DOI: 10.61782/fa.2025.0754

FORUM ACUSTICUM EURONOISE 2025

ROBUST 3D LOCALISATION OF ANOMALIES IN THE REVERBERATION TIME SIGNAL

Thomas Rittenschober
Seven Bel GmbH
Linz, Austria

Antoine Decloux
Seven Bel GmbH
Grenoble, France

ABSTRACT

Reverberation time measurements form the basis for room acoustic optimizations of existing building structures. During the verification of the achieved room acoustic improvements, anomalies may appear in the reverberation time signal which may be hard to spatially localize, especially in spaces with demanding acoustic requirements such as large, open workspaces or concert halls.

This contribution focuses on the application of the Sound Field Scanning method to the fast spatial localization of room reflections. In this process, an omnidirectional sound source is positioned at an observation point in the room and periodically excited with bandlimited pulses. At the same observation point, an acoustic camera system consisting of a rotating linear microphone array is oriented towards the preferred spatial direction. The emitted pulses and associated room reflections are captured on the measurement surface of the rotating microphone array. Acoustic images with high depth resolution are generated in parallel planes to the measurement surface.

In complex situations, the task of spatially localizing anomalies in the reverberation time signal can be reduced to a few measurements from different perspectives, thus, significantly accelerating the problemsolving process with high confidence.

The method is exemplarily described through the room acoustic analysis of a university lecture hall.

Keywords: *acoustic imaging, sound field scanning, reflections, room acoustics*

1. INTRODUCTION

Sound source localization using two-dimensional microphone arrays is attractive for visualizing sound emissions from a distance without disturbing the sound pressure field.

These *acoustic cameras* typically consist of up to 130 microphones arranged over a circular area (35 cm to 2.5 m in diameter) and include an optical camera to overlay acoustic and visual data.

Their imaging performance is primarily determined by the array diameter and the number of microphones. The diameter governs spatial resolution—the larger the array, the better the resolution (see section “Sensor Concept”). The number of microphones affects the dynamic range and minimum detectable sound pressure level, as more microphones improve signal-to-noise ratio.

While stationary sound sources are often easily localized—including reflections—short-duration events (10 ms or less) pose a challenge. Since sound travels 1 m in under 3 ms, high temporal resolution is crucial for resolving reflections in space and time.

State-of-the-art acoustic cameras offer 10 ms temporal resolution, which is insufficient for many room acoustics applications, as reflections overlap in the resulting image.

These limitations motivated the development of a new sensor concept designed for high spatial resolution, dynamic range, and temporal precision, aimed at localizing specific components of the reverberation time signal.

2. SENSOR CONCEPT

The proposed sensor concept is motivated by the underlying physics describing the spatial resolution and dynamic range of a sound imaging system. For simplicity, we consider a sensor with a linear, continuously distributed sensing capability with aperture length L. The corresponding normalized, horizontal directivity pattern D is given by

$$D(\lambda, \theta) = \mathrm{sinc}\left(\frac{L}{\lambda}\cos\theta\right) \tag{1}$$

where θ is the angle of arrival of an incident sound wave and $\lambda = \frac{c}{f}$ is the wavelength, c is the speed of sound in air, and f is the sound frequency [2–4].

The sensor is most sensitive for sound waves coming in at zero degrees, and its sensitivity degrades for waves approaching at other angles.

The spatial resolution of a sound imaging system is typically quantified by the -3 dB beamwidth of the main lobe. An improved spatial resolution can therefore be achieved in two ways: (i) increasing the aperture length L or (ii) increasing the frequency of the sound event. Option (i) essentially translates into an increased size of the sensor which, as we will see later on, requires a higher count of distributed microphones and, thus, impacts the hardware complexity. Option (ii) may potentially be an available parameter in applications where the excitation frequency of the ultrasound transmitter can be controlled. Yet, it shall be considered that the higher the excitation frequency is, the more difficult it becomes to implement an ultrasound transmitter with both omnidirectional characteristics and sufficient sound power.

The side lobes play a special role for real arrays with a finite number of discrete microphones. In fact, the side lobe level quantifies the dynamic range of a sound imaging system. If, for instance, the side lobe level at a certain frequency is 10dB below the main lobe level and assuming that all involved sound sources can be spatially resolved, then the imaging system is still able to localize secondary sources with a pressure level less than 10dB below the most dominant source. The dynamic range can be improved by increasing the number of distributed microphones which, again, impacts the hardware complexity.

