CONTROVERSIES IN DIGITAL SOUND RENDERING

Dr. Thomas Hilton & Chet Barney, Utah State University, Email: Hilton@cc.USU.edu
Dr. Hoon Choi and Dr. Gi-Taek Hur, Dong Shin University

ABSTRACT

A comprehensive model of sound rendering is briefly summarized. Controversies surrounding each element of the model are described, all stemming from tradeoffs in quality vs. file size: sampling rates, compression methods, filtering and effects, dealing with artifacts of the process, and social/ethical issues arising from the tremendous flexibility of the digital medium.

Keywords: multimedia, sound rendering, audio processing

INTRODUCTION

In Hilton et al. [5], a comprehensive model of sound rendering was developed and explained. This paper builds on that one to identify and prioritize controversies in sound rendering. The goal of this paper is to clarify the difficulties and disagreements presently being debated among sound rendering professionals with an eye to addressing each of them in more detail elsewhere.

Sound rendering is defined as the procedure of creating an analog electrical audio signal, sampling it into digital data, storing and retrieving that data, resynthesizing an analog audio signal from the digital data, and smoothing that signal (Hilton, et al.). In the context of this definition, Figure 1 shows the comprehensive sound rendering model.

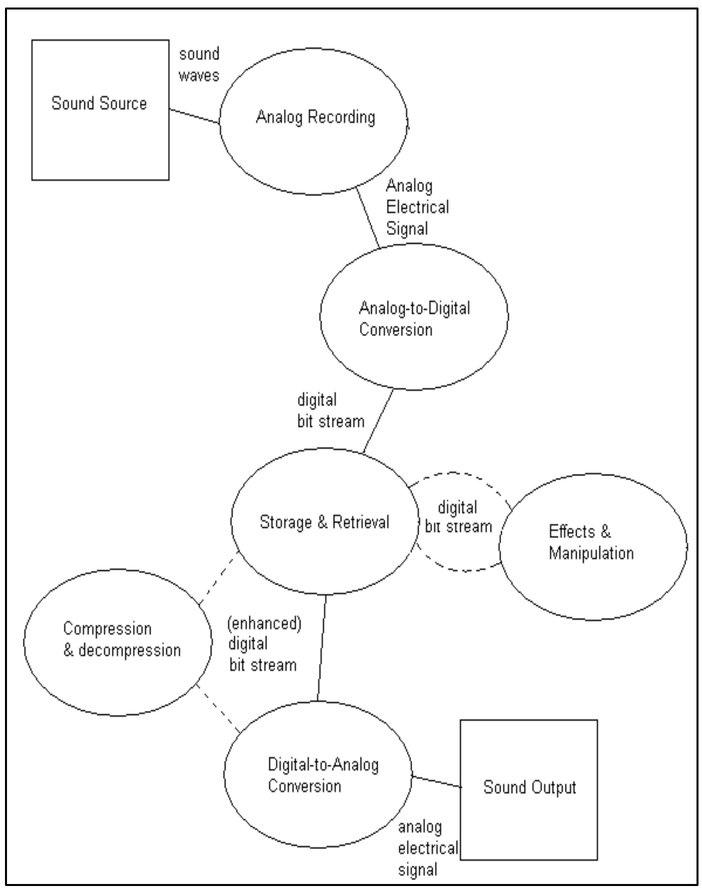

Figure 1. Comprehensive Sound Rendering Model

Figure 1 shows that a sound source is recorded in analog format, digitized, modified if needed, and then stored; for playback it is retrieved, again modified as needed, converted back to analog form, and output as sound. Readers who would like more detail on this process are referred to the earlier paper. Readers who believe the process is straightforward and needs no elucidation are invited to read on and discover that artistic disagreement and technical difficulty exist throughout.

CONTROVERSIES

The major controversies in digital sound rendering can be considered as resulting from the conflict between sound quality and audio file size. In this respect, the type of sound quality that is desired may be sacrificed due to the file size that can be handled by a single listener. Note that, generally speaking, audio processor power (i.e., the number of audio playback instructions executed per second) has more than kept pace with recording and playback requirements and is not considered a controversy in sound rendering.

File Size Requirements

The standard minimum sampling rate for high-fidelity audio is 44,100 samples per second (44ksps). Sixteen bits (2 bytes) are needed per sample to achieve a good signal-to-noise ratio. With this information the amount of data needed for digital audio can be calculated: 41,000 samples per second, times 2 bytes per sample, times 2 channels for stereo, times 60 seconds per minute equals more than 10 megabytes of data per minute for CD-quality audio.

