1. NetSound

NetSound is a sound and music description system based on the pseudocode for encoding sound streams as decompositions from a sound-specific decomposition description (representing arbitrarily complex signal processing algorithms) and event lists (comprising scores or MIDI files).

This description is analogous to the Pustelny® language for image and text information in which construction information for fonts and images is separated from raw ASCII text. As a network sound transmission protocol, NetSound has the advantage of being able to transmit a wide selection of sounds using a descriptive format that does not require a high-bandwidth channel. Since description-based audio represents acoustic events as parameterized units, a great deal of control over the resulting sound is offered. The use of complex instrument descriptions and appropriately parameterized scores makes it possible to specify descriptions of complete sound tracks or musical pieces using a very small amount of data. Other synthesis languages’ instruments, such as the MUSIC-N languages, and commercial synthesizer implementations can be translated into Csound syntax.

2. NetSound as a Sound Specification Protocol

Object-based representations for sound synthesis can be thought of as a series of audio processing building blocks that are threaded into a signal processing network for each class of sound. Each sound instance produces a copy of the signal-processing template for that class of sound. These data structures are constructed on the client side by the Csound compiler. Once constructed and memory resident, the signal processing networks can be executed in real time under the control of a score or MIDI file event list. Csound features a complex dynamic execution environment that adjusts memory requirements as needed and maintains efficiency by optimized allocation and reallocation of memory.

3. Network Advantages of NetSound

Most of the existing network audio protocols rely on lossy audio compression techniques in order to reduce the bandwidth of an audio data stream. There are also protocols that are able to stream and uncompress buffered audio data in real time; for example, RealAudio® is able to deliver 1 channel of compressed sound over a 28.8Kbit communications channel at a resynthesis sampling rate of 11KHz. The quality of these techniques varies as a function of the compression ratio. Realtime compressed audio streams are good for browsing audio material but do not offer a quality that is acceptable for high-fidelity sound reproduction. High-quality compression schemes such as MPEG do not reduce the data enough to make transmission of large quantities of audio data feasible in a small amount of time. All of the existing techniques exhibit a linear relationship between the length of the original audio stream and the size of the compressed file.

NetSound has the advantage of requiring far less server throughput capacity and storage capacity than existing protocols. NetSound also has the potential to represent sound streams with a data packet that is sub-linearly or linearly scaled relative to the size of the original data stream.

4. Client-side computational efficiency vs. Bandwidth

Since NetSound utilizes Csound as a real-time software synthesis engine, issues of computational requirements must be addressed. The decision to exploit client-side computing resources is born out of the observation that current network activity is limited by client/server throughput rather than available processor cycles. As long as this is the case, a tradeoff between processor usage and bandwidth requirements must be made. In terms of processor usage, the most efficient method of audio synthesis is sample playback. However, an algorithmic technique such as FM synthesis or granular synthesis requires far more mathematical operations per audio sample, but also requires much less sound specification information for synthesis. Thus there is an complex relationship between computational efficiency and bandwidth requirement for specification, and the art of network sound design involves the careful consideration of computational resources and bandwidth availability. It is possible to exploit the merits of both when specifying sound streams using NetSound. In situations where bandwidth is plentiful, sample-based synthesis techniques are perhaps preferable. However, when processor cycles are likely to be available, other synthesis techniques, such as additive synthesis, or phase- vocoder synthesis may be incorporated to reduce network bandwidth requirements at the expense of increased client processor load.

5. Conclusion and Future Work

NetSound is currently well suited to synthesizing the types of sound and music that are produced in a modern multimedia production studio. It is the goal of NetSound to eliminate the pre-mastering stage of multimedia sound production in favor of distributing algorithmic synthesis descriptions, any necessary audio samples or analysis signals, and structured event lists for the sound stream. This information is only implicitly represented in a modern multi-media studio, because as yet there are no standards for exporting information relating to the specification of signal-processing networks. The future of software sound synthesis is somewhat dependent on such protocols, and NetSound is perhaps a first in this regard. As software synthesis starts to become embedded in multimedia technologies, we believe that the principles outlined above will become a governing factor in software-based sound design.

Csound is a Registered Trademark of the Media Lab, M.I.T.