

Flexible and intuitive pointing method for 3D auditory localization experiments

Matthias Frank¹, Ludwig Mohr², Alois Sontacchi¹, Franz Zotter¹

¹*Institute of Eletronic Music and Acoustics, University of Music and Performing Arts Graz, Austria*

²*Student, University of Music and Performing Arts Graz, Austria*

Correspondence should be addressed to Matthias Frank (frank@iem.at)

ABSTRACT

This paper presents an intuitive pointing method for measuring the perceived direction in 3D localization experiments. The method uses a motion tracked toy-gun as pointing device and can be used from all positions in any nearly convex surrounding hull or loudspeaker setup, as the pointed direction is computed from the piercing point of the gun's direction and the surrounding surface. The reference point for the pointed direction can be chosen freely. The computation is implemented in real time open-source software and works with any 6 degrees of freedom tracking system. In this paper the accuracy of the method is measured for the static and hand-held case using 2 tracking systems in 2 rooms/surrounding hulls.

1. INTRODUCTION

One important parameter for the perception of spatial sound quality is the match between the presented and perceived direction of an incident sound event. This paper compares a family of particularly suitable and accurate devices for the indication of the perceived localization. Particularly, these devices shall be superior in their accuracy and intuitive in their usage. Both will help us in order to gain a more detailed knowledge about localization by experimental evaluation, to improve and validate theoretical models of human auditory localization, and to evaluate different methods for audio spatialization (e.g. WFS, Ambisonics, VBAP). Depending on the application, a dedicated reference point for the perceived direction can be necessary. Furthermore, the possibility to evaluate at multiple listening positions is sometimes required. This paper proposes a method for auditory localization experiments that offers both options with negligible parallax errors.

2. METHODS FOR LOCALIZATION EXPERIMENTS

There are several methods for measuring the perceived direction in localization experiments, which differ in aspects of accuracy and complexity or intuitivity for the subject. The methods can be roughly divided in 4 categories: verbal methods, graphical methods, methods of

adjustment, and pointing methods.

Verbal methods, as direct naming of a direction in terms of azimuthal and elevation angle [1], are technically easily implemented but lack the feature of intuitivity and are limited by the subject's ability to describe directions accurately [2]. Naming one out of several visible loudspeakers which the subject perceives to be the sound source becomes inappropriate whenever the location of phantom or virtual sources lying between the loudspeakers are considered.

Graphical methods, as drawing the audible direction into a sketch or plan of the surrounding space, require the subject to map a three dimensional space into two dimensions of a piece of paper or a screen and hereby inherit systematic errors [2].

Methods of adjustment are also applicable for localization experiments [3]. The subject's task is to match the direction of a controllable sound source with the direction of a reference source. However it is not sure, whether the subject matches the sources according to their direction or other parameters (e.g. timbre).

Pointing methods have been proven to be most suitable in regard of intuitivity while retaining a high level of accuracy.

2.1. Pointing Methods

There are several different options for a subject to point towards a certain direction (an overview can be found

in [4]). Quite naturally, human gestures offer typical ways of doing so. For instance, looking into a certain direction, turning the head or even the whole body towards that direction, or using a combination of these gestures is intuitive.

However pointing methods exhibit several restrictions. Pointing by eye and head gestures is restricted by the human physiology. Moreover, the tracking of eye movement is a technical challenge [5]. Results obtained by these pointing methods tend to be inaccurate due to a lack of proprioceptual decoupling [4], and other issues [6]. Research suggests that results obtained via these methods are inaccurate by approx. 5° . However they can be improved by adding visual feedback, like pointing with a finger or an object held in one hand. This has been shown to improve the results, especially in the vertical direction, where eye, head, and body movements are particularly limited [6]. Although the improvement reduces the bias, results deviate at a larger scale as the spatial resolution of human hearing [7].

Other methods in which subjects use a joystick, trackball, or knob to control a laser point [4, 8] achieve good proprioceptual decoupling and are accurate. Nevertheless, their handling may be counter-intuitive, and subjects tend to make too small movements with respect to the initially indicated direction [4].

If the lateralization of the perceived direction is to be tested, it is sensible to prevent the subjects from moving their head during listening. This excludes dynamic localization cues or the naturally superior frontal localization cues. Nevertheless, preventing movements by mechanical fixation of the subject's head causes a geometrical bias of the indicated direction towards the subject's line of sight.

