Extension and Accuracy Evaluation of a Pointing Method for 3D Auditory Localization Experiments

Student Project in Sound Engineering

Ludwig Mohr

Supervisors: Dipl.Ing. Matthias Frank, Dr. Franz Zotter

Graz, January 27, 2011

kunst uni graz

institut für elektronische musik und akustik

iem

Abstract

Multichannel audio reproduction systems in two dimensional and even three dimensional setups are encountered with increasing frequency in e.g. concert halls, cinemas and even home theaters for movies and video games. Algorithms and techniques for reproducing an acoustic field accurately are a vital area of research.

One important parameter for the perception of spatial sound quality is the match between the presented and actually perceived direction of an incident sound event. Listening experiments are essential in assessing these performance parameters. For obtaining quantified information on the observed direction methods enabling a human test subject to map their perception onto directional parameters such as azimuth and elevation are necessary. These methods should be as intuitive as possible while at the same time offering high accuracy.

This project extends a promising method for localization experiments for easier employment. This method is highly intuitive as well as accurate. The absolute accuracy is evaluated, both in static and subjective situations.

Contents

1 Introduction

1.1 Motivation

Multichannel sound reproduction systems for both two dimensional and three dimensional loudspeaker setups are getting more and more common. Algorithms for accurate reproduction of acoustic fields are an area of ongoing research. One performance measure common to all available systems is the extent of mismatch between intended and perceived directionality of an acoustic cue. At this time however, all models for acoustic localization are insufficient, especially those able of incorporating the influence of the listening room. For this reason it is inevitable to conduct listening experiments in order to assess the performance of a given sound field reproduction system.

Several methods for quantifying the perceived localization, i.e. infering perceived azimuth and elevation relative to the listening position, exist. A considerable trade-off between ease of use, accuracy, implementation complexity, and obtrusiveness has usually to be accepted.

A system for localization experiments is available at the Institute of Electronic Music and Acoustics (IEM) of the University of Music and Performing Arts Graz from an earlier project [FZS08]. This system holds promise to be able to reduce this trade-off considerably. As of yet, it is operable in one room of the institute only (IEM Cube, [cub]). This project extends this system to be compatible with any device capable of tracking all six degrees of freedom of an object. Furthermore, a simple approximation for the surrounding hull spanned by the loudspeaker positions is implemented.

The project is completed by an evaluation of the accuracy of the system. The exact surrounding hull requiring correct knowledge of the loudspeaker positions as well as the approximate hull are considered.

1.2 Organization of the Report

This project report is organized as follows.

Section 2 starts with an overview of the different method classes for quantifying an indicated direction in auditory localization experiments and discusses requirements for these methods. Subsection 2.1 attends to their inherent advantages and drawbacks. Subsection 2.2 introduces the pointing method for localization experiments available at the IEM.

Section 3 focuses on the technical aspects of the modifications made to the method, attending the topics of hull approximation, filters for input data smoothing, and output buffering.

The method evaluation is discussed in section 4. Static accuracy is evaluated in subsection 4.1, subjective accuracy is attended to in subsection 4.2. This section concludes in subsection 4.3 with a comparison of the two available models for the surrounding hull in terms of relative deviation.

Section 5 summarizes the work accomplished in this project and gives an outline for further modifications.

2 Methods for the Assessment of Directionality in Localization Experiments

Methods for indicating and quantifying perceived localization in the context of sound reproduction face up to the following requirements:

Accuracy is required to reduce the observational error as much as possible. Ideally, the uncertainty intrinsic to the method should be well below 1°. The human auditory system is limited to a spacial resolution of about 1° where most sensitive [Bla74].

Complexity. For reasons of expenditure and feasibility methods that are more easily employed are in preference.

Intuitivity is important in order that test subjects will be less distracted from the main task. Intuitive methods may also reduce instruction time and indication error.

Obtrusion inherent to the method will most certainly have a negative impact on test subject performance. If for reasons of an assured listening position the test subjects head is fixated, the resulting discomfort and reduced mobility will have an adverse effect on the indication error.

2.1 Overview of Method Categories

Methods for measuring perceived directionality in localization experiments can be divided roughly into four categories: verbal methods, graphical methods, methods of adjustment, and pointing methods.

Verbal methods such as stating a direction explicitly in terms of azimuth and elevation angle are easily implemented. Yet they are hardly intuitive and greatly limited by the subjects ability to accurately describe directions [WK97, MFRdB01]. Naming one loudspeaker perceived as source out of several visible loudspeakers is only applicable whenever locations of phantom or virtual sources in between loudspeakers are not of interest. Aiding the indication by markings on an acoustic transparent fabric in front of the speaker setup generates either a bias towards the markings or limits the employable resolution drastically.

An example of the category of graphical methods is drawing the perceived location of the stimulus into a sketch or plan of the surrounding space. Methods of this category require the test subject to map a three dimensional space onto the two dimensions of a piece of paper or a computer screen. This mapping is prone to generating errors [MFRdB01].

