



# Estimation of Reflections from Impulse Responses

Sakari Tervo<sup>1</sup>, Teemu Korhonen<sup>2</sup>, and Tapio Lokki<sup>1</sup>

(1) Aalto University, School of Science and Technology, Department of Media Technology, P.O. Box 15400, FI-00076 Aalto, Finland

(2) Tampere University of Technology, Department of Signal Processing, P.O. Box 553, FI-33101 Tampere, Finland

**PACS:** 43.55

## ABSTRACT

The characteristics of early reflections have a major effect on the acoustics of concert halls. In this article a framework for automatic localization of reflections and their properties is formulated. The framework uses impulse responses measured with multiple microphones. The focus is on the methods that can be used for detecting reflections and the methods that estimate the direction of arrival. Three methods for both tasks are given and their performance is measured using simulated data. Finally an example in a real auditorium is shown using the most reliable methods for detecting and estimating the direction of arrival of the reflections.

## INTRODUCTION

The early reflections and their properties play an important role in the acoustics of an enclosure. The knowledge of the location of early reflections is of interest for example in auralization and when studying concert hall acoustics. In this article a framework for localizing reflections from measured impulse responses is formulated. The presented methods assume that an impulse response is measured from a source to a microphone array.

The framework proposed in this article for the localization of reflections is three folded. A principle illustration of the framework is presented in Fig. 1.

In the first part of the framework the sound source and the microphones are self-localized or automatically calibrated. That is, the location of the sound source with respect to the microphone array is solved. This can be done for example by applying techniques presented in earlier research [1–3]. In addition, the speed of sound can also be estimated when the distances between microphones in the array are known.

In the second part, which is the main contribution of this article, the reflections are detected and localized. This is done using multiple microphones and applying acoustic source localization techniques. Basically any setup of microphones can be used to localize the direction of arrival. Here an open 3-D spherical six-microphone setup is used. Moreover, any of the well developed acoustic source localization techniques can be used, for an overview of them the reader is referred to e.g. [4]. Here three acoustic source localization methods are tested for estimating the direction of arrival, of which two are based on cross correlation and one of them is based on sound intensity vectors.

In the third part, the reflections are illustrated by mapping them in to the geometry of an enclosure or by for example overlaying them on top of an image as in [5]. In addition, the properties of the surfaces can be estimated up to some extent by using, e.g., simple windowing techniques.

A related topic to the localization of reflections is the localization of the reflective surfaces, or the estimation of room geometry. The problem of estimating the location of reflections or reflective surfaces using microphones and loudspeakers has been addressed by several authors with various approaches [6–14].

Günel as well as Antonacci et al. estimate the room geometry by spanning a microphone around a loudspeaker [7, 8]. The basic principle is to measure impulse responses from the loudspeaker to the microphone with different angles. Then the distances to the closest surfaces are estimated from the impulse responses. Reflective surfaces are found since their relative position changes somewhat periodically with respect to the sound source and the microphone over the measured responses. Developments of the method presented in [8] are given in [12].

Kuster uses acoustic imaging for finding the room geometry and other acoustic properties [9, 10]. It 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 [9, 10].

Tervo and Korhonen propose an inverse mapping of the multipath propagation problem, which is then used together with acoustic source localization to estimate a reflective surface from speech signal in an auditorium [11]. Moreover, Tervo *et al.* use a highly directional loudspeaker and a microphone array to trace the reflections in time and space using ray-tracing inversely [6].

In [13] and [14] 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. SIRR in principle can be used directly in the framework of this article.

The article is organized as follows. In the second section the theory is given. The third section describes the methods used

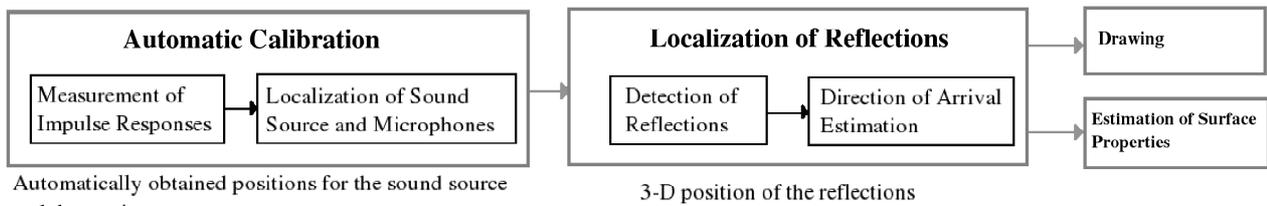


Figure 1: Diagram of the different tasks in automatic localization of reflections or estimation of the room geometry. This article addresses the second part of the diagram, that is the estimation related to the localization of reflections, though some examples of the drawing is also shown.

