BeBeC-2022-D14

ACOUSTIC INVESTIGATION OF WAVE-FIELD-SYNTHESISED VIRTUAL SOUND SOURCES USING A 3D MICROPHONE ARRAY

Luca Liebich, Adrian Philipp, Michael Strobl, Tim-Niklas Wennemann and Gert Herold
TU Berlin, FG Technische Akustik
Einsteinufer 25, 10587 Berlin, Germany

Abstract

Wave field synthesis is a sound reproduction method that attempts to create a predefined sound field in space by driving multiple sound sources in such a way that the desired sound field results from the superposition of the individual sound sources. Due to practically necessary simplifications the synthesis of virtual sound sources comes with undesirable artefacts. These can be identified with the help of simulations, but have apparently not yet been investigated in detail with beamforming methods. In this work, synthesised point sources of a wave field synthesis system as well as real sound sources are recorded with the help of a 3D microphone array. The measurements are compared to each other and to simulated sound sources. For the measurement an L-shaped 3D microphone array with 64 microphones was used. To determine the limitations of the microphone array all sound source positions are also simulated. Two quantities are used to evaluate the sound sources: the deviation of the main lobe from the expected position, and the number of side lobes. In addition, heatmaps are used for qualitative comparison of the different sound sources. With the developed microphone array, virtual and real sound sources within the theoretical cutoff frequencies of the microphone array and the wave field synthesis system can be localized in the frontal plane comparably well.

1 INTRODUCTION

Wave field synthesis is a method of sound reproduction which attempts to generate a virtual sound field by driving a large number of sound sources in such a way that the desired sound field results from the superposition of the individual sound fields. Because of the compromises that must be made for practical implementation, artefacts occur in the synthesis of virtual sound fields. The resulting limitations of wave field synthesis can be determined using simulations.

However, a practical investigation of the sound field using beamforming methods is still largely unexplored. In this work, both synthesized point sources of a wave field synthesis system and real sound sources are measured with a microphone array and are then compared with each other and with simulated sound sources.

1.1 Wave field synthesis

Wave field synthesis was initiated by Berkhout [2] in 1988. It is based on the Huygens' principle [5] and represents an approach to reproduce sound fields spatially. With wave field synthesis, the physically correct reproduction of sound fields is possible. The requirement for wave field synthesis (WFS) is described by Ahrens [1] as follows:

> ”A given ensemble of elementary sound sources shall be driven such that the superposition of the sound fields emitted by the individual elementary sound sources makes up a sound field with given desired properties over an extended area.”

These extended areas can be surfaces or volumes. *Secondary sources* has become the accepted term for elementary sound sources. An ensemble of secondary sources is often called an *array*. Practically, the secondary sources cannot have infinitely small dimensions and therefore cannot be continuously distributed. The loudspeakers are positioned at discrete positions, which in turn means that there is a spatial discretization. This discretization leads to location, frequency and angle dependent errors in the synthesized sound field. These errors are often called *spatial aliasing*.

For the driver function of a linear array the following anti-aliasing condition can be specified

$$f \leq \frac{c}{\Delta x \left(1 + \left|\cos \alpha_{\mathrm{pw}}\right|\right)} \tag{1}$$

Δx is the distance between the individual secondary sources and α_{pw} the radiation angle of the plane wave [10]. In the WFS Modules that are used for this paper [3], the drivers responsible for the upper frequency range are installed at a distance of 10 cm. If this distance is used in Eq. 1 and a plane wave from normal direction is assumed, this results in a cutoff frequency for spatial aliasing of 3430 Hz.

1.2 Beamforming

The goal of microphone array methods is to develop a sensor that can spatially separate sound sources. The main lobe of the directivity pattern of an array is called a *beam*. The beamforming algorithm aligns the spatial filter of the array in the desired directions [6]. This is achieved by simultaneously recording sound pressures at spatially distributed sensor positions and then phase shifting the signals according to sound travel times from the sensors to assumed source positions on a focal grid. The summation of the shifted signals yields a sound pressure at grid points where a source is present, and cancels out at points without a source x[4].
The beamforming equation for the frequency domain is:

$$b\left(x_t\right) = h^{\mathrm{H}}\left(x_t\right) C h\left(x_t\right), \quad t = 1 \ldots N$$

The number of focus points is denoted as N. The *steering vector* h contains the phase shift corresponding to the travel times from the focal point to each microphone as well as the level correction according to the monopole sound propagation model [7].
For the cross-spectral matrix C, the measured time signals are processed using Welch's method [11]. The signal is divided into blocks and transformed to the frequency domain using FFT. The cross spectral matrix can then be calculated by averaging the cross spectra of the recorded channels,

$$C = \overline{pp^{\mathrm{H}}}$$

where H represents Hermitian (conjugate) transpose and p contains the complex spectral sound pressures at each microphone position [4].

