PERCEIVED SYNCHRONY IN A BIMODAL DISPLAY: OPTIMAL INTERMODAL DELAY FOR COORDINATED AUDITORY AND HAPTIC REPRODUCTION

William L. Martens and Wieslaw Woszczyk
McGill University, Faculty of Music
Montreal, QC, H3A 1B4 Canada

[wlm, wieslaw]@music.mcgill.ca

ABSTRACT

The purpose of this study was to determine the range of optimum intermodal delay values for coordinated auditory and haptic reproduction of brief impact events. Indirect psychophysical methods were used to find the intermodal delay that would be most likely to generate the response of perceived synchrony between acoustic and structural vibration components of those events. A recording of a representative impact sound was processed to create bimodal stimuli with varying amounts of intermodal delay between the bimodally reproduced components. The haptic component of the bimodal stimulus was whole-body vibration presented via a platform on which the observer was seated. Using four actuators moving together, users could be displaced linearly upwards or downwards, with a very quick response and with considerable force (the feedback-corrected linear system frequency response was flat to 50 Hz). The auditory component of the bimodal stimulus was presented in an immersive virtual acoustic environment via a multichannel reproduction of simulated indirect sound. The direct sound component matched to the haptic stimulus was reproduced via a frontally-located pair of loudspeakers that included a low-frequency driver capable of reproducing sound with a linear frequency response ranging from 25 to 300 Hz and a high-frequency driver extending well above 20kHz. The intermodal delay was adaptively varied using a two-alternative, forced-choice (2AFC) procedure to track the point of subjective simultaneity (PSS) based upon temporal order judgments with the following response options: 1) haptic sensation seemed to precede auditory sensation; and 2) haptic sensation seemed to follow auditory sensation. Then, in order to avoid sequential response biases in the tracking procedure, a constant stimulus method was used to determine directly the optimal range of intermodal delay values for producing observer responses of intermodal synchrony, with two response options: haptic sensation either seemed to precede or to follow auditory sensation.

1. INTRODUCTION

Multimodal display technology that is used to reproduce a remotely captured and/or recorded event is most effective when the transmitted and reproduced stimulation is synchronized with minimal intermodal delay [1, 2]. Such coordinated display of visual, auditory, tactile, and kinesthetic information can produce for an observer a strong sense of presence in a reproduced environment when asynchrony is below threshold for human detection [3], but even when asynchrony is detectable, there is useful variation in human experience within the tolerable range of asynchrony. The research reported here focused upon asynchrony between display components in just two modalities, the auditory and the haptic, in an attempt to quantify their bimodal integration in isolation from other display modalities. In particular, it was the interaction between acoustic and structural vibration components of selected acoustic events that was of interest here, since first hand experience with bimodal display of these events suggested that physical synchrony between auditory and haptic reproduction was not necessarily required to produce a subjective experience of simultaneity.

One hypothesis under test was that the optimum intermodal delay required for coordinated auditory and haptic reproduction might have the haptic component arriving slightly before the auditory component. Indeed, there is a physical basis for such a hypothesis regarding the auditory and haptic components of actual impact events. For example, when a heavy object is dropped on the floor of the space in which an observer is located, the structure-borne component of the impact event moves more quickly through the floor than does the air-borne component; thus the auditory stimulus is naturally preceded by the haptic stimulus that could be felt in an observer's feet, or via the "seat of the pants." It had been observed informally that some relative delay in the reproduction of the auditory component of a recorded

impact sound produced a more realistic perception of the event. The current study was designed to determine quantitatively just how great the intermodal delay should be for optimizing reproduction realism in this bimodal display, with an emphasis upon producing the optimal sense of presence for the observer. The result of the study will naturally be specific to the particular bimodal display technology utilized in the study, but can be generalized to other such displays via human-centered measurement of the bimodal stimulus (meaning that the measurement should attempt to capture the proximal stimuli near the sense receptors of an observer, rather than the distal stimuli associated with the actual or reproduced event). The utilized bimodal display, described in more detail in the authors' previous paper [4], presented haptic information via a chair upon which an observer was seated, and that chair was mounted on a platform that could be moved quickly with considerable force. The auditory information was displayed via a loudspeaker array that created a virtual acoustic environment with high spatial fidelity.

