SUBSPACE PROJECTION OF MULTICHANNEL AUDIO DATA FOR AUTOMATIC CONTROL OF MOTION-PLATFORM-BASED MULTIMEDIA DISPLAY SYSTEMS

William L. Martens & Wieslaw Woszczyk

Multichannel Audio Research Lab
McGill University
Montreal H3A 1B4
Canada

ABSTRACT

This paper addresses a practical problem associated with multimedia display systems that utilize motion platforms to create for users both haptic and vestibular sensations associated with their movement through a displayed environment. Given audiovisual content for which motion data is not available, the motion data that is required for motion platform control can be generated automatically from multichannel audio data such as that distributed on DVDs presenting popular movie titles. This paper presents initial results of a study designed to test the effectiveness of a subspace projection of multichannel audio data for automatic control of motion-platform-based multimedia display.

1. INTRODUCTION

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. Compared to more conventional media, such immersive audiovisual content produces a higher proportion of user responses indicating higher "sense of presence" or "sense of being transported to the electronically-mediated space." (See, for example, [1]). 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, the illusion is weakened considerably by one factor that often has been ignored in development of advanced multimedia displays. Observers visiting these virtual worlds are not disembodied minds. Regardless of where users' eyes and ears take them, their bodies most often stay put in the physical display space. On the other hand, if simulation includes touch (haptic) sensations and motion (vestibular) sensations that are consistent with what is seen and heard, a heightened "sense of presence" is to be expected. Consequently, there has been a growing interest in providing a means for creating such multimodal experiences for consumers in home theater and computer gaming applications, typically in the form of a moving seat or a motion platform.

The research described in the current paper was focused upon the problem of how to automatically generate motion data for existing multimedia titles for which none is available. The target application was home theater reproduction of popular movie titles currently available on DVD, using a motion platform for which motion data is required. The goal was to produce user motion for the best reproduction of spatial information in synchronized multimodal display. The means was to process the 5.1 channel audio data present on the DVD to create a lower-dimensional signal suitable for control of a motion platform with three degrees of freedom (3DOF).

Though there is a significant body of research on how to create motion cues for users of flight simulation technology via vestibular stimulation (see, for example, [2]), there has been relatively little work done on the most effective integration of motion platforms into home-entertainment-oriented audiovisual content display. Before further explicating the role of audio analysis in generating motion data that is the topic of the current research, however, a brief background on the technology for motion-based display is presented.

1.1. 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 of 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: <u>yaw</u>, <u>roll</u>, and <u>pitch</u>. If a person who is standing upright rotates to the left or right about the vertical axis through their body, this rotation is termed <u>yaw</u>. If an airplane were to dive forward, the rotation about the left/right axis would be termed <u>pitch</u>, and if the airplane were to tilt to one side, the rotation about the front/rear axis would be termed <u>roll</u> (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, <u>sway</u>, <u>surge</u>, and <u>heave</u>.

1.2. A 3DOF Motion Platform for Multimodal Display

The current research is based upon a commercially available motion platform, the Odyssée™ system from the Quebec based company D-BOX Technology (http://www.d-box.com). The Odyssée™ 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.

Roll	Rotation about the front/rear axis ("tilt left")
Pitch	Rotation about the left/right axis ("tilt back")
Heave	Upward/downward translation ("rise up")

Table 1. Definition of terms used in describing three available types of motion, with examples of commands for motion platform movements associated with each.

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 of 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.

1.3. Subspace projection of multichannel audio data

This subsection introduces the concept of a subspace projection of multichannel audio data for automatic control of motion-platform-based multimedia display. The problem that is addressed here is the following: The variation in energy over frequency, time, and space in the audio signals reproduced by five satellite loudspeakers and a single subwoofer provides a basis for driving a motion platform that is synchronized with acoustic events in the DVD program. But this complex variation must be reduced to a much lower dimensional representation in order to provide motion control signals suitable for the motion platform. Audio analysis can estimate the distribution of energy over time and space in multiple frequency bands with resolution matching human sensory capacities. These multidimensional data can be submitted to principal components analysis as a data reduction method to yield just three time-varying control signals, which can be applied to the "Motion Profile Generator" unit for the motion platform.

