2_F = G_{w,w}(\omega), \forall\omega$ is the expected variance of the measurement noise $w(t)$.

6.3. Steered response power

A popular family of TDOA-based acoustic source localization functions is the SRP methods. In these methods, the acoustic source localization likelihood is evaluated as a spatial combination of cross correlation functions $\mathrm{R}_m(\tau)$ for each location candidate, denoted with $\mathbf{x}$ [7],

$$P_{\text{SRP-TDOA}}(\mathbf{x}) = 1/M \sum_{m=1}^{M} \mathrm{R}_m\big(\tau_m(\mathbf{x})\big). \tag{44}$$

The SRP using generalized correlation method with GCC-PHAT function is commonly referred to as SRP-PHAT function.

The signals can be similarly steered using TOAs as the TDOA estimation functions are steered using TDOAs. In steered beamforming the signals are artificially steered by delaying them towards a location. The delay-and-sum beamformer is considered as the most basic case of beamforming [85]. When the delay-and-sum beamformer output is squared the output is SRP [10]

$$P_{\text{SRP-TOA}}(\mathbf{x}) = \int \left| 1/N \sum_{n=1}^{N} h_n\big(t - t_n(\mathbf{x})\big) \right|^2 \mathrm{d}t. \tag{45}$$

This function implements MLE-S in equation (40) in time domain without the noise or signal model. However, if equation (45) is implemented in the frequency domain, the TOA information is lost, since SRP-TOA becomes the same as SRP-TDOA with an additional (constant) energy term [41, 10].

Since the room impulse responses are already directly mapped into the TOAs, the time variable becomes $t = 0$. The time integral over dt then has no effect on the localization function and equation (45) is written as

$$P_{\text{SRP-TOA}}(\mathbf{x}) = \left| 1/N \sum_{n=1}^{N} h_n\big(t_n(\mathbf{x})\big) \right|^2. \tag{46}$$

The TOA and TDOA information can be both used to measure the position of a reflection. Intuitively, the next step is to combine both TOA and TDOA information. The SRP function, when TDOA and TOA information are both used, is here proposed to be calculated as

$$P_{\text{SRP-CM}}(\mathbf{x}) = (1 - W) P_{\text{SRP-TOA}}(\mathbf{x}) + W P_{\text{SRP-TDOA}}(\mathbf{x}), \tag{47}$$

where $W \in (0, 1)$ is a weighting factor, included in this function since the steered response is effectively used twice in SRP-CM. The weight W then emphasizes either the TDOA functions or TOA functions.

6.4. Pseudo-likelihood

Recently it was shown in [8] and [86] that the use of multiplication instead of addition is advantageous when combining the TDOA estimation function. This leads to a pseudo-likelihood function [41, 8, 86]

$$P_{\text{PL-TDOA}}(\mathbf{x}) = \prod_{m=1}^{M} R_m\big(\tau_m(\mathbf{x})\big), \tag{48}$$

where PL stands for pseudo-likelihood. It should be noted that thresholding and shaping has to be done for the TDOA estimation functions so that they are non-negative pseudo-likelihoods [70]. It is straightforward to show that, if the maximum of each TDOA estimation function $R_m(\tau)$ is modeled with a PDF, PL-TDOA and MLE-TDOA methods are the same methods.

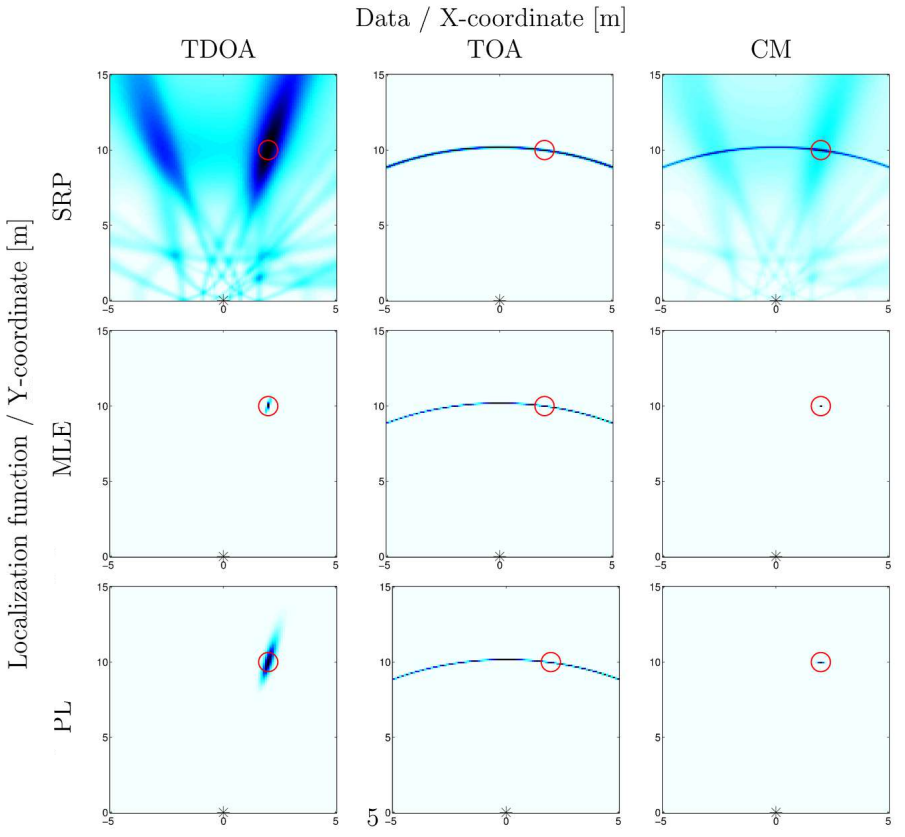

Figure 5. Examples of acoustic source localization functions for a grid of (x, y, z)-locations with $z = 1.5$ m in the case of no noise. The microphone array at (0,0,0) m is denoted with a star, and the reflection at (2,11,1.5) m is denoted with a circle.

Here it is proposed that the PL function for TOA is formed by multiplying the individual TOA estimation functions, i.e.,

$$P_{\text{PL-TOA}}(\mathbf{x}) = \prod_{n=1}^{N} D_n\big(t_n(\mathbf{x})\big). \tag{49}$$

Thresholding and shaping can be done for the TOA estimation functions so that they are non-negative pseudo-likelihoods. In the most simplest case, the function TOA estimation function is the absolute maximum of the room impulse response

$$P_{\text{PL-TOA}}(\mathbf{x}) = \prod_{n=1}^{N} \big|\, h_n\big(t_n(\mathbf{x})\big)\,\big|. \tag{50}$$

The analogy between PL-TOA and MLE-TOA is the same as the analogy between PL-TDOA and MLE-TDOA. If only one maximum is selected in PL-TOA from the impulse response, and the corresponding TOA is modeled with a PDF, PL-TOA and MLE-TOA are the same methods.

The proposed combination of the PL functions is formed by multiplying the individual PL-TOA and PL-TDOA functions

$$P_{\text{PL-CM}}(\mathbf{x}) = P_{\text{PL-TOA}}(\mathbf{x})\, P_{\text{PL-TDOA}}(\mathbf{x}). \tag{51}$$

Again, if the TOA and TDOA estimation functions maxima are modeled with PDFs, then PL-CM is equal to the MLE-CM function.

Note that in PL-CM the weighting of PL-TOA or PL-TDOA similarly as SRP-TOA and SRP-TDOA in SRP-CM has no effect, since the weighting will not change the maximum of PL-CM. However, although the PL-CM can not be weighted, the logarithmic version of it can be, i.e.,

$$\begin{aligned}\lambda_{\text{PL-CM}}(\mathbf{x}) \;=\; & (1-W)\log\big\{P_{\text{PL-TOA}}(\mathbf{x})/N\big\} \\ & + W\log\big\{P_{\text{PL-TDOA}}(\mathbf{x})/M\big\},\end{aligned} \tag{52}$$

where $W \in (0, 1)$ and the log-pseudo-likelihoods of TOA and TDOA are normalized with N and M, respectively.

6.5. Examples of the localization maps

Examples of the localization maps with different methods are provided in Figure 5. The data is a simulated perfect reflection with no noise at (2,11,1.5) m, and the microphone array is located at (0,0,0) m. As can be seen from Figure 5, the TDOA based methods provide good information about the direction whereas the TOA based methods seem to work well in the distance estimation. When the TOA and TDOA are combined a better localization method is made. As seen from Figure 5 SRP localization maps have more "ghosts" than other methods, i.e., local maxima that do no correspond to the true reflection location.

In this example the simplest method for the search of the maximum is presented. That is, the maximum can be found using a predefined grid of locations. However, this is often not very efficient, therefore some other methods for the search of the maximum are discussed next.

6.6. Search of the extremum

Basically any global optimization method can be used for the search of the extremum. In general, there is no way of ensuring that the global optimization method will converge to the global extremum since localization with spherical wave propagation model is a non-linear problem. Therefore there is usually a need for Monte-Carlo simulations to validate the optimization method for a certain problem. Since the literature on the optimization methods is extensive, only some selected methods used for localization are discussed here.

The most naive and straightforward method for the search of the maximum is to use a (predefined) grid of location candidates as in Figure 5. The drawbacks of this approach is the slowness of the computation when the grid size is large. Namely, in a 3-D grid of a volume of say a concert hall, the number of data points is very large, especially if the spacing between the grid points is small. Thus the estimation meets the curse of dimensionality. However, since the evaluation of the ASL function is the same at each selected time instant for any data point, the process can be parallelized as in [87]. Using parallel computation decreases the used time for the evaluation in total, but requires special implementation considerations and special equipment, such as the general purpose graphic processing unit.

In this article, the well-known Nelder-Mead method is used to find the extremum in the ASL functions [88]. The Nelder-Mead method requires a proper initial guess for the source location.

6.7. Computational complexity of the localization methods

Although reflection localization within the framework of this article is always an offline task, some comparison between the complexity of the methods is provided. The complexity is compared with the 'Big O notation', $\mathcal{O}(\cdot)$.

For basic beamforming the complexity is built up from the number of ASL function evaluations E, the length of the signal L, and the number of the microphones N. For cross correlation, the complexity of the estimation function is $\mathcal{O}(L\log\{L\})$ and since all the microphones are used twice in the calculation of the ASL function the complexity increases by $\mathcal{O}(N^2)$. [70]

Moreover, the complexity of the TOA estimation with the simple peak picking method is $\mathcal{O}(L)$. For TOA estimation with AC approach the complexity is $\mathcal{O}(L\log\{L\})$, but that approach is not used here. Since the MLE-S method calculates the localization function over frequency band, its complexity is increased by the number of frequencies used $\mathcal{O}(F)$.

Table II lists the computational complexity of the methods introduced in this section. The TOA-based methods have lower computational complexity than the other methods since the room impulse responses are directly mapped into the TOAs.

