Title	Audio quality and capacity issues in network design
Author(s)	Murphy, David
Publication date	2000-05
Original citation	Murphy, D. (2000) 'Audio Quality and Capacity Issues in Network Design', UK AES 15th Conference: Moving Audio, Pro-Audio Networking and Transfer, MA-03, London, UK 8-9 May. Available at: http://www.aes.org/e-lib/browse.cfm?elib=9055
Type of publication	Conference item
Link to publisher's version	http://www.aes.org/e-lib/browse.cfm?elib=9055 Access to the full text of the published version may require a subscription.
Rights	**© 2000, Audio Engineering Society. All rights reserved. This paper was presented at the UK AES 15th Conference: Moving Audio, Pro-Audio Networking and Transfer, as paper number MA-03. The full published version can be found at: http://www.aes.org/e-lib/browse.cfm?elib=9055**
Item downloaded from	http://hdl.handle.net/10468/10344

Audio Quality and Capacity Issues in Network Design

David Murphy
University College, Cork, Ireland

UK AES 15th Conference: Moving Audio, Pro-Audio Networking and Transfer (May 2000)

INTRODUCTION

This paper addresses some of the areas relating to audio quality and capacity issues when designing networks for professional audio applications. It is designed to provide the reader with the fundamental knowledge necessary to make an informed decision about audio quality when designing networks.

Within the professional audio industry, data networks have been used for quite some time for the transfer of audio files from one location to another. A conventional data network (e.g. Ethernet with TCP/IP) is adequate for this purpose, for instance to store digital audio files on a remote backup system.

When transfer times and data integrity are critical other network topologies need to be considered. A network designed for the real-time transfer of multimedia content, including digital audio data, must address the areas of transfer rates, capacity, and data reliability.

Network technologies, such as ISDN, ADSL, and Cable Modem, are now available to the domestic market whilst larger-capacity network technologies, such as Fiber Channel and 100Base-T Ethernet, are commonplace in the commercial sector. Recent advances in network research have given rise to optimized local network technologies, such as IEEE 1394; these facilitate large capacity transfers over inexpensive local area networks. Within the professional audio industry, a well-planned network infrastructure will enable the use of more efficient methods in the recording, postproduction, and transmission of digital audio. The modern audio production process is almost entirely within the digital domain and any advancement that makes this process smoother will result in higher productivity.

The recording/production facility of tomorrow will probably consist of a distributed system, which will only be limited by the underlying network infrastructure. For instance, the actual recording environment and control room might be co-located whilst the resources for mixing (outboard effects, etc.) might be rented from an external company. In this situation the audio data would be routed in 'real-time' to and from the external site. Alternatively, musicians and engineers might exist in several off-site locations equipped with the appropriate technologies to sustain real-time recording and synchronization of multiple audio and control channels.

Presently the situation is rather different, real-time recording/editing/manipulation is only feasible within constrained environments, of which there are a few examples. The key to the future success of this approach is the level of transparency of the distributed system, i.e. the environment must be completely integrated and there should be no perceivable degradation in the quality of the audio as a result of this new model.

As networking capacity increases one would assume that it would be more than adequate for professional audio applications. However, research in audio has also advanced and consequently professional audio requirements have increased. Therefore network capacity must now handle high-quality audio files with settings of 24bit/96 kHz and multiple channels of sound.

There are several different implementations of networks to be found within the professional audio industry. At a high level these can be categorized into two areas: downloading and streaming of audio data. Briefly, downloading concerns the transmission of packetized audio data in a non-real time mode, whilst streaming involves the transfer of audio data in real time. Audio files that are downloaded are usually transferred from one environment to another for storage purposes or delivery to other sites for further work. Before examining the real-time requirements for audio transfer, it is best to standardize on what is understood by the term "real-time". The Oxford Dictionary defines real-time as:

> "*Computing* **a** (of a system) in which input data is processed within milliseconds so that it is available virtually immediately as feedback to the process from which it is coming…" [1]

From this we can infer that in terms of network transfer, real-time implies the near instantaneous transfer of data. This is a very important concept in professional audio since any latency in the transfer process can be aurally perceived. Real-time applications, such as live recording or live broadcast, are critical situations and cannot tolerate any loss of information or time. Therefore, when planning a network infrastructure one has to consider carefully the audio quality requirements.

To determine the networking requirements for audio, three distinct areas can be identified: minimum level of audio quality, compression schemes, and network-type.

AUDIO QUALITY ISSUES

The standard format for representing digital audio is linear Pulse Code Modulation (PCM) data. PCM contains discrete information that represents various attributes of the audio signal, such as phase and magnitude.

