Acoustic Imaging of Sound Sources with a student-designed Acoustic Camera

Joachim Sigl and René Scheucher
Department of Automotive Engineering
FH Joanneum University of Applied Sciences
Alte Poststr. 149, A-8020 Graz, AUSTRIA

ABSTRACT

An acoustic camera consists of a microphone array, a data recorder and sound analysis- and -visualization software. It creates a color-coded sound map that displays the sound sources overlaid on the visual image of the recorded object. The sound maps are usually produced by analyzing the phase differences of the signals measured by the array microphones. Delay-and-sum beamformer and multiple signal classification (MUSIC) techniques are used in this work for the localization of sound sources. Beamformers are able to determine the amplitude of incident sound, but suffer from poor resolution and from ghost images. MUSIC, on the other hand, is an established technique for efficient and accurate noise source location, which can provide high-resolution source maps, but does not provide any information about the sound level. The combination of both methods gives comprehensive information about the acoustic emission of the system under investigation.

I.   INTRODUCTION

The detection of acoustic sources of plant and machinery or vehicles was usually done by hand with level measuring instruments or sound intensity probes. This manual scanning is nowadays replaced by a stationary array of microphones (e.g. see [1]) that allows the examination of complete structures within one measurement. The output of the array usually results in a noise map, which can be overlaid on a digital image of the test object.

The main disadvantage of this technique is the costs of commercially available microphone arrays. For that reason the authors were entrusted by their supervisors with the development of a home-made acoustic camera within the scope of a student project in their junior year. The system should be portable, allowing for outdoor data collection, and applicable for off-line and real-time signal processing tasks.

The defining features of an acoustic camera essentially are its ability to estimate the direction of arrival (DOA) and the intensity of the incident sound waves. The fundamental principle behind DOA estimation is the use of the phase information present in the acoustic signals picked up by microphones. Due to the microphones' spatial separation the signals arrive time-delayed. For known array geometries these time delays deliver with the help of sophisticated algorithms the DOA of the signal (e.g. see [2]).

The frequency content of the incident signal imposes a restriction on the array geometry. When the phase delay between signals from a pair of microphones exceeds half of the wavelength, spatial aliasing will occur. This causes a wrong interpretation of the time delays, which in turn results in a wrong DOA estimation. This leads to the condition that the spacing s between any pair of microphones in the

array should not exceed half of the smallest wavelength present in the signal

$$s \leq \frac{\lambda_{\min}}{2}. \qquad (1)$$

Several methods are used for processing the signals' phase information for DOA estimation. Beamformer-based methods combine the microphone signals in such a way that the array output emphasizes signals from a certain direction. The power of the array output is high if the direction of incident sound is in "look"-direction and it is attenuated if the sound wave arrives from somewhere else. In delay-and-sum beamformers the signals of each channel are time shifted appropriately to compensate for the delays of the incident signal and subsequently summed up (see Section III).

Subspace based methods divide the cross-correlation matrix of the array signals into signal and noise subspaces using eigenvalue decomposition to perform DOA estimation (see chapter IV). These methods are able to distinguish multiple sources that are located close to each other much better than the beamformer-based methods.

