Acoustic Source Localization in Home Environments - The Effect of Microphone Array Geometry

Timon Zietlow[1]*, Hussein Hussein*[2] *and Danny Kowerko*[2]

[1] *Computer Graphics and Visualization Group, Technische Universität Chemnitz,*
[2] *Junior Professorship Media Computing, Technische Universität Chemnitz,*
D-09107 Chemnitz, Germany

{timon.zietlow,hussein.hussein,danny.kowerko}@informatik.tu-chemnitz.de

Abstract: Localization of acoustic sources is a powerful tool in different applications, for example, in quality assurance by locating the source of noise created by a machine or locating a person calling for help in ambient assisted living environments. This paper presents the first steps in the *localizeIT* project for acoustic source localization by simulation and comparison of different microphone array geometries to determine their effect on the localization performance. The time difference of arrival (TDOA) is used in the acoustic source localization method. The TDOA is estimated through an approximation of the phase shift of different signals calculated with the cross correlation operation. In order to find a good arrangement of microphones, a simulation is performed to compare different microphone arrangements for the presented localization method, with regard to the available space in a typical indoor environment. Those different microphone arrangements are classified into two categories: compact measuring stations positioned inside the monitored area, and larger satellite systems with microphones surrounding the monitored area. In the simulated scenario, the large satellite systems are superior to the compact measuring stations using the same number of microphones. Comparing different compact systems the results suggest, that the arrangement has a measurable and practical influence on the precision of the system.

1 Introduction

Acoustic source localization (ASL) is the estimation of the sound source position using its emitted acoustic signal. ASL is an important task in many near-field and far-field applications such as mobile robots, speech enhancement, surveillance systems, human-computer interaction, and other sensory applications like sonar [1][2][3][4]. ASL is based on the principle of acoustic wave propagation and the reception of acoustic waves by a set of microphones in a specific configuration (microphone array) in different times. The time delay estimation (TDE) of wave arrival to the microphones is usually used to calculate the distance to the sound source. This method calls Time Difference of Arrival (TDOA). The Direction of Arrival (DOA) of the sound source (considering both azimuth and elevation directions) can be estimated from the TDOA [1]. There are many TDE methods which vary in the degree of accuracy and computational complexity. The basic method is the calculation of Cross Correlation (CC) between signals received by two different microphones to find the peak point which corresponded to the time delay. The Generalized Cross Correlation (GCC) method is an improved method of CC. The most widely used method in ASL is the GCC due to its accuracy and moderate computational complexity. Different weighting functions can be used in GCC, for example, PHAse Transform

(PHAT), Maximum Likelihood (ML), ROTH correlation (ROTH) and Smoothed COherence Transform (SCOT). The weighting function is 1 in the standard CC method [3][5].

The classical array theory suggests that the distance between two microphones should be equal or smaller than half-wavelength of the sound source in order to avoid the spatial aliasing effect (spatial sampling theorem). Many microphone array geometries are used for the localization of acoustic sources. There are Uniform Planar Arrays (UPAs), for example, linear (Uniform Linear Array - ULA), rectangular (Uniform Rectangular Array - URA), circular (Uniform Circular Array - UCA) [6], and harmonically nested linear sub-arrays [7]. Furthermore, there are non-planar arrays such as spherical microphone arrays [2]. The azimuth estimation in the ULA causes an ambiguity of $180°$, but the UCA provides a $360°$ azimuthal coverage. The planar arrays provide a poor estimation in the elevation angles. Spherical microphone arrays can yield more accurate results by the azimuth and elevation estimation [2]. In addition to regular arrays, some arbitrary array configurations are used for the source localization [8]. Many microphone spacing schemes are suggested from uniform to logarithmic [6]. The localization performance can be improved if the microphones are close to the sound source [9].

The geometry of the microphone array or the selection of microphone positions plays an important role in the localization performance. We want to investigate the effect of the microphone array geometry on the localization performance by collection the microphones in a specific array (compact systems) or by distribution the microphones in the whole room (large satellite system). In both cases, the ASL algorithm must to yield good results for position estimation of the sound source which is located at any position in the room.

This paper is organized as follows: Section 2 gives an overwiew on the *localizeIT* project for object tracking using audio-visual information. The localization approach and the experiments' designs are described in section 3. Section 4 gives information about the evaluation of the gathered data and reviews the results. Finally, conclusions and future work are presented in Section 5.