Methods of adaption are applicable for localization experiments, too. The subject matches the perceived localization of a controllable sound source with that of the test stimulus. However, it is disputed if a different characteristic, e.g. timbre, is matched rather than matching directionality [PK01]. Systematic deviations, wether they originate from the algorithm or the loudspeaker setup, will not be observable when the two stimuli are generated by the same system. However, if a different playback system for the adjustable source is used, e.g. a separate speaker mounted on a robotic arm (cf. [ROP05]), the difference in topology of the playback systems will distort the result.

Pointing methods have been proven to be most suitable for the task in terms of intuitivity while at the same time retaining a high level of accuracy. Among the several options for indicating a direction (a more complete overview is available in [See03]) the probably most intuitive option is to use human gestures. Usable gestures are, yet not restricted to, looking into a direction, turning head or body towards that direction and using a hand or a handheld object to point at the observed source. Restrictions applying to these methods are e.g. the head and eye movements limited by human physiology. Furthermore is tracking of eye-movement a technologically complex and expensive task [SSJW10]. The lack of proprioceptual decoupling [See03] and further influences [MLGM08] add to the uncertainty of obtained results. Research suggests these results to be uncertain to about 5°. These values may be improved by visual feedback, yet results still deviate greater than the spacial resolution of the human auditory system [MLGM08, Bla74].
Different approaches such as pointing a laser beam controlled by a joystick or trackball are proprioceptually decoupled well and achieve a high accuracy. Yet their handling may be counter-intuitive, research suggests they produce a bias towards the direction initially indicated by the device. [See03, Lew98].

For testing the lateralization of the perceived direction head movement has to be prevented in order to exclude dynamic localization cues as well as frontal localization cues. A mechanical fixation of the subjects head introduces a bias towards the line of sight.

The pointing method designed at the IEM [MLGM08, FZS08] promises to be superior in terms of intuitive handling and the ability to prevent head-movement during listening, if required, while maintaining a high accuracy.

2.2 "WiiVolver": Pointing Method at the IEM

The method implemented at the IEM is based on an infrared tracking system manufactured by VICON [Vic] equipped with 15 M-Series infrared cameras. The system is able to track all six degrees of freedom (DOF) of any object equipped with retro-reflective markers. The coverage extends over almost the whole room, tracking precision can be presumed to be constant for every position covered.

As pointing device a toy-gun equipped with a game controller and reflective markers is used (cf. Figure 1). The gun comes with iron sights and a laser pointer which can be used to give additional visual feedback while aiming. However it is advisable to manually calibrate especially the laser pointer before use. The subject announces the perceived direction by pulling the trigger. The trigger actuation is captured by the game controller and sent to the evaluation software via bluetooth. The trigger signal logs the current direction indicated. This direction is calculated using the piercing point of the projection of the gun's direction with the surrounding hull spanned by the loudspeakers. The direction indicated by the subject is evaluated with respect to an arbitrary reference point. As reference point usually either the position of the subject's head or the center of the loudspeaker setup is used.
The additional buttons on the controller may be used for other tasks such as controlling the experiment sequence or evaluation of scale based attributes.

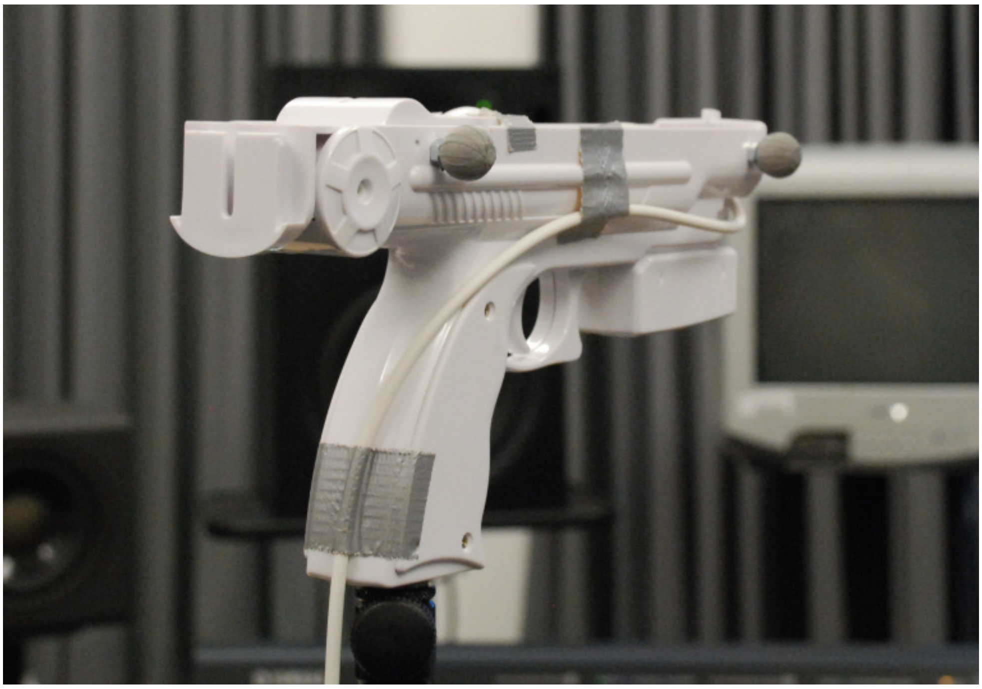


Figure 1: "WiiVolver" toy-gun equipped with retro-reflective markers and a game controller, mounted on a microphone stand.

