AES WHITE PAPER

AESWP—1001

Technology Report TC-NAS 98/1:

NETWORKING AUDIO AND MUSIC USING INTERNET2 AND NEXT-GENERATION INTERNET CAPABILITIES

Technical Council
AUDIO ENGINEERING SOCIETY, INC.
INTERNATIONAL HEADQUARTERS
60 East 42nd Steet, Room 2520 · New York, NY 10165-2520, USA
Tel: +1 212 661 8528 · Fax: +1 212 682 0477
E-mail: HQ@aes.org · Internet: http://www.aes.org

AES White Paper:
Networking Audio and Music Using Internet2 and Next-Generation Internet Capabilities

Robin Bargar, ***University of Illinois***
Steve Church, ***Telos Systems***
Akira Fukuda, ***NHK Sound***
James Grunke, ***Hotz / MIDI Manufacturers Association***
Douglas Keislar, ***Muscle Fish***
Bob Moses, ***PAVO***
Ben Novak, ***Microsoft***
Bruce Pennycook, ***McGill University***
Zack Settel, ***McGill University***
John Strawn, ***S Systems Inc.***
Phil Wiser, ***Liquid Audio***
Wieslaw Woszczyk, ***McGill University***

Executive summary

The current Internet is inadequate for transmitting music and professional audio. Performance and collaboration across a distance stress beyond acceptable bounds the quality of service. Since audio and music are intrinsic parts of human society, it is important to include them in a new Internet proposal. Audio transmission is anyway an important part of other network-based applications, such as teleconferencing. Audio and music provide test cases in which the bounds of a network are quickly reached and through which the defects in a network are readily perceived.

1 Why include music and audio?

1.1 *Role of music in society*

The document http://www.internet2.edu/html/engineering.html foresees participation in Internet2 by medical researchers, physical scientists, and the leading-edge research community. To relegate audio and music to a background function in a future Internet is to make a major if not disturbing judgment.

All known historical societies have engaged in some form of musical activity. A society that neglects music probably does so to its own detriment. After all, music evokes emotion. Music provides identity. Music is intellectually stimulating. Audio is fundamental to many of our most important rituals and events. Memories involving audible phenomena also serve as powerful reminders of personal, national, and racial histories. Music has often been used as a means to communicate across linguistic and cultural barriers. As such, audio and music are core elements of understanding people. Music is also a powerful tool for learning and growth. In fact, recent research [1] has shown that music training can measurably increase performance by children and college students with such tasks as spatial reasoning.

It is commonly accepted that the World Wide Web has changed much of the world experience. The pace of this change exceeds that of any in the history of humanity. New technologies have been developed on this infrastructure as the demand for more compelling experiences continues to grow. Streaming media is the most notable of these technologies. While the concept of packet-based audio was considered at the onset in the Internet, streaming audio did not emerge as a significant Internet "feature" until 1995. The initial implementation relied on standard voice compression technology and had very low fidelity. In fact, many of the transmissions were simply inaudible or were too annoying to be useful. This technology has come far in subsequent years and now provides good fidelity audio at bandwidths in the range of ISDN connections. A new Internet needs to take a conscious step forward in this domain.

1.2 *Audio and music are a very demanding test case*

Musicians are extremely sensitive to very subtle (few-millisecond) shifts of temporal materials. They are also extremely sensitive to localization information. For example, a good orchestral conductor can isolate minor tuning imperfections among twenty or more string instruments. A truly gifted conductor will know which specific instrument in the section is sharp or flat. This is an impressive skill, which is learned by developing the ability to localize the position of competing sound signals using the temporal differences of arrival times of the sound at each of our ears. Audio over Internet2 needs to provide this kind of accuracy. As such, audio and music transmission will provide a critical test case, where errors can be readily perceived by one of our primary senses.

2 Where we are today

As we move from open-air analog-based audio transmission to packetized digital delivery systems, it is evident that there are many difficulties yet to be overcome.

The subtle details contained in audio require very high resolution storage formats. For example, a 3-minute song on standard audio CD requires 32 megabytes of audio data. A higher quality format for audio research could require 260 megabytes. This order of magnitude storage difference makes Internet audio file transfer impractical. To view this problem from a different perspective, Table 1 shows bit rates associated with various forms of storage and transmission.

