Perception of Audio-Generated and Custom Motion Programs in Multimedia Display of Action-Oriented DVD Films

Kent Walker and William L. Martens

Sound Recording Area, Department of Theory, Schulich School of Music,
McGill University, 555 Rue Sherbrooke Ouest, Montréal, QC, H3A 1E9, Canada
kent.walker@mail.mcgill.ca
http://www.music.mcgill.ca/~kentw/

Abstract. This paper addresses a practical problem associated with multimedia display systems which utilize motion-platforms or chairs. Given audio-visual content for which motion data is not available, motion may be automatically generated from multichannel audio, usually from a Low-Frequency Effects channel (LFE) such as that distributed on Digital Versatile Discs (DVDs). Alternatively, custom motion programs may be created to accompany multimedia content. This paper presents the results of a study designed to test the sense of realism, sense of presence, and global preference for multimedia playback in these two distinct cases of platform accompaniment: motion which has been generated automatically from audio, and motion which has been designed expressly for the purpose of stimulating appropriate haptic and vestibular sensations.

1 Introduction

1.1 Research Context and Research Questions

Since the advent of the Digital Versatile Disc (DVD), affordable advances in multimedia display technology have evolved which offer consumers opportunities for enhanced experience in the form of devices capable of haptic stimulation and motion accompaniment (whole-body vibration). These devices are many and fall into several categories but generally may be described as platforms and chairs which are capable of moving and shaking observers. Given the variety of devices available and the absence of standardized practice, multimedia content providers such as Hollywood film houses have yet to supply this niche market with motion programs for home entertainment use.

As a result developers of haptic technology are faced with the question of how to generate motion programs for their customers. Two main options exist: the first is to generate motion programs automatically using existing content; the second is for hardware manufacturers to employ programmers to create custom-coded motion from scratch for content from media houses. Depending on the program material or the multimedia scene presented to the observer, either method of generating whole-body vibration may be sufficient in terms of enhancing an

D. McGookin and S. Brewster (Eds.): HAID 2006, LNCS 4129, pp. 1–11, 2006.

observer's experience. For example, previous investigations by the authors [15] and others [3] have indicated that for music-only material, generation of whole-body vibration using existing audio signals is sufficient for the enhancement of experience provided that the systems are properly coordinated. In other words, for music programs separate vibration recordings or custom motion programs are generally not required. However, in the case of more action-oriented program material this may not be the case. Though there is a significant body of research on how to create motion-cues for users of flight simulation technology via vestibular stimulation [18], there has been relatively little work done on the most effective integration of motion platforms into home entertainment oriented audiovisual content display. Different multimedia scenes may require different kinds of motion programs in order to create within observers the greatest suspension of disbelief.

The experiments described herein investigated multimedia display of commercially available DVD film titles. Motion data generated from audio Low Frequency Effects channels (the *.1* channel on a *5.1* channel disc) was compared to motion generated specifically for individual DVD titles by a human designer. Experiments were designed to determine if existing information present in LFE channels is sufficient for the enhancement of presence and realism, if observers prefer custom-coded motion when watching DVD titles, and if the extra costs involved in the creation of custom motion programs is perceptually justifiable. The tested hypothesis was: for action-oriented DVD titles is custom-coded motion more likely to be preferred over audio-generated motion, and is this preference linked to an observers sense of presence and sense of realism for a displayed multimedia scene?

Before further explaining the experiments conducted, background on the technology for motion-based display and human sensitivity to vibration is presented.

1.2 Enhancing Presence in Multimedia Display

In recent research and development of advanced multimedia display technology, great emphasis has been placed upon multichannel sound and the enhanced consumer experience associated with coordinated display of visuals and spatial audio content. The potential for user immersion in the presented virtual world is one benefit of such multimedia display which is most properly called *television-type telesensation* [13]. Compared to more conventional media, such immersive audio-visual content produces a higher proportion of user responses indicating higher *sense of presence* or "sense of being transported to the electronically-mediated space" [7]; consumers can forget that the virtual world presented to their eyes and ears is an electronic reproduction, and imagine instead that they are experiencing the virtual world first hand. However, this suspension of disbelief is weakened considerably by one factor that has often been ignored in the development of advanced multimedia displays: observers visiting these virtual worlds are *not* disembodied minds. Regardless of where observers' eyes and ears take them, their bodies most often stay put in the physical display space. On the other hand, if simulation includes touch and motion sensations which are consistent with what is seen and heard, a heightened sense of presence is to be expected.

