DOI: 10.61782/fa.2025.0811

SPATIAL LOCALISATION AND LEVEL ACCURACY IN THE DETECTION OF A MOVING SOURCE USING A DELAY-AND-SUM BEAMFORMER

Alejandro Osses[1*] **Bente Ackermans**[1] **Helmer Nuijens**[1]
Rik Sweep[1] **Maarten Hornikx**[2]
[1] Livable Cities Acoustics team, Sorama B.V., Eindhoven, the Netherlands
[2] Department of the Built Environment, Eindhoven University of Technology, Eindhoven, the Netherlands

ABSTRACT

Delay-and-sum is one of the most fundamental signal processing techniques to estimate the direction of a sound source based on measurements from an array of microphones. However, the estimation of distance and/or sound pressure level of the source cannot be accurately retrieved if we do not take into account the properties of the acoustic environment, the sound field, where the measurements take place. In this contribution, we show the performance of a delay-and-sum beamformer in the spatial localisation of a source emitting a sound of constant overall level. The source is gradually moved in an area of 230 m^2 with source-receiver distances up to $\sim$35 m. Results are shown before and after applying a level correction based on the inverse distance law and a transfer function estimate obtained empirically. This transfer function represents a pragmatic simplification for the acoustic environment, which otherwise can be complex to measure due to the more or less influence of absorption, reflection, diffraction, refraction, and diffusion phenomena. Our evaluation is focused on the estimated source location obtained from the maxima of the imaging results and on the estimated sound pressure levels.

Keywords: *Beamforming, localisation, sound pressure level.*

**Corresponding author*: *alejandro.osses@sorama.eu.*

1. INTRODUCTION

The detection of sound events has been extensively investigated in many applications, perhaps with some more interest in radar and underwater applications since the first half of the twentieth century [1]. To monitor the direction of arrival of a sound, more than one microphone—a microphone array—is required, such that differences in the time of arrival of a sound to the different microphones can be used to estimate the direction from which that sound is coming. Such principle is captured by delay-and-sum (DAS) beamforming, a basic but well-established technique in acoustics.

In other applications such as sound source imaging in acoustics, the topic of this paper, it is often the case that the sources under investigation are close to the microphone arrays or that beamforming is applied in controlled situations in a way that a free field can be assumed. However, in complex acoustic environments, sound propagation is influenced by physical phenomena including absorption, reflection, diffraction, refraction, and diffusion. In such complex environments, accurately retrieving the location and the estimation of the actual intensity of a sound source, remains a significant challenge. The sound intensity of a detected sound source is one of our focuses in this paper, which we estimate using sound pressure levels derived from the imaging results. In the past, to tackle these challenges various methods have been employed, including deconvolution techniques and other widely used inversion methods (see, e.g., [2, 3]). Despite their effectiveness, these methods often face limitations such as computational intensity, the need for precise environmental modelling, and difficulties in real-time applications [2,4]. For this reason, a simplified correction method is presented that makes use of an empirically-obtained transfer function to more accurately quantify sound source in non-free field conditions.

In this study, a DAS beamformer is used to analyse a sound source of constant level that was gradually moved within a 230 m^2 area. The evaluation is focused on estimations of the position and level of the sound source. There was a distance of up to 35 m to the measurement devices, two 64-channel microphone arrays. The results are compared before and after the level correction is applied.

2. METHODS

2.1 Delay-and-sum beamforming

An audible sound source can be captured by a sensor—a microphone—after some delay in time due to the distance between the sound source and the sensor. This time delay can be interpreted as a phase shift in the frequency domain. This effect is captured in the Green's function of the Helmholtz equation:

$$x_j = \frac{\exp\left(-2\pi \cdot i \cdot \frac{f}{c} \cdot ||q_j - p_k||\right)}{4\pi \cdot ||q_j - p_k||} \cdot s_k \qquad (1)$$

where i represents the imaginary unit and x_j represents the signal projection of source amplitude s_k onto position q_j of microphone j after travelling from source position p_k with a speed of sound represented by c. The frequency of the sound signal is denoted by f.

