On the Localization of Sound Sources with Moving Microphones
Ortung von Schallquellen mit bewegten Mikrofonen

Dr. Thomas Rittenschober, Seven Bel GmbH, Linz, Österreich, thomas.rittenschober@seven-bel.com

Kurzfassung

Dieser Beitrag beschreibt eine neuartige Methode zur Visualisierung von Schallquellen unter Verwendung eines rotierenden Linear Array mit einigen wenigen digitalen Mikrophonen. Das rotierende Array tastet hierbei ein einfallendes Schallfeld auf einer kreisförmigen Fläche ab. Während die reduzierte Hardware Komplexität das Design eines leichtgewichtigen Handgeräts für die Schallvisualisierung ermöglicht, benötigt der algorithmische Teil des Mess-Systems besondere Aufmerksamkeit. Wohlbekannte Methoden zur Schallvisualisierung, wie z.B. Beamforming oder Nahfeld-Holographie, können nicht ohne weiteres auf Signale, die von bewegten Mikrophonen stammen, angewendet werden. Die vom Autor vorgeschlagene Methode basiert auf einer Kompensation der Doppler Verzerrungen der bewegten Mikrofonsignale und einer Bewertung der Kohärenz der resultierenden Signale in Bezug auf ein unbewegtes Referenz-Mikrofonsignal für jeden Punkt eines akustischen Bildes. Das Messverfahren wird im Kontext der Leckagesuche an Automotive Strukturen näher beleuchtet.

Abstract

This contribution describes a novel method for visualizing sound sources using a rotating linear array of a few digital microphones. The rotating array scans the incident sound field on a circular area. While the reduced hardware complexity allows for the design of a lightweight, handheld sound imaging device, the algorithmic portion of the measurement system requires special attention. In fact, established methods of sound imaging like beamforming and nearfield holography cannot be applied to signals stemming from moving sensors. The proposed method of computing an acoustic image using the described measurement setup is based on compensating the moving microphone signals for Doppler distortions and evaluating the coherence of the resulting signals with a non-moving reference microphone for each point in the acoustic image. The method is evaluated in the context of leakage detection in automotive structures.

1 Introduction

This contribution addresses the topic of localizing sound sources with moving microphones. Special attention is paid to the use case of leakage visualization in automotive structures based on a transmitter-receiver setup where the transmitter is a speaker operating in the ultrasound domain and the receiver is a measurement instrument comprising a multitude of ultrasound microphones.

An overview of existing technologies using such a transmitter-receiver setup is given in section “State of the Art” where the foundation for motivating a receiver design based on a rotating linear microphone array is laid.

In section “Sensor Concept”, the authors describe the hardware related properties of the sensor and the special characteristics of acoustic data captured with the proposed sensor.

Section “Acoustic Image Computation” explains in detail how the acoustic signals captured by the sensor are processed to generate heatmaps showing the spatial position and strength of leakages.
In section “Applications”, the authors give an overview of the areas of a car where leakage detection is typically performed. Actual measurement results from production cars are depicted and analyzed.

2 State of the Art

The ultrasound microphone is a widely accepted sensor device for detecting leakages in industrial infrastructure, e.g. compressed air lines, gas pipelines, pressure vessels. A crack, fissure or porosity in the structure causes the gas to escape and produces a stationary bandlimited sound in the ultrasound frequency domain. The acoustical footprint of the leakage depends on parameters such as defect geometry, pressure differential and gas type but early indications of a leakage can be typically detected in the range of 30-40kHz [1]. To enable humans to search for a leakage, the captured acoustical signal is downconverted to the audible frequency band. When the microphone gets closer to the leakage, the perceived loudness of the downconverted hissing sound increases.

When performing leakage detection for an automotive structure such as a door, window, trunk or bulkhead, the sound escaping through seals, gaskets, grommets and other imperfections has to be generated. For doing so, an ultrasound transmitter is put inside the cavity or cabin of interest. This device typically comprises a multitude of loudspeakers emitting sound in an omnidirectional fashion at multiple distinct frequencies at about 40kHz, thus creating a homogenous sound pressure field at the positions where leakages shall be checked.

ISBN 978-3-8007-5835-7

Instead of logging the corresponding sound pressure levels along the paths of interest, it is common practice to rely on cues from loudness variations in the downconverted acoustical signal when searching for dominant leakages. Since the loudness may only vary within a certain dynamic range, it is required to first dial in on an expected loudness level. Having to focus on loudness variations over a longer period of time may lead to fatigue and lack of concentration.
Ultrasound imaging has, thus, become an attractive alternative to surface bound probing with an ultrasound microphone because visualizing leakages in automotive structures from a distance allows capturing a larger area at once, gives consistent quantification across that area and does not distort the sound pressure field caused by the presence of the measurement device.

