EVALUATION OF A 3D VISUALIZATION BASED MEASUREMENT TECHNIQUE FOR SOURCE LOCALIZATION

Shuai Wang and Deep Sen*†*

*School of Electrical Engineering and Telecommunications
University of New South Wales, Australia
†Qualcomm Inc., Audio Research & Development, San Diego, CA, USA

ABSTRACT

Directional listening tests are an integral part of overall psychoacoustic evaluation of spatial rendering systems as well as general audiological evaluation. A new test paradigm is proposed, where, to reduce the effect of mapping bias, a 3D perspective plot of the visual scene is provided to the listener. Both visual and acoustic stimuli are used to evaluate the proposed technique against an approach where only 2D exocentric scenes are provided to the listener. Results show improved performance in some cases whereas further investigation is warranted for other cases.

***Index Terms*—** Directional listening test, Source localization, Mapping bias, Egocentric, Exocentric.

1. INTRODUCTION

Source localization has been one of the most frequently tested spatial attributes during subjective evaluation of auditory material. It characterizes the process in which listeners identify the perceived direction and distance of a sound source. This is usually subjectively measured to assess the performance of sound/soundfield rendering systems, including audio codec and home entertainment systems. Audiologists also test for source localization when studying subjects' psychological and physiological responses to acoustic events. To measure source localization, directional listening tests are usually deployed. In such tests, stimuli that contain one or more sound objects may be presented either binaurally or using multichannel speaker systems. Listeners are instructed to identify and localize individual sources, and then report the perceived location and/or direction. The means by which they provide such feedback often influence the accuracy and also introduce bias into the results.

A number of directional listening test paradigms have emerged in recent years. They may be categorized into two classes: verbal and non-verbal. The former involves verbal descriptors as feedback for source direction [1]. The non-verbal paradigms, on the other hand, require additional apparatus, such as rulers [2] and 2D auditory scene based diagrams [3, 4], to avoid communication errors and ambiguity in verbal languages. However, existing methods, both verbal and non-verbal, are prone to the inevitable bias in conveying auditory perception which involves expressing an internal 3D perceptual impression in an external 2D report [5]. This paper will thus propose a new 3D visualization based directional listening experiment paradigm, attempting to reduce such bias, and evaluate it against another technique designed from existing 2D auditory scene based methods.

The following sections begin with a brief review of existing source localization measurement techniques. It will be followed by details of the proposed experiment paradigm. Pilot listening tests and their results will be presented in the next section, where results will also be analyzed for comparison. Finally, a discussion is provided on the performance of the proposed approach.

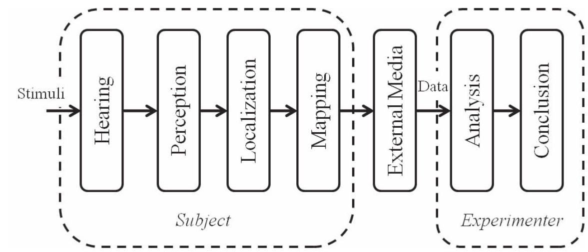


Fig. 1. Generic Structure of Directional Listening Tests [6].

2. SOURCE LOCALIZATION MEASUREMENTS

In order to understand the issues in existing source localization measurement techniques, we need to first look into the directional listening test from a system perspective. Fig. 1 shows the processes involved in a typical directional listening test in block diagrams. As illustrated in this figure, sound captured by the peripheral auditory physiology are represented as physical signals that ultimately stimulate the listener's neurons in the upper auditory pathway. These physiological signals are converted into fundamental internal psychophysical features, such as loudness, pitch, timbre and etc. Based on a subset of these features, subjects form an opinion on which direction the sound originates or how far away the source is located. Such an internal opinion needs to be translated into an external representation that can be recognized and recorded by experimenters for further analysis. The Mapping block in Fig. 1 represents such a complicated psychological process. Since discrepancy often exists between subjects' internal psychological judgement and the experimenter-determined external representation format, the mapping process is most likely to cause undesired bias [5, 6] and inaccuracies. Therefore, the primary focus of designing directional listening tests is to ensure a correct and effective delivery of internal judgements from subjects to experimenters.

