AN INTEGRATIVE MODEL OF DIGITAL SOUND RENDERING

Dr. Thomas Hilton
Utah State University
hilton@cc.usu.edu

Charlotte Andersen
Utah State University

Nathan Walker
Utah State University

Chet Barney
Utah State University

Dr. Hoon Choi
Dongshin University

Dr. Gi-Taek Hur
Dongshin University

Digital sound rendering is the procedure for taking an analog electrical audio signal and sampling it into digital data, storing and retrieving that data, and resynthesizing a smoothed analog audio signal from the digital data. The sound rendering field has developed rapidly with the rest of the multimedia industry, benefiting from contributions worldwide. However, typical of many parts of the information revolution, it suffers from conflicting and ill-defined terms. Therefore, a simplified, comprehensive, integrative model of sound rendering is proposed to standardize and clarify the vocabulary used in the field of sound rendering.

Keywords: sound rendering, digital-to-analog conversion, digital signal processing

With the rest of the information industry, multimedia products have had a revolution in the last decade. However, amazing advances in visual effects have tended to overshadow equally significant developments in sound effects. Interestingly, the auditory environment has been shown to have a potentially greater effect on the believability of a multimedia scene than its visual components [9]. Our ears tell us where we are in the environment as well as where the sound is relative to us [17]. A convincing acoustic track can make the difference in whether or not the audience accepts the virtual world created for them [12].

Effective use of sound rendering technology is the key to making an effective sound environment [12]. The sound environment encompasses everything from the speaker setup to the type of microphone used in recording as well as anything else that contributes to the final effect [19]. Every sound rendering project—everything from movies to websites—exists in a sound environment, and technology convergence has made available similar technology for all those projects. However, academics, professionals, and others interested in sound rendering are separated by a great divide of proprietary information and nonstandard vocabulary. This paper was written to begin to bridge that gap by defining a model of sound rendering that covers the entire process and integrates all its major components. To provide the necessary background, we first present a list of basic definitions of sound rendering concepts. Next we detail the sound rendering model itself and three sub-models. Finally, we identify what we see as the major problems and controversies in the sound rendering state of the art.

Basic Definitions and Sound Rendering Concepts

A main source of confusion in the sound rendering community stems from inconsistent and differing uses of the jargon. Based on an extensive literature review, the following definitions are proposed as standards to help clarify misunderstandings. They are explained in terms of this simplified sound rendering model:

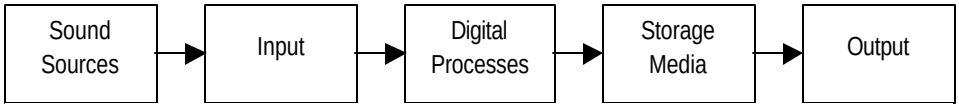

Figure 1: An Initial, Simplified Sound Rendering Model

Sound Rendering

For many people in the industry, sound rendering encompasses the whole process of gathering or creating the sounds, mixing them together and outputting them in an analog form [17]. Other industry professionals use sound rendering to describe "the technique of generating a synchronized soundtrack for animations" [19] or as a synonym for the algorithm that regenerates sound from digital audio data [15]. Still others use different terms altogether such as audio rendering [15, 16], sound synthesis [19], and auditory rendering [3]. Here we suggest the following generally applicable standard definition of sound rendering:

The procedure for taking an analog electrical audio signal and sampling it into digital data, storing and retrieving that data, and resynthesizing a smoothed analog audio signal from the digital data.

Note that this definition does not encompass the sound source (either real or synthesized), the physical recording environment (microphone choice, placement, etc.), the target media (CD, DVD, etc.), conversion from one format to another (e.g., Dolby Digital Surround Sound to MP3), any particular hardware or software configuration, or the final laying down of the soundtrack. This simple definition encompasses many other sub-processes and concepts which will be addressed in this paper.

Sound Sources

This definition of sound rendering does not include the sources of the sound but some discussion is necessary to provide background. The first issue in sound sources is to decide whether to record live clips from the real world or to synthesize the sounds [5]. Sound rendering deals with both types of sound sources [9]. There are many programs available to aid in the process of gathering and enhancing live sounds or to help synthesize new sounds such as Digidesign's ProTools, Mark of the Unicorn's Digital Performer, Ensoniq's Paris, and products from Sonic Solutions, Cedar and others. Sound rendering principles and techniques can be applied to any analog (continuously variable) sound recording.