In order to get a better impression of the actual numbers that the above formulae suggest, we consider an automotive component with a size of 1 m by 1 m and key features lying apart in the range of about $d_1 = 10\,\mathrm{cm}$. The distance at which the measurement is conducted is about $d_2 = 75\,\mathrm{cm}$, which ideally requires the $-3\,\mathrm{dB}$ beamwidth of the main lobe to have an opening angle of less than $\theta_{\max} = 2 \cdot \arctan\left(\frac{d_1/2}{d_2}\right) = 0.13\,\mathrm{rad} = 7.63°$. When localizing impulse-like sound events in the order of a couple of milliseconds, the lowest localizable frequency shall be around $2.5\,\mathrm{kHz}$. Considering that the relation $\cos\theta = \frac{\lambda}{L}$ shall hold, we can derive the minimum array diameter as follows: $L_{\min} = \frac{\lambda}{\theta_{\max}} = 1.06\,\mathrm{m}$, with $\lambda = \frac{343\,\mathrm{ms}^{-1}}{2{,}500\,\mathrm{Hz}}$, see Fig. **??**.

Considering the typical landing pattern of a digital MEMS microphone which is in the range of $4\,\mathrm{mm}$ by $6\,\mathrm{mm}$, the hardware implementation of an array with a high count of microphones for optimum dynamic range can easily become a realization problem.

Based on these insights and trade-offs associated with distributing a high count of microphones across a measurement surface with a diameter of at least 1m, the authors propose a sensor concept which enables high spatial resolution and high dynamic range while targeting minimum weight and complexity of the associated sensor hardware.

2.1 Hardware Implementation

The centerpiece of the sensor concept is a rotating linear array with five distributed microphones which pivots about a non-moving reference microphone. The trajectory of the remaining moving microphones is described by concentric circles. A magnetic rotary encoder which is co-axially aligned with the rotation axis of the array, measures the angular position with respect to the spatially constrained axis of rotation.

The microphones are based on digital MEMS technology and the corresponding signals are acquired over a common signal path using the time division multiplexing (TDM) method. This method enables the straightforward implementation of a microphone multiplexing scheme for data compression and emulation of arbitrary, even non-implementable two-dimensional array geometries. For instance, the data acquisition can be configured such that the reference microphone along with a second channel which periodically switches between the moving microphones, are recorded, see Fig. 1.

It is well known that the directivity pattern of the

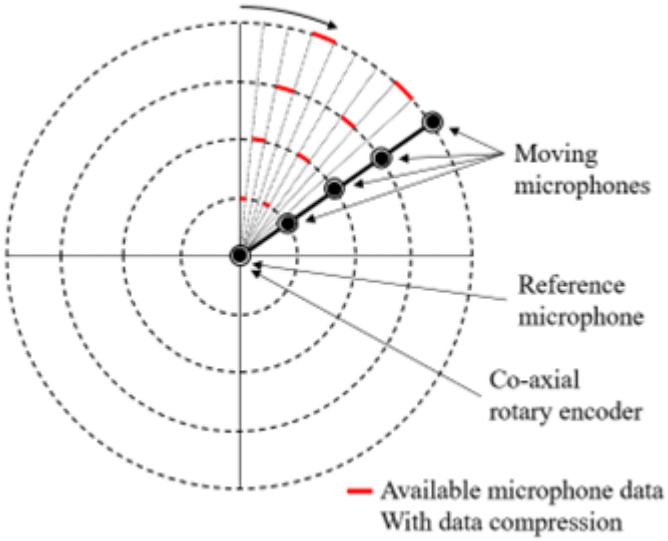

Figure 1. Multiplexing of the moving microphones enables data compression and emulation of arbitrary two-dimensional array geometries.

array and the corresponding position of microphones can be optimized to meet certain performance criteria, e.g. the minimum side lobe level at specific frequencies. While two-dimensional arrays with discrete microphone positions require a complete hardware reconfiguration in terms of repositioning the microphones, the rotating linear array merely requires a software reconfiguration to acquire the data at different positions.

