# AES White Paper:<br>Best Practices in Network Audio

AESTD1003v1 2009-06-04

Nicolas Bouillot, McGill University
Jeremy R. Cooperstock, McGill University
Andreas Floros, Ionian University
Nuno Fonseca, Polytechnic Institute of Leiria
Richard Foss, Rhodes University
Michael Goodman, CEntrance, Inc.
John Grant, Nine Tiles Networks Ltd.
Kevin Gross, AVA Networks
Brent Harshbarger, CNN
Joffrey Heyraud, Ethersound
Lars Jonsson, Swedish Radio
John Narus, VidiPax LLC
Michael Page, Peavey Digital Research
Tom Snook, New World Symphony
Atau Tanaka, Newcastle University
Justin Trieger, New World Symphony
Umberto Zanghieri, ZP Engineering srl

Technical Committee on Network Audio Systems

# Contents

## Executive Summary

Analog audio needs a separate physical circuit for each channel. Each microphone in a studio or on a stage, for example, must have its own circuit back to the mixer. Routing of the signals is inflexible. Digital audio is frequently wired in a similar way to analog. Although several channels can share a single physical circuit (e.g., up to 64 with AES10), thus reducing the number of cores needed in a cable. Routing of signals is still inflexible and any change to the equipment in a location is liable to require new cabling.

Networks allow much more flexibility. Any piece of equipment plugged into the network is able to communicate with any other. However, installers of audio networks need to be aware of a number of issues that affect audio signals but are not important for data networks and are not addressed by current IT networking technologies such as IP. This white paper examines these issues and provides guidance to installers and users that can help them build successful networked systems.

# 1 Background

Network audio technologies find application across a wide number of domains, including studio recording and production activities, archiving and storage, musical performance in theater or concert, and broadcasting. These range in scale from desktop audio production systems to 500 acre theme park installations.

Each domain imposes its own, sometimes conflicting, technical requirements. For example, the low-latency demands of live sound can be at odds with the flexibility and interoperability that is attractive to commercial installations. Similarly, recording studios demand impeccable audio performance and clock delivery that is not required in other domains.

The architectures that support network audio are, at present, largely confined to ethernet-based solutions, with a few alternatives including IEEE 1394. Differences between the various implementations include issues of audio format, number of channels supported, and end-to-end transport latency. In addition to the large number of low-fidelity consumer applications for network audio streaming, various software architectures have been developed to support professional grade audio transport over Internet, albeit without the higher performance that characterizes local area network solutions.

## 1.1 Marketplace

Outside the telecommunications industry, audio networking was first conceived as a means of transferring work units within large studios and post-production facilities. In the 1990s when the digital audio workstation became the dominant audio production platform, standards based Token Ring (IEEE 802.5) and Ethernet (IEEE 802.3) networks were employed to transfer audio files from one workstation to another. Initially these were non-real-time transfers but facility operators, technology providers and networking standards grew to allow real-time playback over the network for the purpose of media storage consolidation, final mix-down and production.

True real-time audio networking was first introduced in installed sound reinforcement applications. By the mid-1990s, digital signal processing was in widespread use in this market segment [12][19]. Digital signal processing improved the flexibility and scalability of sound reinforcement systems. It became possible to infuse facilities such as theme parks and stadiums with hundreds of individually processed audio signals, and thus, the need arose for a flexible, optimized distribution system for these signals. Since audio was now processed in digital format, logically the distribution system should be digital as well.

Live sound applications were the last to adopt digital technology. However, with the introduction of digital consoles and distribution systems targeted to live sound applications, this has changed.

The present marketplace for digital audio networking might be described as fragmented, with multiple technology players. This is the result of disparate performance requirements, market forces and the nature of technology development.

## 1.2 Challenges of audio networks

Despite the maturity of data networks and their capability of transporting a wide variety of media, transport of professional audio often proves quite challenging. This is due to several interrelated characteristics distinct to audio:

the constraint of low-latency, the demand for synchronization, and the desire for high fidelity, especially in the production environment, which discourages the use of lossy encoding and decoding (codec) processes.

A stereo CD quality audio stream (16 bit resolution, 44.1 kHz sampling) requires 1.4 Mbps[1] of data throughput, a quantity easily supported by existing wired LAN technologies although often not by commercial WAN environments. Available bandwidth may be exceeded when the number of channels, resolution, or sampling rate are increased, or when the network capacity must be shared with other applications.

Low latency is an important requirement in audio. Sources of latency include A/D and D/A converters, each of which adds approximately 1 ms, digital processing equipment, and network transport. In a typical chain, audio may need to be transported from a source device to a processing unit, then to a mixer, and from there, to an amplifier. In a digital audio network, latency is increased by the number of such segments that the data must traverse. Many system designers subscribe to a less-is-more philosophy when specifying or evaluating system latency, using zero latency analog systems as a reference point. If necessary, latency can always be added with a delay element, but once introduced, latency can never be removed. This requirement severely restricts the amount of computation and look-ahead that can be performed to achieve data reduction.

Audio networks, as any other digital audio system, need signals to maintain synchronization over the entire system, ensuring that all parts are operating with the same number of samples at any one time. Synchronization signals may also be used to detect the exact moment when A/D and D/A converters should read or write their values. In most digital systems, synchronization signals are transmitted at the sampling rate, i.e., one sync signal for each sample. In audio networking, such an arrangement is often impossible to support, resulting in a more complex clock recovery task.

Flexibility and robustness are further considerations, in particular, when dealing with a large number of nodes. For this reason, recent networks typically adopt a star topology. This provides flexibility in adding, removing, and troubleshooting individual connections, and offers the benefit that, at least for a simple star, traffic between any two devices does not interfere with others.

Each of these requirements may impact the others. For example, latency may be reduced by transmitting fewer audio samples per packet, which thus increases overhead and bandwidth. Alternatively, switches may be avoided by changing the network topology from a star to a bus, which in turn decreases flexibility. The quality of synchronization may be increased with the addition of sync signals over the same medium as used for audio transport, thus rendering the network incompatible with generic data network standards.

# 2 Network technologies

For the purpose of this white paper, a network is defined as a collection of more than two devices, dispersed in space, and connected together via hardware and software, such that any device can communicate with any other connected device. The principal benefit of a network is that resources and information can be shared. Networks are often characterized by a combination of their size (e.g., personal area, local area, campus area, metropolitan area, wide area), transmission technology and their topology (e.g., star, bus, daisy-chain).

Network audio technologies have been based wholly or partially on telecommunications and data communications standards. For example, AES3 is based on an RS 422 electrical interface and AES10 (MADI) is based on the electrical interface developed for the Fiber Distributed Data Interface (FDDI) communications standard.

The remainder of this section provides further discussion of several important aspects of network technologies, both theoretical and practical.

## 2.1 OSI model

The International Organization for Standardization (ISO) created a seven-layer model called the Open Systems Interconnection (OSI) Reference Model to provide abstraction in network functionality. Many popular professional Ethernet-based audio networks only use the Physical and Data Link Layers (layers one and two) of the model, as described in further detail below:

1. The *Physical Layer* provides the rules for the electrical connection, including the type of connectors, cable, and electrical timing.