The position of the subjects head can be tracked by another 6 DOF sensor. In virtually any case, it is important to monitor the subjects head as localization is strongly depending on its position and orientation. If necessary, the subject can be informed if the head position and orientation is out of a defined range, and may hereby autonomously realign. The advantage compared to an otherwise functional mechanical fixation of the head is the minimization of annoyance and the possibility to loosen the head-fixation for aiming when the sound cue has stopped.

All necessary computations are implemented in the real-time audio processing language Pure-Data [pur] and run on a standard desktop computer.

3 Technical Aspects and Modifications

This section discusses the modifications made to the original implementation. All aspects not changed are discussed in-depth in [FZS08].

The system is generally compatible with all tracking devices tracking 6 DOF as long as it is possible to stream the position data directly into the Pure-Data implementation. Depending on the data representation in the stream only small adaption in the code have to be made.

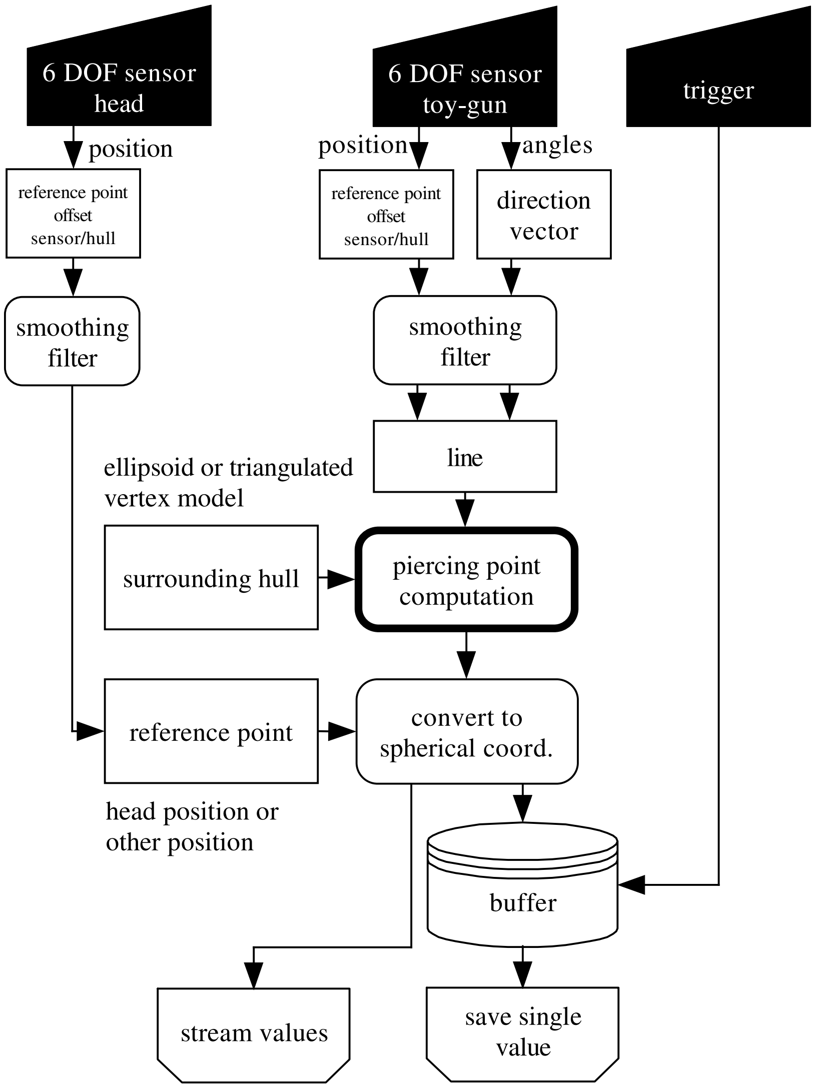


Figure 2: Complete flow chart of the calculation of the indicated direction.

3.1 Triangulated Convex Hull

If the exact positions of all loudspeakers is available the surrounding hull can be modeled exactly. The loudspeaker positions represent the vertexes of the hull, three neighboring vertexes are grouped to form a set defining the edges of a triangle. This is done for every combination of neighboring vertexes. The calculation of the piercing point of the projected line from the toy-gun and the surrounding hull reduces to the calculation of the piercing point of the line with the plane spanned by the triangle which is pierced. The algorithm determining the correct triangle is discussed in [FZS08].

As of now, the hull is hardcoded with the positions of the loudspeakers in the IEM Cube, the vertexes have been assigned manually into the sets defining the triangles. Especially for acoustic reproduction system of high order this task is tedious and prone to errors.