2 localizeIT

The project *localizeIT*[1] is funded by the Federal Ministry of Education and Research in the program of Entrepreneurial Regions. The aim of the project is the tracking of objects by fusion of audio and visual information as well as the analysis of object behavior by using audio and visual information. Inside the tracking area, a number of passive sensors will be installed as shown in Figure 1. There are optical sensors (cameras with mono as well as stereo optics) and microphone arrays which will be distributed inside the area at different positions, several directions and heights to focus on different tasks and optimize theirs effectivity. The optical sensors can be hindered due to different physical properties. In this case, the acoustic sensors can be used instead. The fusion of audio and visual based object tracking can be used to improve the tracking task.

The main tasks in the project according to the audio processing are acoustic source separation, acoustic event detection and source localization. The experiment of classification of acoustic events which are usually occurred in the sector of healthcare are presented in [10]. In this contribution, the first step for acoustic source localization is implemented to detect the best geometry of microphone array by simulation of different arrangements of microphones.

[1] `http://www.localize-it.de`

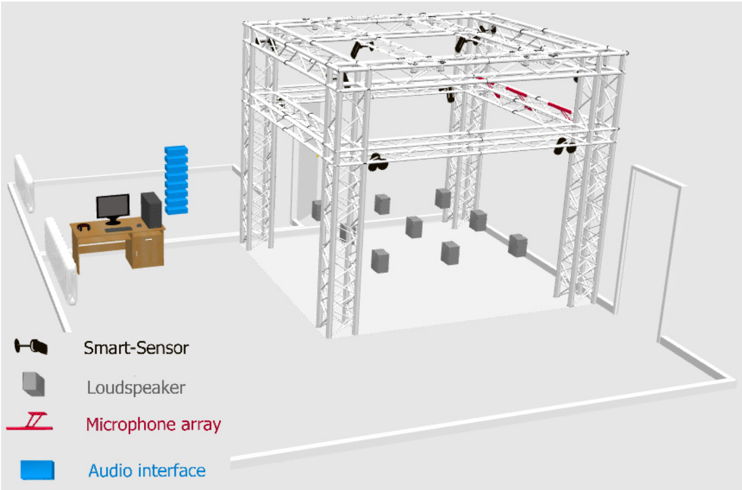


Figure 1 – Sketch of the Media Computing (MC) Audio-Video laboratory.

3 Experiment

We conducted two experiments to examine the effect of microphone arrangements on the localization performance. In the first experiment we tested small centralized microphone arrays observing their surrounding volume. In the second experiment we tried to compare those with large satellite arrays surrounding the observed area.

3.1 Localization Approach

In order to test the different arrangements' suitability for the audio source localization, a basic approach based on the TDOA is implemented. Assuming the two microphones i and j are located at the positions $\boldsymbol{m_i} \in \mathbb{R}^3$ and $\boldsymbol{m_j} \in \mathbb{R}^3$, the audio source is located at $\boldsymbol{s}$ and the TDOA of the signals recorded by the two microphones is $\Delta t_{i,j}$, the following constraint for the location of the audio source $\boldsymbol{s}$ can be formulated:

$$||\boldsymbol{s}-\boldsymbol{m_i}|| = ||\boldsymbol{s}-\boldsymbol{m_j}||+\Delta t_{i,j}\cdot c \tag{1}$$

with c being the speed of sound in the surrounding medium. So basically the distances from the audio source to the both microphones vary by the distance where the sound waves transmit in the TDOA. By considering the squared distances the following equations can be calculated:

$$\begin{aligned}
\delta_{i,j} &:= \Delta_{i,j}\cdot c && (2)\\
(\boldsymbol{s}-\boldsymbol{m_i})^2 &:= r_i^2 && (3)\\
(\boldsymbol{s}-\boldsymbol{m_j})^2 &:= (r_i+\delta_{i,j})^2 && (4)\\
\Rightarrow (\boldsymbol{s}-\boldsymbol{m_i})^2-(\boldsymbol{s}-\boldsymbol{m_j})^2+2\delta_{i,j}r_i &= -\delta_{i,j}^2 && (5)\\
(\boldsymbol{s}^2-2\boldsymbol{s}\boldsymbol{m_i}+\boldsymbol{m_i^2})-(\boldsymbol{s}^2-2\boldsymbol{s}\boldsymbol{m_j}+\boldsymbol{m_j^2})+2\delta_{i,j}r_i &= -\delta_{i,j}^2 && (6)\\
\boldsymbol{s}^2-2\boldsymbol{s}\boldsymbol{m_i}+\boldsymbol{m_i^2}-\boldsymbol{s}^2+2\boldsymbol{s}\boldsymbol{m_j}-\boldsymbol{m_j^2}+2\delta_{i,j}r_i &= -\delta_{i,j}^2 && (7)\\
2(\boldsymbol{m_j}-\boldsymbol{m_i})\boldsymbol{s}+2\delta_{i,j}r_i &= -\delta_{i,j}^2-\boldsymbol{m_i^2}+\boldsymbol{m_j^2} && (8)
\end{aligned}$$

