NAVIGATION PERFORMANCE EFFECTS OF RENDER METHOD AND LATENCY IN MOBILE AUDIO AUGMENTED REALITY

Nicholas Mariette

LIMSI-CNRS
Audio and Acoustics Group
Orsay, France
nicholas.mariette@limsi.fr

ABSTRACT

This paper describes a pilot study and main experiment that assess user performance at navigating to spatialised sound sources using a mobile audio augmented reality system. Experiments use a novel outdoor paradigm with an application-relevant navigation task to compare perception of two binaural rendering methods under several head-turn latencies. Binaural rendering methods examined were virtual, 6-speaker, first-order Ambisonic and virtual 12-speaker VBAP techniques. This study extends existing indoors research on the effects of head-turn latency for seated listeners.

The pilot study examined the effect of capture radius (of 1, 2, 3, 4 and 5 metres) on mean distance efficiency for a single user's navigation path to sound sources. A significant performance degradation was found to occur for radii of 2 m. The main experiment examined the effect of render method and total system latency to head-turns (176 ms minimum plus 0, 100, 200, 400 and 800 ms) on mean distance efficiency and subjective stability rating (on a scale of 1-5), for 8 participants. Render method significantly affected distance efficiency and 800 ms of added head-turn latency significantly affected subjective stability. Also, the study revealed a significant interaction effect of render method and head-turn latency: Ambisonic rendering didn't significantly affect subjective stability due to added head-turn latency, while VBAP rendering did. Thus, it appears rendering method can mitigate or potentiate stability effects of head-turn latency. The study also exemplifies that the novel experimental paradigm is capable of revealing statistically significant performance differences between mobile audio AR implementations.

1. INTRODUCTION

This paper describes a pilot study and main experiment that attempt to measure variations of navigation performance supported by a mobile audio AR system. Examined here are two common technological limitations of such systems: binaural rendering accuracy; and latency between head-turns and corresponding audio output changes. Both parameters are examined in the following experiments on mobile audio AR perception.

The experiments also contribute a new perceptual evaluation paradigm, whereby participants perform an outdoor navigation task using personal, location-aware spatial audio. The task is to navigate from a central base position to the location of a simulated, world-stationary sound source. For each stimulus, system parameters were varied, providing the experimental factors under examination. Simultaneously, body position/orientation and head-orientation sensor data were recorded for later analysis. Objective performance measures were devised to ascertain any participant navigation performance degradation from the ideal, due to tested system parameter values.

The participant task was designed as a generalisation of real-world mobile audio AR applications. For instance, any navigation task can be generalised as a series of point-to-point navigations (A-B) like those in the present experiment, whether it involves a closed circuit (e.g. A-B-C-A) or an open, continuous series of waypoints (e.g. A-B-C-D-...). This experiment thus provides a method of examining a given system implementation's effectiveness at supporting any navigation activity. The same experimental paradigm could also be used to evaluate factors other than rendering technique and latency. For example, it could be used to compare generalised and individualised HRIR filters, or the presence or absence of sophisticated acoustics simulation.

2. BACKGROUND

To date, few mobile audio AR systems have been reported, particularly those with unrestricted, outdoor functionality. There are even fewer published system performance evaluations, possibly due to extensive existing research on *static* spatial audio perception. However, several technical and perceptual factors are unique to the new experience of interacting with sound sources simulated to seem fixed in the world reference frame. Thus, new experiments are required to understand the performance supported by different system designs.

The most relevant prior research appears in a paper by Loomis et al. [1], discovered after having designed and performed the present experiments. Loomis et al. created a "simple virtual sound display" with analogue hardware controlled by a 12MHz 80286 computer, with video position tracking using a light source on the user's head, and head-orientation tracking using a fluxgate compass. Video position measurements occured at a rate of 6 Hz, and covered 15×15 m with a maximum error "on the order of 7 cm throughout the workspace". Head tracking occurred at a sampling rate of 72 Hz, with 35 ms measurement latency, 1.4° resolution and 2° or 3° accuracy in normal conditions, although it occasionally exceeded 20° with poor calibration or head tilts. Spatial audio rendering consisted of azimuth simulation by synthesis of interaural time and intensity differences. Distance simulation was provided through several cues: the "first power law for stimulus pressure"; atmospheric attenuation; a constant-level signal component processed by artificial reverberation (for externalisation and distance cues); and finally, the naturally occurring "absolute motion parallax" cue of changing source azimuth due to body position.

The experiment required participants to "home" into real or virtual sounds placed at the same 18 locations arranged around a circle from the origin, with azimuth and radius slightly randomised. Results measured the time to localise; time for participant satisfaction of successful localisation; distance error at the terminus; and absolute angular change from near the start of a participant's path to near the end. ANOVA testing showed significant individual differences on some measures, but no significant effect of the real/virtual condition on any measure. Mean results for most measures were also quite similar across conditions, except for angular change, which was much larger for virtual stimuli (33°) than real stimuli (14°), despite no significant effect shown by the ANOVA (p<0.05). The researchers concluded that the simple virtual display could be "effective in creating the impression of external sounds to which subjects can readily locomote", but that more sophisticated displays may improve space perception and navigation performance.

