EFFECT OF INCREASING SYSTEM LATENCY ON LOCALIZATION OF VIRTUAL SOUNDS

ELIZABETH M. WENZEL

NASA Ames Research Center, Moffett Field, CA, USA
bwenzel@mail.arc.nasa.gov

In a virtual acoustic environment, the total system latency (TSL) refers to the time elapsed from the transduction of an event or action, such as movement of the head, until the consequences of that action cause the equivalent change in the virtual sound source. This paper reports on the impact of increasing TSL on localization accuracy when head motion is enabled. Five subjects estimated the location of 12 virtual sound sources (individualized head-related transfer functions) with latencies of 33.8, 100.4, 250.4 or 500.3 ms in an absolute judgement paradigm. Subjects also rated the perceived latency on each trial. The data indicated that localization was generally accurate, even with a latency as great as 500 ms. In particular, front-back confusions were minimal and unaffected by latency. Mean latency ratings indicated that latency had to be at least 250 ms to be readily perceived. The fact that accuracy was generally comparable for the shortest and longest latencies suggests that listeners are able to ignore latency during active localization, even though delays of this magnitude produce an obvious spatial "slewing" of the source such that it is no longer stabilized in space.

INTRODUCTION

In a virtual acoustic environment (VAE), the total system latency (TSL) and the effective update rate are distinct parameters although they may be related in practice. The total system latency, or end-to-end latency, refers to the time elapsed from the transduction of an event or action, such as movement of the head, until the consequences of that action cause the equivalent change in the virtual sound source location. Latencies are contributed by individual components of a VAE system, including tracking devices, signal processors, software to control these devices, and communications lines. TSL differs from the "internal latency" [1, 2] of each system component; e.g., in a spatialization device, the internal latency is the delay between acquisition of location data and the rendered audio output. Typically, the TSL is not simply the sum of the various components' internal latencies and it may also vary over time. Update periods (period = 1/rate) in a VAE system refer to various sampling or rendering intervals which may be present; for example, the time elapsed between successive samples of the listener's head motion (1/tracker update rate) and the time elapsed between calculation of one spatial location and a new spatial location by a spatialization engine (i.e., 1/frame rate). Due to differences in sampling rates, the effective update rate usually corresponds to the update rate of the slowest component in a VAE system. As with latency, there is no reason to expect that a system's update rate remains constant over time. Thus, measurements of the mean, standard deviation, and range of the TSL and update rate provide a better characterization of these parameters.

Surprisingly little is known regarding the impact of introducing latency during dynamic localization although it is clearly a critical issue for virtual environments. One recent study [3] investigated the perceptual impact of parameters like system latency, update rate, and spatial resolution. Update rate and spatial resolution were manipulated by independently changing the parameters of a Polhemus Fastrak, while increased latency was achieved by adding 60-Hz multiples of delay to the minimum latency (29 ms). The subjects' task was to point a toy gun that had a tracking sensor mounted on the handle at the apparent location of an anechoic virtual source. Localization performance was measured by the standard errors of the signed azimuth and elevation components of the pointing response and the average time between judgements in a block of trials. It was found that, compared to the best parameter values possible, localization performance did not significantly degrade until the system latency increased to 96 ms or the update rate decreased to 10 Hz. Increasing the spatial resolution to 13°, the largest value tested, had little impact on localization error. However, the psychophysical method that was used to measure localization accuracy was self-terminated by the subjects, resulting in average trial lengths of about 1.75 s. Such stimuli durations may not have been long enough to allow adequate head-motion sampling by the listeners. Also, the average directions of the pointing

responses and the front-back confusion rates (the localization error most affected by enabling head motion) were not reported in [3]. The authors may have chosen not to report such data because of the large individual differences they observed in their data.

A similar, but somewhat slower (greater TSL), virtual audio system than the one used here has also been used in previous studies of localization with and without head motion [4-6]. These studies demonstrated that, compared to static localization, enabling head motion dramatically improved localization accuracy of virtual sources synthesized from non-individualized head-related transfer functions (HRTFs). In particular, average front-back confusion rates decreased from about 28% for static localization to about 7% when head motion was enabled [4]. Confusion rates on the order of 5% are typically observed during static localization of real sound sources [7].