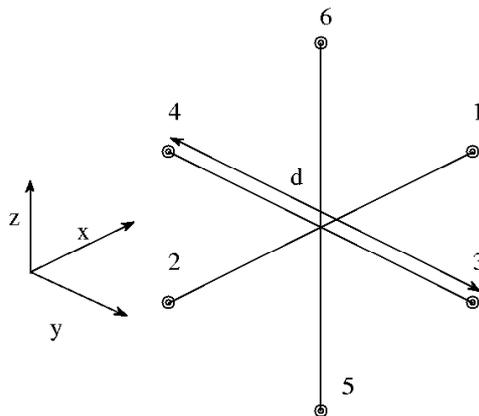


Figure 2: 3-D microphone array suitable for sound intensity estimation. Spacing  $d$  is equal between two microphones on a single axis.

for detecting the reflections and for estimating the direction of arrival. In the fourth section simulations are conducted and results presented. Fifth section shows a practical example from an auditorium using some of the methods introduced in the third section. Sixth section concludes the article with discussion and motivation for future work.

## THEORY

A simplified model of the room impulse used in this article is presented next. In addition, because many methods introduced later on are based on sound intensity vectors, the estimation of them is also presented.

### Impulse Response Model

In a room environment, the sound  $s(t)$  traveling from the sound source to the receiver  $n$  is affected by the impulse response  $h_n(t)$ :

$$p_n(t) = h_n(t) * s(t) + w(t), \quad (1)$$

where  $*$  denotes convolution and  $w(t)$  is measurement noise, independent and identically distributed for each receiver.

In the context of this article, the impulse response can be simplified as

$$h(t) = \sum_{k=1}^K \int \alpha_k(\omega) e^{j\omega(t-\tau_k)} d\omega \quad (2)$$

where  $\alpha_k(\omega)$  is the frequency ( $\omega = 2\pi f$ ) dependent attenuation factor for each reflection  $k$ , and  $\tau_k$  is the time delay related to the distance of the path of the reflection. The attenuation factor is generally known to be dependent on the properties of the surface and air absorption [15].

### Sound Intensity

On a certain axis  $x$ , the sound intensity is given in the frequency domain as

$$I_x(\omega) = \Re\{P^*(\omega)U_x(\omega)\}, \quad (3)$$

where  $P(\omega)$  and  $U_a(\omega)$  are the frequency presentations of the sound pressure and of the particle velocity with angular frequency  $\omega$  [13]. In addition,  $\Re\{\cdot\}$  is the real part of a complex number and  $(\cdot)^*$  denotes the complex conjugate.

The pressure in the middle of the array, shown in Fig. 2, can be estimated as the average pressure of the microphones [13, 16]:

$$P(\omega) \approx \frac{1}{6} \sum_{n=1}^6 P_n(\omega). \quad (4)$$

In the frequency domain the particle velocity is estimated for x-axis as:

$$U_x(\omega) \approx \frac{-j}{\omega\rho_0 d} [P_1(\omega) - P_2(\omega)], \quad (5)$$

where  $d$  is the distance between the two receivers,  $j$  is the imaginary unit, and for example with the speed of sound  $c = 343$  m/s the median density of the air is  $\rho_0 = 1.204$  kg/m<sup>3</sup>.

The sound intensity in (3) is estimated with the approximations in (4) and (5). For obtaining the  $y$  and  $z$ -component of the sound intensity, the microphones 1 and 2 are replaced in (5) with microphones 3 and 4, and 5 and 6, respectively. The overall sound intensity vector for a frequency  $\omega$  is then noted with  $\mathbf{I}(\omega) = [I_x(\omega), I_y(\omega), I_z(\omega)]$ .

Kallinger *et al.* show that the estimation using (4) and (5) leads to a biased direction for the sound intensity vectors [16]. This is due to the fact that the pressure gradient is not constant between the microphones but rather a sinusoidal one. The bias can be corrected up to some frequency. However, here the bias correction is not used since the highest frequency in the experiments is selected to be so low that the bias can be neglected.

### DETECTION OF REFLECTIONS

The average energy flow from a diffuse field is zero [15]. In addition, in an ideal diffuse field the sound intensity vectors are uniformly distributed. In contrast, in a plane wave all the intensity vectors point to the same direction. The early reflections in an impulse response are expected to resemble a plane wave, and the late reverberation should be nearly diffuse field.

Three methods for detecting the reflections in the impulse response in short time windows are presented. Conversely, the methods give estimates of the diffuseness. First of the methods is a simple local energy based method, the second and the third method are based on the distribution and energy of the sound intensity vectors. The processing is done in discrete time-frequency blocks although the equations presented here are in the continuous domains.

## Local Energy Ratio

A simple approach for detecting whether a frame includes a reflection is to compare its energy to the energy in the previous and future observations:

$$D_E(t) = \left[ \frac{1}{2T_g} \int_{t-T_g}^{t+T_g} |h(\tau)| d\tau \right] / \left[ \frac{1}{2T_l} \int_{t-T_l}^{t+T_l} |h(\tau)| d\tau \right] \quad (6)$$

where  $T_l$  and  $T_g$  are the local and global window lengths, given here the values 1.3 ms, and 50 ms. Naturally, the procedure could be done on a certain frequency band as well. Since in this article there are more than one microphone in use,  $D_E$  is averaged over all the microphones.