II. MICROPHONE ARRAY AND INSTRUMENTATION SYSTEM

Part of the project was the development of a stand as a holding device for the 16-channel microphone array. It should be portable, stable and dismountable, and have flexible sensor arrangement. The implemented design has a maximum size of 2 m x 2 m (see Figure 1) which gives a maximum sensor spacing of 0.67 m and a minimum of 20 mm. These dimensions correspond according to (1) to an effective frequency range from 250 Hz up to 7 kHz. For measurements in the low frequency range the array can be enlarged up to 4 m by making use of an additional stand and thus enables measurements down to a frequency of 140 Hz. 16 ¼" measuring microphones (Roga MI-17) were used as sound pressure transducers. These microphones need a supply current of 2 - 6 mA, which was delivered by the DEWE-3010-IPC front end. This front end is also used for data acquisition. All the microphone
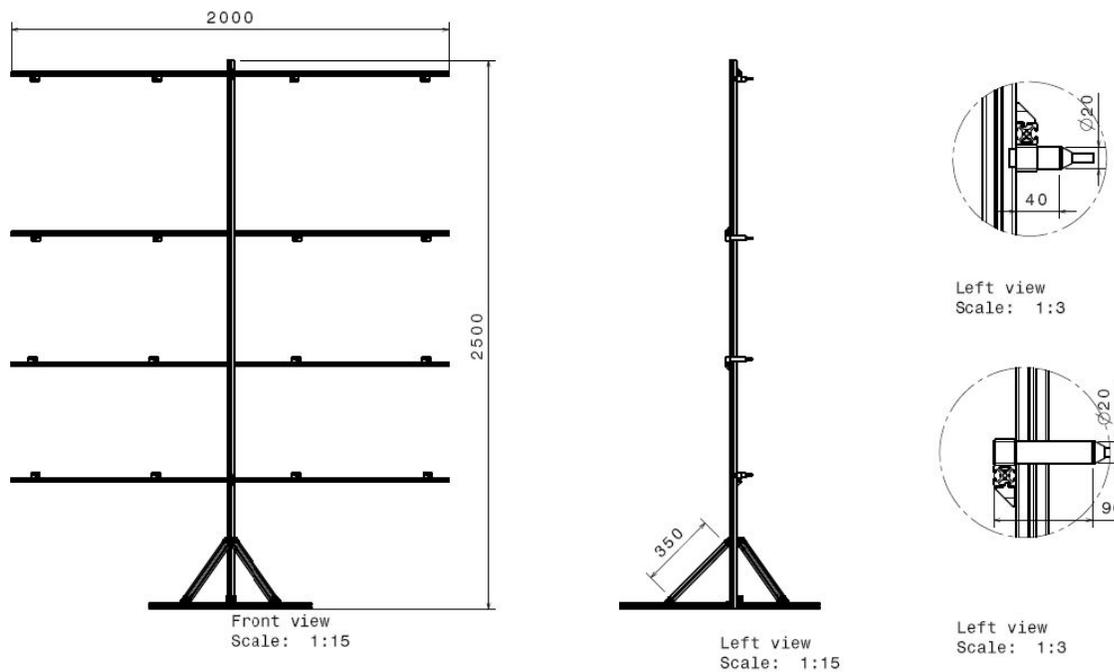
Figure 1. Design drawing of the adaptable microphone array with detailed microphone clamping system

channels have to be sampled simultaneously since the relative phase information present in the signals is crucial for the DOA estimation.

III. DELAY-AND-SUM BEAMFORMER

The job of a beamformer is to enhance signals from a particular direction and to attenuate signals from other directions. This is done by making use of a set of spatially separated microphones and selecting a direction from which to accept signals, while rejecting signals from other directions. In this project a narrowband beamformer is implemented, which assumes that the incident signal that the beamformer is trying to capture has a narrow bandwidth centered at a particular frequency.

The array consist of N (= 16) microphones at the locations $\mathbf{p}_n$, for n = 1, 2,…, N (as indicated in Figure 2). These microphones produce a set of signals denoted by the vector

$$\mathbf{f}(t,\mathbf{p}) = \begin{pmatrix} f_1(t,\mathbf{p}_1) \\ f_2(t,\mathbf{p}_2) \\ \vdots \\ f_N(t,\mathbf{p}_N) \end{pmatrix} \qquad (2)$$

with $\mathbf{p} = (\mathbf{p}_1,\mathbf{p}_2,\ldots,\mathbf{p}_N)^T$ and $\mathbf{p}_n = (\mathbf{p}_{x_n},\mathbf{p}_{y_n},\mathbf{p}_{z_n})^T$.

When a plane wave is propagating in the direction

$$\mathbf{a}(\vartheta,\varphi) = \begin{pmatrix} -\sin(\vartheta)\cos(\varphi) \\ -\sin(\vartheta)\sin(\varphi) \\ -\cos(\vartheta) \end{pmatrix} \qquad (3)$$