This research shows many similarities to the present experiment, however its main thrust is towards basic verification of navigation ability using simple spatialised stimuli, even though, as stated: "homing can be accomplished merely by keeping the sound in the median plane until maximum sound intensity is achieved". In other words, the closed perception-action feedback loop enables determining correct azimuth via head-turns, after which body translation and intensity distance cues can be used to find the sound position. In contrast, the present experiment accepts that a basic system should afford navigation, but it hypothesises that system parameters such as latency and rendering technique might affect performance. The present focus is to determine how system technical performance affects participant navigation performance.

Another relevant study by Walker and Lindsay [2] investigated navigation in a *virtual* audio-only environment, with potential application to navigation aids for the visually impaired. This study focused on the performance effect of waypoint "capture radius", which describes the proximity at which the system considers an auditory beacon to be successfully reached. The simulation environment provided head-orientation interaction but participants navigated while sitting in a chair using buttons on a joystick to translate forwards or backwards in the virtual world, using orientation to steer. Thus, the participants' proprioception and other motion-related senses were not engaged, even though auditory navigation was possible. The present pilot study investigated the same factor, but in augmented reality (that mixed synthetic spatial audio with *real* vision/motion).

Results showed successful navigation for "almost all" participants, with relatively direct navigation between waypoints, and some individual performance differences. The main result identified a varied effect of capture radius on navigation speed and path distance efficiency, with a performance trade-off between the two. For the medium capture radius with the best distance efficiency, navigation speed was worst, and for the small and large capture radii, distance efficiency was lower, but speed was higher. It appears that with a medium capture radius, participants were able to navigate a straighter path through the course, and took more time to do so. This study informed the present pilot study, which sought to find an optimal capture radius to *accentuate* navigation performance differences due to other system parameters, without making the navigation task excessively difficult.

3. METHOD

3.1. Setup

Experiment trials were performed in daylight, during fine weather conditions, on university sport fields, which provided a flat, open, grassy space. The pilot study was conducted with only the author as a participant. The main experiment was performed by eight male volunteers, of unknown exact age, in their 20s or 30s, with one in his early 50s, all with no known hearing problems.

Participants carried the experiment hardware system comprised of a Sony Vaio VGN-U71 handheld computer, interfaced to a Honeywell DRM-III position tracker and Intersense InertiaCube3 using a Digiboat 2-port USB-Serial interface, with audio output to Sennheiser HD-485 headphones. The experiment was managed using custom software (written in C# .NET 2.0) that interfaced the DRM-III and InertiaCube3 and provided a graphical user interface for the participant to respond and control their trial progress. The software also recorded position and orientation tracker data, and controlled the real-time stimulus playback and binaural rendering in Pure Data [3], via the Open Sound Control (OSC) protocol [4].

3.2. Procedure

The experiment was self-paced by the participant using the custom software. For each stimulus, the participant was required to begin by facing in a given direction at a marked *base position* in the centre of a clear space of at least 35 metres radius. Note however that the accuracy of the base position and starting direction was not critical because the software re-zeros its position and direction just prior to starting each stimulus.

When ready, the participant presses a software button to start the current stimulus, which is synthesised to simulate a stationary sound source at a chosen distance and azimuth. The task is to walk to the virtual source position and stop moving when it stops playing, which occurs when the participant reaches a given "capture radius" from the precise source location. If the participant fails to locate the stimulus within 60 seconds, the stimulus stops and a time-out message is displayed. In either case, the software prompts the participant to rate the perceived stability of the source position on a scale of 1 (least stable) to 5 (most stable). In the pilot study, the participant rated "perceived latency" (to head-turns), but this was considered potentially too esoteric for the main study. Finally, the participant must walk back to the central base position, face in the start direction and repeat the process for the next stimulus, and so on until the trial is completed. Figure 1 shows a photograph of a participant walking to locate a stimulus sound source during an experiment trial.

Participants were given the following guidance to prepare them for the experiment:

1. As per previous experiments using the DRM-III position tracker [5, 6], for optimal tracking, participants were asked to walk at a steady, medium pace, only in the direction their torso was facing - i.e. to walk forward and avoid side-stepping or walking backwards.
2. Participants were reminded that the sound is intended to be stationary in the world, positioned at head height (since elevation was not simulated).
3. To help imagine a real sound source, participants were encouraged to look ahead (not at the ground) while they walked, and use head-turns to find the correct source direction.

3.3. Stimuli and experimental factors

Stimuli consisted of real-time spatialised, continuous noise-burst-trains. The raw noise-burst-train itself was a continuously-looped, ten-second sample of Matlab-generated Gaussian white noise enveloped by a rectangular wave with duty cycle of 50 ms on, 100 ms off. The raw stimulus sound pressure level at the minimum simulated source distance of one metre was set to 75dBA per headphone channel with the Vaio sound output at full volume. Thus, the level was repeatable and would never exceed 75dBA even if the participant was positioned directly over a sound source.

Both pilot study and main experiment employed factorial designs, with one of the experimental factors being head-turn latency.

Figure 1: Participant navigating to target during an experiment trial

Latency is expressed as an additional latency value over the baseline total system latency (TSL), found to be 176 ms (±28.9 ms s.d), using a method based on [7].

Pilot study stimuli were rendered using a six-virtual speaker vector-based amplitude panning technique (denoted *VBAP6*) [8]. Each speaker was simulated binaurally by convolving its signal with the appropriate pair of head related impulse responses (HRIRs) from subject three, chosen arbitrarily from the CIPIC database [9]. Distance was simulated only by controlling level in proportion to the inverse square of source distance.