Measurements of the TSL of the system used in [4-6] indicated a mean and standard deviation of 54.3 +/- 8.8 ms, and minimum and maximum values of 35.4 and 74.6 ms [8]. Examination of the head motions that listeners used to aid localization in [4-6] suggests that the angular velocity of some head motions (in particular, left-right yaw) may be as fast as about 175°/s for short time periods (e.g., about 1200 ms). A maximum TSL of 75 ms could potentially result in short-term under-sampling (compression) of relative listener-source motion as well as positional instability of the simulated source. From psychophysical studies of the minimum audible movement angle (MAMA, [9]) for real sound sources (listener position fixed), one can infer that the minimum perceptible TSL for a virtual audio system should be no more than about 69 ms for a source velocity of 180°/s. If one assumes that these thresholds apply to relative source-listener motion in general (e.g., when the source is fixed and the listener is moving), then the positional displacement of the simulated source due to TSL in [4-6] may have occasionally exceeded the perceptible threshold. Although listeners did not report any obvious instability in source position in those studies, it is useful to formally investigate the impact of varying system parameters like latency in order to characterize the dynamic performance needed in a VAE system to achieve adequate perceptual fidelity.

This paper reports on the effect of systematically increasing TSL on localization accuracy when head motion is enabled. The psychophysical method was similar to that used in [4-6]. However, virtual sources were synthesized from individualized HRTFs measured with a blocked ear canal technique. The subjects' task was to estimate the azimuth, elevation and distance of a target source using a graphical response method. Subjects also rated the perceived latency on each trial. It was expected that increasing latency would degrade localization performance, in particular, that front-back confusion rates would increase with longer latencies.

1. METHOD

1.1 Subjects

Five young adults (3 male, 2 female, ages 16-24) served as paid, volunteer subjects. All had normal hearing, verified by audiometric screening at 15 dB HL, and reported no history of hearing problems. None of the subjects had previous experience in auditory localization experiments or virtual environments.

1.2 Stimuli

The basic stimulus consisted of broadband Gaussian noise of 8-second duration with 10-ms, exponential ramps at onset and offset. Independent samples of the noise were computed in real time using a 24-bit DSP card (Spectrum TMS320/C25) in a Pentium computer. The noise signals were then converted to analog form, level-adjusted and low-pass filtered at 20 kHz (Acoustetron LP Amp), input to Convolvotron boards hosted by the same computer, and again converted to digital (16-bit) form.

Each stimulus was digitally processed in real time by the Convolvotron so that it would simulate one of twelve free-field locations. The processing was based on the direction-specific, outer ear characteristics measured for each subject. The HRTF measurement system used was based on a Crystal River Engineering "Snapshot" system. This system uses a blocked meatus technique with a Golay-code pseudo-random signal, along with post-processing to remove the effects of the listening environment, loudspeaker, and microphones. This allows measurement in a non-anechoic environment, since the post-processing windows the direct sound portion of the signal. Minimum phase approximations of the individualized HRTFs were used to render the stimuli. Briefly, the magnitudes of the minimum-phase filters are the same as the original finite impulse response filters, the phase is derived from the magnitude spectra, and the interaural delay is represented by a pure delay estimated from the peaks of the cross-correlations of the left and right-ear HRTFs. The HRTFs were corrected for the headphones used in the study, although the correction was based on an average headphone response derived from many subjects' previous measurements. Filter lengths were 256 points.

The Convolvotron's specifications state an update rate of 33 Hz and latency of 32 ms. It received head-position data from a Polhemus Fastrak at a nominal update rate

of 120 Hz (115.2 kBaud serial line). The host computer was a 90-MHz Pentium running Windows 95. Measurements of the best or minimum total system latency were conducted to assess the overall dynamic performance of the synthesis system, including the head-tracker, using the method described in [8]. TSL values ranged from 21.8 to 45.9 ms, with a mean and standard deviation of 33.8 +/- 5.0 ms.

During each trial, the orientation of the listener's head was tracked and the stimuli were synthesized in real time using the Convolvotron to simulate a stationary external sound source. Only the orientation of the listener's head was utilized to control relative sound position since the subjects were seated and not allowed to move about the room. Also, we did not wish to attempt to simulate the near-field effects that would be required if the listener was allowed to get too close to a virtual source. Synthesis of smooth relative motion was achieved by linear interpolation between impulse responses derived from the four nearest minimum-phase HRTFs, with the interaural delays interpolated separately and inserted at the end of the filtering process [10]. The HRTF map of the Convolvotron has a resolution of 30° in azimuth and 18° in elevation.

On each trial, one of four latency conditions was presented. Delays corresponding to multiples of the tracker sampling interval (8.3 ms) were created by building a "first-in, first-out" queue. The queue then maintained a fixed number of head position samples, with fresh data inserted at the end and latent position data provided to the synthesis chain from the front. The various queue sizes used were 0, 8, 26 and 56 tracker positions, corresponding to average TSLs of 33.8, 100.4, 250.4 and 500.3 ms, respectively. The relationship between the number of skipped tracker samples and TSL was verified using the method in [8].