As the number of evaluations increases, the computational complexity and time of MLE-S increases. This results was also pointed out by Korhonen for the time-domain beamformer [70]. However, when the number of evaluations increases, the computational complexity of the time domain beamformer presented here (SRP-TOA) does not increase as rapidly as the computational complexity of the conventional time-domain beamformer. This is due to direct mapping of impulse responses to TOAs, which does not add any computational complexity.

Table II. Computational complexity of the localization methods in the reflection localization task. E: Number of localization function evaluations, L: Length of the signal, N: Number of the microphones, and F: Number of frequency bins.

Data	Method	Complexity
TOA	MLE-S	$\mathcal{O}(EL\log\{L\}NF)$
TOA	SRP, PL, & MLE	$\mathcal{O}(NL+NE)$
TDOA	SRP, PL, & MLE	$\mathcal{O}(N^2L\log\{L\}+EN^2)$
TOA & TDOA	SRP, PL, & MLE	$\mathcal{O}(NL+NE+N^2L\log\{L\}+EN^2)$

7. Theoretical performance

The positions of the sensors and the source, as well as the signal and the noise have an effect on the localization variance. These effects can be theoretically measured with Cramér-Rao lower bound (CRLB) [26] analysis, which is given by the inverse of the Fisher Information matrix [26, Ch. 3]

$$\mathrm{cov}(\hat{\boldsymbol{\theta}}) \geq \mathbf{J}(\boldsymbol{\theta})^{-1}. \tag{53}$$

The Fisher information matrix is defined as the squared derivative of the log-likelihood of the estimate probability density function [26, Ch. 3]

$$\mathbf{J}(\boldsymbol{\theta}) = \mathrm{E}\left\{\left[\frac{\partial\lambda(\boldsymbol{\chi}(\boldsymbol{\theta});\hat{\boldsymbol{\chi}})}{\partial\boldsymbol{\theta}}\right]\left[\frac{\partial\lambda(\boldsymbol{\chi}(\boldsymbol{\theta});\hat{\boldsymbol{\chi}})}{\partial\boldsymbol{\theta}}\right]^{\mathrm{T}}\right\}. \tag{54}$$

The theoretical boundaries given in this section use the assumption that the source signal and noise signals are white Gaussian noise. This assumption is necessary and required to make the signal model in equation (3) mathematically tractable [21, 82].

7.1. Time difference of arrival estimation

The Fisher information for TDOA estimation is given as [82, 89]

$$J(\tau) = \frac{2T}{2\pi}\int_0^\infty(\omega)^2\frac{C_{h_1,h_2}(\omega)}{1-C_{h_1,h_2}(\omega)}\,\mathrm{d}\omega \tag{55}$$

where the magnitude squared coherence is related to the SNR via [82]

$$\frac{C_{h_1,h_2}(\omega)}{1-C_{h_1,h_2}(\omega)} = \frac{\mathrm{SNR}^2(\omega)}{1+2\mathrm{SNR}(\omega)}, \tag{56}$$

where

$$\mathrm{SNR}(\omega) = \frac{G_{a,a}(\omega)G_{s,s}(\omega)}{G_{w,w}(\omega)} \tag{57}$$

Setting the power spectral densities flat and assuming

$$G_{a,a}(\omega)G_{s,s}(\omega) = \begin{cases} G_{a,a}G_{s,s} \\ \quad \text{if } |\omega| \in \left[\omega_c - \frac{B}{2}, \omega_c + \frac{B}{2}\right] \\ 0 \quad \text{otherwise} \end{cases} \tag{58}$$

with center frequency ω_c, frequency band B, and assuming that the noise spectral densities are equal, i.e. $G_{w_1,w_1}(\omega) = G_{w_2,w_2}(\omega) = G_{w,w}(\omega)$, the Fisher information formulates into [89]

$$J(\tau) = \frac{\mathrm{SNR}^2}{1 + 2\mathrm{SNR}} \frac{T}{\pi}(B\omega_c^2 + B^3/12). \tag{59}$$

This analysis is valid only for $T \gg 2\pi/B$ and for sufficiently large SNR values, for details, see [90, 89].

7.2. Time of arrival estimation

The derivation of CRLB for TOA estimation can be done by following the steps given for TDOA estimation in [82]. The derivation of the AC function in equation (24) with ML weighting in equation (27) gives

$$J(t) = \frac{T}{\pi}\int_0^\infty (\omega)^2 \frac{C_{s,h_1}(\omega)}{1 - C_{s,h_1}(\omega)}\,\mathrm{d}\omega \tag{60}$$

with magnitude squared coherence equal to

$$\frac{C_{s,h_1}(\omega)}{1 - C_{s,h_1}(\omega)} = \frac{G_{a,a}(\omega)G_{s,s}(\omega)}{G_{w,w}(\omega)} = \mathrm{SNR}(\omega) \tag{61}$$

With the same assumptions on the frequency band and noise as with above with TDOA in equation (58), the Fisher information formulates to

$$J(t) = \mathrm{SNR}\frac{T}{\pi}\left(B\omega_c^2 + B^3/12\right). \tag{62}$$

Since $\mathrm{SNR} > 0$, it can be seen that the CRLB is always smaller for TOA estimation since the Fisher information in TOA estimation is higher.

7.3. Localization

The log-likelihood of the localization with respect to signal model is given by equation (40). The Fisher information matrix is formulated as [15, 52, 91]

$$\mathbf{J}(\mathbf{x}) = 2\Re[\mathbb{H}(\mathbf{D}(\omega,\mathbf{x}))^H Q^{-1}\mathbb{H}(\mathbf{D}(\omega,\mathbf{x}))], \tag{63}$$

where

$$\begin{aligned}\mathbb{H}(\omega,\mathbf{D}(\mathbf{x})) &= \left[\frac{\partial A(\omega)S(\omega)D_1(\omega,\mathbf{x})}{\partial\mathbf{x}},\ldots,\right. \\ &\qquad \left.\cdots\frac{\partial A(\omega)S(\omega)D_N(\omega,\mathbf{x})}{\partial\mathbf{x}}\right] \\ &= A(\omega)S(\omega)\left[\frac{\partial \mathrm{e}^{-\mathrm{j}\omega t_1(\mathbf{X})}}{\partial\mathbf{x}},\ldots,\frac{\partial \mathrm{e}^{-\mathrm{j}\omega t_N(\mathbf{X})}}{\partial\mathbf{x}}\right].\end{aligned}$$

For a single microphone and frequency, the differential with respect to location $\mathbf{x}$ is given by

$$\frac{\partial A(\omega)S(\omega)\mathrm{e}^{-\mathrm{j}\omega t_n(\mathbf{X})}}{\partial\mathbf{x}} = -A(\omega)S(\omega)\mathrm{j}\omega\frac{\partial t_n(\mathbf{x})}{\partial\mathbf{x}}\mathrm{e}^{-\mathrm{j}\omega t_n(\mathbf{X})}, \tag{64}$$

where

$$\frac{\partial}{\partial\mathbf{x}}t_n(\mathbf{x}) = c^{-1}\left(\frac{\mathbf{x}-\mathbf{r}_n}{\|\mathbf{x}-\mathbf{r}_n\|}\right). \tag{65}$$

When assuming independent errors and equal error variances, the Fisher information matrix can be expressed as

$$\begin{aligned}\mathbf{J}(\mathbf{x}) &= \left(\frac{2}{2\pi}\int_{\omega=0}^{\infty}(\omega\|A(\omega)S(\omega)\|)^2\,d\omega\right) \\ &\quad\cdot\left[H_{\mathrm{TOA}}^{\mathrm{T}}\left(\boldsymbol{t}(\mathbf{x})\right)\mathbf{Q}^{-1}H_{\mathrm{TOA}}\left(\boldsymbol{t}(\mathbf{x})\right)\right]\end{aligned} \tag{66}$$

where a design matrix is given for TOAs as

$$H_{\mathrm{TOA}}(\boldsymbol{t}(\mathbf{x})) = \begin{bmatrix} \frac{\partial}{\partial\mathbf{x}}t_1(\mathbf{x}) \\ \frac{\partial}{\partial\mathbf{x}}t_2(\mathbf{x}) \\ \vdots \\ \frac{\partial}{\partial\mathbf{x}}t_N(\mathbf{x}) \end{bmatrix} \tag{67}$$

Moreover, when constant spectral densities for noise and signal are assumed on a certain frequency band B and within some time window of length T, the Fisher information formulates into

$$\begin{aligned}\mathbf{J}(\mathbf{x}) &= \mathrm{SNR}\frac{T}{\pi}\left(B\omega_c^2 + B^3/12\right) \\ &\quad\cdot\left[H_{\mathrm{TOA}}^{\mathrm{T}}(\boldsymbol{t}(\mathbf{x}))H_{\mathrm{TOA}}(\boldsymbol{t}(\mathbf{x}))\right].\end{aligned} \tag{68}$$

7.4. Time difference of arrival based localization

The probability density function for TDOAs is given in equation (30). The Fisher information matrix for TDOA is given by [86, 9]

$$\mathbf{J}(\mathbf{x}) = H_{\mathrm{TDOA}}^{\mathrm{T}}\left(\boldsymbol{\tau}(\mathbf{x})\right)\Sigma_{\mathrm{TDOA}}^{-1}H_{\mathrm{TDOA}}\left(\boldsymbol{\tau}(\mathbf{x})\right), \tag{69}$$

where

$$H_{\mathrm{TDOA}}(\boldsymbol{\tau}(\mathbf{x})) = \begin{bmatrix} \frac{\partial}{\partial\mathbf{x}}\tau_1(\mathbf{x}) \\ \frac{\partial}{\partial\mathbf{x}}\tau_2(\mathbf{x}) \\ \vdots \\ \frac{\partial}{\partial\mathbf{x}}\tau_M(\mathbf{x}) \end{bmatrix} \tag{70}$$

is a matrix including the partial derivatives of equation (7),

$$\frac{\partial}{\partial\mathbf{x}}\tau_m(\mathbf{x}) = c^{-1}\left(\frac{\mathbf{x}-\mathbf{r}_i}{\|\mathbf{x}-\mathbf{r}_i\|} - \frac{\mathbf{x}-\mathbf{r}_j}{\|\mathbf{x}-\mathbf{r}_j\|}\right). \tag{71}$$

The minimum variance that TDOA estimation can achieve is given by equation (59). By assuming independent errors the covariance matrix is given by

$$\Sigma_{\mathrm{TDOA}}^{-1} = \frac{1}{\sigma_{\mathrm{TDOA}}^2}\mathbf{I} = J(\tau)\mathbf{I} \tag{72}$$

which then gives:

$$\mathbf{J}(\mathbf{x}) = J(\tau)H_{\mathrm{TDOA}}^{\mathrm{T}}\left(\boldsymbol{\tau}(\mathbf{x})\right)H_{\mathrm{TDOA}}\left(\boldsymbol{\tau}(\mathbf{x})\right) \tag{73}$$

as the Fisher information matrix for TDOA based localization.