Digital Processes

Once the sounds are in analog electrical form, digital processing can be applied. This processing stage involves many complexities and requires some expertise to decide which ones are the most useful for the purpose. The processes of sampling, filtering, and resynthesizing are used in almost every application of sound rendering and will be discussed further [17].

Sampling. The first general process is sampling the recorded sound wave. Sound waves are a combination of amplitude (loudness) and frequency (pitch), and sampling is converting the analog sound wave into a stream of bits by taking periodic "snapshots" of the wave. Each sample is a multi-bit binary number representing the amplitude of the wave at that point in time; the change in amplitude between adjacent samples represents the frequency. The number of samples recorded per second is the sampling rate. Combining large numbers of adjacent samples created at a high rate (44,100 per second is typical) represents the shape of the original wave with some accuracy. Increasing the sampling rate and the number of bits used to describe each sample increases the accuracy of the representation of the original wave.

Filtering. The next process to be applied is filtering. There are many different types of filters used for different purposes, some of which are explained in greater detail later in this paper. Generally speaking, a filter takes the sound in, changes it in some way, and outputs the altered sound to the next process. Audio effects that digital filters can create all depend on somehow adding, removing, or altering samples [13].

Resynthesizing. Resynthesizing is a collection of processes that are the converse of sampling. These processes take the digital bit stream and from it reconstruct an analog electrical signal [13]. This reconstructed analog electrical signal can then be output through speakers.

Output (and Output Problems)

Speakers transform analog electrical signals into sound waves. Speaker systems use different types and sizes of speakers to generate different ranges of sound: high-frequency drivers ("tweeters") for high-pitched sounds, mid-range and bass drivers ("mids" and "woofers") for low sounds, and subwoofers for very low frequency sounds (such as shock wave vibrations). The goal is to produce a sound wave with the highest possible fidelity to the original. However, because of physical limitations, the sound from a speaker is always inaccurate to some degree [13]. Sound may thus need to be rendered more than once with settings being changed each time to get the desired finished product. Although the listener's ear is the final judge of the sound from a speaker, computer tools also exist that analyze signals to identify possible inaccuracies that will appear in the analog output [13]. Several of these inaccuracies, which can be present at the sound source or accidentally created during digital processing, are described next.

Aliasing. One common problem is aliasing (also sometimes called foldover). Aliasing is a sampling error that occurs when the sampling frequency is less than twice the maximum frequency of the sound being sampled. When the sampling frequency is too low, the maximum frequency of the resynthesized sound is reduced from what it was originally. If there is a significant degree of aliasing, the human ear can hear a different pitch than existed in the original sound [13]. Another effect of aliasing is to decrease sound quality: since aliasing occurs most often in the higher frequency ranges, it often changes the harmonics that give an instrument or vocal sound its unique character. Thus, aliasing can dramatically affect the quality of a sound even if a pitch change is not perceived.

Phase distortion. Phase distortion is interaction between sound waves that was not present in the original sound source. It is usually caused by inaccuracies during the recording or playback process. Phase distortion artificially lowers the amplitude of some parts of the sound and raises the amplitude of other parts, thus compromising the fidelity of the output [13, 19].

Quantization. Quantization is the digital misrepresentation of amplitude during sampling [13]. It occurs when the digital value assigned to a sound does not exactly equal the amplitude of the original waveform. In digital sampling, some degree of quantization is unavoidable since the numbers that describe a particular sample must be rounded off to the nearest digit during analog-to-digital conversion. Higher quality digital recording systems (i.e., systems that use longer numbers to represent each sample—32 bits per sample rather than 16 or 8 for example) have fewer problems with quantization, but all digital sampling produces some amount [19]. Quantization noise can be heard as buzzing, humming, or clicking as the sound fades to silence.

Storage Media

Once sound sampling is finished, the bit stream must be stored. There are a variety of different storage media including magnetic tape, magnetic disks, optical disks, and combinations of the three [13]. There are also a variety of media using both sequential and direct access methods [13]. The selection of the type of storage medium is important but is outside the scope of this paper. Suffice it here to say that the bit stream is stored for later retrieval.

Integrative, Comprehensive Model of Sound Rendering

With basic concepts and vocabulary defined, we now introduce the comprehensive model The integrative, comprehensive model of sound rendering consists of one general model (figure 2), which is subsequently broken down into three sub-models. Each model is explained in detail.

Sound Rendering

The first and most general model (figure 2) shows the processes in rendering a sound.