Table 1. Important bit rates

CA*net3 CANARIE–Bell Canada	40 GBit/s	optical Internet Oct.1998
Internet2 US new national backbone	9.6 GBit/s	network for end of 1999
Internet2 US new national backbone	2.4 GBit/s	network for end of 1998
Internet2 US national backbone	622 Mbit/s	infractructure April 1998
ATM OC-12	622 Mbit/s	
ATM OC-3	155 Mbit/s	
100-BaseT / FDDI LAN	100 Mbit/s	
T3	45 Mbit/s	
10-BaseT Ethernet LAN	10 Mbit/s	
T1	1.5 Mbit/s	
Digital HDTV	40-60 Mbit/s	5.1 audio uncompressed
Next generation DVD (Blue Laser)	23 Mbit/s	
DVD-ROM	11.08 Mbit/s	
Digital DVD-Audio (uncompressed)	9.6 Mbit/s	6 channels max. 96 kHz, 24 bit
Digital TV, DVD-Video (NTSC)	6-10 Mbit/s	5.1 audio, NTSC video compr.
Multichannel audio compressed	224-640 kbit/s	5.1 channels, Dolby Digital
Compact disc	1.14 Mbit/s	44.1kHz, 16 bit, stereo
Stereo audio uncompressed	1.536 Mbit/s	48 kHz, 16 bit, stereo
Stereo audio compressed ("MP3")	20-128 kbit/s	MPEG-2 Layer 3
Normal telephone channel	64 kbit/s	mono, limited bandwidth
Telephone modem	14.4- 56 kbit/s	ITU V.90 modem 56 kbit/s
Cable modem (with Ethernet card)	50-200 kbit/s	up to 10 Mbit/s theoretical
ISDN	64-128 kbit/s	FM stereo quality
ISDB (Integrated Services Digital Broadcasting)	150 Mbit/s	NHK trans. 21 GHz/ch
ADSL (a new telephone service)	512 kbit/s	uses standard wires
ADSL high-speed modem	1 Mbit/s	
Program data	10 kbit/s	
Facsimile (fax)	20 kbit/s	
Still picture	70 kbit/s	
Tell Text	100-200 kbit/s	
Audio graphics	800 kbit/s/ch	

NOTE: Audio stream bit rate is expressed in bits/second (bit/s) or thousands of bits/second (kbit/s), or millions of bits per second (Mbit/s). A higher bitrate provides better audio fidelity but requires more Internet bandwidth. The required bit rate transmission capacity depends on the type of service, coding method, error correction, quality (resolution) of signal, number of channels, and other factors.

Time delays are caused by several factors—encoding and decoding latencies, routing and switching latencies and, of course, transmission latencies. To illustrate the implications, we point to an experiment conducted by a Canadian school board that had just installed services at the level of asynchronous transfer mode (ATM) among many high schools. A conductor at one school was to direct a class of singers at another over the system using digital video and audio feeds. The down beat was given and, predictably, the visual cue arrived late enough that the conductor was out of synchronization with the remote choir by a noticeable fraction of a beat. She quickly learned to adapt to the latency and the session continued. However, the system vendor learned at once that audio, and perhaps audio alone, is the most demanding test of temporal precision of the communications medium.

Beyond problems with time delays, the vast majority of Internet users cannot stream Internet audio material that matches the fidelity achievable with common home stereo systems. The required bandwidth far exceeds that of even relatively expensive ISDN connections. The reliability of Internet connections continues to degrade, even for high-bandwidth users, as the population of Internet users increases faster than the supporting network. Many pundits have predicted massive outages and failures due to this increased load. This sort of cataclysmic event is not necessary to render the Internet useless for high-quality audio applications. A single skip or pop will ruin a high-quality audio experience. As such, the network requirements for high-quality audio are difficult to meet.

As for music, at present musicians are exploring interactive performance with MIDI (Musical Instrument Digital Interface) [2] and with streamed audio. MIDI is a serial control protocol used by manufacturers of electroacoustic instruments, audio signal processors, and a host of computer music software products. The maximum bit rate is 32.5 kbit/s per 16 multiplexed channels of information. MIDI is sufficiently rich to permit complex control gestures (playing notes on a keyboard, tracking and encoding pitches from voice or orchestral instruments, moving a slider on a mixing console, turning a knob on a digital signal processor, manipulating banks of lights) to be captured and encoded as a serial data stream. It is also sufficiently fast to permit real-time compositional algorithms or sound signal processing routines running on personal computers to interact with live performance. Even though packet latencies are at least 1 ms and may exceed 20 ms, practical experience and countless live performances have proven that MIDI is a powerful mechanism for interactive music creation.

3 Technical issues in network transmission of music and audio

3.1 Passive or interactive?

Internet2 requires the infrastructure to support not only passive but also interactive audio experiences. A passive experience would be analogous to listening to a concert via streaming audio, or watching a movie with sound effects via an MPEG movie. An interactive experience requires the ability for the listener to affect the course of audio program creation or playback not only locally but even for other remote listeners as well. An example would be a quartet of musicians, each in remote locations, playing synchronous music while monitoring each other over the Internet2 network.

3.2 Latency

Interactive audio applications are dependent on a low-latency connection between network locations. The time between playing a note and its arrival at a remote location cannot exceed a small fraction of a second. It is very difficult to interact on a musical level when latencies exceed this threshold.