The problem with codifying audio data as linear PCM is that the quantity of data used to represent the original is substantial. PCM is so data intensive that it is impractical to transfer more than a pair of stereo channels of uncompressed PCM data over moderate data networks in real-time. Table 1. lists data rates for typical professional linear PCM (data rates for 96 kHz have not been included).

Audio quality is dependent upon a number of factors these include available bandwidth, dynamic range, spatial attributes, etc. The audio material and its intended application are also important in determining a minimum acceptable level of quality. For instance, the quality requirements for speech are different to those for high-fidelity music.

The primary concern in speech coding is the intelligibility of the information. The process of coding and transmission must not affect the comprehension of the information, also known as entropy. Speech has a limited bandwidth and dynamic range; therefore it requires less data to represent the content. Hence it is suitable for transfer over low-capacity networks.

Music, however, contains important elements that are spread across the audible spectrum and covers a relatively large dynamic range – this equates to more content to codify. In order to transfer high-fidelity music over a data network in real-time a compromise has to be made between quantity and quality. To compound the matter further 24bit technology is now commonplace in the professional environment and professional sampling rates are moving from 48 kHz to 96 kHz. This means that more bandwidth is required to transmit professional audio data.

Table 1 Typical linear PCM settings and associated data rates

Sampling Rate	Channel Mode	Wordlength	Bit Rate per second
44.1kHz	stereo	16-bit	1,411,200
48kHz	mono	16-bit	768,000
48kHz	stereo	16-bit	1,536,000
48kHz	stereo	24-bit	2,304,000
48kHz	5-ch.	24-bit	5,760,000

Digital Noise

To quantify quality levels a brief review of noise/error metric systems is beneficial. In digital audio noise is quantified using the Signal-to-Noise Ratio (SNR). This metric measures the ratio of the noise floor to the signal level. In digital systems this is approximately 6 dB for every bit in the data word. For instance, the SNR of an 8 bit word is circa 50 dB and for a 16 bit word the SNR is circa 98 dB (this is the CD Audio standard). The higher the ratio the less perceived the noise is, conversely, the lower the ratio the higher the perceived noise. An equivalent unit of measurement is used in data systems. One of the primary causes of digital noise is lost or incorrect bits in the data word or message. The Bit Error Rate (BER) is used to measure the tolerance level of the system to bit errors – this is also perceived as audible noise.

Analog signals have a continuous waveform and because they are continuous, different types and levels of noise can be tolerated. Digital signals, on the other hand, are less tolerant of noise. The noise encountered in the transfer of a digital signal is caused by wrong bits or lost bits. Consequently the data is either wrong or right and there is no intermediate stage between the two as there is in the analog domain. This can be considered as one of the main strengths and weaknesses of the digital process. An error can be introduced very easily, however when an error is encountered it can be rectified quickly (simply by inverting the bit value). This is very dependent upon the type of error, whether it is a short error (random error) or it is a sustained error (burst errors). Therefore, if the data suffers from a burst error, such as a scratch on a CD or a series of lost packets in data transfer, the error can be quite severe and is realized as audible noise or distortion.

Degradation of audio quality is first encountered at the initial stage of digital recording. During the process of converting an analog signal to the digital domain two potential areas for errors are quantization-noise and jitter. Briefly, quantization is performed after the analog audio has been sampled. It is the process of mapping the dynamic range of the signal into a series of discrete binary values. This leads to a rounding-off of the signal's amplitude value, as there are only a finite number of levels for representing the amplitude. This equates to an error of ± 0.5 of a quantization interval. In an 8 bit system there are 256 levels of quantization and in a 16 bit system there are 65,536 levels available. The more quantization levels available the less noise is introduced by rounding-off.

Jitter occurs when there is a temporal variation in the sampling process, i.e. the time intervals between each successive sample are not equal. As with quantization noise, jitter is perceived as noise and distortion – the type is determined by the properties of the jitter [2]. Periodic Jitter can lead to phase modulation, where audible sidebands are produced above and below the audio signal. These sidebands can be calculated by the formula:

$$f_c-f_m \text{ and } f_c+f_m \quad \text{(Formula 1)}$$

where f_c is the frequency of the carrier and f_m is the frequency of the modulator, which in turn is the frequency of the jitter. In practical terms, jitter of 10ns is audible in an audio CD and, as a rule of thumb, the higher the audio frequency the more susceptible it is to jitter.

Multiple Channels

In a professional environment audio files normally consist of multiple channels, e.g. there are normally a number of tracks in any one project. The introduction of multiple channels increases the capacity requirements for transferring audio data. Another important issue for consideration is synchronization between the channels. When transferring audio tracks as separate data streams or packets, maintaining synchronization is an important aspect of the transfer process. Any delays could potentially destroy the synchronization between channels. Generally, when transferring multiple channels of audio there are three options available: (1) interleave the channels at source, (2) interleave the audio data packets, or (3) transfer the channels as separate streams synchronized together by a control channel.