Recording. The analog signal of the original sound is recorded using a microphone (or it is digitally synthesized). The microphone transforms the incoming sound into an analog

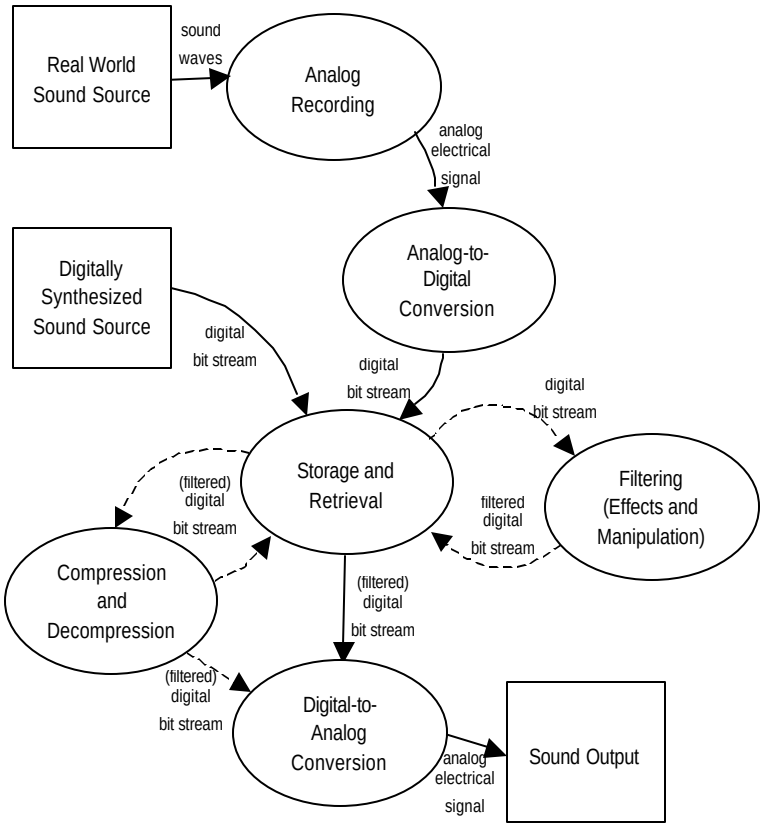

Figure 2. An Integrated, Comprehensive Model of Sound Rendering

electrical signal.

Conversion. The analog-to-digital conversion is accomplished via an Analog to Digital Conversion (ADC) chip that performs analog filtering, sampling, and some digital filtering. The output of the ADC is a bit stream containing all the samples from the sound [13].

Storage. The ADC passes the bit stream to a storage medium to await retrieval. The bit stream is stored in any number of formats on any number of storage media. The choices of media and format depend on the intended use of the sound and the hardware to be used in future playback or manipulation [13]. The compression/ decompression algorithm specific to the intended end format is also identified in this stage along with any special effects that will be in the final sound [14]. Compression inevitably degrades sound quality and should only be applied when it is necessary to save storage space or decrease transmission time. Since compression algorithms can be repeatedly applied, even to already compressed files, be careful not to compress files repeatedly if this can be avoided since sound quality will deteriorate with every round of compression.

Playback. To play a sound, it must be retrieved from the storage medium and passed to the digital-analog converter (DAC). The DAC is firmware that processes the bit stream to reconstruct an analog signal [13].

The DAC resynthesizes the analog signal. This is the converse of sampling. Where sampling took samples of an analog wave, resynthesis recreates the wave from the samples. The more samples taken, the more accurately the wave can be resynthesized [13]. The resynthesized analog wave is then passed through an analog filter to correct anomalies from the conversion [5].

The reconstructed analog electrical signal is then sent to the speakers, which play the sounds. The number, type and arrangement of speakers depends entirely on the desired sound environment. There has been much research in this area, particularly relating to virtual reality environments. However, this is not sound rendering, per se, and is not covered in this document.

Analog-to-Digital Conversion Sub-model

Figure 3 shows in greater depth the processes involved in analog-to-digital conversion. ADC consists of at least three processes and sometimes more. Many different hardware-software setups are available, some contained in a single chip and some requiring many separate components; however, the basic process is the same.