The minimum  $D_E(t) = T_l/T_g$  is found by placing for example a Dirac pulse at time instant  $t$ , and in this case  $\min\{D_E(t)\} = 0.026$ . It follows from the nature of the diffuse field, that the expected value of the maximum of  $D_E(t)$  is 1. The same applies for the other two estimators as well. That is, in a diffuse field the estimation results vary.

## SIRR Diffuseness Estimate

Diffuseness estimate used in Spatial Impulse Response Rendering (SIRR) is given by the relation between the sound intensity vectors length and the total energy [13, 14]:

$$D_{\text{SIRR}} = 1 - \frac{\|\mathbf{I}(\omega)/c\|}{E(\omega)} = 1 - \frac{2Z_0 \|\Re\{P^*(\omega)\mathbf{U}(\omega)\}\|}{|P(\omega)|^2 + Z_0^2 \|\mathbf{U}(\omega)\|^2} \quad (7)$$

where  $\|\cdot\|$  denotes the vector norm,  $|\cdot|$  is the absolute value of a complex number and  $Z_0 = \rho_0 c$  is the acoustic impedance of the medium. SIRR diffuseness estimate gets values between 0 and 1.

## Spherical Variance

In [6] the 2-dimensional spherical variance is used for detecting the reflections. Here 3-dimensional case is shown. The spherical variance is calculated as [17]

$$V = 1 - \|\mathbf{S}\|/W, \quad (8)$$

where

$$\mathbf{S} = \sum_{w_1}^{w_2} \mathbf{I}(w) \quad (9)$$

is the sum of all the individual vectors over the selected discretized frequency band from  $w_1$  to  $w_2$ , and  $W = w_2 - w_1$ . In the preliminary experiments it was noticed that the normalized intensity vectors  $\mathbf{I}(w)/\|\mathbf{I}(w)\|$  give more robust results than the unnormalized ones, therefore the spherical variance is calculated from the normalized ones.

As an example, consider a 2-dimensional case in polar coordinate system, and the angle component  $\theta$  of it. The spherical variance of a set of angle components is close to zero if they are “tightly clustered” [17]. Note that, although  $V = 1$  for an ideally diffuse field, it is also 1 for angles of the form  $[\theta_1, \theta_2 \dots, \theta_K, \theta_1 + \pi, \theta_2 + \pi \dots, \theta_N + \pi]$  [17]. Therefore, in theory  $V$  can not be used directly to estimate the diffuseness of the sound field. However, in practice, situations where the angles occur periodically are very rare.

## LOCALIZATION OF REFLECTIONS

The estimation of the distance of a detected reflection is formulated. In addition, three methods for estimating the direction of arrival of the reflections are presented. One of the methods uses sound intensity vectors and two are so-called steered response power methods. Examples of the methods are given in Fig. 3, in 2-dimensions for simplicity.

## Distance Estimation

The distance that the sound wave has travelled is calculated directly using the time delay of the detected reflection and the speed of sound

$$d = c\tau_k, \quad (10)$$

where  $\tau_k$  is the time delay related to a reflection  $k$ . When several microphones are in use and if the far-field assumption is used then (10) gives the distance to the center point of the array. Using near-field assumption the distance has to be calculated separately for each microphone. This can be done by applying (10) to the so-called time delay difference framework which is used in the cross correlation based methods, and is formulated later on in this section.

The speed of sound can be estimated to some extent if the direction of arrival and the intra-sensor distances are known. The speed of sound is solved with these parameters using basic geometry. For example so-called slowness vector estimation could be perhaps applied inversely for this problem since full dimensional version of it does not require any information on the speed of sound [18, Ch. 7].

## Mean Direction of the Intensity Vectors

The direction of the arriving sound wave can be estimated as the spherical mean (SME) of the sound intensity vectors over a frequency band [17]

$$\hat{\mathbf{I}} = \frac{\mathbf{S}}{\|\mathbf{S}\|}. \quad (11)$$

Conversion from Cartesian to spherical coordinates can be done using basic trigonometric equations. Here, the length of each sound intensity vector is first normalized to 1 before the conversion. This is done based on the results in [19], where the normalized vectors are found to provide more noise robust results than the unnormalized ones.

In [19] four other possibilities for estimating the direction of arrival from the sound intensity vectors are presented and discussed. Although the methods in [19] are given in 2-dimensions they can be extended to 3 dimensional data. Example of the spherical mean is given in Fig. 3(a).