Also, the implementation of large arrays with a diameter of more than one meter does not increase the hardware complexity. In fact, the number of microphones distributed along the linear array can stay the same since the fine spatial sampling along the concentric circles guarantees adherence to the spatial sampling theorem [1, 2].

The rotating linear array is a self-powered system and uses wireless technology for data transmission of the audio and rotary encoder data to a processing unit.

2.2 Properties of signals acquired by moving microphones

In order to better understand the characteristics of a signal acquired by a moving microphone, we consider a point source with harmonic excitation signal $u(t)$ at the frequency f_0:

$$u(t) = \Re\left\{e^{i2\pi f_0 t}\right\}$$

Assuming that the corresponding sound wave is independent from the distance to the source and the rotational speed f_{rot} of the moving microphone is constant, the audio signal $y_1(t)$ captured by the moving microphone is given by

$$y_1(t) = \Re\left\{e^{i2\pi f_0 t} e^{-i2\pi f_0 \frac{d_1(t)}{c}}\right\}$$

where $d_1(t)$ is the time-varying distance between the sound source and the position of the moving microphone along its circular trajectory with radius R, see Fig. 2.

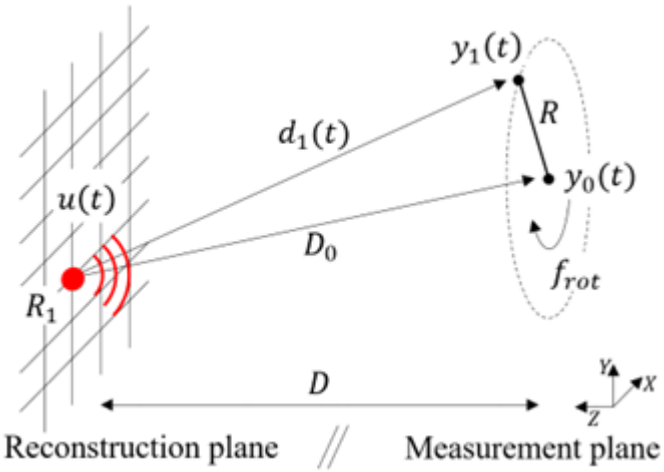

Figure 2. Notations used for describing the setup comprising a point source in the reconstruction plane and a reference microphone and a moving microphone located in the measurement plane.

We further denote the constant distance between the sound source at reconstruction point R_1 and the stationary reference microphone by D_0, the corresponding audio signal by $y_0(t)$:

$$y_0(t) = \Re\left\{e^{i2\pi f_0 t} e^{-i2\pi f_0 \frac{D_0}{c}}\right\}$$

and the distance between the parallel reconstruction and measurement planes by D. The origin of the Cartesian coordinate system with point representation given by (X, Y, Z) is at the position of the reference microphone,

and the XY-plane is the measurement plane. Considering the parameters

$$f_0 = 1\,\text{kHz}, \quad f_{\text{rot}} = 1\,\text{Hz}, \quad R = 1\,\text{m}, \\ D = 3\,\text{m}, \quad D_0 = (-0.1\,\text{m}, -0.1\,\text{m}, D) \qquad (2)$$

the short-time Fourier transformation of the moving microphone signal $y_1(t)$ is a Doppler-shifted version of the original source signal $u(t)$.

As expected, the short-time Fourier transformation of the moving microphone signal $y_1(t)$ is a Doppler-shifted version of the original source signal $u(t)$.

3. ACOUSTIC IMAGE COMPUTATION

The measurement setup depicted in Fig. 2 along with the basic observations on the signal properties of the moving microphone and reference microphone signals now enable us to derive an algorithm for the computation of a map describing the distribution of sound sources in the reconstruction plane.

3.1 The case of perfect Doppler shift compensation

As a first step, we map the signal $y_1(t)$ of the moving microphone to the spatial position of the reference microphone. This transformation involves backpropagating $y_1(t)$ to the point of the sound source in the reconstruction plane using the time-varying distance $d_1(t)$ and then forward propagating the signal to the reference microphone position using the constant distance D_0.

The resulting signal $\underline{y}_1(t)$

$$\underline{y}_1(t) = y_1\left(t + d_1(t) - D_0\right) \\ = \Re\left\{e^{i2\pi f_0 t} e^{-i2\pi f_0 \frac{(d_1(t) - d_1(t) + D_0)}{c}}\right\} = y_0(t)$$

has the obvious property that the Doppler shift previously induced in $y_1(t)$ is fully compensated and is identical to the signal captured at the position of the reference microphone.