towards the centre of the microphone array, $\mathbf{a}(\vartheta,\varphi)$ represents the direction of incidence in a spherical coordinate system centered in the origin of the array. If all the microphones are assumed to be identical and the distance between the source and the array is assumed to be large (far-field assumption), then the gains of each microphone are equal. Then the signals at the microphones are delayed with respect to a reference microphone. If the correct set of time delays τ_n is chosen the sum of the microphone signals reaches a maximum. Thus the sum of all microphone signals divided by the number of microphones, the array output $y(t)$, represents the sound pressure of an incident plane wave
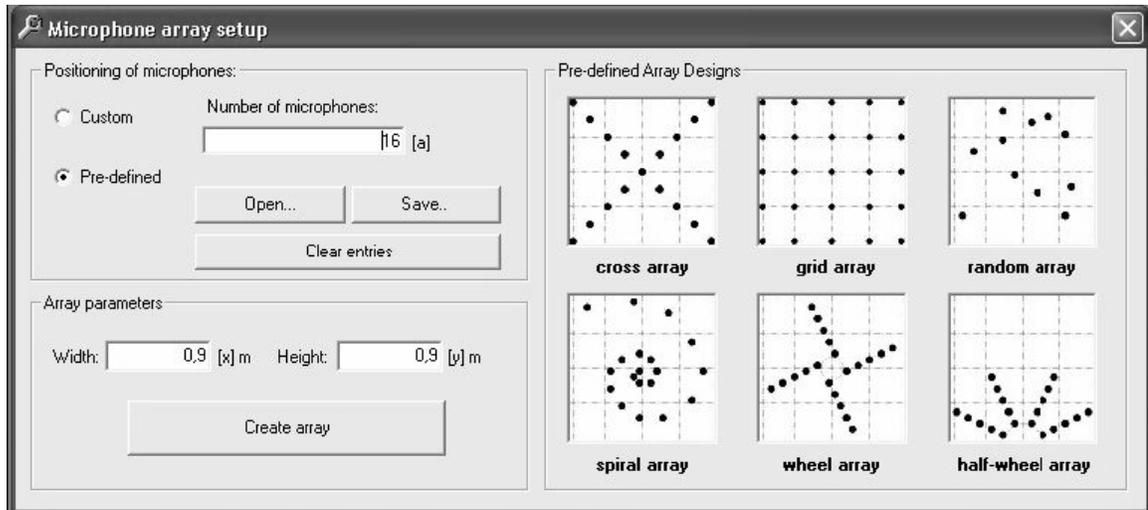

Figure 2. User interface for the setup of the microphone array.

$$y(t) = \frac{1}{N}\sum_{n=1}^{N} f_n(t) = \frac{1}{N}\sum_{n=1}^{N} f(t - \tau_n).$$

$$(5)$$

Every other set of time delays leads to an attenuated array output, which in turn allows an estimation of the direction of arrival. The fundamental principle behind this DOA estimation is to use the phase

information present in the signals picked up by microphones that are spatially separated. Preceding transformation of time signals into frequency domain is most suitable for the processing of phase information.

The frequency domain representation of the sound signals can be obtained by taking the Fourier transform of (4)

$$\mathbf{F}(\omega) = \begin{pmatrix} F_1(\omega) = \int_{-\infty}^{\infty} f(t - \tau_1)\,e^{-i\omega t}\,dt \\ F_2(\omega) = \int_{-\infty}^{\infty} f(t - \tau_2)\,e^{-i\omega t}\,dt \\ \vdots \\ F_N(\omega) = \int_{-\infty}^{\infty} f(t - \tau_N)\,e^{-i\omega t}\,dt \end{pmatrix} = \begin{pmatrix} F(\omega)\,e^{-i\omega\tau_1} \\ F(\omega)\,e^{-i\omega\tau_2} \\ \vdots \\ F(\omega)\,e^{-i\omega\tau_N} \end{pmatrix}. \qquad (6)$$

With the wave number $\mathbf{k} = \dfrac{\omega}{c}\mathbf{a} = \dfrac{2\pi}{\lambda}\mathbf{a}$ an array manifold vector can be defined, which incorporates all of the spatial characteristics of an array

$$\mathbf{v_k} = \begin{pmatrix} e^{-i\omega\tau_1} \\ e^{-i\omega\tau_2} \\ \vdots \\ e^{-i\omega\tau_N} \end{pmatrix} = \begin{pmatrix} e^{-i\mathbf{k}\cdot\mathbf{p}_1} \\ e^{-i\mathbf{k}\cdot\mathbf{p}_2} \\ \vdots \\ e^{-i\mathbf{k}\cdot\mathbf{p}_N} \end{pmatrix}. \qquad (7)$$

Thus we have

$$\mathbf{F}(\omega) = F(\omega)\,\mathbf{v_k}. \qquad (8)$$

Those frequency coefficients are multiplied by appropriate and generally complex weights and then summed up to get the frequency domain representation of the array output:

$$Y(\omega) = \mathbf{W}^* \mathbf{F}(\omega). \qquad (9)$$

The weight vector

$$\mathbf{W} = (W_1, W_2, \ldots, W_N)^T \qquad (10)$$

is in the present case simply $W_j = 1/N$ for all j.

