Perceptual Evaluation of Spatial Audio for "Audio Nomad" Augmented Reality Artworks

Nick Mariette

Audio Nomad is a three-year art/science research collaboration on the creative and technological potentials of location-sensitive, mobile spatial audio. The first Audio Nomad productions were two versions of *Syren* – a ship-based multi-speaker installation using the ship's position from a GPS receiver to render a two-dimensional soundscape. New work including *Virtual Wall* (Berlin) will create a personal location-sensitive spatial soundscape on headphones using a portable computer, GPS receiver and digital compass. The technological intent is to enable the artist to augment real world objects and spaces with sounds perceived to emanate from them. It is important to know the maximum perceivable accuracy of the intended augmented reality effect, given human and technology limitations, even if soundscape design doesn't always require maximum precision. Ultimately, authoring software features will inform the artist of afforded perceptual quality, enabling better utilisation of the medium's potential. Few similar projects have been produced to date and fewer have published quantitative perceptual evaluation research. This paper reviews the field and describes present experimental results and future work on the perceptual evaluation of binaural spatial audio for mobile augmented reality, especially Audio Nomad artworks.

Audio Nomad is a research collaboration between artist Dr. Nigel Helyer (Sonic Objects), Dr. Daniel Woo (Human Computer Interface Lab, UNSW) and Prof. Chris Rizos (Satellite Navigation and Positioning Lab, UNSW), producing art/science research outcomes with location-sensitive spatial audio technology. The author is a PhD candidate working with Audio Nomad, developing spatial audio synthesis and researching perception of audio augmented reality.

Audio Nomad produced two versions of *Syren*, a ship-based location-sensitive spatial audio installation that renders sounds to a multi-channel speaker array in relation to visible landmarks and regions as the vessel navigated waterways, first on the Baltic Sea (Helyer, Woo et al. 2004; Woo, Mariette et al. 2005), then Sydney Harbour (Helyer, Woo et al. 2006; Woo, Mariette et al. 2006). New Audio Nomad works implement personal location-sensitive spatial audio on headphones, for pedestrian users. *Virtual Wall* will trace the now-absent Berlin Wall through Berlin-Mitte, overlaying space with a complex two-dimensional soundscape generated on a mobile device (Helyer, Woo et al. 2006).

The effectiveness and differentiation of Audio Nomad artworks from location-sensitive works using non-spatialised audio depends partly on user perception of spatial audio presented by the system. In particular, the pedestrian system enables user interaction with content by choosing where to walk and how to move through the augmented space. Self-determined interaction with potentially naturalistic sound presentation highlights spatial sound perception. This *perception-action coupling* (Philbeck, Loomis et al. 1997; Afonso, Katz et al. 2005) enables a new creative medium of manipulating perceivable sound behaviours relative to user movements and positions of real objects. The personal Audio Nomad system is the subject of perceptual evaluations described here.

First, audio augmented reality projects in art and science are reviewed, focusing on user perceptual evaluation. Following is an examination of how technical intentions aim to support creative requirements of Audio Nomad works. Subjective experiments are designed and conducted to understand system limitations and develop a perceptually optimal implementation. Particular human and technological limitations are presented and current experiments described. Finally, the potential application of experimental results to the development of Audio Nomad artworks is discussed.

1. Related location-sensitive spatial audio research

Recent advances in consumer portable computing and position sensing technologies enable implementation of increasingly sophisticated, lightweight audio augmented reality systems. However, the concept of linking "virtual" sound objects to visible/audible real-world objects was first proposed no later than 1993 (Cohen, Aoki et al. 1993), demonstrated by spatialising a ringing telephone sound to seemingly emanate from a real telephone. Audio Nomad advances this concept by presenting audio augmented reality outdoors using a mobile device so users can discover and interact with the synthetic spatial soundscape by walking and turning their head.

Human perception of real and synthetic spatial audio has been extensively researched, although mainly using laboratory experiments with little or no visual stimulus and interaction (if any) only by head-turns. Experimental designs are rarely ecologically comparable to the mobile augmented reality setting, adding two important factors not present in traditional sound localisation experiments: stimuli are multimodal (activating visual/aural sensory relationships); and interaction utilises user body movement as well as head-turns. Following is a brief review of location-sensitive spatial audio projects featuring both these factors.