## Cross Correlation based methods

A popular family of acoustic source localization functions is the steered response power methods, though throughout this article the name *cross correlation based methods* is used. In these methods, the acoustic source localization likelihood is evaluated as a spatial combination of cross correlation functions  $R_{x_i, x_j}$  for each location candidate  $\mathbf{x}$  [20]:

$$P(\mathbf{x}) = \sum_{\{i,j\}=1}^M R_{x_i, x_j}(\tau(\mathbf{r}_i, \mathbf{r}_j; \mathbf{x})), \quad (12)$$

where  $\{i, j\}$  denotes a microphone pair,  $\tau(\mathbf{r}_i, \mathbf{r}_j; \mathbf{x})$  is the time delay term, and  $M$  is the number of microphone pairs. The maximum argument of (12) is the location estimate:

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} \{P(\mathbf{x})\}. \quad (13)$$

In traditional acoustic source localization, the time delay difference is calculated as the difference of the distance between the location candidate and the locations of the microphones:

$$\tau(\mathbf{r}_i, \mathbf{r}_j; \mathbf{x}) = c^{-1} (\|\mathbf{r}_i - \mathbf{x}\| - \|\mathbf{r}_j - \mathbf{x}\|), \quad (14)$$

where  $c$  is again the speed of sound. Again, conversion from Cartesian to spherical coordinate system is done to obtain the

direction of arrival. An important difference between the spherical mean and the cross correlation methods is that (14) enables the distance from the source to the receiver to be taken in to account, whereas SME uses the farfield assumption.

The *generalized cross correlation* function between two received signals  $x_i$  and  $x_j$  is given as [20]:

$$R_{x_i x_j}(\tau) = \mathcal{F}^{-1}\{\mathcal{W}(f)C_{x_i x_j}(f)\}, \quad (15)$$

where  $\mathcal{W}(f)$ ,  $C_{x_i x_j}(f)$ , and  $\mathcal{F}^{-1}$ , are the weighting function, cross power spectral density between signals  $x_i$  and  $x_j$ , and inverse Fourier transform, respectively. Two different weighting functions are selected, the direct cross correlation (CC)

$$\mathcal{W}_{CC}(f) = 1 \quad (16)$$

and the phase transform (PHAT)

$$\mathcal{W}_{PHAT}(f) = 1/|C_{x_i x_j}(f)|. \quad (17)$$

Other options for weighting are discussed for example in [4].

In practice, the accuracy of the acoustic source localization is limited by the sampling frequency. To achieve higher accuracy, interpolation can be applied. Interpolation can be done either to the original signals, the cross correlation, or the localization function. Here the cross correlation function is interpolated with *exponential fitting*, which is presented in [21] for several maxima and for a single maximum in [22]. In exponential fitting,  $N_p$  highest maxima are selected, and they are assumed to have an exponential shape

$$f(\tau|\kappa[m]) = \kappa_{2,m} \exp(\kappa_{3,m}(\tau - \kappa_{1,m})). \quad (18)$$

The parameters  $\kappa[m] = \kappa_{1..3,m}$ , associated with a certain maximum  $m$ , are solved with respect to the maximum and the two discrete neighboring points. As a result, the cross correlation function is parameterized with  $N_p \times 3$  coefficients. Here 5 highest maxima of the cross correlation function were selected, that is  $N_p = 5$ . First coefficient,  $\kappa_{1,m}$ , describes the time delay associated with each maximum. The second one,  $\kappa_{2,m}$ , is the height of the maximum, and the third one,  $\kappa_{3,m}$ , is the width. The time delay estimation function is then given by

$$R_{x_i x_j}(\tau) = \max_m \{f(\tau|\kappa[m])\}, \quad (19)$$

i.e. the parameterized exponential giving the maximum value for  $\tau$ .

Examples of steered response function with the interpolation and different weighting for the cross correlation function is given in Fig. 3(b)-(c).

## EXPERIMENTS

Simulations are conducted to study the performance of the detection of reflection and the direction of arrival estimation. In addition, as an example one of the reflection detection and direction estimation methods are demonstrated with real impulse responses from an auditorium.

### Monte Carlo Simulations

The simulations are conducted in 3 dimensions using 6 microphones in a square grid with  $d = 2.5$  cm. A reflection is generated at a distance of 15 m, from the central point of the array, and the azimuth direction of arrival of the reflection is changed from 0 to 360 degrees between every 1 degree. The direction estimation is limited to azimuth angle. That is only the 4 microphones on the x-y plane are used. The results of the comparison can also be generalized to elevation angle. The distance is assumed to be known a priori before the direction estimation. A sinc-function is used as the signal for the reflection.

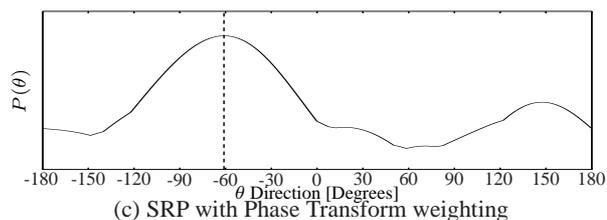
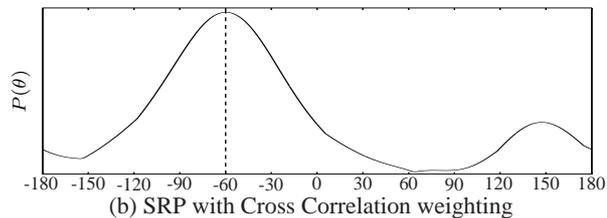
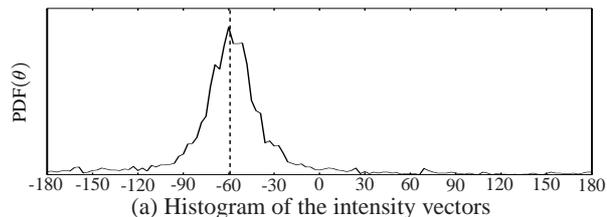


