VIRTUAL ACOUSTICS - APPLICATIONS AND TECHNOLOGY TRENDS

Jyri Huopaniemi
Nokia Research Center
Speech and Audio Systems Lab.
jyri.huopaniemi@nokia.com

Lauri Savioja, Tapio Lokki
Helsinki Univ. of Tech.
Telecommunications Software
and Multimedia Laboratory
{lauri.savioja,tapio.lokki}@hut.fi

Riitta Väänänen
Helsinki Univ. of Tech.
Laboratory of Acoustics
and Audio Signal Processing
riitta.vaananen@hut.fi

ABSTRACT

Virtual acoustics is a general term for the modeling of acoustical phenomena and systems with the aid of a computer. It comprises many different fields of acoustics, but in the context of this paper the term is restricted to describe a system ranging from sound source and acoustics modeling in rooms to spatial auditory perception simulation in humans. In other words, virtual acoustics concept covers three major subsystems in acoustical communication: source, transmission, and the listener. This paper discusses virtual acoustics from the point of view of applications and technology trends.

1 INTRODUCTION

The topic of this paper is virtual acoustics. It can be explained as physically-based or perceptually-based modeling of sound sources, room acoustical systems and human spatial hearing primarily by means of digital signal processing (DSP). Thus, virtual acoustics deals with the three major subsystems in acoustical communication: source modeling, transmission medium modeling, and listener modeling [1]. The goal in virtual acoustics is to deliver any acoustical message (voice, music, noise or combinations thereof) in a virtual reality system from the source to the receiver as it would happen in a real or imaginary acoustical situation.

This tutorial paper gives an overview of current virtual acoustic technology, presents state-of-the art research results, and looks at the current and future technology impact of this research area. The paper is organized as follows. First, background is supplied to a broad virtual acoustics concept and the underlying subproblems. In the following section, we concentrate on virtual acoustic environments, looking at specific application areas of virtual and augmented reality. Applications and technology trends are discussed next, with emphasis on current and emerging technologies. Finally, conclusions are drawn and suggestions for future work are given.

2 BACKGROUND TO VIRTUAL ACOUSTICS

The history of virtual acoustics can be roughly divided into the pre-computer era (i.e. research carried out before 1950's) and the modern research (the main topic of this paper) which relies on the use of digital signal processing. Of the three main categories of virtual acoustics, analog source modeling could be considered historically the forerunner. Research into artificial speech production started already in the 1700's by von Kempelen and Kratzenstein [2] who created the first "talking machines". The great acousticians of 1800's, von Helmholtz and Lord Rayleigh, contributed apart from theoretical research also to many areas of applied acoustics, including room acoustics, musical acoustics, spatial hearing and hearing in general. First computational formulae for describing room acoustics were proposed in early 1900's by Wallace Sabine, and for spatial hearing the first formal theory was proposed at the same time by Lord Rayleigh. It was, however, not until 1960's before the present understanding and concepts of virtual acoustics started to develop with the aid of digital signal processing (DSP) and computer technology. In the late 1980's the first consumer real-time applications featuring source, room and listener models were introduced [3]. Since that time, the advances in virtual acoustics research and applications have been very rapid.

Acoustics as a science has traditionally been approached from a *physical* point of view, investigating the behavior of sound generation, sound propagation in an enclosure, and sound pickup by acoustic or electroacoustic means. On the other hand, acoustics also has a strong *perceptual* component - it is the human ears and brain who at the end process and evaluate the acoustical message. The psychoacoustical and perceptual aspects are increasingly being reflected in research into virtual acoustics. In the following section, the modern virtual acoustic concept is discussed.

2.1 Virtual Acoustic Concept

This section discusses an object-oriented virtual acoustic concept. A communication framework for virtual acoustics is shown in Fig. 1 (modified from [4]). The signal path consists of analysis, transmission, and synthesis parts. Within these parts, the source, medium and listener models are used accordingly. Sound source modeling has been an active area of research in the digital era, traditionally in speech and audio coding, but also in sound synthesis and recently in complex tasks such as *auditory scene analysis* (ASA) [5]. In Fig. 1, the analysis of a sound mixture, carried out in ASA, involves the segregation and recognition of elementary sounds. In the optional representation part, the sound objects are coded and the extracted information can be attached using metadata. For transmission, streaming and download is utilized, aiming at real-time and non-real-time applications, respectively. Search and retrieval techniques are often applied when querying sounds, for example, from a database. In the synthesis framework, the audio objects are decoded and different synthesis methods can be applied to feed the audio signal to the environment and finally to the receiver model.