The power of the beamformer output in the frequency domain, the array output power spectral density, can be written as

$$P(\omega) = Y(\omega)^2 = \big(\mathbf{W}^*\mathbf{F}(\omega)\big)\big(\mathbf{W}^*\mathbf{F}(\omega)\big)^* = \mathbf{W}^*\big(\mathbf{F}(\omega)\mathbf{F}(\omega)^*\big)\mathbf{W} = \mathbf{W}^*\mathbf{R}(\omega)\mathbf{W} \qquad (11)$$

where $\mathbf{R}(\omega)$ is a N by N matrix representing the cross power spectral densities of the channel input signals.

In order to localise the sound sources the array response function is introduced. This function is a generalised form of a frequency

response, since it represents the response of a microphone array to a plane wave of frequency ω incident on the array at an arbitrary $\mathbf{a}(\vartheta, \varphi)$. In steered delay-and-sum beam formers the weights (10) are given by

$$\mathbf{W} = \mathbf{v_k}(\vartheta,\varphi) = \frac{1}{N}\begin{pmatrix} e^{-i\mathbf{k}\cdot\mathbf{p}_1} \\ e^{-i\mathbf{k}\cdot\mathbf{p}_2} \\ \vdots \\ e^{-i\mathbf{k}\cdot\mathbf{p}_N} \end{pmatrix} = \frac{1}{N}\begin{pmatrix} e^{-i\omega\tau_1} \\ e^{-i\omega\tau_2} \\ \vdots \\ e^{-i\omega\tau_N} \end{pmatrix}, \tag{12}$$

which leads to an array output power spectral density of

$$P(\omega,\vartheta,\varphi) = \left(\mathbf{v_k}(\vartheta,\varphi)^*\mathbf{F}(\omega)\right)\left(\mathbf{v_k}(\vartheta,\varphi)^*\mathbf{F}(\omega)\right)^* = \mathbf{v_k}(\vartheta,\varphi)^*\mathbf{R}(\omega)\mathbf{v_k}(\vartheta,\varphi). \tag{13}$$

where $P(\omega,\vartheta,\varphi)$ represents the squared intensity of sound for a narrowband input signal of frequency ω with the direction of arrival ϑ and φ.

IV.   MULTIPLE SIGNAL CHARACTERIZATION

Multiple Signal Classification (MUSIC) [3] is one of the most popular subspace based narrowband methods. In this method the cross power spectral density matrix $\mathbf{R}(\omega)$ is decomposed into eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$ and corresponding eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N$. If K sound source are monitored by a microphone array consisting of N microphones, then the eigenvalue decomposition of the $N \times N$ matrix $\mathbf{R}$ decomposes the N-dimensional space of the matrix into a K-dimensional signal subspace and, if the eigenvalues $\lambda_{K+1} \geq \lambda_{K+2} \geq \ldots \geq \lambda_N$ are all zero, a (N-K)-dimensional null space, which are both spanned by the corresponding eigenvectors. Alternatively, if $\lambda_{K+1} \geq \lambda_{K+2} \geq \ldots \geq \lambda_N$ are merely very small, below the noise level of the system represented by $\mathbf{R}$, the eigenvectors $\mathbf{v}_{K+1}$, $\mathbf{v}_{K+2}, \ldots, \mathbf{v}_N$ span the noise subspace of $\mathbf{R}$. The crucial idea of MUSIC is that, since $\mathbf{R}$ is self-adjoint, the signal subspace and the noise subspace are orthogonal to each other.