In a DAS beamformer, different time delays are tested and compensated. These delays are related to the distances between the beamforming point k and the microphones j in the array, one of each potential source-microphone combination. If the microphone signals happen to result in aligned phases, their addition will amplify the signal or, in other words, the microphone signals will interfere constructively. Conversely, microphone signals from another location (other time delays) with no perfect alignment, are expected to partly cancel each other after addition, interfering destructively. In this latter case, it is less likely that a sound source is located at that position. Point p_k indicated by the red diamond in Figure 1 produces a constructive interference resulting in a local maximum of the heatmaps shown later in Figure 2. Similarly, the blue diamond (only shown in the left panel of Figure 2) indicates a position at which a local minimum is found.

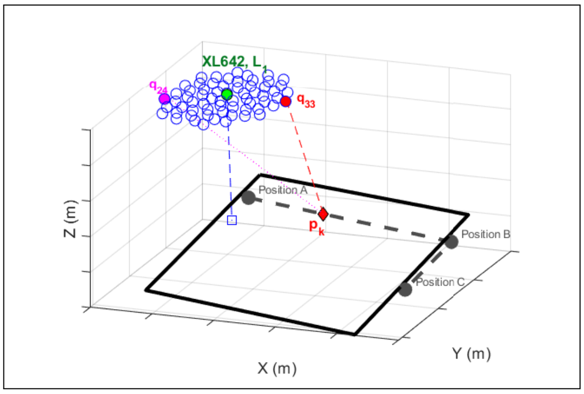

Figure 1: Schematic of one XL642 microphone array (L_1) and the main target area towards which it was beamformed (thick black line). The grey dashed curve indicates the position described by the sound source from position A to C, see the body text for details. The blue open markers indicate a magnified visualisation (x30) of the 64 microphones of L_1, while the green-filled marker indicates its central coordinate. The blue square indicates the projection of the facing direction of the array onto the beamforming plane. The red marker indicates the speaker position p_k at t=7.7 s. The closest and farthest microphones q_j to p_k are highlighted in red (j=33) and pink (j=24), respectively. These microphones are distant by 0.30 m from each other, for which a time "steering" delay of $\sim$150 μs leads to constructive interference (dark red region) at this point p_k, as can be seen later, in Figure 2. Instead, if a steering delay of $\sim$350 μs is used for these microphone signals (related to a position $\sim$16 m to the right of p_k), a region of destructive interference is obtained with a heatmap magnitude 25.4 dB below the maximum, see the magnitudes around the blue marker in the left panel of Figure 2.

In this study we use a DAS beamformer implemented in the frequency domain which receives waveforms recorded with a microphone array (here 64 channels per array) and has as free parameters the minimum and maximum frequency of the analysis (here $f_{\text{min-max}}$ = 6000—6100 Hz) and the parameters that define the size and overlap of the consecutive analysis windows. DAS beamforming is used here as a visualisation technique, where areas of constructive and destructive interference are depicted as "pixels" with different colours, from dark red to white, respectively, as shown in the figures later in this paper (Figures 2 and 4). These colour maps will hereafter be referred to as "heatmaps." More importantly, we express the heatmaps in dB, which after applying an appropriate dB offset related to the sensitivity of the microphones in the array, lead to magnitudes that have a one-to-one relationship with the sound pressure level at