For this quality of audio, a high-density floppy disk holds less than 8 seconds of sound, and a 100 MB Zip cartridge holds less than 10 minutes. Clearly, the memory and storage requirements of digital audio are substantial. Fortunately, a compact disc (or a 550 MB CD-ROM) can hold about an hour of stereo sound, and a hard disk drive of at least 20 gigabytes is standard for audio recording and processing.

Quality Requirements

Although the ear is not particularly effective in measuring absolute sonic values, it is very keen in hearing variations in frequency and amplitude. Audible frequencies are generally between about 4 kHz and 20 kHz, audible amplitudes range from about 0.1 decibels to well over 100 decibels—although immediate physical injury to the ear begins at about 110 decibels [3, p80]. In this context audio quality can be defined as faithfulness or “fidelity” of the playback to the original in frequency and amplitude.

If this were the whole story, there would be little controversy: computer audiophiles would just use CD-ROMs. Unfortunately, two issues complicate the situation. The first is that 44ksps doesn’t actually result in top-quality audio fidelity; notably, that sampling rate allows for significant aliasing (a sort of echo distortion) and quantization noise (usually heard as clicking or humming in the background of quiet recordings). Part of getting rid of aliasing, quantization noise, and other undesirable audio traits of digital recording and playback is to increase the sampling rate. A typical sampling rate actually used in professional recordings is about 96ksps. With 16-bit samples, this yields a data rate of about 23 MB per minute, which would require about 1.3 GB for an hour of audio. This is way more data than will fit on a CD or a CD-ROM.

Added to the sampling rate problem is the sample size problem. Sixteen bits per sample is an acceptable minimum, but high quality demands a larger sample size to more accurately reflect the actual frequency at the instant of sampling. Thus, typical sample sizes in professional recording environments are 32 to 64 bits per sample. This results in a data rate of 46 to 92 MB per second, or about 16.5 to 33 GB per hour. The computer industry is far from supplying that kind of capacity in an inexpensive, portable medium. (It is interesting to note that CDs represent a level of reduction in audio fidelity from these extremely high quality studio recordings; the reduction is necessary to fit 33 GB of audio onto a 550 MB CD.)

A final insult to digital audio quality is the fact that many people want to transmit and receive music via the Internet. High-end consumer modem data rates hover in the mid-40kbps (thousand bits per second) range, or only about 5 KB per second. Even so-called broadband Internet access generally delivers throughput no faster than 200-300 mbps (25 to 40 MBps). Both of these throughput rates pale in comparison to the data rates generated in high quality digital recordings.

Data Compression

One popular approach to solving the conflict between quality and file size is to compress the data for storage and decompress it for resynthesis during playback. Probably the most popular audio data compression format is MP3, although there are others. Many people hear little if any difference between MP3 and CD-audio (probably because they are running their test with sub-par electronic components such as cheap computer speakers). However, if quality hardware is used, significant differences in frequency response and complexity/depth appear. As explained below, this is a direct result of the MP3 algorithm deleting samples during compression.

Compression Formats

Once data compression is deemed appropriate, there is another controversy to address: to implement lossless or lossy compression. Kearns [6] explains that original conceptions of data compression were based on the assumption that 100% accurate decompression was required. Indeed, this is true for many data types (notably byte-based files such as text files, databases, etc.). The “lossless” algorithms developed under this scenario generally use some variation of the run-length encoding concept: if a repeating pattern of bits is found in a portion of the file, the pattern is stored only once along with a counter indicating how many times to repeat it for decompression (this counter is called the “run length” because it expresses how long to “run” the stored pattern). For instance, if a single high C is played on a violin for 2 seconds in a piece of music, taking 44 ksps with stereo 16-bit samples would result in about 352 KB of data. However, since every sample is the same, run-length encoding could be used to reduce it to a mere four bytes: two bytes for one 16-bit sample and two bytes for the run (which would be 88,000). Of course, most files are more complex than that, so a compression factor of 88:1 is rare; still, this illustrates the concept.

However, as Kearns points out, lossless compression doesn’t shrink most audio files as much as one would like. This is for two reasons: 1) they start out huge, and 2) they usually contain relatively few repeating patterns, which means that run-length encoding algorithms don’t shrink them much. Fortunately, audio files also have another trait: the sounds they represent are often so complex that most listeners can’t fully attend to them. This means that some of the data could be completely excluded from the playback and nobody (or almost nobody) would notice. Thus, data compression algorithms have been developed to selectively erase data from the file rather than encoding it more compactly. Such algorithms are called “lossy” because they involve irretrievable data loss, and the most famous of them are the MPEG algorithms. MP3 (MPEG 1, audio layer 3) is the lossy algorithm most often applied to audio files.