Due to the close resemblance of the surrounding hull to a convex hull it is possible to alter an algorithm for convex hull creation to automate this task. Vertexes closer to the center of the hull than all its neighbors will be discarded when presented to a standard convex hull algorithm. To automatically create the triangulated convex hull from the loudspeaker positions all distances to the center have to be set to the same value. These "normalized" values are then presented to the convex hull algorithm. After the hull has been created the correct distances will be used again.

Convex hull algorithms are available for MATLAB at [mat], an open source version is available at [qhu].

3.2 Approximate Ellipsoid Hull

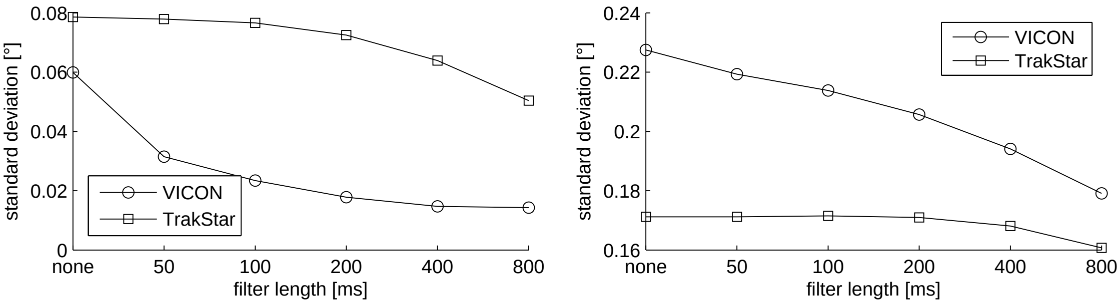
The surrounding hull may be approximated with an ellipsoid defined by the three radii r_1, r_2 and r_3 (1). The tracking system supplies us with the position $\boldsymbol{p} = [p_1, p_2, p_3]^T$ and the three orientation angles ψ, θ and ϕ of the toy-gun. The toy-gun's spin is irrelevant for this application, therefor ψ is discarded. Azimuth θ and elevation ϕ are transformed into the normalized direction vector $\boldsymbol{v}$. Position $\boldsymbol{p}$ and direction $\boldsymbol{v}$ form together the line $L = \{\boldsymbol{p} + \lambda\boldsymbol{v} | \lambda \in \mathbb{R}\}$. The piercing point $\boldsymbol{x} = [x_1, x_2, x_3]^T$ has to lie on line L (2) and on the hull of the ellipsoid (1).

$$\frac{x_1^2}{r_1^2} + \frac{x_2^2}{r_2^2} + \frac{x_3^2}{r_3^2} = 1 \tag{1}$$

$$\boldsymbol{x} = \boldsymbol{p} + \lambda_{\boldsymbol{x}}\boldsymbol{v} \tag{2}$$

This leads to a quadratic equation for $\lambda_{\boldsymbol{x}}$ whose analytic solution is (3)

$$\lambda_{\boldsymbol{x}} = \frac{-b + \sqrt{b^2 - 4ac}}{2a}, \tag{3}$$



(a) stationary toy-gun

(b) hand-held toy-gun

Figure 3: Standard deviation of angle error for different smoothing filter lengths.

with:

$$a = \frac{v_1^2}{r_1^2} + \frac{v_2^2}{r_2^2} + \frac{v_3^2}{r_3^2},$$

$$b = 2\left(\frac{v_1 x_1}{r_1^2} + \frac{v_2 x_2}{r_2^2} + \frac{v_3 x_2}{r_3^2}\right),$$

$$c = \frac{x_1^2}{r_1^2} + \frac{x_2^2}{r_2^2} + \frac{x_3^2}{r_3^2} - 1.$$

Depending on the position of the origin of the loudspeaker array it may be necessary to shift the ellipsoid by a vector $\boldsymbol{s} = [s_1, s_2, s_3]^T$ such that the center of the ellipsoid and the center of the loudspeaker setup match each other.

The triangulated vertex hull model and the approximate ellipsoid hull model are compared in section 4.3.

3.3 Filtering the Tracking Data

Figure 2 shows the smoothing filter applied to the tracking data at the input of the algorithm. This is done in order to reduce the impact of sensor noise present in the tracking device and smooth the tremor of the hand-held toy-gun. Figures 3(a) and 3(b) show the standard deviation of the angle error as a function of filter length for the two available tracking systems. Stationary and hand-held toy-gun are considered.

The filter length is given in *milliseconds* since the actual length in frames varies with the frame rate of the tracking device.