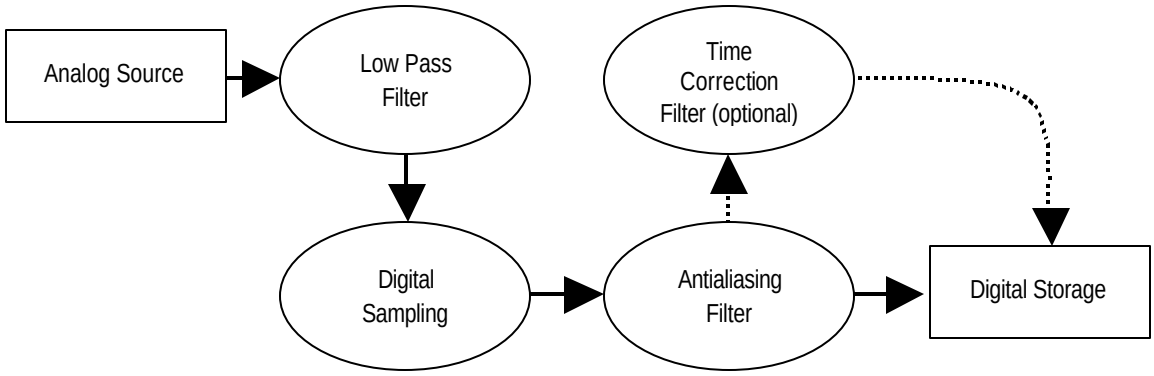

Figure 3. Analog-to-Digital Conversion Sub-model

Low pass filtering. The input to the ADC is an analog electrical signal (usually from an analog recording but possibly directly from a microphone). This signal is first fed into an analog "lowpass filter," a type of filter that removes extremely high frequencies that could be problematic further into the sampling process [13]. Newer lowpass filters let the user set the cutoff frequency. Other filters determine the cutoff frequency with a simple algorithm that takes the average of the current input sample and the previous input sample as shown in Figure 4:

Y[n]	=	(0.5 * x[n])	+	(0.5 * x[n-1])
Current Output		Half of the Current Input		Half of the Previous Input

Figure 4. A Simple Low Pass Filtering Algorithm

Sampling. The filtered analog signal is next passed to an ADC chip that samples it, outputting a digital bit stream that represents the analog signal. The ADC samples the sound many times per second; this is called the sampling frequency (not to be confused with the sound frequency) [13]. As an example, the sampling frequency for CDs is 44,100 samples per second. This can be calculated from Nyquist's sampling theorem: "to reconstruct a signal, the sampling frequency must be at least twice the frequency of the signal being sampled" [13]. In other words, there should be "at least two samples per period" of the original wave, one to sample the high point and one to sample the low point produces a fairly accurate picture of the original sound [13]. To sample the normal human hearing range of frequencies (0 Hz to 22.05 kHz), Nyquist's theorem indicates a sampling rate twice that range: 44.1 kHz. This is a good starting point, but oversampling, the taking of more samples than required by Nyquist's theorem (typically 96 kHz), is now commonly used to compensate for digital recording problems (notably phase distortion) that manifest at the lower sampling rate. More on this below.

Anti-aliasing filtering. The filtered analog signal is next passed to the anti-aliasing filter. The purpose of this filter is to reduce effects of aliasing in the signal. Since too much anti-aliasing filtering can introduce phase distortion into the recording, some (hopefully innocuous) aliasing is generally left in the recording so as to minimize phase distortion [13].

Time filtering. An optional time correction filter is also shown in Figure 3. The time correction filter is to correct phase distortion. Time filtering is often a good idea because high frequencies generally propagate faster than low frequencies, so some phase distortion is usually present from the original analog recording on through the rest of the rendering process. A time filter slightly delays sounds with higher frequencies so as to bring them back into phase with the lower frequency sounds in the recording [13]. Oversampling can also be used to reduce phase distortion. After filtering, the modified bit stream of samples is stored.

Digital-to-Analog Conversion Sub-model

When the stored bit stream is retrieved for playback, the digital-to-analog conversion of the stored samples must occur. Figure 5 shows in greater detail the functions of the DAC and the processes involved in this resynthesis.

