PROCEEDINGS OF SCIENCE

MUSIC and AUDIO – Oh how they can stress your network!

Dr R P Fletcher[1]
University of York
Heslington, York, YO10 5DD, UK
E-mail: r.fletcher@york.ac.uk

Nearly ten years ago a paper written by the Audio Engineering Society (AES)[1] made a number of interesting statements:

1. The current Internet is inadequate for transmitting music and professional audio.
2. Performance and collaboration across a distance stress beyond acceptable bounds the quality of service
3. Audio and music provide test cases in which the bounds of the network are quickly reached and through which the defects in a network are readily perceived.

Given these key points, where are we now? Have we started to solve any of the problems from the musician's point of view? What is it that musician would like to do that can cause the network so many problems? To understand this we need to appreciate that a trained musician's ears are extremely sensitive to very subtle shifts in temporal materials and localisation information. A shift of a few milliseconds can cause difficulties. So, can modern networks provide the temporal accuracy demanded at this level?

The sample and bit rates needed to represent music in the digital domain is still contentious, but a general consensus in the professional world is for 96 KHz and IEEE 64-bit floating point. If this was to be run between two points on the network across 24 channels in near real time to allow for collaborative composition/production/performance, with QOS settings to allow as near to zero latency and jitter, it can be seen that the network indeed has to perform very well.

Lighting the Blue Touchpaper for UK e-Science - Closing Conference of ESLEA Project
The George Hotel, Edinburgh, UK
26-28 March, 200

[1] Speaker

PoS(ESLEA)028

1. Introduction

The Audio Engineering Society (AES) published their white paper (AESWP-1001) in 1998, "Networking Audio and Music using Internet2 and Next-generation Internet capabilities"[1]. In this they made some key observations regarding the current state of audio and music over conventional networks:

1. The current Internet is inadequate for transmitting music and professional audio.
2. Performance and collaboration across a distance stress beyond acceptable bounds the quality of service
3. Audio and music provide test cases in which the bounds of the network are quickly reached and through which the defects in a network are readily perceived.

Additionally they observed that the designers of Internet2 foresaw its use for medical researchers, physical scientists and the leading-edge research community. Audio and music were relegated to a "background function". The AES called this "a major, if not disturbing judgement".

2. Discussion

It is important to understand the importance of the role of music in society. All known historical societies have engaged in some form of musical activity. It provides identity, intellectual stimulation and can evoke powerful memories. It is fundamental to many rituals, events and communication. It has also been shown to be a powerful tool for learning and growth. We ignore these facts at our peril.

Trained professional musicians are very sensitive to small shifts of temporal materials in the order of milliseconds. For example, a conductor can isolate minor tuning problems in an orchestral section, and usually can identify the actual musician as well. This is no trick, and is done by localising the position of competing signals reaching his ears using the temporal differences. This fact lends credence to the need for more and better surround sound systems for electronic reproduction or performance. If audio and music is to be transmitted across the new networks, then they need to provide this functionality and level of accuracy.

The Internet2 Quality of Service (QoS) Working Group published a survey of the "Network QoS Needs of Advanced Internet Applications" in 2002[2]. This paper showed the need to facilitate new frontier applications, to explore complex research problems and to enable seamless collaboration and experimentation on a large scale. The notion of a virtual research space and shared virtual reality was also noted. Additionally, real time access was shown to be a major requirement as was new levels of interactivity with multisensory cues.

From the above it does not require a huge leap to see how music and audio now fit into these "advance internet applications". Indeed, they consume most of the areas where higher

levels of QoS are required. Interestingly this paper did find that audio and video were major requirements. They also (for once) differentiated between high quality audio and professional quality audio. The needs of these two categories necessarily overlap, but the additional constraints placed on the network for the latter place huge demands on the networks.

It is important to differentiate between passive and active audio streams. A passive, delivery only system can use compression, jitter buffers and can adapt to bit rate changes. An active stream will most likely have no compression, or at least lossless compression, will work in real time and for some streams they will be two way, and hence round trip times will need to be factored into the equation. There is a plethora of delivery formats, e.g. mp3, aac, real, wav, wma to name a few. In the main we can deliver stereo as a file (listen later) or as a stream (listen now), but what of the other formats, e.g. 5.1, 7.1, or the many ambisonic types?

The "quality" of any audio is always open to debate, and no more so than that delivered over the net as "near CD quality" at 128 kb/s mp3! But on earphones can you really tell the difference and more importantly, do you care? For "better quality", one can always up the encoding bit rate. We will also have surround sound delivery in the near future thanks to the new mpeg surround sound standard published on 12 February 2007[3]. Work is also in progress to encode some ambisonic types into vorbis .ogg streams.

Another debate concerns compression (or not). There are lossless and lossy compression types. Examples of lossless formats are FLAC, Apple lossless, Dolby TrueHD, Monkey's Audio, TTA, wavpak, WMA lossless, and examples of lossy formats are mp3, adpcm, ATRAC, Dolby Digital, Musepack, TwinVQ, Vorbis ogg and WMA. In the lossless world, FLAC and wavpak are very popular and in the lossy world, mp3, ATRAC, vorbis ogg and WMA are popular. Of course, the Dolby standards are ubiquitous as well.