3.2 The General Case

We now consider the transformation for a point R_2 in the reconstruction plane which is at a constant distance $\underline{D}_0$ from the reference microphone position and away from the point source, see Fig. 3.

An additional Doppler shift is induced in the transformed signal $\underline{y}_1(t)$ given by:

$$\underline{y}_1(t) = y_1\left(t + \underline{d}_1(t) - \underline{D}_0\right) \\ = \Re\left\{e^{i2\pi f_0 t} e^{-i2\pi f_0 \frac{(d_1(t) - \underline{d}_1(t) + \underline{D}_0)}{c}}\right\}$$

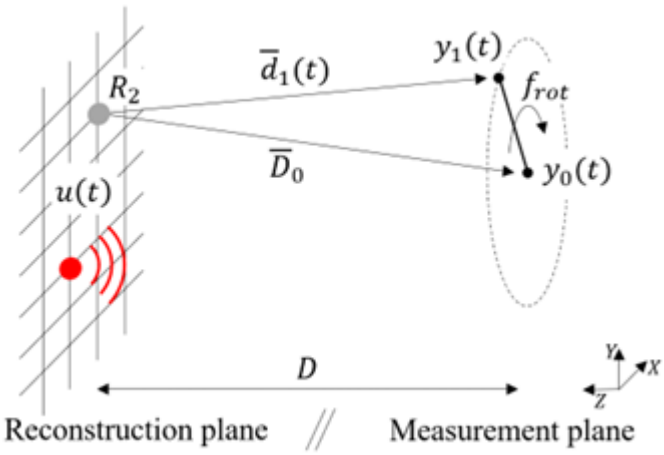

Figure 3. Notations used for describing the setup when mapping the moving microphone signal $y_1(t)$ to the reference microphone position via the point R_2 in the reconstruction plane.

Now, we apply the coherence function $C_{\underline{y}_1 y_0}(f)$ as a frequency-dependent measure of statistical similarity of the transformed signal $\underline{y}_1(t)$ and the signal $y_0(t)$ captured at the reference microphone position:

$$C_{\underline{y}_1 y_0}(f) = \frac{|S_{\underline{y}_1 y_0}(f)|^2}{S_{\underline{y}_1 \underline{y}_1}(f) S_{y_0 y_0}(f)}$$

where $S_{\underline{y}_1 y_0}(f)$ is the cross-spectral density of the signals $\underline{y}_1(t)$ and $y_0(t)$, and $S_{\underline{y}_1 \underline{y}_1}(f)$ and $S_{y_0 y_0}(f)$ are the power spectral density functions of $\underline{y}_1(t)$ and $y_0(t)$, respectively [4]. The coherence function varies in the interval $0 \leq C_{\underline{y}_1 y_0}(f) \leq 1$. We get a high coherence value at a specific frequency f for points in the reconstruction plane where a sound source is actually located, and a low coherence value for points where there is no or little sound radiation.

We can now use this metric to produce a heatmap representing the distribution of sound sources over the entire reconstruction plane. Considering the parameters from (2) and a point source at the spatial position $(-0.1\,\mathrm{m}, -0.1\,\mathrm{m}, -D)$, we get the color-coded representation of the coherence function $C_{\underline{y}_1 y_0}(f)$ evaluated at $f = 1\,\mathrm{kHz}$. With this special set of parameters, the resulting heatmap is also referred to as the point spread function (PSF), which is used to quantify the performance of an imaging system in terms of spatial resolution and dynamic range [1].

4. 3D TRACKING OF REFLECTIONS

The data acquisition and processing described in the previous section can be adapted to the task of localizing impulse-like sound events with high temporal resolution in a straightforward manner.

As a first step, the sound event under investigation shall be made repeatable. While the linear microphone array is spinning, the repeating sound events are recorded at different rotation angles, see Fig. 4.