Ultrasound imaging devices are typically implemented as two dimensional microphone arrays with up to about 130 ultrasound microphones distributed over a circular area with a diameter of less than 15cm and a camera to overlay the acoustical heatmap and the optical image of the measurement scene. The imaging performance of ultrasound imaging devices is mostly governed by two contributors, namely the diameter of the microphone array and the number of microphones distributed over that area. The array diameter predominantly sets the spatial resolution. The larger the array diameter, the higher the spatial resolution is, see section “Sensor Concept” for a detailed explanation. The number of microphones influences the dynamic range of the resulting acoustic image, i.e. the maximum difference in loudness level that can be resolved. Also, a higher count of distributed microphones directly translates into a lower minimum detectable sound pressure level due to an improved signal-to-noise ratio.

Since ultrasound leakages in automotive applications are typically measured at a distance of 0.75 to 1.5 meters and automotive structures exhibit improved insulation with every new generation, the signal-to-noise ratio and, thus, the number of microphones becomes essential. Also, since automotive structures have joints running in parallel with only a few centimeters apart, e.g. B-pillar, spatial resolution at the excited frequency of the ultrasound transmitter is essential, see section “Sensor Concept” for a detailed explanation.

While ultrasound imaging devices enable a fast measurement and analysis, it shall still be noted that – at the time of this publication - the weight of available handheld devices is close to or above the recommended maximum weight of 2.3kg for handheld equipment per standard MIL-HDBK-759C [2]. Using such equipment for an extended period of time may pose the risk of tiring of the hand-arm-system.

These performance and ergonomics criteria have led the authors to propose a new sensor concept targeting at high spatial resolution and dynamic range, low minimum detectable sound pressure level and comfortable use.

3 State of the Art

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

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

where θ is the angle of arrival of an incident sound wave and λ = c/f is the wavelength, *c* is the speed of sound in air and *f* is the sound frequency [4] [5] [6]. The shape of the directivity function is depicted in Figure 1. In this configuration, the distributed sensor is most sensitive for sound waves coming in at zero degree and its sensitivity degrades for waves approaching at other angles.

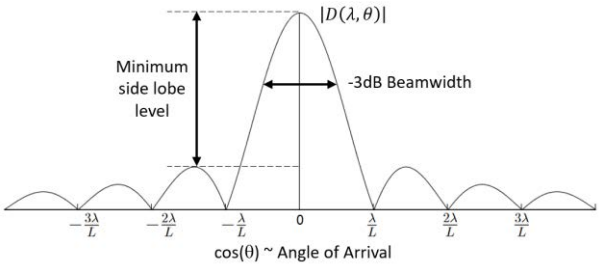

Figure 1. Normalized, horizontal directivity pattern D for a continuous linear array with aperture length L evaluated at wavelength λ.

The spatial resolution of a sound imaging system is typically quantified by the -3dB 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 in Figure 1 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 the B-pillar of a car body where the distance between window and door sealings running in parallel is typically in the range of about $d_1 = 10cm$. Resolving leakages measured at a distance of about $d_2 = 75cm$ ideally requires the main lobe to have an opening angle of less than $\theta_{max} = 2\tan^{-1}\left(\frac{d_1/2}{d_2}\right) = 0.13rad = 7.63°$. When detecting leakages at 40kHz in air, this translates into a minimum array diameter of $L_{min} = \lambda / \sin(\theta_{max}) = 4.1cm$ with $\lambda = 343ms^{-1}/40.000Hz$, see Figure 1.

Considering the typical landing pattern of a digital MEMS microphone which is in the range of 4mm by 6mm, 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, the authors propose a sensor concept which enables high spatial resolution and high dynamic range while targeting minimum array size, weight and complexity of the associated sensor hardware.

3.1 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 coaxially aligned with the rotation axis of the array, measures the angular position with respect to the spatially constrained axis of rotation, see Figure 2.

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 Figure 3.

It is well known that the directivity pattern of the 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.

