Audio Engineering Society
Conference Paper
Presented at the Conference on
Audio for Virtual and Augmented Reality
2016 September 30–October 1, Los Angeles, CA, USA

Spatial Auditory Feedback In Response to Tracked Eye Position

Durand R. Begault

[1]*Human Systems Integration Division, NASA Ames Research Center, Moffett Field, California*

Correspondence should be addressed to Durand R. Begault (Durand.R.Begault@nasa.gov)

ABSTRACT

Fixation of eye gaze toward one or more specific positions or regions of visual space is a desirable feature within several types of high-stress human interfaces, including vehicular operation, flight deck control, target acquisition, etc. It is therefore desirable to have a means to give spatial auditory feedback to a human in such a system about whether or not the gaze is specifically directed towards a desired position. Alternatively, it is desirable to use eye position as a means of controlling a device that provides auditory feedback so that there is a correspondence between eye position and control voltages that manipulate aspects of an auditory cue that includes spatial position, pitch and/or timbre.

1 Introduction

This paper describes the concept, implementation and applications of a novel means of using spatial audio based on the gaze direction of one or more users.

2 Background

About 15 years ago, researchers at the NASA Ames Vision Group collaborated with researchers in the Spatial Auditory Display Laboratory to investigate evidence for a specialized auditory motion processing system in the brain [1, 2, 3]. With the head fixed (or when using a head-mounted display without head tracking), a moving visual target is followed by the eyes either using rapid movement between fixed locations (saccadic eye movements) or by using smooth movement (pursuit). Smooth pursuit only occurs when the image moves smoothly across the retina. Considerable evidence had previously existed on the existence of specialized visual motion-processing system where smooth pursuit eye movements respond to the motion of a visual target, independent of eye position [4]. The collaborative research sought to determine whether or not an independent auditory motion detection mechanism exists.

A virtual acoustic technique was designed to provide stimuli to initiate smooth eye pursuit, following an initial eye fixation position. Virtual auditory stimuli were presented through stereo headphones (Sennheiser HD580). The source sound consisted of a looped recording of a continuously rung small bell, providing a stream of transient attacks considered optimal for localization since it contained broadband spectral energy principally n the region of maximal hearing sensitivity (1-4 kHz). This source was digitally processed using 3-D audio software (SLAB [5]) to simulate binaural audio cues (interaural time, intensity, and spectral cues) that would result from an actual moving sound source. The spectral cues were computed using each individual observer's measured head-related impulse responses (HRIRs) using a blocked meatus technique [6]. Minimum phase approximations of the individualized HRIRs were used to generate the stimuli. Sound trajectories were rendered by dynamically updating the filter coefficients and interaural time differences of the minimum-phase HRIRs in real time using the parameter cross-fading method described in detail in [7].

Participants' eye movements followed the motion of a virtual auditory target with a combination of pursuit and saccades. Figure 1 shows sample eye position and velocity traces for the two observers. Periods of pursuit (vertical arrows) would generally precede the initial saccade (leftward arrows), as has been reported for visually driven pursuit [4]. The instantaneous position of the auditory target at the moment of pursuit onset was in the opposite direction. This supports the proposition that pursuit of a moving auditory target can be driven by the auditory motion signal alone, even when in conflict with any potential auditory position cue [3].

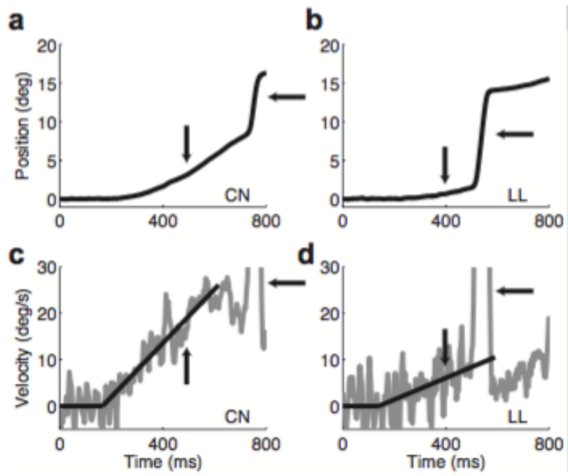

Figure 1. Oculomotor response to auditory motion. Raw eye position (a,b) and velocity (c,d) traces for two observers for an auditory target moving at 60°/s. Time zero is when the fixation spot was extinguished. Vertical arrows indicate initial periods of pre-saccadic smooth pursuit. Leftward arrows indicate saccades evident from their stereotyped rapid change in position and large peak in velocity (from [3]).