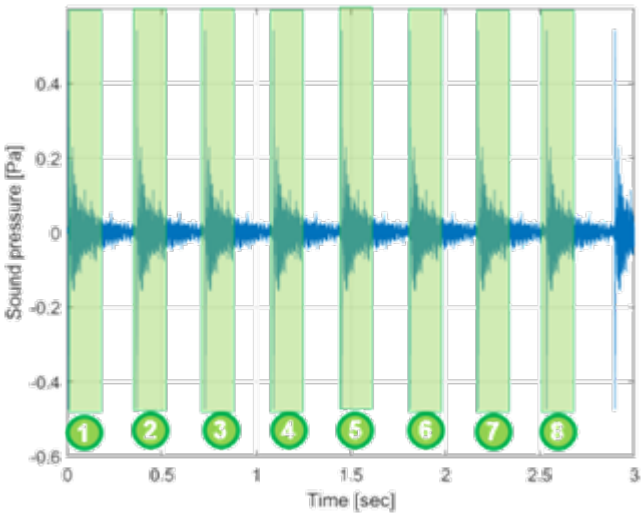

Figure 4. Record of periodic, matching sound events. The user selects a representative impulse and matching impulses are identified based on autocorrelation and frequency domain characteristics.

In a second step, the user selects a representative impulse and all repetitions are searched for in the remainder of the audio signal. The search algorithm uses auto-correlation and frequency domain characteristics in order to robustly identify valid repetitions of the representative impulse.

It is then verified whether the recorded impulses are homogeneously distributed across the measurement surface, see Fig. 5.

For doing so, the measurement surface is decomposed into sectors and one sector shall assigned to one impulse at most. Assigning a sector to more than one impulse or not assigning a sector to an impulse at all would be equivalent to introducing an apodization of the function describing the spatial distribution of microphones.

Based on the reduced set of impulses which have been spatially assigned to sectors of the measurement surface, the best homogeneous distribution of these impulses is computed.

The optimization criterion for deriving a best possible distribution of impulses is achieved by regarding each sector as a unit mass with unit distance from the center of the measurement surface.

A compound metric based on the inertia tensor and center of mass of this distribution can be used to derive an optimal subset of spatially recorded impulses.

Now that the relevant impulses have been identified, we partition the selected impulse of the audio signal into overlapping time intervals in the order of milliseconds. The matching partitions in the other impulses can be found based on their time-offset from the selected impulse, see Fig. 6.

An acoustic image for a specific partition of the selected impulse can now be computed based on audio signals only consisting of matching partitions from all relevant impulses. Since the onset of the direct sound from the speaker to the reference microphone of the linear array is known, the distance between the measurement plane and the reconstruction plane for a specific signal portion of the reverberation signal can be computed accordingly. This means that the reconstruction plane moves away from the measurement plane at the speed of sound. Thus, for every signal portion of the reverberation signal a different distance from the measurement plane is chosen.

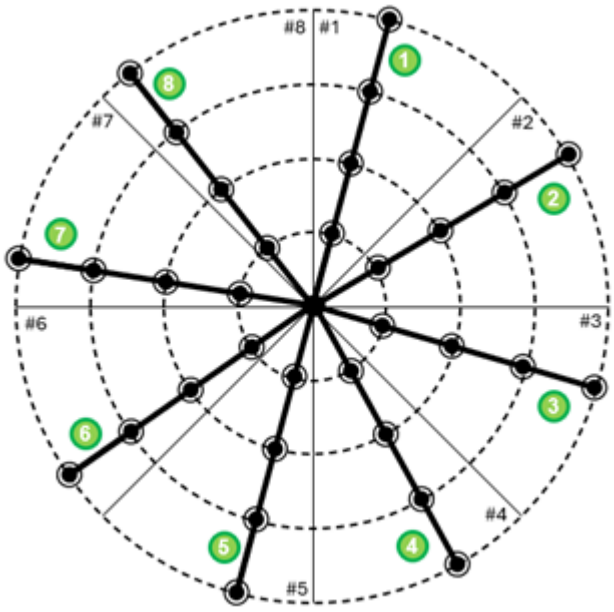

Figure 5. The best possible set of spatially recorded impulses is derived from omitting redundant impulses for specific sectors of the measurement surface (e.g. impulse 4 in sector 3) and considering a compound metric based on the inertia tensor and the center of mass of this spatial distribution.