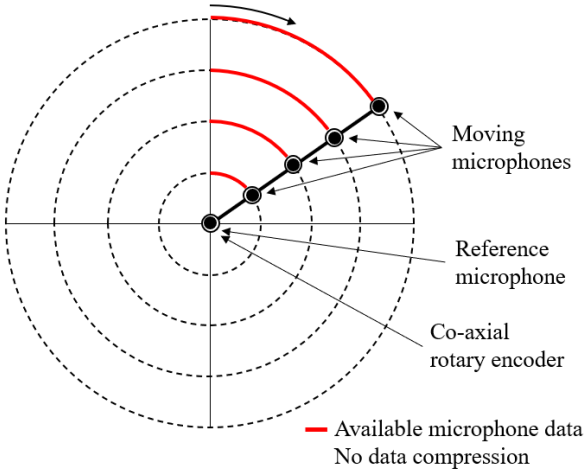

Figure 2. Rotating linear array comprising five microphones pivoting about the reference microphone. The trajectory of the moving microphones is described by concentric circles.

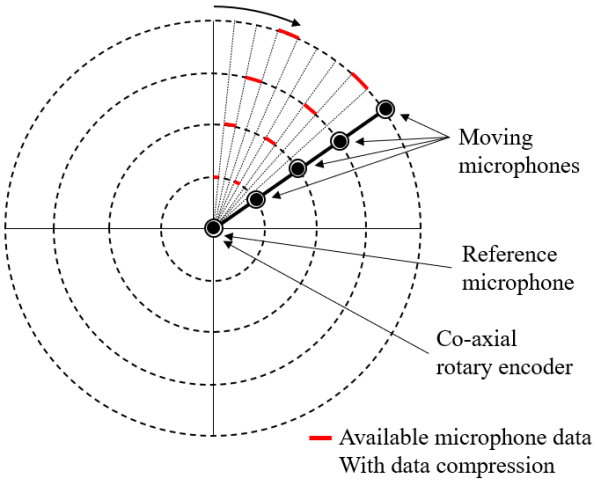

Figure 3. Multiplexing of the moving microphones enables data compression and emulation of arbitrary two-dimensional array geometries.

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 [3] [4].

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.

ISBN 978-3-8007-5835-7

3.2 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

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

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 Figure 4.

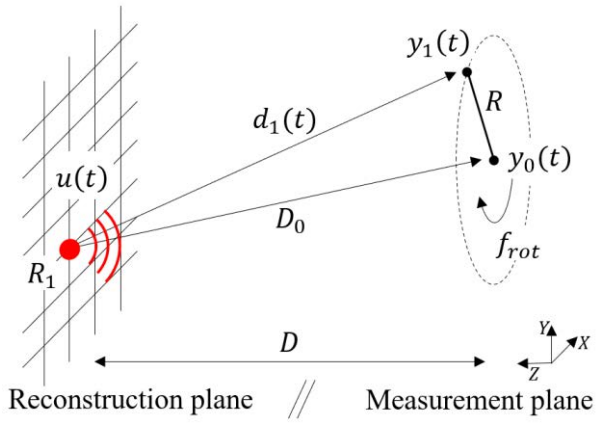

Figure 4. Notations used for describing the setup comprising a point source in the reconstruction plane and a reference microphone and a moving microphone located in the measurement plane.

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

$$\begin{aligned} f_0 &= 1kHz, \\ f_{rot} &= 1Hz, \\ R &= 1m, \\ D &= 3m, \\ D_0 &= (-0.1m, -0.1m, D), \end{aligned} \tag{1}$$

we get the following result for the time-varying distance $d_1(t)$

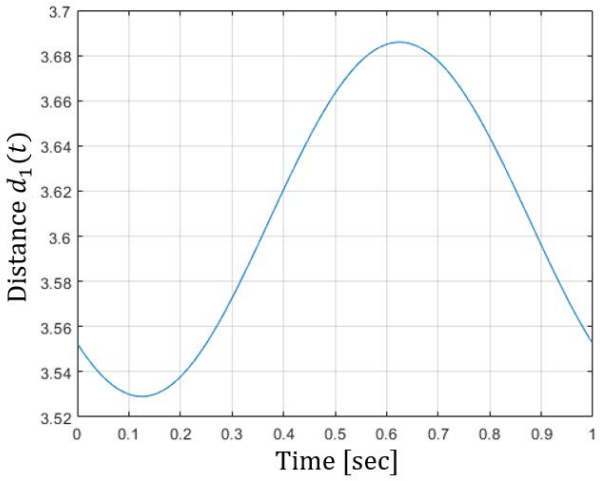

Figure 5. The time-varying distance between the sound source and the position of the moving microphone along its circular trajectory.

and the short-time Fourier transformation of the audio signal $y_1(t)$, see Figure 5 and Figure 6.