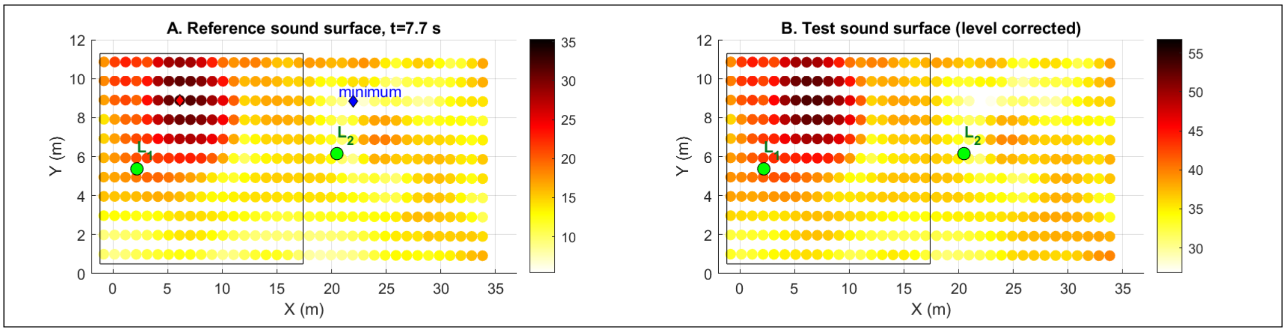

Figure 2: Two dimensional visualisation of the beamformed sound surface (heatmap) data 7.70 seconds after the start of the measurement. The black markers filled in green indicate the location of the two microphone arrays. The dark red area in the heatmap indicates the location of the omnidirectional speaker based on the measurement. For the reference heatmap (left panel), the blue-filled and red-filled diamonds indicate the minimum and maximum levels of the heatmap of 6.3 and 31.7 dB, respectively.

the specific element or "pixel" of the heatmap, as measured by the microphones within the $f_{\text{min-max}}$ range. The heatmaps in dB at this stage are used as the reference (non-corrected) heatmaps that have K pixels with each pixel belonging to one of the target points of the area onto which it will be beamformed. For the beamforming point k in that area, the magnitude will be expressed to as $h_{k,\text{ref}}$.

2.2 From heatmap to sound pressure level: Estimation and distance compensation

The DAS beamformer does not compensate for the attenuation of the sound due to its propagation through the acoustic environment, despite the smart use of steering time delays to identify regions of coincidence by the algorithm. In other words, the direct interpretation of the heatmap magnitudes leads to a lower sound-pressure-level reading than what the sound source actually emitted. To compensate for this decrease, after beamforming, the $h_{k,\text{ref}}$ magnitudes are subjected to a level compensation based on the exact distance d_{km} between each beamforming point k and the location of microphone array m, measured in meters, this compensation C was:

$$C(k,m) = A \cdot \log_2(d_{km}) \tag{2}$$

with A being a fitted parameter that represents the number of decibels that are increased by every doubling of the distance d_{km}. In general, this parameter should not exceed 6, with a value of 6 representing a free-field condition and a compensation of 6 dB per doubling of the distance [5]. In this study, a fixed value A of 5 dB was used in Eqn. (2), based on measurements that we had previously collected in this same acoustic environment, [1] that can be conceptually related to the assessment of a transfer function of the acoustic environment under investigation [6]. In this way, the corrected heatmap is obtained as:

$$h_{k,\text{test}} = h_{k,\text{ref}} + C(k,m) \tag{3}$$

The locations of the two microphone arrays L_m used in this study, two Sorama XL642 units (Sorama, Eindhoven, the Netherlands), are indicated by the green markers in Figures 2 and 4, with heights of 23.21 m (for L_1) and 22.01 m (for L_2). With average d_{km} distances of 22.59 and 21.64 m and maximum distances of 35.33 and 30.70 m, for L_1 and L_2, respectively.

The heatmaps from L_1 and L_2 were generated for the same set of beamforming points k. Both heatmaps were subsequently combined. The combined heatmap was obtained from a point-by-point weighted sum, similar to an arithmetic average [2].

2.3 Measurements: Data collection

A white noise having a statistical flat spectrum between 0 and 7000 Hz was played back through an omnidirectional speaker. The omnidirectional speaker was held at a constant height of 1 m from the floor and it was moved from position A (at $t = 0$ s) to position B (at $t = 20$ s) and then to position C by the end of the measurement, at $t = 25$ s. Positions A, B, and C are highlighted in grey in Figure 1. The sound pressure level L_p of the sound source was constant and measured to be $L_p = 78.5$ dB of continuous equivalent level, read from a Class 1 NTi XL2

[1] The interested reader is referred to Figure 6 in Appendix A.

sound level meter at a distance of 1 m from the speaker. However, due to the beamforming range between 6000 and 6100 Hz (see Section 2.1), only the effective amount of white noise within this band-limited range can be visualised in the heatmaps. For this range, the white noise had 60.0 dB of effective signal. [2] This level of 60.0 dB is, therefore, the target level that we expect to achieve when visualising the resulting DAS heatmaps, if the attenuation by distance is appropriately compensated.