Verbal descriptions, as the obvious choice of conveying information, have been used in several directional listening experiments. For example, in Wightman and Kistler's work [1], listeners were asked to locate sound sources in reference to a spherical coordinate system and describe the perceived direction using absolute azimuth and elevation angles in degrees. The measurement resolution may be as high as 1 degree, but it requires a strong familiarity to the concept of azimuth and elevation from listeners. Similarly, Evans at. el. [7] divided the 2D horizontal plane using 12 directions and named them in a clock pattern. Test records show that while this approach is more user-friendly, the resolution drops significantly to 30^o.

In addition to verbal descriptors, several pointing equipment, such as hands, sticks or even more precise head-tracker and hand-held laser pointer, have been used in listening tests to help subjects report their judgement on source direction and/or distance [2, 8]. Generally speaking, these pointing techniques are more intuitive compared with the verbal only approaches. However, they consequently increase the difficulty in data collection and recording for further analysis. Further, head movement is often restricted for certain listening tests, especially those involving loudspeaker systems. This certainly limits the application of head-trackers.

An alternative group of non-verbal techniques are 2D visualization based. In such tests, subjects are required to report their internal impression of sound locations in an external pictorial representation. In other words, the perceived location needs to be mapped from the internal 3D psycho-physical domain to a 2D plane which is usually the horizontal plane in most existing tests. Such scene-based techniques have been applied in [3] where subjects hand-sketch the perceived location on paper, as well as in [4] where computers are used. Compared with other techniques, these visualization-based approaches allow evaluation of source width and depth as well as providing subjects with a certain degree of freedom regarding the spread/shape of sound sources. Besides, computer-aided techniques can collect and store subject data instantaneously which increases the testing efficiency. However, current 2D visualization-based techniques suffer significantly from the process of mapping. Since, during localization (cf. Fig. 1), subjects orientate the perceived sources and surrounding space in reference to themselves (which is termed egocentric [9]), sources are often "misplaced" when mapped from the egocentric internal space into the exocentric[1] external plane. Further, the internal psycho-physical space is three-dimensional while the horizontal plane is only two-dimensional. Therefore, height/elevational information is lost during the translation of perceived source location into a horizontal representation, and the recorded source distance is often the 2D projection of the target distance. Last but not the least, these visualization-based techniques often lack a reference or a scale. As a result, distance evaluation is merely an approximation, and measurement results may not be comparable across different subjects. To alleviate the reference problem, visual markers have been deployed in several experiments [10, 11]. Listeners can certainly benefit from these markers in space and map perceived sound sources into the same external pictorial representations. However, it is also of concern that the measured source location may be biased towards the markers due to the so-called "re-mapping effect" [4, 8].

3. 3D VISUALIZATION BASED EVALUATION PARADIGM

Since source localization measurements can be significantly affected by mapping errors, the primary task of our new evaluation technique is to reduce the amount of transformation involved in the process of converting an internal egocentric impression into an external representation. An obvious solution to this issue is to present subjects with an external egocentric format that is similar to their internal psycho-physical space. Therefore, a perspective view of the 3D external space is provided to subjects during evaluation. Such a perspective space is centered at the listening point which is similar to what is perceived internally by a listener. Figure 2(b) shows the Graphical User Interface (GUI) for this method, which includes a perspective plot of the 3D surrounding space which we call *egocentric* view or *perspective* view. The following paragraphs will discuss details of this GUI and how it is used by subjects during evaluation.

