LOTTO (LOCALIZE OTTOSONICS): AUTOMATED SPATIAL SPEAKER LOCALIZATION USING TETRAHEDRAL MICROPHONE ARRAY

Benjamin WESCH (benjamin.wesch@kunstuni-linz.at) [1],
Manuel MITTERHUBER (manu@goon-studios.com) [2], and
Martin KALTENBRUNNER (martin.kaltenbrunner@kunstuni-linz.at) [1]

[1] *Tangible Music Lab*, **Kunstuniversität Linz**, Hauptplatz 6, 4020 Linz, Austria
[2] **OTTO Kulturgenossenschaft**, Rodlstrasse 19, 4100 Ottensheim, Austria

ABSTRACT

This paper presents LOTTO (Localize OTTOsonics), an open, affordable and portable system for automated speaker localization in spatial audio setups. Implemented in Pure Data (Pd) [1], it utilizes a 3D-printed tetrahedral microphone array to precisely determine speaker positions through time-of-arrival measurements. The implementation has been successfully tested with a real-world setup, demonstrating its practical applicability for spatial audio installations.

1. INTRODUCTION

Spatial audio systems require precise knowledge of speaker positions for accurate sound field reproduction. For temporary installations and touring setups, traditional manual measurements are impractical and error-prone, especially when venue constraints force last-minute speaker repositioning.

Our automated system efficiently determines speaker positions using acoustic measurements from a tetrahedral microphone array. It generates a brief test signal and analyzes the received signals to precisely measure time-of-arrival at four measurement microphones, then calculates the three-dimensional coordinates of each speaker. After an initial single-distance calibration, the system provides highly accurate position measurements.

This work addresses a practical challenge in spatial audio deployment, offering a robust solution for rapid and precise speaker localization. The implementation has been validated with a hemispheric array of 23 speakers, as shown in Figure 1, demonstrating its effectiveness in real-world applications.

Figure 1. Test setup showing the tetrahedral microphone array positioned within a hemispheric 23-speaker array.

2. RELATED WORK

2.1 Speaker Localization Methods

Traditional methods for speaker localization often rely on manual measurements. Professional equipment for measuring angles in space precisely is expensive. As an alternative, we found solutions like smartphone compasses for azimuth measurement or laser pointers taped to angle meters rather impractical. Also, when using laser distance meters or tape measures to determine speaker positions in cartesian coordinates, it is necessary to project or draw a coordinate system on the floor which sometimes is simply not possible.

Initial experiments used a Rode NT-SF1 ambisonic microphone with four capsules in tetrahedral arrangement. While compact and suitable for direction-of-arrival estimation using intensity-based methods [2], the close capsule proximity limited time-of-arrival precision. Our current implementation uses four separate measurement microphones on a larger tetrahedral structure (edge length $\approx 577\,\mathrm{mm}$), providing superior angular precision through increased spatial separation. Individual budget microphones eliminate the need for precisely matched capsules while achieving high accuracy at lower cost.

Other automated approaches include camera-based systems and acoustic measurement techniques [3]. Acoustic cameras [4] use larger microphone arrays with beamforming to create visual "heat maps" of sound sources, requiring more complex hardware and processing.

2.2 Time-of-Arrival Localization

Time-of-arrival localization uses exact time measurements of the arrival of a signal at multiple receivers to determine a source's position. The technique has been widely used in GPS and acoustic source localization systems [3, 5]. For speaker localization, the controlled nature of the test signal allows for highly precise measurements.

3. IMPLEMENTATION

3.1 Hardware Architecture

The core of the system is built around a tetrahedral array of four Behringer ECM8000 measurement microphones. These omnidirectional condenser microphones were chosen for their flat frequency response (15 kHz to 20 kHz), high sensitivity (-60 dB), and excellent price-performance ratio at €23.90 each. The microphones are mounted on a custom 3D-printed structure that ensures precise tetrahedral positioning with an edge length of approximately 577 mm (other dimensions were tested - but this size yielded good results and allows for comfortable transportation and setup).