In this paper, the source modeling aspects are not discussed in great detail. The reader is referred to [6, 7] for further information on the sound source modeling area. In the following, we elaborate more on the research of virtual acoustic environments, and specifically the room and listener models thereof.

3 VIRTUAL ACOUSTIC ENVIRONMENTS

In this section and for the remainder of the paper, we limit the discussion to a sub-category of virtual acoustics called virtual acoustic environments [1]. The virtual acoustic concept is closely related to other media for many reasons. The audiovisual technology (audio, still picture, video, animation etc.) is rapidly integrating into a single interactive media. Research on audiovisual media modeling has increased dramatically in the last decade. Standardization of advanced multimedia and virtual reality rendering and definition has been carried out for several years. Such examples are the Moving Pictures Expert Group MPEG-4 [9] and the Virtual Reality Modeling Language (VRML) [10] standards. The progress of multimedia has also introduced new fields of research and application into audio and acoustics, one of which is virtual acoustic environments and their relation to graphical presentations.

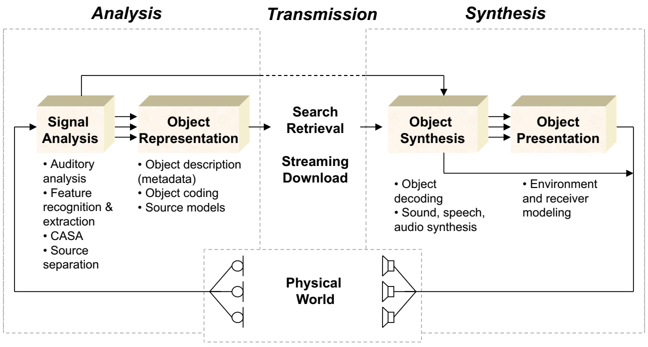

Figure 1: Framework for virtual acoustics research (after [4]).

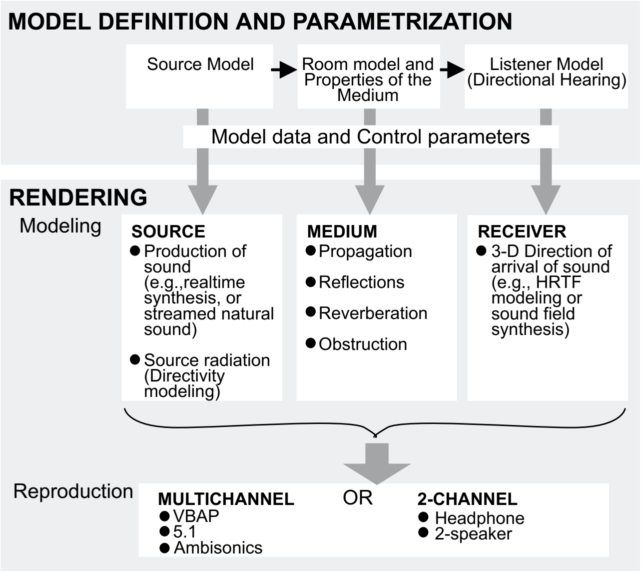

Figure 2: Modeling concept of virtual acoustics [11, 12].

The design and implementation of virtual acoustic displays (VADs) can be divided into two main tasks as depicted in Fig. 2 [11, 12]: a) model definition and parametrization, and b) rendering. The model definition of a virtual acoustic environment includes the prior knowledge of the system to be implemented, that is, information about the sound sources, the room geometry, and the listeners. The definition part is normally carried out off-line prior to the real-time simulation process. In the rendering part the modeling task is divided into source, room and listener modeling as discussed previously. For rendering, there are many choices, normally divided into multichannel or two-channel cases. The term *auralization* [13] is understood as a subset of virtual acoustics referring to modeling and reproduction of sound fields (shown as rendering in Fig. 2).

3.1 Source Modeling

Source modeling consists of methods which produce (and possibly add physical character to) sound in an audiovisual scene. The most straightforward method is to use pre-recorded digital audio. In most auralization systems sound sources are treated as omnidirectional point sources. This approximation is valid for many cases, but more accurate methods are, however, often needed. For example, most musical instruments have radiation patterns which are frequency-dependent. Typically sound sources radiate more energy to the frontal hemisphere

---

[1] Virtual acoustic environments are also called virtual auditory displays, auditory virtual environments, virtual auditory space [1, 8].

whereas sound radiation is attenuated and lowpass filtered when the angular distance from the on-axis direction increases.