The pilot used 40 factor combinations of five capture radii (1, 2, 3, 4, 5 metres) and eight additional head-turn latency values (0, 25, 50, 100, 200, 300, 400 and 500 ms). Two repetitions of each factor combination resulted in 80 stimuli, each spatialised to a random direction at a distance of 20 metres.

The main experiment used only 40 stimuli in total, comprised of four repetitions of the ten combinations of two render methods and five latencies. The first render method (denoted *Ambi-B*) was first-order Ambisonic decoded to binaural via a six-speaker virtual array, using a energy decoder [10]. The second render method (denoted *VBAP12-G*), used a twelve-speaker virtual VBAP array, with a single, distance-variable ground reflection, using low-pass filtering to simulate a grass surface. Additional latency was set to 0, 100, 200, 400 or 800 milliseconds. Stimuli were spatialised to a random distance between 15 and 25 metres at a random azimuth within five sectors of the circle (to avoid direction clustering). Source capture radius was set to 2 metres due to the pilot study results, reported in Section 4.

3.4. Results analysis

For each stimulus, all raw data available from the DRM-III position tracker and IntertiaCube3 orientation tracker was recorded four times a second. This included position easting and northing, the number of steps taken, body heading and head yaw, pitch and roll. Another file recorded all stimulus factor values (latency, render method, azimuth and range) as randomised for that trial. Lastly, a results file recorded the stimulus factors again, with the source position, participant's rating response, time elapsed, and distance walked (integrated along the participant's path). These data were imported into Matlab and analysed with respect to stimulus factors to produce statistical results.

Results were analysed using several performance measures based on the participant's response time, position tracks and head-orientation movements to navigate to the stimulus position. Performance measures were designed, based on the navigation time and path, in terms of the participant-source geometry shown in Figure 2. Several performance measures have been analysed, but for brevity in the present paper, only distance efficiency (*DE*) is presented. DE is calculated as the ratio of the direct path length (*d*+*c*) to the actual (curved) path length (*a*+*c*), as shown in Equation 1. Both distances are measured to the centre of the stimulus capture circle of radius *c*, assuming the last section of the actual walk path would be a straight line if it was completed.

$$DE = \frac{d+c}{a+c} \tag{1}$$

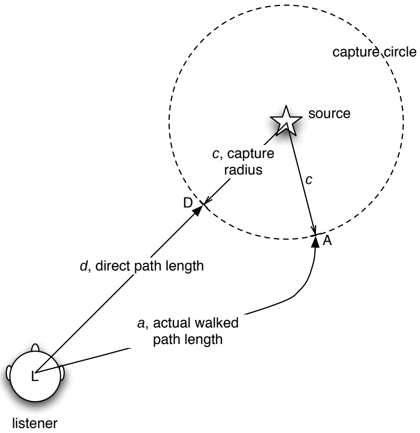


Figure 2: Participant and source geometry, showing source capture radius, direct path and participant's actual walked path to the edge of the source capture circle

Subsequent results analyses mainly use multi-way ANOVA and post-hoc multiple comparison tests using Tukey's Honestly Significant Difference (HSD) at p<0.05. Whenever ANOVA was used, the assumption of normal distribution of the given performance measure was tested, and statistical independence of the experimental factors was checked. Unless stated otherwise, all displayed error bars represent the 95% confidence interval (CI) around data points.

4. PILOT STUDY RESULTS

As noted, the pilot study used only the author as the sole participant and only examined additional head-turn latency and capture circle radius factors. Figure 3 shows the resultant raw data for every stimulus, on a separate plot for each capture radius setting. The plots show the position tracks with a marker at every footstep and a short vector displaying the recorded head yaw data. Also shown is a straight line from the base position to the point source position at the centre of a circle representing the capture radius.

On inspection of the stimulus position tracks, it's clear that some are straighter than others. Ideally the participant would perfectly localise the sound source before they move, then walk directly towards the point source, stopping at the capture radius, where the target is considered reached. In reality, the participant isn't able to perfectly localise the stimulus from the beginning, although thanks to system head-turn interactivity providing perceptual feedback, the initial source bearing determination is usually fairly accurate. Then, while the participant walks in the initial chosen direction, any perceived azimuth errors increase as the participant nears the source, which forces continual reassessment of the momentary source direction. Errors are potentially exacerbated by head-turn latency, especially if the participant keeps walking in a direction based on a delayed head-yaw reading. Thus, added latency was expected to degrade the participant's navigation path from the ideal straight line into a curved or piecewise path with continual corrections along the way.

Initial visual comparison of participant tracks reveals they are straighter for the largest (5 metre) capture circle when compared to the smallest capture circles of 1 or 2 metres. Another basic feature is the apparent tendency for the participant to curve anti-clockwise rather than clockwise as he nears the source position. However, close inspection indicates that anti-clockwise curves are not ubiquitous, for example, see the source at about 30° clockwise from north on the 3 m capture radius plot (Figure 3c). Also, some tracks curve clockwise at the start and end curving anti-clockwise.

Another feature is the tendency for occassional use of larger head-turns, presumably to reassess the source azimuth, e.g. the track just anti-clockwise from north on the 3 m capture radius plot. Larger head-turns mostly occur at the beginning and towards the end of the navigation. At the beginning, the task resembles a traditional stationary localisation experiment, since the participant usually doesn't move away from the base position until they've judged the source direction. Head-turns here enable almost complete mitigation of front-back localisation errors, evidenced by the participant moving away from the base in almost the correct direction in most cases. Greater head movement again becomes useful to adjusts navigation direction along the way to the target.