7.5. Time of arrival based localization

The probability density function of the error is given in equation (34). The calculation of CRLB for TOA proceeds as previously for TDOAs. The difference is that the partial derivation in equation (71) for TOAs has the form given in equation (65). The partial derivatives are re-formulated into a matrix, which has the form given in equation (67).

The minimum variance of the TOA estimation given in equation (62). When independent errors are assumed the covariance matrix is given by

$$\Sigma_{\mathrm{TOA}}^{-1} = \frac{1}{\sigma_{\mathrm{TOA}}^2}\mathbf{I} = J(t)\mathbf{I}. \tag{74}$$

The Fisher information matrix for TOA-based localization formulates into

$$\mathbf{J}(\mathbf{x}) = J(t)H_{\mathrm{TOA}}^{\mathrm{T}}\big(\boldsymbol{t}(\mathbf{x})\big)H_{\mathrm{TOA}}\big(\boldsymbol{t}(\mathbf{x})\big) \tag{75}$$

which is identical to equation (68). That is, in theory localization using MLE-TOA, SRP-TOA, or MLE-S function have the same performance.

7.6. Combination of time difference and time of arrival information based localization

When the errors are independent the inverse of the minimum covariance matrix for the combination of TOA and TDOA estimates is given as

$$\begin{aligned}\Sigma_{\mathrm{CM}}^{-1} &= \mathrm{diag}\left(\frac{1}{\sigma_{\mathrm{TOA}}^2},\ldots,\frac{1}{\sigma_{\mathrm{TOA}}^2},\frac{1}{\sigma_{\mathrm{TDOA}}^2},\ldots,\frac{1}{\sigma_{\mathrm{TDOA}}^2}\right)\\ &= \mathrm{diag}(J(t),\ldots,J(t),J(\tau),\ldots,J(\tau))\end{aligned} \tag{76}$$

where the first values are TOA variances and the rest are TDOA variances.

For notational convenience, it is of use to define a vector including both TOA and TDOA values $\chi = [\chi_1, \chi_2, \ldots \chi_{N+M}] = [t_1, t_2, \ldots, t_N, \tau_1, \tau_2, \ldots, \tau_M]$. That is, the first six values of the vector are TOAs and the rest TDOAs. Then the combined design matrix is given as [68]

$$H_{\mathrm{CM}}(\chi(\mathbf{x})) = \begin{bmatrix}\frac{\partial}{\partial\mathbf{x}}\chi_1(\mathbf{x})\\ \frac{\partial}{\partial\mathbf{x}}\chi_2(\mathbf{x})\\ \vdots\\ \frac{\partial}{\partial\mathbf{x}}\chi_{N+M}(\mathbf{x})\end{bmatrix} = \begin{bmatrix}\frac{\partial}{\partial\mathbf{x}}t_1(\mathbf{x})\\ \frac{\partial}{\partial\mathbf{x}}t_2(\mathbf{x})\\ \vdots\\ \frac{\partial}{\partial\mathbf{x}}t_N(\mathbf{x})\\ \frac{\partial}{\partial\mathbf{x}}\tau_1(\mathbf{x})\\ \frac{\partial}{\partial\mathbf{x}}\tau_2(\mathbf{x})\\ \vdots\\ \frac{\partial}{\partial\mathbf{x}}\tau_M(\mathbf{x})\end{bmatrix} \tag{77}$$

The Fisher information matrix is then given by

$$\mathbf{J}(\mathbf{x}) = H_{\mathrm{CM}}^{\mathrm{T}}\big(\chi(\mathbf{x})\big)\Sigma_{\mathrm{CM}}^{-1}H_{\mathrm{CM}}\big(\chi(\mathbf{x})\big). \tag{78}$$

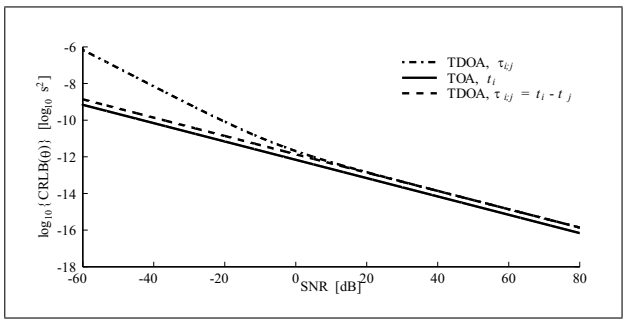

Figure 6. Cramér-Rao lower bound versus signal-to-noise ratio (SNR) for TDOA and TOA.

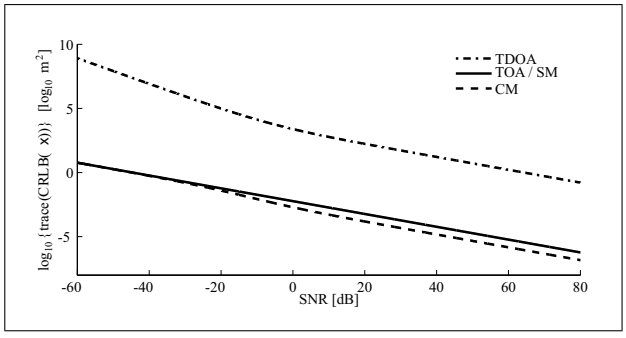

Figure 7. Cramér-Rao lower bound versus signal-to-noise ratio (SNR) for localization using TDOA, TOA, and CM for reflection located at (10.5, 8.2, 2) m.

7.7. Theoretical experiments

7.7.1. Comparison of the performance of the localization methods

In this theoretical comparison, the frequency and the temporal parameters are fixed to $\omega_c/(2\pi) = 12\,\mathrm{kHz}$, $B/(2\pi) = 24\,\mathrm{kHz}$, $T = .004\,\mathrm{s}$. This corresponds to a situation where full bandwidth at 48 kHz sampling frequency and 4 ms time window is used in the analysis. The idea is to compare the localization methods in the same conditions.

Figure 6 presents the CLRB for TOA and TDOA against SNR, calculated using equationsd (62) and (59), respectively. In addition, CRLB for TDOA that is calculated as the difference of two TOA estimates is presented. TOA estimation has smaller CLRB than TDOA estimation. The CRLB of the traditional TDOA estimation approaches the CRLB of the TDOA estimation which is calculated as the difference of two TOAs, as expected from their equations.

In Figure 7, CRLB for TOA, TDOA, and CM, calculated with equations (75), (73) and (78), respectively, are shown with parameters at location (10.5, 8.2, 2) m. As shown in equation (75) and (68) the TOA-based localization and direct localization have the same CRLB. This is noted in Figure 7 with TOA/SM. The microphone array has the same geometry as the one given in Figure 4 with $d_{\mathrm{spc}} = 100\,\mathrm{mm}$ and is located at (0,0,0) m. As mentioned the CRLB for signal model is the same as CRLB for TOA. Clearly, CM has the smallest CRLB and TOA the second smallest. Interestingly, around -25 dB, TOA and CM have

the same performance. This is caused by the increment in the variance of TDOA estimation shown in Figure 6.

7.7.2. Cramér-Rao lower bound versus noise level, distance, and loudspeaker direction

As shown earlier, the CLRB is affected by the SNR. Here the effect of noise level, distance, and the power response of the loudspeaker are studied for a perfectly specular reflection with absorption coefficient of 0 for all frequencies. Here, only the performance of MLE-TOA is considered. The microphone array is located at (0,0,0) m and the reflection is at $(d, 0, 0)$ m from the microphone array where d is the distance. Thus, the reflection in the direction of normal incidence of a surface at the distance of $d/2$ is studied. Air absorption filtering is implemented with a finite impulse response filter with 96 coefficients according to the specifications given in [92]. The sound pressure is set to 1 at 1 meter distance from the center of the microphone array. Moreover, attenuation according to the $1/r$-law is assumed. Then, the signal-to-noise ratio is given as

$$\mathrm{SNR}(\omega) = \frac{G_{s,s}(\omega, \theta_s)\|A_a(\omega)\|^2/d^2}{G_{w,w}(\omega)}, \qquad (79)$$

where $A_a(\omega)$ is the air absorption filter and $G_{s,s}(\omega, \theta_s)$ is the loudspeaker energy response (spectral density) at the direction θ_s. In this example a popular loudspeaker, Genelec 1029A, is studied. The energy response of it is shown in Figure 8. The measurements were taken at every 10 degrees in azimuth angle. When the loudspeaker is facing the measuring microphone the azimuth angle is $\theta_s = 0$. As previously, the sampling frequency is set to 48 kHz, and the time interval to $T = 4$ ms.

The results of the experiment are shown for eight octave bands and for the full band from 0 to 24 kHz in Figure 9a–c. For (a) and (b) the loudspeaker is directly facing the surface, for (b) and (c) the distance is 5 m from the loudspeaker to the reflection, and for (a) and (c) the noise level is −60 dB.

It is evident from Figures 9a–b that as the noise level and distance increase the performance decreases. Since the noise level and the distance directly affect the SNR the result is expected. Moreover, it is noticed that as the center frequency of the octave band increases the performance increases. This increment is caused by the widening of the frequency band, i.e., as the center frequency doubles the frequency band also doubles. The only exception in this behavior is the 16 kHz octave band, and the full band (0–24 kHz). At those frequency bands the average signal energy is lower on average than at 8 kHz octave band, therefore also the performance is weaker. Thus, as shown in equation (60), the Fisher information is integrated over a frequency band; the more the SNR and the wider the frequency band, the better the performance.

As shown in Figure 9c the lack of energy at the directions $\theta_s \in (-180, -50) \cup (50, 180)$, i.e., the back and sides of the loudspeaker, at 1, 2, 4, 8, and 16 kHz octave bands results in the decrement in the performance when compared to the performance at $\theta_s = 0$ direction.

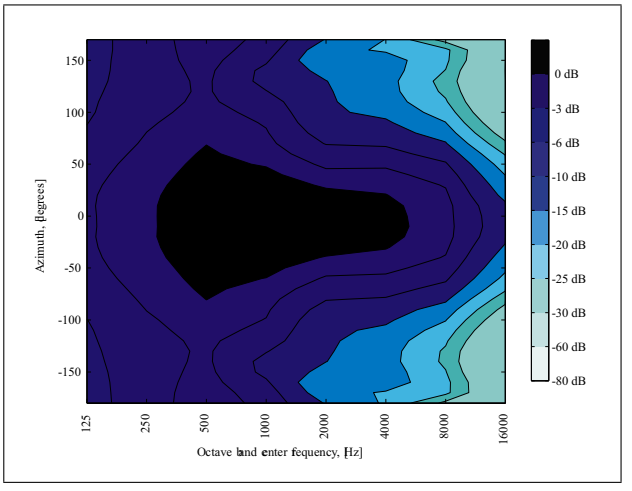

Figure 8. Energy response of Genelec 1029A measured in octave bands at every 10 degrees on the horizontal plane. Linear 2-D interpolation is used for the data in the visualization.