Figure 3: Examples of the direction estimation methods. The sound source is located at  $60^\circ$  and 15 m, and reflection-to-diffuse sound ratio is 0 dB. Estimated directions of arrival are  $-59.1^\circ$ ,  $-60.0^\circ$ , and  $-61.0^\circ$ , for SME, CC, and PHAT, respectively, and they are shown with horizontal dashed lines. The cross correlation based methods are discretized to the accuracy of 0.5 degrees.

Diffuse sound is added to the simulated pressure signals by adding uncorrelated white noise to each microphone. Adding white noise to the signals corresponds to a perfect diffuse sound field. The standard deviation  $\sigma_d$  of the white noise which is simulating the diffuse sound field is altered to study the performance. This quantity is noted here with Reflection-to-Diffuse sound Ratio (RDR), and it is the ratio between the energy of the reflection and of the diffuse sound field. Each condition is simulated 50 times, leading to 18000 Monte-Carlo samples for each RDR condition.

The direction was estimated using the discretized versions of the equations presented in the previous section. For the sound intensity based direction estimation methods, the frequency band was limited from 100 Hz to 3 kHz, due to reasons discussed earlier. The performance of the direction estimation methods is measured with spherical variance and absolute value of the spherical mean, noted here with  $V_e$  and  $|M_e|$ , respectively.

The results from the simulations are shown in Figs. 4 and 5. The results with the different diffusion estimation methods propose that all of the methods are capable of detecting diffuse and non-diffuse sound field. In addition, all have an asymptotic behavior. Spherical variance gives the most consistent results of all the three methods, since it varies between 0 and 1, and gives the value 0.5 when the RDR is 0 dB. SIRR diffuseness estimate is biased in the upper part since it gives values only up to 0.8. Local energy ratio does not decrease as steeply as the other estimates around 0 dB. The percentiles shown in Fig. 4 illustrate the earlier discussed behaviour of the diffuse field. That is, the uncertainty of the estimation increases as the diffuseness increases due to the random nature of the diffuse field.

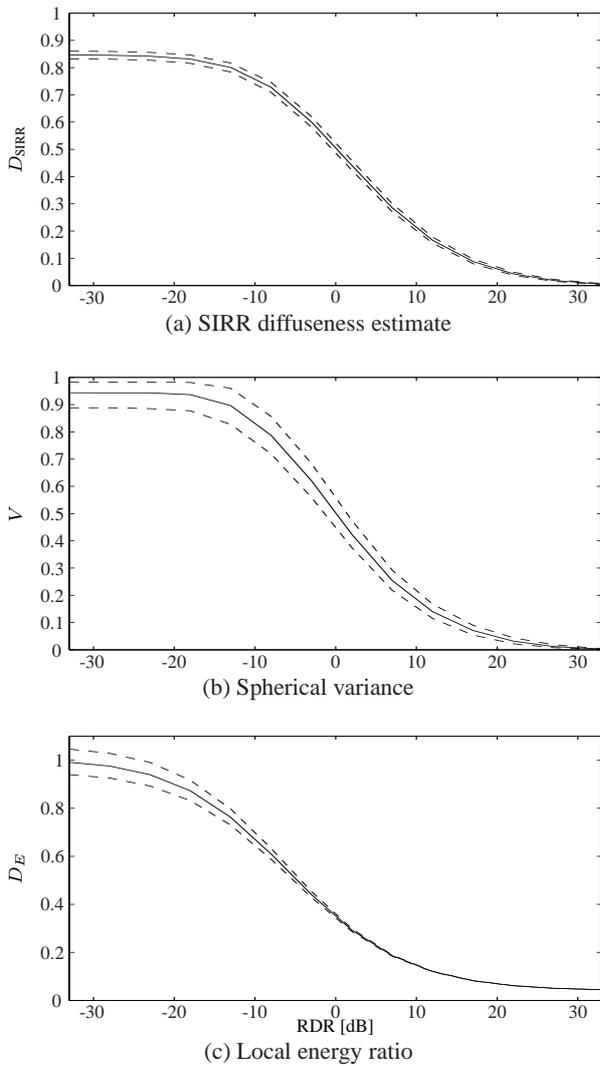


Figure 4: Performance of the diffuseness estimation methods against RDR. Percentiles 50 and  $50 \pm 47.6$  are shown. The uncertainty of the estimation increases as the diffuseness increases due to the random nature of the diffuse field.