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 $u(t)$.

3.3 Acoustic image computation

The measurement setup depicted in Figure 4 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.

3.3.1 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 $\bar{y}_1(t)$

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

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.

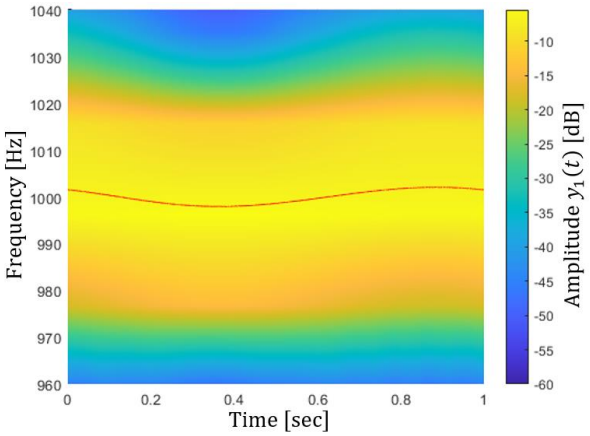

Figure 6. The short-time Fourier transformation of the moving microphone signal for the time-varying distance $d_1(t)$ depicted in Figure 5.

3.3.2 The general case

We now consider the transformation for a point R_2 in the reconstruction plane which is at a constant distance $\overline{D}_0$ from the reference microphone position and away from the point source, see Figure 7.

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

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

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

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

where $S_{\bar{y}_1 y_0}(f)$ is the cross-spectral density of the signals $\bar{y}_1(t)$ and $y_0(t)$ and $S_{\bar{y}_1 \bar{y}_1}(f)$ and $S_{y_0 y_0}(f)$ are the power spectral density functions of $\bar{y}_1(t)$ and $y_0(t)$, respectively [6].

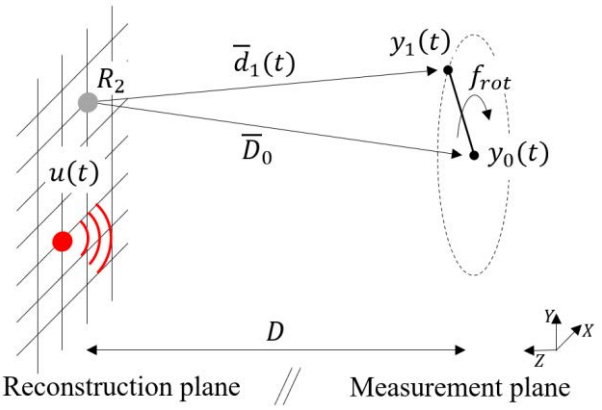

Figure 7. Notations used for describing the setup when mapping the moving microphone signal $y_1(t)$ to the reference microphone position via the point R_2 in the reconstruction plane.

The coherence function varies in the interval $0 \leq C_{\bar{y}_1 y_0}(f) \leq 1$. 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 Equation (1) and a point source at the spatial position $(-0.1m, -0.1m, -D)$, we get the color-coded representation of the coherence function $C_{\bar{y}_1 y_0}(f)$ evaluated at $f = 1kHz$ depicted in Figure 8. With this special set of parameters, the resulting heatmap is also referred to as point spread function (PSF) which is used to quantify the performance of an imaging system in terms of spatial resolution and dynamic range [3].

3.3.3 Comparison with fixed arrangement of microphones

In order to appreciate the image quality achieved with one moving microphone only, we can compute the coherence function for a hardware setup with a fixed spatial arrangement of microphones equally spaced along the trajectory of the moving microphone. Figure 9 depicts the result for 12, 24 and 96 microphones.

It is readily visible that the coherence function for the arrangement of 96 microphones approaches the result from the moving microphone. Using 12 microphones only leads to the well-known artifact of grating lobes caused by spatially undersampling the incident sound field [4].