It is apparent, that the TrakStar system already applies filtering at its output since additional filtering has minimal impact. A filter length of 50ms is enough to reduce the VICONs error by 50% in the stationary case, a filter length of 200ms reduces the error

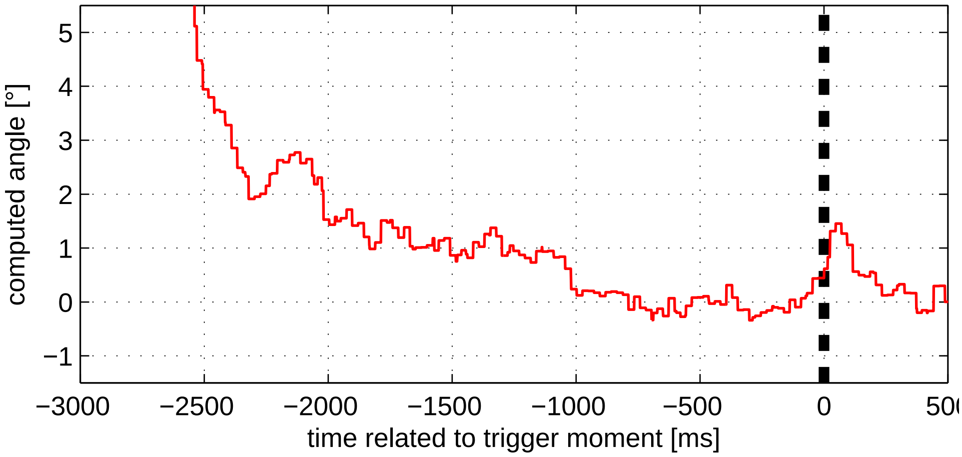


Figure 4: Indicated direction during the process of aiming after listening.

by 10% for the hand-held toy-gun. The VICON system seems to be more error-prone in dynamic tracking compared to the TrakStar system (inside its range) regarding the standard deviation. However, it still offers lower angle errors on average (cf. section 4).

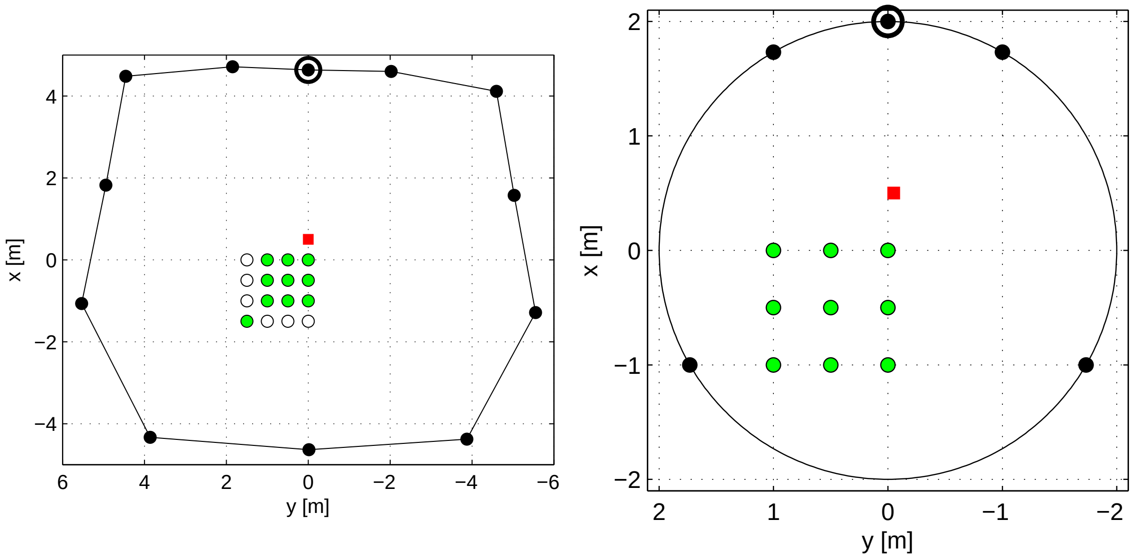
3.4 Output Buffering

Figure 4 shows a typical process of aiming (VICON system in the IEM Cube, no filtering). It is apparent, that the trigger actuation (dashed line) causes a deviation from the indicated direction.

The characteristics of the distortion and its direction vary with subject and position. Most subjects show a preferred direction, which seems to be independent of right-handedness or left-handedness. The exact angle is already shown approximately one second before trigger actuation. Even the fastest subject tested aims at least 200ms at the target before pulling the trigger. This observation is used to compensate for the deviation from the indicated direction during trigger actuation by buffering the output values. The buffer parameters have to be adjusted depending on the latency of the tracking system, the input filtering, and the latency introduced by signal transmission via bluetooth.

4 Evaluation of the Accuracy of the System

The accuracy of the method was evaluated using two tracking systems (VICON, 15 M-Series infrared cameras, 120fps, full room coverage [Vic] and Ascension TrakStar, DC magnetic field, 240fps, 76cm range [tra]) in the IEM Cube using the triangulated vertex hull model and with the Ascension TrakStar in the IEMs production studio using the approximate ellipsoid hull model. Both static (cf. subsection 4.1) and subjective accuracy (cf. subsection 4.2) have been evaluated. In order to collect comparable data the toy-gun was tracked by both tracking systems simultaneously where applicable. The indicated direction is calculated in reference to the center of the respective hull model.



(a) IEM Cube (b) IEM production studio

Figure 5: Measurement setups: measurement positions (green), additional positions for static tests (white), TrakStar transmitter (red square), lowest plane of vertexes/loudspeakers (black), the optical target is encircled.