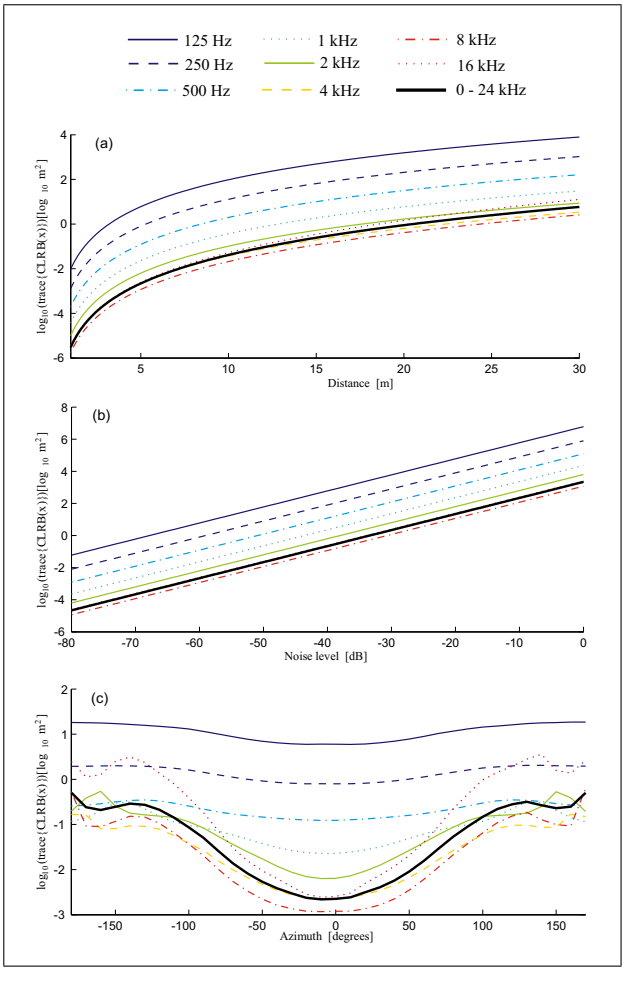

Figure 9. Cramer-Rao lower bound versus (a) distance, (b) noise level, and (c) the direction of the loudspeaker with respect to the normal of the surface. The loudspeaker used in this experiment is Genelec 1029A. As the distance and noise level increase the performance decreases. In addition, when the loudspeaker is directly facing the surface, best performance is achived.

8. Experiments

Simulation and real data experiments are conducted to study the performance of TOA, TDOA, and CM-based localization methods. The CRLB for each estimation task is also presented.

8.1. Error metrics

The performance of the methods is measured here with mean squared error (MSE)

$$\mathrm{MSE}(\hat{\boldsymbol{\theta}}) = \mathrm{E}\left\{\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2\right\}. \tag{80}$$

The other used error metric is the anomaly percentage, which is defined as the ratio between anomalous estimates and total number of estimates, i.e,

$$\mathrm{AN}(\hat{\boldsymbol{\theta}}) = \mathrm{E}\left\{\mathbf{1}\left\{\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\| > \varepsilon\right\}\right\} \tag{81}$$

where $\mathbf{1}\{\cdot\} = 1$ if the condition is true and 0 otherwise.

8.2. Monte-Carlo simulations

The reflection signal model applied in the following Monte-Carlo simulations is of exponential form

$$s_n(t|t_n(\mathbf{x}), \sigma^2) = \mathrm{e}^{-(t-t_n(\mathbf{x}))^2/\sigma^2}. \tag{82}$$

Throughout the simulations the 'variance' parameter of the reflection signal is set to $\sigma = 2/f_s$, where $f_s = 10,000$ Hz is the sampling frequency. The TOA $t_n(\mathbf{x})$ is calculated assuming the spherical wave propagation model. The analysis window length in all the simulations is set to $T = 4$ ms. For simplicity, the attenuation factor is set to $A(\omega) = 1, \forall\omega$ in the simulations.

Since the assumed reflection signal is exponential, the exponential fitting for the TDOA and TOA estimates [24] and for TDOA and TOA estimation functions [25] are applied. As an example, in the case of no noise, the direct cross correlation of two exponential functions is an exponential function. This result is well known for the example with normal distributions.

8.3. Time difference of arrival estimation

Time difference of arrival estimation methods, introduced in Section 4 are compared against signal-to-noise-ratio. The TDOAs are randomized from a uniform distribution between -1 and 1 ms, i.e. $\mathcal{U}(-1, 1)$ ms.

The results of 10,000 Monte-Carlo samples are presented in Figure 10. As expected, the MLE is the most robust against noise having the smallest number of anomalous estimates. ASDF has the smallest number of anomalous estimates when SNR < 20 dB, but this is due to its limitations in the TDOA estimation. That is, the maximum TDOA error with ASDF is half of that of the other methods.

The most accurate method is MLE when SNR <60 dB. When 60 dB < SNR <80 dB CC and ASDF, are the most accurate and when SNR >80 dB, ASDF is the most accurate.

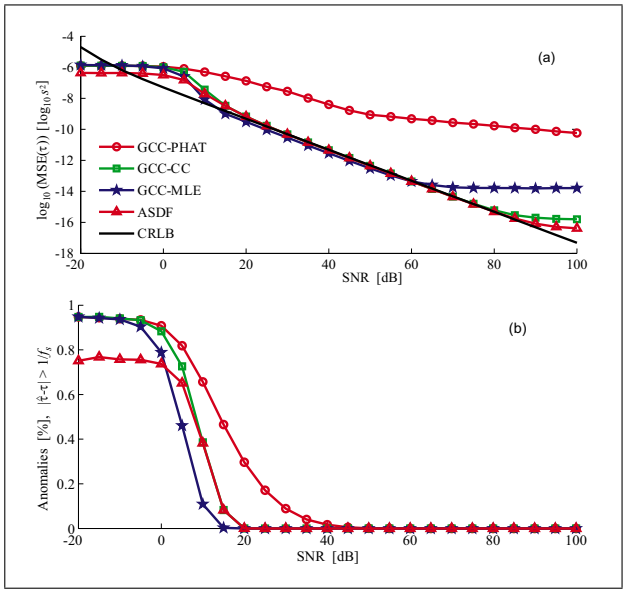

Figure 10. Results for TDOA estimation against signal to noise ratio. (a) Variance, (b) Anomaly.

As shown in Figure 10, ASDF and GCC-CC achieve CRLB when 25 dB < SNR < 75 dB. Moreover, GCC-MLE is lower than the CLRB when 15 dB < SNR < 55 dB. This result indicates that the GCC-MLE TDOA estimation is biased. The bias is a result of the exponential fitting. With very high SNR values the CRLB does not predict the MSE of the methods. This behavior was also noticed in [20]. The reason for this behavior is the truncated window size [20]. The two different windows include two different peaks that have different samples [20]. True zero delay value can therefore only be achieved with auto-correlation and zero noise level.

Direct cross correlation (CC) is the most reasonable selection for TDOA estimation for reflection localization since it does not require a priori information about the noise as MLE. Moreover, CC performs well when compared to the other methods and in addition the calculation of it is straightforward and computationally light. For these reasons CC is used as the TDOA estimation method in the following experiments.

8.4. Time of arrival estimation

Time of arrival estimation methods, introduced in Section 5 are tested against signal-to-noise-ratio. The TOAs are randomized from a uniform distribution between -1 and 1 ms, i.e. $\mathcal{U}(-1, 1)$ ms.

The results of 10,000 Monte-Carlo samples are presented in Figure 11. The simple peak picking method is noted with $\arg\max\{h(t)\}$ in the results of Figure 11. ASDF and CC are the most accurate methods for the TOA estimation. MLE is the most robust against noise, but loses accuracy, due to the fact that the exponential fit does not describe the MLE function shape. The peak picking method, that does not require any a priori knowledge, performs in general better than PHAT and has smaller variance than MLE when SNR > 20 dB. As in TDOA estima-

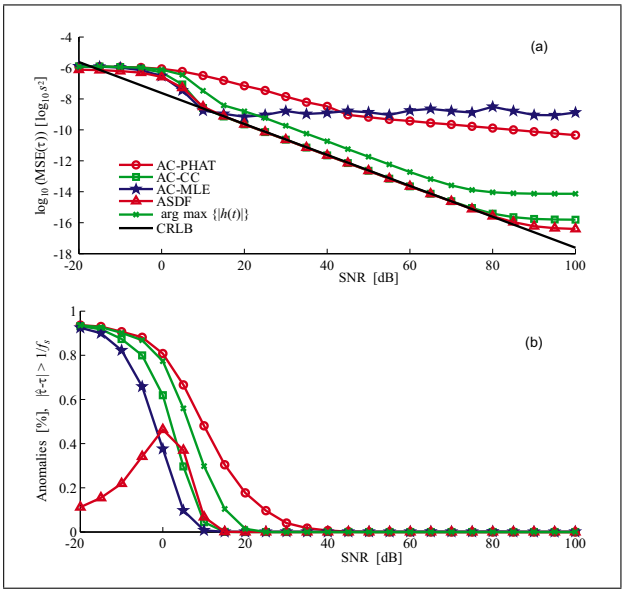

Figure 11. Results for TOA estimation against signal to noise ratio. (a) Variance, (b) Anomaly.

tion, also here the maximum TOA errors for ASDF are half the of the maximum error of the other methods.

As shown in Figure 11, ASDF and AC-CC achieve CRLB when 15 dB < SNR < 75 dB. When SNR < 15 dB, the estimation is saturated as the large number of anomalies suggests. As with the TDOA estimation, also here the MSE of the methods does not achieve CRLB with very high SNR values. The explanation for this behavior is the same as earlier for TDOA estimation.

The TOA estimation with GCC-MLE is not realistic since it would require the knowledge of both source and noise signals. Since the peak picking method is the only blind method, i.e. it does not require the knowledge on the source or noise signal, and has a performance that is comparable to the other methods, it is the most reasonable choice in the general case for the estimation of TOAs. For these reasons, it is used as the TOA estimation method in the following experiments.

8.5. Localization

The reflection location is drawn 1,000 times from a 3-D uniform distribution between −20 and 20 m, i.e.

$$x \sim \mathcal{U}(-20, 20)\,\text{m}, \quad y \sim \mathcal{U}(-20, 20)\,\text{m},$$
$$z \sim \mathcal{U}(-20, 20)\,\text{m}.$$

The microphone array is set to the origin (0,0,0) and the reflection signal is windowed with 4 ms time window around the TOA between the reflection location and the origin.