Hear&There is an outdoors audio augmented reality system developed by the Sociable Media Group at MIT Media Lab (Rozier, Karahalios et al. 2000). It enabled users to leave "audio imprints" in a courtyard for discovery by subsequent users. No significant subjective evaluation was presented although questions were posed about the feasibility and possibilities of navigating a space through sound only.

A Walk in the Wired Woods is an interactive art installation by Hewlett-Packard Labs in Bristol (Hull, Reid et al. 2002), in which a photographic exhibition of woodland views was augmented with location-sensitive music, narration and field recording elements presented on headphones. The sound was stereo, not spatialised, and interaction was via body position and head orientation. The installation received several hundred visitors, observed to typically spend twenty minutes engaged with the work, with anecdotal positive feedback describing a sense of added value from the juxtaposition visual and aural media. A multi-choice questionnaire discovered visitors felt more "immersed", "inspired" and "normal" than "strange", "lonely" or "frustrated", and identified the experience as being closer to a "walk in woods" than a "museum tour".

LISTEN (2001-2003) is an interdisciplinary, multi-institutional audio augmented reality project, probably the most sophisticated of its kind to date, using remote server-based spatial audio rendering, radio frequency positioning with decimetre accuracy, and user devices consisting only of wireless headphones and tracking transmitters (Eckel 2001). It was deployed in several public installations including two art gallery exhibitions and a commercial motor show. Subjective evaluation results were published in the final report (Eckel, Diehl et al. 2003) and other publications. One paper describes possible "perceptual devices" (strategies) for creating coherence between visual and aural content using semantic, spatial and temporal relationships (Goßmann 2004). Tracking and event data was logged for all visitors of an augmented painting exhibition and data visualisations provided evidence that users successfully associated virtual sounds with real visible objects. The final report describes visitor observation and questionnaire results, verifying that most visitors perceived the spatial interaction "devices" to positive effect. One pertinent observation is that questions oriented towards user interaction-mode acceptance were problematic in the art exhibition context, when perception is aesthetically motivated.

The *Personal Guidance System* (PGS) is a project to develop and evaluate a GPS-based spatial audio navigation system for the visually impaired. It began with a concept paper in 1985 (Loomis 1985), and the first prototype was built in 1993. Audio guidance consists of text-to-speech spatialised to navigation waypoint locations retrieved from a geographic information system (GIS) database. Recent research on the PGS focuses on this user interface and its effectiveness as a navigation tool compared to other interface modalities such as non-spatialised speech. This is the most in-depth research identified on a location-sensitive spatial audio system, however it is strongly task oriented and does not address visual stimuli interactions.

Two outdoors location-sensitive spatial audio projects are: *Tactical Sound Garden*, using GPS and relative position without head-turn interaction, installed at the 2006 International Symposium on Electronic Arts, San Jose (Shepard 2004); and *Sonic Landscapes*, a precursor to the Audio Nomad project, which augmented a heritage graveyard in Sydney (Helyer 1999-2001), shown in Figure 1. Neither project published user experience evaluations to date.

In the reviewed projects and research, little has been done to extend established human sound localisation research to the mobile augmented reality situation. Subjective evaluation has been limited to validation or qualitative investigation, not quantitative perceptual experiments.

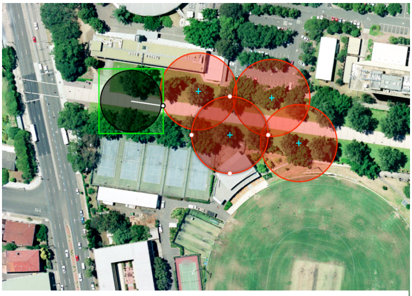
2. Creatively driven technology

Audio Nomad has a primary goal to produce complete artworks as outcomes of science/engineering research and development. This goal necessitates regular creative content development, tests and continuous evaluation of artistic requirements. From the technological development perspective, two users exist: the audience member and the artist.

For the artist, a user-centred design approach (Woo 2005) was followed to develop authoring software enabling a non-programmer to compose location-sensitive spatial audio content as a two-dimensional representation overlaid on detailed aerial photographs or maps. Unlike many artworks with single-use software, the Audio Nomad software is reusable for a deployment anywhere that has GPS reception and availability of appropriate maps or survey images. Authoring software usability evaluation is not discussed here, other than to note that audience perception is dependent on authoring and preview tools used to specify the work. For this reason, repercussions of end-user perceptual experiment results will be analysed for potential incorporation into authoring tool and content preview facilities.