The comparison between the direction estimation methods in Fig. 5a shows that CC suffers from the least variance in the estimation. With CC the bias is less than 5 degrees when RDR is more than -13 dB. For PHAT the corresponding figure is -11 dB and for SME it is -6 dB. SME has the highest variance. At least one of the reasons why the sound intensity vector based method perform worse is that it uses a limited frequency band whereas the cross correlation based methods use the whole frequency band. In the future work, the performance of the cross correlation methods in different frequency bands should be considered. In addition, in the cross correlation based methods the search of the direction maximum is done in a grid of 0.5 degrees. This lowers the variance when the conditions are good, since the estimated direction of arrival is quantized.

In this simulation, the source was in the far field (15 m) and the distance between the microphones was small (2.5 cm). If the source is in near field the estimation with SME becomes biased, since it uses the far field assumption by default. CC or PHAT will not suffer from this since the distance to the source can be taken into account in the time delay difference based source localization. Here the distance was assumed to be known a priori, therefore correct time delay differences could be used for CC and PHAT.

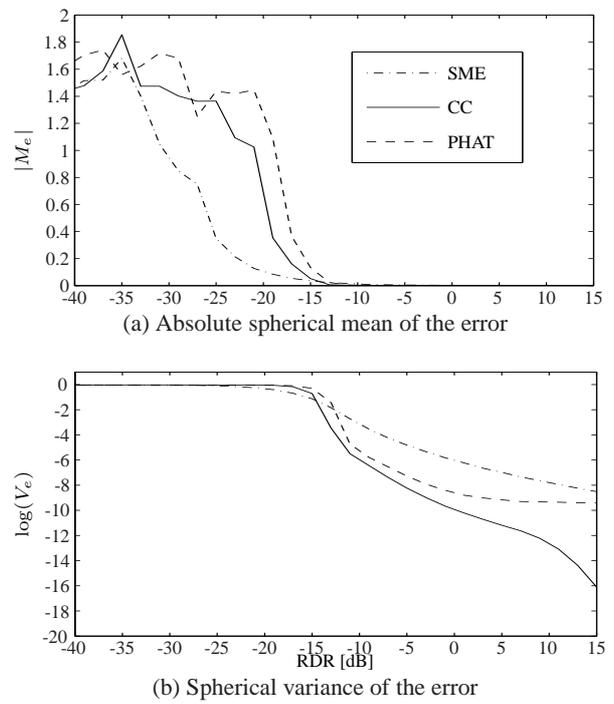


Figure 5: Performance of the direction estimation methods against RDR. Direct cross correlation (CC) has the lowest variance.