2 METHODS

2.1 Microphone array geometry

The microphone array geometry was calculated using the spiral equation presented in [8].

$$r = R\sqrt{\frac{m}{M}}, \qquad m = 1, 2, ..., M \tag{2}$$

$$\phi = 2\pi m \frac{(1+\sqrt{V})}{2} \tag{3}$$

with M the total amount of Microphones, R the radius of the spiral and r and ϕ as polar coordinates. To achieve an even distribution of microphones over all directions V was chosen to be $V = 5$. To prevent reflections from the supporting structure of the microphone array as much as

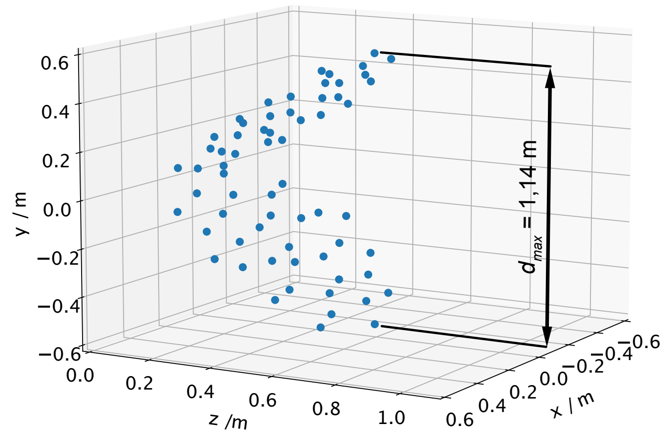


Figure 1: Used microphone positions in a three dimensional coordinate system with a maximum distance between two microphones of $d_{max} = 1.14\,m$.

.

possible, an aviary grid was used as a sound-permeable structure. Due to the maximum microphone distance of $d_{max} = 1.14\,\text{m}$, the lower cutoff frequency of the array can be determined as $f_{min} = 301\,\text{Hz}$ with $f_{min} = \frac{c}{d_{max}}$ and an assumed sound velocity of $c = 343\,\text{m/s}$. This means that sound sources with a frequency lower than the cutoff frequency cannot be reliably positioned with the microphone array used.

2.2 Measurements

The measured WFS system is installed at the Technical University Berlin. As shown in Fig. 2, the WFS modules are arranged in a 2.5 dimensional ring-like geometry [1] around the entire perimeter of the room. Due to the multiple secondary sources along the x-axis the WFS system is capable of creating different virtual sound fields through the superposition of multiple elementary sound fields. In the top view of Fig. 3 for example the sound field of a virtual sound source that is located behind the WFS system is visualised. Nonetheless, this virtual sound field is only valid in the xz-plane. As can be seen in the side view of Fig. 3, the WFS system does not have multiple secondary sources along the y-axis, hence it cannot create virtual sound fields in the yz-plane. The microphone array was positioned in the centre of the WFS system facing the front of the room. Sound sources were positioned in front of the microphone array. A mobile dodecahedron loudspeaker served as a real sound source. A virtual sound source was