Following spatial synthesis, the signals were again converted to analog form, passed through anti-aliasing filters (Krohn-Heit 20-kHz low-pass elliptic filters), and fed to a custom headphone driver. Finally, the stimuli were transduced by headphones (Sennheiser HD-430) and presented at an overall level of about 70 dB SPL.

1.3 Procedure

An absolute judgement paradigm similar to earlier experiments [4-6] was used. However, instead of providing verbal/numerical estimates of location, the subjects' task was to indicate the apparent azimuth, elevation and distance of a virtual source using a graphical response method. (Fig. 1). Using a mouse, listeners moved two vectors so that they corresponded to the apparent azimuth and elevation of the target location. The azimuth and elevation displays were yoked such that the azimuth vector determined the orientation of the representation of the head in the elevation display. Moving a dot along the length of the azimuth vector also indicated the relative distance of the source. Subjects were instructed that the distance scale was anchored by the following categories: 0 inches for a sound at the center of the head, 4 inches for a sound located at the perimeter of the head, and at 1 foot, 2 feet, and greater than 2 feet for externalized sounds. For example, a sound heard three feet away and directly in front would produce a response of 0° azimuth, 0° elevation, and 3 feet (distance category 5). A verged-cranial sound heard directly to the left and somewhat elevated might produce "- 90° azimuth, + 15° elevation, and 4 inches (distance category 2). Subjects also rated the amount of latency on each trial by adjusting the pointer on a slider bar with endpoints labelled "minimum" and "maximum" latency (arbitrary scale values of 0 to 25).

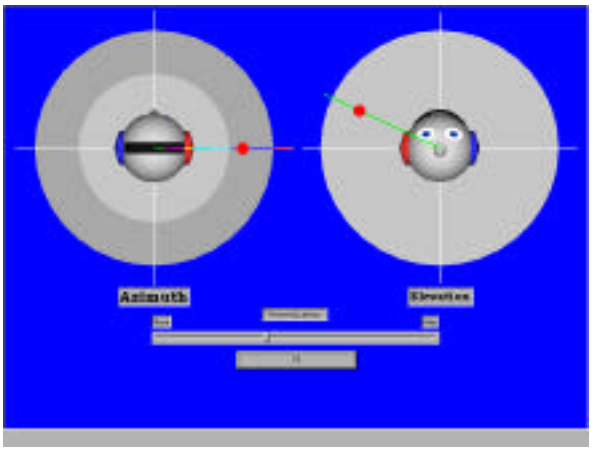

Figure 1: Illustration of the graphical response screen.

Listeners were presented stimuli from twelve different source locations (Table 1), with latency values of 33.8, 100.4, 250.4 and 500.3 ms, for a total of 48 stimuli. Each stimulus was repeated 5 times. The 240 trials comprising the twelve locations and four latencies were randomized and then separated into ten, 24-trial blocks with a different randomized order for each subject. Approximately 5 blocks were run per day with rest-breaks given at least every 2 to 3 blocks. Prior to the experimental runs, a training session was conducted which included a verbal explanation of the response co-ordinates and one to two practice blocks for training on the localization task using only the minimum latency condition. A different block of trials was used to demonstrate the minimum and maximum latency conditions to the subjects at four representative locations; 0°, 90°, 180° and –90° azimuth (0° elevation).

During testing, subjects were seated on a pivoting chair inside a 10-ft square, double-walled soundproof chamber in front of a table with a color monitor, keyboard, and mouse. The Fastrak source was mounted on a wooden rod suspended from the ceiling. The source was about 18 inches from the top of the subject's headphone band where the tracker sensor was mounted. At the beginning of each session, the lights were dimmed in the room and the subjects donned the headphones. At the start of each trial, subjects were required to orient straight-ahead (facing the CRT screen) to within +/-5° azimuth and elevation. Feedback regarding their orientation was given and when they were within the 5-degree limits, they pushed the space bar to begin a trial. At this point, the tracker was calibrated so that the initial position of the subject's head determined the 0°, 0° orientation for each trial. Subjects were instructed to begin each trial by orienting straight ahead and then move (reorient) their heads as much as possible in order to localize the sound source. However, they were also instructed to remain seated and not to lean their heads far forward or to the side in order to stay within the best operating-region of the head-tracker. They then heard the 8-second noise stimulus and provided their estimates of azimuth, elevation, distance, and latency during a self-paced response interval. Feedback was not provided. A record of their head position and orientation was also stored for each trial.

Table 1: Target locations used in the study.