If we now consider any arbitrary vector $\mathbf{v}$, the Euclidean distance of $\mathbf{v}$ from the signal subspace is the length of the projection of $\mathbf{v}$ into the noise subspace. The squared magnitude of this distance is then given by

$$d(\mathbf{v})^2 = \sum_{j=K+1}^{N}\left(\mathbf{v}_j^H\mathbf{v}\right)^*\left(\mathbf{v}_j^H\mathbf{v}\right) \tag{14}$$

where $\mathbf{v}_j$ represents the jth eigenvector of $\mathbf{R}$. In equation (14) the eigenvectors that span the noise subspace are used to compute the distance of $\mathbf{v}$ from the signal subspace. A vector $\mathbf{v}$ that lies entirely in the signal subspace minimizes this squared distance. Another way to minimize the latter is to maximize the reciprocal of the squared distance (e.g. see [4]):

$$P(\mathbf{v}) = \frac{1}{d(\mathbf{v})^2}. \tag{15}$$

If $\mathbf{v}$ belongs to the signal subspace, then the distance d is zero and $P(\mathbf{v})$ approaches infinity. Due to the non-vanishing eigenvalues of the noise subspace, $P(\mathbf{v})$ goes to a very large value.

The procedure just described reveals one of the disadvantages of the MUSIC algorithm. It requires a priori knowledge of the number of sources in order to estimate the dimension of the signal subspace.

The array manifold vector $\mathbf{v_k}(\vartheta,\varphi)$ (12) represents the spatial sampling of a plane wave with wave number $\mathbf{k}$ arriving from $\mathbf{a}(\vartheta,\varphi)$. Thus if ϑ and φ happen to
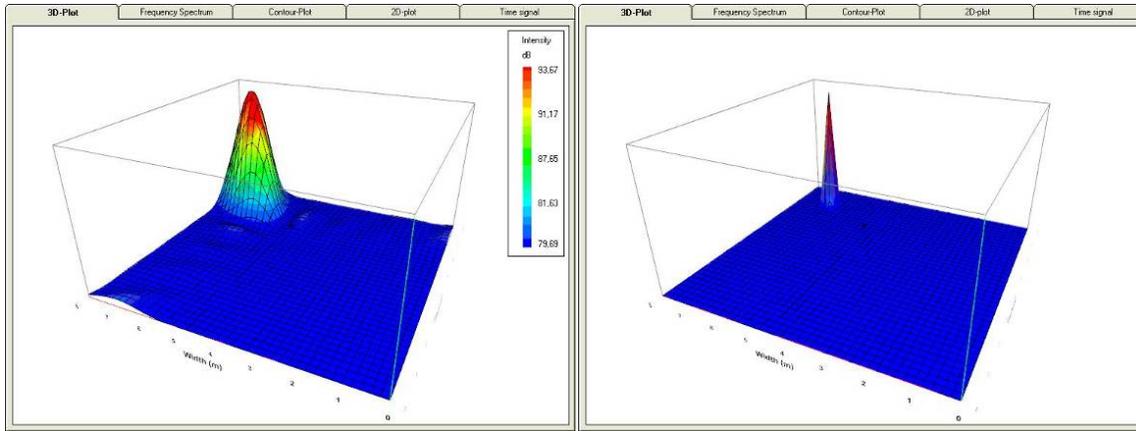
Figure 3. The left picture represents the beamformer's response; the right picture the MUSIC algorithm's output of a single sound source.
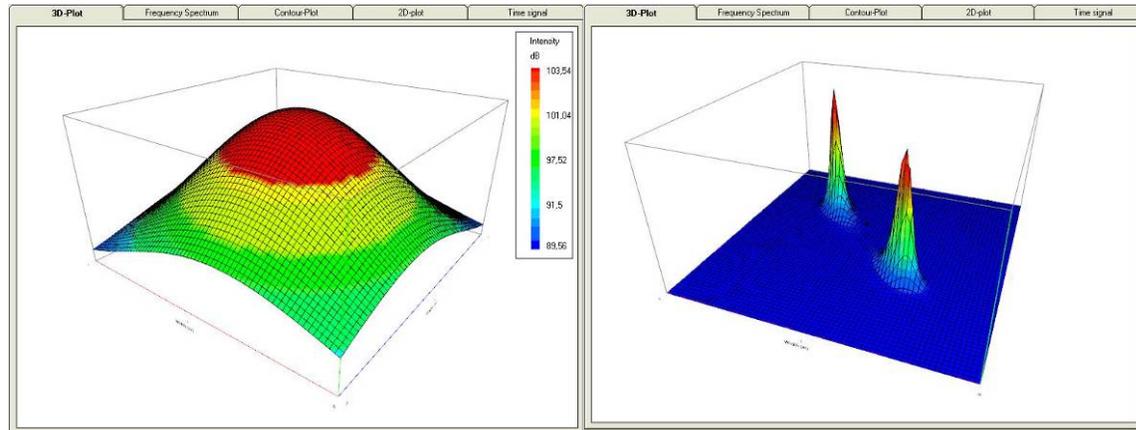
Figure 4. The beamformer cannot distinguish the two sound sources (left picture). The MUSIC algorithm (right picture) clearly recognizes two different sound sources.

