IMMERSIVE SPATIAL SOUND FOR MOBILE MULTIMEDIA

V. Ralph Algazi[1] Richard O. Duda[2]

CIPIC Interface Laboratory

University of California, Davis

508 2nd Street, Suite 107

Davis, CA 95616, USA

http://interface.cipic.ucdavis.edu

vralgazi@ucdavis.edu

rod@duda.org

ABSTRACT

While mobile technology precludes large electronic displays for visual immersion, sound heard over headphones – a widely accepted technology – is ideally suited for mobile applications. In this paper, we report on a newly developed immersive headphone-based approach that opens new opportunities for mobile multimedia. This new motion-tracked binaural sound technology (abbreviated as MTB) provides and exploits the strong perceptual cues that are created by the voluntary motions of the listener's head. A head-tracker is used to modify dynamically the sound heard by the listener, recreating the natural experience. After presenting an overview of the key concepts and of the challenges of the implementation of MTB, we describe examples of MTB spatial sound applications. Finally, we outline new mobile multimedia applications that would combine immersive spatial sound, head-tracking, and small visual displays.

Keywords: binaural, head tracking, HRTF, immersive, mobile, spatial sound, 3-D sound

[1] Corresponding author, IEEE member
[2] IEEE member

1. INTRODUCTION

The rapid expansion of mobile electronics is straining the ability of the user interface to provide information to the user. The reduced capacities of small optical displays limit the presentation of ordinary text and graphics, and preclude anything approaching visual immersion. By contrast, spatial sound delivered over headphones can provide an impressively realistic experience of auditory immersion, making it a natural choice for mobile applications.

In this paper we describe a new immersive spatial sound technology called MTB, which stands for Motion Tracked Binaural. MTB is equally applicable to live, recorded, and computer synthesized sound, or to augmented-reality mixtures of such sounds. Thus, it provides a "sonic canvas" that can be populated with a full palette of spatially localized information. MTB employs a head tracker to sense motions of the listener's head, using that information to stabilize the sound field and to eliminate front/back confusion. This not only generates a strong sense of presence, but also returns valuable feedback to the information provider.

In Section 2, we describe the use of binaural methods for live recording and computer generated spatial sound. In Section 3, we provide an overview of MTB and its advantages. In Section 4, we discuss important implementation issues. In Section 5 we describe some existing and planned applications of this immersive sound used on its own. In Sections 6 and 7, we speculate on some scenarios suitable for mobile multimedia that exploit the ability to generate a highly realistic and stable soundscape.

2. BACKGROUND

Binaural technology is a standard approach for providing spatial sound over headphones [1, 2]. Although it is capable of producing vividly realistic results, sound sources that are on or near the median plane tend to be heard inside the listener's head. The source of this problem is the lack of response to head motion. Median plane sources produce essentially the same sound in each ear. When the sounds continue to be the same in each ear no matter how the listener turns his or her head, they are perceived as being inside the head. Because the most interesting sources are usually in the median plane, this is a serious problem.

The need for spatial sound in virtual reality (VR) systems stimulated research interest in how binaural sound can be synthesized, and to studies of head-related transfer functions (HRTFs) [3]. The HRTF relates the sound signal at the ear to the sound signal at the source. Its inverse Fourier transform is called the head-related impulse response (HRIR). If the sound signal is convolved with the listener's own left-ear and right-ear HRIRs and the results are presented over properly compensated headphones, the listener will experience the perception of the source at the corresponding location in space.

Because the sound has to be registered with the graphics, VR systems typically employ head trackers, modifying the binaural audio in real time to keep the spatial locations of the virtual sources fixed when the listener moves. This greatly reduces the problem of in-head localization. Unfortunately, the HRTF-based approach requires knowing the location and the sound signal for each source separately, making it difficult to impossible to use this approach for recording live sound. MTB solves this problem.