Nine different localization methods are tested. In detail, SRP, MLE, and PL with TOA, TDOA, and CM data are used for localization of reflections. The formulation for the methods is given in section 5. Direct cross correlation and direct peak picking methods with exponential fitting provided in sections 3 and 4 are used for TDOA and TOA estimation, respectively. Since MLE-S will lead to the same localization result as SRP-TDOA, as shown in [10], it is not tested here.

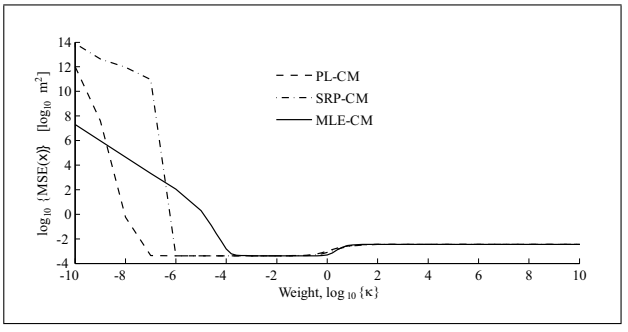

Figure 12. Optimization results against weighting parameter κ with signal-to-noise ratio of 60 dB and with 1,000 Monte-Carlo Samples for each SNR condition.

The reflection signal model is the one presented in equation (82). The location is searched from the localization function using the Nelder-Mead simplex method implemented in MATLAB function `fminsearch`. The initial location value for the optimization method is set to the vicinity of the true location.

8.5.1. Optimization of the parameters

The weighting parameters for the combined methods are optimized. The question is, which weight produces the best result for each method? For MLE the weighting factor κ is defined as the relation between the TOA and TDOA variance, as

$$\kappa = \frac{\sigma_{\text{TDOA}}}{\sigma_{\text{TOA}}} \tag{83}$$

This selection sets the following limitations [12]:

$$\lim_{\kappa\to\infty} P_{\text{MLE-CM}}(\mathbf{x}) = P_{\text{MLE-TOA}}(\mathbf{x}), \tag{84}$$
$$\lim_{\kappa\to 0} P_{\text{MLE-CM}}(\mathbf{x}) = P_{\text{MLE-TDOA}}(\mathbf{x}). \tag{85}$$

For PL-CM and SRP-CM the weighting is limited to $0 < W < 1$. This gives the following obvious limits for SRP-CM function

$$\lim_{W\to 0} P_{\text{SRP-CM}}(\mathbf{x}) = P_{\text{SRP-TOA}}(\mathbf{x}), \tag{86}$$
$$\lim_{W\to 1} P_{\text{SRP-CM}}(\mathbf{x}) = P_{\text{SRP-TDOA}}(\mathbf{x}). \tag{87}$$

and for PL-CM

$$\lim_{W\to 0} \lambda_{\text{PL-CM}}(\mathbf{x}) = \log\{P_{\text{PL-TOA}}(\mathbf{x})/N\}, \tag{88}$$
$$\lim_{W\to 1} \lambda_{\text{PL-CM}}(\mathbf{x}) = \log\{P_{\text{PL-TDOA}}(\mathbf{x})/M\}. \tag{89}$$

The weight factor κ is given the values according to $\log_{10}\{\kappa\} = -10, \ldots, 10$. For MLE-CM the variance of the TOA error is set to $\sigma^2_{\text{TOA}} = 1$, and the variance of the TDOA error is altered as $\sigma_{\text{TDOA}} = \kappa\sigma_{\text{TOA}}$. The weight for SRP-CM and PL-CM is $W \in (0, 1)$, and it is calculated through κ as $W = 1/(10^{\kappa} + 1)$.

The results of this experiment are shown in Figure 12. All the combined methods achieve equally good performance with some weighting.

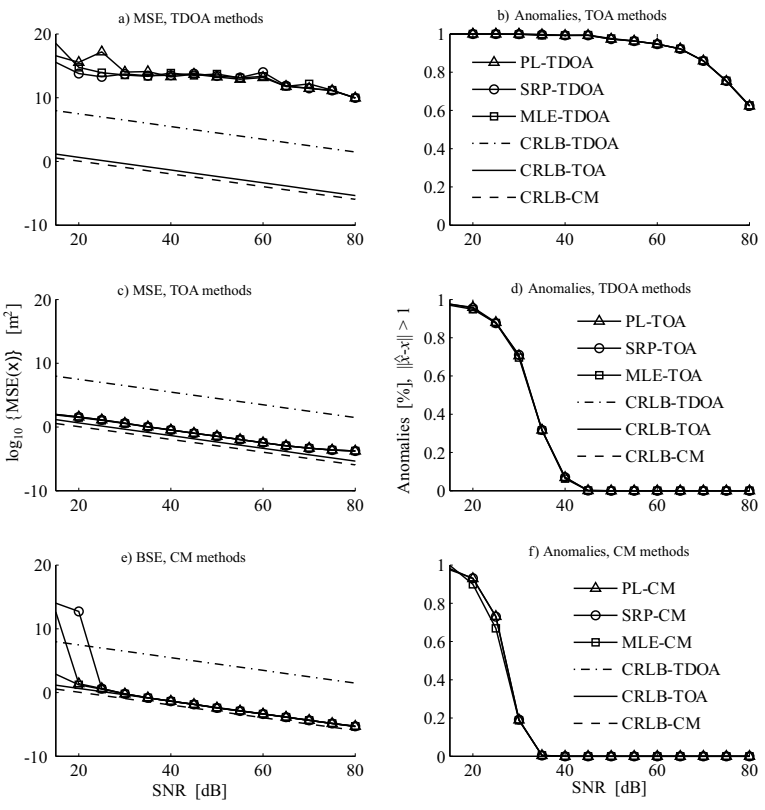

Figure 13. Results for localization against signal-to-noise ratio (SNR) from 1,000 Monte-Carlo samples. MSE and percentage of anomalies are presented for TOA-based methods in a) and b), for TDOA-based methods in c) and d), and for CM in e) and f). In each subfigure the CRLB for TOA, TDOA, and CM is presented. In overall, the CM methods have the best performance.

As shown in Figure 12, the optimal weight for SRP-CM is $W \in (1/(10^{-6}+1), 1/(10^{-1.2}+1))$, for PL-CM $W \in (1/(10^{-7}+1), 1/(10^{-1.4}+1))$, and for MLE-CM $\log_{10}\{\kappa\} \in (-3.4, -0.2)$. A reasonable choice for MLE-CM weighting factor is $\log_{10}(\kappa) = -2$ since it is close to the middle region of the optimal values. For MLE-CM this means that TOA variance is about 100 times the TDOA variance. As shown in Figure 12, for SRP-CM, and PL-CM the value $W = 1/(10^{-2}+1) = 0.99$ for the weight is a reasonable choice. This means that SRP-TOA has a weight $W = .01$. and the SRP-TDOA has the weight $W = 0.99$. The same applies for PL-TOA and PL-TDOA in the PL-CM method. These optimized values are used in the following experiments.

8.5.2. Simulation results for localization methods

Results from simulations with the localization methods are presented in Figure 13a–f. As can be seen from Figure 13, and as expected from the theoretical point of view, the CM-based methods have smaller MSE than TOA or TDOA-based methods. In overall, MLE-CM has the smallest and SRP-CM the highest MSE, out of the combined data methods. At 15 dB MLE-CM has smaller MSE than SRP-CM or PL-CM. TOA based methods have clearly smaller MSE than TDOA based methods. MLE-TOA has the smallest MSE and SRP-TOA the highest out of the TOA-based methods. As shown in Figure 13, the methods achieve CRLB for TOA but not the CRLB for CM. As with TOA and TDOA estimation, the CRLB is best achieved when 15 dB < SNR < 75 dB.

8.5.3. Discussion

The reason why, for example, MLE-CM performs better than MLE-TOA or MLE-TDOA in the experiments is that the TOA and TDOA errors do not correlate with the selected TOA and TDOA estimation methods. Then combining these two pieces information, TOA and TDOA estimates, increases the overall information. If the selected TOA estimation method would be AC-CC, ASDF, or AC-MLE, MLE-TOA would also achieve the CLRB for localization with TOA. The aforementioned TOA estimation methods require knowledge of the source and noise signal. If these pieces of information are available, then it is beneficial to use MLE-TOA with AC-CC, since it is the computationally most efficient method and achieves the CRLB. As mentioned, this article focuses on the blind estimation methods, where it is assumed that the source and the noise signals are unknown.

If MLE-CM would be used with AC-CC and, say, GCC-CC, the performance would be the same as with MLE-TOA with AC-CC. This is due to the fact that the errors of AC-CC and GCC-CC estimations correlate highly. Therefore, the theoretical CRLB for CM can never be achieved in practical situations with the current setup, unless the measurement for TOA estimation and TDOA estimation is done separately. The maximum Fisher information available in six microphones is already used in theory in CRLB for localization with TOA. Nevertheless, the theoretical CRLB for CM is a useful tool for investigating the performance when the TOA and TDOA errors do not correlate.

8.6. Real data experiments

Real data experiments were conducted in two different enclosures, a class room, and an auditorium to study the performance in real conditions. The methods that are solely based on TDOA data are not tested here due to their poor performance in the simulations.

8.6.1. Test setups

The first experiment was conducted in a shoebox-shaped class room (7.09 m×9.35 m×3.76 m) stripped of chairs and tables. A skeleton model of the room is shown in Figure 14. As illustrated in Figure 14, there was a closet on the west wall (WW), an extrusion, a window, and a door on the south wall (SW), and a whiteboard and a door on the east wall (EW). The walls of the room are of painted sheet rock and the floor (FL) is concrete covered with a plastic mat. These materials have a reasonably low absorption coefficient and it is expected that they produce clearly identifiable reflections to the impulse responses. However, lamps, ventilation, and other equipment typical for a modern class room are hanging from the ceiling (CE), which reduce the number of localizable reflections from the ceiling.

The loudspeaker was located in the back and the microphone array in front of the room. The experiment was repeated four times with different locations for the loudspeaker and the microphone array. The microphone array was the G.R.A.S microphone array, introduced in section 2.2, with $d_{\text{spc}} = 100\,\text{mm}$ spacing. Moreover, to have different conditions for the reflections, the loudspeaker was rotated around its z-axis between every 10 degrees from 0 to 360 degrees. The height of the loudspeaker and array was from 1 m to 1.5 m. For each different condition ($36 \times 2 \times 2$) six reflections and the direct sound are localized leading to a total of 1008 location estimates for each of the localization methods.

The second experiment is done in an auditorium. The floor plan of the auditorium is shown in Figure 15. The auditorium has a volume of $1800\,\text{m}^3$, with an inclination of about 10° in the audience area. The loudspeakers were located in 4 different positions at 1.5 m height. The microphone array used in this experiment was a custom made microphone array (TKK-3D, see [93] for details), with 6 electret microphones. The same dimensions, i.e., geometrical setup was used for this array as for the G.R.A.S array. The microphone array, was located every 20 cm in 45 different positions, as shown in Figure 15.