The pointing method used in [6] and [9] seems to be superior in its intuitive handling and with respect to the design of listening experiments requiring a given static head position during listening. The following sections extend the method to convex listening rooms, and evaluate the method by measuring its accuracy.

3. PROPOSED POINTING METHOD

As pointing device, the proposed method uses a toy-gun (see figure 1) with laser and iron sights for optical aiming. The position and orientation of the gun is captured by a tracking system with a 6 degrees of freedom (DOF) sensor. The pointed direction is computed from the projection of the gun's direction on the surface of the sur-

rounding hull or loudspeaker setup. Thus the method works at any listening position within the range of the tracking system, for every nearly convex shaped hull or loudspeaker setup, and offers a full 3D directional coverage. The pointed direction can be referred to the position of the subjects head or to any other reference point; e.g. for spatial audio reproduction system it is common to refer the pointed direction to the center of the loudspeaker system.

The trigger moment is captured by a game controller and transmitted to the computer. The buttons on this controller can be used for additional tasks like controlling the experiment sequence or evaluation of scale based attributes.

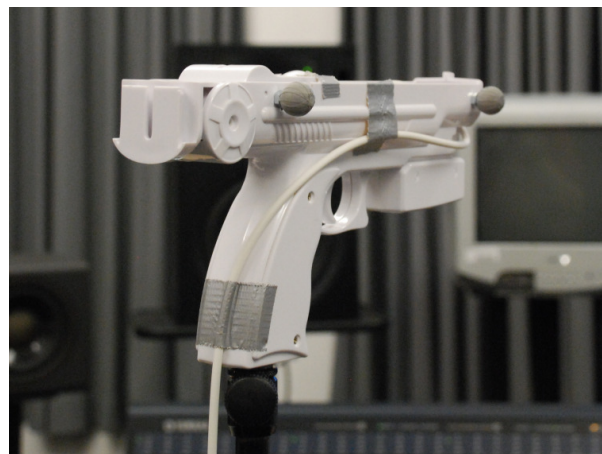


Fig. 1: Mounted toy-gun with reflectors for infrared tracking

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 the localization is strongly dependent 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 align. 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 the aiming task after listening.

As a model of the surrounding hull, we use either a convex hull algorithm or a simple ellipsoid. In the first case the vertexes of the hull (e.g. the loudspeaker positions) have to be defined, whereas in the second case only the three elliptic radii are needed. These two

variants cover all cases from spherical or linear arrays (as commonly used for WFS, Ambisonics and VBAP) to more complex setups.

As the pointing method can be also used to move virtual sources in real time, localization experiments using a virtual source as acoustic target can be realized, too.

3.1. Computation of the pointed direction

This section explains the computation of the direction pointed to by the subject (see also figure 2). The complete computation is implemented in Pure Data [10], a real time open-source software, running on a personal computer. In the first stage, the data of the 6 DOF sensors is received. The data consists of the position $\mathbf{p} = [p_1, p_2, p_3]^T$ and the 3 orientation angles ψ , θ and ϕ of the toy-gun, and as additional parameters the position $\mathbf{h} = [h_1, h_2, h_3]^T$ and orientation of the subject's head (the orientation of the subject's head is only needed for monitoring but not for the computation of the pointed direction). The orientation angles of the toy-gun are converted into a normalized direction vector \mathbf{v} , disregarding its spin ψ . In order to reduce measurement noise (e.g. the shaking of the hand and arm of the subject), the vectors \mathbf{p} , \mathbf{h} , and \mathbf{v} are temporally smoothed by filtering (see section 5.1). From the filtered data, a line $L = \{\mathbf{p} + \lambda \mathbf{v} | \lambda \in \mathbb{R}\}$ is constructed. The piercing point $\mathbf{x} = [x_1, x_2, x_3]^T$ of this line and the surrounding hull is computed according to the hull model (see 3.1.1 and 3.1.2 for details). If the reference frame of the hull model and the tracking system are not identical, an offset can be considered in the computation.

As most spatial sound reproduction setups use their center as the reference point, the piercing point \mathbf{x} is converted to spherical coordinates, after setting up the origin to any user-defined reference point. If the piercing point is required to be head-related, the reference point is set to the current head position \mathbf{h} .

The pointed direction can either be streamed continuously or stored as a single value at the trigger moment. In order to avoid deviations of the pointed direction during trigger actuation (see section 5.2), the series of computed values is stored in a buffer. Reading out the buffer values that precede the trigger moment provides reasonable suppression of shaking.