3. BINAURAL SOUND CAPTURE AND MTB

MTB is a generalization of traditional binaural sound capture and reproduction. It uses spatial sampling to solve the problem of responding to voluntary head motion [4, 5]. Instead of using two microphones in a dummy head, MTB employs M microphones. A typical MTB microphone array is shown in Fig. 1. In this example, eight microphones are uniformly spaced around the equator of a head-sized sphere. The listener wears a head tracker, so that at any instant the system can determine the locations of the listener's ears. If the ears happen to be coincident with a pair of microphones, the signals from those microphones are directed to the listener's headphones. In general, the listener's ears will be between two microphones, and an interpolation procedure is used to determine the headphone signals.

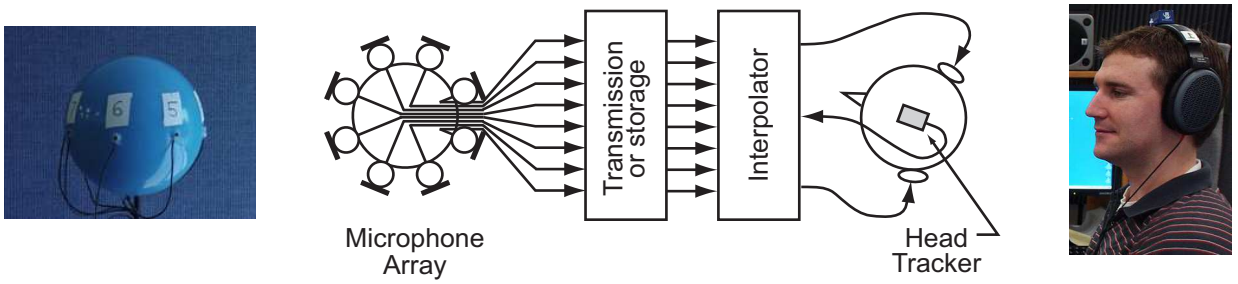

Fig. 1. The basic MTB system.

We have analyzed and experimented with a variety of interpolation procedures [5, 6]. In principle, one should have at least two samples per wavelength, which calls for about 128 microphones to cover the full audio bandwidth. With a practical number of microphones, the simplest approach of merely switching between microphones produces positional discontinuities and distracting clicks. Direct linear interpolation of microphone signals produces comb-filter spectral notches, where the frequency of the first notch is inversely proportional to the spacing between the microphones. This produces annoying "flanging" sounds when the listener turns his or her head. Neither of these procedures is satisfactory.

An effective compromise is to divide the microphone signals into low-frequency and high-frequency components. The high-frequency components are switched and the low-frequency components are linearly interpolated [5]. The frequency f_{max} that separates low- from high- frequency components is chosen to be below the frequency of the first comb-filter notch. This procedure exploits the fact that most of the energy in audio signals is in the low-frequency band, and high-frequency signals tend to be episodic rather than sustained, making the switching transients less audible. In practice, we find that this procedure produces high-quality results using 8 microphones for speech and 16 microphones for music.

Although MTB produces highly-realistic, well externalized spatial sound, the signals produced by this method only approximate the exact experience, and critical listening tests have revealed various audible defects [7]. We have developed methods to correct for these problem, if corrections are required, and refer the interested reader to [7] for an extended discussion of this topic.

4. IMPLEMENTATION ISSUES

In Section 5, we describe a number of potential applications for MTB. Each class of applications raises important practical considerations. These include such things as (a) the number of microphones needed, (b) the spatial locations of the microphones, (c) whether the sound is to be heard live, recorded, synthesized, or as a mixture, (d) compatibility with conventional audio systems, (e) conversion of conventional recordings to the MTB format, (f) whether conversion must be done in real time or can be done off line, (g) whether the sound is to be delivered to a single listener or broadcast to an audience of listeners, (h) the bandwidth requirements, (i) the storage requirements, (j) the possibilities for bandwidth and/or storage compression, and (k) the usual tradeoffs between computation, bandwidth, and storage. In this section, we discuss some of these considerations and describe some alternative approaches to implementation.

4.1 Spatial Sampling Strategies

Different applications have different requirements for spatial sampling. In the systems that we have built to date, the microphones are mounted uniformly around the equator of a sphere or a horizontal section of a vertical cylinder. This is appropriate for a common situation in which the sound sources of interest lie in or close to a horizontal plane, and the listener rotates his or her head about a vertical axis. However, other situations call for other spatial sampling strategies.