3.3 Time stamping

Continuously flowing audio streams, video streams, and discrete events between locations must be synchronized with each other.

3.4 Quality of service

From our own experience we know that the quality of service requirements for audio streams are particularly demanding. Lost or late data packets can easily destroy the continuity of the audio signal stream, thus seriously degrading the quality of the sound. Video streams are less sensitive to millisecond-range breaks in content continuity. Based on current practice, Internet2 will need to exceed a minimum point-to-point service at the level of OC-3 (ATM @ 155 Mbit/s). Transmitting IP packets directly, as is anticipated in the recently announced Canadian CANARIE network (www.canarie.ca), will alleviate the overhead associated with ATM and SONET (synchronous optical network) layers. This alternative should also be explored by Internet 2.

3.5 Audio formats

To maintain cost-effective efficient access to Internet2, the protocols selected for streaming audio must be compatible with conventional digital audio equipment. Today’s digital audio equipment streams linearly encoded PCM audio streams in a multiplexed two-channel format conforming to the AES3 (AES/EBU) standard. Additional audio channels are typically transported over multiple instances of AES3 (not multiplexed into one big stream). Today an emerging technology named IEEE 1394 is promising to replace AES3 as the multichannel digital audio interconnect, carrying multichannel data in packets conforming to the IEC 61883-6 standard. Internet2 audio protocols should be consistent with either AES3 or IEC 61883-6 protocols to enable low-cost reliable bridge devices.

The emergence of multichannel digital audio utilizing some form of perceptual coding (such as Dolby Digital, DTS, and MPEG audio) is another important consideration in specifying audio delivery protocols. Compressed audio formats are typically transported over a IEC 60958 (SPDIF) bus, in accordance with the IEC 61937 standard. Again, it is recommended that Internet2 utilize the packet formats specified by these standards to ensure compatibility with legacy digital audio equipment. Another consideration is the selection of audio data formats, sample rates, and encoding methods. Standard sampling rates include 32 kHz, 44.1 kHz, 48 kHz, 88.2 kHz, and 96 kHz. Standard data formats include linear PCM with 16-, 18-, 20-, 24-, and 32-bit word lengths. Internet2 should support all permutations of these sampling rates and data word formats, and in fact should support future formats when they someday appear. Support of 32-bit single-precision and 64-bit double-precision IEEE floating-point PCM samples is also recommended for optimal compatibility with personal computers.

Synchronization between the sending and receiving devices on the network is also important. It is necessary to convey the sampling rate of the digital audio signals as they are transmitted from one remote site to another. This is typically achieved by transporting time stamps with the audio samples, or by referencing sample rates to a global clock such as GPS. Time alignment is also necessary in some applications (such as a simulcast between disparate audio, video, and other data streams). The Internet2 must provide a quality of service (QOS) that ensures that audio data are delivered at a rate suitable for recovering sample rate clocks without jitter or dropouts, and maintaining proper time alignment between the signals.

Failure to observe these industry standard digital audio transport standards when specifying digital audio protocols for Internet2 could result in prohibitively complex bridge device implementations, thereby limiting Internet2 to high cost, highly complex installations. By following industry standards the Internet2 world can easily bridge to today's digital audio systems and institutions and provide a highly valuable service to the digital audio community.

3.6 Number of channels

For distributed performance to be meaningful, the auditory experiences of the performers and/or conductor must be simulated accurately at the remote site. The simulated space requires techniques quite different from those used in sound recordings. While we know little about truly realistic simulation, it is certain that many channels of audio and video will be required and that these channels must be carried in precise temporal synchronization for the remote experience to be accurate.

Each sound transmission should be identified according to the number of channels and the arrangement. (Consider that even in commercially available compression techniques such as Dolby Digital [AC-3], channel 2 is not always the right channel.)

If a tradeoff will have to be made between the number of channels and the quality of the signal within the available bitrate, the user should be given a choice of which characteristic to maintain at the highest level of priority, and which at the lower levels. In some situations, depending on the program material, the multichannel attribute may be the most important characteristic of sound from the music and sound perception point of view. Thus there should be the option to preserve the multichannel characteristic of the original at all costs in the transmission.

3.7 Client–server issues

Many tools have been developed and are still being developed that promise to deliver audio over networks for playback. For the most part, these systems use a client–server approach in which the materials to be delivered are contained on one or several servers of various types and delivered over a network to a client. The client is usually some type of personal computer or, in some cases, a dedicated playback device or yet another server.

3.7.1 Decompression requirements

Typical characterizations of decompression processing requirements specify some percentage of some commodity CPU such as a Pentium running at some specific clock speed. This does not always equate to being able to actually use a given decompression technique on a given machine and in most cases represents not the peak usage but an average CPU utilization. There is also the assumption of an entire machine dedicated to the task of decoding audio when in fact there may be background or foreground processes competing for resources. These competing processes can be unrelated operations, operating system functions, or even overhead from the audio receiver process.