Four reflections, two side wall reflections, one floor, and one ceiling reflection, are localized from the data measured in the auditorium. These surfaces can be considered to be reasonably flat and rigid. The floor is made of concrete and is covered with hard-wood, the ceiling is painted concrete and the walls are wood material, which has circular holes of size 5 mm between every 2 cm. The ceiling of the auditorium is flat at the stage area. In the audience area the ceiling is made of reflectors that are about 2 m by 13 m in area that are in a saw-waveform-shaped arrangement.

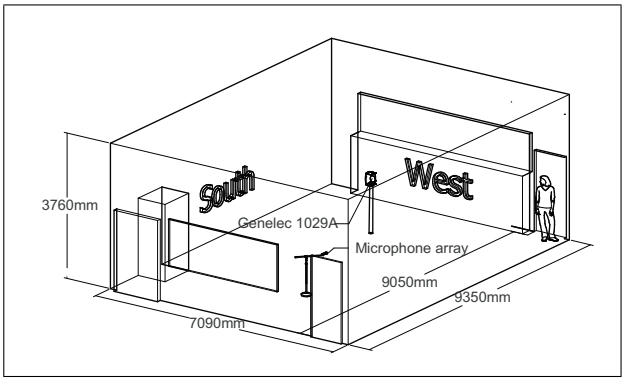

Figure 14. Experimental setup in the class room.

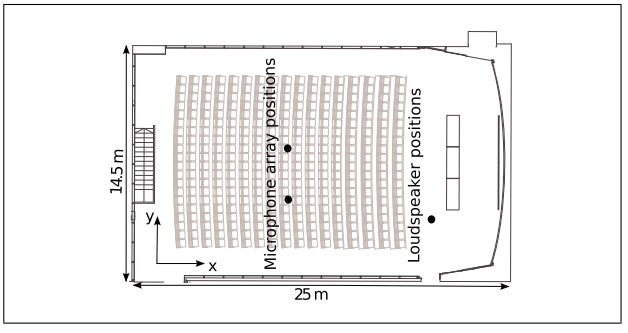

Figure 15. Experimental setup in the auditorium.

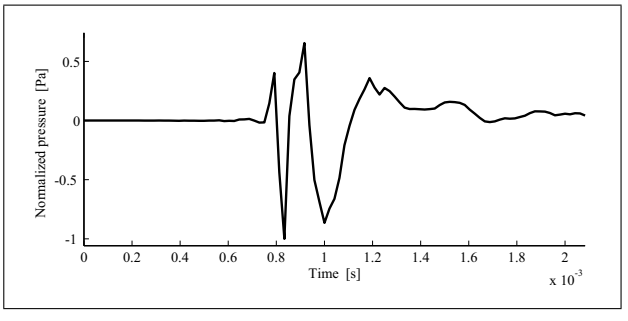

Figure 16. Time domain impulse response of Genelec 1029A at 5 m in front of the loudspeaker. A typical impulse response of Genelec 1029A has two peaks.

For each condition (4×45), the direct sound and four reflections are localized leading to 900 estimations for each method.

In all the experiments, the impulse responses were measured with the sine-sweep technique from 40 Hz to 24 kHz at 48 kHz [94]. Moreover, the loudspeaker used in all the experiments was of type Genelec 1029A. Some parameters describing the experimental setups in more detail are shown in Table III. The speed of sound was estimated in both experiments based on temperature and humidity and the analysis window size was set to $T = 1.3\,\text{ms}$.

8.6.2. Compensation of the loudspeaker impulse response

An impulse response of the loudspeaker used in the experiments, Genelec 1029A, measured directly in front of the loudspeaker at 5 m, is shown in Figure 16. This two-

Table III. Conditions, parameters, and some quantities used in different experimental setups. The reverberation time is calculated for wide band from 500 Hz to 24 kHz. Gen.: Genelec.

Parameter	Classroom	Auditorium
Speed of sound c [m/s]	345.2	344.6
Volume [m^3]	250	1800
Analysis window size [ms]	1.3	1.3
Sampling frequency [kHz]	48	48
Microphone array	G.R.A.S	TKK-3D
Loudspeaker	Gen. 1029A	Gen. 1029A
No. of measurements	144	180
No. of localized reflections	1008	900
Reverberation time [s]	1.4	1.2

peaked impulse response is typical for the loudspeaker used in the experiments and the shape of the impulse response stays similar to directions ±50 degrees from the center plane of the loudspeaker. The two peaked impulse response of Genelec 1029A is caused by two issues. Firstly, the loudspeaker consists of two elements that are separated by approximately 10 cm. This causes some differences in the delays for low and high frequencies, depending on the direction of the loudspeaker with respect to the microphone. Secondly, the low-frequency-element of the loudspeaker has a higher mass, and it does not respond to the voltage changes in the coil as quickly as the tweeter, thus causing the low frequencies to be delayed. Due to the two-peaked impulse responses, the reflections do not introduce sharp peaks in the localization function, and the intersection of the spheres is "blurred".

The fact that the impulse response consists of more than one peak, makes localization using this particular loudspeaker difficult, since in additive noise either of the peaks might be the absolute maximum. Therefore, the shape of the impulse response is compensated for. The compensation is done by deconvolving the measured room impulse response with the loudspeaker impulse response measured in front of the loudspeaker. This is an intermediate compensation for the loudspeaker impulse response. With this approach at least the reflections that are in the most prominent direction of the loudspeaker should be more identifiable. The compensation of the loudspeaker impulse response to all directions would require the knowledge of the orientation of the loudspeaker and the directions of each arriving sound wave. If this can be done, then one can use the MLE-TOA with AC-CC, ASDF, or AC-MLE TOA estimation methods, for localizing the reflections, since they take into account the impulse response of the source.

8.6.3. Results

The results of the experiments are shown in Table IV and in Table V. Two figures are presented in the results, the number of non-anomalous estimates (K) and the root of the mean squared error (RMSE) of non-anomalous estimates.

In the class room experiment, the reflections from the floor, which produces shortest paths from the loudspeaker to the microphone array and has a low absorption coefficient, is the easiest to localize. Other walls have similar absorption coefficients and they are on average at approximately equal distance from the array. Therefore the localization of the reflections from them has similar performance. The pieces of equipment hanging from the ceiling make the localization of the reflections from it more difficult than other surfaces in the enclosure.

In the auditorium experiment, the reflections from the closest surface, the north wall are the easiest to localize. Although the south wall is made of same material as north wall, the reflections from it are more difficult to localize since the path length via it is longer than that of the north wall. The reason why the floor of the auditorium is more difficult to localize than the walls is that the chairs of the audience area obstructing parts of the direct path from the floor reflection to the microphones. The localization of the reflections from the ceiling is made difficult by the reflectors that might diffract some parts of the sound, and therefore obstruct a direct reflection from the ceiling.

As can be seen from the results, in favorable conditions, i.e., for the direct sound and the for the first reflections PL-CM and PL-TOA have the most non-anomalous estimates and are the most accurate. That is, they are the most robust methods in good conditions. However, SRP-CM performs the best in overall in both of the experiments. PL-CM does not have the same advantage over PL-TOA as SRP-CM has over SRP-TOA. This can be explained with the differences between PL-CM and SRP-CM. PL-CM, even if only one TOA or TDOA estimation function fails, i.e. produces likelihood that is close to 0 to the true location, the whole estimation fails. In SRP-CM, this kind of failure in the TOA or TDOA estimation functions does not corrupt the whole localization function. In other words, SRP-CM is more robust against the individual errors in the TOA or TDOA estimation than PL-CM.

MLE-TOA and MLE-CM have the worst performance in the real experiments in overall. These methods only use the maximum of the TOA and TDOA estimation functions for localization. Since the other methods use the whole TOA and TDOA estimation functions for localization, it seems that they also include some additional information. Moreover, since only the maximum is selected in the MLE-TOA and MLE-TDOA from the TOA and TDOA estimation functions it is possible that the wrong maximum is selected.

8.6.4. Discussion

One probable reason for the errors in the localization with all the methods is the fact that there are other acoustic phenomena, e.g., diffraction from the chairs in the auditorium, disturbing the localization.

The case where there are more than one reflection present within one analysis window was not studied in this article. The situation where the reflections arrive during the same analysis window occurs for example when the microphone array and the loudspeaker are placed on the center longitudinal section of the room. Since the microphones of the microphone array are in spatially different

Table IV. Results from the class room experiments. K is the number of estimates that have RMSE < 1 m. SRP-CM performs the best in overall. DS: Direct sound, FL: Floor reflection, SW: South wall reflection, EW: East wall reflection, WW: West wall reflection, NW: North wall reflection, CE: Ceiling reflection.

	Measure	DS	FL	SW	EW	WW	NW	CE	Total
PL-TOA	K	142	128	72	64	64	56	28	554
	RMSE [m]	0.06	0.07	0.11	0.08	0.15	0.14	0.43	0.11
PL-CM	K	142	128	72	64	64	56	28	554
	RMSE [m]	0.06	0.07	0.11	0.08	0.15	0.14	0.42	0.11
SRP-TOA	K	142	128	72	63	66	57	31	559
	RMSE [m]	0.06	0.06	0.12	0.07	0.17	0.17	0.40	0.11
SRP-CM	K	143	128	72	63	66	58	34	**564**
	RMSE [m]	0.06	0.06	0.12	0.07	0.17	0.16	0.40	**0.11**
MLE-TOA	K	144	125	82	62	60	53	22	548
	RMSE [m]	0.07	0.05	0.16	0.08	0.13	0.14	0.57	0.11
MLE-CM	K	144	125	82	62	60	53	22	548
	RMSE [m]	0.07	0.05	0.16	0.08	0.13	0.14	0.57	0.11

Table V. Results from the auditorium experiments. K is the number of estimates that have RMSE < 1 m. SRP-CM performs the best in overall. DS: Direct sound, FL: Floor reflection, SW: South wall reflection, NW: North wall reflection, CE: Ceiling reflection.

	Measure	DS	NW	SW	FL	CE	Total
PL-TOA	K	180	177	125	90	58	630
	RMSE [m]	0.19	0.33	0.48	0.35	0.47	0.34
PL-CM	K	180	176	125	91	57	629
	RMSE [m]	0.19	0.33	0.48	0.35	0.46	0.34
SRP-TOA	K	180	172	130	91	60	633
	RMSE [m]	0.19	0.32	0.48	0.36	0.42	0.33
SRP-CM	K	180	171	132	98	68	**649**
	RMSE [m]	0.19	0.32	0.49	0.38	0.42	**0.34**
MLE-TOA	K	180	168	117	85	33	583
	RMSE [m]	0.19	0.33	0.46	0.36	0.40	0.32
MLE-CM	K	180	168	117	85	33	583
	RMSE [m]	0.19	0.33	0.46	0.36	0.40	0.32

locations the different reflections arrive to them at different time. In theory and practice, the localization of two reflections that arrive within the same analysis window is the same problem as the multi-source localization problem. Some of the methods used for multi-source localization problem, e.g. the one presented in [95], are applicable also for the multi-reflection localization problem. The MLE method presented in this article can not be directly applied for the multi-reflection localization problem. However, the SRP and PL methods are directly applicable. Therefore in the future work the PL and SRP methods are preferred.