3.2 Room Modeling

Computers have been used to model room acoustics from the 1960's. The first ray-tracing program was published in 1968 by Krokstad et al [14]. Since that there has been lots of progression in the area. The main simulation techniques have been the ray-tracing and the image-source methods which both are based on geometrical room acoustics. In the late 1980's, auralization was introduced for allowing to hear the simulation results [15]. During the last ten years there has been a major breakthrough in using computers to help design concert halls. They have replaced the traditional scale models in most of the design cases, although the scale models are still used in some major constructions together with computer models.

The task of modeling sound propagation behavior in acoustical spaces (Fig. 2) is an integral part of virtual acoustic simulation. The goal in most room acoustic simulations has been to compute an energy time curve (ETC) of a room (squared room impulse response) and, based on that, to derive room acoustical attributes such as reverberation time (RT_{60}). The ray-based methods, the ray tracing [14, 16] and the image-source methods [17, 18], are the most often used modeling techniques. Recently computationally more demanding wave-based techniques such as finite element method (FEM), boundary element method (BEM), and finite-difference time-domain (FDTD) methods have also gained interest [19]. These techniques are, however, suitable only for low-frequency simulation [13]. In real-time auralization, the limited computational capacity calls for simplifications, modeling only the direct sound and early reflections using geometrical acoustics and rendering the late reverberation by recursive digital filter structures [20, 21, 22, 23].

It should be noted that although detailed modeling of sound transmission (medium in Fig. 2) has an important role, many applications do not necessarily target for simulating of a reverberating, enclosed space such as a room or a concert hall. It may also be desirable to model simply the sound transmission in the air (inlcuding effects such as distance-dependent attenuation, air absorption, or Doppler effect), or effects such as occlusion caused by obstructing objects between the sound and the listening point, or individual reflections, echoes, or reverb. This is specifically the case in modern real-time virtual reality applications, such as computer games, where there may be dynamic movement of the sound source and the listener, for example, that result in time-varying acoustics. In this type of applications it may not be necessary to hear detailed room acoustics but just simple environmental acoustic effects that increase the immersiveness of an audiovisual application.

In [12, 11, 19], issues in real-time modeling and rendering of room acoustics are discussed. Conceptually, a division can be made between a *physical modeling* approach and a *perceptual modeling* approach. The physical modeling approach aims at capturing both the macroscopic (reverberation time, room volume, absorption area, etc.) and microscopic (reflections, material absorption, air absorption, directivity, diffraction, etc.) room acoustical features by a geometrical acoustics approach, and diffuse late reverberation using statistical means. The goal in perceptual parametrization, on the other hand, is to find an orthogonal set of parameters using which a virtual acoustic rendering algorithm can be controlled to produce a desired auditory sensation. Therefore the output is user-controllable, not environment-controllable as in the physical approach. The perceptual approach uses physical properties for the direct sound, macroscopic room acoustical features and a static image-source method for early and late early reflections, and a statistical late reverberation module for late reverberation [22].

This division clearly separates the two main application goals for parametric room impulse response rendering techniques. The first approach is more suitable for accurate simulation of spaces such as concert halls, auditoria, and for auditorily accurate audiovisual rendering, such as modeling of environemntal audio effects that are consistent visual objects in virtual reality scenes. The latter approach provides a possibility to create a high-quality room acoustic response perhaps more efficiently based on intuitive, perceptually controlled parameters. Therefore the perceptual rendering scheme suits better for applications aiming at creating room acoustic effects, such as, post-processing of sound or music perfromances, and teleconferencing.

3.3 Listener Modeling

In listener modeling (Fig. 2), the properties of human spatial hearing are considered. Simple means for giving a directional sensation of sound are the interaural level and time differences (ILD and ITD), but they cannot resolve the front-back confusion[2]. The head-related transfer function (HRTF) that models the reflections and filtering by the head, shoulders and pinnae of the listener, has been studied extensively during the past decade. With the development of HRTF measurement techniques [26, 27] and efficient filter design methods [28, 11] real-time HRTF-based 3-D sound implementations have become applicable in virtual environment synthesis.

3.4 Methods for 3-D Sound Reproduction

The illusion of three-dimensional sound fields can be created using various methods. The goal in 3-D sound simulation is to recreate any static or dynamic natural or

---

[2] See [1, 24, 25, 13] for fundamentals on spatial hearing and auralization.

imaginary sound field using a desired amount of transducers and proper signal processing techniques. The reproduction methods can be divided into three main categories due to their inherently different processing requirements: 1. multichannel reproduction, 2. two-channel reproduction, and 3. monophonic reproduction In the following, the first two categories for reproduction are shortly described (the third one is trivial, containing only one channel for playback).