[1] Note that this figure ignores packet headers and tails, control signals and possible retransmissions.

2. The *Data Link Layer* defines the rules for sending and receiving information across the physical connection of devices on a network. Its primary function is to convert a stream of raw data into electrical signals that are meaningful to the network. This layer performs the functions of encoding and framing data for transmission, physical addressing, error detection and control. Addressing at this layer uses a physical Media Access Control (MAC) address, which works only within the LAN environment.

3. The *Network Layer* defines protocols such as Internet Protocol (IP) and X.25 for opening and maintaining a path to the network between systems. At this layer, endpoints are specified by inter-network addressing. This layer is responsible for the connectionless transfer of data from one system to another across a single hop, but not for reliable delivery. It is also responsible for fragmenting data into suitably sized packets as necessary for transport.

4. The *Transport Layer* provides additional services that include control for moving information between systems, additional error handling, and security. The Transport Control Protocol (TCP) is a well-known connection-oriented Transport Layer protocol that ensures transmitted data arrives correctly at the destination. The User Datagram Protocol (UDP) is a connectionless protocol, which does not provide any acknowledgements from the destination. Although UDP is considered less reliable than TCP because it provides no guarantees of packet delivery, it is often preferred for streaming media applications because of its superior real-time performance. The popular Real-time Transport Protocol (RTP), used in many streaming media applications, defines a packet format for network delivery of audio or video, which itself is typically carried inside of UDP. RTP is often used in conjunction with the RTP Control Protocol (RTCP) to provide feedback to the sender regarding quality of service as required for session management.

5. The *Session Layer* coordinates the exchange of information between systems and is responsible for the initiation and termination of communication *dialogues* or *sessions*. The Session Initiation Protocol (SIP), used commonly in Voice over Internet Protocol (VoIP), streaming multimedia, and videoconferencing systems, resides at this layer.

6. The *Presentation Layer* protocols are a part of the operating system (OS). These negotiate the use and syntax that allows different types of systems to communicate, for example, by dealing with the conversion between application-level representations and network data representations.

7. The *Application Layer* includes a range of network applications such as those that handle file transfers, terminal sessions, and message exchange, including email.

## 2.2 Transmission schemes

Data transmission may take place over wired media, such as copper wire and fiber optic connections or wireless media, such as terrestrial radio and satellite. In either case, the transmission scheme used can be classified as asynchronous, isochronous or synchronous.

*Asynchronous* communications are non-real-time such as web browsing and e-mail transport. The general purpose nature of asynchronous systems gives these transport technologies a wide market, high volumes and low costs. Examples of asynchronous communications include Ethernet, the Internet, and serial protocols such as RS-232 and RS-485.

*Synchronous* communications systems are specialized, purpose-built systems. They can be the best solution for a well-focused data transport application, but are ill-suited for supporting a mixture of services. Carrying asynchronous data on a synchronous transport is inefficient and does not readily accommodate bursty traffic patterns. Examples of synchronous communications systems include AES3, AES10, ISDN, T1 and the conventional telephone network.

*Isochronous* communication is required to deliver performance that is quantified in a service agreement on the connection between communicating nodes. The service agreement specifies parameters such as bandwidth, delivery delay and delay variation. Isochronous transports are capable of carrying a wide variety of traffic and are thus the most versatile networking systems. Examples of systems supporting isochronous communications include ATM, IEEE 1394, and USB.

## 2.3 Topology

Common network topologies include:

*Bus:* A single common wire connects all devices. Although this topology is rarely used now, early Ethernet devices used coaxial cable that was connected to each device with a T-connector or "viper" tap. This should not be confused with modern *virtual* bus systems, such as IEEE 1394 or USB.

*Star:* Devices are interconnected through centrally located network distribution hubs, such as repeaters, switches, or routers. Hubs may also be interconnected to create a star-of-stars topology. Most Ethernet connections today are designed in some form of a star configuration.

*Daisy-chain:* Devices are connected end-to-end. Setup and wiring is simple as no separate network equipment is required. However, failure of a single device or connection in a daisy chain can cause network failure or split the system into two separate networks.

*Ring:* Similar to a daisy-chain, but with the last device connected to the first to form a ring. On failure of a single device or connection, the ring reverts to a still functional daisy-chain.

*Tree:* Similar to a daisy-chain, but devices are allowed to make connections to multiple other devices, provided that no loops are formed.

*Spanning-tree:* Any network topology is permitted. The network features distributed intelligence to deactivate individual links to produce a working spanning-tree star-of-stars configuration. Upon failure of network links or components, the network can automatically reconfigure, restoring deactivated links to route around failures.

*Mesh:* Any network topology is permitted. Routing algorithms insure that traffic moves forward to its destination efficiently utilizing all links and does not get caught in any of the loops in the topology.

## 2.4 Routing

The routing options supported by a network are an important measure of its usefulness and flexibility.

*Point-to-point* or *unicast* routing, allows direct communications between one sender and one receiver. Point-to-point connections may be defined statically by the network configuration or dynamically, as is the case when placing a telephone call or retrieving a page from a web server.

*Point-to-multipoint* or *broadcast* routing supports asymmetric communication from a sender to one or more receivers. This scenario is exemplified by radio broadcast or a conventional analog cable television distribution system. A variation known as *multicast* allows specified devices to join a virtual group, such that a message sent by one member of the group goes to all other members of the group, but not to any other devices. This allows some of the benefits of broadcasting without swamping other network devices with irrelevant messages.

*Multipoint-to-multipoint* routing allows every node to broadcast or multicast to all others.

## 2.5 Data transport management

Devices that are connected to audio networks can be sources and/or destinations of digital audio. These devices may generate audio themselves, as in the case of studio synthesizers, or they could have a number of plugs from which audio is sourced. As we have seen in previous sections, various technologies provide the means to transmit and receive the audio samples over a physical network. Data transport management deals with the management of addressing, encapsulation and subsequent extraction of the networked audio.

Data transport management technologies allow the user to select a particular audio channel from a cluster of channels transmitted over the network. Management protocols such as SNMP (IP-based) or AV/C (specific to IEEE 1394) allow differing degrees of flexibility, for example, limiting the number of audio channels that may be transmitted, the size of packets containing the audio samples, or the association between channel numbers and output plugs at the destination. The routing of audio streams can be presented to users in the form of graphic workstation displays, in a manner appropriate to the particular application domain.

## 2.6 Network service quality

Quality of Service (QoS) is a measure of how reliably the data transmitted on the network arrives at its destination. The parameters by which QoS is measured include the data rate, expressed in bits per second or packets per second, the latency, which is the time interval between a data item being transmitted by the source and received by the destination, and the proportion of data that is lost (never arrives at the destination). Both the long-term value and short-term variations are significant in each case.