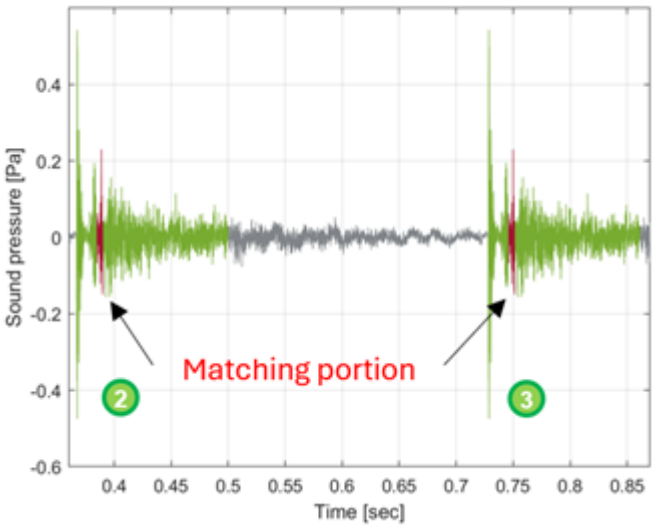

Figure 6. Matching partitions for impulses 2 and 3 which are spatially assigned to sectors 2 and 3, respectively.

5. APPLICATIONS

The proposed sensor concept is now applied to localizing an anomaly in the reverberation signal which is recorded in the stage area of a concert hall.

5.1 Setup

The measurement setup comprises the following devices:

- a rotating sensor with a total of five microphones sampled at 21.3 kHz and distributed over a length of 66 cm with one reference microphone at the center of rotation and four microphones moving along circular trajectories on a disc with a maximum diameter of 132 cm
- a mobile device (model: Samsung Galaxy Tab S9) for capturing the audio as well as rotary encoder data and sending the data to
- a high performance laptop (model: Dell Precision 7680) for computing the acoustic images and
- an omnidirectional speaker (model: Norsonic Nor276 including amplifier Nor 280).

The omnidirectional speaker is positioned in the center of the stage along with the rotating linear array, see Fig. 7 and Fig. 8.

Figure 7. Arrangement of omnidirectional source and rotating linear array in the stage area of a concert hall.

The sensor rotates at a speed of two revolutions per second. The impulse-like sound event is emitted at a repetition frequency of 2 Hz. A total of 20 sound events are captured. The optical image is taken with a horizontal field of view of 69.5° and the optical camera—facing away from the auditorium—is pointed in the direction of the left half of the stage area.

Figure 8. Co-located arrangement of omnidirectional source and rotating linear array.

5.2 Results

Fig. 9 depicts the recorded reverberation time signal of the reference microphone where the horizonal axis is expressed as a distance in meters from the measurement plane. The impulse from the direct sound event is visible at the beginning of the chart, i.e. at a distance of 0 m away from the measurement plane. At a distance of 17 m away from the measurement plane, we can detect an anomaly in the reverberation time signal.

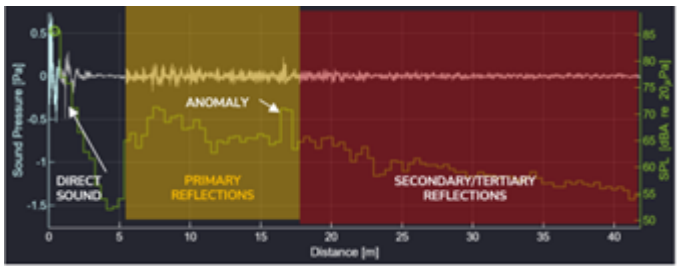

Figure 9. Recorded reverberation time signal with anomaly at a distance of 17 m away from the measurement plane.

Fig. 10 shows the corresponding acoustic image which localizes the anomaly in the top left ceiling area of the stage. The ceiling is equipped with diffusor panels.

The arrangement of omnidirectional source and rotating linear array can be modified for the analysis of sound propagation from one point in space to an observation point, see Fig. 11.