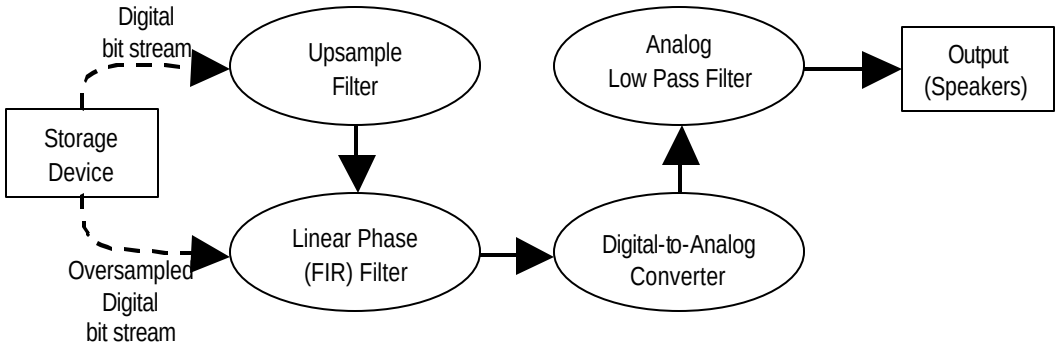

Figure 5. Digital-to-Analog Conversion Sub-model

Oversampling. The first digital sound recordings (for CDs when they first appeared on the market) contained significant problems with phase distortion that were caused by stringent anti-aliasing, which had been made necessary by the use of a barely adequate (44.1 kHz) sampling rate. This problem was solved by oversampling during the ADC process (explained above). An oversampled digital recording, when converted back to analog form, contained significantly higher fidelity than did a normal digital recording. Unfortunately, although this is fine for new digital recordings, it does not do anything for the old ones.

Upsampling. Enter upsampling, invented at Data Conversion Systems (dCS) in Saffron Waldon, England in 1998 [22]. An upsampling filter is firmware that creates and inserts more digital samples into the bit stream than were originally stored. The first upsampling systems used linear or quadratic interpolation to create and insert several (typically seven) "samples" between each pair of real samples. However, experimentation revealed a very interesting (and at least partially inexplicable) thing about upsampling: it actually works best when binary zeros are inserted between the recorded samples. One would think that these "null samples" would have no effect at all on the sound, but as Dr. John Culling of Cardiff University asserts, "Interpolating zeroes rather than linearly or quadratically interpolated values appears odd to the uninitiated. However, this method *is* the most appropriate" [21]. Moreover, dCS spokespeople have confirmed that "we have found that upsampling reveals information that is present in the master source, but which is not audible when the CD is played back normally. Upsampling cannot increase the amount of information in a signal and the exact mechanism behind the perceived sonic improvements is currently not entirely clear. We are continuing our research into this subject" [21]. Beyond filling the minute silences between stored samples, oversampling is also useful for reducing quantization noise and phase distortion [5]. Thus, upsampling DACs give lackluster 16-bit/44.1 kHz digital recordings the audio richness of 24-bit/96 kHz recordings.

Generally speaking, there are two types of upsampling: multi-bit and single-bit. Multi-bit upsampling was used in linear or quadratic interpolation systems since entire samples needed to be interpreted to calculate the intermediate samples. However, when zero-only upsampling was found to be superior, one-bit upsamplers quickly became preferred. This is because one-bit upsamplers are much simpler to implement (although they must operate many times faster than multi-bit upsamplers) [13].

Linear phase filter. After upsampling (if needed), the digital bit stream is usually passed through a linear phase (FIR) filter. A linear phase filter removes every sample that is outside a predetermined range (usually the range of normal human hearing, 0 Hz to 20 kHz). This process reduces phase distortion and noise, but without the problems introduced by an analog filter [13].

Conversion. Once this is accomplished, the bit stream is finally fed into a digital-to-analog converter that produces an analog electrical current virtually identical to the original.

Output. After being resynthesized, the freshly produced analog signal is fed through one last filter. This time it is an analog low pass filter that removes any high-frequency harmonics generated in the resynthesis. After that, the signal is finally output to the speakers.

Problems and Issues Pertaining to the Model

Now that we have explained a comprehensive model that integrates many of the heretofore separately treated elements of sound rendering, it is only fair to admit that things in the industry are not nearly this tidy. Multiple issues drive controversy among sound rendering professionals. Some of the more obvious and important ones are

- All-at-once downloads vs. real-time streaming
- File compression vs. maintenance of playback quality
- Unpredictable network transmission latencies
- Integrating voice and data networks
- Low-power digital signal processing for portable applications.

Conclusion

After reviewing some of the concepts related to sound rendering, general and specific models, some problems and issues, and what the industry is working on, a general picture of the current state of the sound rendering environment comes into focus. Continued research and development in digital sound; it's creation, implementation, and distribution, is of vital importance in today's digital world. A convincing auditory environment has been proven to be more effective in creating a believable environment than visual cues. A good soundtrack is integral to any virtual reality simulation. Without good sound, it is just a pretty picture.

REFERENCES

[1] "DD Clip Pro." Version 3.03. April 2000. http://www.softlab-nsk.com/ddclipro/ (30 June 2000).