be the incident angles of arrival, $\mathbf{v_k}(\vartheta, \varphi)$ belongs to the signal subspace and thus $P(\mathbf{v_k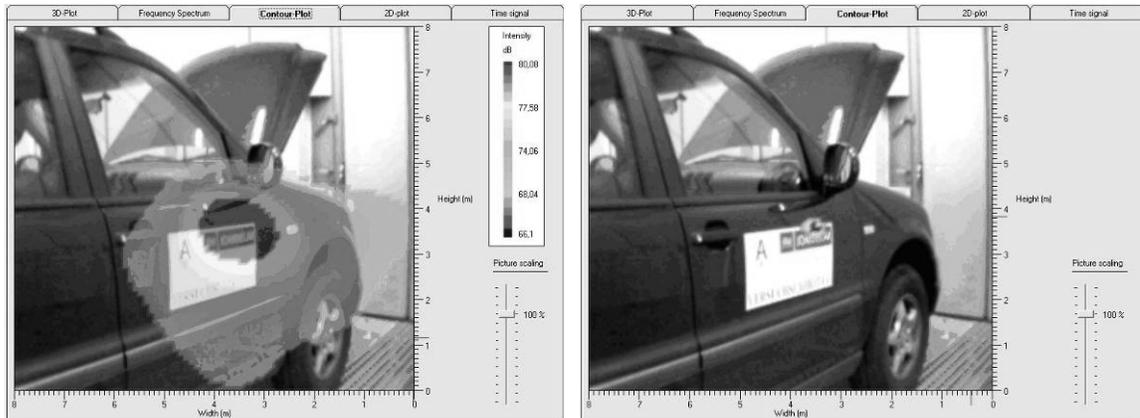}(\vartheta, \varphi))$ approaches a large value. If $P(\mathbf{v_k}(\vartheta, \varphi))$ is computed for all possible angles of arrival

$$P(\mathbf{v_k}(\vartheta, \varphi)) = \frac{1}{\sum_{j=K+1}^{N}\left(\mathbf{v}_j^H \mathbf{v_k}(\vartheta, \varphi)\right)^*\left(\mathbf{v}_j^H \mathbf{v_k}(\vartheta, \varphi)\right)} \tag{16}$$

then the true angle of arrival produces a sharp peak in $P(\mathbf{v_k}(\vartheta, \varphi))$ and this feature can be used to determine the position of the sound sources.

V.   IMPLEMENTATION OF
     ALGORITHMS

The program consists of five main modules, which are

6

Figure 5. Noise emissions of an SUV's combustion engine as measured by the microphone array. In the left picture delay-and-sum beamforming and in the right picture the MUSIC algorithm was employed. (For a full-color version of this figure, see the cover illustration for this issue.)

- Input of measured or generation of simulated time signals
- Fast Fourier Transform (FFT)
- Delay-and-Sum Beamforming
- Calculation of eigenvalues and eigenvectors of spectral density matrix
- Direction of Arrival estimation by Multiple Signal Classification.

The time signal is either measured or generated by the program in its simulation mode. If the data are sampled via the front-end device the application offers a comfortable read-in algorithm that allows easy data handling. All data are transferred into a flex-grid sheet for inspection.

While the user interface is programmed in Visual Basic (VB), the computations are performed by dynamic link libraries (DLL) written in ANSI C. Calling C routines from VB provides a convenient way to quickly develop graphical user interfaces and packages that use mathematical routines (see e.g. [5]).