In general, participant tracks show that navigation to the target was usually very successful, even for the most difficult 1 m capture radius. Further conclusions require derivation of performance measures and statistical analysis with respect to experimental factors. Thus, ANOVA tests were performed to look for significant effects of source capture radius and added head-turn latency on distance efficiency.

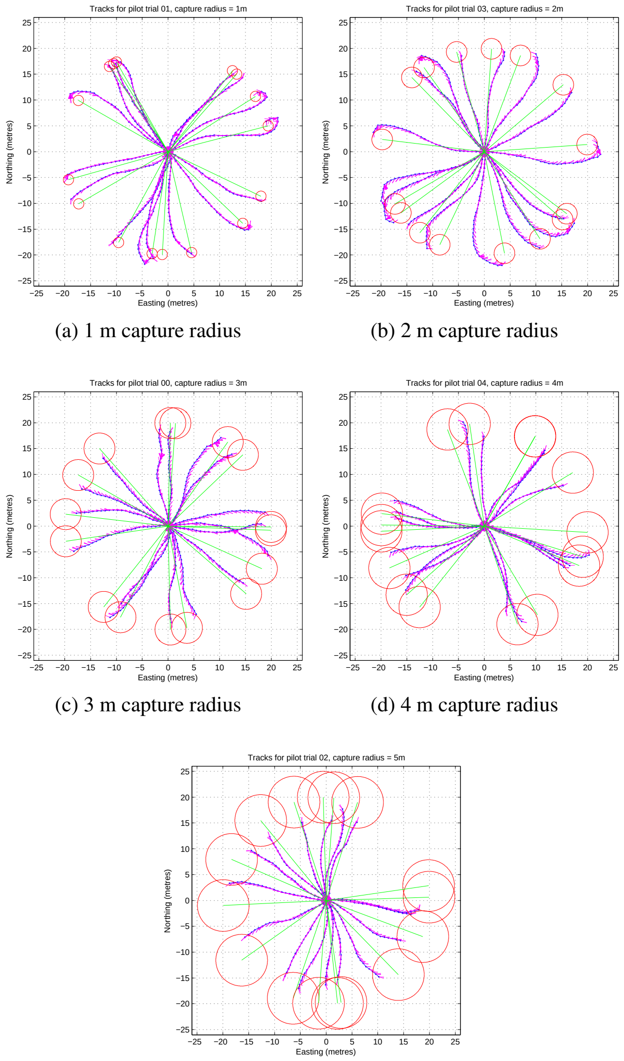
For the pilot study, the ANOVA assumption of normal distributed independant variables was adequate for distance efficiency, but not for latency rating, so it was not analysed. Results show significant effects of capture radius on distance efficiency ($F(4,40) = 6.3, p = 0.00048$). Post-hoc multiple comparisons of capture radius values were then performed using Tukey's HSD (p<0.05).

Figure 4 shows distance efficiency versus capture radius, with significant differences between the 2 m radius and 3, 4 and 5 m radii, but insignificant differences between other value pairs. This insignificant effect of the 1 m radius is unexpected, but might be explained by a weak effect of capture radius, and the small number of repetitions using only one participant. Nevertheless, there appears to be a trend of increasing distance efficiency with increasing capture radius. This is expected since most path curvature occurs close to the source position, so a larger circle allows less opportunity for azimuth localisation errors and head-turn latency to cause navigation deviations. Conversely, smaller capture circles include a greater proportion of the critical final stage of homing-in to the source. A similar conclusion was reached in [2], using a virtual auditory environment.

5. MAIN RESULTS

5.1. Raw position and head-yaw tracks

Figure 5 shows the raw data for every stimulus, on a separate plot for each participant, in which the author was participant 1. As for the pilot study, navigation tracks show each footstep with head-yaw data at that position, and a straight line to the target. The capture radius was set to 2 metres for all stimuli as a result of the pilot study, since this was the largest radius that significantly affected distance efficiency.



(a) 1 m capture radius

(b) 2 m capture radius

(c) 3 m capture radius

(d) 4 m capture radius

(e) 5 m capture radius

Figure 3: Pilot study tracks, showing participant body position, head orientation, source position, bearing and capture radius.

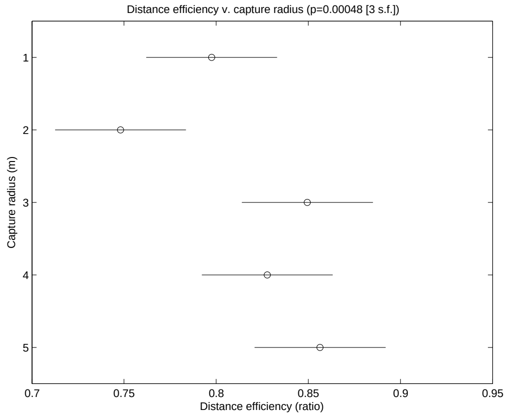


Figure 4: Post-hoc Tukey HSD multiple comparison of capture radius vs. distance efficiency

Participants varied significantly in their ability to navigate directly to the target, and individual navigation performance also varied between stimuli. For all participants, some stimuli were reached almost directly, while others were overshot or reached via a curved path that often became more curved towards the source position. Participants usually headed away from the base position in approximately the right direction, rarely in the opposite direction. Then, during navigation, static localisation errors and possibly head-turn latency affected the momentary perceived target azimuth. Momentary localisation errors such as this increase due to the source/listener geometry as the participant nears the source. This tends to prompt more drastic path corrections, or pauses with head-turns, as participants proceed towards the target. Effectively, the participant's ability to walk an ideal straight path towards the source is reduced depending on their navigation strategy in combination with their perceptual ability, the inherent rendering resolution and system latency.

