Proceedings, 5th International Conference on Multimodal Interfaces, ACM, Vancouver, Canada, 2003, pp. 73-76.

Sensitivity to Haptic-Audio Asynchrony

Bernard D. Adelstein, Durand R. Begault, Mark R. Anderson[1], Elizabeth M. Wenzel
NASA Ames Research Center, [1]QSS Group Inc.
Moffett Field, CA 94035-1000
E-mail: Bernard.D.Adelstein@nasa.gov

ABSTRACT

The natural role of sound in actions involving mechanical impact and vibration suggests the use of auditory display as an augmentation to virtual haptic interfaces. In order to budget available computational resources for sound simulation, the perceptually tolerable asynchrony between paired haptic-auditory sensations must be known. This paper describes a psychophysical study of detectable time delay between a voluntary hammer tap and its auditory consequence (a percussive sound of either 1, 50, or 200 ms duration). The results show Just Noticeable Differences (JNDs) for temporal asynchrony of 24 ms with insignificant response bias. The invariance of JND and response bias as a function of sound duration in this experiment indicates that observers cued on the initial attack of the auditory stimuli.

Categories and Subject Descriptors

H.5.2 [**Information Systems**]: User Interfaces – *haptic I/O, auditory (non-speech) feedback.*

General Terms

Measurement, Performance, Experimentation, Human Factors.

Keywords

Haptic, audio, time delay, latency, cross-modal asynchrony, multi-modal interfaces, virtual environments.

1. INTRODUCTION

Sound is a natural feature of the mechanical impacts and vibrations that we experience in everyday manual interactions with our real surroundings. Likewise, similar interactions with high performance computer haptic interfaces can produce comparable acoustic signals, simply as a byproduct of the interface's structural response to simulation-driven forces. Research on multimodal haptic-audio perceptual performance points to the benefit of deliberately adding sound to haptic interfaces [3], [7], [8], [9]. Moreover, effort has recently been directed specifically toward modeling and simulating sounds from the physics that arise in response to real and virtual mechanical interactions [2], [5].

Because sound can be a normal consequence of both real and simulated haptic events, the relative low-cost and ubiquity of computer audio suggests the value of employing sound generation technologies for augmentation of haptic displays. However, in order to effectively implement real-time acoustic modeling methods and allocate available computational resources, knowledge of perceptually tolerable haptic-audio asynchrony is needed.

It has been postulated that the more properties shared between two modalities, the stronger will be the observer's "unity assumption" that information from different sensory channels can be attributed to the same distal event or object [14]. Among these properties are spatial location, motion, and temporal patterning or rate [15], all of which would be impacted by temporal (a)synchrony in a multi-channel display system.

Temporal features that have been examined in experimental studies related to uni- and multi-modal synchrony include perceived temporal order (e.g., [1]), simultaneity or fusion (e.g., [6]), and the empty interval (gap) between successive stimuli (e.g., [13]). While the temporal aspects of unimodal visual, tactual, and aural perception have been widely reported in the literature, cross-modal combinations have focused almost exclusively on the visual-auditory combination (See [10] for a concise summary). With the exception of an investigation by Levitin et al. [10], the haptic-auditory combination appears not to have been explored.

Levitin et al. [10], in reviewing prior work, noted excessively high temporal thresholds for visual-haptic asynchrony and proposed that these might have been due to perceptual experiment tasks that were "strange or without ecological validity." In this vein, a self-generated haptic event such as a hammer strike, for which the observer's attention may be primed to perceive the naturally expected auditory consequence [12], might lead to a different result than an exogenously applied experimental vibrotactile stimulus (e.g., [6], [13]). Likewise, we propose that a realistic percussive sound response to a hammer strike might be more appropriate than the brief clicks used in many auditory synchrony studies (e.g., [6]).

The objective of this study is to quantify the perception of asynchrony between successive haptic and auditory events. As opposed to the perceived simultaneity measured by Levitin et al. [10], here we examine observers' perceptual sensitivity to differences in the gap between haptic and auditory events (i.e., auditory lag). Additionally, we consider the effect of the time duration (and therefore one aspect of the realism) of the auditory stimulus to understand which portion of that signal defines the relevant portion of the temporal cue.