ISBN 978-3-8007-5835-7

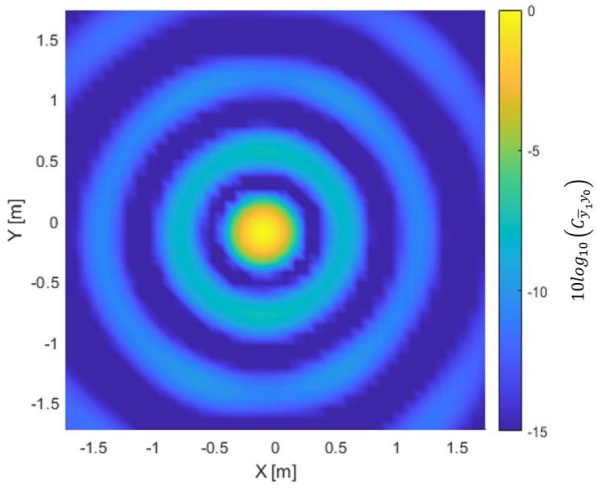

Figure 8. Color-coded logarithmic representation of the coherence function $C_{\bar{y}_1 y_0}(f)$ evaluated at $f = 1kHz$.

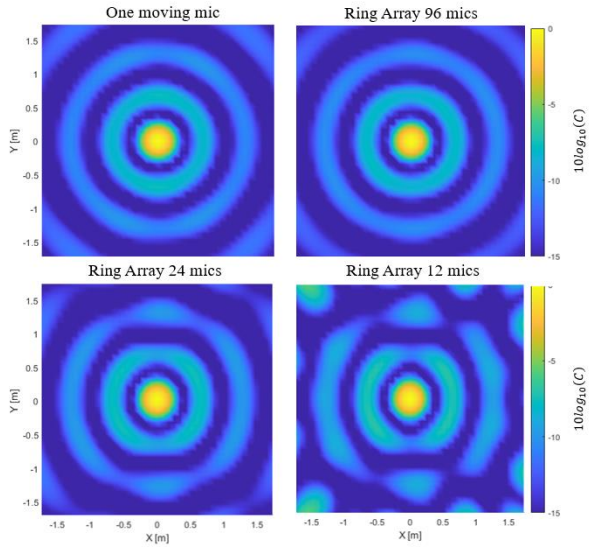

Figure 9. Coherence function $C_{\bar{y}_1 y_0}(f)$ for fixed spatial arrangement of 12m 24 and 96 microphones evaluated versus one moving microphone.

3.3.4 Multiple distributed sound sources

Before we evaluate the capabilities of the proposed sensor concept in real world applications, we finish this section by computing the distribution of multiple sound sources of equal strength emitting a tone at 1kHz and located at positions $P_1 = (-1m, 0, D)$, $P_2 = (0{,}0, D)$, $P_3 = (1m, 1m, D)$ in the reconstruction plane, see Figure 10.

We shall note that the artifacts surrounding the three sound sources are coming from the mutual, positive interference of the point spread functions at the three spatial positions, thus, degrading the useable dynamic range.

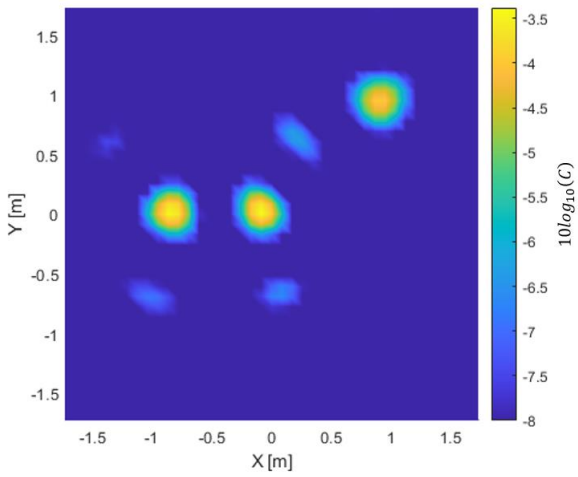

Figure 10. Coherence function $C_{\bar{y}_1 y_0}(f)$ for multiple sound sources emitting a tone at 1kHz and located at positions $P_1 = (-1m, 0, D)$, $P_2 = (0{,}0, D)$, $P_3 = (1m, 1m, D)$ in the reconstruction plane.

4 Applications

The proposed sensor concept is now applied to leakage and tightness testing of actual automotive structures in a quality control context.

4.1 Measurement setup

The measurement setup comprises the following devices:

- a rotating sensor with a total of eight microphones sampled at 130kHz and equally distributed over a length of 6cm with one reference microphone at the center of rotation and seven microphones moving along circular trajectories on a disc with a maximum diameter of 12cm ,
- an omnidirectional ultrasound loudspeaker (manufacturer: SDT, model:T-Sonic9) placed inside the car cabin and generating ultrasound at 39,9kHz and 40,1kHz at a stable sound pressure level of 113dB measured at a distance of 1m under free field conditions,
- a mobile device (model: Samsung A71) for capturing the audio as well as rotary encoder data and sending the data to
- a high performance laptop (model: Dell precision 7550) for computing the acoustic images.