Azimuth	Elevation
0	0
-30	-36
-45	0
-90	36
-135	0
150	36
180	0
135	0
120	36
90	-36
60	-36
45	0

2. RESULTS AND DISCUSSION

2.1 Localization Data

Localization judgements tend to be corrupted by two kinds of error, relatively small errors on the order of 10 to 20° and the special class of errors known as confusions (sounds heard with a front-back or up-down error across the interaural or horizontal axes). When confusion rates are low, as with real sound sources, confusions are usually corrected or eliminated during data analysis. However, confusion rates tend to be high with virtual sources under some conditions and must be dealt with in some other way. Here, the triple-pole plotting technique described in [11] has been adopted in combination with the method used for computing front-back and up-down confusion rates in [7]. Briefly, the triple-pole method represents the azimuth judgement in terms of two angles. The left-right angle is formed by the judgement vector and the median plane (i.e., the laterality of the judgement: -90° left and +90° right). The front-back angle is formed by the judgement vector and the vertical plane passing through the two ears and distinguishes judgements in the front vs. rear hemispheres (-90° rear and +90° front). The up-down angle is simply equivalent to the elevation judgement.

Confusion rates and triple-pole representations of the raw data were computed separately for each subject in each latency condition. Triple-pole plots for a representative subject are shown in Figure 2. In general, the pattern of the judgement angles appears to be largely unaffected by increasing latency.

The pattern of the front-back and up-down judgement angles is also reflected in the mean confusion rates summarized in Figures 3 and 4. Mean azimuth (front-back) confusion rates were low overall, ranging from 5.2 to 8.8%, and the effect of latency was non-significant. Subjects differed in that some showed primarily front-to-back confusions while others exhibited both front-to-back and back-to-front confusions. Mean elevation (up-down) confusion rates were higher, ranging from 11.3 to 21.3%. The effect of latency was also significant, with confusion rates tending to increase with increasing latency. Generally, subjects showed a predominance of down-to-up confusions, suggesting that these confusions are the result of a general upward bias in judgements rather than true confusions in elevation.

The results of the distance category estimates are summarized in Figure 5 which plots the mean % of externalized judgements (judgements > 4 inches). The data indicate that all subjects externalized the majority (96 to 98%) of the stimuli in all latency conditions. (The average distance category rating was about 3.5 or about 1 to 2 ft.) The effect of latency was non-significant.

To provide some notion of the variability or localization blur of the location judgements, error angles were computed for each trial and averaged over the 5 repetitions for each stimulus condition. The error angle is the unsigned angle between each judgement vector and the vector to the target location (relative to the origin in a spherical co-ordinate system). Thus, the error angle represents the distance between two points on the surface of a sphere and does not distinguish between the