According to manufacturer's data the accuracy of the TrakStar system is depending mostly on the distance between transmitter and sensor. The accuracy and coverage of the VICON system is assumed to be uniform within the measurement area. Therefore, the measurement is performed for a part of the considered area. As the accuracy can be assumed to be independent of the toy-gun's orientation and the target position, the measurement is performed with a single optical target. The measurement is concentrated on azimuthal angles in order to simplify the illustrations. Figures 5(b) and 5(a) illustrate the measurement positions for the IEM Cube and the IEM production studio, respectively.

In subsection 4.3 the deviation of the approximate ellipsoid hull model from the triangulated vertex hull is evaluated empirically.

4.1 Static Accuracy

The static accuracy was measured with the toy-gun mounted on top of a microphone stand. The gun was adjusted to indicate towards the target from each of the positions on the 50cm grid (cf. green and white circles in Figure 5) using both laser and iron-sights. For each position 10s of output data of the algorithm have been recorded. No additional filtering has been applied to the data.

Figures 7 and 8(a) show the absolute angle error in terms of absolute value of the mean (big bars) and standard deviation (small bars). The bar color indicates whether the sum of mean and standard deviation lies above (red) or below (green) the spatial resolution of human hearing (approximately 1°, [Bla74]).

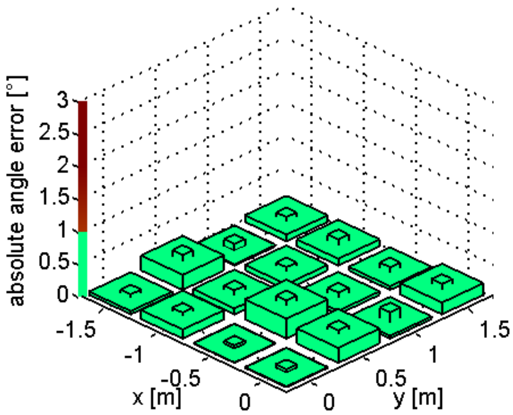
Figure 7(a) suggests that the assumption of uniform coverage and accuracy for the VICON system is true at least for the center of the room. Errors well below 1° are achieved. The error introduced by the Ascension TrakStar shows an explicit dependency of the distance of sensor and transmitter (cf. figures 7(b) and 8(a)). In the IEM Cube absolute errors below 1° are achieved even for some positions beyond the 76cm range from the manufacturer's data (cf. figures 7(b)). In the production studio errors below 1° are only achieved for positions within the operational range. This is due to the larger distance of the transmitter to the center of the room resulting from technical restrictions as well as interference by ferromagnetic objects close to the tracking system (e.g. the mixing desk present in the production studio, c.f. figure 6).

4.2 Subjective Accuracy

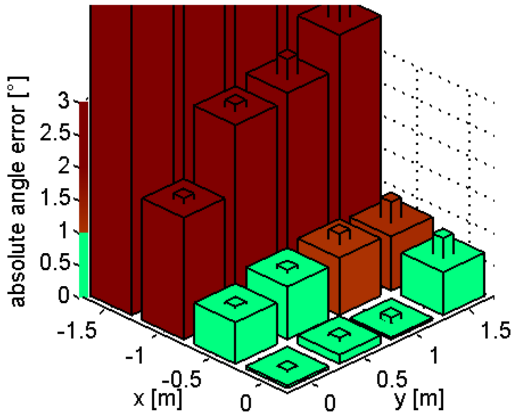
When used in listening tests the toy-gun will be hand-held, thereby affecting the indicated position by the subject's visual acuity and motor skill. In order to assess the system error involving this effect the experiments from subsection 4.1 are repeated with five test subjects holding the toy-gun and aiming at the target. The number of tested positions has been reduced to those denoted by green circles due to the already enormous static error for more distant positions.

Figures 8(b) and 10 show the results in the same fashion as for the static evaluation. Mean error and especially standard deviation have increased notably. The accuracy of the VICON system remains below the 1° mark, the TrakStar now has more positions with an error below or close to the 1° limit. This is due to the toy-gun being extended up to approximately 60cm towards the target when hand-held, thus the distance of the TrakStar's emitter and sensor is reduced.

Figure 6: Measurement setup for static accuracy in the production studio, measurement positions marked on the floor.

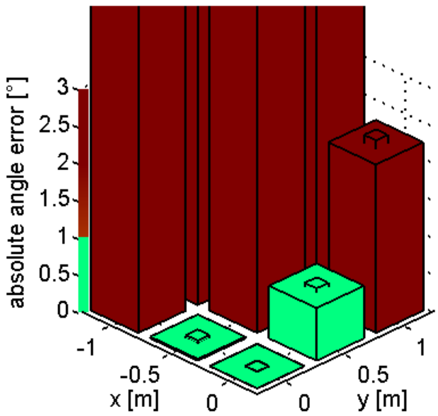


(a) VICON

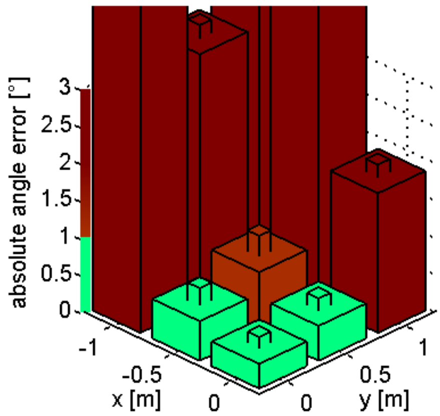


(b) TrakStar

Figure 7: Absolute error for the stationary toy-gun in terms of mean (big bar) and standard deviation (small bar), sum of mean and standard deviation indicated by color.