For activities such as web surfing, QoS is relatively unimportant. Delays of a second or two will be masked by random delays in other parts of the chain such as waiting for access to a server, or for the retransmission of lost or corrupted packets [11]. However, for live audio, all aspects of QoS are important. The available data rate must be sufficient to convey the bits as fast as they are produced by the source, without interruption. In any system, a sudden, unexpected, increase in latency can cause a drop out in the signal at the destination. Latency is most critical where the output is related to live sound (see Sections 4.4 and 4.7), as there is no opportunity to retransmit lost data.

If a defined QoS it required for a particular flow such as an audio stream, network resources must be reserved before the destination begins receiving it. This is straightforward with connection-oriented technologies such as ISDN and ATM. Less comprehensive QoS capability is available on Ethernet and IP via 802.1p, DSCP, RSVP and the like. A more comprehensive QoS solution for Ethernet is currently under development as part of the IEEE 802.1 working group's AVB initiative.

All communication over ISDN is via fixed-rate channels with fixed latency and high reliability. If the required data rate is more than the channel capacity, several channels are aggregated together. On an ATM network, there is a call set-up procedure during which the source equipment specifies what QoS is required. Both these circuit-switched technologies are perceived as obsolescent and beginning to be withdrawn in favour of IP [24][5]. In this direction, Multiprotocol Label Switching (MPLS) over Ethernet may be considered a viable replacement.

A service that has no defined QoS is described as "best effort": the network tries its best to deliver the data but can make no guarantees. Private networks may be overspecified to ensure an adequate bandwidth margin and may prevent congestion by segregating other traffic. In this case, throughput and reliability will be as good as on a network with QoS, although latency may remain higher on a store-and-forward packet routing network than on a circuit-switched network.

However, under best effort conditions, audio streaming must allow for sudden changes in latency by providing adequate buffering at the receiving end. Redundant information (e.g., Reed-Solomon forward error correction) can be introduced into the stream to mitigate against the problems of packet loss, but this increases the transmitted data rate which may make packet loss more likely and increase jitter.[2]

Strategies for achieving stream-specific QoS, including differentiated service and MPLS, rely on resource reservations. This requires that networked audio applications are deployed in a centrally configured and fully controlled network, hence restricting such solutions to relatively small networks.

At a larger scale, or when the network configuration remains uncontrolled, QoS can be optimized at the Transport Layer only, i.e., by end-to-end protocols. As an example, TCP includes a congestion control algorithm that evaluates network resources and adjusts sending rate in order to share available bandwidth with coexisting streams. This is fine for file transfer, where data are received ahead of the time they are required, and played out locally. However, for purposes of time synchronization, networked audio over Internet is often carried over UDP or RTP/UDP, which do not include native congestion control algorithm.[3]

## 2.7 Wireless

Digital audio streaming over packet-based networks may benefit from the adoption of wireless networking technologies, most obviously by the elimination of interconnection cables.

The first popular wireless standard for large-scale digital audio communication was that of cellular phone networks. However, these employ heavy lossy compression, geared toward the transport of speech data, typically operating in the range of 5.3 or 6.3 kbps for G.723.1 encoding and 8 kbps for G.729A encoding.

On a smaller physical scale, a Bluetooth network or *piconet* of up to eight active members is characterized by random hopping between 79 frequencies, allowing for simultaneous operation of multiple such networks in the same

[2]The EBU committee N/ACIP (`http://wiki.ebu.ch/acip/Main_Page`) is studying this issue.

[3]The TCP-Friendly Rate Control protocol (TFRC), currently under specification by the IETF Audio/Video Transport (avt), may offer a solution to this problem.

location. Piconets carry both asynchronous data and audio, the latter transmitted over synchronous connection oriented (SCO) channels [1]. The specification supports a maximum of three full-duplex 64 kbps SCO links. These employ either logPCM (A-law or u-law) speech coding or Continuous Variable Slope Delta (CVSD) modulation coding. Despite enhanced quality and latency compared to G723.1 and G.729A codecs, limited audio quality and bluetooth range make this technology currently unsuitable for professional audio applications.

For uncompressed audio transmission, the IEEE 802.11 WLAN specification [13] represents one of the most promising networking formats, due to its wide adoption in many digital consumer electronic and computer products, as well as the continuous ratification of enhancements in many state-of-the-art networking aspects, such as high-rate transmission, security, and adaptive topology control. Typical theoretical bitrate values currently supported by the 802.11 families of protocols include 54 Mbps (802.11a/g) and 100-210 Mbps for 802.11n, rendering them applicable for uncompressed quality audio.

Audio delivery in WLAN environments introduces a number of implementation issues that must be considered for the realization of high-fidelity wireless audio applications. For example, wireless transmission range is currently limited to distances up to 100 m, while non-line-of-sight transmission is further limited. Electromagnetic interference is also a significant factor that affects both throughput and delay performance. QoS enhancements are provided by the recently ratified 802.11e specification [8].

The IEEE 802.16 (WiMAX) standard [14] provides native QoS capabilities and connection-oriented forwarding. Typical bitrate values currently supported vary between 32 and 134 Mbps. Transmission range and power requirements are approximately triple and 10 times that of 802.11, respectively.

# 3 Current audio networking systems

Table 1 presents an overview of various audio networking systems available today. It should be stressed that obtaining an objective comparison of latency measurements is problematic, given that the various technologies operate with different constraints in terms of the underlying network medium and the limits imposed by hardware outside of their control. Where relevant information was available to elaborate on the measurements, these are indicated as notes. Unless otherwise specified, all measurements are intended to be exclusive of delays resulting from analog/digital conversion and DSP and all channel counts are assumed to be bidirectional (input and output).

Certain point-to-point technologies, such as AES3 and AES50, can appear as networks through the use of intelligent routers. Taking advantage of such hardware, these technologies can and are being used for effective large-scale multipoint-to-multipoint audio interconnection. In this case, the distinction between networked and point-to-point communication is somewhat blurred. However, a dedicated router implies higher cost, less flexibility, a potential single point of failure, as well as a need to communicate with one or more routers to affect routing changes.