There has been a small but growing interest in providing a means for creating such high-quality *multimodal* experiences for consumers in home theater and computer gaming applications, typically in the form of moving seats or motion platforms. Such multimodal content has the potential to not only enhance experience, but to spawn new markets for both entertainment and electronics corporations in a commercial age fraught with the terror of economic uncertainly, caused largely by the real threat of entertainment piracy. To date, the film industry's reaction to piracy has mostly been the so-called *high-definition* technological revolution of home-theater, resulting most recently in an emerging format war between High-Definition Digital Versatile Disc (HD-DVD) and Blu-Ray Disc (BD). Nevertheless, experts in these new formats question the ability of these systems to offer extra value over what's currently available to consumers via existing popular DVD standards, as many consumers have lesser quality sound reproduction systems and televisions that will benefit perhaps only marginally from these higher-definition media [12]. However, if DVD houses embrace and promote haptic and motion transducer technology together with their new formats, they may have a greater degree of success with the high-definition revolution: the addition of a mode of display is more immediately gratifying to consumers than just-noticeable differences in display quality for audio and video alone.

1.3 Haptics and Motion Sensation

Haptic comes from the Greek *hapt esthai* which may be literally translated as *feel sense* [1]. Haptic sensation can result from many forms of stimulation, ranging from local vibrations on the skin to whole-body movement, and relies on a variety of sense organs and receptors from multiple sensory systems. Cutaneous sensory organs such as *Meissner's Corpuscles* (used for the sensation of vibration below approximately 40 Hz), *Pacinian Corpuscles* (used above 40 Hz), and *Merkel Discs* (used for the sensation of pressure on the skin) [4]. The haptic senses also include the organs of human kinesthesia such as muscles, tendons, and joints (used for the sensation of movement below 30 Hz), although such vibration-related stimulation is distinguished from the stimuli for the vestibular organs of the inner-ear, namely the *semi-circular canals* (used for detection of angular acceleration of the head), and the *otoliths* (used for detection of linear acceleration). It is important to note that there is a great deal of overlap in our sensory systems with regards to vibration, as our hearing sensitivity extends down to approximately 20 Hz, while our vibration sensitivity grows with frequency to peak at approximately 250-300 Hz; therefore, low-frequency sound sources are often *both heard and felt.*

1.4 Motion Degrees of Freedom (DOF)

In control of motion platforms, it is necessary to distinguish how freely the platform may be made to move through space. Complete freedom of motion in space admits Six Degrees of Freedom (6DOF), which include the possibility of three directions of displacement and three angular gyrations. The terms that are used in the technical literature to describe an object or observer's motion through space are not in such popular use, and so these terms are related to more common language in this section. To begin with, there are three terms that are used to describe the rotation of an observer whose spatial position does not change: yaw, roll, and pitch. If a person who is standing upright rotates to the left or right about the vertical axis through their body, this rotation is termed yaw. If an airplane were to dive forward, the rotation about the left/right axis would be termed pitch, and if the airplane were to tilt to one side, the rotation about the front/rear axis would be termed roll (not to be confused with the acrobatic maneuver termed a *barrel roll* in which an airplane not only rotates along its longitudinal axis, but also follows a trajectory along the surface of a cylinder or barrel). The remaining set of three terms is used to describe the spatial translation of an observer whose orientation in space does not change. These three translational motion terms, describing movement leftward, forward, or upward, are respectively: sway, surge, and heave.