5.2. Analysis of Variance

ANOVA tests were performed for distance efficiency and subjective stability against the factors of latency, render method and participant, after testing all assumptions. After initial visual inspection of the normal distribution, data transformation was determined necessary for distance efficiency. Several potential transformations were visually inspected and satisfactory results achieved by taking the arcsine value, considered suitable for proportion data [11].

Individual participant differences were highly significant for distance efficiency, but not stability ratings, which were normalised between participants for comparison of the unanchored stability scale. Individual differences were expected, so further significant effects of participant are not mentioned unless they are not trivial.

Distance efficiency was significantly affected by render method ($F(1,258) = 41, p = 8.7 \times 10^{-10}$) and the interaction between head-turn latency and participant ($F(28,258)$ = 1.5, p = 0.048). Inter-participant normalised stability ratings were significantly affected by head-turn latency ($F(4,258) = 8.2, p = 3.2 \times 10^{-6}$), the interaction of latency with render method ($F(4,258)$ = 5.8, p = 0.00018), and the interaction of render method with participant ($F(7,258)$ = 3.8, p = 0.00059).

Interestingly, render method significantly affected distance efficiency, but not stability rating, even though stability was affected by the paired interactions of render method with both head-turn latency and partipant. The interaction with participant is interesting in itself since stability rating was normalised between participants, effectively cancelling individual differences on average. Thus, some participants rated stability significantly differently depending on render method. Head-turn latency alone only significantly affected stability rating. Post-hoc multiple comparison tests (Tukey HSD, p<0.05) are presented next, with error bars on all plots representing 95% confidence intervals.

5.3. Participant differences

Individual differences (displayed in Figure 6) were prominent for distance efficiency, but not stability rating, which was normalised between participants. That said, for distance efficiency, except for Participant 1 (the author), participants were not significantly different, with results averaging around 0.75 efficiency. The fact that the author performed significantly better suggests a possible learning effect, given that the author had more experience with the experimental task, the system, and the binaural rendering used.

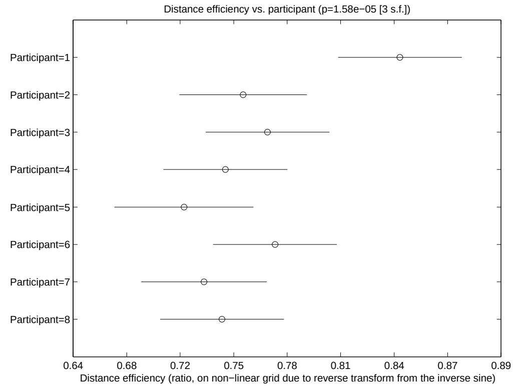


Figure 6: Post-hoc, Tukey HSD multiple comparison of participant vs. distance efficiency

5.4. Effects of rendering method

Render method significantly affected distance efficiency (DE) but not stability rating. VBAP12-G rendering achieved approximately 0.81 DE, significantly better (p>0.05) than 0.71 DE for Ambi-B (Figure 7). This result might be explained by the different angular resolution or localisation blur resulting from the differences between the rendering technologies. VBAP12-G used 12 virtual speakers (with 12 HRIRs, assuming head-symmetry), and a panning method that only interpolates between adjacent virtual speakers. In contrast, Ambi-B used first-order, 2D Ambisonic rendering with energy decoding to 6 virtual speakers (using 6 filters by combining the 2D Ambisonic speaker decode with the HRIRs). A separate, unpublished static localisation experiment using these rendering methods showed better mean localisation errors for the VBAP12-G method (20.5°) compared to Ambi-B (23.1°) [12]. Similar relative results have been obtained in a localisation performance comparison between amplitude panning and Ambisonic rendering to a real 6-speaker array [13].

Overall, VBAP12-G rendering seems to significantly assist navigation to a sound source in terms of distance efficiency, under any head-turn latency conditions. Source stability was not affected by render method on its own, but interaction of render method with head-turn latency was significant, (see Section 5.6).