Table 1: Audio Network Technologies Matrix

| Technology | Transport | Transmission scheme | Mixed Use Networking | Control Communications | Topology[1] | Fault Tolerance | Distance[2] | Diam. | Network capacity | Latency | Max available sampling rate |
|---|---|---|---|---|---|---|---|---|---|---|---|
| AES47 www.aes.org www.ninetiles.com | ATM | isoch. | coexists with ATM | Any IP or ATM protocol | Mesh | provided by ATM | Cat5=100m, MM=2km, SM=70km | ∞ | ∞ | 125 $\mu$s per hop | 192 kHz |
| AES50 www.aes50.com | Ethernet physical layer | isoch. or synch. | dedicated Cat5 | 5 Mbps ethernet | Point-to-point | FEC, redundant link | Cat5=100m | ∞ | 48 channels | 63 $\mu$s | 384 kHz and DSD |
| Note: Ethernet transport is combined with a proprietary audio clock transport. AES50 and HyperMAC are point-to-point audio connections, but they bridge a limited bandwidth of regular Ethernet for the purpose of control communications. An AES50/HyperMAC router contains a crosspoint matrix (or similar) for audio routing, and an Ethernet switch for control routing. The system topology may therefore follow any valid Ethernet topology, but the audio routers need a priori knowledge of the topology. While there are no limits to the number of AES50 routing devices that can be interconnected, each hop adds another link's worth of latency, and each router device needs to be controlled individually. | | | | | | | | | | | |
| AudioRail www.audiorail.com | Ethernet Physical Layer | synch. | Cat5 or fiber | Proprietary | daisy chain | none | Cat5=100m, MM=2km, SM=70km | ∞ | 32 channels | 4.5 $\mu$s + 0.25$\mu$s per hop | 48 kHz (32 ch), 96 kHz (16 ch) |
| Aviom Pro64 www.aviom.com | Ethernet Physical Layer | synch. | dedicated Cat5 and fiber | proprietary | daisy-chain (bidirectional) | redundant links | Cat5e=120m, MM=2km, SM=70km | 9520 km | 64 channels | 322 $\mu$s + 1.34 $\mu$s per hop | 208 kHz |
| Note: The network diameter figure is the largest conceivable network using fiber and 138 Pro64 merger units; derived from maximum allowed response time between control master and furthest slave device. Pro64 supports a wide variation range from the nominal sample rate values (e.g., 158.8 kHz - 208 kHz). | | | | | | | | | | | |
| CobraNet www.cobranet.info | Ethernet data-link layer | isoch. | coexists with Ethernet | Ethernet, SNMP, MIDI | Spanning tree | provided by 802.1 | Cat5=100m, MM=2km, SM=70km | 7 hops, 10 km | ∞ | 1.33 and 5.33ms | 96 kHz |
| Note: Indicated diameter is for 5.33 ms latency mode. CobraNet has more stringent design rules for its lower latency modes. Requirements are documented in terms of maximum delay and delay variation. A downloadable CAD tool can be used to validate a network design for a given operating mode. Network redundancy is provided by 802.1 Ethernet: STP, Link aggregation; redundant network connections (DualLink) and redundant devices (BuddyLink) are supported. | | | | | | | | | | | |
| Dante www.audinate.com | any IP medium | isoch. | coexists with other traffic | IP | any L2 or IP | provided by 802.1 + redundant link | Cat5=100m, MM=2km, SM=70km | ∞ | 700 channels | 84$\mu$s | 192 kHz |
| Note: Channel capacity is based on 48 kHz/24-bit sampling and operation on a 1 Gbps network. The latency value is based on 4 audio samples with this configuration. Note that latency is dependent on topology and bandwidth constraints of the underlying hardware, for example, 800 $\mu$s on a 100 Mbps Dolby Lake Processor. | | | | | | | | | | | |
| EtherSound ES-100 www.ethersound.com | Ethernet data-link layer | isoch. | dedicated Ethernet | Proprietary | star, daisy chain, ring | fault-tolerant ring | Cat5=140m, MM=2km, SM=70km | ∞ | 64 | 84–125 $\mu$s + 1.4 $\mu$s/node | 96 kHz |
| Note: EtherSound allows channels to be dropped and added at each node along the daisy-chain or ring. Although the number of channels between any two locations is limited to 64, depending on routing requirements, the total number of channels on the network may be significantly higher. | | | | | | | | | | | |
| EtherSound ES-Giga www.ethersound.com | Ethernet data-link layer | isoch. | coexists with Ethernet | Proprietary | star, daisy chain, ring | fault-tolerant ring | Cat5=140m, MM=600m, SM=70km | ∞ | 512 | 84–125 $\mu$s + 0.5 $\mu$s/node | 96 kHz |
| Note: EtherSound allows channels to be dropped and added at each node along the daisy-chain or ring. Although the number of channels between any two locations is limited to 512, depending on routing requirements, the total number of channels on the network may be significantly higher. | | | | | | | | | | | |
| HyperMAC | Gigabit Ethernet | isoch. | dedicated Cat5, Cat6, or fiber | 100Mbps+ Ethernet | Point-to-point | redundant link | Cat6=100m, MM=500m, SM=10km | ∞ | 384+ channels | 63 $\mu$s | 384 kHz and DSD |
| Livewire www.axiaaudio.com | Ethernet data-link layer | isoch. | coexists with Ethernet | Ethernet, HTTP, XML | Spanning tree | provided by 802.1 | Cat5=100m, MM=2km, SM=70km | 10 km | ∞ | 2.75ms + 1ms/hop | 48 kHz |
| Note: Network redundancy is provided by 802.1 Ethernet: STP, Link aggregation. | | | | | | | | | | | |
| mLAN www.yamaha.co.jp | IEEE-1394 | isoch. | coexists with IEEE-1394 | IEEE-1394, MIDI | Tree | provided by 1394b | 1394 cable (2 power, 4 signal): 4.5m | 100 m | 63 devices (800 Mbps) | 354.17 $\mu$s | 192 kHz |
| Note: Many mLAN devices have a maximum sampling rate of 96 kHz, but this is a constraint of the stream extraction chips used rather than the core mLAN technology. | | | | | | | | | | | |
| Nexus www.stagetec.com | Dedicated fiber | synch. | dedicated fiber | Proprietary | Ring | provided by FDDI | MM=2km | 10 km | 256 channels | 6 samples | 96 kHz |
| Note: Fault tolernace is provided by the counter-rotating ring structure of FDDI, which tolerates single device or link failure. | | | | | | | | | | | |
| Optocore www.optocore.com | Dedicated fiber | synch. | dedicated Cat5/fiber | proprietary | ring | redundant ring | MM=700m SM=110 km | ∞ | 512 channels at 48 kHz | 41.6 $\mu$s | 96 kHz |
| Note: These entries refer to the classic fiber-based Optocore system; no information has yet been obtained regarding the Cat5e version. Confirmation is being sought for the figure of 110 km max distance. | | | | | | | | | | | |
| Rocknet www.medianumerics.com | Ethernet Physical Layer | isoch. | dedicated Cat5/fiber | proprietary | ring | redundant ring | Cat5e=150 m, MM=2km, SM=20 km | 10 km max, 99 devices | 160 channels (48 kHz/24-bit) | 400 $\mu$s @ 48 kHz | 96 kHz |
| UMAN | IEEE 1394 and Ethernet AVB | isoch. and asynch. | coexists with Ethernet | IP-based XFN | daisy chain in ring, tree, or star (with hubs) | fault-tolerant ring, device redundancy | Cat5e=50m, Cat6=75 m, MM=1 km, SM=>2 km | ∞ | 400 channels (48 kHz/24bit) | 354 $\mu$s + 125 $\mu$s per hop | 192 kHz |
| Note: Transport is listed for media streaming and control. Ethernet is also for control. Base latency measurement is provided for up to 16 daisy-chained devices. UMAN also supports up to 25 channels of H.264 video. | | | | | | | | | | | |

[1] The reader is referred to Section 2.3 for formal definitions of the various topologies.

[2] MM = multi-mode fibre, SM = single-mode fibre

## 3.1 Compatibility

With few exceptions, the audio networking systems described in Table 1 are characterized by the transport of uncompressed audio in PCM format, which in principle could be reformatted as requested. Most standards support 24-bit PCM at the usual sample rates (44.1, 48, 88.2, 96 kHz), with several of the standards extending to 192 kHz.