3.1.1. Ellipsoid hull model

The 3 radii r_1, r_2 and r_3 define an ellipsoid that is applicable as a spatial model of the surrounding hull.

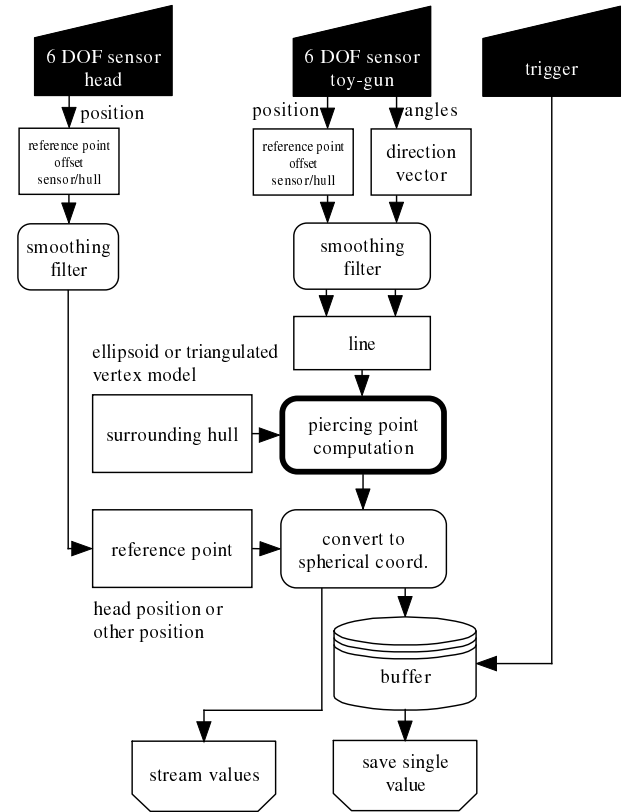


Fig. 2: Computation of the pointed direction.

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

$$\mathbf{x} = \mathbf{p} + \lambda \mathbf{v} \quad (2)$$

The piercing point \mathbf{x} has to lie on the hull of the ellipsoid defined by equation (1), while satisfying equation (2). This leads to a quadratic equation for λ , which is solved analytically.

3.1.2. Triangulated vertex hull model

The definition of the triangulated vertex hull model is more difficult. At first, the triangles between neighboring vertexes are determined. For this purpose, we use the convex hull algorithm in Matlab [11] (this algorithm is also available as open-source software [12]). This algorithm uses only those vertexes which build up a strictly convex hull. To outflank this possible restriction, all ver-

tex radii can be set to the same value for feeding the convex hull algorithm. With the resulting triangles around each vertex, the piercing point of the line L and the hull can be computed.

At first, the algorithm sorts the vertices according to their distances to L . It selects the nearest vertex \mathbf{a} and the first corresponding triangle consisting of \mathbf{a} , \mathbf{b} and \mathbf{c} . In the next step, the algorithm computes the piercing point of L and the plane spanned by \mathbf{a} , \mathbf{b} and \mathbf{c} . For this purpose the following linear system of equations is solved:

$$\mathbf{p} + \lambda \mathbf{v} = \mathbf{a} + \alpha(\mathbf{b} - \mathbf{a}) + \beta(\mathbf{c} - \mathbf{a}) \quad (3)$$

If a piercing point of the line and the triangle exists, i.e. the correct vertex and triangle has been chosen, the following inequalities for λ , α and β must hold:

$$\lambda \geq 0 \quad (4)$$

$$0 \leq \alpha, \beta \wedge \alpha + \beta \leq 1 \quad (5)$$

If (4) does not hold, the piercing point lies outside the indicated direction, and the next vertex \mathbf{a} will be tested. If the direction is correct, but (5) does not hold, the vertex \mathbf{a} can be correct, but the piercing point lies outside the observed triangle. In this case, the next associated triangle will be tested. If no valid triangles have been found, another vertex will be selected (see figure 3).

4. MEASUREMENT OF ACCURACY

This section presents the measured accuracy of the pointing method when aiming at optical targets in 2 different rooms using 2 different 6 DOF tracking systems. The triangulated vertex hull model is considered for the first room and for the second the ellipsoid model is applied. The computation of the piercing points is referred to the origin of those models. As tracking systems, we used an Ascension TrakStar (DC magnetic field, 240 frames per second (fps), 76 cm range) [13] and a VICON system equipped with 15 M-series cameras (infrared, 120 fps, covering most of the room) [14], which mark both edges of the price range. In order to collect comparable results, the toy-gun has been tracked by both tracking systems simultaneously.