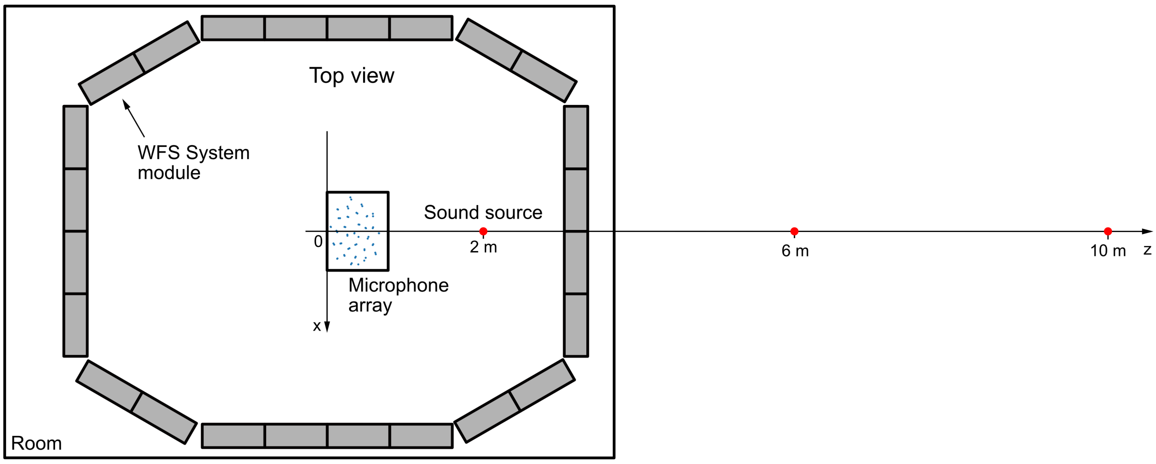


Figure 2: Proportional drawing of the WFS system in top view at the TU Berlin with the position of the microphone array as well as the sound sources at distances of 2 m, 6 m and 10 m.

measured centrally in front of the microphone array at distances of 2 m, 6 m and 10 m. The distance of 2 m was located inside the WFS system and the two distances with 6 m and 10 m were located outside. Due to the spatial limitations of the room, the dodecahedron as a real sound source was measured only at the distance of 2 m. All positions were also displayed and evaluated using a simulated source generated with the Python library Acoular [9]. In order to be able to evaluate the entire auditory spectrum, a continuous white noise over a duration of 60 s is used as the measurement signal and the sound pressure level was subsequently averaged over the entire duration. For the computation of the measurements, the Python library Acoular

[9] is also used. By beamforming in the frequency domain, the time signals were converted into sound source distributions, on which the further evaluation was based. The FFT block size is 4096 samples for all beamforming computations. For the display of sound source distributions of single frequencies, the focus grids are resolved with 1 cm. In total, a frequency range of 50-20000 Hz is evaluated in one-third octave bands in order to cover approximately the entire audible spectrum.

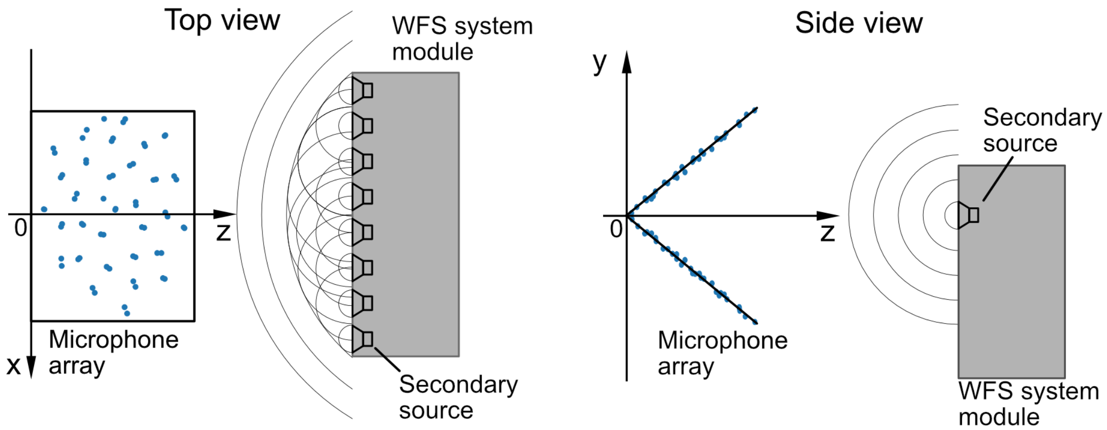


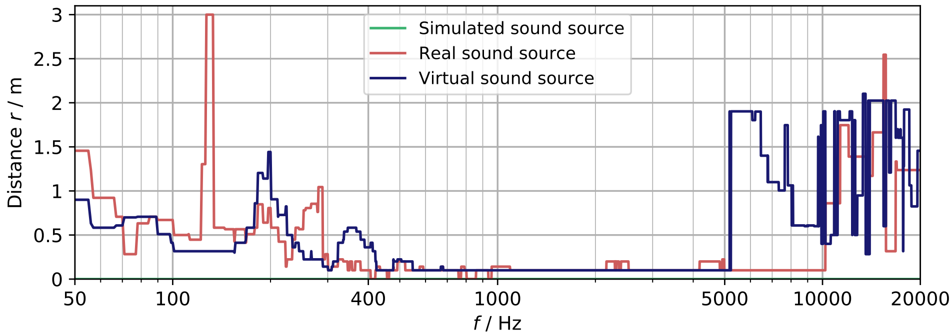
Figure 3: Schematic depiction of the orientation of the microphone array and the WFS system. The top view shows a spherical virtual sound field created by the WFS system for a virtual source behind the WFS system in the xz-plane. The side view shows the spherical sound field created by the secondary sources.