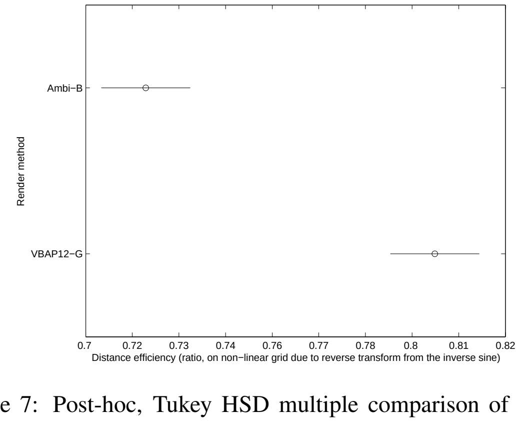
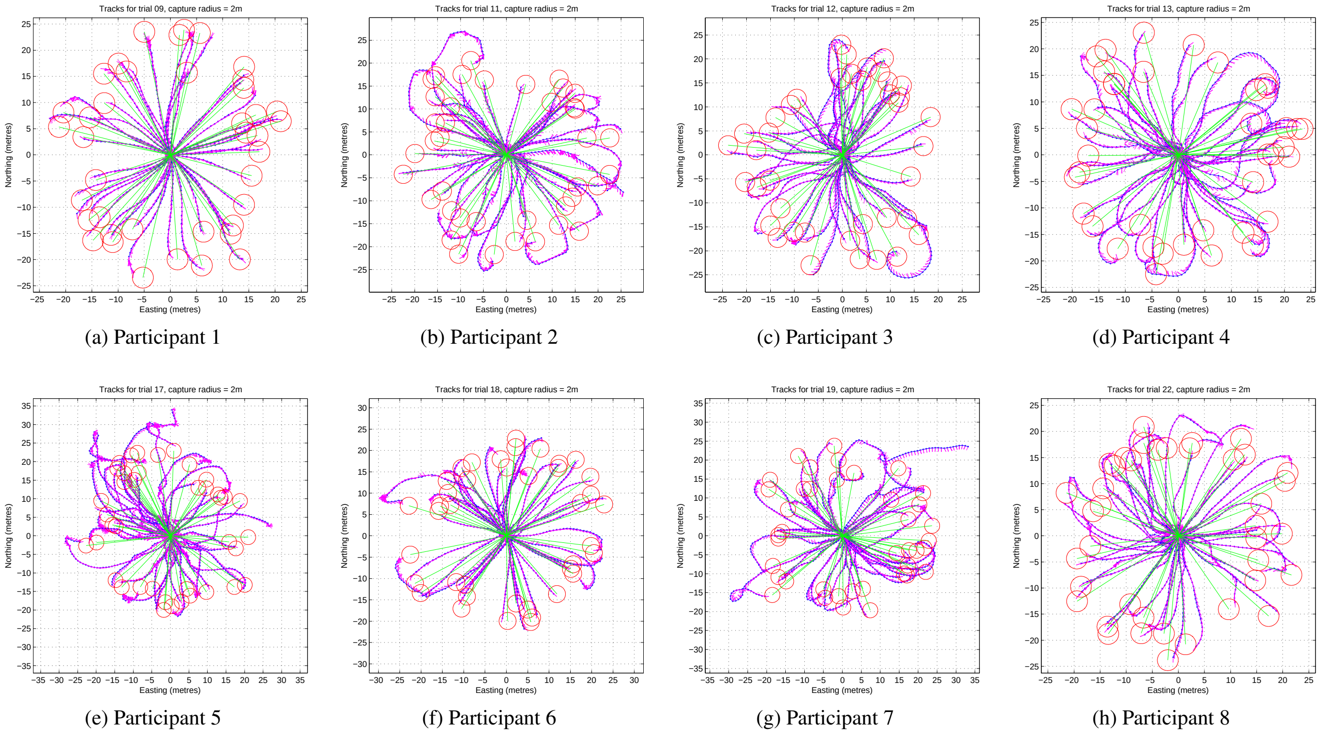


Figure 7: Post-hoc, Tukey HSD multiple comparison of render method vs. distance efficiency



(a) Participant 1

(b) Participant 2

(c) Participant 3

(d) Participant 4

(e) Participant 5

(f) Participant 6

(g) Participant 7

(h) Participant 8

Figure 5: Experiment record of all participant tracks

5.5. Effects of added head-turn latency

Added head-turn latency significantly affected the subjective stability rating (as expected), but not distance efficiency. Post-hoc multiple comparison (Figure 8) shows that only the 800 ms added latency (976 ms TSL) was significantly different to other values (400 ms, 200 ms, 100 ms and 0 ms). The stability ratings (scale: 1-5) ranged from approximately 2.75 for 0 ms added latency to below 2.1 for 800 ms added latency. This appears to be a monotonic trend of decreasing subjective source stability ratings as latency increases.. This conclusion is reinforced by the linear regression presented in Figure 9, which reveals a highly significant (p=0.00104) trend of decreasing stability rating versus added latency.

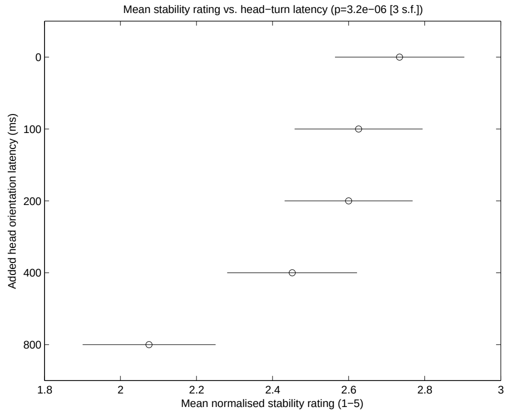


Figure 8: Post-hoc, Tukey HSD multiple comparison of added head-turn latency vs. inter-participant normalised stability rating