3. RESULTS AND DISCUSSION

One frame of the obtained beamforming results are shown in Figure 2 for the reference heatmap, without level compensation (left panel) and for the test heatmap (right panel), after a level compensation based on the adopted inverse distance law was applied.

The heatmaps in Figure 2, show for the selected analysis frame at $t = 7.7$ s, that both sound surface heatmaps have the same lobes, evidenced by the similarities in the dark red regions in both panels. However, as expected for the compensated representation, the magnitudes of the right panel are similar in relative terms but the overall values are higher than those of the reference heatmap, as depicted by the difference of about 20 dB in the corresponding colour scales.

To provide further support to the previous statements, the maximum levels $L_{p,\text{max}}$ were extracted for all 250 frames (25 seconds, one frame every tenth of a second), and are shown in Figure 3.

The results in Figure 3 depict the systematic increase in the predicted maxima from the beamformed lobes, that are on average 22 dB closer to the target band-limited level of 60 dB. However, there is still a range of about 10 dB that needs to be bridged or, in other words, that seems to require a more accurate level compensation. So far, the test compensation method does not seem to affect the position of the sound source for one analysis frame (Figure 2) and in fact it brings the maxima from the beamforming results closer to the target level of 60 dB.

The same observation can be extended to all frames in the heatmaps. To do so, the trajectory of the sound source throughout the whole measurement is shown in Figure 4.

[2] The band-limited level $L_{p,\text{band-limited}}$ is obtained from $L_{p,\text{band-limited}} = L_p - 10 \cdot \log_{10}(\text{BW}/\text{BW}_{\text{band-limited}})$. For the white noise that was used, the total level L_p was 78.5 dB, the bandwidth (BW) was 7000 Hz and $\text{BW}_{\text{band-limited}}$ is 100 Hz, i.e., the frequency range between 6000 and 6100 Hz. This formula is valid because of the statistical flat spectrum of a white noise.

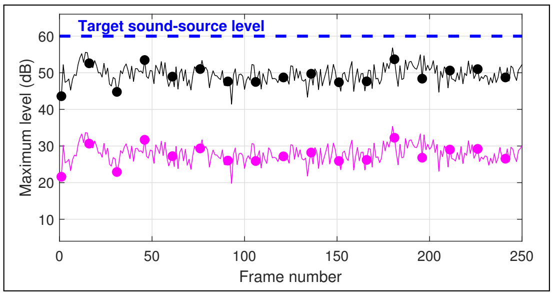

Figure 3: Maximum levels $L_{p,\text{max}}$ for each analysis frame. The pink and black traces indicate the maxima retrieved from the reference heatmap and from the inverse-distance-law corrected heatmap, respectively. The trace with corrected levels (black) is on average 22 dB above the pink trace, however, it is still about 10 dB below the target band-limited level of 60 dB. The filled markers indicate the maximum levels obtained every 1.5 s (15 analysis frames), which match the source positions highlighted by the thick open squares with the same colours in Figure 3.

The trajectory was estimated from the maximum lobe of each analysis frame and is superimposed to the heatmaps of the last analysis frame of the measurement, at $t = 25$ s. The locations of the sound source are indicated by the pink and black traces in the left and right panels of Figure 4, respectively. The filled symbols represent the locations after an arbitrary time step of 1.5 s or, considering the beamforming parameters, every 15 analysis frames. Highlighting filled symbols gives a clearer idea of the speed at which the speaker was moved within the area of interest [3] and also for comparison with the maximum levels at those locations that as highlighted in Figure 3. The similarity between the pink and black traces supports the fact that the sound source is located at the same virtual position, for the reference and test heatmaps, without or with the level compensation, respectively.