3 Conceptual Approach

Fundamentally, spatial auditory feedback is potentially useful not only for motivating smooth pursuit in experimental contexts. That success motivated development of several applications concepts that someday may prove potentially useful. Since the time of our initial experiments, the use of eye-position sensitive technologies has become more mainstream (such as used on some Samsung Galaxy PDAs). There has been related nascent research in applications for the blind and for perceptual research [8, 9]. The concepts presented here have motivated the development of rough prototypes and demonstrations over the years aimed primarily at improving performance in a human factors context. They include intra-viewer feedback; human-machine interaction; and collaborative viewer feedback.

Fixation of eye gaze toward one or more specific positions or regions of visual space is a desirable feature within several types of high-stress human interfaces, including vehicular operation, flight deck control, target acquisition, etc. It is therefore desirable to have a means to give feedback to a human in such a system about whether or not the gaze is specifically directed towards the desired position. Alternatively, it may be desirable to use eye position as a means of controlling a device that provides auditory feedback or auditory sound in general, so that there is a correspondence between eye position and control voltages that manipulate aspects of an auditory cue delivered by a sound synthesis engine, including spatial position, pitch and/or timbre.

A general description of an implementation approach we have used is shown in Figure 2. Eye tracking hardware and software provide a data stream at a relatively rapid rate. The update rate and consequent moment-to-moment position data is typically greater and "jittery" than a desired rate of change in virtual audio position for indicating a gaze direction (as opposed to smooth pursuit). Since saccades can cause rapid changes in position whose velocity exceeds perceptually useful auditory movement velocities, the first component of our approach involves a temporal-spatial smoothing algorithm that low-pass filters and/or integrates positional-temporal information. In this way the averaged gaze direction can be discerned over an interval instead of "instantaneous" eye position.

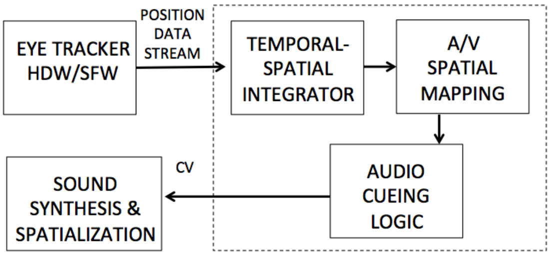

Figure 2. System block diagram.

A non-linear mapping from the smoothed eye position to the position of the auditory cue is the second component of our approach. A particular individual's perceptual mapping of auditory space in relationship to auditory cues heard in actual space may require re-mapping from "intended" to "perceived" positions. In many cases it may be more practical to actually 'exaggerate' the auditory positional cue relative to the eye position. For instance the eyes may have a maximal lateral displacement of +/- 40 degrees, but cueing to +/- 90 degrees may provide a more reliable cue. More complex mappings may be appropriate depending on the task.

A final component includes audio cueing logic that provides feedback only when certain circumstances are met. For instance, when the user's gaze has drifted away from an intended position, the logic may be to provide auditory feedback, but otherwise, to be silent. The sound playback logic also contains means for communicating control voltages to an audio tone generator, or alternatively, a spatial sound-rendering device.

Data input regarding head position is necessary when eye position must be oriented towards a target in the external environment. In collaborative user scenarios, head tracking data from multiple sources must be accounted for.

4 Applications

Deficiencies in smooth pursuit and visual scanning in general are evident in persons afflicted by schizophrenia, autism, trauma, or drugs and alcohol. Autism spectrum disorder can include traits such as disengagement manifested by lack of eye contact in gaze direction. Eye tracking technologies have been proposed as a research tool and possible diagnostic for early autism in [10]. It may or may not be possible to use auditory feedback as part of a behavioral therapy approach to provide positive feedback when eye contact is made towards a particular target.

Fixation of eye gaze toward one or more specific positions or regions of visual space is a desirable feature within several types of high-stress human interfaces, including vehicular operation, flight deck control, target acquisition, etc. The efficacy of using spatial auditory cues for target acquisition in a "head-up" auditory display is described in [11]. For target acquisition, feedback can be provided to a user about whether or not the gaze is specifically directed towards a desired direction. For example, a tone representing a target can be generated from a specific position, and another tone can be generated from the user's gaze direction; when the tones overlap in virtual space, the user is cued that they are looking in the intended direction.