Content Identification and Localizability (Fig. 2(a)). After entering few non-identifying personal details, listeners are directed to play the auditory stimuli and identify individual sources by selecting from a list of options. Subjects also need to indicate the overall localizability of the sound objects on a continuous scale, which measures the difficulty in locating the object. The localizability scale only appears on the GUI if the listener indicates they can perceive the object. The number of options listed is thus greater than the actual sound present in the stimuli. Listeners may play the stimuli multiple times to familiarize themselves with the auditory objects in stimuli.

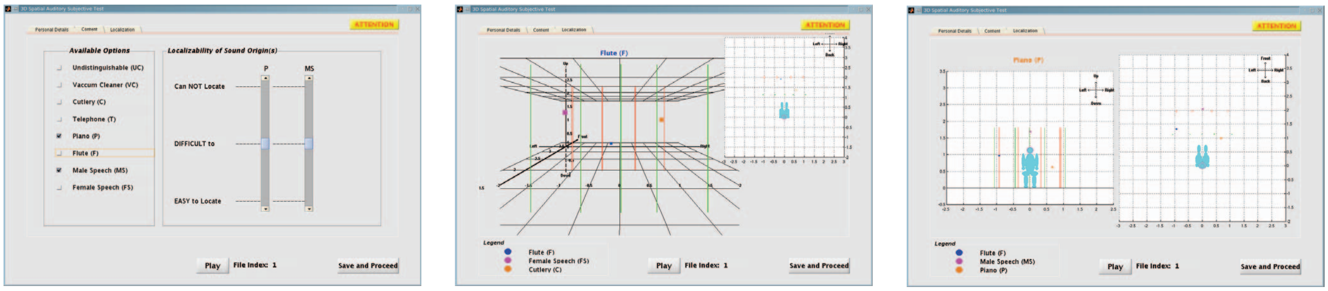
Localization (Fig. 2(b)). This part of the GUI presents listener with two plots. One is a 3D perspective view of the surrounding space and the other is a 2D orthographic top view of the listening area. During experiments, subjects are asked to indicate the location of perceived sound source by positioning the corresponding marker in the 3D perspective view. Control of markers in the 2D plot is disabled which means marker movements can not be initiated from this view. Figure 2(b) shows an example where three sources have been identified earlier on the previous "Content/Localizability" tab. Those markers are color coded with legends displayed at the bottom left corner of the GUI, and they can be positioned anywhere in the 3D perspective space using a mouse with a center wheel. Subjects may press and hold the left button of the mouse over a marker and drag it across the x-z plane, while the front-back (y axis) dynamics is controlled by the center wheel. As subjects move the marker around in 3D perspective view, the 2D top view plot will be simultaneously synchronized to provide extra guidance in y-direction.

There are a few things worth noting for this 3D method. Firstly, visual references are used in the test. As shown in the GUI- Fig. 2(b), five green and four red perspex rods are lit and placed in the testing area. These visual cues not only provide subjects with extra reference which is similar to the visual marker approach reviewed early, but also bear the benefit of enhancing the perspective view in that 3D plot. Secondly, the chair used in this technique is custom made with an adjustable seat and a fixed chin-holder. The chin-holder prevents subjects from moving their heads, which is very important for certain localization experiments. At the same time, the adjustable seat ensures the chair fit various subjects with different heights. In addition, this GUI is efficient, user friendly and runs completely on MATLAB. Automatic collection and recording of evaluation results is enabled once the "Save and Proceed"(cf. Fig. 2) button is pressed, and a "Play" button is also included for repeatable playback allowing subjects to control the pace of the evaluation.

4. PILOT TEST AND COMPARISON

Pilot tests were conducted using the proposed egocentric 3D Visualization Based Evaluation Method. Results were compared against those obtained from an exocentric approach (Fig. 2(c))- a method we designed from the traditional 2D visualization based approach which employs only 2D orthographic diagrams. More details and analysis will be provided in the following subsections.

[1] In this paper, exocentric is used to describe an orthographic space, as oppose to egocentric for perspective space. Despite the fact that polar coordinate systems are often used in 2D visualization-based tests, they still bear orthogonal properties and they can not be taken as an egocentric plane.