For the audience member, the development goal is to achieve technological transparency (unless the artist wishes to draw attention to the process). While true technological transparency is unlikely, this aim provides a baseline technical specification on which to build creative tools. Research in human sound localization has established the baseline perceptual performance for a stationary body, and present research begins to establish baseline perceptual performance for human bodies in motion.

A better understanding of afforded perception performance will assist in developing new modes of audience engagement with location-sensitive spatial audio works. For example, the artist may design for user investigation, discovery, learning and playfulness with the content. Another example is non-verbal sound/interaction design for directing a user towards content as they stray away from an augmented region.



The authoring software's graphical user interface provides a powerful metaphor for the idea of a precisely located augmentation (Figure 2). A circle can be placed on an aerial photograph to indicate the desired location of a sound, with the centre of the circle representing an omni-directional point sound source. This suggests that by using a detailed aerial photograph, a sound may be placed so precisely, (e.g. on the image of a statue) that the user perceives it to convincingly emanate from the real statue. Whether or not this is true depends on many factors, discussed in the following section.

3. Technological and human limitations to perception

Many technological and human factors affect the perceptual result afforded by any practical location-sensitive spatial audio system. Ideally, the user should hear all synthetic sounds consistently at the intended direction and distance. If the artist wants a less deterministic means of specifying the content in a region, the software should support this within consistent, descriptive control bounds. Consistency in both cases depends on understanding the baseline perceptual ability afforded by a system.

Technological limitations include the need to make the user device portable, lightweight, affordable and self-contained, which limits computation power and spatial audio rendering quality. The user then perceives sound locations with reduced directional accuracy, greater incidence of front-back confusions, (the user locates a sound behind them when it should appear in-front), and externalisation errors, (the sound seems to originate inside the user's head, rather than outside at the intended distance).

Other technological limitations are user position/orientation sensor accuracy and latency (time) before this data affects the spatial audio output. Accuracy contributes to sound source position and direction errors, relative to the user. Latency has a more complex effect, leading to increased localisation errors for brief sounds and slower localisation for longer sounds (Brungart, Simpson et al. 2004).

Human limitations include inevitable issues with *generalised* binaural spatial audio synthesis (for headphones), which works by filtering sound using *non-individual* models or measurements of the external ears. For most applications, including Audio Nomad, models/measurements specific to each individual user are an impossibility, so a generalised measurement is used. Unfortunately, generalised synthesis leads to the same localisation errors as those described above, caused by limiting rendering quality (Møller, Sørensen et al. 1996). Other limitations come from *psychoacoustic* phenomena, for example, the effect of visual stimuli co-present with aural stimuli has been identified to produce the "ventriloquist effect" (Choe, Welch et al. 1975), where the visual biases the auditory image – which could be either advantageous or detrimental. Another human factor is interaction by self-motion: perception/action coupling, which has complex effects, for example, listener motion with inaccurate distance perception can make static sounds seem to move, or cause false perception of moving sound trajectories (Loomis 1995).

4. Perceptual testing

Several subjective experiments will characterise perception of multi-modal, location-sensitive spatial audio; the effect of system design options; and comparison to mono-modal perception of spatial audio without position interaction, from traditional human sound localisation research. Initial results are presented, with aims and implications of future experiments. All experiments take place in an open, grassy outdoors space – a semi-controlled setting ecologically valid to the Audio Nomad artwork application.

The first experiment (Mariette 2006) used a factorial design, and had two aims: first, to validate a novel subjective judgment response method that involves multi-modal perception/action coupling; second, to compare perception performance afforded by two render techniques: ambisonic and discrete binaural rendering, using three binaural filter lengths (64, 128 and 200 samples). Six participants listened to pre-rendered (non-interactive) spatial sound stimuli on headphones and indicated the apparent source position by walking from a base position until the auditory image coincided with a real visual position marker (a camera tripod fifteen metres from the base, shown in Figure 3b). The participant's position was logged four times a second using a Honeywell Dead Reckoning Module (DRM-III) inertial tracking device (Figure 3b).

Data analysis extracts perceived azimuth and distance from position tracks for each stimulus, and localization error is calculated. Results were discarded for one participant with significantly greater azimuth error attributable to poorly followed instructions. Analysis of Variance (ANOVA) and a post-hoc multiple comparison test (Tukey's HSD) of five participants' results revealed significantly lower mean azimuth localisation error for discrete binaural rendering (13 degrees error), compared to ambisonic rendering (17.5 degrees), with under 5% probability of wrongly rejecting the null hypothesis that the mean errors were not significantly different ($p<0.05$). Binaural filter length had no significant effect on localisation errors. The render techniques' afforded localisation performance corresponds closely to traditional sound localisation experiments using seated participants (Strauss and Buchholz 1999), thus validating the novel experimental method.