Since the SRP-CM adds the squared impulse responses and TDOA estimation functions it is possible that the true reflection location gets less evidence than a "ghost" or a competing reflection. This behavior is recognized in the speech source localization [8]. However, the problem was not present in the experiments of this paper.

One reason for the anomalous estimates with all the methods is that the arriving sound wave from the direction of the reflection is not as "impulse-like" as the sound wave in front of the loudspeaker. Thus, the magnitude of the emitted sound wave in the direction of the reflections is lower, and does not contain as much high frequency energy as the impulse in front of the loudspeaker.

One possibility to get around the problems related to the loudspeaker non-idealities is to use only the phase information of the signal. However, this decreases the SNR in the frequencies that have a low magnitude and as a result decreases the performance, as seen in the simulations with PHAT which uses only the phase information.

Another possibility to obtain more accurate TOA information is to measure the impulse response of the loudspeaker to a grid of directions in free-field conditions. Then the impulse response of the loudspeaker can be compensated for in the room impulse response by deconvolving the reflection with the corresponding direction free-field impulse response. This however would require a large data set of a priori measurements of the loudspeaker and the knowledge of the locations of the reflections.

As mentioned, TOA estimation can be also improved by applying the sparse impulse response technique presented in [67]. The higher the directivity of the loudspeaker is, the better the TOA estimation can be achieved with the sparse impulse response technique.

Summarizing the above notes on the loudspeaker directionality and impulse responses, there is a requirement for a loudspeaker or a sound source that can produce close to perfect dirac impulses. Preferably, the loudspeaker should also be perfectly omni-directional or then have a infinitely narrow directivity.

9. Conclusions

Localization of reflections is a pre-task for example in estimation of the geometry of the enclosure or the acoustic properties of the reflective surfaces. This paper presented a framework for localizing reflections and several methods were compared with theoretical, simulation, and real data experiments.

Three different localization functions for TDOA, and TOA were studied. In addition, it was proposed that by combining both TDOA and TOA information a better localization result can be achieved. For the proposed combination also three localization functions were tested. This leads to a total of nine different localization functions.

It was shown by theoretical developments that it is advantageous to use both TOA and TDOA information to localize the reflection. Simulations also verified this finding. Moreover, out of the three localization methods, MLE, SRP, and PL, the MLE method that combines the TOA and TDOA information were found to achieve the best performance. The TDOA-based methods had the worst performance in the simulations.

Real data experiments included localization of reflections in two enclosures. The real data experiments verify the observations of the theoretical development and simulation results. When the TOA and TDOA information are combined, the best performance is achieved. Out of the methods that combine TOA and TDOA information, the worst performance in the real situations was with the MLE method. This was found to be due to the fact that in the MLE methods only the maximum of the TOA and TDOA estimation functions are used, whereas in other methods the whole estimation functions are considered. The best performance was with the SRP method that combine the TOA and TDOA information.

In some applications, computational efficiency might be a more important aspect than robustness against noise. Then the use SRP-TOA is recommended since it is computationally lighter than SRP-CM, and the performance of it is only slightly worse than that of SRP-CM.

Future work includes integrating the localization methods with the sparse impulse response measurement technique [67] to estimate the locations of the reflective surfaces. In addition, the estimation of the absorption coefficient using the introduced methodology and the localization of the reflective surfaces in all kind of rooms is left as the future work.

Acknowledgement

The authors are grateful to the anonymous reviewers for their valuable comments and to Mr. Philip Robinson for proofreading the manuscript.

The research leading to these results has received funding from the Academy of Finland, project nos. [218238 and 140786], the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no. [203636], and Helsinki Graduate School in Computer Science and Engineering. In addition Nokia Foundation is thanked for the promotion.

References

[1] D. Aprea, F. Antonacci, A. Sarti, S. Tubaro: Acoustic reconstruction of the geometry of an environment through acquisition of a controlled emission. 17th European Signal Processing Conference, 2009, 710–714.

[2] F. Antonacci, A. Sarti, S. Tubaro: Geometric reconstruction of the environment from its response to multiple acoustic emissions. IEEE International Conference on Acoustics, Speech and Signal Processing, 2010, 2822–2825.

[3] B. Günel: Room shape and size estimation using directional impulse response measurements. 3rd EAA Congress on Acoustics, Forum Acusticum, 2002.

[4] E. Mommertz: Angle-dependent in-situ measurements of reflection coefficients using a subtraction technique. Applied Acoustics **46** (1995) 251–263.

[5] C. Nocke: In-situ acoustic impedance measurement using a free-field transfer function method. Applied Acoustics **59** (2000) 253–264.

[6] V. C. Raykar, I. V. Kozintsev, R. Lienhart: Position calibration of microphones and loudspeakers in distributed computing platforms. IEEE Transactions on Speech and Audio Processing **13** (2005) 70–83.

[7] M. Omologo, P. Svaizer, R. De Mori: Spoken dialogues with computers (ed. e. de mori), chapter acoustic transduction, page 61. Academic Press, London, UK, 1998.

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

[9] Y. T. Chan, K. C. Ho: A simple and efficient estimator for hyperbolic location. IEEE Transactions on Signal Processing **42** (1994) 1905–1915.

[10] J. M. Valin, F. Michaud, J. Rouat: Robust localization and tracking of simultaneous moving sound sources using beamforming and particle filtering. Robotics and Autonomous Systems **55** (2007) 216–228.

[11] K. W. Cheung, H. C. So, W. K. Ma, Y. T. Chan: Least squares algorithms for time-of-arrival-based mobile location. IEEE Transactions on Signal Processing **52** (2004) 1121–1130.

[12] S. Tervo, T. Lokki, L. Savioja: Maximum likelihood estimation of loudspeaker locations from room impulse responses. Journal of the Audio Engineering Society **59** (2011) 845–857.

[13] R. I. Reza: Data fusion for improved TOA/TDOA position determination in wireless systems. M.Sc. Thesis, 2000.

[14] I. Ziskind, M. Wax: Maximum likelihood localization of multiple sources by alternating projection. IEEE Transactions on Acoustics, Speech and Signal Processing **36** (1988) 1553–1560.

[15] J. C. Chen, R. E. Hudson, K. Yao: A maximum-likelihood parametric approach to source localizations. IEEE International Conference on Acoustics, Speech and Signal Processing **5** (2001) 3013–3016.

[16] J. Benesty: Adaptive eigenvalue decomposition algorithm for passive acoustic source localization. The Journal of the Acoustical Society of America **107** (2000) 384–391.

[17] M. S. Brandstein, H. F. Silverman: A robust method for speech signal time-delay estimation in reverberant rooms. IEEE International Conference on Acoustics, Speech and Signal Processing **1** (1997) 375–378.

[18] J. Chen, J. Benesty, Y. Huang: Performance of gcc-and amdf-based time-delay estimation in practical reverberant environments. EURASIP Journal on Applied Signal Processing (2005) 25–36.

[19] J. Chen, J. Benesty, Y. Huang: Time delay estimation in room acoustic environments: an overview. EURASIP Journal on Applied Signal Processing (2006) Article ID 26503.

[20] G. Jacovitti, G. Scarano: Discrete time techniques for time delay estimation. IEEE Transactions on Signal Processing **41** (1993) 525–533.

[21] C. Knapp, G. Carter: The generalized correlation method for estimation of time delay. IEEE Transactions on Acoustics, Speech and Signal Processing **24** (1976) 320–327.

[22] E. A. Lehmann: Particle filtering approach to adaptive time-delay estimation. IEEE International Conference on Acoustics, Speech and Signal Processing **4** (2006) 1129–1132.

[23] X. Lai, H. Torp: Interpolation methods for time delay using cross-correlation for blood velocity measurement. IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control **46** (1999) 277–290.

[24] L. Zhang, X. Wu: On cross correlation based discrete time delay estimation. IEEE International Conference on Acoustics, Speech and Signal Processing **IV** (2005) 981–984.

[25] S. Tervo, T. Lokki: Interpolation methods for the SRP-PHAT algorithm. Proceedings of the 11th International Workshop for Acoustic Echo and Noise Control (IWAENC 2008), Seattle, WA, USA, September 14-17 2008, Article ID 9037.

[26] S. M. Kay: Fundamentals of statistical signal processing. Volume 1: Estimation theory. Prentice-Hall, New Jersey, USA, 1998.

[27] J. Abel, J. Smith: The spherical interpolation method for closed-form passive source localization using range difference measurements. IEEE International Conference on Acoustics, Speech and Signal Processing **12** (1987) 471–474.

[28] Y. Huang, J. Benesty, G. W. Elko, R. M. Mersereati: Real-time passive source localization: A practical linear-correction least-squares approach. IEEE Transactions on Speech and Audio Processing **9** (2001) 943–956.

[29] K. W. K. Lui, J. Zheng, H. C. So: Particle swarm optimization for time-difference-of-arrival based localization. European Signal Processing Conference, 2007, 414–417.

[30] C. W. Reed, R. Hudson, K. Yao: Direct joint source localization and propagation speed estimation. IEEE International Conference on Acoustics, Speech and Signal Processing **3** (1999) 1169–1172.

[31] H. Schau, A. Robinson: Passive source localization employing intersecting spherical surfaces from time-of-arrival differences. IEEE Transactions on Acoustics, Speech and Signal Processing **35** (1987) 1223–1225.

[32] J. Smith, J. Abel: Closed-form least-squares source location estimation from range-difference measurements. IEEE Transactions on Acoustics, Speech and Signal Processing **35** (1987) 1661–1669.

[33] H. C. So, S. P. Hui: Constrained location algorithm using TDOA measurements. IEICE Transactions on Fundamentals of Electronics Communications and Computer Sciences **86** (2003) 3291–3293.

[34] P. Stoica, J. Li: Lecture notes-source localization from range-difference measurements. IEEE Signal Processing Magazine **23** (2006) 63–66.

[35] M. Walworth, A. Mahajan: 3D position sensing using the difference in the time-of-flights from a wave source to various receivers. 8th International Conference on Advanced Robotics, 1997, 611–616.

[36] P. Aarabi: The fusion of distributed microphone arrays for sound localization. EURASIP Journal on Applied Signal Processing (2003) 338–347.

[37] A. Johansson, G. Cook, S. Nordholm: Acoustic direction of arrival estimation, a comparison between Root-MUSIC and SRP-PHAT. TENCON, IEEE Region 10, volume B, 629-632, 2004.

[38] B. Mungamuru, P. Aarabi: Enhanced sound localization. IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics **34** (2004) 1526–1540.