2 Methods

2.1 The Loudspeaker Array

The current study used a calibrated multichannel stereophonic sound system compliant with ITU standards [6] yet with full range capabilities at all standard angles (in degrees relative to the median plane the angles were -110, -30, 0, 30, and 110). Five low-frequency drivers (ranging from 35 Hz to 300 Hz) and five higher-frequency drivers (ranging from 300 Hz to well over 20,000 Hz). The low-frequency drivers were Mini-Mammoth subwoofers manufactured by Quebec-based D-BOX Technology, which were placed at standard locations for the 5 main speakers in surround sound reproduction. The higher-frequency drivers were dipole radiating, full range transducers featuring the Planar Focus Technology of Level 9 Sound Designs of British Columbia. These speakers were placed at the same azimuth angles as the 5 low-frequency drivers, but in a ring positioned 15 degrees above the observer's ear level. All listening tests were conducted in a treated room [11].

2.2 The Visual Display

A Panasonic model TH-50PHD6UY HDTV 50-inch plasma display with a resolution of 1366 x 768 and contrast ratio of 3000:1 was used to display the video portion of all DVD titles. A Pioneer model DV 563A DVD player was also used. Titles were of the widescreen variety and observers sat at a distance of 2 meters from the screen.

2.3 The Motion Platform

The research presented here used a commercially available motion platform, the Odyssée system from the Quebec based company D-BOX Technology [9]. This

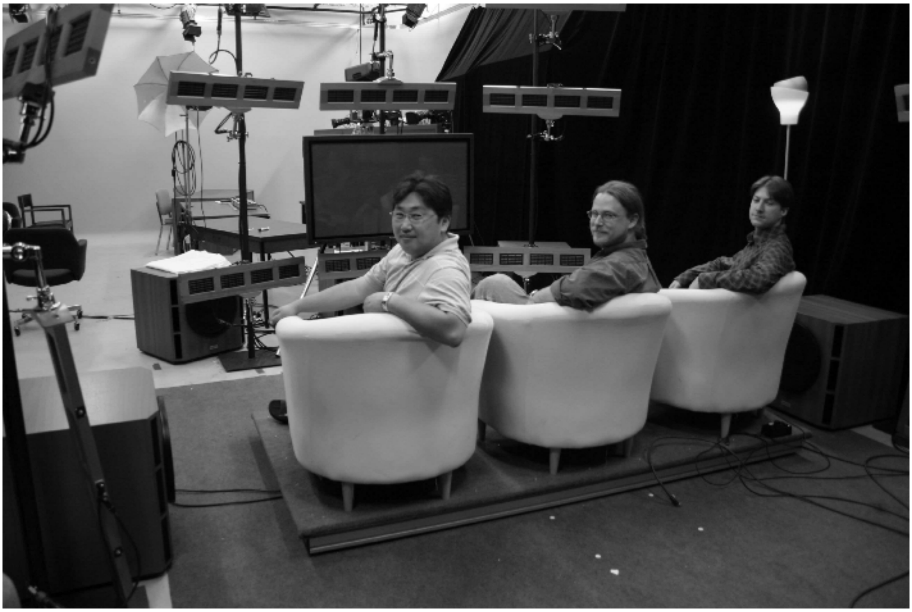

Fig. 1. A photograph of the installation at the Immersive Presence Lab at McGill University. The loudspeaker array, visual display, and motion platform are visible. Three participants (not the individuals shown) sat at the positions indicated.

system uses four coordinated actuators to provide multiple users with motion in three Degrees of Freedom (3DOF) for a home theater setting. No rotation about the vertical axis is allowed by this system, so no change in heading (or yaw) can be simulated. But two other rotations are enabled, and also translation along a single axis. The three terms describing these control possibilities are summarized in Table 1, along with an example of a command that could be sent to the motion platform to produce each type of motion.

The motions generated by the Odyssée system are based upon its use of four actuators positioned at the corners of the motion platform, each of which is independently controlled but only capable of motion along a vertical axis. When all four actuators move together, users can be displaced linearly upwards or downwards with a very quick response and with considerable force (the feedback-corrected linear system frequency response is flat to 100 Hz). The two angular motions are enabled as follows: upward motion of the two left actuators coupled with downward motion of the two right actuators enables pure roll, and contrary motion of the front and rear actuators allows for pitch control.