The two primary needs for audio analysis in generating motion data that is the topic of the current research are the following: First, an improved method for realtime generation of motion signals from DVD audio is needed to enhance the motion-platform-based reproduction of DVD titles for which no motion data is available. This need is not yet met at the current stage of this research using a subspace projection of multichannel audio data. Second, an improved method for non-realtime generation of motion signals is needed for use in the authoring of motion profiles that are to be distributed on CD-ROM to accompany DVD titles, so that reproducing these DVD titles will not need automatic realtime motion profile generation. This latter case is the preferred mode of synchronized reproduction of audiovisual content with motion, since it allows for human intervention in the process, with superior results expected. This need is met by the results achieved at the current stage of this research, as three tracks of motion data can be supplied along with the 5.1 channel audio data from a DVD selected by a human designer of motion profiles. These three tracks can be read into the multitrack mixing application that has been used to produce motion profiles for more than 300 DVDs by designers at D-BOX Technologies. The results of the current research provide potential benefits in terms of the time required for motion profile designers to produce quality products from their labor, since they are given a good starting point for the analyzed DVD.

2. METHODS

A detailed description of the audio analyses involved in producing a subspace projection for motion control is beyond the scope of this short paper, and so only a brief summary of the methods employed are provided here. A more thorough description of the basis of the analysis considering human auditory capabilities is provided in the first author's paper on related work in spatial hearing [4]. The principles specific to the current research project are the following:

1. The envelope function over time for each of 6 channels of digital audio from a DVD is computed with each of 24 frequency bands simulating the critical bands of human hearing (incorporating Matlab functions from Slaney's [5] "Auditory Toolbox").
2. The resulting data was composed into matrices spanning 4-second segments of the audio program from the DVD, and these matrices contained 24-column partitions for the five satellite loudspeaker channels, and an additional (201st) column corresponding to the single subwoofer channel.
3. These matrices were submitted to principal components analysis to derive an orthonormal basis for energy variation over frequency, time, and space.
4. The matrices were post-multiplied by the first three columns of the derived orthonormal basis matrix to produce three column vectors of principal component scores, at the time resolution fixed by the rows of the input data matrices.
5. These vectors were lowpass filtered to produce slowly varying control signals suitable for use as motion profiles.

The result constitutes a subspace projection of the input multichannel audio data that contains in three vectors a simple reduction of a large proportion of the variance in the input matrices, which can be envisioned as a 210-dimensional space. Providing more detail on this method is beyond the scope of the current paper, and only a brief review of current results is given in the remainder of the text.

3. RESULTS

Fig. 1 shows the results of the method for a 4-second segment of multichannel audio sampled from a popular DVD title. The upper panel of the figure shows for reference the normalized summed envelope functions for all 24 frequency bands separately for just two of the audio input channels, those destined for the left and right satellite loudspeakers. The lower panel of the figure shows two of the automatically generated motion signals destined for control of the platform on which users are positioned, one for roll angle (positive values are for leftward roll), and one for upward translation termed heave (positive values are for upward heave).

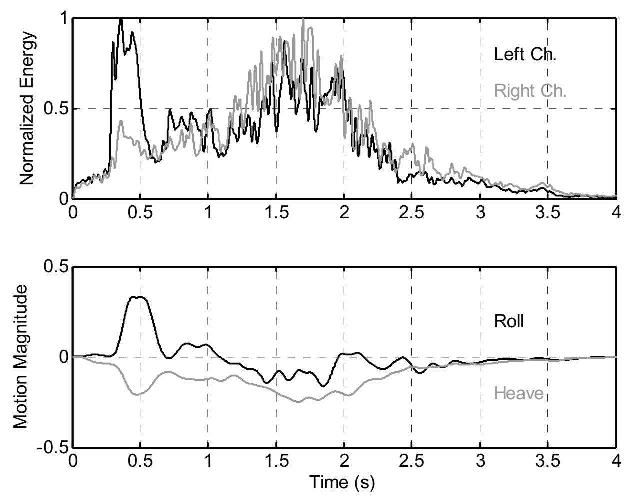

Figure 1. Results of the subspace projection method for a 4-second segment of multichannel audio sampled from a popular DVD title. See text for details.