3.5 Two-Channel Reproduction

The two-channel stereophonic reproduction methods enable replication of natural human spatial hearing cues. This is achieved using binaural technology, which utilizes HRTFs measured on human subjects or dummy heads (or derived by mathematical models).

3.5.1 Headphone Reproduction

In the case of binaural reproduction for headphones, the ear signals provided by HRTFs may directly be used as digital filters if proper equalization is carried out. Known advantages of binaural reproduction are the (trivial) facts that the acoustics of the listening room and the positioning of the listener do not affect the perception. On the other hand, head-tracking is required to compensate for head movements, and possibly to reduce front-back localization errors. Furthermore, it has been found that individual HRTFs should be used to create the best performance in localization and externalization [29, 30]. It can be concluded that headphone 3-D audio gives remarkable added value to traditional stereo by externalization of the sound image.

3.5.2 Loudspeaker Reproduction

Loudspeaker reproduction of binaural information is critically different from headphone reproduction. The directional characteristics that are present in binaural signals are exposed to crosstalk when loudspeaker reproduction is used. Crosstalk occurs because the sound from the left loudspeaker is heard at both left and right ears, and vice versa. The theory of crosstalk canceling for binaural reproduction was presented in the 1960's by Bauer [31] and practically implemented by Atal and Schroeder [32], and has been later investigated intensively by the research community and industry. A taxonomy of crosstalk canceling technologies is given in [33]. Crosstalk canceled binaural systems have two main limitations: 1) critical listening position, and 2) critical listening room conditions. The full spatial information can be retained only in anechoic chambers or listening rooms and the "sweet spot" for listening is very limited. On the other hand, at its best crosstalk canceled binaural reproduction is capable of auralizing a sound field in a very convincing and authentic way.

3.6 Multichannel Techniques

A natural choice to create a two- or three-dimensional auditory space is to use multiple loudspeakers in the reproduction. With this concept the problems in retaining spatial auditory information are reduced to placement of the N loudspeakers and panning of the audio signals according to the direction [34]. The problem of multiple loudspeaker reproduction and implementation of panning rules can be formulated in a form of vector base amplitude panning (VBAP) [35] or by using decoded 3-D periphonic systems such as Ambisonics [36, 37]. The VBAP concept introduced by Pulkki [35] gives the possibility of using an arbitrary loudspeaker placement for three-dimensional amplitude panning. The rapid progress of multichannel surround sound systems for home and theater entertainment during the past decade has opened wide possibilities also for multiloudspeaker auralization. Digital multichannel audio systems for movie theaters and home entertainment offer three-dimensional spatial sound that has been either decoded from two-channel material (such as Dolby ProLogic) or uses discrete multichannel decoding (such as Dolby Digital). The ISO/MPEG-2 AAC audio coding standard offers 5.1 discrete transparent quality channels at a rate of 320 kb/s (compression rate 1:12) [38]. Another multichannel compression technique that is already widespread in the market is Dolby Digital which provides similar compression rates to MPEG-2 AAC [38].

A general problem with multichannel ($N > 2$) sound reproduction especially in domestic setups is the amount of needed hardware (6 speakers for a discrete 5.1 surround setup) and their placing. On the other hand, intuitively, the listening area should be larger and the localization effect more stable than with two-channel binaural loudspeaker systems. Another advantage of multichannel reproduction over binaural systems is that no listener modeling (use of HRTFs) is necessary, and thus it is in general computationally less demanding.

4 TECHNOLOGIES AND APPLICATIONS

Virtual acoustics has had a significant impact in multimedia technology in recent years. The applications that have gained most attention are in entertainment and games. Such examples are a multitude of software for 3-D audio for different computer platforms (including the corresponding Application Programming Interfaces, API's), as well as home theater systems utilizing virtual surround technologies, and emerging virtual reality applications. On the other hand, research into virtual auditory displays in user interfaces and telecommunications applications has also increased rapidly.

In this section, some of the technologies are briefly discussed that have made possible the increasing use of virtual acoustics in many of the afore-mentioned applications. First an overview of the API's included in various programming language developer applications for low-level control over the spatial sound functions is given. The *Scene Description API's* that allow a higher-level presentation and hierarchically structured programming of interactive audiovisual scenes are also discussed. Available commercial applications of virtual acoustics are presented and some more scientifically interesting projects are overviewed.

