# GAZE AS A MEASURE OF SOUND SOURCE LOCALIZATION

**ROBERT SCHLEICHER, SASCHA SPORS, DIRK JAHN, AND ROBERT WALTER**

[1] *Deutsche Telekom Laboratories, TU Berlin, Berlin, Germany*
{robert.schleicher,sascha.spors}@tu-berlin.de

We present a study on utilizing eye movements for acoustic source localization tests. Test subjects had to indicate the presumed location of a hidden sound source with their head unconstrained by either fixating or additionally pointing with a laser pointer. Stimuli varied only in the horizontal plane from +45° (left) to -45° (right). Fixation error was always smaller than error in pointing and remained constant for all source positions, whereas pointing error showed a clear relation to source position with more eccentric positions leading to a higher error. Based on these results we conclude that gaze constitutes a useful measure for sound localization tests.

## INTRODUCTION

Current audio systems try to render the spatial audio scene as realistic as possible to increase user immersion. For this purpose it can be necessary to create *virtual* or *phantom* audio sources i.e. simulate a sound source at a position where actually no loudspeaker is located by manipulating properties of the signal that is emitted from the existing loudspeakers [1]. Various approaches have been proposed to achieve this goal, for example wave field synthesis [2]. A common way to assess the effectiveness of a certain spatial audio rendering technique is to conduct a listening test with human test subjects.

## 1 LOCALIZATION TESTS

In a localization test, test subjects are asked to specify the perceived origin of a sound that was played to them, either by selecting among a set of fixed source locations or by freely indicating the presumed direction using a graphical user interface (GUI), head movements, or a laser pointer [3-5]. In all these methods it is not clarified whether the translation of an auditory stimulus into a movement introduces an error in sound localization [6]. Although there exists rich literature on various aspects of eye movements and audiovisual integration [7, 8], gaze direction has been seldom used as dependent variable in localization tests. If authors consider this measure, they rather do it to let the subjects fixate a predefined target to distinguish between central and peripheral (with respect to vision) auditory localization performance [4, 6]. To our knowledge [9] is one of the first studies that examined the use of gaze as an indicator of source localization systematically. The main findings were substantial variations across subjects in localizing accuracy and large errors for localizing eccentric targets in the horizontal plane.

[9] used a scleral search coil to record eye movements. While this method allows very precise measurements, wearing a contact lens with a protruding thin wire imposes considerable discomfort on the subject. The necessary cornea anaesthetization requires the presence of a medical professional, thus making it less suitable for most audio research laboratories. However, the advantages of gaze as an indicator of sound source localization summarized by [9]], namely the high ecological validity and the negligibility of an additional training phase for the test subjects encouraged us to try an alternative, less obtrusive method to record eye movements.

Thus the objective of this study is to test the applicability of gaze as a measure of auditory localization performance and its precision compared to pointing with a laser pointer. This is realized using a combination of a head mounted eye tracker and a motion tracker for head tracking

## 2 METHOD

### 2.1 Experimental Setup

The experiment took place in a acoustically treated room of size 5.2*5.3*3 m. The reverberation time RT60 was lower than 0.45 s.

An array of seven loudspeakers was placed in front of the subject's chair at distance of r = 2 m. One loudspeaker was straight in front of the subject's face at eye level, the others were positioned on a circle centered around the subject's seat with an angular spacing of 15° in a range from +45° (left) to -45° (right). The actual position of the loudspeakers was concealed by a grey acoustically transparent curtain to the subjects as they otherwise would serve as a strong visual anchor (see Figure 1). On the curtain a continuous 1 cm broad white strip was attached at the height of 1.28 m from complete left to right. This strip served as a horizontal reference line both for fixations and laser pointing. Vertical lines and a intermittent dot were printed on the strip every 1

cm to facilitate constant fixation. The scenery visible to the subject (i.e. curtain with white line on it) was recorded by a camcorder on an elevated position behind the subjects. The general setup can be seen in figure 1.

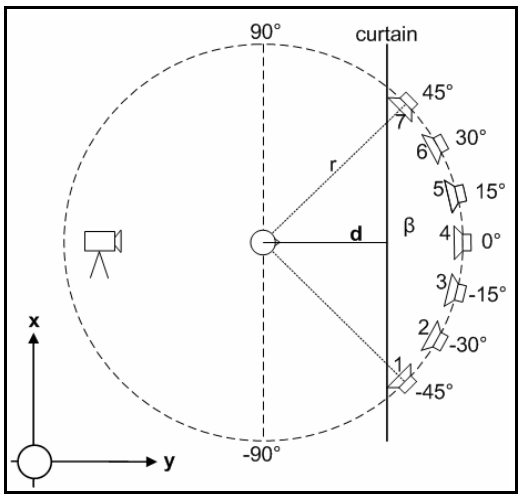


Figure 1: Experimental setup: 7 loudspeakers are centered around the subject's seat in steps of 15° from +45° left to -45° to the right in a radius of r = 2m. Their specific position is hidden by a grey acoustically transparent curtain which is d = 1.2m in front of the subject. A camcorder in the back of the subject served to record the whole scenery. The coordinate system in the lower left shows the coordinate system of the head tracker referred to in the text.