The specific coordinates of the four microphones, normalized to unit edge length, are:

$$\vec{p}_0 = [0 \quad \frac{1}{\sqrt{3}} \quad -\frac{1}{2\sqrt{6}}]^T \quad \text{(FD)} \tag{1}$$

$$\vec{p}_1 = [-\frac{1}{2} \quad -\frac{1}{2\sqrt{3}} \quad -\frac{1}{2\sqrt{6}}]^T \quad \text{(BLD)} \tag{2}$$

$$\vec{p}_2 = [\frac{1}{2} \quad -\frac{1}{2\sqrt{3}} \quad -\frac{1}{2\sqrt{6}}]^T \quad \text{(BRD)} \tag{3}$$

$$\vec{p}_3 = [0 \quad 0 \quad \frac{3}{2\sqrt{6}}]^T \quad \text{(T)} \tag{4}$$

These coordinates place the origin at the center of the tetrahedron, with one microphone positioned forward and slightly down (FD), two microphones spread in the back and down with opposite x-coordinates (BLD and BRD), and one microphone at the top (T). In the actual implementation, these coordinates are multiplied by the measured edge length to obtain the real-world microphone positions.

The tetrahedral array is implemented using a 3D-printed structure that can be reproduced using files available through the repository of the software implementation [6]. As illustrated in Figure 2, the microphones are oriented outward to ensure optimal coverage of the surrounding space.

3.2 Software Architecture

The Pure Data (Pd) implementation consists of several interconnected modules working together to achieve precise speaker localization with a sample rate of 48 kHz. At its core, it utilizes a windowed single-cycle sine burst with a frequency of 4800 Hz (negative to positive peak distance of 5 samples), optimized for precise time-of-arrival detection while remaining reproducible by standard speakers. The detection system achieves sub-sample precision through parabolic interpolation of the received signal's positive peak.

Figure 2. 3D-printed tetrahedral structure with outward-facing measurement microphones, designed to fit standard microphone stands.

Position calculation is handled by a Cramer's rule solver module, which processes the time differences to determine three-dimensional coordinates. A calibration module manages the single-distance reference measurement necessary for absolute positioning. LOTTO provides real-time feedback through an interactive interface where users can monitor and adjust measurements as needed (see Figure 3).

3.3 Signal Design

The system employs a carefully designed test signal in shape of a single-cycle sine burst windowed with a Hann function [7] as shown in Figure 4. This provides an optimal balance of

- Minimal duration reducing interference from room reflections
- Well-defined onset
- Spectral characteristics suitable for measurement microphones
- Sufficient energy for reliable detection while remaining unobtrusive
- Reproducible by standard speakers

Several alternative approaches were considered, following established acoustic measurement techniques [8]:

- Cross-correlation of a chirp signal: While offering good detection properties, this approach introduced unnecessary length, as our system only requires a single reliable peak.

ISBN 978-3-200-10642-0

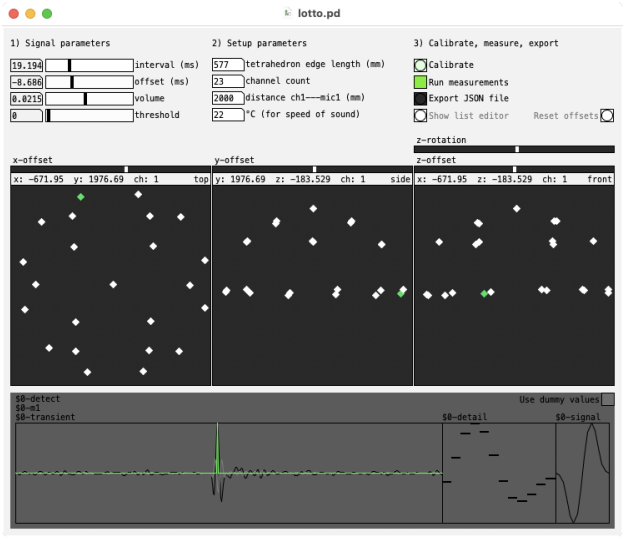

Figure 3. Pure Data GUI showing: (top) system parameters with calibration controls and export options; (middle) top, side, and front views with real-time updates; (bottom) measurement area with signal processing and detection markers. A separate text window can be opened to display all channels' spherical coordinates.

- Window function with mid-point sign flip: This creates a strong detectable transient. However, its high-frequency content makes it impractical for speaker reproduction.
- Detection based on zero-crossing between positive and negative peak. We did not follow this approach, as the peak detection already yields reliable sub-sample precision.
- Overlapping sine sweeps played back on each speaker with minimum offsets that can be deconvolved and analyzed. This would additionally give us impulse responses for each speaker/microphone pair and could be an interesting approach for future work, but adds unnecessary complexity for simple speaker localization requirements.