In practice, there are several issues for compatibility between formats that should be addressed and solved with specific implementations, most notably:

- configuration and control protocols are incompatible between formats, and bridging among them is non-trivial; some protocols are sophisticated while others are fairly simple
- auxiliary data communication channels have varying available bitrates, ranging from less than 100 kbps to several Mbps; as an example, AES50 provides a 5 Mbps LAN control channel embedded in its transmission, while other protocols provide asynchronous serial port tunnelling
- propagation of the global sync reference is done with various techniques, which cannot be mixed easily; most solutions recover the audio clock from packet time of arrival, while AES50 uses a dedicated twisted pair for synchronization

For these reasons, audio network nodes are typically interfaced with other nodes using a single format; change of format is accomplished with back-to-back specific implementations, with AES3 being the most common interface protocol between different solutions.

The main issue is currently related to licensing. A node that is made compatible with more than one standard would need to apply to different licensing schemes. For these reasons, interoperability at the node level is currently neither implemented nor foreseen. However, single-box format conversion is available in certain cases (e.g., the Cobranet/Ethersound/Aviom conversion box by Whirlwind).

## 3.2 Implementation notes

Most audio networking formats are based on 100 Mbps or 1 Gbps full-duplex links, as the available bandwidth is considered sufficient for streaming a significant number of channels. Practical implementations of transport formats are usually realized with specialized logic.

Commercially available physical layer (PHY) devices are used for the critical interface to the physical medium. Specialized implementations on FPGAs or ASICs are used to implement the Media Access Control interface to the PHY and conversion to synchronous interfaces, in order to guarantee the necessary continuous data rate.

Some systems are also available on dedicated integrated circuits, allowing for tight integration in custom form factors. Licensing is an important aspect of audio networking formats; most are patented at least partially by the proposing companies.

Current systems include:

- module-level solutions, which generally provide audio channels in digital form and the recovered master clock reference when applicable
- intellectual property (IP) block solutions, in which a functional logic block can be purchased and integrated into an FPGA; this approach leads to tight integration but can cause high non-recurring costs, due to both licensing of the IP and integration of same into the custom project
- ASIC-level solutions, which simplify licensing management and allow integration into space-constrained designs; some solutions, such as CobraNet devices, also include internal signal processing capabilities. Only a few audio networking formats have ASIC solutions available.

In the near future, DSP-based solutions are likely to become more common, given their performance, support for peripherals, and cost-effective integrability with signal processing functions.

### 3.3 Computer interfaces

Of particular relevance for network audio hardware is the interface between the audio equipment and the computers to which they are connected. Common interfaces include Peripheral Component Interconnect (PCI), IEEE 1394 (FireWire), and USB. PCI, while slightly older and less convenient than the others, due to the need to install special cards inside the computer chassis, traditionally offered the advantage of low latency. However, more recent versions of the various interfaces (PCI Express, IEEE 1394b, and USB 2.0) offer significant improvements in both throughput and latency.

Compliance with standards remains a concern. In the case of PCI audio, no interoperability standards exist and every manufacturer is left to develop a proprietary solution, including OS drivers. USB 2.0 audio standards are gradually being adopted on certain platforms. Similarly, FireWire audio relies on standard protocols, although not all manufacturers fully support them. The added advantage of FireWire is the option for direct communication between peripheral devices, without needing the host computer.[4]

USB uses a centralized protocol, which reduces the cost of peripherals. USB moves the network processing tasks to the host computer, placing fewer demands on peripheral devices and making the device-side technology less expensive. The multi-channel USB 2.0 audio specification was released in the Fall of 2006, finally establishing an interoperability standard for device drivers and audio communications.

The table below compares the data throughput and resulting maximum channel count for different technologies. 100BaseT and Gigabit Ethernet are included here for reference

Table 2: Audio Interface Technologies

| **Format** | **Bandwidth** | **Max. Audio Channels** (@ 24 bits/sample, 96 kHz) |
|---|---:|---|
| USB 1.1 | 12 Mbps | 3 (theoretical limit) |
| USB 2.0 | 480 Mbps | 80 (per end point) |
| FireWire 400 | 400 Mbps | 85 (theoretical limit) |
| FireWire 800 | 800 Mbps | 170 (theoretical limit) |
| Legacy PCI | 1 Gbps | 100 (typical) |
| PCI Express (PCI-E) x1 | 2.5 Gbps | 800 (theoretical limit) |
| 100BaseT Ethernet | 100 Mbps | 32 (typical) |
| Gigabit Ethernet | 1 Gbps | 320 (typical) |

## 4 Case studies and best practices

### 4.1 Commercial Audio

Commercial audio systems span a large range of applications in terms of scale and complexity. On the low end are sports bars and music systems in retail establishments. More complex examples include the installed sound reinforcement systems in stadiums and arenas, various forms of audio distribution in theme parks, background music and/or paging systems for hospitals, cruise ships, convention centers and transportation facilities. In some cases, these systems may be linked together over great distances, for example, an interlinked network of train stations, or sport campuses such as those built for the Olympics. For commercial applications involving large distance or channel counts, audio networking is a well-suited technology, offering reliability and the ability to diagnose problems remotely. Further discussion of networked audio in commercial applications can be found in other AES publications [10][11].

### 4.2 Recording studios

The audio transmission needs of a recording studio are different from those of theme parks or stadium installations. While in the latter case distributed audio is sent long distance, studio operations are usually confined to smaller

[4]However, this functionality has not yet been fully utilized for audio devices. This topic is being addressed by the AES Standards Subcommittee on Digital Audio (Working Group on Digital Input-Output Interfaces SC-02-12).

spaces, with limited needs for live multi-zone distribution. Recording studios range in size from a single room in the case of a home studio, to larger multi-room facilities dedicated to recording orchestral works or television and film soundtrack production. Postproduction studios can be as simple as a cubicle with a computer and all-in-one portable multi-track recorders occupy the low end of this spectrum.

A recording studio involves shorter cable runs (under 100 m) and point-to-point network communication topologies; communication is typically bidirectional. Microphone signals are carried from each instrument in the studio to the mixing console and/or audio recorder located in the control room. Individual headphone mixes for each musician are sent from the control room back into the studio for monitoring purposes. Channel count varies by job complexity. A setup feeding 32 channels into the control room and returning a dozen stereo mixes (24 mono channels) covers a majority of cases in professional recording, aside from large symphonic works, which may require more than 100 channels. Smaller project and home studios based around a computer, a control surface and an audio interface, have also gained prominence for music recording. In these setups, recording one instrument at a time and overdubbing are typical techniques. This reduces the channel count to 4x4 or less.

When recording to a digital audio workstation (DAW), latency may be introduced due to computer processing. Since this can affect musicians' ability to hear themselves during the recording process, zero-latency monitoring, using a local analog feed, is often employed. This is not subject to digital delays and therefore offers instantaneous feedback.