4.1 Technologies

The use of virtual acoustics in various audio or audiovisual applications has increased in recent years, due to research as well as computational advances. In the dawn of real-time 3-D audio and virtual acoustic applications (in late 1980's and early 1990's), most of the applications used dedicated hardware. What has clearly affected the most the popularity of virtual acoustics in commercial applications, is the increase of computational capacity of computers. This has resulted in the broad range of processing capabilities that sound cards and other sound processing modules nowadays provide. In most cases today the audio signal processing capabilities of multimedia PC sound systems comprise support of 3-D sound and modeling of virtual acoustic effects to some detail. Although sound cards and other dedicated hardware devices still use separate DSP, 3-D audio features can be easily realized even with a general-purpose CPU of a standard PC.

What is also required for practical implementations of virtual acoustics are the higher-level controls such as API's. These have been created in order to easy-to-use interfaces to the actual sound processing routines. This is a necessity when applications are built by non-experts that simply want to increase the immersiveness of their 3-D audiovisual software products, for example. By 3-D sound API's we mean ones that consist of a set of methods for including sounds to 3-D applications. These methods are given spatial sound attributes as parameters that are used to control the sound processing events when the program is executed. Low-level API's include those that are formed of a set of methods within a programming language such the DirectSound3D within the Microsoft DirectX API for C++. *Scene Description Languages*, on the other hand, are higher-level API's than those that are implemented as programming language libraries, and are used to build 3-D audiovisual scenes composed of hierarchically structured objects often called *nodes* that represent functional units in the scene that is finally presented at run time. These API's can be characterized as interfaces for building up interactive 3-D applications with a small set of parameters, and to be used intuitively also by non-expert programmers. Typical examples of Scene Description API's are the Virtual Reality Modeling Language, VRML97 [10], and the Binary Format For Scenes (BIFS) in MPEG-4 standard [9], discussed in more detail below. See Table 1 for some of the available 3-D APIs with references.

Typically the 3-D sound programming and scene description API's nowadays allow features such as assigning a 3-D spatial position and simple non-uniform directivity characteristics with a sound source, as well as adding other spatial effects such as distance dependent attenuation or reverberation to sounds. This type of processing framework allows for many different interactive applications. For example, the user can navigate in the 3-D scene with the aid of an input device. Then it depends on the relative positions of the source location and the virtual listening point and interactive movements in the local coordinate system of the renderer, how the sound is perceived. The actual sound that is processed spatially is usually stored in a file and streamed from it at run-time. However, as a specification for compressed presentation and decoding of audio and visual data, MPEG-4 enables associating streamed and real-time decoded audio to 3-D scene objects. The reader is referred to a companion paper [39] for more details on the MPEG-4 virtual acoustic processing specification.

Table 1 lists some of the APIs and projects that involve immersive 3-D sound in audio and audiovisual applications.

4.2 Applications

The applications of virtual acoustics can roughly be divided to non-realtime, realtime and interactive ones depending on the need of processing power and the types of technologies involved in these applications. The division can also be done according to the type of application. Entertainment can be considered the biggest application area, but telecommunications and auditory display (including user interface) areas are also emerging.

The *non-realtime* applications might use very elaborated processing methods, but processing power is not usually a restricting issue, because the simulations can be done as batch jobs. Such applications of virtual acoustics include, e.g., room acoustics modeling software systems used as design tools, and spatial effects processors when used for offline post-processing of sound recordings. Also, high-quality virtual reality rendering applications would fall into this category.

In *realtime* applications such as teleconferencing or interactive TV, the sound is streamed from a server or audio data storage, such as compressed audio file or a DVD. In these applications the sound has to be spatially processed and played back in real time, although with a latency depending on the input and the output buffers of the processing device. The emphasis is placed on efficient implementation of the effects, and depending on the available processing power compromises may have to be made in the quality or detailness of the modeled spatial features. Another group of realtime applications is pro audio effects, for example, reverbs and other spatial processors used in studios.

The biggest and from a technology point of view maybe the most challenging group of realtime applications are those that are controlled *interactively.* These differ from the above mentioned realtime applications in that they essentially assume controllability of the sound processing routines with an immediate response without audible artifacts when the processing parameters are changed. Typical applications are games and augmented or virtual reality where the user can affect his/her viewpoint in the virtual world and thus also the relative position with respect to the sound sources. In virtual reality applications important issues are consistency between the visual and the sound scene, low enough a latency in response to user or source movements, and careful design in the way that the parameters of the spatial sound processing routines are updated.

5 FUTURE CHALLENGES

Virtual acoustics has still many challenges both in research and utilization of the research results in applications. In this section, a brief look at some of the challenges is provided.