According to manufacturer information of the TrakStar system, the accuracy is basically depending on the distance between transmitter and sensor. For the VICON system, a uniformly coverage and accuracy

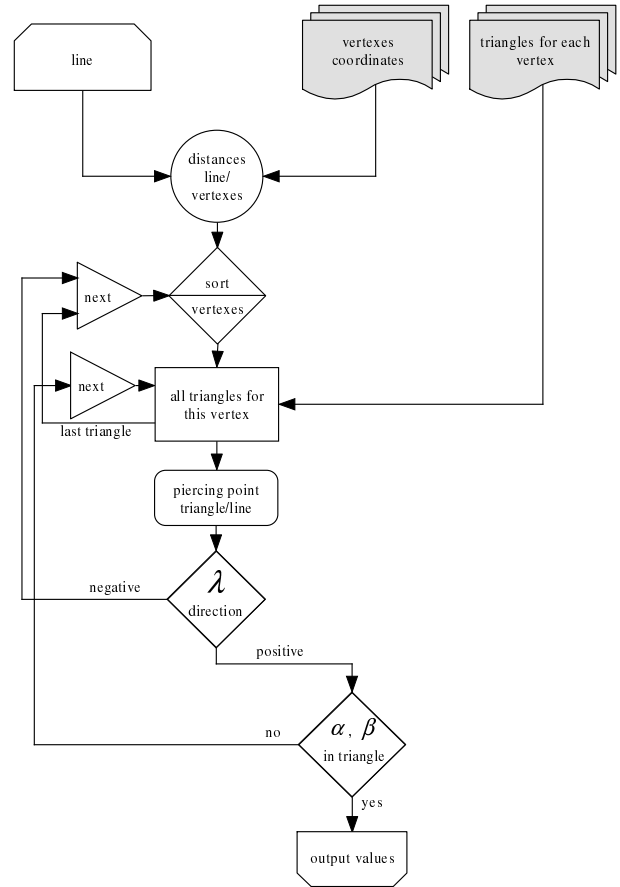


Fig. 3: Piercing point computation for vertex triangulated hull.

can be assumed for the considered measurement area. Therefore, the measurement positions can be limited to a part of this area. As the accuracy can be assumed to be independent of the toy-gun's orientation and the target position, the measurement conditions can be focused on a single optical target. In order to simplify the illustration, we concentrate on the 2D (only azimuthal angles) case. For the purpose of separating the influence of systematic errors (e.g. from tracking systems) from the errors done by the subjects (e.g. visual acuity and motor skill), the measurements are divided into two parts.

4.1. Static Accuracy (fix mounted toy-gun)

In order to measure the static accuracy, the toy-gun was mounted on a stand. The measurement positions are distributed from the origin of the rooms with a grid of

50 cm (see figures 4 and 7). From each position, the toy-gun is adjusted towards the optical target using the laser and iron sights. For each position, 10s of output data (shown angle) has been recorded. Note that no filtering has been applied to the tracking data, as the effect of filtering will be discussed later in section 5.1.

The results (figures 5, 6 and 8) show the absolute angle error, i.e. the deviation of the pointed direction from the direction of the optical target (0°). The mean values of the shown angles for each particular position are illustrated as big bars and the corresponding standard deviations as small bars. Due to illustration matters, the big bars show the absolute values of the mean showed angles. The colors indicate the sum of absolute mean values and the standard deviations. As the spatial resolution of human hearing is at best 1° [7], particular attention is paid to angle errors below that value. Note that for a better readability of the small values, the coordinate systems in figures 5, 6 and 8 are rotated by 135° compared to 4 and 7.

4.1.1. Static Accuracy: Room 1 (IEM CUBE): triangulated model, both tracking systems

Figure 4 shows the measurement setup. The measurement area consists of 16 positions in a $1.5\text{m} \times 1.5\text{m}$ square spreading south west from the center. To provide an idea of the room dimensions, the horizontal loudspeakers ring is plotted. The TrakStar's transmitter (depicted as a red square) is positioned $(0.29, -0.02)$ from the center.

Figure 5 indicates, that the VICON systems covers all measured positions uniformly and leads to angle errors clearly smaller than 1° . The TrakStar system shows an explicit dependence of the error from the distance between the measurement position and the position of the transmitter. Both the absolute mean values and the standard deviation rise when increasing the distance. Nonetheless, there are positions whose angle errors remain below 1° , even though the distance between transmitter and sensor is bigger than the manufacturer-specified range of 76 cm.