For a fixed microphone i the equation (8) gives a system of non linear equations with one equation for each microphone $j \neq i$ and four unknown variables, which is solved using the Levenberg–Marquardt algorithm [11]. This requires at least five equations, which lead to the requirement, that our microphone array contains at least six microphones. In order to increase the precision of the algorithm and inure it against errors in the estimated $\Delta t_{i,j}$ values, we evaluated the system of nonlinear equations N times using every microphone as the fixed microphone i and averaging the results (N being the number of microphones in the array).

A similar effect could be achieved by formulation a larger system of nonlinear equations with $N+3$ unknown variables and $N*(N-1)$ equations. This approach can also be used to decrease the number of required microphones to four (for $N=4$, there are 12 equations and seven unknown variables).

3.2 Simulation of the Input Data

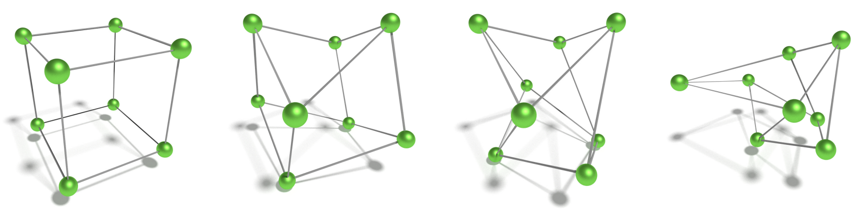
To test the different arrangements, only the position estimation as described above is simulated and not the estimation of the TDOA. So the TDOAs for a specific location of the audio source $\boldsymbol{s}$ are calculated using the following equation:

$$\Delta t_{i,j} = |||\boldsymbol{m_i} - \boldsymbol{s}|| - ||\boldsymbol{m_j} - \boldsymbol{s}||| \cdot c^{-1} \cdot n \qquad (9)$$

where n is a noise factor used to introduce some error to the system, which could be present due to signal transition times in the cable or the processing hardware, variations in air pressure, etc. In our simulations a factor of $n \in [0.99, 1.01]$ is chosen randomly for each $\Delta t_{i,j}$. In order to get meaningful data the experiments are repeated 25 times and the resulting errors (see section 4) are averaged.

3.3 Experiment I

In order to determine how good a microphone arrangement performs, the position estimation is done for simulated audio sources located in a regular grid surrounding the microphone array. To compensate the size of the array itself, the microphone arrays are scaled to a common size. The size of a microphone array is determined by the maximal distance between two microphones $d = argmax_{i,j} ||\boldsymbol{m_j} - \boldsymbol{m_j}||$. We used a $41 \times 41 \times 41$ grid with a $0.25m$ spacing between the neighboring points resulting in a $X \times Y \times Z = 10 \times 10 \times 10m$ volume. The evaluated microphone arrays are scaled to a size of $d = 0.25m$ and consisted of eight microphones in 4 different arrangements: Cube, Twisted I, Twisted II, and Random as shown in Figure 2. The microphone arrays are located in the middle of the evaluated volume.

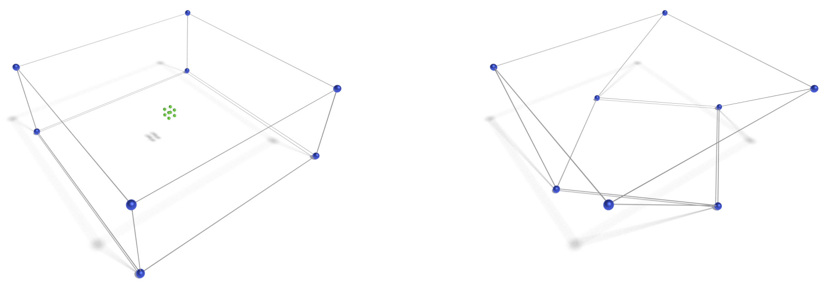


(a) Cube **(b)** Twisted I **(c)** Twisted II **(d)** Random

Figure 2 – The tested small microphone arrays.

3.4 Experiment II

As aforementioned the second part of the experiments is to compare compact centralized microphone arrays with the large distributed arrays. Since the focus of our work is in living room environment, the large microphone array is scaled to fit the living room ($X \times Y \times Z = 5 \times 5 \times 2.3m$). However the evaluated grid was the same as in Experiment I, allowing a direct comparison of the evaluated data. We tested two large arrangements. The first one with all microphones placed in the corners of the room as shown in Figure 3a. In the second arrangement, the top microphones are located in the corners of the room and the bottom microphones on the edges half way between the corners (Figure 3b).