A detailed set of measurements was performed on the platform system using an Agilent 35670A Dynamic Signal Analyzer as the measurement device, as well as a PCB Piezoelectronics Model 356 416 tri-axial accelerometer. This was done in order to characterize general aspects of the performance of the platform, as well as its reaction to the stimuli used in perceptual experiments of motion quality.

Table 1. Motion Possibilities of Platform Used in Experiments

Motion Label	Description and Command Example
Roll (Y)	Rotation about the left/right axis ("tilt left")
Pitch (X)	Rotation about the front/back axis ("tilt back")
Heave (Z)	Upward/downward translation ("rise up")

It was noticed at this time that the measured results were affected by the location on the platform where the measurements were taken. Different areas had differing modal response when stimulated. It was also noticed during measurement that the weight on the platform had drastic effects on active-servo performance. As a result, measurements were taken with one of the participants sitting on the platform in the center seat used for perceptual testing. For these reasons, the majority of measurements were taken directly on the platform where the participants placed their feet, as well as directly from the hand of one participant. While measurements taken at the platform surface demonstrated that it was indeed moving only mostly with regards to heave (up and down) when heave was input to the system, measurements taken at the hand displayed movement in multiple dimensions; this makes sense as the observer was not asked to hold his body completely still during measurement (indeed given the force of acceleration applied to the platform this may have been impossible). The peak acceleration values along the heave axis were found to range from approximately 100 *milli-g* to 197 *milli-g* ($1g = 9.8\,m/s^2$) depending on the stimulus and the measurement location. Motion measured at the hand of an observer seated on the motion platform for an excerpt from the Hollywood blockbuster *Master and Commander* is

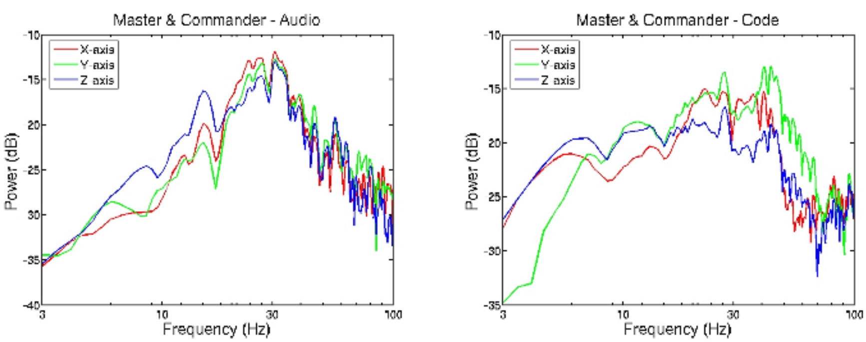

Fig. 2. Motion measured at the hand of an observer seated on the motion platform for an excerpt from the Hollywood blockbuster *Master and Commander*. Motion generated from the Audio LFE channel is shown on the left. Motion resulting from human-designed FX-Codes that were provided by D-Box Technologies are shown on the right. The Code motion measurement displays a greater variety of motion, as well as a lower frequency-extension, as frequencies below approximately 20 Hz are not normally part of multichannel audio programmes.

shown in Figure 2. Measurements were taken under two conditions: *Audio* (shown on the left) where motion was automatically generated by the LFE channel, and *Code* (shown on the right) where motion codes were custom crafted by a human motion-designer. D-Box Technologies offers more than 300 custom crafted motion codes to customers of its whole-body vibration products. While the *Audio* measurement displays axes which appear to be relatively more correlated than the *Code* measurement, the apparent decorrelation of *Audio* X Y Z are created by the human body's natural response to whole-body vibration (which has 6 degrees of freedom). The *Code* motion measurement displays a greater variety of motion, as well as a lower frequency extension, as frequencies below approximately 20 Hz are not normally part of multichannel audio programmes.