(a) Content Page (b) Localization (ego/3D) (c) Localization (exo/2D)

Fig. 2. 3D and 2D Graphical User Interface used in pilot tests.

4.1. *Methodology*

The 3D egocentric pilot tests took place in a dark anechoic room of size 5.44 x 3.66 x 2.54 m, where a total of 25 Genelec 8030A and 8130A loudspeakers were situated. A 32-channel RME-M-32-DA device was used to drive the loudspeakers. The custom-made chair was placed at the center of the loudspeaker array with its chin-holder fitted and seat adjusted for every listener. In order to conceal these loudspeakers from test subjects, dark acoustically transparent curtains were placed around the chair. As shown in Fig. 2(b), 9 illuminated perspex rods were set up inside the room as reference, with a row of 5 green ones positioned closer to the chair. A workstation that supports Matlab was used to host the GUI, and it was placed outside the room for remote control and elimination of potential noise source. Inside the dark room, subjects were provided with a LCD monitor and a mouse to interact with the GUIs.

Ten engineering students were recruited for the pilot tests after passing an audiological pre-screening process. None of them had previous experience with either egocentric or exocentric source localization experiments which may reduce the potential bias caused by their familiarity to either interfaces. Training was conducted first to ensure subjects know how to use the GUIs properly. Since visual perception is more reliable than acoustic perception in localization [12], visual stimuli were deployed to facilitate the primary comparison. Subjects were tasked with having to localize visual objects. For this, three different sized boxes were placed in the room at different positions. Each box was lit up by a LED at its center. Subjects were instructed to use the egocentric GUI (described above) to indicate the positions of the boxes. Each subject repeated the evaluation three times to complete the visual session.

For audio testing, five sets of stimuli comprising of mono anechoic recordings of cutlery noise, male speech, female speech, flute, and broadband Gaussian white noise were used. To simplify the pilot test, each stimuli contained only one acoustic object. The loudspeaker used for playback was varied across all materials. Subjects were asked to locate each sound source and indicate it on the GUI. The test was double-blind with the 5-stimuli playlist automatically randomizing the order in which the playback occurred. The loudness level of the playback was normalized to be consistent across all subjects.

For comparison, the same set of tests, both visual and audio, were also conducted using the exocentric scene based method. The GUI used for this method is shown in Fig. 2(c). It includes two 2D plots- a back view and a top view. Different from the new 3D approach, both of these plots are exocentric/orthographic. Using this GUI, subjects have to click and drag markers in both plots to completely localize the perceived sound. However, the left-right (x-axis) dynamics in the top view are disabled, which means the x-axis localization is only completed in the back view diagram. This increases the efficiency of using this GUI as both 2D orthographic diagrams are simultaneously synchronized along x-axis just like in the egocentric approach. Apart from the GUI, the other experimental conditions are identical to the proposed egocentric method.

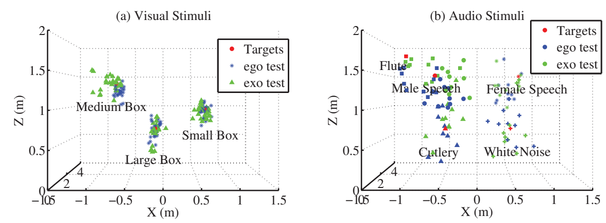


Fig. 3. Pilot Test Results: red, blue and green markers show target, egocentric and exocentric results respectively

4.2. *Results and analysis*

Pilot test results are shown in Fig. 3 - 5. Figure. 3 plots the target locations (in red) for both visual and audio tests. Measurement results from egocentric (ego) tests are marked in blue, while the green indicates results from exocentric (exo) experiments. As expected, compared to the visual results, the audio results show an increased spatial distribution in both ego and exo tests. This observation supports the use of visual stimuli as the primary excitation for comparison of the two methods. It also suggests employing visual stimuli for trainings in directional listening tests, which allows analysis of how much of the error can be attributed to the GUI/interface and how much to fundamental auditory acuity or rendering inaccuracies.