A method is needed to provide an accurate profile and guarantee resources for audio decode. Resources come in two classes, negotiable and nonnegotiable. Typically nonnegotiable items refer to hardware capabilities. For instance, if a computer's sound system can handle only audio sampled at 22.05 kHz, 32 kHz, or 44.1 kHz, this profile information is nonnegotiable. The negotiable component of this may be the inclusion of an in-line real-time resample filter. The profile should include at a minimum the following items:

- Memory requirements
- CPU type
- Operating system
- Peak CPU utilization
- Maximum duration of peak CPU utilization
- Average CPU utilization
- Playback capabilities (16 bit 8 kHz to 48 kHz, stereo, etc.)
- Availability of hardware-assisted decode or encode
- Compression technique capability sets.

Once the audio server knows these things, it can determine the type of audio stream to supply to the client. In some cases this information will indicate that the various streams available from the server cannot satisfy the audio delivery request. In other cases, the server will use this information in order to secure resources for audio playback on the client machine that match one of the audio profiles within the server's range of delivery options. By delivering an audio stream based on negotiated parameters that are then guaranteed by the client to exist, the audio consumer can be guaranteed an experience that is trouble free while meeting the minimum standards set by the content author.

3.7.2 Standardization of content on server side

While major headway has been made in the area of advanced standard file formats for audio delivery, adoption of these formats is by no means settled. As an industry, producers of media should demand that their network broadcast equipment adhere to a standard format. Until server vendors have achieved an appropriate standard, we will be faced with the following problems:

- Content life cycles of 6 to 12 months
- Constant need for consumers to download new players
- Incompatibility between existing content and new servers from the same or different vendors

- Lowest common denominator content encoding
- Duplication of nearly identical applications from multiple vendors
- No standard way to utilize existing HTML or other code with embedded clients when changing from one server to another.

Formats such as Advanced Streaming Format contain advanced capabilities and can be readily extended to encompass new technologies as they are conceived. Using a single robust standard such as this will support competition while introducing some measure of compatibility.

3.7.3 Client architecture

Lack of a single standard client architecture is becoming a problem as more and more server vendors appear. While it can be argued that independent architectures provide a platform for rapid enhancement and expansion of capabilities by each server vendor, it is also clear that this moving target has stifled proliferation of content, rendering the advancements largely meaningless. It is possible to deliver clients that have the same interfaces without stifling competition. The modularization of network layers, compression techniques, and file read and write capabilities is advised. These modular components can then be plugged in to a standardized client as needed.

Many systems already use a modular design that allows decompression techniques from any vendor to be utilized by the existing client. Currently this modularity is on a per-client basis rather than a per-system basis. In other words, it is not enough to know what your CPU and operating system are. You also need to know which of many clients you will be using in order to obtain the proper decompression technique. This raises the complexity considerably for all parties involved. Once you have simplified component selection, automatic component download can be accomplished in a much more straightforward manner.

3.7.4 Copyright protection

By developing a standard client it will also become possible to implement some level of compatibility for audio copyright management and protection. A particularly compelling application that would benefit from this advancement includes pay-per-play streaming of high-quality audio. Consumers could access a much larger library of music than is currently feasible with physical media. Such an improvement in content access would surely provide new growth potential for the music industry. This lowered barrier to musical exploration would also result in a more diverse and creative musical landscape.

Current standardization efforts for secure systems have focused on building frameworks that support a diverse application space. This approach should also be taken within Internet2. This type of framework promotes interoperability between systems while not restricting the development of diverse secure applications. Current security systems have no level of interoperability, which results in incompatible content for the consumer. By employing such a generic security framework, clients could download and embed secure modules for specific applications in the same manner that was described previously for audio compression algorithms. The result would be an integrated client that could render content protected by multiple security systems.

4 Envisioned applications

We ride roughshod here over the question of whether Internet2 shall remain limited to the academic community. Since so many advances in audio and music are aimed at the consumer, surely the needs of the consumer and the commercial market will also push the envelope of the musical and audio capabilities of Internet2.

4.1 Digital libraries

We agree that "images, audio and video can, at least from a delivery point of view, move into the mainstream currently occupied almost exclusively by textual materials. This will also facilitate more extensive research in the difficult problems of organizing, indexing, and providing intellectual access to these classes of materials." [3]

One of the largest obstacles to the adoption of digital audio libraries is the lack of easy and rapid content location. There are many simple query-string-based approaches to this problem, but searching through a recorded audio library requires quick and high-quality rendering of content. One example of this application is the research into classical melodic structures. A student using this library needs to quickly compare and contrast melodic segments contained in a library. The reason for this quick comparison lies in the limitations of human music memory. Humans can retain certain aspects of a music experience for only a few seconds. As a result, the digital library system must be able to deliver each musical segment consistently and rapidly. Otherwise the research and learning experience will be significantly degraded. Furthermore, the quality of these musical pieces must be very high to allow distinction of performances and recordings.