5.1 Psychoacoustic aspects

One of the topics that has not yet been extensively examined when looking for optimized algorithms for spatial sound processing, is the psychoacoustic or subjective relevance and the corresponding accuracy of the modeled features. In many cases when modeling a room response that is based on a geometrical description of the acoustic space the target is to model the response in as detailed a fashion as possible (e.g., with maximum numbers of early reflections with best possible late reverberation). However, a single room acoustic response for a sound includes many features that can be controlled independently with a large set of parameters. The psychoacoustical relevance of some of these have been examined independently, but to be able to define what components in an interactive audio- or audiovisual scene are audible and relevant at each moment of time, more psychoacoustic research need to be carried out. Generally, optimization of signal processing systems using auditory criteria and understanding of the underlying psychoacoustics and human hearing is likely to be a continuing hot topic in virtual acoustics (as well as other signal processing) in future research.

5.2 Computational and practical aspects

New applications that take advantage of virtual acoustics evolve as the technology matures and the research in the area results in more efficient and optimized algorithms and implementations. What still restricts the use of advanced virtual acoustics modeling is the processing power available for sound DSP, as well as processing and operating systems issues such as latency related to the control input and buffering. Although processing power of computers continuously increases, detailed sound environment modeling usually requires more power than is currently available in a single computer. Also, in the case of audiovisual applications the trend traditionally has been to use a larger amount of the increasing processing power in image than sound processing. Thus, optimization of sound processing routines is one of the key issues when more modeled features, or several sound sources are included in the same virtual 3-D sound environment.

The rendering and reproduction of 3-D sound is also not a straightforward task since many of the methods, such as the HRTF modeling or even multichannel panning techniques may require individualized design for different listeners, and also adaptation to the reproduction technique and the listening space. Generalized rendering models are therefore an active future research area. This also includes the user interaction with the rendering tools, for example, using advanced controls like head-tracking and other sensor-based data.

One of the issues that restricts the popularity is the platform-dependency of implementations of the existing sound processing API's and programs. This remains the case as long as there is diversity in the products for playing and processing the sound (soundcards, platform dependent sound libraries). However, the developing standards and specifications tend to overcome this problem so that even in different platforms similar interfaces would exist to enable the portability of programs including 3-D sound processing.

5.3 Industry collaboration

The collaborative work of the research and industry is crucial to the field, because it often results in common practices and specifications, reducing the divergence of end-user platforms and software choices. As an example of a functional working group, an industry collaboration called *Interactive Audio Special Interest Group* (IASIG) has been established to develop and propose to the industry common guidelines and recommendations that should be included in modern sound applications and API's [40]. One of the groups of the IASIG is the 3-D Audio Working Group (IASIG 3DWG), which has published two sets of guidelines for sound and acoustic environment modeling in 3-D audio API's and applications. The 3-D audio rendering guidelines, Levels 1 and 2, include parametric representation of sound location, sound source radiation model, and sound propagation in reverberant environments. One of the major future challenges also for 3-D sound is the emerging of integrated media for use in interactive applications and over computer networks and mobile telecommunications. To be able to have virtual acoustics and 3-D audio in such applications, the transmission, storage, and decoding (decompression and presentation of data) have to be efficient, i.e., the data has to be low-bandwidth, and the decoding (including 3-D sound processing) must run in real time and with an acceptable latency. The parametric presentation of both the sound content, and

Scene description APIs	
VRML	http://www.vrml.org/
MPEG-4	http://www.cselt.it/mpeg
3D sound APIs	
JAVA3D	http://java.sun.com/products/java-media/3D/
OpenAL	http://www.openal.org/home/
Microsoft DirectX	http://www.microsoft.com/directx/
A3D	http://www.a3d.com/
EAX	http://www.env-audio.com/
QSound's Q3D	http://www.qsound.com/
Sensaura	http://www.sensaura.co.uk/
Research projects on real-time systems	
DIVA [12]	http://www.tcm.hut.fi/Research/DIVA/
Spatialisateur [22]	http://www.ircam.fr/produits/logiciels/log-forum/spat-e.html
VAS [41]	ftp://ftp.aic.nrl.navy.mil/pub/VAS/
AudioLab/CyberStage [42]	http://viswiz.gmd.de/~eckel/audiolab/index.html
Sound LAB [43]	http://vision.arc.nasa.gov/~bwenzel/index.html
Uni-Bochum [44]	http://www.ika.ruhr-uni-bochum.de/indexeng.htm

Table 1: Information on 3-D sound API's and research projects on real-time virtual acoustics.

the source model and virtual acoustic space associated with it, provide flexible possibilities for creating virtual acoustic applications for future media purposes.