4.1.2. Static Accuracy: Room 2 (IEM production studio): ellipsoid model, TrakStar only

As the VICON system is permanently installed in room 1, the measurements in second room use only the TrakStar system. The measurement area consists of 9 positions in a $1\text{m} \times 1\text{m}$ square spreading south west from the center. In this room, a standard 5.1 surround setup

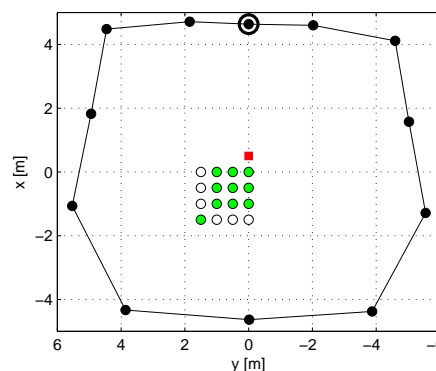


Fig. 4: Measurement setup in room 1: measurement positions (green), positions, which are not used for subjective tests (white), TrakStar transmitter (red square), lowest plane of vertexes/loudspeakers (black), optical target is encircled.

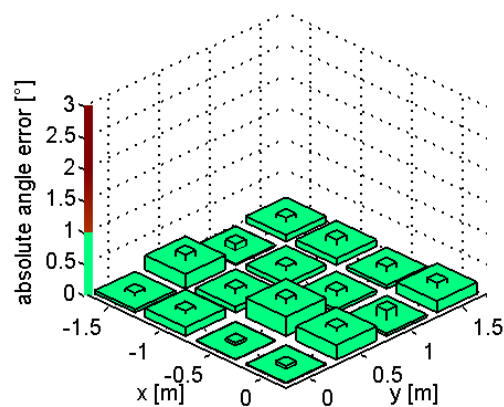


Fig. 5: Room 1, VICON (fixed toy-gun): absolute means (big bars) and standard deviations (small bars) of the angle error. Colors indicate the sum of both values.

(see figure 7) is available, so the ellipsoid hull model is applicable. The TrakStar's transmitter (depicted as a red square) is positioned $(0.50, -0.05)$ from the center.

There are only 3 positions where the angle errors are smaller than 1° . This is due to the bigger distance of the TrakStar transmitter to the central position. Another reason is the interference by the surrounding ferromagnetic metal (e.g. mixing console). For better results, this facts should be considered.

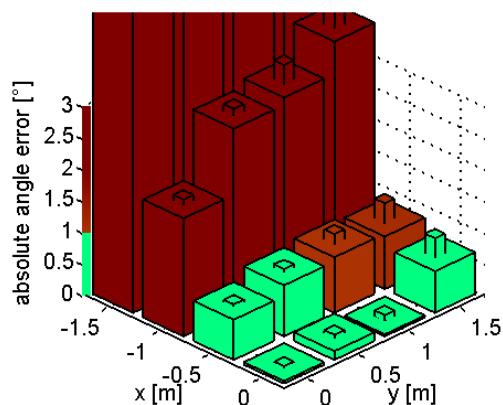


Fig. 6: Room 1, TrakStar (fixed toy-gun): absolute means (big bars) and standard deviations (small bars) of the angle error. Colors indicate the sum of both values.

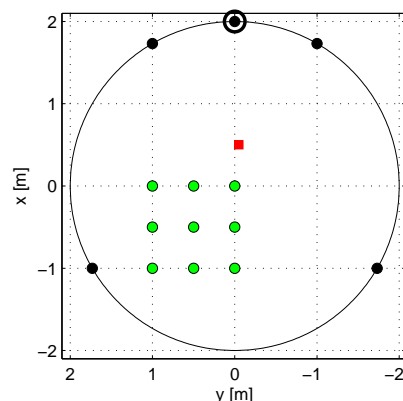


Fig. 7: Measurement setup in room 2: measurement positions (green), TrakStar transmitter (red square), loudspeakers of surround setup (black), optical target is encircled.