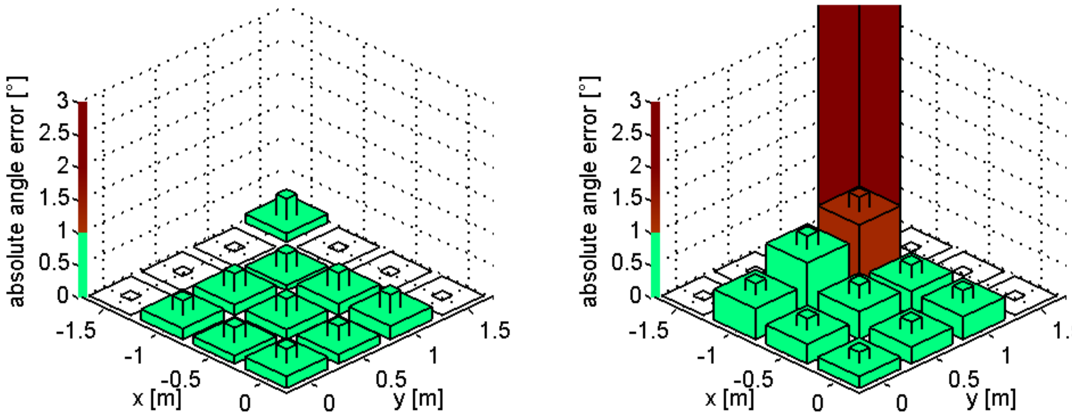
(a) static measurement



(b) subjective measurement

Figure 8: Absolute error of the Ascension TrakStar for both stationary and subjective tests. Depiction of standard deviation and mean as in Fig. 7.

Figure 9: Aiming using the iron-sights during the subjective accuracy measurements.



(a) VICON (b) TrakStar

Figure 10: Absolute error for the hand-held toy-gun in terms of mean (big bar) and standard deviation (small bar), sum of mean and standard deviation indicated by color.

4.3 Model Comparison

In order to assess the uncertainty introduced by the approximation of the surrounding hull with an ellipsoid the static accuracy has been evaluated again for the positions shown in figure 11 using the VICON system. The indicated direction has been calculated for the triangulated vertex hull model and an ellipsoid hull model with $\boldsymbol{r} = [4.25, 5, 3]^T$ m. The absolute deviation of the calculated angle with respect to the triangulated hull model is shown in figure 12 as function of the distance between measuring position and origin. The model deviation increases with greater distance to the origin due to an increasing parallax error.

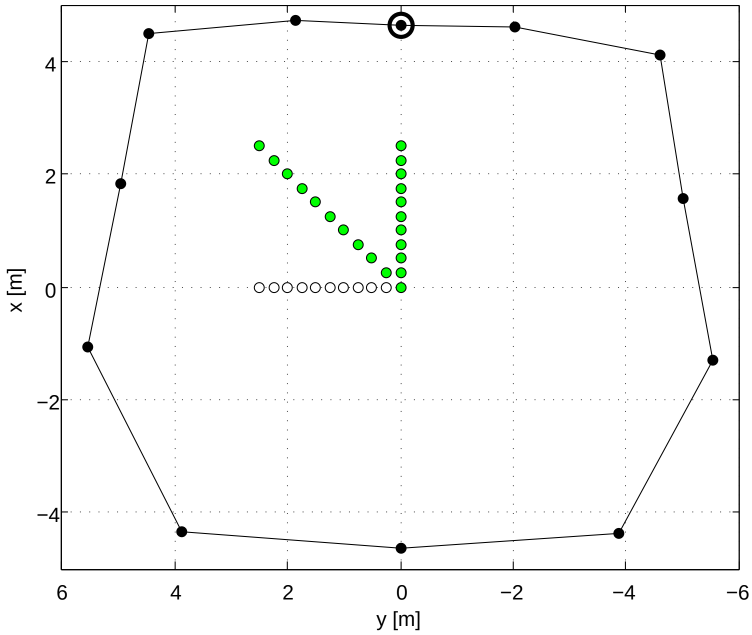


Figure 11: Measurement positions for the model comparison. Figure 12 shows the results for the points along the x-axis and along the diagonal.