Beyond text searches and content rendering, the capabilities for doing searches based on auditory and musical content have recently begun to be studied. For example, how do you search for a sound track containing a bell sound?

One company [4] has developed technology for content-based retrieval of audio [5]. Such a system can analyze all the audio in a digital library, representing each sound by a compact set of parameter values, which are then stored in a database. The user can construct various types of queries based on the audio content, including "query by example." With query by example, the user indicates the desired type of sound by selecting all or part of an existing sound file in the library, or even by recording a sample sound into a microphone. The system displays a list of all sounds in the library that are similar to the given sound, ordered by degree of similarity.

Such technology can also be useful for segmenting and classifying audio (whether in a file or in a live streaming input). For example, a sports broadcast can be segmented into useful chunks by identifying and distinguishing moments of cheering, narrative by different sportscasters, and other distinctive sorts of sounds. Thus, long, sonically heterogeneous recordings can be broken into segments of more or less homogeneous sound, either upon entry into the database or upon retrieval.

These technologies are clearly relevant not only in the context of a self-contained digital library, but also with distributed, Web-based resources. For example, an Internet2 search engine could allow a user to search the Web for sounds that are similar to a specified sound. To build the indexes for such a search engine requires periodic and automatic downloading of massive quantities of audio, a function that would benefit from the increased bandwidth of Internet2—as would the per-user auditioning of search results.

4.2 Three-dimensional audio

In addition to a rapid comparison of musical segments as mentioned, the system will need to render many audio segments simultaneously and in a complex manner. Users of interactive audio applications will require three-dimensional audio attributes like head-related transfer functions (HRTF) to provide a realistic and useful audio delivery platform. For example, in an educational VRML replication of an orchestra, children can move about the stage and visit the different instrument families. As they get closer to the brass section, they will hear and understand the sound of that family of instruments; off to the left they will hear the timpani, and they can choose to move in that direction based on the three-dimensional audio source.

Internet2 must support persistent environments where a dynamic change to the audio source is recognizable to all participants in real time. These changes must be persistent in that if a sound generator is affected by one user and later experienced by another user, the changes are audible to the second user.

4.3 Beyond virtual reality

Network-based real-time immersive audio-visual applications go a step beyond virtual reality that stretches the capabilities of current technologies. They provide an immersive (content-rich) framework, permitting one or more individuals to observe and interact with things or individuals elsewhere in remote locations. While somewhat similar to today's applications in teleconferencing, these applications differ in three important ways: they require high-resolution audio and video content streamed in two directions across a network, they depend on precise synchronization of their events and audio-visual content streams, and finally, they require a significantly higher degree of network "quality of service." Such applications share certain basic requirements with teleconferencing systems. In teleconferencing, systems using ISDN-based technology have proven to be effective, while for immersive audio-visual applications, they have not; bandwidth requirements and quality of network service are almost always the determining factors.

4.4 Broadcasting (audio webcasting)

We are in a very early stage with regard to radiolike services on the current Internet—more or less the equivalent of the very early crystal radio phase in the development of radio. Audio quality is decidedly low fidelity. Transmissions are interrupted for no discernible reason. Competing transmission systems use incompatible players. Computers are not ideal radio receivers. And so on. But the Internet has proven the concept that people want to communicate with more variety than is available from the “old” mass media. With the bandwidth of Internet2, broadcastlike audio content can be distributed in a viable fashion. This is as powerful an enabling technology for the electronic media as the World Wide Web has been for the print media.

Audio webcasting follows an important technology trend from analog through digital to packetized digital. In line with this, the existing mass-broadcasting paradigm may be replaced by something modeled after what has happened in the book and magazine publishing business. The megabookstores now popular in the United States carry tens of thousands of books and a few hundred magazines, some of which are aimed at a mass audience, but most of which are targeted to very specific reader needs and interests. The old mass magazines (*Life*, *Look*, *Saturday Evening Post*, and so on) are gone. Newsletters and “small-press” publications are growing rapidly, their cheap production enabled by the new technology of desktop publishing. That relatively few sources of audio programming are available currently is the result of a technology limitation, not for lack of a receptive market. Radio frequencies are a scarce commodity, and because they are valuable, they must be used in a way that optimizes numbers of listeners. Because they are dependent upon advertiser support, radio stations must appeal in the broadest possible fashion. There is no place, with the present limitations in the radio-frequency spectrum, to provide programming with anything like the wide variety the public is demanding from a modern bookstore. That is why Internet delivery is so compelling: there is no limit to the number of channels. As a source of information, the Web is superior to radio and television today; as a source of entertainment, the Web has already overtaken competing media for many “early adopters.”