2. METHODS AND MATERIALS

2.1 Apparatus

Participants were seated alone in a darkened sound isolation booth facing a 17-inch diagonal CRT display that was connected

ICMI'03, November 5–7, 2003, Vancouver, British Columbia, Canada.
Copyright 2003 ACM 1-58113-621-8/03/0011…$5.00.

to the computer controlling the experiment. Participant state was monitored via a closed circuit infrared camera system. Participants grasped the metal handle of a rubber-tipped 7.5 inch (19 cm) Taylor percussion (reflex test) hammer in their dominant hand and used it to tap lightly on the side of a brick that was affixed to the table top in front of the CRT. Participants' arm posture was not constrained, but the lateral strike of the brick nominally entailed less than 10 cm of travel at the hammer tip. In their other hand, participants held a two-button mouse-like device that they used to input their response to the computer and control their progress through the experiment. Subjects received diotic aural stimuli via circumaural headphones (Sennheiser HD545).

2.2 Haptic Stimulus

The Taylor percussion hammer was instrumented with a single-axis accelerometer (Analog Devices ADXL 150EM1). The accelerometer signal resulting from the hammer strike of the block was conditioned via custom circuitry to produce a single 600 µs TTL pulse at the initial acceleration rise that then locked out any other hammer input for the subsequent 500 ms. The TTL pulse was then converted to a MIDI signal (MIDI Solutions Footswitch Controller) for introduction into the audio portion of the equipment.

2.3 Auditory Stimuli

The three different auditory stimuli in the experiment were composed of digitally synthesized signals modeled on the acoustical characteristics of a struck wooden idiophone with a hollow cavity. All three signals had the same initial 1 ms strike signature, but their decay envelopes were varied to produce sound stimuli that had overall durations of either 1, 50 or 200 ms. The resulting sound characteristics of the stimuli were similar to a struck hollow wood block for the 50 and 200 ms version, and more akin to an impulsive click for the 1 ms version.

The auditory stimuli were rendered in the experiment by a digital sampler (Roland S-760). Time delay of the auditory stimuli with respect to the hammer tap was introduced by a digital delay unit (Sony DPS-D7) that received the output of the digital sampler. The delay unit's output was mixed with pink noise and then delivered to the headphones via a headphone amplifier. The pink noise was presented continuously throughout the experimental trials in order to mask the sound produced by the actual hammer-brick impact.

Output levels at the headphones were measured using a calibrated binaural microphone (Neumann KU-100) and real-time analyzer (Agilent 35670A). The three pre-recorded experiment stimuli were played back at a peak A-weighted level of 96 dB. The pink noise was set to an A-weighted rms level of 57 dB to mask the sound generated by the physical hammer strike (A-weighted peak level of 52 dB) and other ambient sounds.

2.4 Experiment Latency Control

The fixed latency overhead of the experiment system was ~7 ms (measured range: 6.6-7.8 ms). All additional delays, set by the delay unit, as well as selection of the acoustic stimuli, were controlled by the experiment computer via a USB-to-MIDI interface (Roland UA-100).

2.5 Participants

Twelve participants (10 M and 2 F; age range 18-33 yrs, mean 25.2 yrs) were selected from among laboratory colleagues and their associates. All participants had normal hearing and were free of neuromotor impairment. All were naïve to the details of the experiment.

2.6 Procedure

Participants tapped the brick with the rubber tip of the hammer and, in response, heard one of the three acoustic stimuli stored by the sampler. They tapped a second time and heard the same stimulus again. In one interval, the sound followed the hammer tap by the system's baseline delay of 7 ms (i.e., the reference level). In the other, an experimentally controlled amount of latency between 0 and 256 ms (i.e., the probe) was added to the baseline by the delay unit. The order of presentation for the probe and reference intervals was randomized for each stimulus pair. Participants employed a two-alternative forced-choice protocol to judge which of the paired intervals had less delay between its tap and the consequent sound (i.e., which of the two intervals was the reference) and then entered their choice by the appropriate button press on the input response device. Participants were required to keep their eyes closed while tapping the hammer, but could open their eyes to be reminded by the CRT display as to the meaning of the mouse buttons.