3 RESULTS

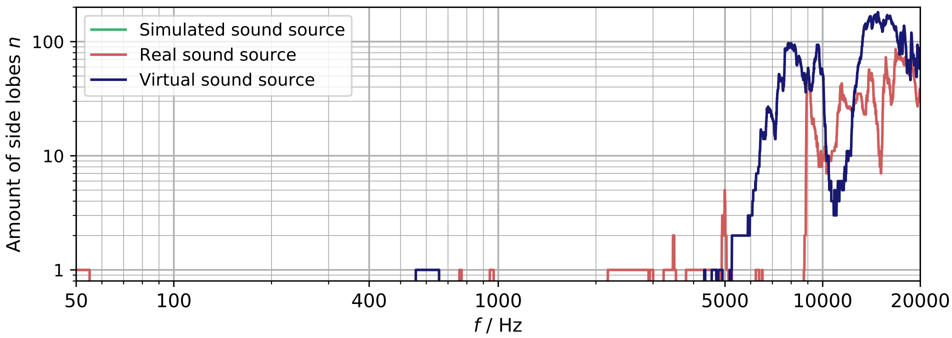
The results are divided into the front view and the top view. In the front view, the sound source is located in front of the microphone array in the xy-plane. The top view, as visualised in Fig. 2 and 3, shows the xz-plane. Two parameters are defined for the quantitative evaluation of both planes. On the one hand, the measured and simulated positions of the main lobes are compared with the expected position of the sound sources, which results in a distance r in meters. The position of the main lobe is determined by the maximum sound pressure level. On the other hand, the number of side lobes is determined, with the aim of examining the wave field synthesis system for spatial aliasing effects. For this purpose, the number of relative maxima with a sound pressure level of $L_P > L_{P\max} - 3\,\mathrm{dB}$, i.e. the amount of side lobes n, within the beamforming results are counted in dependence of the frequency.

3.1 Front view (xy-plane)

Fig. 4a shows the distance from the position where a sound source is expected to the position where the maximum sound pressure level is determined in respect to the frequency with an accuracy of 0.1 m. In the evaluated measurements of the real and virtual sound source, the position of the sound source in each case is at a distance of 2 m in front of the microphone array. Accordingly, the focal plane for the beamforming is also set at 2 m. Above 300 Hz, the real sound source can be localised with a tolerance of 0.2 m distance. For the virtual sound source, this is possible for frequencies above 400 Hz. Above a frequency of 5 kHz, the deviation from the expected position for the virtual sound source increases abruptly to over 2 m and remains at a strong deviation above this frequency. The real sound source deviates strongly from the expected value above a frequency of 10 kHz. Fig. 4b shows the number of sidelobes that are within $\Delta L_p = 3\,\mathrm{dB}$ range of the maximum sound pressure level of the measurement. From a frequency of 5 kHz the side lobes within the tolerance range increase visibly for both the real and the virtual sound source.



(a) Distance r from the position of the maximum sound pressure level to the expected position of the sound source



(b) Number of side lobes n with a sound pressure level of $L_P > L_{Pmax} - 3\,\mathrm{dB}$ calculated in respect to the frequency

Figure 4: Frequency-dependent quantities for quantitative evaluation of measurements of real, virtual and simulated sound source at 2 m distance.

In order to visualise the limits of the system in addition to the quantities for quantitative evaluation of the measurement results, sound source distributions in the focal plane at 2 m are shown in Fig. 5. It can be seen that for a frequency of 200 Hz, the sound source cannot be localised at the expected position for either the real or the virtual sound source. At a frequency of 1600 Hz, representative of the frequency range in which the localisation functions, the sound source can be localised well in all cases. Here, the images of the real, virtual and simulated sound source show strong similarities. In the sound source distributions of the real sound source, a white cross can be seen, which represents a determined relative maximum. To make the sound