Eventually it may even be possible to extend the Internet's reach to portable receivers. An analogy can be made to the existing print orientation of the Web: the Internet's variety compensates for lack of portability. Newspapers and magazines are often read nowhere near an Internet connection, so portability is an advantage there also, but this has not prevented the growth of the Web for text and graphics. Home, work, and university listening is likely to rise as the content becomes more varied and interesting. After all, the computer is not necessarily a bad audio receiver device. Audio can be listened to while a user is doing something else with the computer. It can be used as background, just as radio often is.

4.5 Audio production: standardization of Internet methods

Once the specifications for standard control protocols for studio mixing and production have been established, it will be possible to participate in the production process from a distance using the Internet. For example, mixing a sound file from a distance by sending mixing instructions over the Internet will be possible, while auditioning the results in real time.

An artist, engineer, producer, or teacher will be able to correct a mix from a distance by specifying a number of changes in the three-dimensional architecture of the mixdown file. This will be similar to the composer's ability to change a composition or performance (for example, by correcting a MIDI file). Soft access and control of any place in the mix will be available. Needless to say, a teacher located hundreds of miles away will be able to demonstrate changes to a mix to a group of students, or adjust a mix for each student separately.

4.6 Collaborative work

The standardization of audio compression techniques, for example, is a lengthy and expensive process. There is no purely objective measure for audio quality. As a result, audio compression algorithms must go through many stages of subjective listening tests. The listening tests take place at many worldwide locations and require the convergence of many professionals from around the world. In addition, the candidates for these tests come from a select group of "golden ear" listeners. These special listeners have the ability to detect problems in the audio that most humans would not hear. As such, it is critical to have them participate in these events.

If these listening tests could be conducted remotely throughout the world through a high-quality network, much time and expense could be saved. The audio material could be distributed to the remote locations via the network. The audio rating could also be logged in real time to a central scoring facility. This connection would remove many of the planning delays that are common in audio standards testing.

4.7 Music education

We note the following in the on-line documentation of Internet2:

> Interesting examples of learningware for the appreciation of music have been developed at several institutions. Migrating exemplars, such as those developed at ... Purdue University ..., to a Web-based environment is constrained by current limitations on the quality of streaming audio. Internet2 services could remove these constraints ... [6]

and:

> In an Internet2 environment, moreover, studio instruction in music also would have new opportunities. World-class musicians could be invited to offer their insight and expertise. For example, a two-way video/audio connection might link a high school jazz band with an artist-in-residence at a university. The high quality of the communication link would allow demonstration and critical review to occur. In addition, the students would literally be able to "jam" with the university-based instructor. This connection could be extended to musicians (whether students or professional artists) at additional locations. The instruction could be enriched by introducing recorded audio and video performances drawn from a network-based server. The student interaction with the instructor could be recorded for later review, either by the instructor or for practice by the students. [6]

This kind of application provides a real-time interactive high-quality audio-visual framework within which the player in one remote location is coached by an expert ("coach,” or even the composer) in another. Coach and player are separated by an arbitrary distance. The use of the application is effectively decoupled from notions of geographic distance, despite its degree of immersion. The implications of this on a cultural level are profound.

Even when coaching players of average ability, the coach's evaluation of the player's performance depends on the discernment of auditory detail operating on extremely fine time and pitch scales (on the order of milliseconds or cents). Though secondary, fine visual observations regarding fingering technique, posture, or breathing, for example, can also play an important role in the evaluation. The coach's keenness of discrimination relies on the stability and the time–frequency resolution of the audio and video signals received. The system must also provide the player with the same ability to evaluate the coach's musical performance (demonstrating technique).

In addition to the music curriculum, universities and colleges will be able to offer some of their sound recording curricula on the Internet, where participants will find entire lectures, experiments, demonstrations, exercises, or training programs fully accessible from a distance. For example, a student in a distant location will be able to learn about an MS microphone, audition and control the stereo MS microphone from a distance, audition the microphone's on- and off-axis responses, as well as learn about the design and construction of the MS microphone without ever seeing one in real life. The student will also use an equalizer or a compressor, or reverberate a sound, without having to purchase the equipment.