3.4 Time-of-Arrival Detection

The detection process operates in two stages, building on established time-delay estimation methods [9]. First, it identifies the initial transient by monitoring the second derivative (acceleration) of the incoming audio stream, detecting the first value above a threshold. This approach effectively captures sudden changes in the signal while being robust against background noise.

At a sampling rate of f_s and speed of sound c, the basic sample-rate-limited precision would be:

$$d_{min} = \frac{c}{f_s} \tag{5}$$

For example, with our system operating at $48\,\mathrm{kHz}$ sampling rate and at room temperature ($20\,^\circ\mathrm{C}$ where $c \approx 343\,\mathrm{m\,s^{-1}}$), the basic resolution would be limited to approximately $7.15\,\mathrm{mm}$.

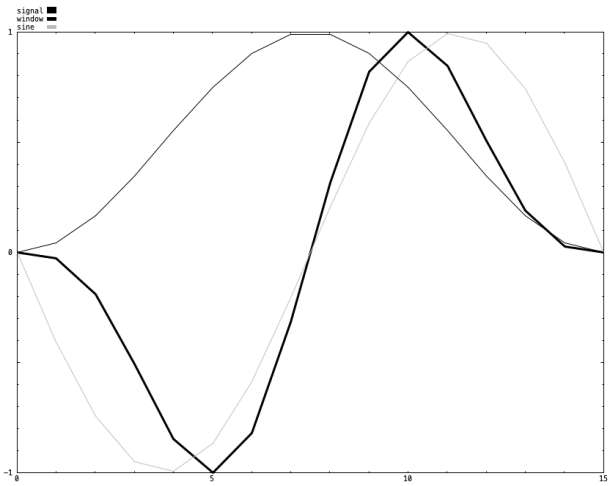

Figure 4. Test signal components showing the single-cycle sine burst, Hann window function, and the normalized resulting windowed burst overlaid in a single graph.

To improve upon this sample-rate limitation, we implement sub-sample precision through parabolic interpolation. Once the initial transient is detected, the algorithm identifies the highest sampled value (discrete peak) within a small time window. While this sampled peak only approximates the true peak position, we can calculate a more precise location of the actual peak by fitting a parabola through this sample and its two neighbors. The interpolation is given by:

$$\begin{aligned} a &= \frac{y_{n-1} + y_{n+1} - 2y_n}{2} \\ b &= \frac{y_{n+1} - y_{n-1}}{2} \\ x_{\text{offset}} &= -\frac{b}{2a + \epsilon} \end{aligned} \tag{6}$$

where y_n represents the y-value of the peak sample, $y_{n\pm1}$ its neighbors, x_{offset} gives the calculated offset from the sampled peak to the actual peak position in fractional samples. $\epsilon = 10^{-5}$ is a small constant to prevent division by zero. This interpolation process is visualized in Figure 5.

Through sub-sample interpolation of the peak positions, we can achieve distance measurement resolutions below $1\,\mathrm{mm}$ between each microphone and the speaker.

3.5 Directional Runtime Compensation

While using the center of a microphone's membrane as a pivot point, we discovered systematic deviations within the measured time of arrival (and therefore calculated distance) of the signal based on the angle between the microphone and the source direction. This is probably a combined effect caused by the longer path the pressure wave has to travel and diffract around the microphone to reach and move the membrane, and the geometry of the opening of the membrane itself. Through careful measurements of these deviations at different angles (0°, 22.5°, 45°, 67.5°, 90°, 112.5°, 135°, 167.5°, and 180°), we derived a third-order polynomial correction function:

ISBN 978-3-200-10642-0

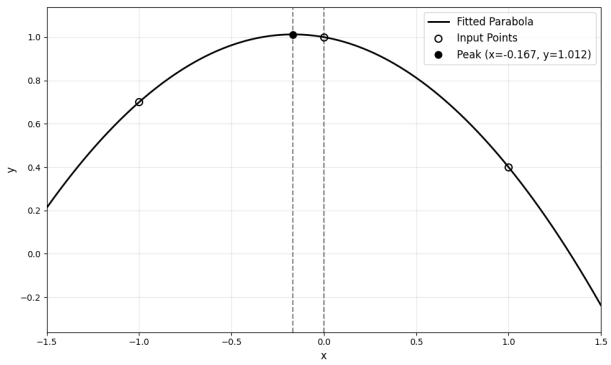

Figure 5. Illustration of parabolic peak interpolation showing how sub-sample precision is achieved by fitting a parabola through three adjacent samples around the peak.

$$d_{offset} = \alpha^3 \cdot 1.646 \cdot 10^{-6} - \alpha^2 \cdot 1.0347 \cdot 10^{-4} + \alpha \cdot 0.0193 \quad (7)$$

where α is the angle in degrees, and d_{offset} gives the distance offset in millimeters that needs to be subtracted from the measured distance. The measured offsets and fitted correction function are shown in Figure 6.