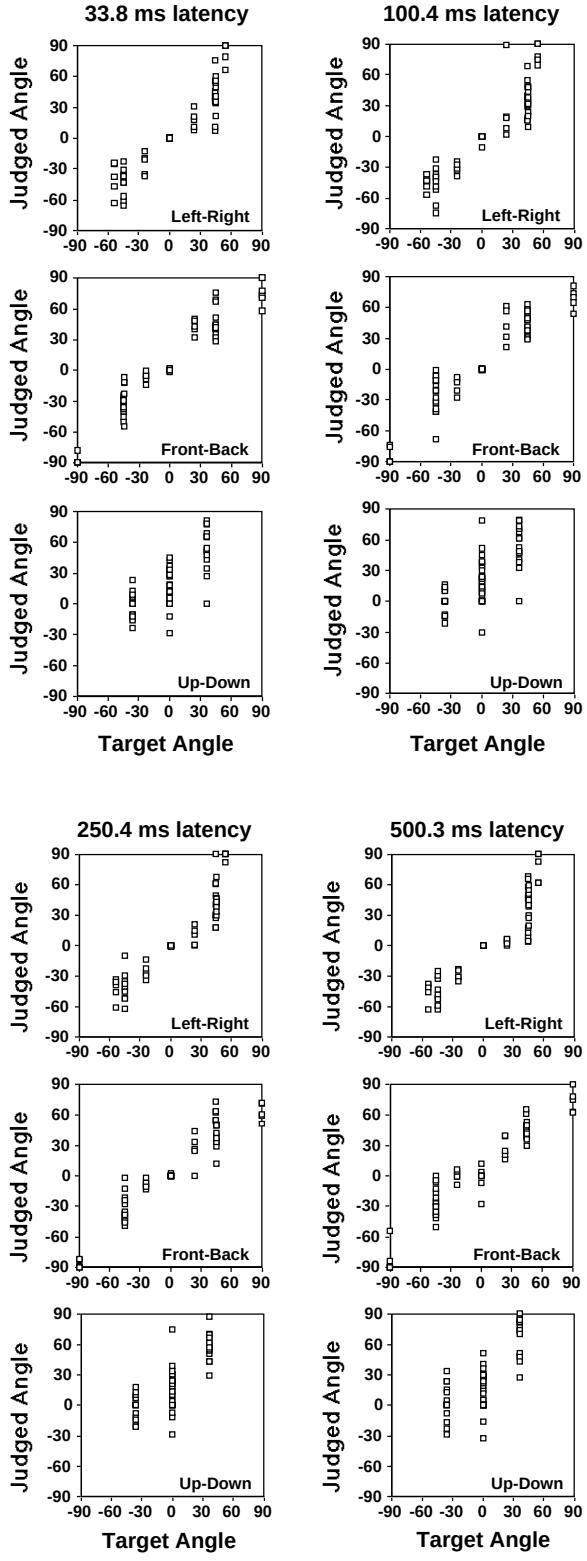

Figure 2: Triple-pole plots of raw data for subject KJ.

azimuth and elevation components of a Cartesian co-ordinate system as reported in [3]. Figure 6 summarizes average error angles as a function of latency condition. The effect of latency is significant with error angles tending to increase (26.2° to 36.3°) with the latency of the stimulus.

Figure 7 summarizes the results of the latency ratings averaged over all 5 subjects and 12 target locations as a function of latency condition. Listeners were asked to rate the perceived latency of each stimulus because it was observed during pilot studies that the subjects did not readily notice even rather large latencies. Thus, in addition to localization performance measures, it was thought useful to have some assessment of whether the subjects actually heard the latencies in the stimuli. The overall effect of latency on latency ratings was significant. The data in Figure 7 indicate a moderate ordinal relationship between actual and perceived latency. Mean latency ratings were near the "minimum" scale value for both the 33.8 and 100.4 ms latencies (1.7 to 1.8 rating), at 5.0 for 250.4 ms, and at 10.9 for 500.3 ms. Apparently, latency was not obvious to the subjects until it reached 250 ms. Even for the largest latency tested, the subjects never utilized the maximum scale value on the slider bar. Thus, even though they had training on examples of the minimum and maximum latencies, when the latency conditions were randomly intermixed during the experiment, the subjects apparently developed an internal scale with different subjective endpoints.

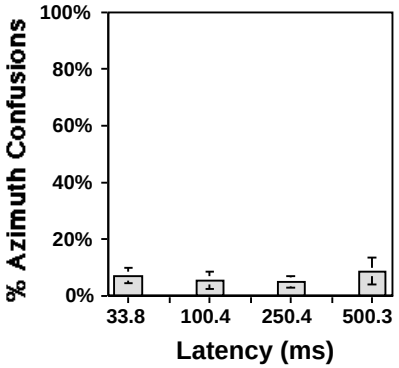

Figure 3: Mean % front-back confusions as a function of latency averaged across all applicable positions and subjects. The error bars represent standard errors for 5 subjects. The effect of latency is non-significant.

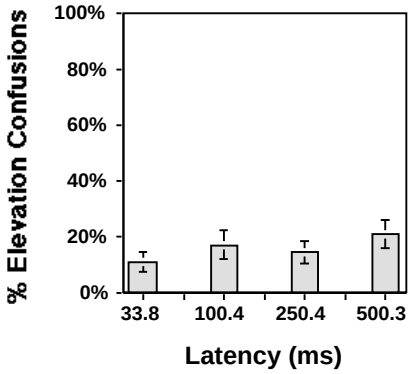

Figure 4: Mean % up-down confusions as a function of latency averaged across all applicable positions and subjects. The error bars represent standard errors for 5 subjects. The effect of latency is significant (1-way ANOVA, F (3, 12) = 4.18, p = .031).

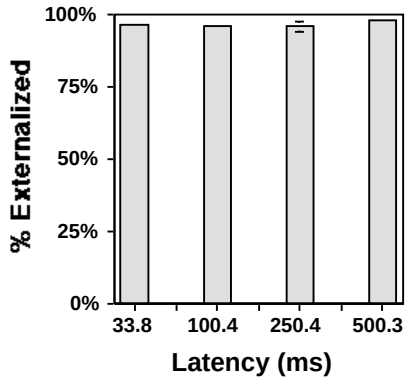

Figure 5: Mean % externalized judgements as a function of latency averaged across all positions and subjects. The error bars represent standard errors for 5 subjects. The effect of latency is non-significant.

In general, the pattern of the raw data judgement angles, confusion rates, externalization data, and error angles (Figs. 2-6) agree with previous studies that have examined dynamic localization of both real and virtual sources. A number of studies have indicated that head motion further reduces or eliminates the already low confusion rates observed for real sound sources [e.g., 12, 13]. Wightman and colleagues have also observed that, compared to static localization (without head motion), confusions are nearly eliminated when head motion is enabled for virtual sources synthesized from individualized HRTFs [14, 15].