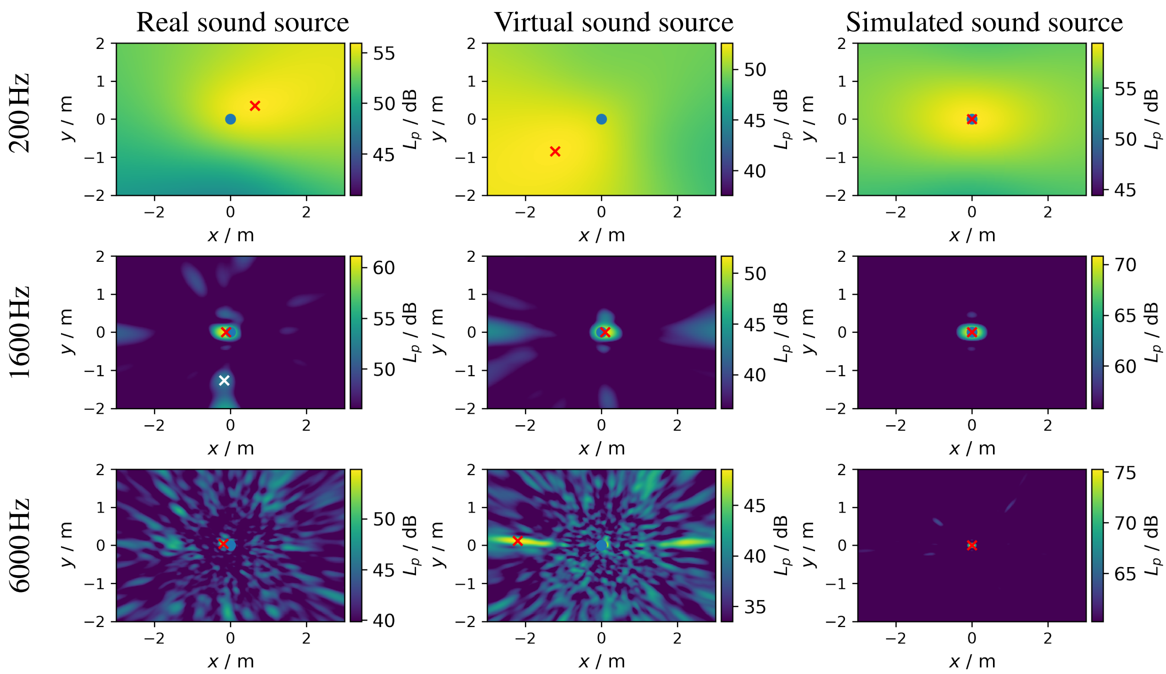
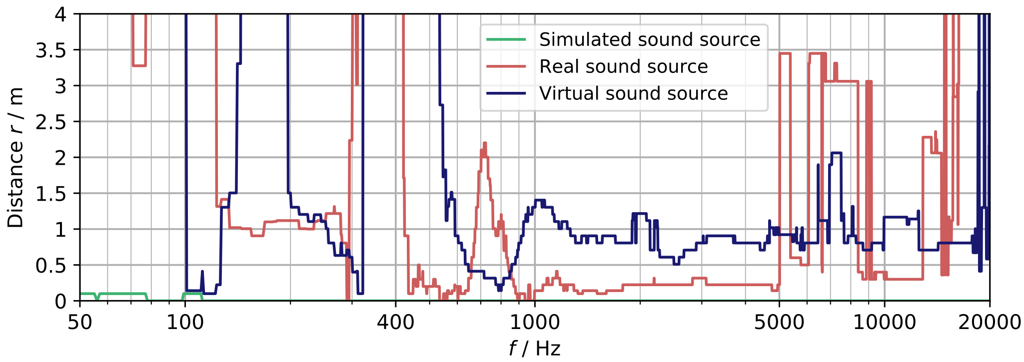


Figure 5: Sound source distributions of real and virtual sound source, as well as a simulation for three different frequencies in the front view (xy-plane). The blue dot symbolises the position of the expected sound source, the red cross shows the determined sound source with maximum sound pressure level in the indicated area.