Binocular eye movements were recorded with a head mounted Eyelink II eye tracker (SR Research, Canada) with a sampling rate of 250 Hz. Head position was tracked with a Fastrak motion tracker (Polhemus, USA) at a sample rate of 120 Hz. Its magnetic transmitter was mounted above the participant chair and one sensor was attached to the eye tracker helmet. The stimulus material comprised of a sequence of four pink noise impulses of 1000 ms duration with 400 ms breaks played at a level of 60 dBA (SPL). For response a standard bimanual game pad with a laser pointer attached to it was used and subjects were asked to hold it with both hands to avoid any discrepancy due to handedness.

### 2.2 Procedure

The experiment consisted of two blocks: in the first block, subjects were asked to listen to the sound, look in its direction, and press a button on the game pad when they thought they were fixating its presumed origin on the white line ('gaze' block).

In the second block, subjects were asked to mark the presumed origin on the white strip with the laser pointer attached to the game pad and again press the button when they thought both their gaze and the laser pointer where targeting the sound origin ('point' block). Each block consisted of 10 stimuli from each loudspeaker played in a random order to the subject, leading to altogether 70 trials per block.

Upon arrival, subjects were seated and the chair was adjusted to align the subject's eye level with the horizontal white reference strip on the curtain. While they were allowed to move their head horizontally (i.e. left to right, the x-axis in figure 1), subjects were asked not to move it vertically (up and down, the z-axis that would be pointing into figure 1) or in the sagittal plane (back and forth, the y-axis in figure 1). To fulfill this they were requested to keep the back of the head in continuous contact with a small cushion. They were then acquainted with the stimulus material and the game pad. Participants could only listen and respond to a trial once. After subjects were familiar with the setup, the eye tracking helmet was fitted and the device calibrated. The same was done for the head tracker, and the experiment began. After completion, subjects were thanked for their participation and paid 15 Euros as a compensation. Altogether the experiment lasted around 45 minutes. Ten normal hearing subjects (six female) with a mean age of 25 +/- 3.8 years participated.

Only seven of these absolved the second block of the experiment ('point') as the experiment was announced to last not longer than 60 minutes which was would have otherwise been exceeded. For five of these seven subjects, additional eye movements while pointing were available. Due to the small sample size and to avoid confusion, these recordings were excluded from the statistical analysis that focus on the comparison *gaze* vs. *pointing*. We will only briefly mention the result of visual inspection of these *gaze while pointing* data in the discussion.

To summarize, there were ten data sets for the 'gaze' block available, and laser pointer data for seven of these subjects ('point' block) .

### 2.3 Data Processing

Head position was obtained by analysing the x and y coordinates of the motion tracking sensor on the eye tracking helmet. The z coordinates were disregarded as subjects had been prepared not to move their head up and down. Eye movement data were available as rotation angle of the eye relative to the head. As the stimuli varied only in the horizontal plane and the subjects were seated to be at eye level with the reference line, vertical eye movements were of lesser interest here. Testwise incorporating the vertical coordinates rather increased measurement error, probably due to the fact that their recording is more affected by lid movements. For this reason, only horizontal eye position values of both eyes averaged over the last 10 samples at the moment of button press were used to determine final gaze direction in combination with head rotation.

The laser pointer position was obtained from the video clips. A customized software developed in MATLAB was used to annotate the relative position of the laser dot on the white reference strip at the moment of button press[1]. The audio track of the film recording served to synchronise it with the other data sources. Unfortunately, the laser dot in the outermost positions of the white reference strip could not be identified as precisely due to reflections and overexposure. Therefore, the laser pointer data for the two the outermost source locations (+/- 45°) had to be excluded.

Localization error was defined as the difference between the true location and indicated location, where *indicated* means *looked at* for the gaze and *pointed at* for the laser pointer data. For both data types, trials with errors larger than Mean +/- 3*Stdev were marked as outliers. Altogether fourteen trials for the *gaze* and five trials for the *point* condition had to be excluded.