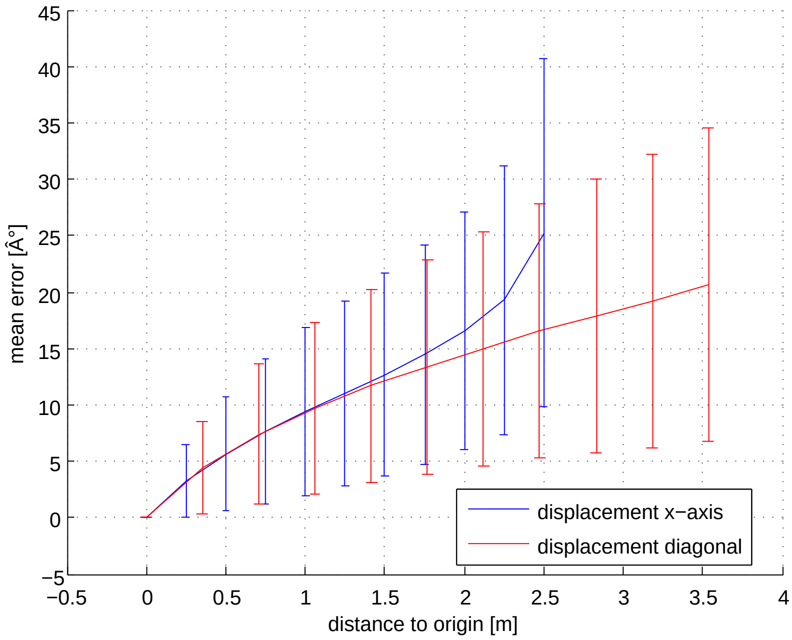


Figure 12: Deviation of the resulting direction when using the approximate model relative to the values of the triangulated vertex hull model.

5 Conclusion

The pointing method can now be used with either an accurate model for the surrounding hull using triangulated vertexes or with a simple approximation of the hull as an ellipse. Creation of the triangulated hull model given the exact positions of the loudspeakers has been automated based on convex hull algorithms.

It has been shown that smoothing filter and output buffer have to be adjusted based on the characteristics of the tracking system in use. Otherwise however it is possible to employ this pointing method in three dimensional localization experiments using any tracking system able to track 6 degrees of freedom and providing the tracking results as output stream to another application.

Measurements suggest that systematic errors introduced by the pointing method stay well below the mark of 1° affiliated with the spatial resolution of the human auditory system, as long as the operational range of the tracking system is not exceeded. It is possible to receive these results with a comparably low-priced tracking system (Ascension TrakStar).

Further extensions possible are a complete open-source implementation in Pure-Data to provide the possibility to create the vertex hull model in the software suite without the need of other (proprietary) tools. The extension of the triangulated vertex hull model to a rhombic hull model via an additional vertex far below the speaker setup is planned. With this every direction indicated will have a defined output.

References

[Bla74] J. Blauert, *Räumliches Hören*. S. Hirzel-Verlag, 1974.

[cub] "Iem cube at the institute of electronic music and acoustics, graz." [Online]. Available: http://iem.kug.ac.at/en/services/jahresberichte/200001/cube.html

[FZS08] M. Frank, F. Zotter, and A. Sontacchi, "Localization experiments using different 2d ambisonics decoders," presented at the 25th Tonmeistertagung, Leipzig, Germany, 2008.

[Lew98] J. Lewald, "The effect of gaze eccentricity on perceived sound direction and its relation to visual localization," in *Hearing Research*, vol. 115, 1998, pp. 206–216.

[mat] "The mathworks, matlab." [Online]. Available: http://www.mathworks.com/

[MFRdB01] R. Mason, N. Ford, F. Rumsey, and A. de Bruyn, "Verbal and nonverbal elicitation techniques in the subjective assessment of spatial sound reproduction," *Journal of the Audio Engineering Society*, vol. 49, no. 5, pp. 366–383, 2001.

[MLGM08] P. Majdak, B. Laback, M. Goupell, and M. Mihocic, "The accuracy of localizing virtual sound sources: Effects of pointing method and visual environment," presented at the DAGA 2010, Berlin, Germany, 2008.

[PK01] V. Pulkki and M. Karjalainen, "Localization of amplitude-panned virtual sources i: Stereophonic panning," *Journal of the Audio Engineering Society*, vol. 49, no. 9, pp. 739–752, 2001.

[pur] "Pure-data." [Online]. Available: http://www.puredata.info/

[qhu] "Qhull." [Online]. Available: http://www.qhull.org/

[ROP05] B. Razavi, W. O'Neill, and G. Paige, "Both interaural and spectral cues impact sound localization in azimuth," in *Neural Engineering, 2005. Conference Proceedings. 2nd International IEEE EMBS Conference on*, Dept. of Biomed. Eng., Rochester Univ., NY. IEEE, 2005.

[See03] B. Seeber, "Untersuchung der auditiven lokalisation mit einer lichtzeigermethode," PhD Thesis, University of Technology, Munich, 2003.

[SSJW10] S. Spors, R. Schleicher, D. Jahn, and R. Walter, "On the use of eye movements in acoustic source localization experiments," presented at the AES 124th Convention, Amsterdam, The Netherlands, 2010.

[tra] "Ascension technology corporation." [Online]. Available: http://www.ascension-tech.com/

[Vic] "Vicon motion systems." [Online]. Available: http://www.vicon.com/

[WK97] F. Wightman and D. Kistler, "Monaural sound localization revisited," *The Journal of the Acoustical Society of America*, vol. 101, no. 2, pp. 1050–1063, 1997.