Because MTB interpolates between microphones that are close to the location of the ears, the general principle is that microphones should be located more densely in places where the listener's ears are more likely to be. For applications such as underwater exploration, all head orientations are presumed to be equally likely, and the microphones should be distributed uniformly over the surface of the sphere. Such applications are said to be omnidirectional (see Fig. 2a). When the sources of interest are primarily in the horizontal plane but the listener is equally likely to turn to face any azimuth, the microphones should be uniformly distributed around a horizontal equator. Such applications are said to be panoramic (see Fig. 2b). For applications such as reproducing musical performances, the sound sources define a

preferred facing direction, and the sampling around the equator can be sparse and non-uniform. Such applications are said to be frontal (see Fig. 2c).

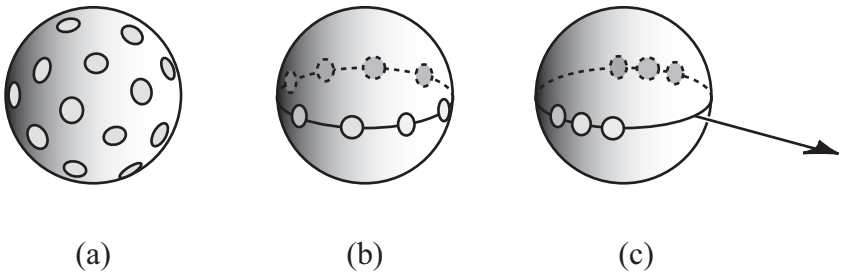

Fig. 2 Spatial sampling patterns. (a) omnidirectional, (b) panoramic, (c) frontal

Clearly, the number of microphones required for omnidirectional applications is greater than the number required for panoramic or frontal applications. In an earlier paper, we showed that the number of microphones needed for quarter-wavelength sampling was $(128\ \pi / 3)(a\ f_{max} / c)^2$ for omnidirectional sampling and $8\ \pi\ a\ f_{max} / c$ for panoramic sampling [5]. Here a is the radius of the sphere, c is the speed of sound, and f_{max} is the cutoff frequency for the low-pass filter used for interpolation, as described in Section 3. For the numerical values a = 0.0875 m, c = 343 m/s and f_{max} = 2.5 kHz, these formulas call for 55 microphones for omnidirectional and 16 microphones for panoramic sampling. With frontal sampling, the required number of microphones is still further reduced. In fact, a six-microphone solution appears to be adequate for this common situation, and – remarkably – the total required bandwidth is no more than that needed for 2.5 microphones [8].[3]

4.2 Virtual MTB

Up to this point, we have tacitly assumed that physical microphones would be used to capture the sounds of interest. However, a simulated MTB microphone array can be used to "capture" the sounds of

[3] Frontal applications also lend themselves to simple methods for introduced personalized corrections for optimum performance [8]. Thus, they represent a particularly attractive class of MTB applications.

synthesized sources. Let $h_m(t)$ be the impulse response of the m-th microphone in an MTB array to a sound source S. Then if $s(t)$ is the sound signal from the source, the microphone output is given by the convolution $x_m(t) = h_m(t) * s(t)$. For good externalization and musically pleasant sound, $h_m(t)$ should include the effects of the room. In simple cases (such as a spherical MTB array in a rectangular room), $h_m(t)$ can be computed. Alternatively, $h_m(t)$ can simply be measured.

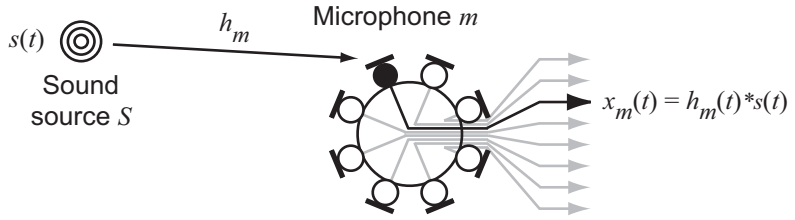

Fig. 3. Spatial sound synthesis using the MTB impulse response h_m for Microphone m.