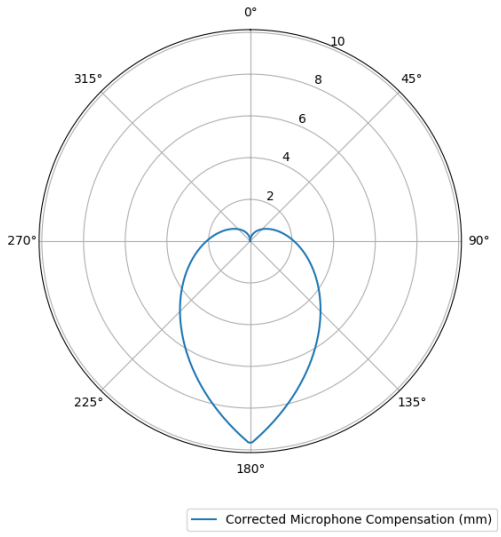

Figure 6. Polar plot showing the measured distance offset (in millimeters) as a function of angle between microphone axis and source direction. The solid line shows the fitted polynomial correction function that is subtracted from the measurements.

3.6 Position Triangulation

Our approach first measures the absolute time-of-arrival at each of the four microphones:

$$d_i = c \cdot \text{TOA}_i \quad \text{for } i = 0, 1, 2, 3 \quad (8)$$

where TOA_i is the time of arrival at microphone i and c is the speed of sound. This yields four absolute distances from the source to each microphone.

To solve for the three-dimensional source position $\vec{x}$, LOTTO uses microphone 0 as reference and forms three geometric equations using the distance differences:

$$\|\vec{x} - \vec{p}_i\| - \|\vec{x} - \vec{p}_0\| = d_i - d_0 \quad \text{for } i = 1, 2, 3 \quad (9)$$

Through algebraic manipulation (squaring both sides, expanding, and eliminating quadratic terms), each equation transforms into a linear equation in $\vec{x}$. The three resulting linear equations form a system that can be solved using Cramer's rule [10] to obtain the three-dimensional coordinates of the source position.

3.7 Distance Estimation Refinement

While Cramer's rule provides the initial position solution, we improve distance estimation accuracy by leveraging the known tetrahedral microphone geometry. With precisely positioned microphones in a zero-centered tetrahedral arrangement (edge length $\approx 577\,\text{mm}$), the final distance estimate is calculated as follows:

First, we compute the average of all four individual distance measurements:

$$d_{avg} = \frac{1}{4} \sum_{i=1}^{4} d_i \quad (10)$$

Due to the geometric properties of a regular tetrahedron with edge length e, the actual distance r from the origin to the source can be derived as:

$$r = \sqrt{d_{avg}^2 - \left(\frac{e}{2}\right)^2} \quad (11)$$

This relationship emerges from the tetrahedral symmetry and becomes increasingly accurate as the distance from the origin increases. For distances greater than twice the radius of the circumscribed sphere, the formula achieves mean errors below 0.5% with a maximum error of approximately 1%. The correction maintains the azimuth and elevation angles from the Cramer's rule solution while providing highly accurate distance estimates.

3.8 System Parameters and Interface

The LOTTO system provides several adjustable parameters through its Pure Data interface:

3.8.1 Signal Parameters

- Measurement interval: adjustable to longer intervals for environments with strong reflections
- Signal offset: Allows temporal alignment of the detection signal within the measurement window of 512 samples
- Volume control: For optimizing signal-to-noise ratio
- Peak threshold: Filters out inactive or malfunctioning channels