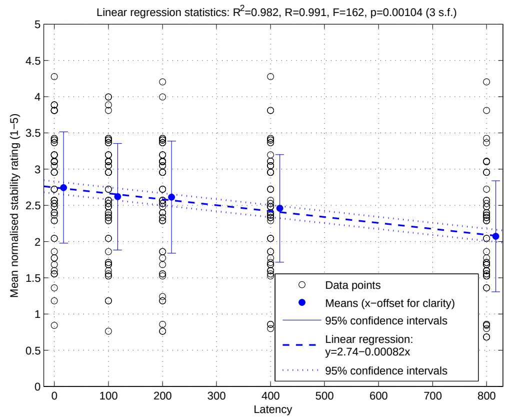


Figure 9: Linear regression of inter-participant normalised stability rating vs. added head-turn latency

5.6. Interaction effects of added head-turn latency and render method

ANOVA results (Section 5.2) revealed that (inter-participant normalised) subjective stability rating was significantly affected by the interaction of added head-turn latency with render method (F(4,258) = 5.8, p = 0.00018). Figure 10 displays the post-hoc multiple comparison of latency and render method versus stability rating. This shows no significant difference for Ambi-B rendering between any added latency (0 - 800 ms, or 176 - 976 TSL), while VBAP12-G produced significantly different stability ratings between several latencies. The range of stability ratings for VBAP12-G rendering spanned from approximately 2.9 at 200 ms added latency, to approximately 1.8 at 800 ms, which was significantly different to the three lowest latencies (0, 100 and 200 ms).

Linear regression analysis displayed in Figure 11 provides further evidence that Ambisonic rendering shows a weak, insignificant (*p*=0.301) trend of decreasing stability rating with increased head-turn latency, while VBAP12-G shows a stronger, significant (*p*=0.0269) trend. Apparently the greater part of the significant trend of stability rating versus latency (Figure 9) can be attributed to VBAP12-G stimuli.

This interesting result could be interpreted as showing that the lower-resolution, higher-blur Ambisonic rendering mitigates the impact of increasing head-turn latency on subjective source stability, which is evident for VBAP12-G rendering. By reducing the azimuth resolution and increasing localisation blur, head-turn latency becomes less perceptible, effectively increasing the upper threshold of head-turn latency before measureable performance degradation. Conversely, higher resolution, lower-blur rendering requires lower head-turn latencies to avoid a perceptible lag.

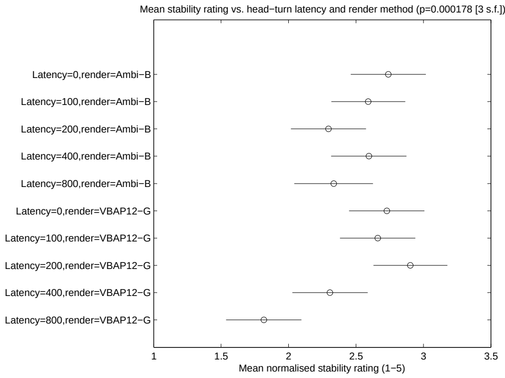
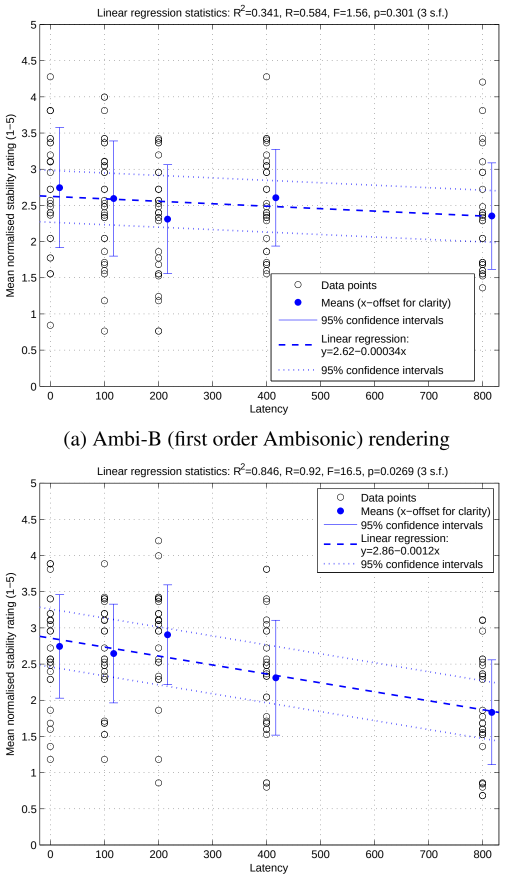


Figure 10: Post-hoc, Tukey HSD multiple comparison of added head-turn latency and render method combinations vs. inter-participant normalised stability ratings



(a) Ambi-B (first order Ambisonic) rendering

(b) VBAP12-G rendering

Figure 11: Linear regression of inter-participant normalised stability rating vs. added head-turn latency, across render methods

6. SUMMARY, DISCUSSION AND CONCLUSIONS

This experiment was designed to study measurable impacts on mobile audio AR user performance due to the most critical system limitations of binaural rendering method and total system latency to head-turns. Rendering is limited by mobile computing power, battery life, and generalised HRIRs, while head-turn latency is potentially limited by the operating system and sound output buffer delays. The experiment task to navigate from a single, central base position to the location of multiple virtual sound sources was designed to generalise any potential navigation task directed by positional auditory beacons.