(a) Room

(b) Twisted Room

Figure 3 – The smaller cube green array represents one of the compact arrays. The two large microphone arrays are presented in blue points.

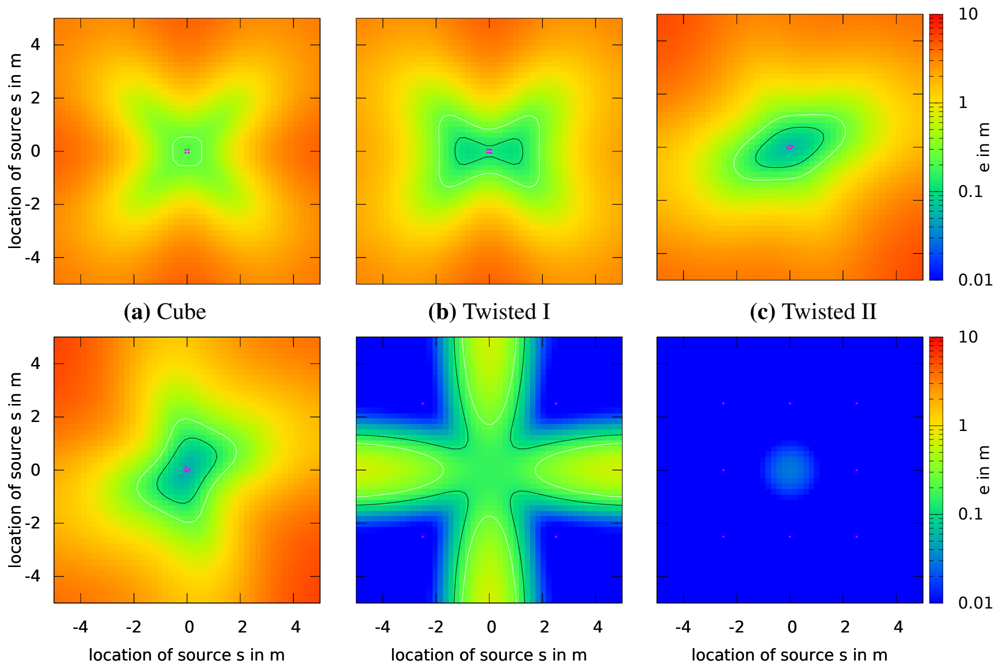
4 Results

In order to compare the different arrays, the resulting errors are evaluated by considering the distance of the calculated location (l_c) and the actual location (l_a) as follows:

$$e = ||l_c - l_a|| \tag{10}$$

The localization is considered as failure when $e > 0.125m$, since the l_c is then actually closer to the next evaluated point than to the actual location l_a. Overall the arrangement of the microphones has a huge influence on the shape and size of the well tracked area as illustrated in Figure 4. The compact arrays show that the error ratio increases with the distance rising between acoustic source and microphone array. The error ratio varies according to the shape of the compact array (see Figure 4: a, b, c, and d). If the microphones are placed in the corners of the room, the localization results are better close to these corners in comparision to the middle of the room (Figure 4: e), whereas the localization results are good in the whole room by using the twisted room arrangement (Figure 4: f).

The number of well enough estimated locations with $e < 0.125m$ shall be used to determine how good the array actually performed. The compact microphone arrays worked poorly (about 3%), especially for audio sources located far away. The larger arrays however worked very well with over 99% of the tested points resulting in an error of $e < 0.125m$ as shown in Figure 5.



(a) Cube (b) Twisted I (c) Twisted II

(d) Random (e) Room (f) Twisted Room

Figure 4 – Top view of the errors made due to the introduced error in meters. The black and white outlines mark an error of $0.125m$ and $0.25m$, respectively. The magenta dots mark the location of the microphones.