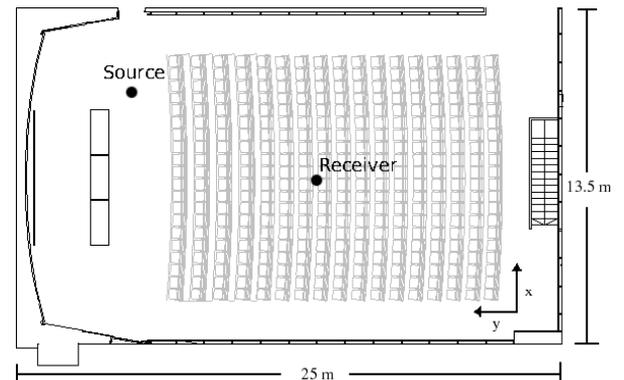


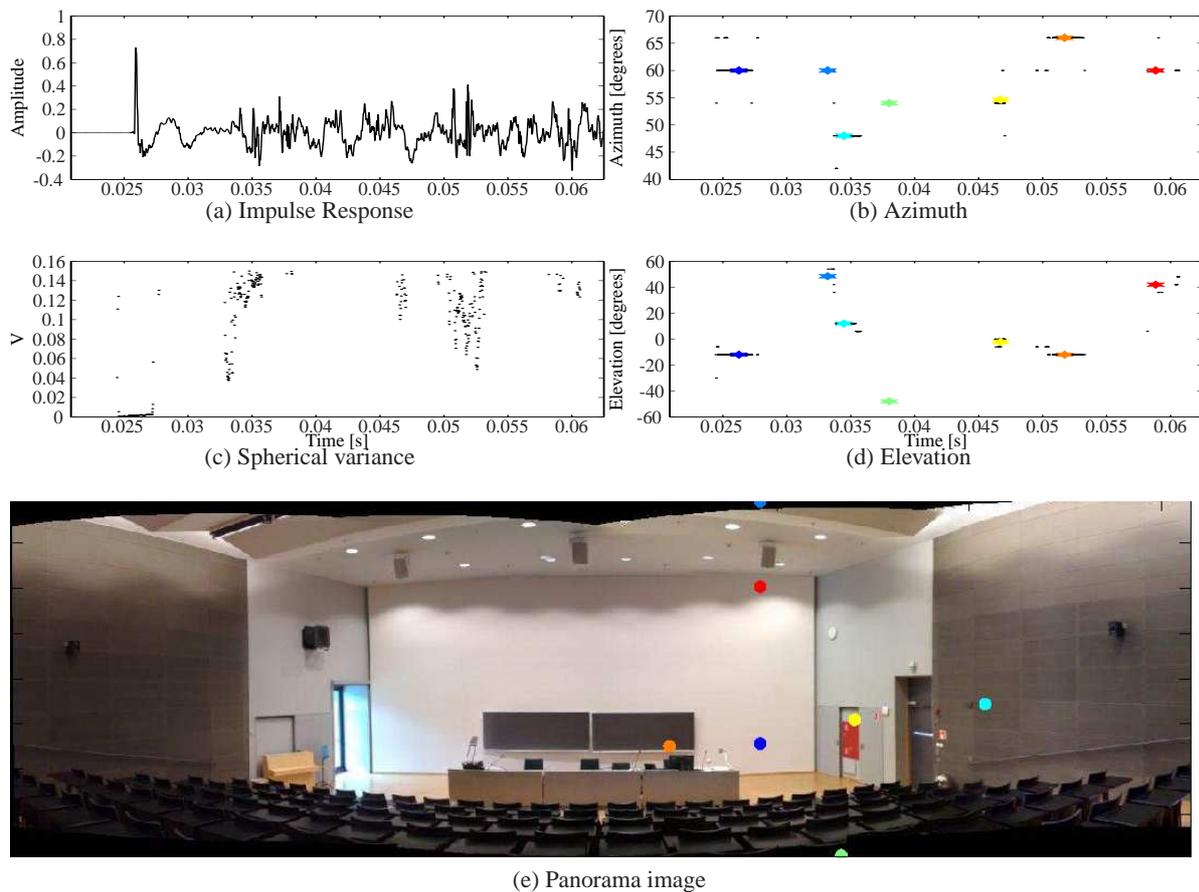
Figure 6: Setup in the auditorium and the overall floorplan.

### Example in an Auditorium

An example of reflection localization with a room impulse responses measured in an auditorium is shown. The measurement setup and floorplan of the auditorium are depicted in Fig. 6. The audience area has an inclination of about 10 degrees, as the height of the auditorium decreases from about 8 m to 5 m, leading to volume  $1800 \text{ m}^3$ . One source position and two receiver positions were used in the experiments. Three dimensional microphone grid (see Fig. 2) with  $d = 1 \text{ cm}$  and a loudspeaker which has an omnidirectional directivity pattern up to 8 kHz were used. The sampling frequency was set to 48 kHz, and the speed of sound was estimated to be  $c = 342.1 \text{ m/s}$ . Here the estimation of the speed of sound is based on the generally known relationship with the temperature.

Impulse responses were measured from 40 Hz to 24 kHz using the sine-sweep technique [23]. The impulse responses are analyzed using a rectangular window of length 128 samples with 99 % overlap with sampling frequency of 48 kHz.

The localized reflections and the impulse response from mi-



(e) Panorama image

Figure 7: Impulse response, spherical variance, and the direction of arrival of the reflections. Direction of arrival of the 6 manually selected reflections overlaid on top of a panorama image of the auditorium seen from the receiver location. The dark blue dot is the direct sound. Other colored dots represent the reflections.

crophone number 1 are shown in Fig. 7. The most consistent diffuseness estimate in the simulations, the spherical variance is used for detecting the reflections. If the spherical variance is below 0.15 the current frame is detected as a reflection. This threshold value is arbitrarily selected. The direction estimation is done with SRP using direct cross correlation weighting. Since the time of arrival is always known, the distance for the time differences in the SRP function in (14) can be used correctly at each time instant using (10).

The 99 % overlap and the selected threshold for spherical variance leads to a situation where a single reflection is detected more than once as can be seen in Fig. 7(c). An option to get rid off this problem is to group the estimates using e.g. Gaussian Mixture Model, according to their location and diffuseness value. Here a manual selection of the reflection was used to group them.

In Fig. 7(e) direct sound and 5 selected reflections are overlaid on a panorama image of the auditorium. It follows the idea of visualization of impulse responses shown in [5]. **Direct sound** is on arriving from the right hand side of the table, the source position is also illustrated in Fig. 6. **The first reflection** to arrive at the receiver is a ceiling reflection. **Second reflection** is a wall reflection from the right side wall. **Third reflection**, is a second order reflection via floor and the back wall, and **fourth reflection** is via floor and side wall. **Fifth reflection** is a second order reflection also via floor and back wall, and **sixth reflection** is via ceiling and back wall. Somewhat surprising is that none of the selected reflections are directly via back wall. It seems that the curvy shape of the wall does not produce any

first order reflections for this specific source-receiver combination.

## CONCLUSIONS

A framework for localizing reflections was proposed. The framework consists of three main parts of which one was studied here, the localization of reflections. This category can be further divided into two subcategories, detection of reflections and direction of arrival estimation.

The simulation results showed that a cross correlation based direction of arrival estimation methods outperform the sound intensity vector based methods. The framework was demonstrated also in a real auditorium using the methods that performed the best in the simulations.