The user can set up the microphone array either by making use of pre-defined array configurations (see Figure 2) or by placing the sensors individually (random array option). The 16 sampled or generated time signals are then transformed into the frequency domain by FFT and a number of array manifold vectors (7) are generated according to the requested resolution of the incident angles ϑ and φ. For the beamforming part of the program the cross power spectral density matrix $\mathbf{R}(\omega)$ is established and with it the squared intensity of sound for a narrowband input signal of frequency ω, the output power spectral density array $P(\omega, \vartheta, \varphi)$, is calculated (13). It is used for the illustration of the frequency dependent sound intensity.

For the MUSIC algorithm the Hermitian matrix $\mathbf{R}(\omega)$ is first reduced into tridiagonal form by Householder reduction [6] and then decomposed into its eigenvectors. Then the array $P(\mathbf{v_k}(\vartheta, \varphi))$ (16) is calculated and provided for the visualisation part of the program. In order to facilitate the comparison of the measured sound pattern with the sound sources of the recorded object a webcam is employed. The digital picture produced by the camera is displayed together with the contour plots of the beamformer's and MUSIC algorithm's outputs.

VI.   CASE STUDIES

A single sound source with a given frequency and intensity is generated in the upper left region of the observation window with the help of the program's simulation mode (see Figure 3). In the left picture the beamformer's response to the signal is illustrated, which gives both information about the position of the source and the amplitude of the incoming sound wave (color code on the right hand side of the picture). On the other hand, the MUSIC algorithm's output (right picture in Figure 3) contains no

information about the sound intensity. Nevertheless, a comparison of the two pictures shows the supremacy of MUSIC in terms of spatial resolution.

This circumstance is illustrated even more impressively in Figure 4. Two sound sources having the same frequency were generated in the simulation mode close to each other (equivalent to a distance of about half a meter in a real measurement). In the dedicated frequency range the beamformer is obviously not able to resolve the two sources (Figure 4, left picture) and produces a blurred sound intensity pattern covering the locations of both sources. The MUSIC algorithm (Figure 4, right picture) on the other hand shows its ability to identify two sources with same frequency and amplitude, but suffers from the shortcoming of giving no information about the sound intensity. Apparently only the combined output of both algorithms gives detailed information about the acoustic emission of the test object.

In another study involving real measurements, the combustion engine of a sports utility vehicle (SUV) was used as sound source. Although the V6 Diesel engine running at ~2500 rpm was hidden in the engine compartment the microphone array could localize it (Figure 5). The measurements were taken in the department's anechoic chamber.

VII.   CONCLUSION

In this paper the development of a home-made acoustic camera was presented. At the start of this student project 16 microphones and a data acquisition front end with the ability of sampling 16 signals simultaneously were available. In order to build an acoustic camera out of it, the hardware for the microphone array and the software for the generation of an acoustic image had to be developed. The portable stand for the array was designed to be continuously adjustable for the prevention of spatial aliasing in a broad frequency range. For the acoustic imaging two different algorithms were employed. Delay-and-sum beamforming was used for the determination of the intensity of the incident sound as a function of the direction of arrival. Multiple signal classification was additionally implemented in order to enhance the microphone array's

spatial resolution. Case studies on both virtual and real sound sources confirmed the functionality of the acoustic camera.

REFERENCES

1. M. Brandstein, *Microphone Arrays*, Springer Verlag, Berlin (2001).
2. P.-J.Chung, "Fast algorithms for parameter estimation of sensor array signals," PhD thesis, Ruhr-Universität Bochum, Bochumer Universitätsverlag (2002)
3. R.O.Schmidt, "Multiple emitter location and signal parameter estimation," *Proceedings of the RADC Spectral Estimation Workshop*, pp. 243-258 (1979).
4. K.Varma, "Time-Delay-Estimate Based Direction-of-Arrival Estimation for Speech in Reverberant Environments," Master's Thesis, Virginia Polytechnic Institute and State University (2002).
5. A.Harrich, S.Jagsch, S.Riedler and W.Rosinger, "Computerized Data Reduction and Analysis in Positron Annihilation Doppler Broadening Coincidence Spectroscopy," *American Journal of Undergraduate Research*, Vol. **2**, No. 3, 13 (2003)
6. W.H.Press, S.A.Teukolsky, W.T.Vetterling and B.P.Flannery, *Numerical Recipes in C – The Art of Scientific Computing*, Cambridge University Press (1988)