2.4 Measurements of Crosstalk Between Display Systems

Crosstalk between subwoofers and the motion platform was also measured, as subwoofers are capable of inducing structural vibration as well as haptic sensation in listeners. This was measured to be less than 20 *milli-g*. The reverse was also true as the motion platform did induce airborne vibration experienced as sound, however, this was measured to be approximately 40 dB C below the sound level of the loudspeakers.

2.5 Tests of Perceived Motion Quality

In a blind test, three observers were shown a total of 30 excerpts from seven different Hollywood DVD titles shown in Table 2. Each excerpt was presented with both audio-generated motion as well as custom motion in random order. Subjects were asked to rate motion on a 10-point scale (using pen and paper) with regards to three perceptual attributes: sense of realism, sense of presence, and global preference. As can be seen the measurements shown in Figure 2, the frequency content of the audio-generated and custom motion were significantly different. Observers could, therefore, easily discriminate a difference between each type of motion, however, they were not told which type of motion they were experiencing.

Table 2. Hollywood action-oriented DVD titles used for the generation of whole-body vibration and motion

DVD Title	Film Studio
Cast Away	20th Century Fox
Charlie's Angels (2000)	Sony Pictures Entertainment
I-Robot	20th Century Fox
Finding Nemo	Walt Disney Video
Master and Commander - The Far Side of the World	20th Century Fox
Perfect Storm	Warner Home Video
Terminator 3	Warner Home Video

3 Results

Figure 3 displays mean values for 30 LFE-derived and coded-motion excerpts of 1 minute in length from the seven action-oriented Hollywood DVD titles (see Table 2). There is a significant difference in the mean ratings observed for each subject (with regards to each attribute), and the mean *differences* between subjects for each attribute were relatively consistent. With regards to all attributes, subjects' mean ratings for coded-motion (Code) were higher than those for LFE-derived motion (Audio). Mean differences were greatest with regards to global preference (Pref) and smallest for sense of presence (Pres). It is clear that the coded motion was greatly preferred over audio-generated motion.

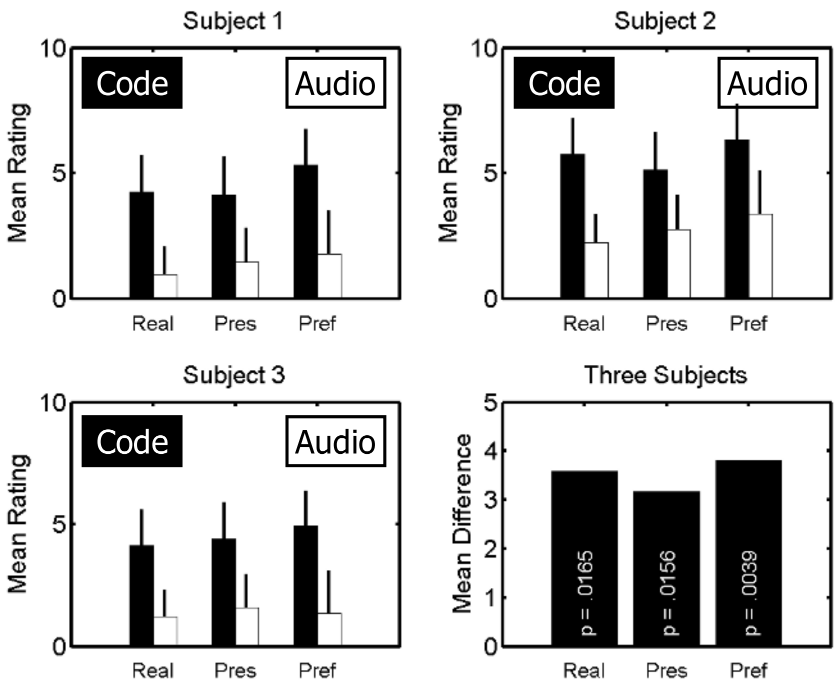

Fig. 3. Ratings of sense realism (Real), sense of presence (Pres), and global preference (Pref) from three observers exposed to both motion generated automatically from an LFE audio program and custom generated coded-motion. Observers were shown 30 excerpts of approximately 1 minute each, from seven different Hollywood DVD titles (see Table 2). Although there is a significant difference in the mean ratings each subject made with regards to each attribute, the mean *differences* between subjects is relatively consistent.