## 4.3 Archives

The most apparent benefit of a shared network archive is the efficiencies provided to workflow and access. The archivist no longer has to deal with cataloging and moving physical media. With increased ease, audio data can be located and accessed quickly, therefore allowing for increased productivity and workflow.

Data networks used for storage and access to audio archives must be centralized, scalable, fault-tolerant, and easily accessible to workstations throughout the network. Centralized storage allows ease of administration, providing one point of control. Significant price decreases in networking technology and storage have allowed for affordable network audio storage for all sized archives, particularly in the NAS (Network Attached Storage), SAN (Storage Area Network) or DMSS (Digital Mass Storage System) environments.

SAN or DMSS allow for centralized storage and commonly employ a combination of redundant disk arrays, load balancing, fault-tolerance, integrated network security permissions, and ideally a tape library backup system. From the perspective of the end-user, DMSS functions as if it were a storage medium local to the workstation. The most common host-bus connection uses fibre channel for increased throughput and performance. Workstations access the DMSS via gigabit Ethernet or gigabit fiber switches.

There are several critical decisions that determine the network design and operation of the audio archive, subject to cost constraints. First, disk space must be available to house the current collection of digital audio data, the amount of analog audio to be digitized, and for long-term growth as more content is acquired. Storage may also be desired for high-resolution preservation copies along with lower-resolution preview or access copies. Second, the number of concurrent users needing access to the digital audio content and whether such access must be real-time should also be considered. Bandwidth requirements may be determined by the need for streaming of audio content for previewing or research purposes, access to content for editing purposes and the creation of derivative files. Calculations based on resolution can then be made to estimate both the discrete and aggregate bandwidth needed for the shared storage solution.

Network archives can readily be made available to the public. A common implementation of this can be seen in kiosks found in libraries, museums, as well as private industry. Kiosks can be integrated so that content is accessible across the network from the centralized mass storage system. A second method for public access is in an Intranet or Web environment where, typically encoded, content is available for streaming or uncompressed content is available for download. Once networked audio content is available locally on a LAN, an archivist or systems administrator can easily use a combination of FTP, IIS, and/or port forwarding to make that content available via the Internet. This access method is critical to researchers, students and scholars. The end-user in this case can be located outside of the DMSS, yet still gain access to the content from any location.

With network audio archives, physical security of media is not a primary concern, given that the physical infrastructure is only accessed by a Systems Administrator, but there are a new host of security issues within this environment. These include unauthorized access and distribution, file integrity, and malicious code. The network architect must provide security permissions that will allow access to network audio objects strictly on a per-user or group basis. Furthermore a security system must be employed to prevent users from copying and distributing content in an unauthorized manner.

The first line of defense against unauthorized access to an audio archive is at the entry point of a network, the router or gateway. The use of a firewall in front of the router or gateway provides security such as stateful packet inspection, tarpits, blacklisting, port scan detection, MAC address filtering, SYN flood protection, network address translation (NAT), and typically a secure channel via VPN for remote users.

Critical to maintaining the integrity of the audio data are antivirus and spyware solutions, available as both hardware and software forms. These protect against Trojan horses, worms, spyware, malware, phishing and other malicious code. Both hardware and software solutions are available. Similarly, for archive material distributed to the public through the Internet, there is a need for technology to provide security through such means as watermarking, access-control, and copy-protection to prevent unauthorized use of the content.

## 4.4 Live performance

Networked audio systems have found important application in the context of live performance. In the case of mobile sound reinforcement systems, simplicity of setup is critical, as the process must be repeated frequently. Networked audio technologies considerably relieve the problems of wiring, allowing the transport of numerous audio channels and supplementary control data over optical fibers, coaxial or Cat5 cables. These are significantly thinner and lighter than the traditional multi-conductor audio snakes, facilitating both their transport and handling.

Moreover, networked audio systems offer channel routing possibilities and remote control of the equipment that enable fast reconfiguration without any physical user intervention. These technologies enable the creation of an audio bus between key points of the installation, allowing exchange of signals between the monitor mixer, the front of house and a production control room, which offers a substantial improvement over traditional analog systems.

Latency, also described in Section 4.7, is a critical problem in live events. The generally agreed upon latency limit for the entire system, between source and restored signal, is approximately 15-20 ms [16]. Assuming the incompressible latency resulting from A/D converters, D/A converters, console and front of house DSP can be as high as 20 ms, this leaves very little (if any) tolerance for additional delays in network audio transport. To guarantee optimal alignment of the loudspeaker systems and avoid phasing, this latency must be constant in time at each point of the network.

Optimum reliability of the sound reinforcement system is a constant concern. Multiple point-to-point connections of a traditional analog system may provide a greater level of overall reliability, even though the probability of a single failure is higher. Indeed, a connection breakdown or the malfunction of a network audio device may cause a system failure on a larger scale. This explains the variety of solutions proposed to ensure redundancy in data transmission. These include individual device redundancy for each connection, as is the case for star topologies, and global redundancy on the scale of the entire system, as in ring topologies. It is worth noting that the manufacturers of network infrastructure devices such as cables, optical fibres and connectors have further developed and significantly enhanced their products in order to meet the demands for robustness of mobile systems and applications. Examples include the EtherCon connector, armored fiber and Cat5 cables.

## 4.5 Radio

Networked audio over IP has become common in radio operations for streaming of radio programs from remote sites or local offices into main studio centers. The infrastructure can be well managed private networks with controlled quality of service. The Internet is increasingly also used for various cases of radio contributions, especially over longer distances. Radio correspondents may choose to use either ISDN or the Internet via ADSL to deliver their reports.

As of this writing, a number of manufacturers provide equipment for such applications. These include AEQ (Spain), AETA (France), APT (Ireland), AVT (Germany), Digigram (France), LUCI (NL), Mandozzi (CH), Mayah (Germany), Musicam (US), Orban (US), Prodys (Spain), Telos (US), Tieline (Australia), and Youcom (NL).

Until recently, end units from one manufacturer were not compatible with those from other companies. Based on an initiative of German vendors and broadcasters, in early 2006, the European Broadcasting Union (EBU) started a project group, N/ACIP, Audio Contribution over IP.[5] to suggest a method to create interoperability for audio over IP, leading to the endorsement of a standard [6].

[5]Further information can be obtained from Mathias Coinchon, e-mail: coinchon@ebu.ch

The requirements for interoperability are based on the use of RTP over UDP for the audio session and Session Initiation Protocol (SIP) for signalling. IETF RFC documents for commonly used audio formats in radio contribution, such as G.722 speech encoding, MPEG Layer 2, and PCM, define the packet payload audio structure. An additional tutorial document covers basic network technologies and suitable protocols for streaming audio over IP [7].

## 4.6 Distance education

The use of live networked audio systems is enabling communication across great distances and bringing artists and audiences closer together. Musicians spend much of their lives practicing in isolation as they perfect their art. With the advent of high-speed networking, they are now able to interact with each other in practice rooms, classrooms, and on the stage, while physically located across the country or even on different continents. This enables the exploration of pedagogy, styles and techniques, not previously available. Teachers and students have moved beyond the traditional exchange of raw audio files of musical performance for comment and critique to live interaction over networks, and with it, witnessed the potential for greater pedagogical accomplishment in a specified time period. This has spawned the use of network technology, often involving videoconferencing, in several music programs to involve living composers in rehearsals and concerts, provide conducting seminars, and interactive coaching sessions [15].