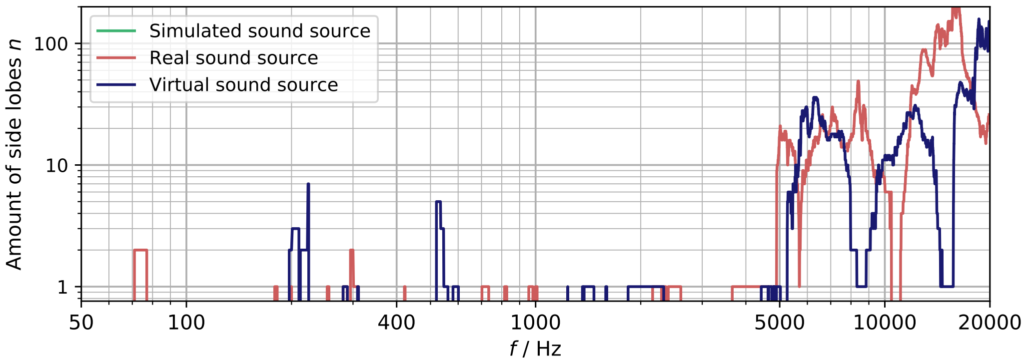
source distributions in Fig. 5 clearer, they are only shown exemplary at 1600 Hz of the real sound source. At 5 kHz the side lobes in the heat map of the virtual sound source dominate so that the true position can no longer be localised correctly. The localisation of the real sound source with the microphone array, however, is still possible. The height at which the virtual sound source is localised corresponds approximately to the height of the secondary sources.

3.2 Top view (xz-plane)

In addition to the evaluation in the front view, the measurement data was also evaluated in the top view, i.e. the room depth. Fig. 6a shows that, with a few exceptions, the virtual sound source cannot be localised correctly. For the real sound source, a frequency range between 1 kHz and 5 kHz emerges, in which it can be localised almost constantly with a relative distance of $r = 0.2$ m. Extreme outliers that exceed the limits of the illustration and occur with both the real and the virtual sound source are conspicuous. The number of side lobes shown in Fig. 6b increase above 5 kHz, similar to the front view. In the simulation, again no side lobes occur, which is why the characteristic curve cannot be seen on the logarithmised ordinate axis.
Fig. 7 shows the sound source distribution of three virtual sound sources in the third octave band



(a) Distance r from the position of the maximum sound pressure level to the expected position of the sound source



(b) Number of side lobes n with a sound pressure level of $L_P > L_{Pmax} - 3\,\mathrm{dB}$ *calculated in respect to the frequency*

Figure 6: Frequency-dependent quantities for quantitative evaluation of measurements of real, virtual and simulated sound source at 2 m distance.

with the centre frequency of 1600 Hz. The figure illustrates that the determined intersection point with maximum sound pressure level along the z-axis does not correspond to the expected position. It is noticeable that the maximum sound pressure level is determined at a distance of 3-4 m, which corresponds approximately to the distance of the secondary sources.
Fig. 8 also shows sound sources in the top view, but one real sound source at a distance of 2 m and two simulations for the distances 6 m and 10 m. Since these distances already go beyond the spatial dimensions of the room, the comparison cannot be made with real sound sources. In contrast to the virtual sound source shown in Fig. 8, the position of the maximum sound pressure level matches the expected position of the sound source.

4 DISCUSSION

Within the range between 400 Hz and 5000 Hz, a strong similarity of virtual and real sound source can be observed for the front view, seen in Fig. 4. From 5000 Hz, and up, many side