The pilot study examined the impact of source *capture circle* radius and head-turn latency on performance measures of distance efficiency and head-turn latency rating. Decreasing capture radius significantly reduced distance efficiency, showing that straighter navigation paths were supported by larger capture circles. Pilot study results informed the main experiment, in which capture radius was set to 2 m, the largest circle with significant detrimental impact on distance efficiency. For better navigation performance in mobile AR applications, capture radius should be 3 metres or more for a system with similar binaural rendering techniques, tracking accuracy, and latency specifications to the experiment system.

The main experiment examined the performance impact of head-turn latency and render method (Ambi-B or VBAP12-G), which differ mainly in terms of inherent azimuth resolution and localisation blur. Performance measures were distance efficiency and subjective source *stability* rating. Overall results showed that regardless of sometimes severe system performance degradation, all eight participants successfully navigated to most source positions within 60 seconds. Thanks to system interactivity to position and head-turns, participants set out towards a sound source in approximately the correct direction, despite high front-back confusion rates and azimuth localisation errors common in static (non-interactive) experiments.

VBAP12-G rendering performed significantly better than Ambi-B for distance efficiency, with no significant difference for stability rating. In contrast, added head-turn latency didn't affect distance efficiency, but significantly worsened stability rating, for 800 ms added latency (976 ms TSL). Added latency caused no significant effects for 200 ms (376 ms TSL) or less.

Possibly the most important result was that the interaction between render method and latency significanctly affected subjective stability, despite render method alone having no effect. Significant stability degradation due to increased head-turn latency only occurred for the VBAP12-G rendering, not for the Ambi-B rendering. The lower resolution, higher blur of Ambi-B rendering apparently mitigated the detrimental effect of high head-turn latency on perceived stability. Conversely, VBAP12-G rendering apparently exacerbated latency's effect on stability.

In conclusion, several key findings for implementing mobile audio AR systems were identified. First, even systems with high head-turn latency and/or relatively low resolution (high blur) rendering can afford successful user navigation to positional sound sources, but degradation of both specifications does damage objective and subjective participant performance. Lower resolution, higher-blur Ambisonic rendering decreased navigation distance efficiency, but reduced the detrimental effects of latency on source stability rating. Higher resolution, lower-blur VBAP rendering improved distance efficiency but enabled high latencies to decrease source stability. Navigation performance is thus best supported by improving both system specifications, as might be expected. Finally, mid-range latency (of up to 200 ms) can be tolerated, regardless of rendering method.

7. ACKNOWLEDGEMENTS

This research took place in 2007 and 2008, during the author's PhD candidacy at University of New South Wales, Sydney, Australia. It occurred within the AudioNomad project lead by Dr. Daniel Woo, Dr. Nigel Helyer and Prof. Chris Rizos. AudioNomad was supported by the Australian Research Council as a Linkage Project (LP0348394) with the Australia Council for the Arts, under the Synapse Initiative.

8. REFERENCES

[1] Loomis, J., Hebert, C. and Cicinelli, J., "Active localization of virtual sounds," *J. Acoust. Soc. Am.*, vol. 88, no. 4, pp. 1757–1764, October 1990.

[2] Walker, B. N. and Lindsay, J., "Auditory navigation performance is affected by waypoint capture radius," in *The 10th Int. Conf. on Auditory Display*, Sydney, Australia, July 6-9 2004.

[3] Puckette, M., "Pure data," in *Int. Computer Music Conf.*, San Francisco, 1996, pp. 269–272, Int. Computer Music Association.

[4] Wright, M. and Freed, A., "Open sound control: A new protocol for communicating with sound synthesizers," in *Int. Computer Music Conf.*, Thessaloniki, Greece, September 25-30 1997.

[5] Mariette, N., "A novel sound localization experiment for mobile audio augmented reality applications," in *16th Int. Conf. on Artificial Reality and Tele-existence*, Ed., Hangzhou, China, 2006, pp. 132–142, Springer, Berlin/Heidelberg.

[6] Mariette, N., "Mitigation of binaural front-back confusions by body motion in audio augmented reality," in *Int. Conf. on Auditory Display*, Montreal, Canada, June 26-29 2007.

[7] Miller, J. D., Anderson, M. R., Wenzel, E. M. and McClain, B. U., "Latency measurement of a real-time virtual acoustic environment rendering system," in *Int. Conf. on Auditory Display*, Boston, MA, USA, 6-9 July 2003.

[8] Pulkki, V., "Virtual sound source positioning using vector base amplitude panning," *J. Audio Eng. Soc.*, vol. 46, no. 6, pp. 456–466, 1997.

[9] Algazi, V. R., Duda, R. O., Thompson, D. M. and Avendano, C., "The cipic hrtf database," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics*, Mohonk Mountain House, New Paltz, NY, October 21-24 2001, pp. 99–102.

[10] Benjamin, E., Lee, R., Aficionado, L. and Heller, A., "Localization in Horizontal-Only Ambisonic Systems," *131st AES Convention, preprint*, vol. 6967, pp. 5–8, October 8 2006.

[11] McDonald, J., "Data transformations," 2007.

[12] Mariette, N., *Perceptual Evaluation of Personal, Location Aware Spatial Audio*, Ph.D. thesis, School of Computer Science and Engineering, University of New South Wales, 2009 [in examination].

[13] Strauss, H. and Buchholz, J., "Comparison of virtual sound source positioning with amplitude panning and ambisonic reproduction," in *The 137th regular Meeting of the Acoustical Society of America*, 1999.