4.8 Broadening musical participation

Interactive music can open up music enjoyment and appreciation to the greater population who never had the time or opportunity to learn to play music. One company [7] has developed computer-assisted music performance technology that enables people of a wide range of musical abilities, from professionals to beginners, to perform and interact with music. By enabling an Internet2 browser with such technology, the following benefits can be realized:

- By providing an adaptable musical control layer, such technology empowers musicians to increase their physical performance abilities exponentially in terms of dexterity, harmonic depth, and rhythmic accuracy.
- Equipped with an extensive musical database, this technology enables musicians to exploit sophisticated chord voicings and scalar patterns well beyond their current musical vocabulary.
- The software assists musicians in original and authentic expression while performing new and unfamiliar stylistic material.
- The program allows musicians a migration path to other musical disciplines, that is, percussion players using their rhythmic skills to perform dynamic harmonic material in real time.

On a larger scale, the Internet2 network could facilitate collaborative television production and simultaneous performance of music by artists placed anywhere in the world. A very ambitious and successful project linking musicians from seven locations spread all over the globe was conducted during the opening ceremony of the Nagano Winter Olympics. Using ISDN (384 kbit/s) and satellite television technology for audio and video signals, a chorus in Berlin, Cape Town, Beijing, New York (United Nations), and Sydney performed the final movement of Beethoven's *Ninth Symphony* together with the Winter Orchestra (a full symphony orchestra, chorus, and soloists) at the Nagano Kenmin Cultural Center Hall. Furthermore, a 2000-person chorus located at the Olympic Stadium joined the performance, which was mixed for a live broadcast to viewers around the world.

The setup coordinated by NHK involved the use of seven satellites, twelve uplinks with encoders and decoders, and multiple-delay compensation. Audio and video signals were sent from Nagano's International Broadcasting Center via an optical fiber network to Tokyo and further via the KDD High-Reliability Network to satellite uplink stations at Yamaguchi and Ibaragi. There the signals were beamed to seven satellites which retransmitted the signals to stations in Germany, South Africa, China, the United States, and Australia, as well as return signals back to Japan. The Line-Operation Centre receiving the signals added delays to all signals until they matched the delay of the longest transmission to and from Berlin. Thus the video and sound of the orchestra playing in the Nagano Kenmin Cultural Center Hall and the five choruses were projected on a big screen in the Olympic Stadium and via a public-address system, and also sent to the broadcasting center where pickup of the chorus from the Stadium was added.

From the technical standpoint, the event demonstrated the absolute need for precise compensation of transmission delay of audio and video signals to achieve perfect synchronicity. As well, the issues of reliability of telecommunication systems and compatibility between audio, video, digital formats, compression systems, interfaces and standards became very apparent. The high cost of this operation and its complexity demonstrated the need for more effective and economical ways of networking high-quality audio and video on a worldwide basis.

4.9 Forensic applications

Internet-based forensic audio laboratories could offer services of decomposing and segregating audio signals, removing reverberation or noise from sound recordings, plus other processes involving pattern recognition, using complex signal processing engines and advanced software.

4.10 Automatic update of software tools

Manufacturers of audio equipment will offer new software tools for sound processing, making their new product available to the consumers faster than it takes to build a hardware version. These "plug-in" software devices will be "auditioned" over the Internet, and used as "demo" units limited only by their expiration date, finally purchased directly from the manufacturer using the Internet. All modifications or upgrades to these soft tools will be automatically sent to every registered user. Learning of a new product that has just become available will be as quick as finding a new icon on the desktop, ready for testing at a click of a button.

4.11 Scientific and engineering data representation

One part of the Internet2 proposal [6] foresees, in the context of scientific experiments, "tools for synchronizing temporal data (such as music) with related text and images (such as musical scores)." Actually, the field of sonification (or audification) is active [8] [9] if young. Our ears are adept at hearing many places in space at once, including receiving information around corners and through solid walls, and detecting events at distant locations. Moving this concept to the Internet, we should be able to model, render, and transmit an auditory portrait of our virtual surroundings. Sound can depict a composite of information surrounding us in cyberspace, whereas our visual browsing is accomplished only one page at a time. Sound can provide vast amounts of information while we are still waiting for a few image frames to download. And when audio is synthesized locally, only the control instructions need to be transmitted from a remote location, further upping the interactive ante of audio over video. The Internet is not only a topology of personalized sources and sinks replacing the impersonal radio stations and record stores of yesterday. The topology is itself a metaphoric space for audio signals to propagate and resonate, capable of informing a listener of other listeners and players on-line, and indicating the relative locations of sites with desired information. The subsystems that we specify and build should be capable to treat the Internet as a meta-instrument that resonates with sound propagation, and provides immediate feedback from interactive interface devices, as well as transmitters and receivers of prepackaged media. The parallel emphasis of local computation on the one hand and streaming of remote signals and control messages on the other captures both the broadcast metaphor and the music performance metaphor. The construction of a dual-metaphor cross-platform tiered approach to an "audio operating system" can sustain a new Internet well into the next century.

5 Acknowledgment

Elizabeth Cohen, Chair of the Future Directions Committee and past president of the AES, provided the initial inspiration and impetus for the preparation of this paper.