## 3 RESULTS

The following section examines the localization error with regard to the localization method and source position. It will focus on the comparison of *gaze* vs. *pointing*. All statistical analyses reported were done using the 'mixed model' function of SPSS. Amongst other things, mixed models have the advantage that they can also handle data sets with missing entries in a repeated-measurement design, which is not possible in classical analysis of variance (ANOVA), and at the same time report similar statistics (i.e. F-value and significance level). For a detailed description of mixed linear models or multilevel modeling approaches see [10].

Figure 2 shows that the mean error strongly increases when comparing pointing with fixation data. A mixed models analysis of variance with *method* (*gaze* vs. *pointing*) as fixed factor and *trial number* as a covariate to account for training- or time-on-task effects yielded a significant effect of *method* ($F_{1,316}$=56.258; p<0.000), but none for *trial number* ($F_{1,335}$=1.258; p=0.263).

By taking a closer look at the distribution of errors across the different sound sources (see figure 3), it can clearly be seen that subjects tend to overshoot for lateral targets when using a pointer, i.e. locate the sound source more eccentrically than it actually is (negative error for sources on the left, i.e. positive azimuth and positive error for sources on the right, i.e. negative azimuth).

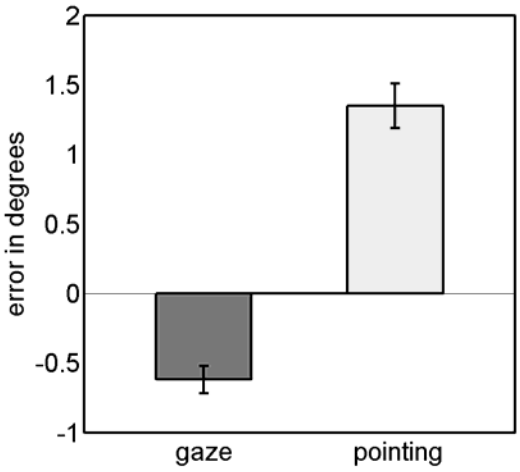


Figure 2: mean error over all subjects for both types of localization data obtained in the experiment: fixation and laser pointer data. Whiskers denote standard errors.

The mean error of eye movements in the *gaze* condition however remains more or less constant for all source locations. Computing an additional analysis for the error in pointing and the error in the gaze with regard to source location reveal that there is a strong effect of source location on pointing error ($F_{6,827}$=27.990; p<0.000), and a significant interaction of *method*source location* ($F_{4,835}$=37.982; p<0.000).

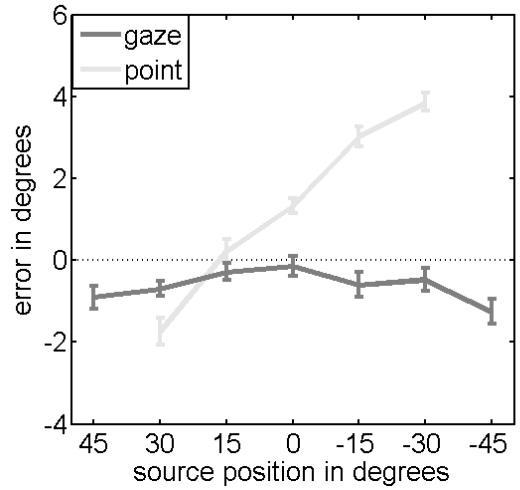


Figure 3: mean error in relation to source position (+45°=left, -45= right) across all subjects for both types of localization data obtained in the experiment: fixation only in block 1 and laser pointer data both from block 2. Whiskers denote standard errors. Please note that for positive azimuth degree source positions (to the left) negative errors mean overshoot as do positive errors for negative azimuth degree sources (to the right).

[1] The annotation tool for Matlab can be downloaded at: http://www.mathworks.com/matlabcentral/fileexchange/25950-mmplayer

## 4 DISCUSSION

Using a head mounted eye tracker in combination with head tracking we were able to utilize eye movements as an indicator of horizontal auditory source localization in subjects whose head was unconstrained. The findings are quite promising and also outperform the most common approach for localization tests of pointing to the presumed source. Compared with eye movements, pointing leads to an increased error for more eccentric targets which has also been reported by other authors [5]. However, the specific type of error, namely overshooting, has only been reported by Lewald et al. [2000] for trials where the head was fixated.

Looking appears to be more accurate even while pointing as visual inspection of the available eye movement recordings during the pointing block revealed. Although these data have to be interpreted with caution as they only are based on 5 subjects and were thus not analyzed statistically, it appears that people first localize the source with their eyes and then try to lead the pointer to that area of fixation, but are less successful in hitting its center. We will discuss possible reasons for that further below, but would like to point the readers to potential shortcomings first.