As an example of lossy audio data compression, consider a typical rock band consisting of guitars, drums, and vocals. When digitally sampling a band performance, each sample will contain sounds from each band member, but some samples will be dominated by one sound. Thus, if two samples where the predominant sound is from a loudly squealing lead guitar have a sample between them dominated by a softer bass drum thump, most listeners will never hear the middle sample; it could be completely dropped and replaced on playback with a sample interpolated from the remaining two without generally noticeable loss in fidelity. Taking advantage of this fact, the MP3 data compression algorithm simply excludes samples below relative thresholds of amplitude (loudness) and frequency (pitch). When the sound is resynthesized from the data, the excluded samples result in lost sound, but the sounds are lost in places where most people aren't listening anyway.

Another trait of lossy compression that helps with file size is that it can allow user control of recording quality. For instance, MP3 "ripping" software such as Real Media's Real Jukebox can record at data rates (fidelity levels) as low as 28 kbps or as high as 128 kbps. Thus, a typical 3-minute song that would occupy about 30 MB of disk space in CD format can be compressed in 128 kbps MP3 format, stored in 2 to 3 MB, and still yield an acceptable playback fidelity level. Alternatively, a 3-minute (nonmusical) speech that would also occupy 30 MB in CD format can be compressed in 28 kbps MP3 format, stored in 630 KB, and still yield acceptable playback fidelity (because the speaker is clearly intelligible despite obviously compromised musicality).

Intended Audience

For many listeners, then, data compression seems to be a workable compromise between file size and audio fidelity. However, many listeners (notably audio recording professionals whose livelihood depends on their excellence) remain dissatisfied because of what they view as unacceptable fidelity losses from compression. It thus appears that this question cannot be adequately addressed for everyone at once, but rather different storage/quality tradeoffs are appropriate for different audiences.

McCarthy [7] found that the most effective traits for distinguishing types of audio listeners are available network bandwidth and disk storage space. If the target audience has robust connections (100 mpbs or higher actual throughput per user) and plenty of storage space (100's of GB per user), little audio data compression may be needed; typical listeners in this category are moviegoers. On the other hand, if the target audience has low bandwidth access and limited storage, audio data compression becomes necessary; typical listeners in this category are PC users. For most computer users, some type of compression is usually appropriate.

Conclusion

With the use of digital audio rising among corporations, educational systems, and individuals, the need to expand the efficiency of digital sound rendering has also risen. Controversies in sound rendering technology have arisen as the technology has evolved. The central question is one of audio playback fidelity versus file size. Various storage media have been developed to contain large audio files, but file size remains a problem. Both lossless and lossy data compression techniques have been used to shrink audio files, and lossy algorithms such as MP3 are presently the more popular. However, lossy algorithms take a much greater toll on audio quality than do lossless algorithms such as run-length encoding. This means that the quality versus file size controversy will continue for the foreseeable future. The authors hope the time will come when a happy medium will be reached in which inexpensive, portable storage media will have sufficiently large-capacity to contain audio files that have been compressed efficiently but without noticeable quality loss.

Bibliography

[1] 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.” (30 June 2000).

[2] Dobrian, Christopher, “Digital Audio.” MSP: The Documentation Cycling ’74 and IRCAM, December 1997.

[3] Everest, F. Alton, *Master Handbook of Acoustics*. McGraw-Hill, 2001.

[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] Hilton, Andersen, Walker, Barney, Choi, Hur. “A Comprehensive and Integrative Model for Digital Sounding Rendering.” *Issues in Information Systems*, IACIS, 2001.

[6] Kearns, Randy. “What is File Compression?” December 2001, Available online: http://searchstorage.techtarget.com/ateQuestionNResponse/0,289625,sid5_cid425550_tax286193,00.html.

[7] McCarthy, Mark, “Web Based Streaming Media.” Academic Technology and Networks, University of North Carolina, 1998.

[8] Newquist, H.P., “Music & Technology.” Watson-Guptill, 1989.

[9] Piccialli, De Poli, and Roads. “Representations of Musical Signals.” Cambridge: The MIT Press, 1991.

[10] Pohlmann, Ken C., “Principles of Digital Audio.” McGraw-Hill, 1995.

[11] Roads, Curtis. "The Computer Music Tutorial." Cambridge: The MIT Press, 1996.

[12] Takala, Tapio, and James Hahn. “Sound Rendering.” *Computer Graphics 26* (July 1992): 211-219.

[13] Watkinson, John, “The Art of Digital Video.” Focal Press, 2000.