The amount of added latency was controlled according to an adaptive two-down, one-up staircase algorithm [11]. Two consecutive correct responses indicating which interval had less delay decreased the amounted of added latency; one incorrect response caused the latency setting to increase. The staircases were made adaptive by halving their initial 64 ms step size at each reversal until a step size of 4 ms was reached. Staircases concluded following a total of 10 direction reversals. Each staircase employed only one of the three (1, 50, or 200 ms) acoustic stimulus durations.

Eight staircase runs were completed for each acoustic stimulus duration. Four of these started at the minimum added latency (0 ms) and four at the maximum (256 ms). Staircases were run two at a time in interleaved pairs to prevent response biases that would otherwise be caused by the participants' ability to track their progress between successive stimuli. Staircase conditions for each of these pairs were selected at random without replacement from 24 possibilities (3 sound durations X 2 starting levels X 4 repetitions). The entire experiment, including breaks between staircase runs, could be completed by a participant within a single 1-hour sitting.

3. RESULTS

Rather than simply examining the 70.7% thresholds for haptic-to-audio latency (i.e., asynchrony) that could be obtained directly from the average latency of the final reversals for each staircase run [11], detection rates were analyzed in order to estimate each subject's psychometric function at each sound duration condition.

Detection rates were accumulated from the proportion of correct responses at each of the latency levels encountered during the eight staircases for an individual sound duration. A total of at least eight observations at a particular latency level were required for that latency's inclusion in the analysis. An example of detection rates for one participant at a single sound duration is shown in Figure 1. Error bars at each sample point in the plot correspond to the binomially distributed standard error of proportion. A Probit procedure [4] was then used to fit a cumulative Gaussian distribution that models the psychometric function for detection of hammer tap to auditory stimulus latency.

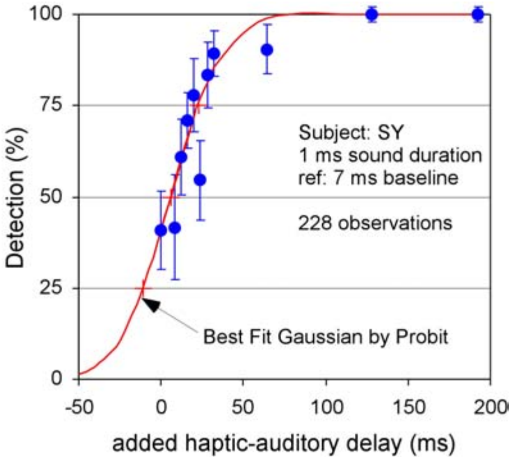

Figure 1. Example of detection rates (% correct responses) and fitted Gaussian psychometric function for one subject at 1 ms sound duration.

Figure 2 illustrates the two important psychophysical parameters of a general psychometric function. The Point of Subjective Equality (PSE) is defined as the stimulus level (in this case the amount of haptic-to-auditory delay) that produces a detection rate corresponding to equiprobable random guessing (50% for this particular experiment design). The difference between the PSE and the stimulus reference level (in this case, the 7 ms system baseline) represents judgment bias that may be due to experiment method, stimulus presentation, individual observer preferences, etc. The change in stimulus between the PSE and the 75% threshold is defined as the Just Noticeable Difference (JND). For this Gaussian psychometric function, the JND is directly proportional to the density's variance; the PSE is the density function's mean.

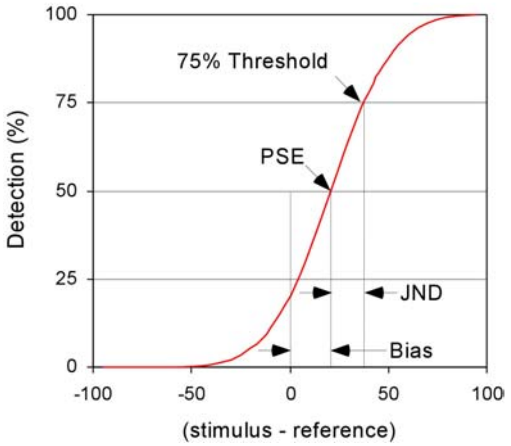

Figure 2. Characteristics of a typical psychometric function.