Similarly, in studies comparing static and dynamic localization for stimuli synthesized from non-individualized HRTFs, Wenzel [4-6] demonstrated that head motion dramatically reduced confusion rates. For example, average front-back confusion rates for six subjects were reduced from 27.6% to 6.8% in [4, 6] and 22.7% to 6.5% in [5]. The advantage due to head motion also applied to stimuli in which the interaural time and level cues were purposely put into conflict, although the effect was not as large and overall confusion rates were higher. Here, azimuth confusion rates were comparable to the dynamic conditions of the previous studies but apparently, adding latency to the stimuli was not enough to disrupt the cues used in discriminating the front from rear locations.

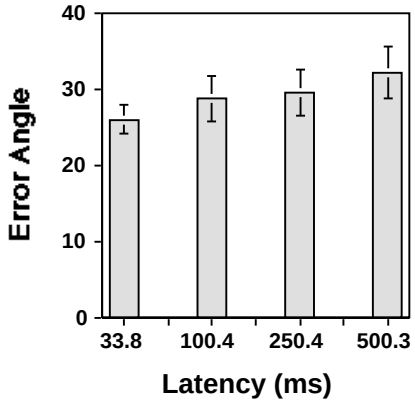

Figure 6: Mean error angles as a function of latency averaged across all positions and subjects The error bars represent standard errors for 5 subjects. The effect of latency is significant (1-way ANOVA, F (3, 12) = 5.21, p = .016).

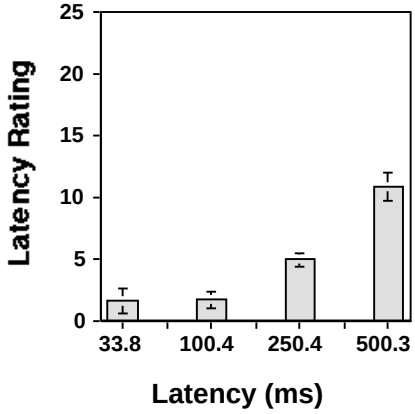

Figure 7: Mean latency ratings as a function of actual latency averaged across all positions and subjects. The error bars represent standard errors for 5 subjects. The effect of latency is significant (1-way ANOVA, F (3, 12) = 33.94, p = .0001).

Elevation confusions were also observed in the previous studies [4-6], with average rates ranging from 21 to 43% and 23 to 37% for static and dynamic conditions, respectively. Again, these confusions appeared to be the result of a general upward bias in elevation. Up-down confusion rates tended to increase somewhat with head motion, particularly for the stimuli with conflicting interaural cues. Here, elevation confusion rates were lower overall (11.3 to 21.3%) and tended to increase with latency. However, planned comparisons showed that only the difference between latencies of 33.8 and 500.3 ms was significant ($F(1,12) = 11.9$, $p < .01$). While the individualized transforms probably provided better overall cues for elevation, increasing the latency apparently reduced their utility for discriminating up vs. down locations. A possible explanation is that with virtual sounds, the horizontal plane is often perceived as tilted upward, with sources in the front appearing higher than those in the rear. Such an effect may be exacerbated with increased latency by making it difficult to track the elevation of a sound source over time.

Relatively few studies have formally examined externalization of virtual sources. Here, externalization rates were uniformly high, probably because of the superior pinna cues provided by the individualized HRTFs. With non-individualized HRTFs, Wenzel (4, 6] observed lower overall externalization rates that significantly increased when head motion was enabled (e.g., 62% vs. 75% for static vs. dynamic conditions). Begault [16] has also shown that the addition of reverberant cues can dramatically increase externalization when using non-individualized HRTFs.

The average error angles observed here are rather large (26.2° to 36.3°) but generally consistent with previous studies of localization of virtual sources [4-7] using an absolute judgement paradigm. The standard errors for azimuth and elevation measured by Sandvad and colleagues [3] were about 5 to 10°, suggesting a similarly large variability in their 16 subjects' localization data. (Since the standard error is the standard deviation of the azimuth and elevation error data divided by the square root of the number of subjects). The increase in error angles with latency observed here is also consistent with [3]. The authors concluded that latency increased azimuth standard errors (but not elevation errors) beginning with a latency of 96 ms. Here, error angles increased gradually with latency, although planned comparisons showed that only the difference between latencies of 33.8 and 500.3 ms was significant ($F(1,12) = 15.4$, $p < .01$). Thus, while the latency ratings indicated that a latency of 250.4 ms was noticeable, a latency of 500.3 ms was required to significantly affect error angles.