Another human factors scenario concerns situational awareness and visual scanning. It is possible to not only direct the gaze of a user with spatial audio, but also to inform the user when they have drifted from an intended position, as well as to provide information about the current position of the eye. Visual scanning out-the-window or across a horizon is an important safety factor in everyday navigation tasks such as driving a car, and in specialized tasks such as air traffic control. Fatigue can cause a lack of visual scanning. One possible application method for improving vigilance would be to periodically create a moving spatial auditory cue that a user would be trained to visually follow, either with or without a secondary auditory cue that tracked their own eye position (for instance, when not following the cue tone for a certain time interval).

It may be desirable to use eye position as a means of controlling a device that provides auditory feedback or auditory sound in general, so that there is a correspondence between eye position and control voltages that manipulate aspects of an auditory cue delivered by an audio tone generator, including spatial position, pitch and/or timbre. Complex interactions within applications such as gaming sound design or new electronic music instruments are limited only by imagination.

In a "collaborative" scenario, one person is informed by spatial auditory cues of the direction and duration of the gaze of a second person.

Consider an example of two security officers, located at different vantage points, who are assigned to monitor a large crowd of people in an airport or at a music concert. Currently, officer one would communicate the location of a suspicious person (target) to officer two using a verbal description over a radio channel: e.g., "look at the person in the black jacket to your left, over by doorway number 5." In a collaborative application of spatial audio and eye tracking, officer one would look in the direction of target one and activate an alert tone to officer two. Through the use of head tracking for both officers, the alert tone is spatially processed to come from the proper relative direction for officer two to direct their gaze. "Fine tuning" could occur for officer two as described above for target acquisition, in that their gaze direction could be signaled spatially and brought to overlap the signal from officer one's direction of gaze. Presumably, given idealized hardware and software, this would allow target acquisition in a fraction of the time needed for verbal communication.

5 Summary

This paper has described several concepts for implementation and application of spatial audio processing driven by eye position, for one or more users. The practical implementation challenges for both eye and head tracking are obvious, but overall the approach has promise and is clearly an interesting area for future research for improved performance in human-machine interfaces as well as for entertainment.

References

[1] Krukowski, A. E., D. R. Begault, E. M. Wenzel, and L. S. Stone (2001). Human oculomotor responses to virtual auditory motion. *Society for Neuroscience 31st Annual Meeting*, *San Diego, CA*.

[2] Krukowski, A. E., D. R. Begault, E. M. Wenzel, and L. S. Stone (2002). Human smooth pursuit eye movement responses to visual, auditory, and imagined target motion. *The Society for the Neural Control of Movement Twelfth Annual Meeting, Naples, Florida,* poster 404.15.

[3] Krukowski, A. E., D. R. Begault, E. M. Wenzel, and L. S. Stone. Pursuing the invisible: Human Oculomotor Response to Auditory Motion. Unpublished manuscript.

[4] Rashbass, C. (1961). The relationship between saccadic and smooth tracking eye movements. *Journal of Physiology (London)* v. 159, pp. 326-338

[5] Miller, J. D. & Wenzel, E. M. (2002) Recent developments in SLAB: A software-based system for interactive spatial sound synthesis. in *Proceedings of the International Conference on Auditory Display, ICAD 2002* pp. 403-408.

[6] Anderson, M., Begault, D. R., Godfroy, M., Miller, J. D., Roginska, A., & Wenzel, E. (2006, May). Design and verification of HeadZap, a semi-automated HRIR measurement system. *Audio Engineering Society 120th Convention,* Paper 6655.

[7] Wenzel, E. M., Miller, J. D. & Abel, J. S. (2000). Sound Lab: A real-time, software-based system for the study of spatial hearing. *Audio Engineering Society 108th Convention,* Paper 5140.

[8] Jones, F. (2012). Auraleyes: investigating the fusion of eye-tracking and spatial audio in electronic sensory aids for the blind. Ph.D. Thesis, University of Colorado at Boulder.

[9] Xiao, M., Wong, M., Umali M., Pomplun, M. (2007). Using eye-tracking to study audio-visual perceptual integration. *Perception* v. 36, pp. 1391-5.

[10] Falck-Ytter, T., Bölte, S., and Gredebäck, G. (2013). Eye tracking in early autism research. *Journal of Neurodevelopmental Disorders* v. 5:28.

[11] Begault, D. R. (1993). Head-up Auditory Displays for Traffic Collision Avoidance System Advisories: A Preliminary Investigation. Human Factors, v. 35, pp. 707-717