4. CONCLUSION AND FURTHER WORK

In this study we have presented data collected with two microphone arrays installed in a nearly free-field acoustic environment where sound propagation seemed to be well predicted when assuming an attenuation of 5 dB every time the distance was doubled. For the analysis and visualisation of the results, delay-and-sum (DAS) beamforming was applied using a frequency-domain implementa-

[3] A rough estimate of the speed at which the sound source was moved is 1 m/s (or 5.4 km/h) or, in other words, two contiguous filled markers being approximately 1.5 m apart in Figure 4.

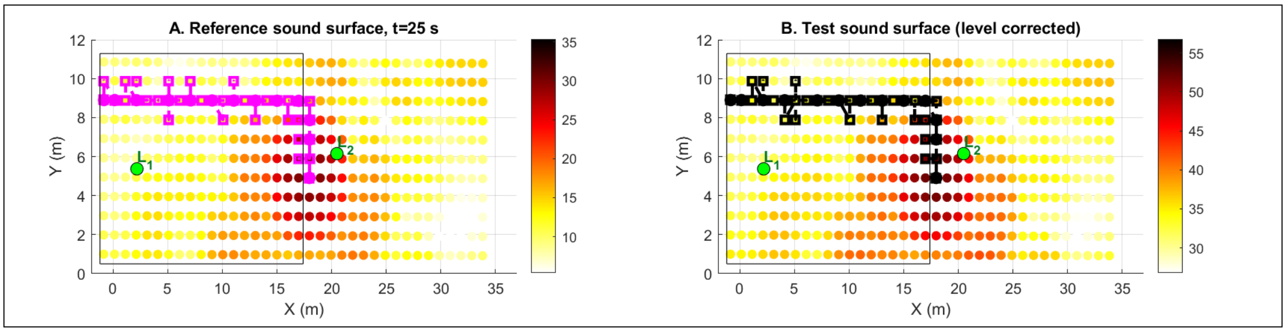

Figure 4: Sound surface heatmaps before and after an inverse-distance-law compensation was added to the obtained magnitudes in dB. The traces in pink and black represent the locations of the maximum lobes in the course of the measurement. The maxima were found gradually from the left to the right of the abscissa and at around time $t = 20$ s (frame 200), the source was found to go down the ordinate. The lobes in both panels correspond to the source location in the last analysis frame, at $t = 25$ s. Note that the colour scale has a different range in both heatmaps.

tion. The white noise sound was analysed using a frequency range between 6000 and 6100 Hz, which is equivalent to analyse a band-passed filtered version of the signals captured by the microphone arrays.

Despite the basic nature of the DAS beamforming technique [7], we emphasise the transparency of its processing, allowing a one-to-one relationship between the estimated magnitudes and the actual sound pressure level of the source. Our experience indicates that such transparency extends not only to visualise results as sound surface heatmaps, but also extends to the synthesis of beamformed (steered) signals, although we do not show this in the current study.

Additionally, we present a simple way to compensate for the distance between the sound source being measured and the microphone arrays based on an inverse distance law. Based on our evaluation, the method improves the level accuracy bringing levels (for distances up to 35 m) closer by 22 dB with respect to the non-corrected heatmaps. However, there are still missing factors in this simple compensation. This level compensation does not seem to affect negatively the localisation of sound sources, at least for a nearly free-field environment as the one evaluated in this study. For this reason, it is important to mention some of the limitations that we did not address and may require further attention:

- **We used an omnidirectional speaker**: This means that the source radiated sound at roughly constant levels in all directions. Given that everyday sound sources are, in general, directional, a more comprehensive analysis of the directionality of sound sources to be used in the acoustic environment is required. The sound power of an important number of sources is known and could be used to investigate this aspect, e.g., supported by dedicated acoustic modelling software such as the open source package OpenPSTD [8] or commercial packages such as SoundPLAN and CadnaA. Alternatively, the sound power of specific sound sources could be estimated from ad-hoc measurements (see, e.g., [9]).
- **The acoustic environment was only roughly characterised**: Under the premise that the distances between the target points towards which we beamformed and the microphone arrays were more critical than the exact than the comprehensive acoustic characterisation of the acoustic environment, we captured the environment properties only pragmatically. This pragmatic approach allowed us to fit the attenuation factor per distance doubling, the factor A of the inverse distance law, which normally adopts values between 3 and 6 dB [see Eqn. (2) and Appendix A]. Other simplifications were: (a) We used a constant speed of sound and the factors of humidity and temperature were not accounted for. No classical acoustic measurements were conducted either, to evaluate the amount of reflections near the sound source or whether the fluctuations in magnitude from the heatmaps were actually produced by other sources in the background, and (b) localisation in reverberant environments is in general more difficult, with wider lobes for the same source-array distance compared to free field conditions [10].

From the points listed above, a particularly challenging case is that of reverberant environments, where prop-

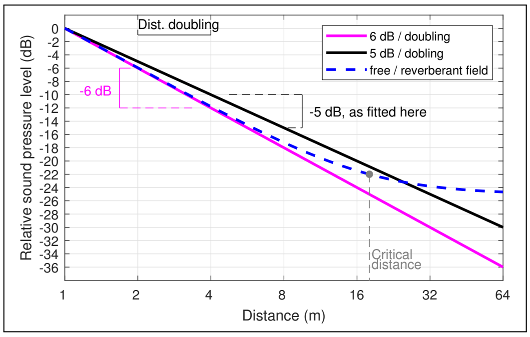

Figure 5: Schematic showing a free field attenuation as a function of the doubling of the distance (pink), using the factor $A = 5$ dB/doubling as used in this study (black), and using a more elaborate attenuation rule, where there is a different attenuation factor for the free field an for the reverberant parts of the curve (blue), delimited by a critical distance.

agation as a function of distance can be better simulated by considering, e.g., a direct and a reverberant path, instead of assuming a constant logarithmic decay (here of 5 dB/doubling) and, alternatively, also to introduce a frequency-dependent approach. In these cases, point-to-point impulse response measurements could be provide a more exact correction factor, which could also be derived as a function of frequency. A schematic of this more elaborate attenuation rule is given in Figure 5.

The study of these shortcomings could only translate in further improvements of the results that we have presented in this study. Still, we believe that the presented approach of distance compensation can apply for an important number of situations where the microphone arrays are at long distances from the sound sources of interest.

5. ACKNOWLEDGEMENTS

We are grateful to Timo Garritsen and Maaren Wiermans, for their support in collecting the measurements used in this study. This project received funding from the VIP-NOM project. Author A.O. received support from the project SINTRA to present the current contribution.

6. REFERENCES

[1] B. Van Veen and K. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Magazine*, vol. 2, pp. 4–24, 1988.

[2] R. Merino-Martínez, P. Sijtsma, M. Snellen, T. Ahlefeldt, J. Antoni, C. Bahr, D. Blacodon, D. Ernst, A. Finez, S. Funke, T. Geyer, S. Haxter, G. Herold, X. Huang, W. Humphreys, Q. Leclère, A. Malgoezar, U. Michel, T. Padois, A. Pereira, C. Picard, E. Sarradj, H. Siller, D. Simmons, and C. Spehr, "A review of acoustic imaging methods using phased microphone arrays," *CEAS Aeronautical Journal*, vol. 10, pp. 197–230, 2019.

[3] L. T. Lima Pereira, R. Merino-Martínez, D. Ragni, D. Gómez-Ariza, and M. Snellen, "Combining asynchronous microphone array measurements for enhanced acoustic imaging and volumetric source mapping," *Applied Acoustics*, vol. 182, p. 108247, 2021.

[4] R. Merino-Martínez, B. von den Hoff, D. Morata, and M. Snellen, "Three-dimensional acoustic imaging using asynchronous microphone array measurements," in *9th Berlin Beamforming Conference*, pp. 1–25, 2022.

[5] F. A. Everest and K. Pohlmann, "Sound in the free field," in *Master Handbook of Acoustics*, pp. 33–38, sixth edition ed., 2015.