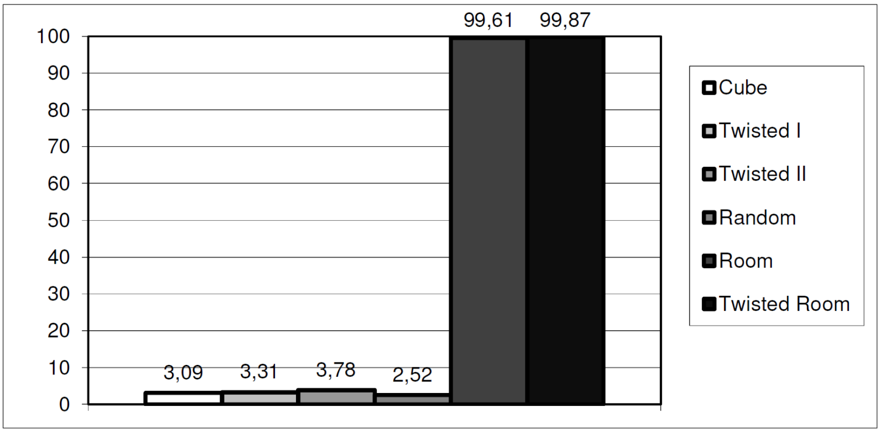


Figure 5 – Comparison of the different arrangements' performance.

5 Conclusion and Future Work

The paper presents the simulation and comparison of different geometries of microphone arrays in order to determine their effect on the localization performance. The presented experiments show, that the arrangement of the microphone array has a measurable influence in the precision of the presented method of audio source localization. It seems, that the size does matter in case of microphone arrays. Even with a constant number of microphones large arrays perform better in a larger area than with small compact microphone arrays. Therefore, it would be reasonable to spread microphones in the room. This approach is somehow impractical for mobile or non-stationary systems, but it is possible in many applications such as the acoustic observation of a patients in home environments. The localization results show that the randomized array is outperformed by any other arrangement. It appears, that the difference in distance between an audio source and different microphones should be maximized to get a stable indication for the localization of the audio source.

The next step in our research would be to find the best arrangement for a given number of microphones, array size or room geometry. In addition, the test of those arrangements in a real live setup in our laboratory.

6 Acknowledgements

This work is funded by the program of Entrepreneurial Regions InnoProfile-Transfer in the project group *localizeIT* (funding code 03IP608).

References

[1] ASTAPOV, S., J. BERDNIKOVA, and J. PREDEN: *Optimized Acoustic Localization with SRP-PHAT for Monitoring in Distributed Sensor Network*. *International Journal of Electronics and Telecommunications*, 59(4), pp. 383–390, 2013.

[2] HUANG, Q. and T. WANG: *Acoustic Source Localization in Mixed Field Using Spherical Microphone Arrays*. *EURASIP Journal on Advances in Signal Processing*, 2014(90), pp. 1–16, 2014.

[3] RITU and S. K. DHULL: *Review on Acoustic Source Localization Techniques*. *European Journal of Advances in Engineering and Technology*, 2(9), pp. 72–77, 2015.

[4] SALVATI, D., C. DRIOLI, and G. L. FORESTI: *On the Use of Machine Learning in Microphone Array Beamforming for Far-Field Sound Source Localization*. In *26th IEEE International Workshop on Machine Learning for Signal Processing, MLSP*, pp. 1–6. Salerno, Italy, 2016.

[5] FENG, M., Y. DIANGE, W. RUJIA, W. JUNJIE, W. ZITENG, and L. XIAOMIN: *A Triangulation Method Based on Phase Difference of Arrival Estimation for Sound Source Localization*. In *Proceedings of the 21st International Congress on Sound and Vibration (ICSV)*. Beijing, China, 2014.

[6] TELLAKULA, A. K.: *Acoustic Source Localization using Time Delay Estimation*. Master's thesis, Bangalore , India, 2007.

[7] ELKO, G. W. and J. MEYER: *Springer Handbook of Speech Processing*, chap. Microphone Arrays, pp. 1021–1041. Springer-Verlag, Berlin Heidelberg, 2008.

[8] COSTA, M., V. KOIVUNEN, and A. RICHTER: *Low Complexity Azimuth and Elevation Estimation for Arbitrary Array Configurations*. In *2009 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2185–2188. 2009.

[9] WARD, D. B.: *Performance of Microphone Array Geometries in a Reverberant Room*. *European Transactions on Telecommunications*, 13(2), pp. 133–138, 2002.

[10] HUSSEIN, H., M. RITTER, R. MANTHEY, J. SCHLOSSHAUER, E. FABIAN, and M. HEINZIG: *Acoustic Event Classification for Ambient Assisted Living and Healthcare Environments*. In O. JOKISCH (ed.), *Proceedings of the 27th Conference on Electronic Speech Signal Processing (ESSV)*, vol. 81 of *Studientexte zur Sprachkommunikation*, pp. 271–278. TUDpress, Leipzig, Germany, 2016.

[11] LEVENBERG, K.: *A Method for the Solution of Certain Non-Linear Problems in Least Squares*. *Quarterly Journal of Applied Mathmatics*, 2(2), pp. 164–168, 1944.