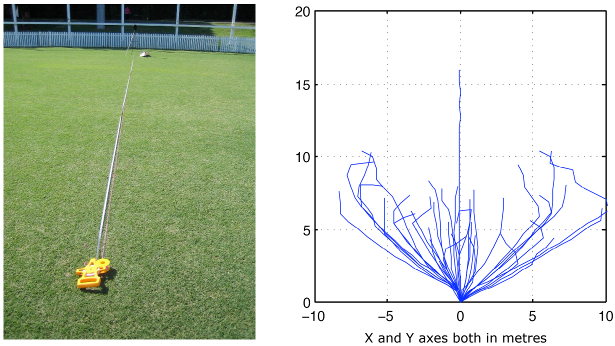


Figure 3 (a) – Experimental setup; (b) – Logged position tracks for one subject

The second experiment (in progress), investigates the hypothesis that human perceptual limitation of front-back errors can be mitigated by position interaction with synthetic spatial audio, again involving perception/action coupling. Stimuli will be rendered in real-time, interactive via position but not head-turns.

The third experiment will investigate the effect of system latency on navigation efficiency to locate a virtual sound source position, with no visual marker. Spatial sound stimuli will be real-time rendered, interactive via position and head-orientation, closing the perception/action feedback loop. The system intends to accurately mimic the perceptual situation for real-world sound sources. Synthetic spatial audio stimuli are effectively equivalent to real stimuli except for intervening technological limitations such as latency.

The final planned experiment will investigate proposed methods to minimise unintentional perceived movements of static sources. The motivation is to mitigate the technological limitation of jitter in the user position data, which otherwise causes unintentional sound source movement relative to the user.

5. Application of perceptual evaluation results

The described experiments aim to create new fundamental knowledge about human perception of synthetic spatial audio in an outdoors, multi-modal setting, with interaction via both head position and orientation, thus involving perception/action-coupling. They also have immediate application potential to the Audio Nomad location-sensitive spatial audio artworks for a pedestrian audience.

Experiments comparing afforded perceptual performance for different rendering techniques (with different computation requirements) enable identification of the perceptually optimal implementation given device price, weight and size limitations.

Experiments that characterise perceptual effects of position and orientation latency enable identification of the perceptually optimal architecture *distribution*. The user device must provide headphone audio output and user position/orientation sensing, however, spatial audio rendering could occur entirely on the user device, or partly or entirely on a remote server, streamed wirelessly back to the user (e.g using WiFi).

Experiments that characterise distance and movement perception might further optimise the user experience by identifying the most appropriate render technique for near and far sound sources. Position jitter could be mitigated for near sounds by reducing directionality and increasing envelopment, while directionality is maximised using different rendering for distant sounds, important to help the user discover further content.

Finally, experiment results might help to design new authoring software features that informs the artist of expected user perception parameters. For example, expected sound localisation accuracy could be visualised by blurring the markers representing sound object locations. Application of perceptual results to user interface features is fertile ground for future applied research.

6. Conclusions

Perceptual evaluation is important to Audio Nomad artworks like *Virtual Wall* for several reasons. Firstly, to verify user ability to perceive spatial sounds using an implementation limited by practical size, weight and cost constraints. Without this quantitative perceptual performance verification, the Audio Nomad system is functionally no different from other location-sensitive audio artworks. Secondly, experiment results inform system design choices such as rendering technique, architecture and different processing for near/far sounds, resulting in a perceptually optimal implementation that best mitigates human and technological limitations. While artworks like *Virtual Wall* don't intrinsically require such detailed perceptual evaluation, it improves and validates user experiences, and simultaneously extends scientific sound localisation research into the mobile audio augmented reality setting.

Acknowledgments

Audio Nomad is a project by Nigel Helyer, Daniel Woo and Chris Rizos, supported by an Australian Research Council Linkage Project (LP0348394) with the Australia Council for the Arts under the Synapse Initiative.

References

Afonso, A., B. F. Katz, A. Blum, C. Jacquemin and M. Denis (2005). A Study of Spatial Cognition in an Immersive Virtual Audio Environment: Comparing Blind and Blindfolded Individuals. 11th Int. Conf. on Auditory Display. Limerick, Ireland.