4 Discussion

Automatically derived LFE generated motion was taken directly from the *.1* LFE channel of the Dolby Digital signal; this channel of the AC-3 encoded signal is band-limited to frequencies below 120 Hz and is sampled at a rate of 240 Hz [14]. While automatically derived LFE motion moved the platform only up and down in the Z dimension, coded-motion was present in all 3 dimensions (Z, X, and Y).

Arguably, a significant step towards enhancing presence (and perhaps haptic sensation) in sound reproduction was taken with the development of the LFE channel for moving pictures [10]. This channel was originally designed for 70 mm film in the 1970s to allow for the addition of effects with lots of low-frequency energy [5]. This technology has since been disseminated to the consumer in the form of 5.1 multichannel (also known as 3-2-1) home theater systems of varying quality which are also used for music reproduction. It is important to note that a *dedicated channel* for LFE (the *.1* channel) should not be taken to be synonymous with the *number of subwoofers* in a reproduction system; both the ideal number and position of subwoofers in multichannel stereophonic speaker arrays (incorporating various bass handling processes) is a question still under investigation by many interested in acoustics and applied psychoacoustics.

That the LFE channel was originally designed for effects and that it is almost exclusively used for this purpose in home theater is of great importance to haptic-vestibular-motion systems which use this channel to provide motion to the observer. It is an assurance that only information relevant to the haptic experience will be transferred. In other words, low-frequency sounds which are not effect related but are present in the five full-range channels (perhaps music, plosives in speech, or such extraneous sounds) will not be routed to the vibration system in question provided that bass-management is not used. The LFE channel, therefore, represents a channel of information which provides haptic systems with a viable source for motion generation.

That said, the results of this study indicate that coded-motion is globally preferred and perceived as being more realistic for action-oriented film content. Coded-motion is more likely to create a stronger suspension of disbelief within the observer. However, given that mean differences for ratings of sense of presence between coded-motion and LFE-derived motion were the smallest of the three examined factors, LFE-derived motion appears to evoke within the observer an adequate degree of presence (more than would otherwise be achieved without any vibration).

While the frequency content of motion codes is determined by the type of motion simulated, generally speaking, frequencies below approximately 20 Hz may be associated with environmental motion commonly found in everyday situations, such as motion experienced in a boat (caused by waves) or vehicle (caused by changes in acceleration), whereas, higher frequency motion is experienced as vibration or shaking. Depending on the frequency and waveform of the motion in question, humans appear to have a natural ability to discern certain environments simply through the sensation of motion. For example, during the physical measurements of coded-motion shown in Figure 2, the observer remarked that he could tell which scenes were water scenes, and which were not, even if audio and video were *not* present (simply from the motion codes). This observation is mostly likely due to the fact that the coded-motion was a three-dimensional motion program (as opposed to audio-generated motion which manifested itself in the platform in the form of heave only). Future publications include a psychophysical study of simpler stimuli which will attempt to verify whether or not this is indeed the case.

5 Conclusions

In the absence of codes designed by hand (custom motion programs), low-frequency audio can generate some appropriate motion especially when taken directly from an LFE channel. Although this audio-generated motion is typically not very satisfying, it does somewhat enhance an observer's sense of presence. Nonetheless, as a first pass for later careful selection and editing, audio-derived motion data can provide savings in terms of human labor costs when programming custom motion. Of course, human intervention in the form of amplification and elimination of audio-derived motion data will almost always produce superior results. The conclusions for this paper are thus :

- Custom motion programs most successfully create in observers a sense of realism related to a virtual environment and are generally preferred to motion generated automatically from audio programs for action-oriented DVD films.
- The existing information in LFE channels are generally useful for subtle enhancement of observer presence, however, custom motion programs are globally preferred, evoke a greater sense of observer presence, and a sense of realism which is not possible with audio-generated motion.
- Manufacturers and media houses might profit from investing in the creation of custom motion programs for home theater applications.