The process of computing the microphone outputs can be computationally demanding. If there are N sources and M microphones, the process requires NM convolutions. However, this procedure provides an effective way to generate virtual auditory space,[4] and is particularly attractive for mixed-reality applications in which a small number of synthetic sources are combined with live or recorded spatial sound.

4.3 Real-time Conversion

Audio input can come from a variety of sources, such a live telephone conversations, computer synthesized sound, or pre-recorded material. In particular, there are a vast number of so-called legacy recordings of music in various formats – monaural, stereo, and surround-sound. The methods used to

[4] The number of convolutions required for HRTF-based methods is only $2N$. However, the impulse responses will change with head orientation, which requires more complex signal processing. Equally important, as we will shortly see, with MTB the convolutions can often be performed once and stored, whereas HRTF-based methods always require real-time convolution. Finally, MTB is decidedly more efficient when there are multiple listeners, because HRTF-based methods require separate real-time convolutions for each listener, whereas an unlimited number of listeners can be supported by the same set of MTB microphone signals.

generate virtual auditory space for MTB can also be used to convert their signals to the MTB format and gain the advantages that MTB provides.

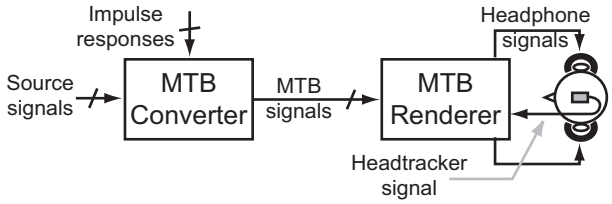

Fig. 4. Converting external source signals for MTB rendering in real time.

The basic process for real-time conversion is illustrated in Fig. 4. For legacy recordings, the source signals are the signals that were intended to drive the loudspeakers. Each impulse response converts the signal from a particular loudspeaker to the signal from a particular microphone. The MTB converter performs the required convolutions. The MTB renderer performs the head-motion-sensitive interpolation.

In general, convolution is much more computationally demanding than interpolation. Real-time conversion requires an MTB converter with the computational capacity to keep up with the input data rate. Furthermore, in some applications (such as home theater or computer games) there is an additional requirement for very low latency.

These issues are particular important for mobile multimedia. It is conceptually simplest to have all of the operations performed on the mobile device. In that case, a wireless link would be used to transmit the source signals to the mobile device. However, if the computational power of the mobile device is inadequate, it may be necessary to partition the operations differently.

One solution is to perform the conversion on a server, and to transmit the MTB signals to the mobile device for rendering (see Fig. 5). Another alternative would be to place the wireless link between the MTB renderer and the listener. This alternative solution requires that the head-tracker signal be transmitted back to the renderer, but it significantly reduces the overall transmission bandwidth. However,

the solution shown in Fig. 5 has the advantage that the MTB signals can be broadcast to any number of simultaneous listeners.

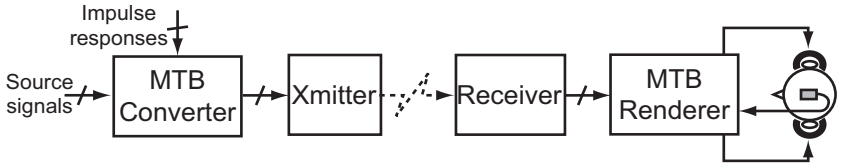

Fig. 5. Separating the conversion and the rendering processes by a wireless link.

4.4 Off-line Conversion

If real-time conversion is not required, the conversion process can be done off-line and the results can be stored. This might be appropriate for the conversion of legacy music recordings, or for the conversion of sound effects used in computer games.