Brungart, D. S., B. D. Simpson, et al. (2004). The Interaction between Head-Tracker Latency, Source Duration, and Response Time in the Localization of Virtual Sound Sources. Int. Conf. on Auditory Display, Sydney, Australia.

Choe, C. S., R. B. Welch, R. M. Gilford and J. F. Juola (1975). "The "Ventriloquist Effect": Visual Dominance or Response Bias?" Perception and Psychophysics 18: 18, 55-60.

Cohen, M., S. Aoki and N. Koizumi (1993). Augmented Audio Reality: Telepresence/AR Hybrid Acoustic Environments. IEEE International Workshop on Robot and Human Communication.

Eckel, G. (2001). Immersive Audio-Augmented Environments - the Listen Project. 5th Int. Conf. on Information Visualization (IV2001), Los Alamitos, CA, USA.

Eckel, G., R. Diehl, et al. (2003). Listen – Augmenting Everyday Environments through Interactive Soundscapes: 132.

Goßmann, J. (2004). Connecting Virtual Sound and Physical Space in Audio-Augmented Environments. 10th Int. Conf. on Auditory Display, Sydney, Australia.

Helyer, N. (1999-2001) "Sonic Landscapes." Accessed: 22/8/2006. http://www.sonicobjects.com/index.php/sonicobjects/more/sonic_landscapes/

Helyer, N., D. Woo and C. Rizos (2004). Syren - Sonic Cartography Becomes Audio Reality. International Symposium on Electronic Art, Helsinki, Stockholm and Tallin.

Helyer, N., D. Woo and C. Rizos (2006). Syren for Port Jackson. New Constellations, Museum of Contemporary Art, Sydney, Australia.

Helyer, N., D. Woo and C. Rizos. (2006). "Virtual Wall." Retrieved 18th October 2006, from http://www.audionomad.com/projects/vwall.phtml.

Hull, R., J. Reid and A. Kidd (2002). Experience Design in Ubiquitous Computing. Bristol, Client and Media Systems Laboratory, HP Laboratories Bristol.

Loomis, J. M. (1985). Digital Map and Navigation System for the Visually Impaired. Santa Barbara, Department of Psychology, University of California, Santa Barbara.

Loomis, J. M. (1995). Some Research Issues in Spatial Hearing. IEEE Workshop, Applications of Signal Processing to Audio and Acoustics. New Paltz, NY, USA.

Mariette, N. (2006). A Novel Sound Localization Experiment for Mobile Audio Augmented Reality Applications. 16th Int. Conf. on Artificial Reality and Telexistence, Zhejiang University of Technology, Hangzhou, China, Springer LNCS.

Møller, H., M. F. Sørensen, C. B. Jensen and D. Hammershøi (1996). "Binaural Technique: Do We Need Individual Recordings?" Journal of the Audio Engineering Society 44(6): 451-469.

Philbeck, J. W., J. M. Loomis and A. C. Beall (1997). "Visually Perceived Location Is an Invariant in the Control of Action." Perception & Psychophysics 59(4): 601-612.

Roberts, C. (2000) "Craig Robert's Report on How Rtk-GPS Was Used in a Virtual Audio Reality System Demonstration." Accessed: 22/8/2006. http://www.gmat.unsw.edu.au/snap/new/sonic_demo.htm

Rozier, J., K. Karahalios and J. Donath (2000). Hear & There: An Augmented Reality System of Linked Audio. ICAD 2000, Atlanta, Georgia, April 2000.

Shepard, M. (2004) "Tactical Sound Garden [Tsg] Toolkit." 4 Accessed: 28/10/2004. http://www.andinc.org/tsg/

Strauss, H. and J. Buchholz (1999). "Comparison of Virtual Sound Source Positioning with Amplitude Panning and Ambisonic Reproduction." The Journal of the Acoustical Society of America 105(2): 934.

Woo, D. (2005). Discussion of User-Centred Design Approach. N. Mariette. Sydney.

Woo, D., N. Mariette, N. Helyer and C. Rizos (2005). "Syren – a Ship Based Location-Aware Audio Experience." Journal of GPS 4(1-2): 41-47.

Woo, D., N. Mariette, J. Salter, C. Rizos and N. Helyer (2006). Audio Nomad. ION GNSS 2006, Fort Worth, Texas.