This study employed classic psychological measurement methods that have been used in intermodal timing studies for many years [5], but the design of this study was also informed by recent methodological review [6,7]. Here, the optimal intermodal delay for human responses of simultaneity was first estimated by the method of constant response, using a two-alternative, forced-choice (2AFC) procedure for tracking the desired response probability point. Then, in order to avoid sequential response biases in the tracking procedure, the method of constant stimuli was used to generate a "simultaneity" response distribution from which the optimal intermodal delay range could be inferred. Results using these methods provide a firm basis for further studies of the optimum intermodal delay for coordinated auditory and haptic reproduction of the selected type of impact sound events.

2. METHODS

This section describes both the stimulus generation methods and the experimental methods used in the experimental tests. First, an overview of the employed auditory and haptic display systems is presented, along with a description of the selected bimodal stimuli.

2.1. Auditory Display System

The auditory display system positioned a single virtual sound source into a virtual acoustic environment via a spherical loudspeaker array consisting of 5 low-frequency drivers (ranging from 25 to 400 Hz) and 32 high-frequency drivers (ranging from 300 to well over 20,000 Hz). The low-frequency drivers were "Mini-Mammoth" subwoofers manufactured by the Quebec-based company D-BOX Technology, and these were placed at standard locations for the 5 main speakers in surround sound reproduction (the speaker angles in degrees relative to the median plane were –110, -30, 0, 30, and 110). The high-frequency drivers were dipole radiating, full range transducers featuring the "Planar Focus Technology" of Level 9 Sound Designs, Inc. of British Columbia, and these 32 loudspeaker panels were placed pairwise in 16 locations lying on the surface of an imaginary sphere of 2-meter radius. Besides 4 locations at extreme high elevation, the spatial organization of the high-frequency drivers was defined by 2 planes at elevation angles of –15 and 25 degrees relative to the horizontal plane. Within each plane, 6 speakers were placed at azimuth angles of –110, -30, 0, 30, 110, and 180 degrees).

The bimodal stimuli were selected the most representative from a number of transient sound sources that were recorded in a rectangular shaped music hall (Redpath Concert Hall) at McGill University using a Schoeps CCM 21H wide-cardioid microphone pointing at the stage. The most satisfying recording was that made by dropping a stack of 3 telephone books from above the stage onto the floor, at a distance of 2 meters from the microphone.

2.2. Haptic Display System

Just as the auditory display system could provide extremely high spatial fidelity for the reproduced sound field, the haptic display system was capable of generating multidimensional vibration stimulation, providing users with motion having three Degrees of Freedom (3DOF) in a home theater setting [8]. For the current study, however, only motion along the vertical axis was enabled, and so for the displayed virtual sound source, a vibrational stimulus was presented via a translation along a single vertical axis. The motion was generated by the Odyssée™ system, a commercially available motion platform manufactured by D-BOX Technology. The Odyssée™ system uses four coordinated actuators to enable control over pitch and roll of the platform on which the user's chair was fixed. When all four actuators move together, users can be displaced linearly upwards of downwards, with a very quick response and with considerable force (the feedback-corrected linear system frequency response is flat to 50 Hz). The haptic stimulus was generated by gating to a 30 ms duration the initial portion of the audio signal (which was a highly reverberant recording a the impact of a phone book on a wooden stage), and then applying a lowpass filter with a cutoff frequency of 50 Hz. The RMS value of the motion platform acceleration was adjusted to provide a realistic experience of vibration appropriate to the 82 dB(A) sound level of the auditory stimulus. The selected vertical

acceleration RMS value was 1.3 m/sec^{2,} measured at the observer's foot position (using a B&K Type 4500 accelerometer and a Type 2239B controller). The platform exhibited only negligible acceleration along other axes of motion. Of course, the haptic display produced an air-borne sound stimulus as well as a structure-borne vibratory stimulus, but this "cross-modal crosstalk" stimulus measured roughly 40 dB lower than the sound stimulus associated with the auditory display component, and as such should probably be regarded as having a negligible influence on intermodal delay sensitivity.