ISBN 978-3-200-10642-0

3.8.2 *Setup Parameters*

- Tetrahedral array edge length: Default $577\,\mathrm{mm}$, determined by the 3D-printed structure and Behringer ECM8000 measurement microphones
- Channel count: Configurable according to the speaker array
- Reference distance: Single-point calibration between channel 1 and microphone 1 (microphone or speaker can be moved freely for calibration)
- Temperature: Used for accurate speed of sound calculation according to the formula $c = 331.3 + 0.606T\ \mathrm{m\,s^{-1}}$, where T is the temperature in degrees Celsius [11]

3.9 Workflow and Export

The measurement process follows a straightforward workflow:

1. Initial calibration using reference distance
2. Continuous measurement runs
3. Adjustment options for center position and orientation:
   - Z-axis rotation (vertical axis)
   - Translation offsets for all axes
4. Export to JSON format compatible with IEM Plugins' AllRaDecoder for ambisonic applications

The interface provides interactive visualization where users can select individual channels to view their precise coordinates. The measured positions can be exported and visualized in IEM Plugins' AllRaDecoder, as demonstrated in Figure 7.

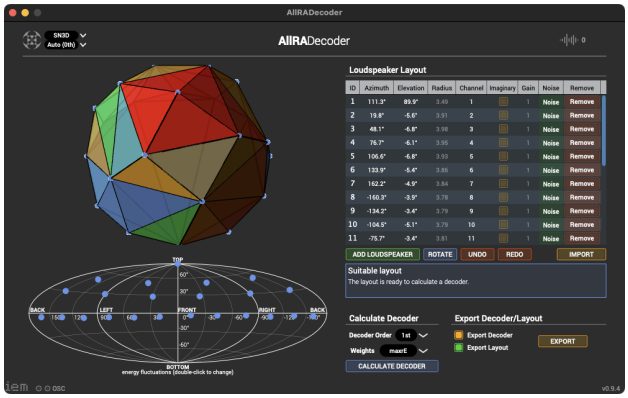

Figure 7. Screenshot of the measured speaker positions imported into IEM Plugins' AllRaDecoder through JSON export.

4. RESULTS

The system was evaluated using a hemispheric array of 23 speakers at a radius of ca. $2\,\mathrm{m}$. While this setup demonstrated the system's practical applicability with results visible in Figure 7, we conducted additional, more precise individual measurements to assess the system's accuracy. We focused on two measurement series (initially calibrated with a distance of $60\,\mathrm{cm}$ to the first microphone):

1. Azimuth accuracy: Measurements at a fixed radius ($2\,\mathrm{m}$) with varying azimuth angles from 0° to 60°. This range was chosen because the tetrahedral geometry of our setup means that measurement characteristics repeat or mirror beyond this range due to the array's symmetry.
2. Distance accuracy: Measurements at a fixed azimuth with varying distances from $1\,\mathrm{m}$ to $4\,\mathrm{m}$.

Table 1. Comparison of manual vs. LOTTO measurements at $2\,\mathrm{m}$ radius with varying azimuth

Test Point	Radius (m)			Azimuth (°)		
	Man.	LOTTO	Δ	Man.	LOTTO	Δ
1	2.00	1.99	-0.01	0	0.9	0.9
2	2.00	1.99	-0.01	5	4.6	-0.4
3	2.00	2.00	0.00	10	8.6	-1.4
4	2.00	2.00	0.00	20	18.4	-1.6
5	2.00	2.00	0.00	30	29.2	-0.8
6	2.00	2.00	0.00	40	39.2	-0.8
7	2.00	2.00	0.00	45	45.8	0.8
8	2.00	2.00	0.00	60	60.8	0.8

Table 2. Comparison of manual vs. LOTTO measurements at varying distances

Test Point	Radius (m)			El.* (°)
	Man.	LOTTO	Δ	
9	1.00	0.99	-0.01	2.3
10	1.50	1.49	-0.01	2.2
11	2.00	2.00	0.00	1.7
12	3.00	3.00	0.00	1.3
13	4.00	4.00	0.00	1.1

* Elevation was not properly aligned to 0°.
Azimuthal alignment was ignored for these measurements.

The results demonstrate high accuracy in both distance and azimuth measurements:

- Distance measurements show excellent precision with maximum deviations of ±1 cm. Note that while the system's internal calculations operate at higher precision, our JSON export format rounds distances to centimeters, limiting the precision of the reported results
- Azimuth angle measurements exhibit accuracy within ±1.6 degrees, as verified using a standard angle stop tool from our lab's workshop. While this tool limited our ability to verify even higher precision, the consistency of measurements suggests that the actual system accuracy might be better

ISBN 978-3-200-10642-0

- While precise elevation measurements were not performed, the LOTTO system consistently reported small positive elevation angles (1.1° to 2.3°), which could be attributed to imprecise alignment of the measurement setup

These measurements confirm that the system achieves sub-centimeter precision in distance measurements and degree-level accuracy in angular measurements, making it suitable for practical speaker localization tasks.