Future work includes the estimation of the absorption coefficient, the reflective surfaces, the automatic calibration of source and microphone positions, and the estimation of speed of sound from measured impulse responses.

## ACKNOWLEDGEMENTS

The research leading to these results has received funding from the Academy of Finland, project nos. [119092, 218238, 5213462], the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no. [203636], Helsinki Graduate School in Computer Science and Engineering, and Graduate School in Electronics, Telecommunications and Automation.

## REFERENCES

- 1 V.C. Raykar, I.V. Kozintsev, and R. Lienhart. Position calibration of microphones and loudspeakers in distributed computing platforms. *IEEE Trans. on Speech and Audio Processing*, 13(1):70–83, 2005.
- 2 B.C. Ng and C.M.S. See. Sensor-array calibration using a maximum-likelihood approach. *IEEE Trans. on Antennas and Propagation*, 44(6 Part 1):827–835, 1996.
- 3 Y. Rockah and P. Schultheiss. Array shape calibration using sources in unknown locations—Part II: Near-field sources and estimator implementation. *IEEE Trans. on Acoustics, Speech and Signal Processing*, 35(6):724–735, 1987.
- 4 P. Pertilä. *Acoustic Source Localization in a Room Environment and at Moderate Distances*. PhD thesis, Tampere University of Technology, 2009.
- 5 A. O’Donovan, R. Duraiswami, and D. Zotkin. Imaging concert hall acoustics using visual and audio cameras. In *Proc. of ICASSP*, pages 5284–5287, 2008.
- 6 S. Tervo, J. Pätynen, and T. Lokki. Acoustic Reflection Path Tracing Using A Highly Directional Loudspeaker. In *Proc. of WASPAA*, pages 245–248, 2009.
- 7 B. Günel. Room shape and size estimation using directional impulse response measurements. In *Proc. of 3rd EAA Congress on Acoustics, Forum Acusticum*, 2002.
- 8 F. Antonacci, D. Aprea, A. Sarti, and S. Tubaro. Acoustic reconstruction of the geometry of an environment through acquisition of a controlled emission. In *Proc. of EUSIPCO*, pages 710–714, 2009.
- 9 M. Kuster, D. de Vries, E.M. Hulsebos, and A. Gisolf. Acoustic imaging in enclosed spaces: Analysis of room geometry modifications on the impulse response. *J. Acoust. Soc. of Am.*, 116:2126–2137, 2004.
- 10 M. Kuster and D. de Vries. Modelling and Order of Acoustic Transfer Functions Due to Reflections from Augmented Objects. *EURASIP J. Advances in Signal Processing*, 2007. Article ID 30253.
- 11 S. Tervo and T. Korhonen. Estimation of Reflective Surfaces from Continuous Signals. In *Proc. of ICASSP*, 2010.
- 12 F. Antonacci, Sarti A., and S. Tubaro. Geometric Reconstruction of the Environment from its Response to Multiple Acoustic Emissions. In *Proc. of ICASSP*, 2010.
- 13 J. Merimaa. *Analysis, Synthesis, and Perception of Spatial Sound: Binaural Localization Modeling and Multichannel Loudspeaker Reproduction*. PhD thesis, Helsinki University of Technology, 2006.
- 14 J. Merimaa and V. Pulkki. Spatial Impulse Response Rendering I: Analysis and Synthesis. *J. Audio Eng. Soc.*, 53(12):1115–1127, 2005.
- 15 H. Kuttruff. *Room acoustics*. Taylor & Francis, New York, NY, USA, 2000.
- 16 M. Kallinger, F. Kuech, R. Schultz-Amling, G. del Galdo, J. Ahonen, and V. Pulkki. Enhanced Direction Estimation Using Microphone Arrays for Directional Audio Coding. In *Proc. Hands-Free Speech Communication and Microphone Arrays*, pages 45–48, 2008.
- 17 K.V. Mardia, P.E. Jupp, and KV Mardia. *Directional statistics*. Wiley New York, 2000.
- 18 T. Pirinen. *Confidence Scoring of Time Delay Based Direction of Arrival Estimates and a Generalization to Difference Quantities*. PhD thesis, Tampere University of Technology, 2009.
- 19 S. Tervo. Direction Estimation Based on Sound Intensity Vectors. In *Proc. of EUSIPCO*, pages 700–704, 2009.
- 20 P. Pertilä, T. Korhonen, and A. Visa. Measurement combination for acoustic source localization in a room environment. *EURASIP J. Audio, Speech, and Music Processing*, 2008. Article ID 278185.
- 21 S. Tervo and T. Lokki. Interpolation methods for the SRP-PHAT algorithm. In *Proc. of 11th IWAENC*, 2008.
- 22 L. Zhang and X. Wu. On cross correlation based discrete time delay estimation. In *Proc. of ICASSP*, volume IV, pages 981–984, 2005.
- 23 A. Farina. Advancements in impulse response measurements by sine sweeps. In *122th AES Convention*, 2007. Paper 7121.