To understand the value of these results the reader needs to be given some details about the content of the audiovisual program. At about .25 seconds from the beginning of the 4-second segment, a lion steps on a piano keyboard which is located to the left of center. Note that the corresponding audio event is panned to the left in the multichannel mix at this point. This event shows up in the automatically generated motion profile as a roll to the left, akin to a tilting of the whole room due to the lion's weight. At the same time, the motion platform exhibits a slight downward drop, as indicated in the negative direction taken by the heave signal shown in Fig. 1. Though the roll angle quickly returns to a value near zero, the heave signal wavers a bit, but continues to a yet lower point, producing a characteristic "sinking feeling in the stomach" of those seated on the motion platform.

4. DISCUSSION

The above example illustrates the nature of the goals for the automatic processing of multichannel audio signals. It is most important to note that the human designer of motion profiles has the visual content as another contribution to the process, and this input contains critical information that is often not present in the audio program. Other roles for the output of the audio analysis (not addressed in detail here) would be the automatic marking of key events that might require further attention from a human designer. Besides directing the designer's attention to key events, there could also be further event-related calculations, such as those that could be done in a second pass on the audio analysis results. A primary advantage of making a second pass would be to add anticipatory motion data to the resulting motion profile, so that the motion platform is prepared for large motions prior to the onset of a marked event. The marked events also provide a basis for shifting the onset time for analysis windows, and potentially expanding window durations as well. Such intelligence in the post-processing of the audio analysis result is highly desirable for non-realtime generation of motion profiles for DVD titles, for which CD-ROM data will be distributed.

Other details worth noting here are that some portion of the audio signal will almost always be mixed back into the motion channels as well. This is especially true for the subwoofer channel signal, which is naturally added to provide vibrations in the motion platform that are removed in the subspace projection process. The reason for this is that the subspace projection processing is limited to large motions of the platform with significant energy below 10 Hz. The subwoofer signal, on the other hand, has significant energy up to 80 Hz, which is fed to the motion platform at much lower level. This is done so that the actuators stay within their linear range, which is only for displacements up to about 1/10 of their total travel (due to power and speed limitation of the actuator motors).

5. CONCLUSION

Though not yet submitted to any systematic user testing, the results of the subspace projection seem promising as a means for automatically generating motion data. An important caveat is that human intervention in the form of amplification and elimination of some results will almost always be necessary. Nonetheless, as a first pass for later careful selection and editing, the current method provides clear potential benefits. A continuing research study is underway that has been designed to find which heuristics work best for post-processing the audio analysis output, with user evaluations providing the means for generalizing the results of this work.

6. ACKNOWLEDGMENTS

The authors would like to thank Dr. Bruno Paillard of D-BOX for many helpful discussions, and to thank the team at D-BOX who demonstrated for us the "art" of motion profile design.

7. REFERENCES

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

[2] Wu, W., Development of Cueing Algorithm for the Control of Simulator Motion systems, MS Thesis, State University of New York at Binghamton, May 1997.

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

[4] Martens, W. L. Principal components analysis and resynthesis of spectral cues to perceived direction. In: Sever Tipei & James Beauchamp, (Eds.), Proceedings of the International Computer Music Conference, Champaine-Urbana, IL. Sept., 1987.

[5] Slaney, M., Auditory Toolbox: A MATLAB Toolbox for Auditory Modelling Work, Version 2, Technical Report 1998-010, Interval Corp., Palo Alto, CA, 1998.