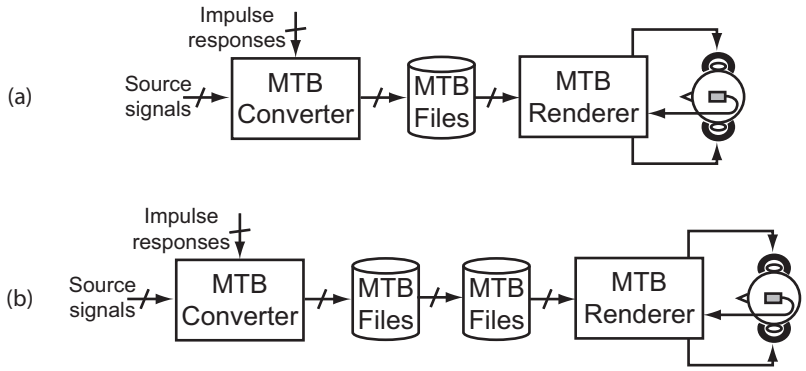

Fig. 6. Off-line MTB conversion. (a) Storage on the mobile device only. (b) Buffered storage.

As Fig. 6 illustrates, the conversion could be both done and stored on the mobile device. Alternatively, a large number of recordings could be done on a server, and a selection could be uploaded

to the mobile device. For music applications, the former solution would be suitable for recordings already owned by the listener, and the latter would be appropriate for a music service.

In all cases, the use of data compression is relevant, both for data transmission and for data storage. The simplest approach would be to use standard techniques to compress each channel independently. Because there are such strong correlations between the channels in MTB recordings, there are clear opportunities for additional savings in bandwidth. Although multichannel compression has received considerable attention, we have not investigated specific MTB compression algorithms, and this remains a topic for future research.

5. IMMERSIVE SPATIAL SOUND APPLICATIONS OF MTB

The potential value of audio interfaces for mobile systems has long been recognized [9, 10], and a number of research investigations have explored its potential use [11-14]. To fit on mobile devices, most of these systems have employed simple techniques, usually without head tracking, and the quality of the spatial audio has been limited. The new capability of "being there" that is provided by MTB for sound capture and reproduction enables a number of immersive sound applications. They can be broadly classified into three categories: (a) remote listening, (b) recording, and (c) broadcasting. We consider each of these briefly in turn.

5.1 Remote Listening

Audio teleconferencing is an obvious application for MTB, as is video teleconferencing, whenever the use of headphones is acceptable. MTB has a similar role in business presentation systems and collaborative-work systems. These are typically frontal applications, which can readily be customized to individual listeners for optimum performance. For all of these applications, the main advantages of MTB are the simplicity of sound capture and the significant improvement in clarity.

MTB can also be useful under water. In particular, it is well known that divers have difficulty locating each other by sound because the higher speed of sound under water leads to small interaural time

differences. If the radius of the array can be scaled appropriately, an MTB array could prove useful in underwater activities.

5.2 Recording

Home theater (including personal home theater) and surround sound for musical entertainment provide major potential applications for MTB. Both of these are frontal applications, and thus both can be customized to individual listeners for optimum performance. Furthermore, because the locations of the sound sources are known exactly, either generalized or individualized corrections can be added to control elevation and enhance static front/back discrimination.

With the growing popularity of digital camcorders and video editing software, amateur recording activities are expanding rapidly. The simplicity of capturing spatial sound with MTB makes it a natural candidate for the home video enthusiast.

When new recordings are made using MTB microphone arrays, new techniques will be developed for the professional production and post-production of audio material with MTB reproduction in mind. Until then, the techniques described in Section 3 can be used to convert recordings in standard formats. Thus, the valuable assets of legacy recordings can be preserved and enhanced.

5.3 Broadcasting

The extension of MTB to radio broadcasting is conceptually quite simple. We are currently engaged in the development of several versions of MTB with options as to the sound source material, the number of microphones, the head tracker algorithm, and the extent of head rotation. For some of these options the requirements with respect to bandwidth are very modest, and may be suitable for broadcast applications [8]. Further, broadcast applications may be considered as modifications of the portable MTB player where most of the storage and processing requirements have been moved from the portable player to the transmitter. The processing requirement with respect to spatial sound rendering with head-tracking, which is quite small, remains the same. A new requirement is the ability to transmit and receive high-quality

multichannel sound signals. This technical problem has already been solved for FM radio, but new digital technology may emerge that would be suitable for satellite radio and provide a higher quality. Note that a radio link is not necessary in broadcasting, and that any audio streaming technology, such via the Internet, may be used. By contrast to “on-the-spot” format conversion, streaming MTB does not require local format conversion or extensive storage, and is therefore quite similar to a radio-link usage.