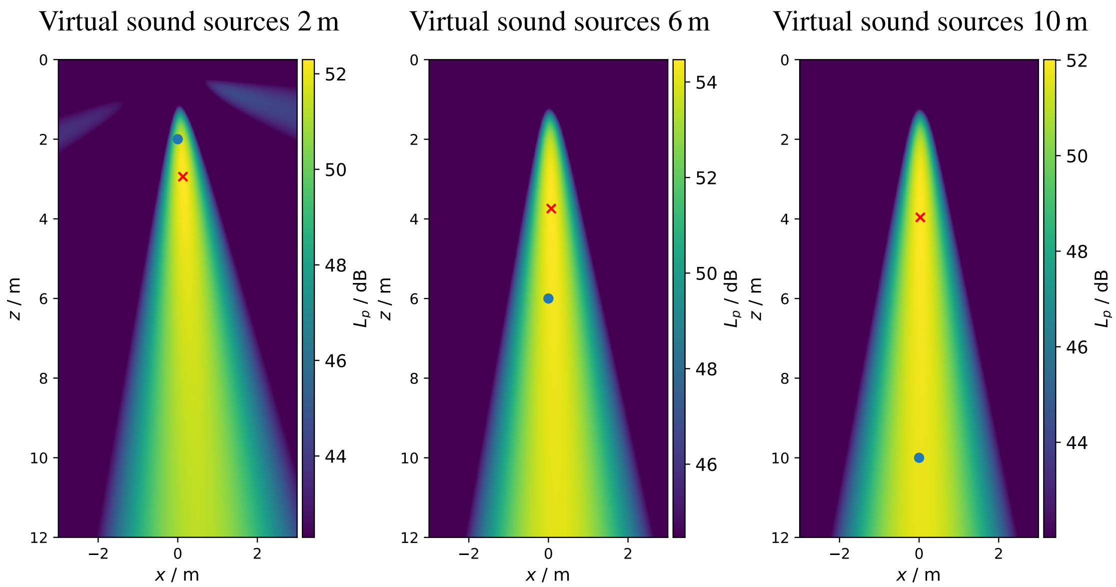


Figure 7: Virtual sound sources at three different distances along the z-axis in the third octave band with centre frequency 1600 Hz. The blue dot symbolises the position of the expected sound source position, the red cross shows the determined sound source position with the maximum sound pressure level in the indicated area.

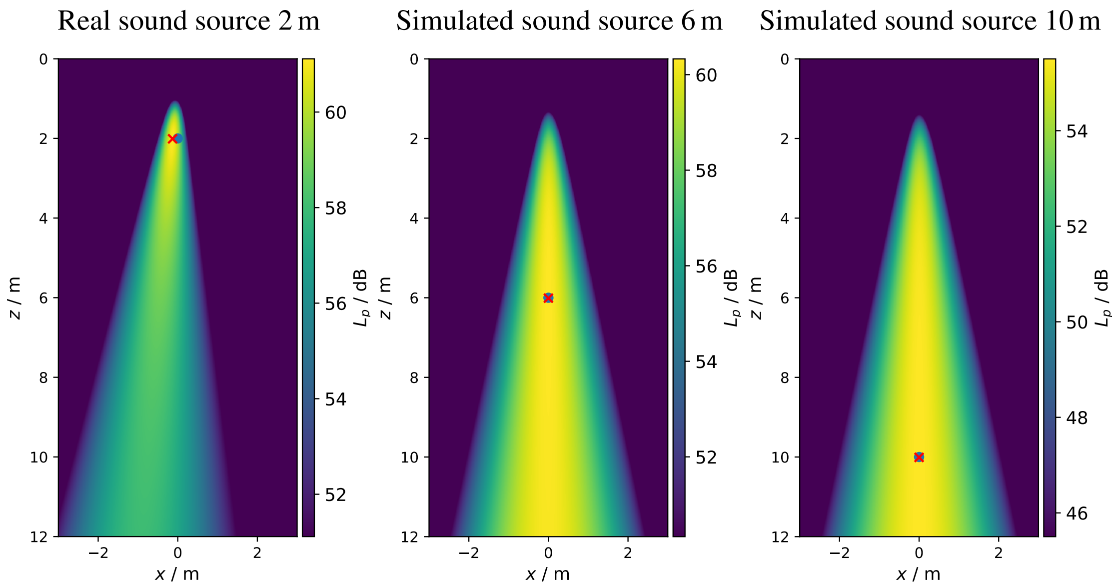


Figure 8: Real and simulated sound sources at three different distances along the z-axis in the one-third octave band with the centre frequency 1600 Hz. The blue dot symbolises the position of the expected sound source position, the red cross shows the determined sound source position with the maximum sound pressure level in the indicated area.

lobes start to appear in the sound field of the virtual sound source. Furthermore, the deviation of the measured position from the expected position increases rapidly. Both effects are probably due to the spatial aliasing of the WFS system that occurs above 5000 Hz. However, this would exceed the anti-aliasing condition of 3430 Hz described in chapter 1.1, which could be attributed to the fact that the microphone array only measures a small area in the middle of the wave field synthesis system. The theoretical anti-aliasing condition in chapter 1.1 applies to the whole area covered by the WFS system. Reducing the correctly synthesised sound field area can lead to better anti-aliasing [1]. When comparing the two plots in Fig. 4, a correlation between the number of side lobes and the distance from the maximum sound pressure level to the expected position can be assumed. However, this correlation was neither investigated in more detail nor checked for causality.