Fig. 12 depicts the recorded reverberation time signal of the reference microphone. The time chart can be decomposed into a time interval of $20 - 60$ ms after the

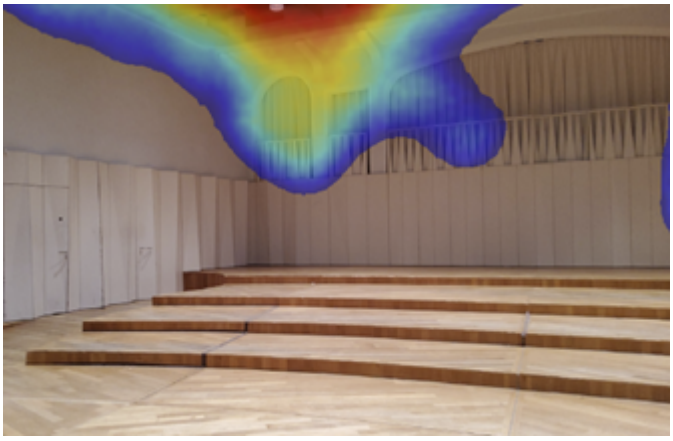

Figure 10. Anomaly in the reverberation time signal at a distance of 17 m away from the measurement plane. Selected frequency band is 250Hz – 3.1kHz at a dynamic range of 3dB.

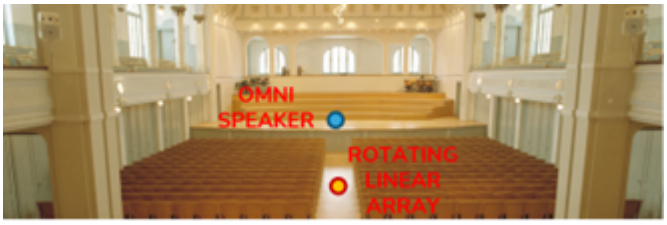

Figure 11. Arrangement of omnidirectional source and rotating linear array for the analysis of sound propagation from the stage area to an observation point in the auditorium.

direct sound event (primary reflections) and a second time interval of 60 – 170 ms after the direct sound event (critical reflections).

While the primary reflections control the level of the received audio signal, the critical reflections govern the delayed perception of the audio signal.

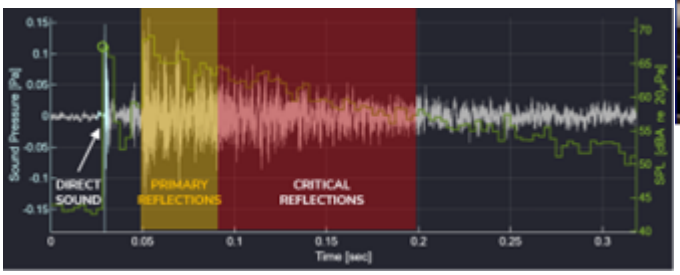

Figure 12. Recorded reverberation time signal with primary reflections (20 – 60 ms after direct sound event) and critical reflections (60 – 170 ms after direct sound event).

The corresponding spatial areas where these reflections are received from at the point of observation are depicted in Fig. 13 and Fig. 14.

Figure 13. Primary reflections located at the point of observation. Selected frequency band is 250Hz – 3.1kHz at a dynamic range of 2dB.

In case the delayed perception of audio signals at the point of observation is an issue, the diffusor ceiling in the stage area can be singled out as the primary reflector while the diffusor wall of the stage area can be considered as a secondary effect with sound pressure levels 6 dB below.

Figure 14. Critical reflections located at the point of observation. Selected frequency band is 250Hz – 3.1kHz at a dynamic range of 6dB.

6. SUMMARY

This contribution addressed the application of the Sound Field Scanning method to the fast spatial localization of room reflections.

The underlying measurement method is derived and its performance properties are described. A measurement setup comprising the above mentioned sensor, a mobile device, a high performance laptop and omndirectional speaker is used to produce acoustic images for localizing anomalies in the reverberation signal which is recorded both in the stage area and auditorium of a concert hall.

Future work will analyse the effectiveness of the method for optimizing the room acoustics of different environments, e.g. office space and production environment.

7. REFERENCES

[1] M. Brandstein, D. Ward: Microphone Arrays, Springer, 2001.

[2] S. Haykin: Array Signal Processing. Prentice-Hall, 1985.

[3] L. J. Ziomek: Fundamentals of Acoustic Field Theory and Space-Time Signal Processing. CRCPress, 1995.

[4] Shin. K, Hammond. J.: Fundamentals of signal processing for sound and vibration engineers. John Wiley & Sons, 2008.