6. IMMERSIVE INTERACTIVE MULTIMEDIA

Immersive spatial sound plays an obvious role in telepresence applications [3, 15-18]. The following multimedia applications are practical extensions of video technology that could capitalize on the sound immersion capabilities of MTB.

- Teleoperations. This is potentially valuable for the remote operation of equipment. An MTB array mounted on a remote-controlled land vehicle would allow one or more operators to hear the events in the neighborhood of the vehicle just as if they were actually present.
- Surveillance. MTB can expand the functionality of surveillance and security systems. These are panoramic or even omnidirectional audio/video applications, where spatial sound can be used to alert the user to unexpected activities that are not currently being viewed. In using MTB to locate the direction to a sound source, the angular resolution can be increased by using two concentric MTB arrays, the first with normal head size and the second with a larger radius. Once the general direction is determined by the smaller array, the listener can “home in” on the source by turning his or her head to null the larger interaural time difference produced by the larger array.
- Games. MTB offers a simple and effective way to enhance video and computer games, whether single-player or multi-player games. This is potentially one of the largest commercial applications for MTB technology.

- Augmented reality. Systems that combine remote listening and virtual auditory space provides a particularly attractive application of MTB technology [19-22]. Live sounds acquired directly by an MTB array, and combined with the virtual sounds that can be efficiently rendered in MTB format can be combined with video and computer graphics for applications such as training.

Computation, transmission, or data storage requirements may preclude these applications from being implemented on mobile platforms. However, it is possible to envision a new class of applications that would capitalize on the comparatively low bandwidth and modest processing requirements of MTB technology, and augment the immersive audio with images capable of being displayed on mobile platforms.

7. A PANORAMIC INFORMATION SYSTEM: SPATIAL SOUND FOR SMALL IMAGE WINDOWS

Although the display of stored or transmitted video on very small screen may be of some appeal for entertainment that the provision of spatial sound would enhance further, we focus on some more utilitarian applications of mobile multimedia. We will expand on such applications in the context of spatial sound immersion and focused vision.

It is common experience that we are aware of the immersion in the sounds that surround us, but that we focus our visual attention to a small sector of space. For television for instance, the display window is typically eight degrees of cone angle. For a mobile display, the subtended angle of the screen is even smaller. A natural role for sound is to direct the listener's attention to objects or events that are not currently being viewed. Once a person turns to face a new direction, the presentation on a small display of limited visual information by graphs, sketches, 3-D graphics, and still or moving images may be sufficient. Because the selection of information is determined by the application, we will expand on examples to illustrate the concepts.

In the following examples, we will augment MTB spatial sound with wireless and portable technologies already in widespread use:

- GPS. The global positioning system provides a reliable positioning and low accuracy orientation system that uses small and inexpensive receivers. GPS also provides a local index or remote wireless access to geographic information such as electronic maps.
- Electronic map databases. These are in widespread use and are now available for download to cell phones.
- MPEG4. A low-bandwidth standard for images and graphical objects, MPEG4 is well suited for the very efficient representation of multimedia information such as images of arbitrary shape, 2-D and 3-D meshes, and textures that can be simplified to adapt to the available bit rate. Audio representation and compression and the creation of synthetic music and sounds are also included in the MPEG 4 standard.

We consider three applications that combine spatial sound with maps, 3-D graphics, images and video.

