HUMAN-CENTERED DESIGN OF ACOUSTIC AND VIBRATORY COMPONENTS FOR MULTIMODAL DISPLAY SYSTEMS

William L. Martens
Faculty of Music, McGill University, 555 Sherbrooke Street W., Montreal, CQ, Canada H3A 1E3
wlm@music.mcgill.ca

1. INTRODUCTION

It is well established that human tolerance for latency between audio and video reproduction for teleconferencing is an important design consideration. Deployment of the multimodal display system described here required the coordination of signals for three sensory modalities, auditory, visual, and vibratory. Though a great deal of research has been done investigating audio/visual interaction, relatively little is known about interaction between reproduced acoustic and vibratory components of remotely captured events. Therefore, a study was undertaken to determine the intermodal delay required for brief acoustic and structural vibrations to be perceived as synchronous. The structure-borne component of recorded impact event was presented via a motion platform on which the observer was seated. The air-borne component of the event was presented via a multichannel loudspeaker array, with simulated indirect sound arriving from all around the observer. By varying the relative level and intermodal delay of the vibratory (structure-borne) components and the acoustic (air-borne) components, conditions allowing successful time order judgment (TOJ) were estimated using a two-alternative, forced-choice (2AFC) tracking procedure. Then, in order to avoid sequential response biases in the tracking procedure, the method of constant stimuli was used to determine the range of intermodal delay values associated with observers' reports of perceived simultaneity as a function of the relative level of the vibratory stimulus. Since the results of this investigation provide a basis for deployment of multimodal display technology that is generated through perceptual experimentation with relative levels and intermodal delay values, they are said to enable human centered design (Martens, 1999). In contrast to displays developed using conventional engineering approaches, then, these results may lead to the creation of more satisfying and convincing virtual environments for applications such as teleconferencing and realistic reproduction of remote musical performances.

1.1 Multimodal display technology

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 (Barfield, et al., 1995). 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 (Miner & Caudell, 1998), but even when asynchrony is detectable, there is useful variation in human experience within the tolerable range of asynchrony (Martens & Woszczyk, 2004). Other recent work (Woszczyk & Martens, 2004) has focused upon asynchrony between acoustic and vibratory display components in an attempt to quantify their multimodal integration in isolation from other display modalities. First hand experience with such bimodal display of these events suggested that physical synchrony between display components was not necessarily required to produce a subjective experience of simultaneity. The novel aspect of the research reported here is that it examined the impact of the relative level of vibration upon the perceived realism and naturalness of remotely reproduced impact events.

2. METHOD

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

2.1 Acoustic component display

The acoustic component was presented 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.

The stimuli were selected as 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. For the current study, the level of the acoustic stimulus was held constant at 82 dB(A).

2.2 Vibratory component display

Only vibration along the vertical axis was presented in this study, although the employed vibration transducer was capable of generating multidimensional vibration stimulation, providing users with motion having three Degrees of Freedom (3DOF) in a home theater setting (Paillard, et al., 2002). 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 displace the observer 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 vibratory 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. For the current study, the maximum vertical acceleration RMS value was 1.3 $m/sec^{2,}$ measured at the observer's foot position (using a B&K Type 4500 accelerometer). The vibration level was attenuated from this maximum RMS value in 7 steps, each of -3 dB, to cover a vibration range of 18 dB. The vibratory stimulus was also delayed relative to the acoustic stimulus in 7 steps of 10 ms, reaching a maximum of 40 ms, but also leading the acoustic stimulus in two cases. Observers made time order judgments, and also reported when the two components seemed to occur simultaneously.

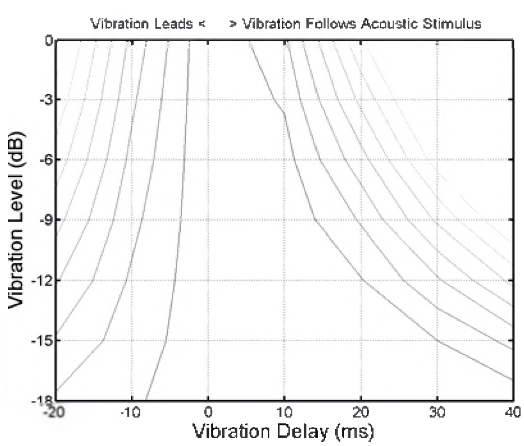

Fig. 1. Contour plot showing acceptability of a multimodal reproduction as a function of two parameters of the vibration display component, delay and level (see text).

3. RESULTS AND CONCLUSIONS

The results of this investigation can be summarized using the single contour plot acceptability of the multimodal reproduction shown in Figure 1. The darkest contours in the plot surround the region of highest acceptability on the plane defined by the 7 by 7 factorial combination of presented vibration delay and level values. At the lowest vibration level presented, the range of most acceptable vibration delay values extends from –10 ms (vibration leading the acoustic stimulus) to the maximum tested delay of 40 ms. As the vibration level was increased, the acceptable range of delay decreased, so that only true physical synchronization produced a reliable response of perceived simultaneity. When criteria other than strict simultaneity were employed, such as "naturalness," the range of acceptable delay values grew to include longer vibration delay values, but for the most natural impression, the vibration could not lead the acoustic stimulus. Also worth noting is the variation in subjective intensity of the impact event. For example, at the lowest vibration level presented, one observer reported that the impact event seemed more "powerful" when the vibration followed the acoustic stimulus by 20 ms, even though this was combination was not associated with the greatest sense of simultaneity.

REFERENCES

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

Martens, W. L., & Woszczyk, W. (2004) "Perceived synchrony in a bimodal display: Optimum intermodal delay values for coordinated auditory and haptic reproduction," Proceedings of the 10th International Conference on Auditory Display, Sydney, Australia, July.

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

Miner, N., & Caudell, T. (1998) "Computational Requirements and Synchronization Issues for Virtual Acoustic Displays," Presence: Teleoperators and Virtual Environments, 7 (4), 396-409.

Woszczyk, W., & Martens, W. L., (2004) "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.

Paillard, B., Roy, P., Vittecoq, P., & Panneton, R. (2002) "Odyssée: A new kinetic actuator for use in the home entertainment environment." Proceedings of DSPFest 2000, Texas Instruments, Houston, Texas, and July, 2000.

ACKNOWLEDGEMENTS

The author would like to thank Dr. Bruno Paillard of D-BOX Technologies for many helpful discussions. This research was supported by Valorisation-Recherche Québec.