Employing these definitions, JNDs and PSEs were computed individually at each of the three sound durations for each of the 12 participants. The JNDs and PSEs for individual participants along with means and standard errors are plotted in Figures 3 and 4. Individual JNDs ranged between 5 and 70 ms; PSEs ranged between –25 and 44 ms. Analyses of variance (ANOVAs) did not reveal a significant dependence of either JND ($F_{2,22} = 2\times10^{-5}$; $p < 1.0$) or PSE ($F_{2,22} = 1.170$; $p < .33$) on sound duration. When pooled across all subjects and sound durations, the average JND for the latency between the haptic event (i.e., hammer tap) and the resultant auditory signal was 24.1 ± 2.2 ms (mean ± std error). From the pooled PSEs, the average response bias of 4.8 ± 2.1 ms was not significantly different from zero ($F_{1,11} = 2.294$; $p < .12$) as expected for judgments with randomized probe-reference order.

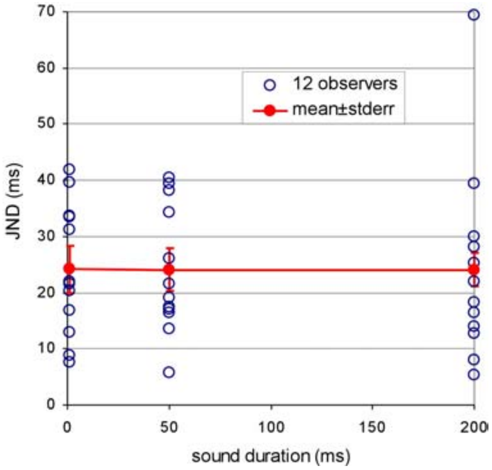

Figure 3. JND for haptic-to-audio delay (asynchrony).

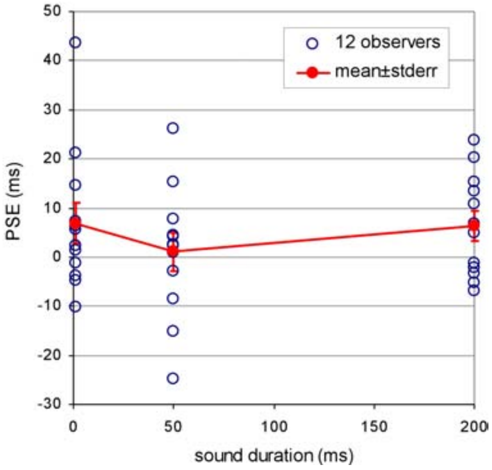

Figure 4. PSE for haptic-to-audio delay (asynchrony).

4. DISCUSSION

The average results from the 12 participants exhibited uniform JNDs and low biases (PSEs) for all three sound durations. The 75% thresholds, obtained by adding the JNDs and biases, averaged between 25 and 31 ms for the three durations. For two of the subjects, the individual 75% thresholds were notably higher for all three sound durations.[1] If the data from these two subjects are removed as indicated in Figure 5, the average 75% thresholds drop to between 18 and 25 ms. These values are somewhat lower than the 42 ms threshold reported by Levitin et al. [10] for the comparable stimulus order (i.e., haptic event first). However, a more complete comparison between the studies cannot be made without the breakout of JND and bias from Levitin et al.'s data.

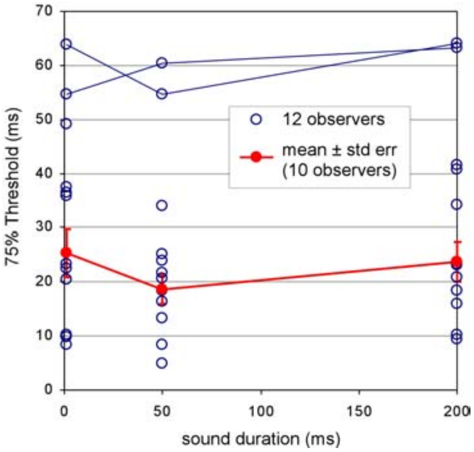

Figure 5. 75% thresholds for 12 subjects. The upper two lines show the data for the two subjects with highest thresholds. The filled circles and lower line indicate the means for the 10 remaining subjects.

The lack of significant variation in JND and PSE with sound duration indicates that participants relied on the sounds' onset as their cue, rather than other features occurring later in the signal, to mark the end of the interval initiated by the hammer strike.