- Multimedia map service and navigator for pedestrians. This augmented navigation system takes advantage of the orientation of the user. The geographic information or map can now be segmented to provide a user-centered view of the information. First, it now becomes possible to display a more detailed sector of the map with its apex at the position of the user and with an orientation matching the direction of his or her head as determined by the head-tracker. Panoramic audio could provide the names of the major streets or landmarks on demand, with the sound coming from the proper direction. This capability can also be augmented by 3-D graphics. Since the maps are built around the user's viewpoint, a flat map view can be replaced by a 3-D labeled view with outlines of buildings and streets or other landmarks.
- Tour guides. This is a variation on the previous application, where the modes alternate between navigation and narrative, the switching between modes being based on the behavior of the user. For instance, if the user is moving, aural information is provided on notable landmarks surrounding the user. If the user faces a landmark more than a preset time, the mode changes to a prepared narrative about the landmark.

- Multisite video conferencing. Although the transmission of single head-and-shoulder video is now well within the reach of mobile systems, a conference with several participants is not currently possible. In a multi-user video conferencing, spatial sound lets the audio information be properly localized and presented to each participant in a spatially well differentiated fashion. This can be achieved because MTB provides an efficient method for rendering a number of spatially distinct sounds. However, in the system that we propose here the video displayed on a small portable screen is the view of a single participant that is automatically chosen by the recipient when he or she faces the direction of that person's voice. This style of interaction is quite natural, in that participants can converse naturally and pay attention in turn to other participants of their choice in a mode that improve both the participation in the group and speech comprehension.

8. DISCUSSION

The applications we have discussed illustrate the combined use of panoramic spatial sound and other media in new mobile applications. The sources of information include low-bandwidth data (such as electronic maps, spatial sound, and limited video) provided by sources such as a base station, a GPS system, and databases stored on the mobile system. Spatial sound rendering and 3-D graphics generation will probably be done on the mobile platform. As usual, there will be different transmission/storage tradeoffs depending on the transmission bandwidth and the storage or processing capabilities of the portable system. Clearly, substantial creativity is required to convert this vision to reality. We present it merely to point out that the ability of MTB to produce convincing immersive audio with modest computational requirements provides new opportunities for immersive portable and interactive multimedia.

9. CONCLUSIONS

By exploiting spatial sampling and the powerful perceptual cues provided by voluntary motion, MTB can generate a highly effective, immersive, headphone-based sound experience with relatively small computational cost, making it particularly attractive for use in mobile systems. The technology applies

equally well to virtually any type of sound source – live, recorded, synthesized, or a mixture of such sounds.

For real sounds, a limitation of the technique is that it applies only to head rotation but not head translation, corresponding to the physical situation of a listener standing at the position of the MTB array. Recordings can be made in which the listener is moved through a space, but the listener cannot choose an arbitrary path interactively. From a commercial standpoint, the need for a head tracker that is sufficiently accurate, drift-free, light weight, small, rugged, reliable, efficient in its use of electrical power, and inexpensive presents a short-term challenge, but one that will be rapidly solved by advances in sensor technology.

Besides sensor technology, the multimedia systems that we have reviewed or proposed are well within the reach of the transmission and processing performance becoming available for portable multimedia systems, such as advanced cell phones and personal digital assistants. The central role that we propose for spatial sound in structuring multimedia into an natural and intuitive experience thus depends on the effectiveness of the MTB technology in providing a stable immersion in spatial sound and on head-tracking as a simple control interface.

10. ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation under grants IIS-00-97256 and ITR-00-86075. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the view of the National Science Foundation.

11. REFERENCES

[1] Rumsey, F., *Spatial Audio*, Focal Press, Oxford, England, 2001.

[2] Davis, M. F., "History of spatial coding," *J. Aud. Eng. Soc.*, vol. 51, no. 6, pp. 554-569, June 2003.

[3] Begault, D. B., *3-D Sound for Virtual Reality and Multimedia*, AP Professional, Boston, MA, 1994.

[4] Algazi, V., Duda, R. O. and Thompson, D. M., "MTB – A method and means for dynamic binaural sound capture and reproduction, " US patent application 20040076301, April 22, 2004.

[5] Algazi, V., Duda, R. O. and Thompson, D. M., "Motion-tracked binaural sound," *J. Aud. Eng. Soc.*, vol. 52, no. 11, pp. 1142-1156 , Nov. 2004.