A number of audio compression schemes take advantage of the similarities in stereo channels. This is known a Joint Stereo Coding and is based upon the M/S Stereo (sum and difference) principle. This technique is very successful and results in a coding gain (reduction in the number of bits to be transferred).

Spatial Information

Information about the spatial attributes of a sound are normally contained in high frequencies, transients, and in the relationships between multiple audio channels. A delay in receiving a channel could very well break down the spatial image within the file. High lossy compression rates or lost data packets can also remove important components used in the creation of the spatial field.

COMPRESSION

Considering the substantial amount of audio data that is required to reproduce the original sound one can conclude that some form of compression is necessary for the real-time transfer of audio data. The goal of compression is to attempt to reduce the file size without modifying the information. Data compression is not a new field. Early uses of information compression involved primitive, yet effective, techniques such as smoke signaling by American Indians, bon-fires in Europe, and, more recently, Morse code. These examples are representative of the two types of compression techniques available, loss-less and lossy. Table 2. contains a list of common compression schemes used within multimedia.

Loss-less Compression

In general, all signals contain the required information, also known as the entropy, and an element of redundancy. Examples of redundancy would typically include superfluous information, or repeated occurrences of a particular piece of information. Loss-less coding compresses the signal without detracting from the entropy. In certain cases the redundant information can be removed without affecting the entropy.

In loss-less compression redundancy can take the form of statistical redundancy, for instance a symbol that appears several times within the data – this is very common in computer data. When redundancy is removed from a signal the entropy increases and, correspondingly, the system's error sensitivity increases. In effect, this means that if any data bits are lost, or, errors occur, the required information is more likely to be compromised. This is not tolerated in a computer environment where every byte and bit is important, for instance in a company's payroll system. This type of information is crucial and cannot be lost.

Loss-less coding however is limited and as a result the coding gain for audio data would not be sufficient to enable it to be transferred in real-time over a data network. An alternative approach can be found in compression schemes using lossy compression techniques.

Table 2 **Different coding techniques and various implementations**

Entropy Encoding	Source Coding		**Hybrid Coding**
	Coding Type	Name	
Run-length Coding	Prediction	DPCM	JPEG
Huffman Coding		DM	MPEG
Arithmetic Coding			H.261
	Transformation	FT	DVI
		DCT	
	Perceptual	Sub-band Coding	
	Layered Coding	Bit Position	
		Subsampling	

Lossy Compression

Multimedia systems, which deliver information for human consumption and not computer systems, are more tolerant of errors; this is primarily due to the mechanisms employed in human perception. For instance, when visual information is processed by the visual system, a lot of the information is not processed, e.g. very high frequencies and colour definition. Lossy compression involves either perceptual encoding or predictive coding. Predictive coding is based upon the coding of the difference information between successive samples or frequency coefficients. A common example of differential coding in audio is the ADPCM coding scheme. While there is a reduction in the quantity of data it is not sufficient for the transfer requirements of professional audio data.

By virtue of the high coding gain achieved, perceptual codecs (compression/decompression) are the de facto choice for multimedia applications, particularly in the field of audio. Selective adjustment or even elimination of particular components of the audio data can accomplish the necessary coding gain. This process should increase the entropy level of the signal while at the same time removing redundancy. However, this is a subjective process of data reduction and in certain cases trained listeners can notice a reduction in quality.

Perceptual compression is based upon a psychoacoustic model of hearing. To fully appreciate how perceptual encoding is used in audio codecs an understanding of the hearing system is necessary.

Hearing System

The hearing system is a sophisticated, finely tuned system that has matured over mankind's evolution. It is a selective mechanism that emphasizes wanted information and de-emphasizes, or even discards irrelevant information. One can conclude that the natural hearing process is a non-linear system [3]. Below is a brief account of the selective attributes of the hearing system.

After a series of transductions the audio signal is passed to the Basilar Membrane. The Basilar Membrane is one of two membranes situated in the inner ear, more correctly called the cochlea, which also contains fluid and is bounded by rigid walls of bone. When sound arrives at the Basilar Membrane the energy is transmitted from the Base of the membrane to the Apex. The membrane functions as a quasi spectrum analyzer, where high frequencies have maximum displacement (amplitude peak) at the base and low frequencies reach maximum displacement just before the apex. In between the base and the apex of the membrane all other frequencies within the audible range register their respective amplitude peaks. Each point on the Basilar Membrane has a Characteristic Frequency (CF). This can be defined as the frequency that has maximum displacement at a particular point on the membrane.