2.3. Listening Experiment

The method of constant response was utilized to estimate the point of subjective simultaneity (PSS) with regard to the auditory and haptic stimuli. This method required the listener to complete a two-alternative, forced choice staircase tracking the intermodal delay at which there was equal probability of choosing one component as earlier than the other. The haptic signal delay was incremented by 10 ms if the auditory sensation was chosen as "Earlier" and decremented by this amount if the auditory sensation was chosen as "Later." Five staircase turnarounds were completed before the set of trials was terminated, each staircase beginning with a randomly selected delay value ranging from -40 to +40 ms. This procedure was completed for each of two observers, and the median intermodal delay was used as a guide in choosing the delay values employed in stimulus generation for a subsequent listening session employing the method of constant stimuli. In contrast to the session using the previously described method of constant response, the latter method did not involve adaptively adjusted delay values for stimulus presentation, but rather presented 20 trials containing stimuli at 7 haptic signal delay values ranging from -30 to +30 ms. Furthermore, in this latter session, observers were required only to indicate whether the auditory and haptic sensations were experienced as simultaneous or not.

3. RESULTS

For both observers completing 5 adaptive staircases tracking the PSS, the median haptic signal delay value was 0 ms. In effect, there was no evidence that the haptic stimulus should be slightly delayed relative to the auditory stimulus, at least for the temporal order judgments providing the basis for the 2AFC procedure employed here. The more telling question for the range of intermodal delay values presented, was for which values the auditory and haptic sensations were experienced as simultaneous and for which they were not. Figure 1 shows the proportion of simultaneity responses of two observers presented with bimodal stimuli at 7 intermodal delay values. The abscissa in each graphs shows the delay of the vibration signal (the haptic stimulus) relative to the audio signal. Note that negative vibration delay values indicate intermodal delays at which the arrival of the vibration signal preceded the arrival of the audio signal (noted in the figure as "Vibration Leads Audio").

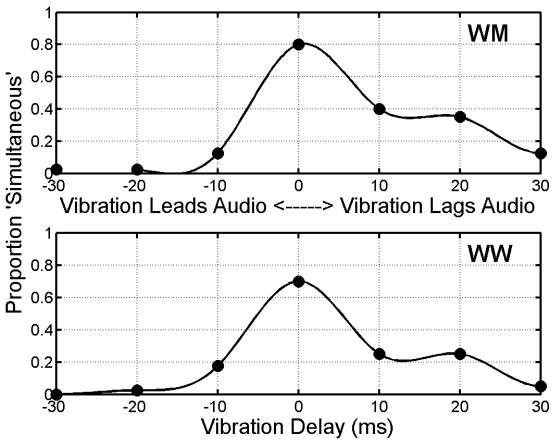

Figure 1. Constant stimulus results for two observers (upper panel shows results for observer **WM**, and lower panel shows results for observer **WW**). The graph plots as a function of the relative delay of the vibration signal (the haptic stimulus), the proportion of trials on which the auditory and haptic sensations were reported as "Simultaneous." These proportions were the result of exactly 40 presentations of the bimodal stimuli at each of 7 intermodal delay values.

When the audio signal was presented simultaneously with the vibration signal, the response of "Simultaneous" was most frequently for both observers. But neither observer showed perfect detection of the trials on which there was no intermodal delay. Such a result reveals that both observers employed a fairly conservative response criterion for reporting simultaneity, since the peak proportions were only 0.8 and 0.7 for observers WM and WW, respectively. Nonetheless, the trials on which vibration lagged behind audio by 20 ms still gave rise to responses of simultaneity. In contrast, a vibration signal leading the arrival of an audio signal was shown to give practically no responses of simultaneity, despite the observation that in reality the structure-borne component of an actual impact event must arrive earlier than the air-borne component (since the impact wavefront moves more quickly through the floor than it does through the air). This "skewness" in the distribution of the simultaneity responses for each observer can be quantified as the third central moment divided by the cube of the standard deviation, which would be equal to zero if the distribution were symmetrical. Using the skewness routine found in the Matlab software [9], the calculated skewness values were 1.03 and 1.29 for the simultaneity responses of observers WM and WW, respectively.