The vehicle under test is a production car (model: Skoda Octavia station wagon, model year 2021) with all doors and windows closed. The sensor is positioned at a distance of 1m and two viewing angles are considered – one targeting at the tailgate and the other focusing on the right side of the vehicle.

The sensor rotates at a speed of two revolutions per second producing twelve acoustic images per second. The incident sound field is spatially sampled by the moving multiplexed microphone at 180 sectors per revolution,

ISBN 978-3-8007-5835-7

also see Figure 3. The optical image is taken with a horizontal field of view of 69,5°.

4.2 Measurement process

A measurement process which is suited for quality control purposes, would typically involve a nominal acoustic footprint. Based on an expected sound pressure level distribution for a well-sealed car, we can quantify acceptable deviations. For the measurement scenario depicted in Figure 11, we see a leakage at the interface between the tailgate and the sill at 53,5dB with a dynamic range of 6dB.

Figure 11. Leakage at the interface between the tailgate and the sill at 53,5dB with a dynamic range of 6dB.

Since this is a production car, we expect this behavior to be nominal. As part of a rigorous, data based quality control scheme, each car can be tested end of line and assessed in terms of pass-fail criteria with respect to tightness of critical structures. Reasonable criteria can be the maximum allowable sound pressure level in a certain region of interest, e.g. along the sill-tailgate interface, or quantified deviations from an expected distribution of leakages with associated pressure levels. Any deviation with respect to these criteria can be an indication of a faulty assembly and, thus, compromise drive comfort.

Figure 12 shows multiple distributed leaks along the seal of the front passenger's window. In this situation, the maximum sound pressure level is at 48,7dB at 3dB dynamic range.

Another noteworthy automotive use case is the identification of dominant leakages in prototype cars. Associated structures and components may experience larger variations in terms of tightness than those seen in series production. Yet, the assessment of acoustic drive comfort is critical during these stages of the development process and feedback from test drivers needs to be assessed properly. In such a situation, acoustic snapshots from multiple viewing angles enable test and verification engineers to assess the capability of their vehicles and derive appropriate counter measures.

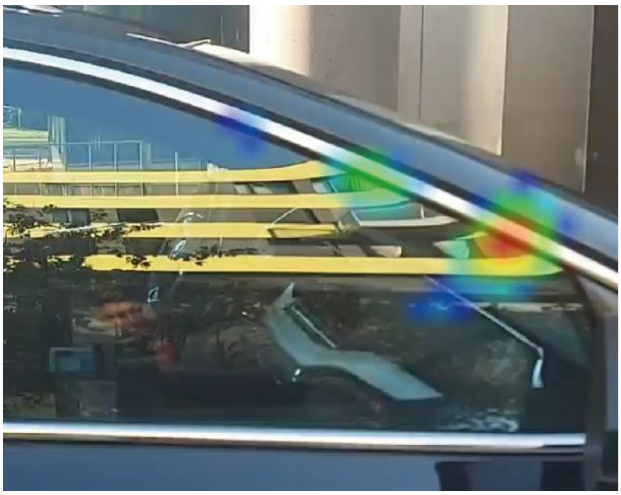

Figure 12. Multiple leakages along the front passenger's window seal at 48,7dB with a dynamic range of 3dB.

5 Summary and conlusions

This contribution addressed the topic of leakage visualization in automotive structures based on a new sensor concept which involves a rotating sensor with multiple ultrasound microphones scanning a circular area. The underlying measurement method is derived and its performance properties are described. A measurement setup comprising the above mentioned sensor, an ultrasound transmitter, a mobile device and high performance laptop is used to produce acoustic images for quality control in a series production and product development context.

6 References

[1] A. Pregelj, M. Drab, „Leak Detection Methods and Defining the Sizes of Leaks“, The 4th International Conference of Slovenian Society for Nondestructive Testing "Application of Contemporary Nondestructive Testing in Engineering" 24 - 25 April 1997, Ljubljana, Slovenia.

[2] Human Factors Guidance for the Use of Handheld, Portable, and Wearable Computing Devices https://hf.tc.faa.gov/publications/2005-human-factors-guidance-for-the-use-of-handheld/full_text.pdf

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

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

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

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