4.2. Subjective Accuracy (hand-held toy-gun)

In the intended application of listening tests, the toy-gun is hand-held by the subjects. In this case, the visual acuity and the motor skill of the subjects affect the accuracy the result. In order to reproduce this situation, the above made measurements have been repeated, but with 5 subjects holding the toy-gun in their hands. Again, 10s of output data (shown angle) has been recorded. As the results for the TrakStar in the static case showed much more than 1° of angle error for some positions, the number of distant positions has been reduced. In contrast to the previous measurements, all positions are now listening positions. The position of the toy-gun, which is important for the accuracy of the TrakStar system, can vary up to approx. 60 cm from these. The results are presented in the same form as for the static case, but the averaging is done both temporarily and over all subjects. The coordinate systems in figures 9, 10 and 11 are again rotated by 135° compared to 4 and 7.

4.2.1. Subjective Accuracy: Room 1 (IEM CUBE): triangulated model, both tracking systems

The same setup as for the static measurement has been used (see figure 4). The measurement area is reduced to a $1\text{m} \times 1\text{m}$ square spreading south west from the center. Additionally, the most distant position at $(-1.5, 1.5)$ is also explored. Those positions which are only used for the static but not for the subjective case are labeled white in the results (figures 9 and 10).

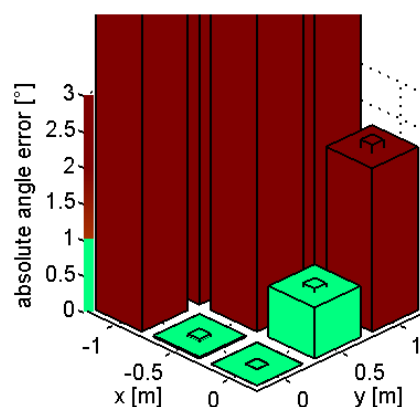


Fig. 8: Room 2, TrakStar (fixed toy-gun): absolute means (big bars) and standard deviations (small bars) of the angle error. Colors indicate the sum of both values.

The results of the VICON system show the same independence on the listening position as for the static case. All angle errors are well below 1° .

The dependence on the listening position for the TrakStar system is the same as for the static case. But as the subjects stretch out their arms into the direction of the transmitter, the distance to the sensor becomes smaller and so do the angle errors. Thus, there are more positions with angle errors below 1° than for the static case.

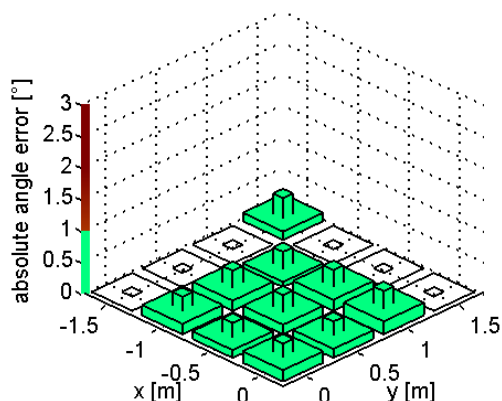


Fig. 9: Room 1, VICON (hand-held toy-gun): absolute means (big bars) and standard deviations (small bars) of the angle error. Colors indicate the sum of both values.

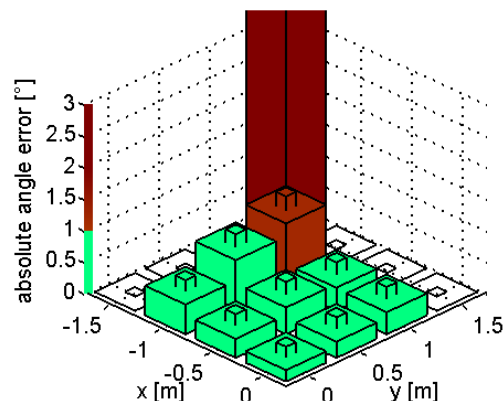


Fig. 10: Room 1, TrakStar (hand-held toy-gun): absolute means (big bars) and standard deviations (small bars) of the angle error. Colors indicate the sum of both values.

4.2.2. Subjective Accuracy: Room 2 (IEM production studio): ellipsoid model, TrakStar only

For room 2, the same setup is used as for the static case ($1\text{m} \times 1\text{m}$ square measurement area (9 positions) spreading south west from the center, see figure 7). The results do not differ much from the static case. The positions with angle errors below 1° are the same. At position $(-0.5, 0.5)$ the error is clearly decreased, but still a little bit above 1° .

Comparing the subjective case to the static case, the results are similar concerning the positions with angle errors below 1° . For the TrakStar system in room 1, the number of these positions increase in the subjective test. This is due to the decreased distance between transmitter and sensor, while stretching out ones arm into the direction of the target.