The final debates are about how many bits are required and at what sample rate. Many will argue that 16 bits is enough and can encode all the frequencies we can hear. But, high frequencies (even those we cannot hear) will colour lower frequencies through interference. Whether we can all hear these subtle changes is debatable. However, it is common to use a minimum of 24 bits when recording to allow for headroom and 32 bits is commonly used at the mixing stage. However, the newer range of hardware mixing consoles and software mixing programs have moved to 64-bit pathways. In fact, 64-bit IEEE floating point format is often advocated. There are endless arguments about significant bits in the various representations of floating point numerical data in computers. None more so than the effects of dithering when moving from a full 64-bit mix down to a 16-bit mix ready for a CD!

And, what of sample rates? Again, there those who would contend that 44.1 KHz is acceptable for all, i.e. at 16-bits, this is what is on a CD. However, 48 KHz is used by DAT and audio on DVD-Video (and let us not forget that many consumer grade soundcards run natively at higher rates and down sample from 48 KHz to 44.1 KHz, often with quite poor results!). 96 KHz is used by recording engineers and is the rate on DVD-A, from stereo to 5.1 surround formats. Higher rates can be found on DVD-A, 192Khz for mono and stereo, and interestingly, 88.2Khz and 176.4 KHz is also seen (heard?) on DVD-A.

Musicians have been collaborating over the net for many years exchanging files, and in some cases actively collaborating with "jamming" systems, e.g. RocketNetworks using midi data streams and recently with NINJAM using audio streams. Lossy compression schemes

should not be used due to inherent signal degradation when going through many encode/decode cycles. Lossless compression can be used but the codecs will introduce more latency into the system. Smart systems which detect silence and only send metronome signals for synchronisation are beginning to be used.

From the network point of view the engineers need to look at the worst case data rates, if the network can cope with these, then everyone will be satisfied. Therefore, 192 KHz at 64-bits is the highest data format likely to be used at present – this is 11.7 Mb/s in uncompressed format per channel. Therefore, if we scale up to a basic 24 channel mix, then the data rate works out at 281.25 Mb/s. Thus, for composers to work together at remote sites in near real time we would need the audio to flow between the two sites, most likely at half the data rate (i.e. 96 KHz). The composers would need to trade off compression against extra latency. In the end, the latency due to the laws of physics has to be dealt with in some form, and having a predictable latency with close to zero jitter is the goal.

The new photonic networks open up many possibilities for such collaborative work. 1 Gb/s connections are commonplace. In the UK, the UKLight network (part of SUPERJanet-5 infrastructure) can be used to interconnect sites at rates up to 10 Gb/s. In turn lightpaths can be provisioned to sites in Europe via GEANT2 and to the USA via STARLight and other country-wide infrastructures (e.g. National Lambda Rail) and to Canada via CANARIE.

To date, and unsurprisingly, most of the use of such connectivity has been for "big science" to have access to very large datasets or high performance computing, aka Grid computing. However, there have been a number of successes in the Arts and Humanities fields with musicians taking master classes remotely, using HDTV video streams and full surround sound audio. In a similar vein, remote collaborations have taken place with jazz ensembles and increasingly, cultural exchanges via HDTV and audio across multiple sites across the globe have been happening.

The next "big thing" to encompass all of the above is the push towards digital cinema, and the cinegrid project[4] is just one of these. The video is at least 4096 x 2048, and may use up to 24 channels of surround sound. The data rates for this format are not inconsiderable, as are the constraints put on the QoS of the networks in use. Add to this the requirement for remote collaboration when creating content for this next generation of cinema experience, we have applications in the audio (and video) world which will really stress the best networks.

3. Conclusion

From the above discussion it should be clear that audio data can indeed make considerable demands on the networks on which it will be running. It should be noted that these demands become more critical as the applications in use move more towards the real time environment. The viability of collaborative composition and performance environments and their adoption into mainstream use in the Performing Arts arena will be in part driven by the ability of the network engineers and designers to provide the QoS required by the audio streams. Add the video dimension into this equation, and the requirements become even more demanding. It may be that for many of those engaged in these new exciting areas more dedicated network services

will have to be provided. To date we see this happening in the broadcasting domain where the main players in the field have their own networks for the transmission of TV and radio content. In the academic research domain we are lucky to have access to the new generation of high speed networks, although it is the case that most of the activity on these networks is associated with "big science", e.g. high energy physics, astronomy, medical science etc. Gaining access to these networks is not trivial, and provisioning the appropriate links to a performance space often requires considerable work and expense. It also requires gaining access to a number of professionals not usually engaged in working with "artists", and this in itself can present a considerable "challenge" for all parties concerned.

It is worth noting that the tools required to create content need to be written and we also need a new breed of artists to design the content both in the video (which may be multiscreen) and audio domains (which will be in surround sound and even 3D). Add to this the challenge of remote collaboration/composition/performance, it is clear that these new tools must embrace the network technologies from the ground up. Also, these e-artists need to interwork with the e-scientists to draw upon their existing expertise.

Finally, it is heartening to see these issues being aired, and even more important to be able to work alongside established e-scientists who now appreciate the issues associated with these media on the network. To put it simply, it's just data, lots of it, and it needs to run quickly, very quickly, and we really must not lose any of it.

References

[1] R. Bargar et al. (1998). "Networking Audio and Music Using Internet2 and Next-Generation Internet Capabilities." TC-NAS/98/1. Audio Engineering Society.

[2] Dimitrios Miras. (2002) "Network QoS Needs of Advanced Internet Applications – A Survey.

[3] See www.mpegsurround.com

[4] See www.cinegrid.org