This study has several limitations, the most obvious one being the restriction to stimuli in the horizontal plane from +45° to -45°. We did this mainly for two reasons: to keep the setup as simple as possible and to obtain data of high quality under optimal conditions for the beginning. Moving the head to more eccentric sources as well up and down would clearly increase measurement error for eye movement data. In addition, localizing horizontal sources is achieved in the auditory system by exploiting interaural time differences (ITD) and interaural level differences (ILD), whereas the localization of vertical sources is based on additional pinna and head shadowing effects and is less precise [1]. The data of Populin's [2008] subjects for example show considerable variation.

The second shortcoming is the analysis of laser pointer data, which was restricted to pointing performance in the range from +30-30° as more eccentric points were not clearly visible in the video files. If the larger error observed in pointing was only due to annotation imprecision, it should be distributed in a more random fashion and not show such a systematic tendency to increase with stimulus eccentricity. Lewald et al. [2000] used a laser LED attached to the subject's head or a swivel pointer with a potentiometer to let the subjects indicate the perceived source location, which may provide more accurate pointing data, but might also require a training phase for some subjects. In that regard, standard laser pointers are surely the more intuitive pointing device. We are currently examining the possibility to synchronize the eye and motion tracking using the open source software LibGaze[2], which might enable us to also record the laser pointer position with an additional sensor of the motion tracking system and thus obtain higher precision for the pointing data. Whether it then will reach the precision of a high quality eye tracker is still in question. In addition, we believe that the higher error for pointing is more or less immanent to the human motor system – at least for normal test subjects, the eyes are much better trained to perform very small and accurate movements at a high speed (e.g. reading) than the hands. The ability to quickly localize the source of an unknown sound and check whether it is a predator or prey has developed over a long time in the course of evolution [11], whereas precise pointing was much less important.

While monitoring data recording, we repeatedly observed that subjects were immediately looking in the presumed direction, focussing a certain area and then trying to match the laser dot onto that fixation area. Here, pointing would add an additional error to visual localization which apparently increases with eccentricity. The fact that gaze data were the average of two eyes might have added to its precision in our case. If that assumption holds true, binocular eye movement can be a very promising candidate for localization tests in addition to the various pointing devices mentioned at the beginning of this article. This is especially true if the task would be extended to localizing moving sources, where also the latency of localizing would become more important.


## ACKNOWLEDGEMENTS

We thank Dr. Shiva Sundaram for his useful suggestions on an earlier version of this paper.


## REFERENCES


[1] S. Sundaram and C. Kyriakakis, *Phantom audio sources with vertically seperated speakers*. in *119th AES Convention*. 2005. NYC, NY October 7–10: Audio Engineering Society (AES).

[2] S. Spors, R. Rabenstein, and J. Ahrens, *The theory of wave field synthesis revisited*. in *124th AES Convention*. 2008. Amsterdam, The Netherlands, May 2008: Audio Engineering Society (AES).

[3] J. Lewald and W.H. Ehrenstein, *Auditory-visual spatial integration: A new psychphysical approach using laser pointing to acoustic targets.* Journal of the Acoustical Society of America, 1998. 104(3): p. 1586-1597.


[2] http://sourceforge.net/projects/libgaze/

[4] B. Razavi, W.E. O'Neill, and G.D. Paige, *Auditory spatial perception dynamically realigns with changing eye position.* The Journal of neuroscience : the official journal of the Society for Neuroscience, 2007. **27**(38): p. 10249-58.

[5] J.C. Makous and J.C. Middlebrooks, *Two-dimensional sound localization by human listeners.* Journal of the Acoustical Society of America, 1990. **87**(5): p. 2188-2200.

[6] J. Lewald, G.J. Dorrscheidt, and W.H. Ehrenstein, *Sound localization with eccentric head position.* Behavioural Brain Research, 2000. **108**(2): p. 105-25.

[7] T.J. Van Grootel and A.J. Van Opstal, *Human sound-localization behaviour after multiple changes in eye position.* The European journal of neuroscience, 2009. **29**(11): p. 2233-46.

[8] D. Zambarbieri, G. Beltrami, and M. Version, *Saccade Latency Toward Auditory Targets Depends on the Relative Position of the Sound Source with Respect to the Eyes.* Vision Research, 1995. **35**(23/24): p. 3305-3312.

[9] L.C. Populin, *Human sound localization: measurements in untrained, head-unrestrained subjects using gaze as a pointer.* Experimental Brain Research, 2008. **190**(1): p. 11-30.

[10] H. Quené and H. van den Bergh, *On multi-level modeling of data from repeated measures designs: a tutorial.* Speech Communication, 2004. **43**: p. 103-121.

[11] A. Öhman, A. Flykt, and F. Esteves, *Emotion drives attention: Detecting the snake in the grass.* Journal of Experimental Psychology: General, 2001. **130**: p. 466–478.