4. CONCLUSIONS

For bimodal display systems in which realistic reproduction of impact events is desired, it has been confirmed, for a representative impact sound, that the structure-borne component of the bimodal stimulus should not precede the air-borne component (at least as revealed by two psychophysical tasks employing the 10-ms resolution on intermodal delay values employed here). In contrast, vibration delay values generally between 10 and 20 ms were found to be tolerable as an intermodal asynchrony for coordinated reproduction of auditory and haptic display components of a reproduced impact event (cf. [10]). Of course, these results are limited to whole-body haptic stimulation, and should not be expected to apply to local haptic stimulation, such as a vibratory stimulus presented to a finger tip (as in [11]). Furthermore, the results support quantitatively the conclusions drawn from less formal experience, such as that which motivated this investigation. These results add a newly validated technique to the sets of tools that contribute to a set of guidelines for enhancing an observer's sense of presence in virtual acoustic environments, recently summarized by the authors [12].

5. ACKNOWLEDGMENTS

The authors would like to thank Dr. Bruno Paillard of D-BOX Technologies for many helpful discussions during the initial stages of this project. This research was supported by Valorisation-Recherche Québec

6. REFERENCES

[1] Barfield, W., Hendrix, C., Bjorneseth, O., Kaczmarek, K. A., and Lotens, W. "Comparison of Human Sensory Capabilities with Technical Specifications of Virtual Environment Equipment," Presence: Teleoperators and Virtual Environments, 4(4), 1995.

[2] Miner, N., & Caudell, T. "Computational Requirements and Synchronization Issues for Virtual Acoustic Displays," Presence: Teleoperators and Virtual Environments, 7 (4), 396-409.

[3] Martens, W. L., "Human-centered design of spatial media display systems", Proceedings of the 1st International Workshop on Spatial Media, University of Aizu, Aizu-Wakamatsu, Japan, 1999.

[4] Woszczyk, W., & Martens, W. L., "Intermodal delay required for perceived synchrony between acoustic and structural vibratory events," To appear in: Proceedings of the 11th International Congress on Sound and Vibration, St. Petersburg, Russia, July, 2004.

[5] Hirsh, I. J. and Sherrick, C. E., "Perceived order in different sense modalities," J. Exp. Psychol., 62, 423-432, 1961.

[6] Spence, C, Baddeley, R., Zampini, M., James, R., and Shore, D. I. "Multisensory temporal order judgments: When two locations are better than one," Perception & Psychophysics, 65(2), 318-378, 2003.

[7] Kohlrausch, A., "Perceptual consequences of asynchrony in audio-visual stimuli," In: IPO Annual Progress Report 35 (2000).

[8] Paillard, B., Roy, P., Vittecoq, P., & Panneton, R., "Odyssée: A new kinetic actuator for use in the home entertainment environment." Proceedings of DSPFest 2000, Texas Instruments, Houston, Texas, and July, 2000.

[9] The Math Works, Inc., Matlab Statistics Toolbox, Natick, MA,, 2002.

[10] Altinsoy, M. E. "Perceptual Aspects of Auditory-Tactile Asynchrony," Proceedings of the 10th International Congress on Sound and Vibration, Stockholm, Sweden, July, 2003.

[11] Altinsoy, M. E., Blauert, J., & Treier, C., "Inter-Modal Effects of Non-Simultaneous Stimulus Presentation," A. Alippi (Ed.), Proceedings of the 7th International Congress on Acoustics, Rome, Italy, 2001.

[12] Martens, W. L., & Woszczyk, W. "Guidelines for Enhancing the Sense of Presence in Virtual Acoustic Environments." H. Thwaites (Ed.), Proceedings of the 9th International Conference on Virtual Systems and Multimedia, pp. 306-313, Montreal, October, 2003.