Acknowledgements

The authors would like to thank: Dr. Atsushi Marui and Mr. Sungyoung Kim for their input, The Schulich School of Music at McGill University, as well as Dr. Durand R. Begault from the Human Factors Research and Technology Division of the NASA Ames Research Center for his direction and 3-D measurements of the platform motion and vibration. This research was supported in part by Valorisation-Recherche Québec (www.vrq.qc.ca) and the Centre for Interdisciplinary Research in Music Media and Technology (www.music.mcgill.ca/cirmmt).

References

1. Begault, D.(2006). Haptic Perception and the Development of New Tools for Digital Studio Technology. Unpublished lecture given to McGill University.
2. Bouguila, L., Ishii, M., Sato, M. (2000). What Impact Does the Haptic-Stereo Intergration Have on Perception in Stereographic Virtual Environment? In: S. Brewster and R. Murray-Smith (Eds.) Haptic Human-Computer Interaction: First International Workshop, Glasgow UK, 135–150.

3. M. Daub and M. E. Altinsoy (2004). Audiotactile simultaneity perception of musical-produced whole-body vibrations. In: Proceedings of the Joint Congress CFA/DAGA.
4. Griffin, M.J. (1990). Handbook of Human Vibration. Elsevier Academic Press.
5. Holman, T. (1999). 5.1 Surround Sound Up and Running. Focal Press.
6. International Telecommunications Union (1994). Recommendation BS.775-1, Multichannel stereophonic sound systems with and without accompanying picture.
7. Martens, W. L., Woszczyk, W. (2003). Guidelines for Enhancing the Sense of Presence in Virtual Acoustic Environments. In: H. Thwaites (Ed.) Proceedings of the 9^{th} International Conference on Virtual Systems and Multimedia, 306–313.
8. McGee, M.R., Gray, P., Brewster, S. (2000). The Effective Combination of Haptic and Auditory Textural Information. In: S. Brewster and R. Murray-Smith (Eds.) Haptic Human-Computer Interaction: First International Workshop, Glasgow UK, 118–126.
9. Paillard, B., Roy, P., Vittecoq, P., Panneton, R. (2000). Odyssée: A new kinetic actuator for use in the home entertainment environment. In: Proceedings of DSPFest 2000, Texas Instruments, Houston, Texas.
10. Rumsey, F. (2001). Spatial Audio. Focal Press.
11. Ryan, T., Martens, W.L., Woszczyk, W. (2005). The effects of acoustical treatment on lateralization of low-frequency sources. In: Proc. of 149th Meeting of the Acoustical Society of America **117**, 2392.
12. Taylor, J., Johnson, M.R., Crawford, C.G. (2006). DVD Demystified. McGraw-Hill.
13. Terashima, N. (2002). Intelligent Communication Systems. Academic Press.
14. Todd, C. *et al.* (1994). Flexible perceptual coding for audio transmission and storage. In: Proceedings of the 96^{th} Convention of the Audio Engineering Society.
15. Walker, K., Martens, W.L (2005).Haptic-Auditory Distance Cues in Virtual Environments: Exploring Adjustable Parameters for Enhancing Experience. In: Proceedings of the IEEE International Workshop on Haptic Audio Visual Environments and their applications (*HAVE 2005*), Ottawa, Canada, 101–103.
16. Walker, K., Martens, W.L., Kim, S. (2006). Perception of Simultaneity and Detection of Asynchrony Between Audio and Structural Vibration in Multimodal Music Reproduction. In: Proceedings of the 120^{th} Convention of the Audio Engineering Society, Paris, France.
17. Woszczyk, W., Cooperstock, J., Roston, J., Martens, W. (2005). Shake, Rattle, and Roll: Getting Immersed in Multisensory, Interactive Music via Broadband Networks. Journal of the Audio Engineering Society, **53**, No. 4, 336–344.
18. Wu, W.(1997). Development of Cueing Algorithm for the Control of Simulator Motion systems. MS Thesis, State University of New York at Binghamton.