5. CONCLUSIONS

This paper presented LOTTO (Localize OTTOsonics), a Pure Data (Pd) implementation for automated speaker localization in spatial audio setups, using a tetrahedral array of four measurement microphones. The system demonstrates several key advantages: portable and simple setup due to the tetrahedral structure design, rapid automated measurements taking approximately $20\,\mathrm{ms}$ per speaker, high accuracy through sub-sample peak detection, and robust performance across different room acoustics.

The complete implementation is available as open-source software, including model files for the 3D-printable tetrahedral structure [6]. Assuming a multichannel audio system with 4 microphone inputs is already available, along with XLR cables and a microphone stand, the main costs for implementing a LOTTO system are primarily the measuring microphones. With each microphone costing €23.90, the total system can be assembled for approximately €100.00, making it a cost-effective solution for acoustic measurements.

6. FUTURE WORK

Several extensions could enhance the system's capabilities. Simultaneous multi-speaker measurements, alternative signal analysis methods like cross-correlation, and compensation for varying environmental conditions represent promising research directions. A key improvement would be removing the current 512-sample window limitation, which currently restricts all speakers to within $3.6\,\mathrm{m}$ range difference.

The current implementation uses Cramer's rule for computational simplicity, but this linearization approach reduces the overdetermined system (4 distance constraints, 3 position unknowns) to a 3×3 linear system. Future work could explore nonlinear least-squares optimization to preserve the overdetermined nature of the problem, potentially improving position accuracy by reducing error propagation and better utilizing measurement redundancy for noise reduction.

Another promising direction would be eliminating calibration requirements entirely by implementing a pure time-difference-of-arrival (TDOA) approach. Additionally, the system could be extended to use sine sweep measurements, capturing both speaker positions and impulse responses for each speaker-microphone pair. This would provide valuable acoustic information for system configuration, though requiring batch processing of all speakers rather than individual real-time localization as each speaker plays its test signal.

Acknowledgments

This work was developed as part of the OTTOsonics project, an open hardware system for low-budget spatial sound setups. The authors would like to thank Miller Puckette for developing Pure Data.

7. REFERENCES

[1] M. S. Puckette *et al.*, "Pure data," in *ICMC*, 1997.

[2] S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki, "Spatial decomposition method for room impulse responses," *Journal of the Audio Engineering Society*, vol. 61, no. 1/2, pp. 17–28, 2013.

[3] J. H. DiBiase, H. F. Silverman, and M. S. Brandstein, "Microphone arrays: signal processing techniques and applications," in *chapter Robust localization in reverberant rooms*. Springer Verlag, 2001, pp. 157–180.

[4] A. Meyer and D. Döbler, "Noise source localization within a car interior using 3d-microphone arrays," in *Proceedings of the BeBeC*, 2006.

[5] V. C. Raykar and R. Duraiswami, "Automatic position calibration of multiple microphones," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4. IEEE, 2004, pp. iv–iv.

[6] B. Wesch, "ben-wes/pd-lotto," May 2025. [Online]. Available: https://github.com/ben-wes/pd-lotto

[7] F. J. Harris, *On the use of windows for harmonic analysis with the discrete Fourier transform*. IEEE, 2005, vol. 66, no. 1.

[8] G.-B. Stan, J.-J. Embrechts, and D. Archambeau, "Comparison of different impulse response measurement techniques," *Journal of the Audio engineering society*, vol. 50, no. 4, 2002.

[9] J. Benesty, J. Chen, and Y. Huang, "Time-delay estimation via linear interpolation and cross correlation," *IEEE Transactions on speech and audio processing*, vol. 12, no. 5, pp. 509–519, 2004.

[10] C. D. Meyer, *Matrix analysis and applied linear algebra*. SIAM, 2023.

[11] G. S. Wong, "Speed of sound in standard air," *The Journal of the Acoustical Society of America*, vol. 79, no. 5, pp. 1359–1366, 1986.

ISBN 978-3-200-10642-0