Appendix
Audio Networking Standards

Since NMP represents a niche field in the academic literature, it often takes advantage of existing technologies for signal transmission. Currently digital audio transmission is covered by a large number of protocols. Of particular interest for NMP applications are those standards that fall under the umbrella terms of audio over IP (AoIP) and audio over Ethernet (AoE). A number of very mature and developed protocols belong to this family, which are used in broadcasting studios, recording studios and such, for the delivery of musical content at very low latency and high reliability.

The efforts in the direction of digital audio delivery started at least 30 years ago with the ratification of the AES3 standard, also known as AES/EBU as it was developed by the Audio Engineering Society with the contribution of the European Broadcasting Union [1]. The AES3 protocol defines the synchronous transmission of a stereo PCM signal over several transmission media, and is incorporated into IEC 60958 standard. In principle, it defines a time-slotted technique for sending stereo audio data in 24-bit or 20-bit sample words. The data is uncompressed and the time slotting depends on the audio sampling rate. Allowed sampling rates are 32, 44.1, and 48 kHz. A simple representation of the time slotting employed by AES3 is provided in Fig. A.1. Commercial variants are S/PDIF (which only transmits stereo data) and ADAT optical interface. ADAT optical interface or ADAT Lightpipe introduces a larger bandwidth, allowing up to 8-channel 48 kHz 24-bit audio. Similarly, MADI introduces multichannel transmission following the AES10 standard [2], and employs coaxial or fiber optic cables. It employs asynchronous time slotting as in the previous standards, so that time slots are not synchronized to the audio sampling rate. The latest amendment to the standard allows up to 64 channels at 48 kHz. Synchronization of the frames is not done by the frame overhead, but by symbols external to the frames. All the aforementioned protocols are widely used in practice for digital transport of audio between specialized hardware and computer audio peripherals; however, they are not suitable to switched packet networks. AES47 and AES51 [3, 4], for instance, respectively, define how to pack AES3 audio data on generic asynchronous networks and specifically on Ethernet hardware. Such communication technologies are leveraged daily in the music broadcasting and distribution fields [5]:

L. Gabrielli and S. Squartini, *Wireless Networked Music Performance*,
SpringerBriefs in Electrical and Computer Engineering,
DOI 10.1007/978-981-10-0335-6

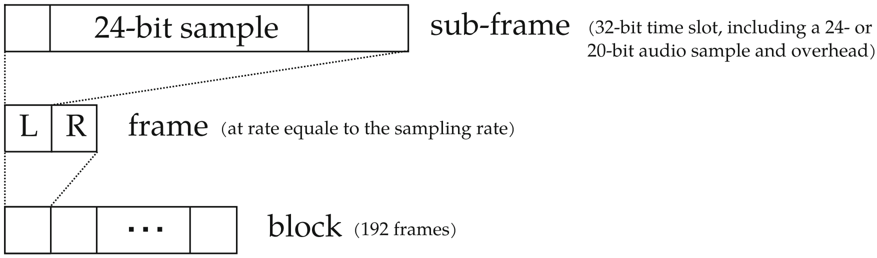

Fig. A.1 A simplified representation of the AES3 or AES/EBU data format and time multiplexing

audio-over-Ethernet technologies are widely used for the local area [6] (broadcasting studios, conservatories, universities, large rehearsal studios, arenas and sport venues, train and metropolitan area alert and voice diffusion, etc.). Commercial products from several manufacturers employ open or proprietary protocols that encapsulate audio over standard Ethernet packets, such as CobraNet and EtherSound. Other protocols encapsulate data in IP packets and are independent of the underlying communication layers. Among these IEEE 802.1BA (also known as AVB), RAVENNA, Q-LAN, LiveWire, and Dante can be counted. The former is taken as an example to enumerate several of the typical features of an audio-over-IP technology. The Dante protocol is based on UDP-encapsulated audio transmission over existing networking infrastructures such as Fast Ethernet, Gigabit Ethernet, or fiber optic IEEE 802.3 links. It also shares a master clock, synchronized between all the Dante-enabled machines using precision time protocol (PTP, standard IEEE 1588), enabling misalignment between the networked audio devices of typical 1μs. Devices are automatically discovered and can be routed or controlled through the network, by employing a software tool provided by the manufacturer. Audio, timing, and connection traffic can coexist with an existing networking infrastructure employed for other non-audio related traffic (as it would be in a school or conservatory local area network), provided that the infrastructure allows for traffic prioritization through DiffServ code points (DSCP), reserving, thus the maximum priority to timing messages (for clock synchronization through PTP), and a medium priority to audio traffic. One advantage of employing general-purpose networking technologies is the ability of employing a personal computer as a Dante device. The protocol does not support IP networking over the Internet or over wireless networks for reasons of reliability and latency; however, it enables fiber optic links even at very distant geographical locations.

IEEE 802.1BA or AVB was originally meant for audio and video, and developed by the Audio Video Bridging Task Group of the IEEE. Later in 2012, it was renamed to Time-Sensitive Networking task group, and the scope of the standard was broadened, thus, potentially increasing the number of adopters of the protocol. AVB provides a synchronization mechanism coherent to the stringent specifications of low latency streams, a bandwidth reservation protocol, delay reduction mechanisms on IEEE 802 networks, and interoperability profiles. Although targeting IEEE 802 networks, the protocol only provides synchronization support for wireless IEEE 802.11 networks, based on OSI layer 2 primitives providing time stamping.

Given the relatively large number of proposed standards and commercial technologies, an interoperability standard for audio networking was produced recently known as AES67 [7]. This standard proposes a set of minimum features all the producer must adhere to provide interoperability between the aforementioned technologies. At the moment of writing many suppliers of audio-over-IP systems are providing an AES67 compatibility layer for interoperability with other suppliers' equipment. AVB also provides interoperability features.

To the best of our knowledge, there is no report of networked music performance employing the aforementioned audio-over-IP protocols. One reason is their restriction to local area networking. The issues introduced by the Internet or a wireless medium require specific solutions that need to be addressed properly and all the current solutions described in literature were implemented by researchers by writing software from scratch.

References

1. Specification of the Digital Audio Interface (The AES/EBU interface) Tech 3250-E, 3rd edn. European Broadcasting Union, 2004
2. AES10-2008: Recommended Practice for Digital Audio Engineering-Serial Multichannel Audio Digital Interface (MADI). Audio Engineering Society, 2008
3. IEC62365: Digital audio—Digital input-output interfacing—Transmission of digital audio over asynchronous transfer mode (ATM) networks, 2009
4. AES51: AES standard for digital audio—Digital input-output interfacing—Transmission of ATM cells over Ethernet physical layer. Audio Engineering Society, 2006
5. Rumsey F (2011) Audio in the age of digital networks. J Audio Eng Soc 59(4):244–253
6. Shuttleworth T (ed) (2011) AES Technical Committee on Network Audio Systems. Emerging Technology Trends Report
7. AES standard for audio applications of networks: high-performance streaming audio-over-IP interoperability. Audio Engineering Society, 60 East 42nd Street, New York, NY, USA, 2013