It is also worth noting that the large individual differences in judgement angles, confusion rates, externalization rates, and error angles typically observed in the previous studies using non-individualized HRTFs were not present in this experiment. With the individualized HRTFs used here, the subjects were remarkably consistent in their behavior.

2.2 Head Motion Data

Examination of the head motion traces recorded for each trial shows that the subjects did as instructed and actively reoriented their heads in order to localize the virtual sources. As has been observed in previous studies of the role of head motion [4-6, 13, 15, 17], the listeners primarily utilized a yawing, or left-right, motion to localize sounds. Pitching (up-down tipping) and rolling (pivoting sideways) motions were also used, but to a lesser extent. These motions are illustrated in Figure 8.

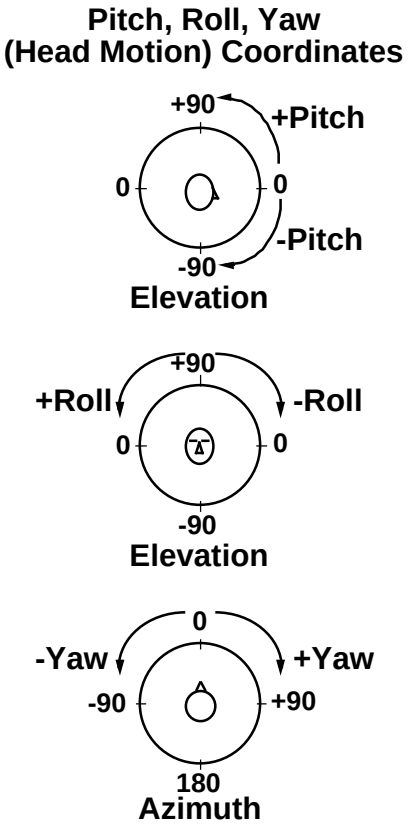

Figure 8: Illustration of the pitch, roll and yaw co-ordinates used to summarize subjects' head motions.

In order to get an idea of the overall head motion behavior of the subjects, the maximum signed deviations for yaw, pitch, and roll were computed from the head motion traces for each trial. Mean values for all subjects

and all stimulus conditions were: yaw, -86.3° and +94.9°; pitch, -14.7° and +16.8°; roll, - 5.4° and +14.1°. ANOVAs were performed for target azimuth by latency and target elevation by latency for each of the maximum signed deviations for yaw, pitch, and roll (a total of eight 2-way ANOVAs). The effect of latency on yaw was non-significant, although there was a suggestion that yaw tended to increase slightly with longer latencies. Main effects for target azimuth ($F(11,44) = 6.0$ & 6.4, $p = .0001$ & .0001, +/- yaw, respectively) and elevation ($F(2,8) = 4.5$ & 5.3, $p = .03$ & .049) were observed which suggested that yawing motions increased for targets in the rear and for higher elevations. On the other hand, increased latency resulted in significant decreases in pitching (azimuth $F(3,12) = 7.2$ & 10.8, $p = .005$ & .001; elevation $F(3,12) = 6.2$ & 8.5, $p = .009$ & .003) and rolling (azimuth $F(3,12) = 7.1$ & 5.6, $p = .005$ & .012; elevation $F(3,12) = 6.2$ & 5.1, $p = .009$ & .017) motions. In general, the main effects and interactions with target position were non-significant. The exceptions were that higher elevations ($F(2,8) = 4.6$, $p = .047$) and some azimuth locations ($F(11,44) = 3.05$, $p = .004$) resulted in increased positive pitching motions.

Overall, it appears that subjects' localization strategies were only moderately affected by latency. Yawing motions, presumably the best method for disambiguating front from rear locations, remained the primary strategy for the listeners in all conditions. Pitching and rolling motions appear to be moderately inhibited by increased latency in the stimuli. Although the individual head motion traces have not yet been examined in detail, it appears that the maximum angular velocities of some head motions (in particular, yaw) are similar to those observed in previous studies [4-6], e.g., about 175°/s for short time periods (e.g., 1200 ms)

3. CONCLUSIONS

Data from five subjects indicated that localization was generally accurate, even with a latency as great as 500.3 ms. Front-back confusions were minimal and almost all stimuli were externalized by all subjects. Both azimuth confusions and externalization were unaffected by latency. Elevation confusions and error angles increased with latency, although the increases were significant only for the largest latency tested, 500.3 ms. Mean latency ratings, on the other hand, indicated that a latency of 250.4 ms was noticeable to the subjects.