6 References

1. F. H. Rauscher, G. L. Shaw, L. J. Levine, K. N. Ky, and E. L. Wright, "Music and Spatial Task Performance: A Causal Relationship," presented at the American Psychological Association 102nd Annual Convention, Los Angeles, CA (1994). http://gopher.tmn.com:70/0/Artswire/AMC/MUSBRAIN/RESEARCH/intro; http://www.amc-music.com/srstudies.htm.

2. MIDI Specification, MIDI Manufacturers Association, PO Box 3173, La Habra, CA (1995). http://www.midi.org/specinfo.htm.

3. www.internet2.edu/html/digital_libraries.html.

4. www.musclefish.com.

5. E. Wold, T. Blum, D. Keislar, and J. Wheaton, "Content-Based Classification, Search, and Retrieval of Audio," *IEEE Multimedia*, vol. 3, no. 3, pp. 27–36 (1996).

6. www.internet2.edu/html/learningware.html.

7. www.hotz.com.

8. G. Kramer, Ed., *Auditory Display: Sonification, Audification and Auditory Interfaces* (Addison-Wesley, Reading, MA, 1994).

9. http://www.santafe.edu/~kramer/icad/Biblio/Bibliography.html.

7 Contact information

Robin Bargar
Audio Development and Virtual Environments
NCSA, Beckman Institute
University of Illinois at Urbana-Champaign
rbarger@ncsa.uiuc.edu

Steve Church
Telos Systems
2101 Superior Avenue
Cleveland, OH 44114
Phone: 216-241-7225
steve@telos-systems.com
http://www.zephyr.com

Elizabeth Cohen
Cohen Acoustical Inc.
Los Angeles, CA 90004
Phone: 213-939-9825
akustik@primenet.com

Akira Fukada
NHK Broadcasting Engineering Department
Sound Section
NHK Broadcasting Center
2-2-1, Jinnan, Shibuya-ku, Tokyo, 150
Japan
akira_f@ta2.so-net.ne.jp

James Grunke
President, Hotz Interactive
Chairman, MIDI Manufacturers Association
142 Marylinn Drive
Milpitas, CA 95035
Phone: 408-942-8494
mapleinc@best.com

Douglas Keislar
Muscle Fish, LLC
2550 Ninth Street, Suite 207 B
Berkeley, CA 94710
Phone: 510-486-0141
doug@musclefish.com
http://www.musclefish.com/

Bob Moses
PAVO
95 Yesler Way
Seattle, WA 98104
Phone: 206-682-7705
bob@pavo.com

Benjamin P. Novak
Developer Relations Group
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052
Phone: 425-271-8748
bennovak@microsoft.com

Bruce Pennycook
Information Systems and Technology
McGill University
845 Sherbrooke Street West
Montreal, PQ, H3A 2T5
Canada
Phone: 514-398-3862
pennycook@ist.mcgill.ca

Zack Settel
Faculty of Music
McGill University
555 Sherbrooke Street West
Montreal, PQ, H3A 1E3
Canada
Phone: 514-398-4535 ext. 5633
zack@music.mcgill.ca

John Strawn
S Systems Inc.
15 Willow
Larkspur, CA 94939
Phone: 415-927-8856
ssys@netcom.com

Philip R. Wiser
Liquid Audio, Inc.
810 Winslow Street
Redwood City, CA 94063
pwiser@liquidaudio.com
http://www.liquidaudio.com

Wieslaw Woszczyk
Faculty of Music
McGill University
555 Sherbrooke Street West
Montreal, PQ, H3A 1E3
Canada
Phone: 514-398-4535 ext. 0507
wieslaw@music.mcgill.ca

About the Technical Council of the Audio Engineering Society

The Technical Council and its Technical Committees respond to the interests of the membership by providing technical information at an appropriate level via conferences, conventions, workshops, and publications. They work on developing tutorial information of practical use to the members and concentrate on tracking and reporting the very latest advances in technologies and applications. Technical Committees strive to create and maintain awareness of the problems, conditions, standards, etc., in the areas vital to the activities of the AES, so that valuable work is performed and is made known.

The Technical Council coordinates the activities and policies of the Technical Committees, advises the Board of Governors by formal resolutions on matters of policy involving technical considerations, and serves as a liaison body for technical contacts and exchange between groups in the Society and those of other organizations with similar or related interests.

Technical Committees are currently appointed for the following fields of interest (and will vary as the art evolves):

Acoustics and Sound Reinforcement
Coding of Audio Signals
Loudspeakers and Headphones
Microphones and Applications
Multichannel and Binaural Audio Technologies
Network Audio Systems
Optical Recording
Perception and Subjective Evaluation of Audio
Signal Processing
Studio Practices and Production
Transmission and Broadcasting