6 CONCLUSIONS

In this paper, we have discussed the applications and technologies of virtual acoustics. The field of virtual acoustics has expanded rapidly in the past decade. Applications such as games and virtual environments, 3-D audio for consumer hi-fi and home theater applications, and physical and perceptual real-time rendering of room acoustics are examples of virtual acoustics in practice nowadays. The future challenges are mostly in further understanding of the psychoacoustical and perceptual aspects of virtual acoustics. Increased knowledge of the human auditory system will enable more efficient signal processing means for virtual acoustics, leading to more immersive applications. Industry collaboration is needed to produce standards and specifications enabling the wide use of virtual acoustic technologies.

ACKNOWLEDGEMENTS

This work has been supported by Nokia Research Center, the Helsinki Graduate School in Computer Science and Engineering, and the Pythagoras Graduate School of the Academy of Finland.

REFERENCES

[1] D. R. Begault. *3-D Sound for Virtual Reality and Multimedia.* Academic Press, Cambridge, MA, USA, 1994.

[2] R. T. Beyer. *Sounds of Our Times: Two Hundred Years of Acoustics.* Springer Verlag, New York, NY, USA, 1998.

[3] S. Foster, E. Wenzel, and R. Taylor. Real-time synthesis of complex acoustic environments. In *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Mohonk Mountain House, New Paltz, New York, 1991.

[4] M. Karjalainen. Immersion and content - a framework for audio research. In *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 71–74, Mohonk Mountain House, New Paltz, New York, 1999.

[5] A. S. Bregman. *Auditory Scene Analysis. The Perceptual Organization of Sound.* MIT Press, Cambridge, MA, USA, 1990.

[6] B. L. Vercoe, W. G. Gardner, and E. D. Scheirer. Structured audio: Creation, transmission, and rendering of parametric sound representations. *Proceedings of the IEEE*, 86(5):922–940, 1998.

[7] T. Tolonen. *Object-Based Sound Source Modeling.* PhD Thesis Proposal, Espoo, Finland, 2000.

[8] B. Shinn-Cunningham, H. Lehnert, G. Kramer, E. Wenzel, and N. Durlach. Auditory displays. In R. Gilkey and T. Anderson, editors, *Binaural and Spatial Hearing in Real and Virtual Environments*, pages 611–663. Lawrence Erlbaum Associates, Dahwah, New Jersey, 1997.

[9] ISO/IEC International Standard (IS) 14496 (MPEG-4). Information technology – Coding of audiovisual objects. 1999.

[10] ISO/IEC International Standard (IS) 14772-1. Information technology – Computer graphics and image processing - The Virtual Reality Modeling Language (VRML97). 1997.

[11] J. Huopaniemi. *Virtual Acoustics and 3-D Sound in Multimedia Signal Processing.* PhD thesis, Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, report 53, 1999.

[12] L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen. Creating interactive virtual acoustic environments. *Journal of the Audio Engineering Society*, 47(9):675–705, September 1999.

[13] M. Kleiner, B.-I. Dalenbäck, and P. Svensson. Auralization – an overview. *Journal of the Audio Engineering Society*, 41(11):861–875, November 1993.

[14] A. Krokstad, S. Strom, and S. Sorsdal. Calculating the acoustical room response by the use of a ray tracing technique. *Journal of Sound and Vibration*, 8(1):118–125, 1968.

[15] D. van Maercke and J. Martin. The Prediction of Echograms and Impulse Responses within the Epidaure Software. *Applied Acoustics*, 38(2-4, Special Issue on Computer Modelling and Auralisation of Sound Fields In Rooms):93–114, 1993.

[16] A. Kulowski. Algorithmic representation of the ray tracing technique. *Applied Acoustics*, 18(6):449–469, 1985.

[17] J. B. Allen and D. A. Berkley. Image method for efficiently simulating small-room acoustics. *Journal of the Acoustical Society of America*, 65(4):943–950, 1979.

[18] J. Borish. Extension of the image model to arbitrary polyhedra. *Journal of the Acoustical Society of America*, 75(6):1827–1836, 1984.

[19] L. Savioja. *Modeling Techniques for Virtual Acoustics*. PhD thesis, Helsinki University of Technology, Telecommunications Software and Multimedia Laboratory, report TML-A3, 1999.

[20] M. R. Schroeder. Natural-sounding artificial reverberation. *Journal of the Audio Engineering Society*, 10(3):219–223, 1962.

[21] J. A. Moorer. About this reverberation business. *Computer Music Journal*, 3(2):13–28, 1979.