Overall, subjects' localization strategies were only moderately affected by latency. Yawing motions, presumably the best method for disambiguating front from rear locations, remained the primary strategy for the listeners in all conditions. Pitching and rolling motions appear to be moderately inhibited by increased latency in the stimuli. Although the individual head motion traces have not yet been examined in detail, it appears that the maximum angular velocities of the head motions (in particular, yaw) are similar to those observed in previous studies [4-6], e.g., about 175°/s.

Together with the results of previous studies (4-6, 14, 15], these data support the notion that head motion can provide robust and powerful cues for localization of virtual sounds. These dynamic cues apparently mitigate the impact of many disrupting factors in the stimulus, including the use of non-individualized HRTFs, conflicting interaural cues, and increased latency.

The fact that accuracy was generally comparable for the shortest and longest latencies tested here suggests that listeners are largely able to ignore latency during active localization. Apparently, this is possible even though latencies of this magnitude produce an obvious spatial "slewing" of the sound source such that it is no longer stabilized in space as the head is reoriented. It may be that the localization task per se is not the most sensitive test of the impact of latency in a virtual audio system. Other tasks that are more directly dependent on temporal synchrony, such as tracking an auditory-visual virtual object, may be much more sensitive to latency effects.

ACKNOWLEDGEMENTS

Work supported by NASA and by the Navy (SPAWARSYSCEN, San Diego). Thanks to Mark Anderson, Alex Lee, and Joel Miller for technical assistance.

REFERENCES

[1] Adelstein, B. D., Johnston, E. R., and Ellis, S. R. 1996. Dynamic response of electromagnetic spatial displacement trackers. Presence: Teleoperators & Virtual Environments, 5, pp. 302-318.

[2] Jacoby, R. H., Adelstein, B. D., and Ellis, S. R. 1996. Improved temporal response in virtual environments through system hardware and software. Proceedings of SPIE: Stereoscopic Displays & Virtual Reality Systems," San Jose, CA.

[3] Sandvad, J. 1996. Dynamic aspects of auditory virtual environments. 100th Convention of the Audio Engineering Society, Copenhagen, preprint 4226.

[4] Wenzel, E. M. 1996. What perception implies about implementation of interactive virtual acoustic environments." 101st Convention of the

Audio Engineering Society, Nov. 8-11, Los Angeles, CA, preprint 4353.

[5] Wenzel, E. M. 1996. Effectiveness of interaural delays alone as cues during dynamic sound localization. Journal of the Acoustical Society of America, 100, p. 2608.

[6] Wenzel, E. M. 1995. The relative contribution of interaural time and magnitude cues to dynamic sound localization. Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio & Acoustics, New Paltz, NY.

[7] Wenzel, E. M., Arruda, M., Kistler, D. J. & Wightman, F. L. 1993. Localization using non-individualized head-related transfer functions. Journal of the Acoustical Society of America, 94, pp. 111-123.

[8] Wenzel, E. M. 1998. The impact of system latency on dynamic performance in virtual acoustic environments. Proceedings of the 16th International Congress on Acoustics and 135th Meeting of the Acoustical Society of America, Seattle, WA, pp. 2405-2406.

[9] Perrott, D. and Musicant, A. 1977. Minimum audible movement angle as a function of signal frequency and the velocity of the source. Journal of the Acoustical Society of America, 62, pp. 1463-1466.

[10] Wenzel, E. M., and Foster, S. H. 1993. Perceptual consequences of interpolating head-related transfer functions during spatial synthesis, Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio & Acoustics, New Paltz, NY.

[11] Kistler, D. J. and Wightman, F. L. 1991. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. Journal of the Acoustical Society of America, 91, pp. 1637-1647.

[12] Wallach, H. 1940. The role of head movements and vestibular and visual cues in sound localization, Journal of Experimental Psychology, 27, pp. 339-368.

[13] Thurlow, W. R., and Runge, P. S. 1967. Effect of induced head movements on localization of direction of sounds. Journal of the Acoustical Society of America, 42, pp. 480-488.

[14] Wightman, F. L. and Kistler, D. J. 1997. Factors Affecting the Relative Salience of Sound Localization Cues. In R. Gilkey & T. Anderson (Eds.), pp. 1-23, Lawrence Erlbaum, Hillsdale, NJ.

[15] Wightman, F. L., Kistler, D. J., and Anderson, K. 1994. Reassessment of the role of head movements in human sound localization. Journal of the Acoustical Society of America, 95, p. 3003-3004.

[16] Begault, D. R. 1992. Perceptual effects of synthetic reverberation on three-dimensional audio systems. Journal of the Audio Engineering Society, 40, pp. 895-904.

[17] Thurlow, W. R., Mangels, J. W., and Runge, P. S. 1967. Head movements during sound localization. Journal of the Acoustical Society of America, 42, pp. 489-493.