[2] "Linear Interpolation: How it Works." Brown University. http://www.cs.brown.edu/stc/outrea/greenhouse/nursery/interpolation/itworks.html (30 June 2000).
[3] Dai, Ping, Gerhard Eckel, Martin Göbel, Frank Hasenbrink, Vali Lalioti, Uli Lechner, Johannes Strassner, Henrik Tramberend, and Gerold Wesche. 1997. "Virtual Spaces: VR Projection System Technologies and Applications." http://viswiz.gmd.de/~eckel/publications/eckel97c/Tutorial.V4.html (30 June 2000).
[4] Funkhouser, Thomas, Ingrid Carlbom, Gary Elko, Gopal Pingali, Mohan Sondhi, and Jim West. "A Beam Tracing Approach to Acoustic Modeling for Interactive Virtual Environments." *Computer Graphics Proceedings, Annual Conference Proceedings* (1998): 21-32.
[5] Horbach, Ulrich, and Marinus M. Boone. "Future Transmission and Rendering Formats for Multichannel Sound." *AES 16th International Conference on Spatial Sound Reproduction* (April 1999).
[6] Jot, Jean-Marc. "Synthesizing Three-Dimensional Sound Scenes in Audio or Multimedia Production and Interactive Human-Computer Interfaces." *5th International Conference: Interface to Real & Virtual Worlds* (1996).
[7] Mahon, Michael. Sound Editor v2.2. "Resampling Algorithms." http://www.grin.net/~cturley/gsezine/GS.WorldView/v1997/Mar/sound22.shk.info.txt (30 June 2000).
[8] McNab, Scott. "Digital Mixing Techniques." 1997. http://www.it.net.au/oxygen/mixing/section4.html (30 June 2000).
[9] Pellegrini, Renato. "Comparison of Data- and Model-Based Simulation Algorithms for Auditory Virtual Environments."
[10] Pope, Jackon, and Alan Chalmers. *Pre-rendering Acoustics and Illumination for Archaeological Reconstructions*, J. A. Barcelo, M. Forte, and D. H. Sanders, editors, pages 105--110. ArcheoPress, Oxford, April 2000.
[11] Pope, Jackson, and Alan Chalmers. "Multi-sensory Rendering: Combining Graphics and Acoustics." In *Proceedings of the 7th International Conference in Central Europe on Computer Graphics*, pages 233--242. University of West Bohemia, Czech Republic, February 1999.
[12] Pope, Jackson, David Creasey, and Alan Chalmers. "Realtime Room Acoustics Using Ambisonics." In *The Proceedings of the AES 16th International Conference on Spatial Sound Reproduction*, pages 427--435. Audio Engineering Society, April 1999.
[13] Roads, Curtis. *The Computer Music Tutorial*. Cambridge: The MIT Press, 1996.
[14] Skywalker Sound. "The Making of a Movie Soundtrack." http://www.thx.com/skywalker/skywalker.html (30 June 2000).
[15] Songlab Studio. "What is Audio Rendering?" http://home.snafu.de/rubo/songlab/midi2cs/m2cfaq.htm#Q2 (30 June 2000).
[16] Staccato Systems. "What is Audio Rendering Technology?" http://www.staccatosys.com/technology/faq.html#2 (30 June 2000).
[17] Takala, Tapio, and James Hahn. "Sound Rendering." *Computer Graphics 26* (July 1992): 211-219.
[18] Texas Instruments. "IT Breaks Industry's DSP High Performance and Low Power Records with New Cores."
http://www.ti.com/sc/docs/news/2000/99085a.htm (6 June 2000).
[19] Tonneson, Cindy, and Joe Steinmetz. "3D Sound Synthesis." http://www.hitl.washington.edu/scivw/EVE/I.B.1.3DSoundSynthesis.html (30 June 2000).
[20] Tsingos, Nicolas, and Jean-Dominique Gascuel. "Soundtracks for Computer Animation: Sound Rendering in Dynamic Environments with Occlusions." http://www.dgp.toronto.edu/gi/gi97/proceedings/papers/TsingosGascuel/ (30 June 2000).
[21] Culling, John. "Dictionary of DSP." http://www.cf.ac.uk/psych/CullingJ/dictionary.html (10 July 2001).
[22] Data Conversion Systems, Ltd. "Official dCS Home Page." http://www.dcsltd.co.uk/ (12 July 2001).

This research was funded by a research grant to the Software Research Center at DongShin University from the Information and Communication Research Organization of the Information and Communication Department of the Federal Government of South Korea