These distance music education applications raise various technical considerations regarding protocol choice (e.g., encoding format, bandwidth, inherent latency), production elements (microphones, lighting, far-site sound reproduction), and possibly the most important, how to make the technology as transparent as possible for the participants. The choices are often dictated by the respective objectives. For example, fidelity of the audio signal and the perceived latency by the participants become paramount when dealing with collegiate and post-collegiate musicians. Protocols that are capable of transmitting (perceptually) uncompressed and minimally delayed audiovisual signals are often favored, although this choice comes at the cost of high bandwidth requirements. Moreover, many systems lack the support for echo-cancellation, thus requiring additional production elements to improve the quality of interaction. Proper microphones, cameras, lights, sound reproduction elements, and a working understanding of live sound reinforcement principles, are essential. Network infrastructure can prove troublesome as well for uncompressed data transmission, as many networks have security measures in place, in particular in the "last mile", that block ports, throttle, or filter the network traffic associated with these protocols.

In projects with the pre-collegiate community, institutions often lack the technical personnel, the financial resources, or the desire to learn how to use the more advanced software protocols for higher fidelity communication. As such, existing standards such as H.323 [23] and its associated codecs are generally used for musical interactions. Fortunately, these have improved in recent years, now incorporating higher bit-rate audio formats such as Polycom's SirenStereo. These codecs offer superior fidelity to the previous G.722 [22] standard, and tend to suffice for discussion of musical elements such as pitch, intonation, and tempo that are typically discussed in these interactions. Additional benefits of H.323 are its widespread adoption, standardized format, and support for remote manipulation. However, cost vs. quality vs. ease-of-use remain trade-offs. Software-based systems are free or nearly free, but rarely deliver comparable quality to dedicated hardware, and are often manipulated by complex GUIs, which may be imposing on the average user. In either case, the latency inherent in such protocols generally precludes distributed performance, discussed in the next section. Finally, while some codecs support multiple audio channels for surround sound reproduction, it is often hard to find a partnering institution capable of reproducing the surround signal faithfully.

As network technology continues to improve, the richness and fidelity of the distance education experience will make these interactions even more valuable. The future holds the potential for musicians to collaborate with a virtual acoustical environment that provides the realism of an in-person experience and, as discussed in the following section, possibly even to perform before "remote" live audiences around the world.

## 4.7 Networked musical performance

Artists have often pushed the envelope of possibility of technologies for making music. The use of networks in musical activity is no exception. Telecommunications infrastructures have been used by artists in creative musical works of great diversity and continues to this day. The use of networks in creative musical works include:

- the conduit through which a performance takes place
- a virtual space that is an assembly point for participants from remote sites

- a musical instrument of sorts with its own sonic characteristics
- the musical object, or canvas onto which sound is projected

The move to wireless and mobile systems has not been lost on artists, with applications including a participative soundwork for ringtones [18] and musical composition using GPS tracking with bidirectional streaming over UMTS mobile phone networks [21].

The vision of networked musical performance (NMP) has been a holy grail of telecommunications and videoconferencing technologies with respect to its demands on both signal quality and minimized latency. Actual distributed performance, making use of network audio systems, has a history going back at least to 1966, with the work of Max Neuhaus in his pieces for public telephone networks. Experiments continued over radio, telephone, and satellite networks, progressing in the 1990s to ISDN and ATM-based systems, with some efforts incorporating video as well.[6] Although early demonstrations were limited to one-way events, effective two-way interaction was seen in 1996 with the *Distributed Rehearsal Studio*, which succeeded in achieving a one-way delay of approximately 85 ms over an ATM circuit [17].

The fundamental challenge preventing effective distributed musical performance is typically one of latency, as musical synchronization becomes increasingly difficult to attain under increased delay. In co-present situations, visual cues such as the movement of a drummer's hand or a conductor's wand allow the musicians to preserve timing. However, in a computer network situation, visual and auditory cues may be equally delayed. Although the International Telecommunication Union defines 150 ms as the threshold for acceptable quality telephony, musicians perceive the effects at much lower values, often less than 25 ms, depending on the style of music they are playing. Experiments conducted at CCRMA and USC [2][3] have investigated these limits. Chew found that latency tolerance in piano duet performance varied depending on the tempo and types of onset synchronization required; for a particularly fast movement, this figure was as low as 10 ms.

Research systems such as those from Stanford[7] and McGill University[8] have demonstrated high quality, low-latency audio distribution over Internet. Until recently, commercially available systems, as described in Section 3 were limited to LAN distribution and thus unsuitable for the demands of low-latency spatially distributed performance over larger distances. In the last few years, several commercial products with network layer addressing capability (see Section 2.1) have become available. The latency achievable by such systems is, of course, affected both by the compression algorithm employed (if any) and the network distance the signal must travel.

Current latencies achievable over the research Internet (e.g., Internet2 Abiline, National LambdaRail, and Canarie CA*net4) impose a one-way delay of approximately 50 ms[9] across North America, which is borderline tolerable for most small group performances. Awareness of these limitations has motivated investigation of the alternative approach of *adding* delay to networked communication in order to ensure that the musicians are synchronized on the beats of a common tempo.

As an often important addition to network audio, simultaneous video transmission may introduce additional latency, either due to characteristics of the acquisition and display hardware, or the use of signal compression to reduce the bandwidth requirements. Under these conditions, improved interaction may be obtained by decoupling the audio and video channels, transmitting the former at minimal latency without concern for audiovisual synchronization [4]. The decoupling of visual cues from audio, however, has musical consequences. When given the choice in the tradeoff between image quality and smoothness, musicians preferred a pixelated image with smooth motion over a high resolution image at lower frame rate decoupled from the audio channel [20].

In the context of NMP, sources of delay may include audio coding, acoustic propagation in air, packetization, network transport, and jitter buffering. Traditionally, in particular when bandwidth is a limiting factor, audio coding has been an obvious target for improvement. An analysis of algorithmic delay in popular consumer audio codecs, such as MP3 and AAC [9] found that as of 2004, there existed no MPEG-standardized audio coding system that introduced less than 20 ms delay. Recent developments, namely AAC Enhanced Low Delay (ELD), further reduce the inherent encoding delay to 5 ms, although an additional 10 ms is required for packetization. Applying a completely different coding paradigm compared to most other audio codecs, Fraunhofer's Ultra Low Delay Audio codec (ULD)

[6] See `http://www.cim.mcgill.ca/sre/projects/rtnm/history.html` for a history of milestones in spatially distributed performance.

[7] `http://ccrma.stanford.edu/groups/soundwire/software/`

[8] `http://ultravideo.mcgill.edu`

[9] Due both to switching and routing delays as well as the signal propagation time, whose speed through copper or fiber is approximately $2/3c$.

typically reaches an inherent encoding delay of 2.6 ms with only another 2.6 ms required for audio packetization at 48 kHz sampling rate. At this setting, quality and data rates are comparable to those of MP3.