Furthermore, the standard deviation increases (compare figures 12 and 13) when the toy-gun is held in hands. This can be assigned to the visual acuity and the motor skill of the subjects.

The results show that the low-priced TrakStar system is comparable to the expensive VICON system in means of angle errors if the measurement is done inside the sensor range and if the system does not interfere with ferromagnetic material. The constraint of the sensor range could be avoided by mounting the transmitter under the subjects chair and thereby take it along, when moving to other listening positions. There is an extended range version of the TrakStar available which has approx. twice the price and triple the range.

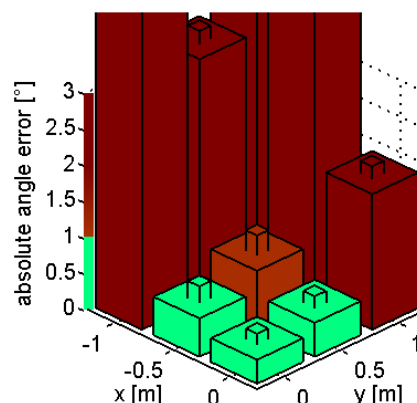


Fig. 11: Room 2, TrakStar (hand-held toy-gun): absolute means (big bars) and standard deviations (small bars) of the angle error. Colors indicate the sum of both values.

5. PRE- AND POST-PROCESSING

All results above are based on measurements without filtering of the tracking data at the input of the algorithm or postprocessing of the output data. This section discusses the effect of both for the improvement of the results.

5.1. Filtering the tracking data

As shown in figure 2, our algorithm provides a filtering of the tracking data at the input. For the following discussion on the effect of the applied median filters, the data of the measurements from room 1 with both tracking systems (for the TrakStar only the results inside the

range, i.e. for angle errors below 1°) are used. Note that for this study, the filters are applied to the output data, but simulations show that the order of piercing point computation and filtering has no effect.

All filter lengths are given in [ms]. Due to the different frame rates of the tracking systems, the filter length in ms leads to different filter lengths in frames (e.g. a length of 100 ms corresponds to 12 frames for the VICON and 24 frames for the TrakStar system). Note that the filters cause a delay of half the filter length added to the input data.

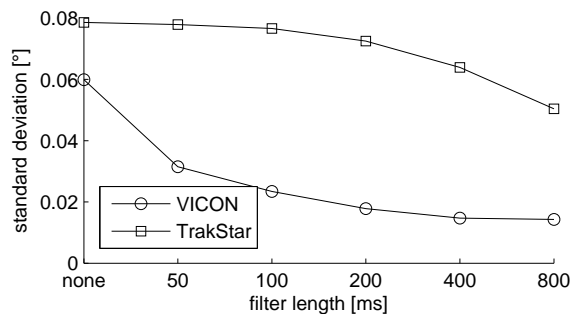


Fig. 12: Standard deviation of angle errors (fixed toy-gun) for different median filter lengths, averaged over positions.

Filtering of the input data with a length of 50 ms halves the standard deviation for the VICON system in the static case (see figure 12) compared to no filtering. Applying longer filters yield no reasonable improvement and increases the latency of the pointing method. The standard deviation of the TrakStar system is slightly higher without filtering, but filtering shows almost no effect. Obviously this system already provides filtered data.

If the toy-gun is held in hands, the standard deviation increases for both tracking systems (see figure 13). Again, the TrakStar system shows again little dependence on filtering. The impact on data of the VICON system is greater, but a length of at least 200 ms is needed to reduce the standard deviation at more than 10%. The VICON system seems to be more error-prone for dynamic tracking compared to the TrakStar system (inside its range) regarding the standard deviation, but still offers lower average angle errors (see. section 4.2.1).

5.2. Buffering the output

In a real listening test, subjects will have to pull the trigger of the toy-gun in order to save the showed angle. In

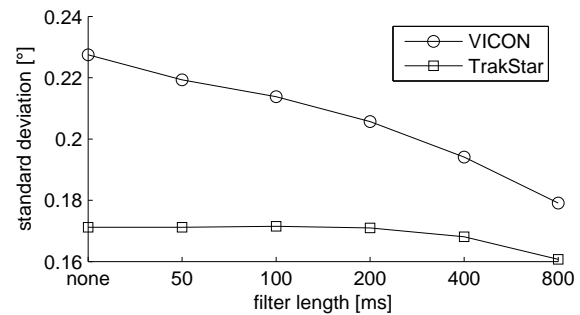


Fig. 13: Standard deviation of angle errors (hand-held toy-gun) for different median filter lengths, averaged over positions and subjects.