While the average JNDs measured are relatively small, one participant in particular had 5-8 ms JNDs with 75% thresholds of only 8-10 ms. As a practical consideration for the design of multimodal haptic-auditory displays and auditory enhancements to haptic interfaces, the data for this individual observer suggest synchronization requirements that may undercut those for unimodal tactile temporal separation (10-30 ms) and begin to approach those for auditory fusion (1-2 ms) [6].

5. ACKNOWLEDGMENTS

This work was supported by a NASA-Ames Director's Discretionary Fund Award and the NASA Intelligent Systems Human Centered Computing Program. Mitch Clapp and Bryan McClain assisted in monitoring the experiment participants.

6. REFERENCES

[1] Craig, J.C., and Baihua, X. Temporal order and tactile patterns. Perception & Psychophysics, 74, 1990, 22-34.

[2] Doel, van den, K., and Pai, D. K. The sounds of physical shape. Presence, 7(4), 1998, 382-395.

[3] DiFranco, D. E., Beauregard, G. L., and Srinivasan, M. A. The effect of auditory cues on the haptic perception of stiffness in virtual environments. Proceedings, ASME Dyn. Syst. & Control. (Dallas, TX, 1997). Vol-DSC-61, 17-22.

[4] Finney, D.J. Probit Analysis, 2nd Edition. Cambridge University Press, Cambridge, UK, 1962.

[5] Fouad, H., Ballas, J.A., and Brock, D. An extensible toolkit for creating virtual sonic environments, Proceedings, International Conference on Auditory Displays, ICAD 2000 (Atlanta, GA, 2000).

[6] Gescheider, G.A. Auditory and cutaneous temporal resolution of successive brief stimuli. J. Exp. Psych. 75(4), 1967, 570-572.

[7] Hendrix, C.M., Cheng, P., and Durfee, W.K. Relative influence of sensory cues in a multi-modal virtual; environment. Proceedings, ASME Dyn. Syst. & Control. (Nashville, TN, 1999). Vol-DSC-67, 59-64.

[8] Lederman, S.J., Klatzky, R.L., Morgan, T., and Hamilton, C. Integrated multimodal information about surface texture via a probe: relative contributions of haptic and touch-produced sound sources. Proceedings, Haptic Interfaces for Virtual Environment and Teleoperator Systems (Orlando, FL, 2002), 97-104.

[9] Lederman, S.J., Martin, A.M., Tong, C., and Klatzky, R.L. Relative performance using haptic and/or touch-produced auditory cues in a remote absolute texture identification task. Proceedings, Haptic Interfaces for Virtual Environment and Teleoperator Systems (Los Angeles, CA, 2003), 151-158.

[10] Levitin, D.J., MacLean, K., Mathews, M., Chu L, and Jensen, E. The perception of cross-modal simultaneity. Proceedings, Computing Anticipatory Systems (Liege, Belgium, 1999). AIP Conf. Proc. 517, 2000, 323-329.

[11] Levitt, H. Transformed up-down methods in psychoacoustics. J. Acoustical Soc. of America, 49(2), 1970, 467-477.

[12] Spence, C., and Driver, J. Cross-modal links in attention between audition, vision and touch: implications for interface design. Internat. J. Cognitive Ergonomics. 1(4), 1997, 351-373.

[13] Van Doren, C.L., Gescheider, G.A., and Verillo, R.T. Vibrotactile temporal gap detection as a function of age. J. Acoustical Soc. of America, 87(5), 1990, 2201-2206.

[14] Welch, R.B. Meaning, attention, and the unity assumption in the intersensory bias of spatial and temporal perceptions. In G. Aschersleben, T. Bachmann, and J. Musseler (Eds.), Cognitive Contributions to the Perception of Spatial and Temporal Events. Elsevier, Amsterdam, 1999, 371-387.

[15] Welch, R.B, and Warren, D.H. Immediate perceptual response to intersensory discrepancy. Psych. Bulletin, 88, 1980, 638-667.

[1] The same two subjects also stood out in an unreported preliminary study employing single interval (i.e., probe only, no reference) judgments because they were unable to converge their staircases within the 256 ms upper bound for that experiment.