From Fig. 6a it is clear that the localization of real sound sources in the top view is only possible for a smaller frequency range than in the front view. This behavior fits the fact that the results on the z-axis are generally less accurate than on the other axes, because here the sound source and the microphone array have the largest distance between each other. Apart from a strong deviation around 700 Hz which is not further investigated in this paper but maybe due to room acoustic phenomena, the real sound source is locatable between 500 Hz and 5000 Hz. For the virtual sound sources, the deviation from the expected position is generally too high to be sufficient for localization. In addition there are extreme outbursts in the frequency range below 600 Hz and above 15 kHz. Since the virtual sound sources in these frequency ranges are localized far outside the range relevant for the evaluation, these outbursts can be categorized as "not localizable". Considering the measurements in Fig. 7 the general distance deviation in Fig. 6a of the virtual sound source between 1000 Hz and 5000 Hz can be attributed to the distance between the expected sound source position and the secondary sound source positions which is about 1 m. This assumption is further supported by the fact that all three measurements for virtual sound sources in Fig. 7 are localized at 3 m instead of the expected 2 m, 6 m and 10 m. As a reason for the detection of the secondary source position can be assumed that the wave field synthesis system does not create virtual sound fields in all dimensions as it is only 2.5 dimensional. The microphone array and the used beamforming computations on the other hand expect three dimensional point sources. A good representation of the 2.5 dimensional WFS behaviour can be seen in the Fig. 3.

5 SUMMARY

In summary it can be stated that with the compiled microphone array virtual and real sound sources can be localized comparably well in the front view within a frequency range of 400 Hz to 5 kHz. This exceeds the theoretical anti-aliasing condition of the wave field synthesis system. The sound source distributions of the virtual and real sound sources also show strong similarities to each other. However, clear differences can be observed in the localization of the sound sources in the top view. For the real sound source, localization is possible for the frequency range between 1 kHz to 5 kHz, which is a narrower frequency band than in the front view. This is to be expected because of the larger distances between sound source and microphone on the z-axis. The localization of the virtual sound source in the top view is not possible sufficiently well, which can be attributed to the technology of the wave field synthesis system in combination with the chosen microphone array geometry and beamforming calculation model. For further investigations of wave field synthesis systems using beamforming techniques it is of interest to choose microphone array and beamforming combinations that can better evaluate depth information in virtual sound fields, since creating depth information is one of the unique capabilities of wave field synthesis.

References

[1] J. Ahrens. *Analytic methods of sound field synthesis*. T-Labs series in Telecommunication Services. Springer, Berlin u.a., 2012. ISBN 9783642257421.

[2] A. J. Berkhout. "A holographic approach to acoustic control." *Journal of the Audio Engineering Society*, 36(12), 977–995, 1988.

[3] FourAudio. "WFS 8-kanal Aktivlautsprecher für die Wellenfeldsynthese." URL `www.four-audio.com/produkte/wfs/`, accessed on 17.05.2022.

[4] G. Herold, T. Geyer, P. Markus, and E. Sarradj. "Simultaneous sound power measurement of engine components." *SAE International Journal of Passenger Cars - Mechanical Systems*, 9(3), 974–979, 2016. ISSN 1946-3995. doi:10.4271/2016-01-1774.

[5] C. Huygens. *Traité de la lumière*. published in Leyden by Van der Aa, 1 edition, 1690. doi:10.5479/sil.294285.39088000545160.

[6] D. H. Johnson and D. E. Dudgeon. *Array Signal Processing: Concepts and Techniques*. Prentice Hall, 1 edition, 1993. ISBN 9780130485137.

[7] E. Sarradj. "Three-dimensional acoustic source mapping with different beamforming steering vector formulations." *Advances in Acoustics and Vibration*, 2012:1-12, 2012. doi:10.1155/2012/292695.

[8] E. Sarradj. "A generic approach to synthesize optimal array microphone arrangements." *Berlin Beamforming Conference*, S4, 2016.

[9] E. Sarradj and G. Herold. "A python framework for microphone array data processing." *Applied Acoustics*, 116, 50–58, 2017. ISSN 0003-682X. doi:doi.org/10.1016/j.apacoust.2016.09.015.

[10] S. M. Spors. *Active Listening Room Compensation for Spatial Sound Reproduction Systems*. Ph.D. thesis, Technische Fakultät der Friedrich-Alexander-Universität Erlangen-Nürnberg., 2005.

[11] P. Welch. "The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms." *IEEE Transactions on Audio and Electroacoustics*, 15(2), 70–73, 1967. doi:10.1109/TAU.1967.1161901.