More detailed results of the visual experiments are shown in Fig. 4. Fig. 4(a), (b) and (c) are the respective Mean Absolute Error (MAE) plots of measured azimuth and elevation angles for small, medium and large boxes, while the distance measurements MAE for all three boxes are collectively drawn in Fig. 4(d). As illustrated in these figures, most subjects have produced comparable localization results using the two methods. Resulting MAEs in ego tests appear small and close to that of the exo tests for estimations of azimuth and distance. In particular, better performance has even been recorded consistently across all subjects for azimuth measurement of the small box, which is positioned close to eye level, with the egocentric test (Fig 4(a)). However, such cross-subject improvement is

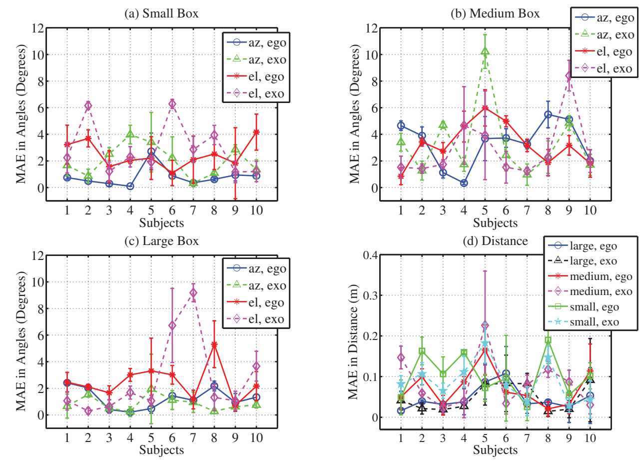


Fig. 4. Visual test results: (a), (b) and (c) are the respective MAE plots of measured azimuth and elevation angles for small, medium and large boxes, and (d) plots MAEs of distance.

not demonstrated in localization of the other boxes. This may be due to variation in subjects' preference of GUI. Post-experiment interviews suggest, although all from the engineering faculty, some subjects prefer to use the GUI designed for egocentric test as it feels more intuitive and realistic, whereas others prefer the orthographic view because the perspective view does not seem identical to their perception of the 3D space and they feel comfortable mapping egocentric observation to exocentric graphs. Thus, the egocentric GUI may need modification to better replicate the frontal 3D space for improvement. At the same time, subjects score slightly higher MAE when assessing elevations of the three boxes (Fig. 4(a), (b) and (c)). This may be because the colored rods and subject's eye level are the only elevational reference provided during localization. In this case, subjects are prone to misplace visual objects on GUIs for those away from the eye level. Such misplacement could also have resulted in the cross-subject inconsistency in performance of egocentric measurements of azimuth as well. As indicated by the result, more elevational reference may be deployed to improve the proposed egocentric test, and further testing with visual stimuli is warranted.

Similarly for the audio tests, comparable results have been achieved between the two GUIs (Fig 5). As shown in Fig 5(a) and (b), no consistent trend was observed across ten subjects in angular and distance measurements. However, MAE values from both methods are very close to each other. Comparing the results of the audio tests with visual tests, the MAE calculated from the former is much higher than the latter. This may be attributed to both the difference in visual and acoustic acuity as well as the fact that subjects are far more trained in visual localization in everyday life. Therefore, subject training is essential for future formal listening experiments.