Compression is often necessary to satisfy the bandwidth constraints of typical Internet distribution architectures. In general, the choice of whether to employ audio compression, and if so, what kind of compression, is usually a trade-off between quality, data rate, latency, error robustness and complexity. Additionally, there is the choice between standardized algorithms, which provide a long-term guarantee to be decodable, or a proprietary scheme, which in many cases can be hand-tailored to a specific application.

## 4.8 Consumer

At present, the most popular method for audio distribution in the home involves expensive, proprietary, analog or digital systems. These systems use custom wiring to carry the audio and control signals. However, two technologies are causing a revolutionary change. The first is the continuing growth and popularity of broadband internet connections into the home, analogous to broadcast television, except that the information only enters the home at one point, and is typically connected to a single PC.[10] This leads to the second technology, that of inexpensive computer networking, either wired or wireless, which allows for distribution of the broadband connection to many PCs and network appliances in the home.

This same home network can now also be used to distribute audio and video in an inexpensive, non-proprietary manner. Underlying technology standards, such as Universal Plug and Play (UPnP) are being used for connection management while HTTP GET functions can be used to transfer the data itself. Interoperability standards are being created to make sure that different manufacturers' devices will connect and work together.

Another related trend is the convergence of PC and Consumer Electronics (CE) technologies. PC makers, driven by Intel and Microsoft, and adding more audio and video capabilty to the PC platform in an effort to win the *living room electronics* opportunity. Meanwhile, CE manufacturers are adding networking capability to their products. One can now buy a home theatre amplifier that performs the usual audio/video switching, decoding and amplifier functions, and can also stream and play internet radio stations, as well as audio stored on any PC in the home.

# Acknowledgement

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

# References

[1] Bluetooth Audio Video Working Group. *Bluetooth Specification: Advanced Audio Distribution Profile*. Bluetooth SIG Inc. 2002, 2002.

[2] C. Chafe, M. Gurevich, G. Leslie, and S. Tyan. Effect of time delay on ensemble accuracy. In *International Symposium on Musical Acoustics*, March–April 2004.

[3] E. Chew, A. A. Sawchuk, R. Zimmerman, V. Stoyanova, I. Tosheff, C. Kyriakakis, C. Papadopoulos, A. R. J. François, and A. Volk. Distributed immersive performance. In *Annual National Association of the Schools of Music (NASM)*, San Diego, November 2004.

[4] J. R. Cooperstock, J. Roston, and W. Woszczyk. Broadband networked audio: Entering the era of multisensory data distribution. In *18th International Congress on Acoustics*, April 2004.

[5] European Broadcasting Union. Audio contribution over IP @ IBC 2006. `http://www.ebu.ch/en/technical/ibc2006/ibc_2006_acip.php`.

[6] European Broadcasting Union. EBU - TECH 3326 audio contribution over IP: Requirements for interoperability. `http://www.ebu-acip.org/Main_Page`.

[10] See `http://news.bbc.co.uk/1/hi/technology/4736526.stm`, `http://www.websiteoptimization.com/bw/0607`, and `http://www.platinax.co.uk/news/29-03-2007/uk-leads-western-europe-in-broadband-penetration/`

[7] European Broadcasting Union. EBU - TECH 3329 a tutorial on audio contribution over IP document 2006. http://www.ebu-acip.org/Main_Page.

[8] A. Floros, T. Karoubalis, and S. Koutroubinas. Bringing quality in the 802.11 wireless arena. In *Broadband Wireless and WiMax*. International Engineering Consortium, January 2005.

[9] M. Gayer, M. Lutzky, G. Schuller, , U. Krämer, and S. Wabnik. A guideline to audio codec delay. In *116th AES Convention, AES preprint 6062*, pages 8–11. Audio Engineering Society, May 2004.

[10] K. Gross. Audio networking: Applications and requirements. *Journal of the Audio Engineering Society*, 54(1/2):62–66, February 2006.

[11] K. Gross and D. J. Britton. Deploying real-time ethernet networks. In *UK 15th Conference on Moving Audio: Pro-Audio Networking and Transfer*. Audio Engineering Society, May 2000.

[12] Harmony Central. BSS Audio's Soundweb offers flexible system for distributing and routing sound, September 1997. http://namm.harmony-central.com/Newp/103AES/BSS_Audio/Soundweb-9088.html. Accessed: 2009-02-22. (Archived by WebCite®at http://www.webcitation.org/5emZkXC1U).

[13] IEEE Standards for Information Technology. *ISO/IEC 8802-11:1999 Telecommunications and Information Exchange between Systems – Local and Metropolitan Area Network - Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Higher-speed Physical Layer (PHY) extension in the 2.4GHz band*. IEEE Standards Office, New York, NY, USA, September 1999.

[14] Institute of Electrical and Electronics Engineers. *IEEE Standard 802.16-2004 for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems*. IEEE Standards Office, New York, NY, USA, 2004.

[15] Internet2. Internet2 member community education initiatives. http://www.internet2.edu/arts/member-education.html.

[16] A. Keltz. Opening Pandora's Box? the "L" word - latency and digital audio systems. *Live Sound International*, 14(8), August 2005. http://www.whirlwindusa.com/wwlatart.html.

[17] D. Konstantas, Y. Orlarey, O. Carbonel, and S. Gibbs. The distributed musical rehearsal environment. *IEEE MultiMedia*, 6(3):54–64, 1999.

[18] G. Levin. A personal chronology of audiovisual systems research. In *New Interfaces for Musical Expression (NIME)*, pages 2–3, Singapore, Singapore, 2004. National University of Singapore.

[19] R. Stamberg. Back to basics. *AV Video Multimedia Producer*, July 2001. URL: http://www.highbeam.com/doc/1G1-77874253.html.

[20] A. Tanaka. von telepräsenz su co-erfahrung: ein jahrzehnt netzwerkmusik. (From telepresence to co-experience: A decade of network music.). *Neue Zeitschrift für Musik*, 5:27–28, September/October 2004.

[21] A. Tanaka, P. Gemeinboeck, and A. Momeni. Net_dérive, a participative artwork for mobile media. In *Mobile Media*, 2007.

[22] Telecommunication Standardization Sector of ITU. *Recommendation G.722: 7 kHz audio-coding within 64 kbit/s*. International Telecommunication Union, November 1988. Available from http://www.itu.int/rec/T-REC-G.722.

[23] Telecommunication Standardization Sector of ITU. *Recommendation H.323: Packet-based Multimedia Communications Systems*. International Telecommunication Union, June 2006. Available from http://www.itu.int/rec/T-REC-H.323.

[24] M. Ward. Broadband kills off consumer ISDN, BBC News. http://news.bbc.co.uk/1/hi/technology/6519681.stm.

# Contact information

The authors of this White Paper can be reached through the Audio Engineering Society's Technical Committee on Network Audio Systems. See `http://www.aes.org/technical/nas` for up to date contact information.