[6] B. Ackermans, *Optimizing sound pressure level calibration with machine learning for enhanced accuracy in smart stadiums*. Bachelor thesis, Fontys University of Applied Sciences, 2024.

[7] V. Perrot, M. Polichetti, F. Varray, and D. Garcia, "So you think you can DAS? A viewpoint on delay-and-sum beamforming," *Ultrasonics*, vol. 111, p. 106309, 2021.

[8] M. Hornikx, T. Krijnen, and L. van Harten, "OpenPSTD: The open source pseudospectral time-domain method for acoustic propagation," *Comput. Phys. Commun.*, vol. 203, pp. 298–308, 2016.

[9] ISO 3740, "Acoustics—determination of sound power levels of noise sources—guidelines for the use of basic standards," Standard, International Organization for Standardization, 2019.

[10] J. Fischer and C. Doolan, "Beamforming in a reverberant environment using numerical and experimental steering vector formulations," *Mechanical Systems and Signal Processing*, vol. 91, pp. 10–22, 2017.

A. TRANSFER FUNCTION MEASUREMENT

In this appendix gives a brief description of the measurements that led to the attenuation factor A=5 dB/doubling used in the inverse distance law adopted in this study. The measurements were conducted by one of the authors [6]. The measurements were collected in the same venue as in this study, but the "area of interest" was not exactly the same. The dimensions of the tested area were however comparable, having similar horizontal and vertical dimensions as the are the area delimited by the black rectangle from Figure 1, where points A, B, and C were located.

A.1 Data collection

A speaker was located at a fixed position near one of the corners of the area of interest, comparable to point B from Figure 1. The UE Megaboom 3 BT speaker was used. This speaker was not omnidirectional but it was calibrated to produce a fixed sound pressure level of 56 dB at a distance of 1 m from the speaker, for the frequency range between 1 and 3 kHz. The sound used in the measurements was a recorded "stadium sound" of people cheering during a football match. For the purposes of reproducibility of the experiment, we chose a sound segment that maintained an approximate overall sound pressure level. When the sound was being played back in loop (with the speaker the fixed position), the experimenter moved to different measurement positions along the horizontal and vertical directions from the speaker and annotated the sound-pressure-level readings as registered by a Class 1 Bedrock SM90 sound level meter. Because for this project, the frequency range of interest was between 1 and 3 kHz, only the sound level meter data for that frequency range (as one-third octave band levels) was used.

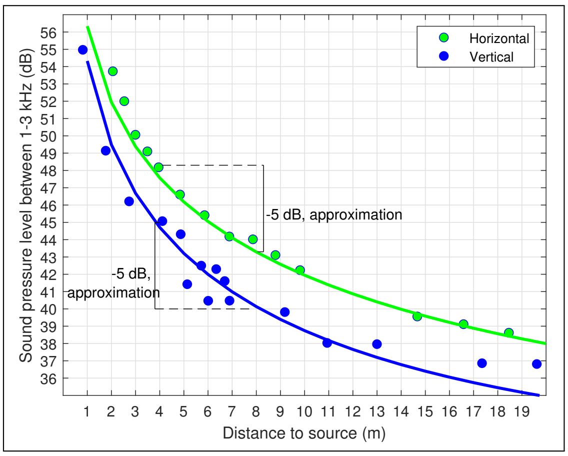

Figure 6: Sound pressure levels as a function of distance for measurements in the horizontal (green) or vertical direction (blue) as obtained from the measurements. For both curves, a decrease of 5 dB/doubling is indicated by the insets in black.

B. RESULTS AND BRIEF DISCUSSION

The measured levels as a function of the distance between the speaker and the sound level meter are indicated by the filled markers in Figure 6. In the figure, 30 labelled measurements are shown, 14 for collected in the horizontal direction and 16 collected in the vertical direction. The distances between the speaker and the sound level meter was between 1 and 19 m. The continuous lines represent the best logarithmic fits to the data that we could obtain. For the curve obtained for the horizontal (green) and vertical directions (blue), insets are shown to highlight the approximate decrease of 5 dB/doubling for the curves between 4 and 8 m.