5. CONCLUSION

A brief review of existing techniques for auditory localization testing revealed that mapping bias is one of the most important issues in subjective testing. Attempting to reduce such bias, a new 3D egocentric visualization based design was proposed in this paper. Pilot tests were conducted with both visual and auditory stimuli to evaluate the proposed egocentric method against an exocentric method originated from the traditional 2D visualization based approach. Testing with visual stimuli served as both a novel way of familiarizing participants with GUI and the primary excitation for comparison of these two methods. Results show better performance when using the proposed 3D-egocentric paradigm for azimuth measurement of the visual stimuli close to eye level. However, this observation is not consistent across all cases. The inconsistency may be attributed to variation of individual subject's preference of GUI, and the lack of elevational reference in experimental set-up. Overall, the new 3D egocentric technique is a good alternative to existing 2D visualization based approach. Further investigation is warranted with improved user interface and employment of additional vertical reference. At the same time, the correlation between subjects' preference of GUI and the performance of testing methods is worth studying as well.

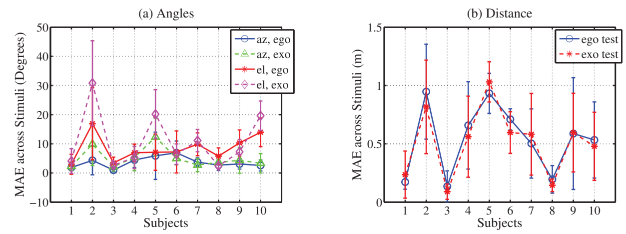


Fig. 5. Audio Test Results:(a) is the plot of angular MAE across stimuli vs subjects; (b) is the plot of MAE in distance across stimuli vs subjects.

6. REFERENCES

[1] D. J. Kistler and F. L. Wightman, "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," *The Journal of the Acoustical Society of America*, vol. 91, no. 3, pp. 1637–1647, 1992.

[2] R. H. Gilkey, T. R. Anderson, and D. H. Mershon, *Phenomenal geometry and the measurement of perceived auditory distance*, pp. 257–274, Erlbaum, 1997.

[3] N. Ford, F. Rumsey, and T. Nind, "Creating a universal graphical assessment language for describing and evaluating spatial attributes of reproduced audio events," in *Audio Engineering Society Convention 115*, 10 2003.

[4] J. Usher, *Subjective evaluation and electroacoustic theoretical validation of a new approach to audio upmixing*, Ph.D. thesis, Schulich School of Music, McGill University, Montreal, QC, Canada, 2006.

[5] R. Mason, N. Ford, F. Rumsey, and B. de Bruyn, "Verbal and nonverbal elicitation techniques in the subjective assessment of spatial sound reproduction," *J. Audio Eng. Soc*, vol. 49, no. 5, pp. 366–384, 2001.

[6] S. Zielinski, F. Rumsey, and S. Bech, "On some biases encountered in modern audio quality listening tests-a review," *J. Audio Eng. Soc*, vol. 56, no. 6, pp. 427–451, 2008.

[7] M. J. Evans, A. I. Tew, and J. A. S. Angus, "Relative spatialization of ambisonic and transaural speech," in *Audio Engineering Society Convention 104*, 5 1998.

[8] B. G. Shinn-Cunningham, N. I. Durlach, and R. M. Held, "Adapting to supernormal auditory localization cues. i. bias and resolution," *The Journal of the Acoustical Society of America*, vol. 103, no. 6, pp. 3656–3666, 1998.

[9] R. A. Hart and G. T. Moore, *The development of spatial cognition: a review*, pp. 226–234, University of Chicago Press, 1973.

[10] P. A. Nelson, O. Kirkeby, T. Takeuchi, and H. Hamada, "Sound fields for the production of virtual acoustic images," *Journal of Sound and Vibration*, vol. 204, no. 2, pp. 386 – 396, 1997.

[11] H. Mller, M. F. Srensen, C. B. Jensen, and D. Hammershi, "Binaural technique: Do we need individual recordings?," *J. Audio Eng. Soc*, vol. 44, no. 6, pp. 451–469, 1996.

[12] D.H. Mershon, D.H. Desaulniers, T.L. Amerson, and S.A. Kiefer, "Visual capture in auditory distance perception: proximity image effect reconsidered.," *J Aud Res*, vol. 20, no. 2, pp. 129–36, 1980.