[22] J-M. Jot. Real-time spatial processing of sounds for music, multimedia and interactive human-computer interfaces. *Multimedia Systems*, 7:55–69, 1999.

[23] W. G. Gardner. Reverberation algorithms. In Mark Kahrs and Karlheinz Brandenburg, editors, *Applications of Digital Signal Processing to Audio and Acoustics*, chapter 3, pages 85–131. Kluwer Academic Publishers, Norwell, MA, USA, 1998.

[24] J. Blauert. *Spatial Hearing. The Psychophysics of Human Sound Localization.* MIT Press, Cambridge, MA, USA, 1997.

[25] H. Møller. Fundamentals of binaural technology. *Applied Acoustics*, 36(3-4):171–218, 1992.

[26] F. L. Wightman and D. J. Kistler. Headphone simulation of free-field listening. I: stimulus synthesis. *Journal of the Acoustical Society of America*, 85(2):858–867, 1989.

[27] H. Møller, M. Sørensen, D. Hammershøi, and C. Jensen. Head-related transfer functions of human subjects. *Journal of the Audio Engineering Society*, 43(5):300–321, May 1995.

[28] J. Huopaniemi, N. Zacharov, and M. Karjalainen. Objective and subjective evaluation of head-related transfer function filter design. *Journal of the Audio Engineering Society*, 47(4):218–239, April 1999.

[29] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman. Localization using nonindividualized head-related transfer functions. *Journal of the Acoustical Society of America*, 94(1):111–123, July 1993.

[30] H. Møller, M. Sørensen, C. Jensen, and D. Hammershøi. Binaural technique: Do we need individual recordings? *Journal of the Audio Engineering Society*, 44(6):451–469, 1996.

[31] B. B. Bauer. Stereophonic earphones and binaural loudspeakers. *Journal of the Audio Engineering Society*, 9:148–151, 1961.

[32] B. Atal and M. Schroeder. Apparent sound source translator *U.S. patent* no. 3,236,949, 1966.

[33] W. G. Gardner. *3-D Audio Using Loudspeakers.* Kluwer Academic Publishers, Boston, 1998.

[34] M. Gerzon. Panpot laws for multispeaker stereo. In *the 92nd Convention of the Audio Engineering Society*, Vienna, Austria, 1992. preprint 3309.

[35] V. Pulkki. Virtual sound source positioning using vector base amplitude panning. *Journal of the Audio Engineering Society*, 45(6):456–466, June 1997.

[36] M. Gerzon. Periphony: with-height sound reproduction. *Journal of the Audio Engineering Society*, 21(1/2):2–10, 1973.

[37] D. Malham and A. Myatt. 3-D sound spatialization using ambisonic techniques. *Computer Music Journal*, 19(4):58–70, 1995.

[38] K. Brandenburg and M. Bosi. Overview of MPEG audio: Current and future standards for low-bit-rate audio coding. *Journal of the Audio Engineering Society*, 45(1/2):4–20, January/February 1997.

[39] R. Väänänen and J. Huopaniemi. Spatial processing of sound in MPEG-4 virtual worlds. In *Proceedings of the X European Signal Processing Conference (EUSIPCO)*, Tampere, Finland, September 5-8 2000.

[40] 3D Working Group of the Interactive Audio Special Interest Group. Interactive 3D Audio Rendering Guidelines, Level 2. Technical report, MIDI Manufacturers Association, Los Angeles, USA, Sept. 1999. Available at http://www.iasig.org/.

[41] F. Hesham, J. A. Ballas, and D. Brock. An Extensible Toolkit for Creating Virtual Sonic Environments. In *Proc. Int. Conf. Auditory Display (ICAD'2000)*, pages 32–37, Atlanta, GA, April 1-4 2000.

[42] G. Eckel. Applications of the CyberStage Sound Server. In *Proc. AES 16th Int. Conf. on Spatial Sound Reproduction*, pages 478–484, Rovaniemi, Finland, April 10-12 1999.

[43] E. M. Wenzel, J. D. Miller, and J. S. Abel. Sound Lab: A Real-Time, Software-Based System for the Study of Spatial Hearing. In *the 108th Audio Engineering Society (AES) Convention, preprint no. 5140*, Paris, Feb. 19-22 2000.

[44] J. Blauert, H. Lehnert, J. Sahrhage, and H. Strauss. An Interactive Virtual-Environment Generator for Psychoacoustic Research. I: Architecture and Implementation. *Acustica united with Acta Acustica*, 86(1):94–102, 2000.