figure 14, a typical aiming process is shown (VICON system, room 1, no filtering). The dashed vertical line illustrates the trigger moment. Apparently, the trigger

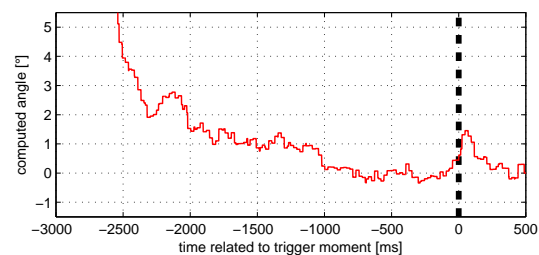


Fig. 14: Showed angles during aiming process after listening.

actuation moves the toy-gun a bit and distorts the measurement. The characteristics of the distortion and its direction varies from subject to subject and from position to position. However most subjects show a preferred direction. This direction seems to be independent on the fact, whether the subject is left-handed or right-handed. It is also visible that the exact angle is already shown approx. one second before the trigger is pulled. Even the fastest of the tested subjects, aims at least 200ms at the target before pulling the trigger. This fact can be used to compensate for shakes of the toy-gun in the trigger moment by buffering the output values (see section 3.1). The settings of this buffer depend on the latency of the tracking system, the applied filter at the input, and on the latency of the trigger.

6. CONCLUSION

This paper presented a flexible and intuitive pointing method for measuring the perceived direction in 3D localization experiments with a toy-gun. Using 2 different hull models to model the surrounding room or loudspeaker arrangement, this method can be used in nearly every surrounding hull/room and for every listening position. The accuracy of this method has been studied in 2 rooms with 2 tracking systems when aiming at an optical target. Even with the low priced system, the angle error remain below 1°, if the sensor range is kept. Furthermore, as the computation can be done in real time using open-source software on a personal computer, this method offers an affordable tool for localization experiments.

A previous version of this method has already been used in listening tests [9].

7. ACKNOWLEDGEMENTS

Thanks to Thomas Musil for his support with the implementation and to all subjects. This project is funded by the Austrian Research Promotion Agency (FFG), the Styrian Government and the Styrian Business Promotion Agency (SFG) under the COMET programme.

8. REFERENCES

- [1] F.L. Wightman und D.J. Kistler, "Monaural sound localization revisited," *J. Acoust. Soc. Am.*, 101(2):1050–1063, 1997.
- [2] R. Mason, N. Ford, F. Rumsey und A.B. de Bruyn, "Verbal and Nonverbal Elicitation Techniques in the Subjective Assessment of Spatial Sound Reproduction," *J. Audio Eng. Soc.*, 49(5):366–383, 2001.
- [3] V. Pulkki and M. Karjalainen, "Localization of Amplitude-Panned Virtual Sources I: Stereophonic Panning," *J. Audio Eng. Soc.*, 49(9):739–752, 2001
- [4] B. Seeber, "Untersuchung der auditiven Lokalisation mit einer Lichtzeigermethode", PhD Thesis, Technischen Universität München, Germany, 2003.
- [5] S. Spors, R. Schleicher, D. Jahn und R. Walter, "On the use of eye movements in acoustic source localization experiments", presented at the DAGA 2010, Berlin, Germany, 2010 March 15–18.
- [6] P. Majdak, B. Laback, M. Goupell, and M. Mi-hocic, "The Accuracy of Localizing Virtual Sound Sources: Effects of Pointing Method and Visual Environment", presented at the AES 124th Convention, Amsterdam, The Netherlands, 2008 May 17–20.
- [7] J. Blauert, "Räumliches Hören", Hirzel, 1974.
- [8] J. Lewald, "The effect of gaze eccentricity on perceived sound direction and its relation to visual localization", *Hearing Research*, 115:206–216, 1998.
- [9] M. Frank, F. Zotter and A. Sontacchi, "Localization Experiments Using Different 2D Ambisonics Decoders," presented at the 25th Tonmeistertagung, Leipzig, Germany, 2008
- [10] Pure Data, <http://puredata.info> .
- [11] The MathWorks, "Matlab", <http://mathworks.com>.
- [12] Qhull, <http://qhull.org>.
- [13] Ascension Technology Corporation, <http://ascension-tech.com>.
- [14] Vicon Motion Systems, <http://vicon.com>.