[39] J. M. Peterson, C. Kyriakakis: Hybrid algorithm for robust, real-time source localization in reverberant environments. IEEE International Conference on Acoustics, Speech and Signal Processing **4** (2005) iv/1053–iv/1056.

[40] D. N. Zotkin, R. Duraiswami: Accelerated speech source localization via a hierarchical search of steered response power. IEEE Transactions on Speech and Audio Processing **12** (2004) 499–508.

[41] E. A. Lehmann: Particle filtering methods for acoustic source localisation and tracking. PhD thesis, Australian National University, 2004.

[42] G. Defrance, L. Daudet, J.-D. Polack: Detecting arrivals within room impulse responses using matching pursuit. 11th International Conference on Digital Audio Effects (DAFx-08), 2008, 1–4.

[43] G. Defrance, L. Daudet, J.-D. Polack: Using matching pursuit for estimating mixing time within room impulse responses. Acta Acustica united with Acustica **95** (2009) 1082–1092.

[44] C. Falsi, D. Dardari, L. Mucchi, M. Z. Win: Time of arrival estimation for uwb localizers in realistic environments. EURASIP Journal on Applied Signal Processing (2006) 152–152.

[45] J. Usher: An improved method to determine the onset timings of reflections in an acoustic impulse response. The Journal of the Acoustical Society of America **127** (2010) EL172–EL177.

[46] J. P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, M. B. Sandler: A tutorial on onset detection in music signals. IEEE Transactions on Speech and Audio Processing **13** (2005) 1035–1047.

[47] W. Kim, J. G. Lee, G. I. Jee: The interior-point method for an optimal treatment of bias in trilateration location. IEEE Transactions on Vehicular Technology **55** (2006) 1291–1301.

[48] P. Pertilä, M. Mieskolainen, M. S. Hämäläinen: Closed-form self-localization of asynchronous microphone arrays. Joint Workshop on Hands-free Speech Communication and Microphone Arrays, 2011, 139–144.

[49] E. Xu, Z. Ding, S. Dasgupta: Source localization in wireless sensor networks from signal time-of-arrival measurements. IEEE Transactions on Signal Processing **59** (2011) 2887–2897.

[50] S. Gezici, H. V. Poor: Position estimation via ultra-wide-band signals. Proceedings of the IEEE **97** (2009) 386–403.

[51] G. P. Yost, S. Panchapakesan: Automatic location identification using a hybrid technique. IEEE Vehicular Technology Conference, 1998, volume 1, 264–267.

[52] J. C. Chen, R. E. Hudson, K. Yao: Maximum-likelihood source localization and unknown sensor location estimation for wideband signals in the near-field. IEEE Transactions on Signal Processing **50** (2002) 1843–1854.

[53] B. Mungamuru, P. Aarabi: Joint sound localization and orientation estimation. IEEE International Conference on Information Fusion, 2003, 81–85.

[54] C. Zhang, Z. Zhang, D. Florêncio: Maximum likelihood sound source localization for multiple directional microphones. IEEE International Conference on Acoustics, Speech and Signal Processing **1** (2007) 125–128.

[55] A. O'Donovan, R. Duraiswami, J. Neumann: Microphone arrays as generalized cameras for integrated audio visual processing. IEEE Conference on Computer Vision and Pattern Recognition, 2007, 1–8.

[56] A. O'Donovan, R. Duraiswami, D. Zotkin: Imaging concert hall acoustics using visual and audio cameras. IEEE International Conference on Acoustics, Speech and Signal Processing, 2008, 5284–5287.

[57] A. Farina, A. Amendola, A. Capra, C. Varani: Spatial analysis of room impulse responses captured with a 32-capsules microphone array. 130th Audio Engineering Society Convention, London, 13-16 May 2011, Paper 8400.

[58] E. van Lancker: Localization of reflections in auditoriums using time delay estimation. 108th Audio Engineering Society Convention, 2000, Paper 5168.

[59] J. Merimaa, V. Pulkki: Spatial impulse response rendering i: Analysis and synthesis. The Journal of the Audio Engineering Society **53** (2005) 1115–1127.

[60] J. Merimaa: Analysis, synthesis and perception of spatial sound-binaural localization modeling and multichannel loudspeaker reproduction. PhD thesis, Helsinki University of Technology, 2006.

[61] M. Kuster, D. de Vries, E. M. Hulsebos, A. Gisolf: Acoustic imaging in enclosed spaces: Analysis of room geometry modifications on the impulse response. The Journal of the Acoustical Society of America **116** (2004) 2126–2137.

[62] M. Kuster, D. de Vries: Modelling and order of acoustic transfer functions due to reflections from augmented objects. EURASIP Journal on Advances in Signal Processing (2007) Article ID 30253.

[63] J. Filos, E. A. P. Habets, P. A. Naylor: A two-step approach to blindly infer room geometries. International Workshop on Acoustic Echo and Noise Cancellation, 2010.

[64] A. Canclini, M. R. P. Thomas, A. Antonacci, F. Sarti, P. A. Naylor: Robust inference of room geometry from acoustic impulse responses. 19th European Signal Processing Conference, 2011, 161–165.

[65] E. A. Nastasia, F. Antonacci, A. Sarti, S. Tubaro: Localization of planar acoustic reflections through emission of controlled stimuli. 19th European Signal Processing Conference, 2011, 156–160.

[66] A. Canclini, F. Antonacci, M. R. P. Thomas, J. Filos, A. Sarti, P. A. Naylor, S. Tubaro: Exact localization of acoustic reflectors from quadratic constraints. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2011, 17–20.

[67] S. Tervo, J. Pätynen, T. Lokki: Acoustic reflection path tracing using a highly directional loudspeaker. Proceedings of the 2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 2009), New Paltz, NY, USA, October 18-21, 2009, 245–248.

[68] S. Tervo, T. Korhonen, T. Lokki: Estimation of reflections from impulse responses. Journal of Building Acoustics **18** (2011) 159–174.

[69] D. Ba, F. Ribeiro, C. Zhang, D. Florencio: L1 regularized room modeling with compact microphone arrays. 35th IEEE International Conference Acoustics, Speech and Signal Processing, 2010, 157–160.

[70] T. Korhonen: Acoustic source localization utilizing reflective surfaces. PhD thesis, Tampere University of Technology, 2010.

[71] A. E. O'Donovan, R. Duraiswami, D. N. Zotkin: Automatic matched filter recovery via the audio camera. IEEE International Conference on Acoustics Speech and Signal Processing, 2010, 2826–2829.

[72] A. Canclini, P. Annibale, F. Antonacci, A. Sarti, R. Rabenstein, S. Tubaro: From direction of arrival estimates to localization of planar reflectors in a two dimensional geometry. IEEE International Conference on Acoustics, Speech and Signal Processing, 2011, 2620–2623.

[73] E. Mabande, H. Sun, K. Kowalzcyk, W. Kellermann: On 2D-localization of reflectors using robust beamforming techniques. IEEE International Conference on Acoustics, Speech and Signal Processing, 2011, 153–156.

[74] H. Sun, E. Mabande, K. Kowalczyk, W. Kellermann: Joint DOA and TDOA estimation for 3D localization of reflective surfaces using eigenbeam MVDR and spherical microphone arrays. IEEE International Conference on Acoustics, Speech and Signal Processing, 2011, 113–116.

[75] J. M. Jot, L. Cerveau, O. Warusfel: Analysis and synthesis of room reverberation based on a statistical time-frequency model. 103th Audio Engineering Society Convention, 1997, Paper 4629.

[76] G. Defrance, J.-D. Polack: Estimating the mixing time of concert halls using the extensible Fourier transform. Applied Acoustics **71** (2010) 777–792.

[77] T. Hidaka, Y. Yamada, T. Nakagawa: A new definition of boundary point between early reflections and late reverberation in room impulse responses. The Journal of the Acoustical Society of America **122** (2007) 326–332.

[78] J. B. Allen, D. A. Berkley: Image method for efficiently simulating small-room acoustics. The Journal of the Acoustical Society of America **65** (1979) 943–950.

[79] M. Kuster: Reliability of estimating the room volume from a single room impulse response. The Journal of the Acoustical Society of America **124** (2008) 982–993.

[80] H. Kuttruff: Room acoustics, 4th ed. Spon Press, NY, USA, 2000.

[81] P. Welch: The use of Fast Fourier Transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. IEEE Transactions on Audio and Electroacoustics **15** (1967) 70–73.

[82] G. Carter: Coherence and time delay estimation. Proceedings of the IEEE **75** (1987) 236–255.

[83] P. Robinson, N. Xiang: On the subtraction method for in-situ reflection and diffusion coefficient measurements. Journal of the Acoustical Society of America (2010) EL99–EL104.

[84] R. Stewart, M. Sandler: Statistical measures of early reflections of room impulse responses. 10th International Conference on Digital Audio Effects (DAFx-07), 2007, 59–62.

[85] R. Mucci: A comparison of efficient beamforming algorithms. IEEE Transactions on Acoustics, Speech and Signal Processing **32** (1984) 548–558.

[86] P. Pertilä: Acoustic source localization in a room environment and at moderate distances. PhD thesis, Tampere University of Technology, 2009.

[87] L. G. da Silveira, V. P. Minotto, C. R. Jung, B. Lee: A gpu implementation of the srp-phat sound source localization algorithm. International Workshop on Acoustic Echo and Noise control, 2010, Paper ID 1062.

[88] J. A. Nelder, R. Mead: A simplex method for function minimization. The Computer Journal **7** (1965) 308–313.

[89] B. M. Sadler, R. J. Kozick: A survey of time delay estimation performance bounds. IEEE Workshop on Sensor Array and Multichannel Processing, 2006, 282–288.

[90] E. Weinstein, A. Weiss: Fundamental limitations in passive time-delay estimation. Part II: Wide-band systems. IEEE Transactions on Acoustics, Speech and Signal Processing **32** (1984) 1064–1078.

[91] J. C. Chen, K. Yao, R. E. Hudson: Acoustic source localization and beamforming: theory and practice. EURASIP Journal on Applied Signal Processing (2003) 359–370.

[92] ISO Standard 9613-1: Acoustics - Attenuation of sound during propagation outdoors - Part 1: Calculation of the absorption of souns by the atmosphere. 1993.

[93] T. Peltonen, T. Lokki, B. Gouatarbes, J. Merimaa, M. Karjalainen: A system for multi-channel and binaural room response measurements. 110th Audio Engineering Society Convention, Amsterdam, The Netherlands, May 12-15 2001, preprint no. 5289.

[94] A. Farina: Advancements in impulse response measurements by sine sweeps. 122nd Convention Audio Engineering Society, 2007, Paper 7121.

[95] B. Alessio, O. Maurizio, S. Piergiorgio: Multiple source localization based on acoustic map de-emphasis. EURASIP Journal on Audio, Speech and Music Processing (2010) Article ID 147495.