[6] Hom, R., "On the interpolation of microphone signals in motion-tracked binaural sound," MS Thesis, Dept. of Elec. and Comp. Engr., UC Davis, Davis, CA, 2005.

[7] Melick, J., Algazi, V. R., Duda, R. and Thompson, D., "Customization and personalized rendering of motion-tracked binaural sound," Paper 6225, 117th Convention of the Audio Engineering Society, San Francisco, CA, Oct. 2004.

[8] Algazi, V. R., Dalton, R. J., Duda, R. O. and Thompson, D. M., "Motion-tracked binaural sound for personal music players," Paper 6557, 119th Convention of the Audio Engineering Society, New York, NY, Oct. 2005.

[9] Kramer, G., ed., *Auditory Display: Sonification, Audification, and Auditory Interfaces.* Addison Wesley, Reading, MA, 1994.

[10] Gaver, W. W., "Auditory Interfaces," in Helander, M., Landauer, T. K. and Prabhu, P., eds., *Handbook of Human-Computer Interaction, 2nd Ed.*, pp. 1003-1041, Elsevier Science, The Netherlands (1997).

[11] Mynatt, E. D., "Audio Aura: light-weight audio augmented reality," *UIST 97* (Tenth ACM Symposium on User Interface Software and Technology), Banff, Alberta, Canada, 1997.

[12] Tannen, R. S., "Breaking the sound barrier: designing auditory displays for global usability," Proc. 4th Conference on Human Factors and the Web, http://www.research.att.com/conf/hfweb/ proceedings/tannen/, June 1998.

[13] Lyons, K., Gandy, M. and Starner, T., “Guided by voices: An audio augmented reality system,” *ICAD 00* (International Conference on Auditory Display), Atlanta, GA, April 2000.

[14] Sawhney, N. and Schmandt, C., “Nomadic radio: Speech and audio interaction for contextual messaging in nomadic environments,” *ACM Transactions on Computer-Human Interaction*, vol. 7, pp. 353-383, Sept. 2000.

[15] Baker, H. H., Tanguay, D., Sobel, I., Gelb, D. and Goss, M. E., “The Coliseum immersive teleconferencing system,” ITP2002, Juan Les Pins, France, Dec. 2002.

[16] Aoki, S., Cohen, M. and Koizumi, N., “Design and control of shared conferencing environments for audio telecommunication using individually measured HRTFs,” *Presence*, vol. 3, no. 1, pp. 60-72, 1994.

[17] Hollier, M. P., Rimell, A. N. and Burraston, D., “Spatial audio technology for telepresence,” *B. T. Technol J.*, vol. 15, no. 4, pp. 33-41, 1997.

[18] Rimell, A., “Immersive spatial audio for telepresence applications: system design and implementation,” Proc. AES 16th International Conference on Spatial Sound Reproduction, Rovaniemi, Finland, Apr. 1999.

[19] Cohen, M., Aoki, S. and Koizumi, N., “Augmented audio reality: Telepresence/VR hybrid acoustic environments,” Proc. 2nd IEEE International Workshop on Robot and Human Communication, pp. 361-364, New York, NY, 1993.

[20] Billinghurst, M. and Kato, H., “Collaborative mixed reality,” in Ohta, Y. and Tamura, H., eds., *Mixed Reality: Merging Real and Virtual Worlds*, pp. 261-284, Ohmsha Springer Verlag, 1999.

[21] Eckel, G., “Immersive audio-augmented environments - The LISTEN Project,” Banissi, B., Khosrowshahi, F., Sarfraz, M. and Ursyn, A., eds., *IV2001* (Proc. 5th International Conference on Information Visualization), IEEE Computer Society Press, Los Alamitos, CA, 2001.

[22] Härmä, A., Jakka, J., Tikander, M., Karjalainen, M., Lokki, T., Hiipakka, J. and Lorho, G., “Augmented reality audio for mobile and wearable appliances,” *J. Aud. Eng. Soc.*, vol. 52, no. 6, pp. 618-639, June 2004.