The Basilar Membrane's frequency selectivity is based upon the same principle as conventional filtering. Each CF has an associated bandwidth, however these bandwidths are not uniform across the Basilar Membrane and are therefore frequency dependent. The CF bandwidth has been described as a Critical Band by Fletcher [4], and later developed by Zwicker [5]. Zwicker subdivided the audible frequency range from 20 Hz to 15 kHz into 24 bands. The critical bands are not uniform across the frequency range, for instance critical bands are much narrower in bandwidth at low frequencies. Consequently, three quarters of the critical bands are below 5 kHz. In practice, a 25th Critical Band is included to encompass the frequencies from 15 kHz to 20 kHz. Each Critical Band is bounded by upper and lower cutoff frequencies and the narrower its bandwidth the higher the Q factor.

Masking

Masking is a phenomenon that occurs within the hearing process. It is a natural mechanism for the removal of redundant elements in an audio signal. In general, if two frequencies of differing amplitudes fall within the same critical band, the frequency with the lower amplitude level will be masked by the dominant frequency. In effect this means that the listener will not perceive the frequency with the lower amplitude level. The dominant frequency is normally called the masking tone.

The scope of the masking effect, in terms of the range of frequencies, is characterized by the masking curve – this is also known as the masking threshold. The masking curve is asymmetrical about the masking frequency. The masking threshold extends above the frequency of the masking tone much more than it extends below it. Again, these properties are also frequency dependent.

Masking can also occur as a function of time, this is known as non-simultaneous masking. In short, the effect of masking can extend forward in time after the masking tone has sounded. This has potential for masking new sounds upto 100ms after the masking tone. There is also an element of backward masking where sounds preceding the masking tone can be masked. However, this affect is not as pronounced as frontal masking and normally only extends to 5ms before the masking tone.

Codecs

There are several perceptual codecs available that are currently in use within the professional audio industry, these include the ISO standardized MPEG codecs and proprietary codecs such as Dolby's AC3 and Lucent Technology's PAC, amongst others.

Perceptual codecs exploit the non-linearity of the hearing process. In general, perceptual codecs split the input signal into a series of sub-bands by means of a filter bank (32 sub-bands in the case of MPEG 2 Layer III and 4 in apt-X100). Sophisticated filtering techniques, such as (polyphase) quadrature mirror filtering (QMF), are used in more advanced professional audio codecs. Sub-bands are based upon the Critical Bands of the hearing system. However, unlike the Critical Bands, sub-bands are generally uniform across the spectrum.

The output from the sub-bands is then analyzed to find the masking threshold within each band. A Fast Fourier Transform or a Discrete Cosine Transform is normally used to perform the spectral analysis of the signal. The output of this is then passed through a psychoacoustic model to determine the masking threshold of each band. Based upon the outcome of this analysis a variable wordlength is assigned to each band that represents the dynamic level within (a form of requantization). There is normally a substantial coding gain at this stage as the number of bits required for each sub-band varies according to its Noise to Masking Ratio (NMR). Watkinson maintains that "professional devices should have a greater NMR than consumer or programme delivery devices" in order to preserve the high-fidelity quality of the audio [8]. The output values are then multiplexed with control data and transmitted as a serial stream. Decompression involves a reversal of the above procedure.

MPEG 2 is an example of a codec that has been received well by the professional audio industry. This codec consists of three layers, Layer I, II, and III. Layer I is the most basic compression scheme of the three while Layers II and III employ more sophisticated compression methods. Of the Layers, Layer III is the most interesting. It is comprised of a combination of both the MUSCIAM and ASPEC algorithms and is known as AAC (Advanced Audio Coding). Briefly, AAC incorporates a 32 sub-band polyphase QMF, the outputs of which are further processed by a Modified Discrete Cosine Transform (MDCT), an advanced psychoacoustic model with non-uniform quantization, and finally Huffman encoding. According to Brandendburg and Junz [7], "upto transparent audio quality" can be achieved. This is an important consideration for professional audio applications where a trade off between quantity and quality must be achieved.

Cascading Codecs

In terms of quality, the cascading of perceptual codecs will affect the integrity of the audio data. The higher the number of lossy compressions the data has to undergo the greater the noise level. In practical terms this poses a dilemma within broadcast/transmission centers for example. For instance, when an MPEG stream is sent to a transmission center to be broadcast in a DAB multiplex. One of the methods for doing this is to take the MPEG stream, which has already had the redundancy removed, convert it back to PCM and then re-encode it so that it can be multiplexed into an DAB carrousel of channels.

This does not apply, however when a perceptually compressed file is further compressed using a loss-less codec. For instance, MPEG 2 Layer III incorporates Huffman coding after the perceptual encoding process is complete. This is acceptable, as it has no affect upon the audio quality – this is merely entropy compression. The most negative aspect of multiple compressions, or cascading codecs, is the introduction of additional latency. For each number of compressions undertaken there has to be a corresponding decompression.

In real-time applications latency is a major concern, especially when it is incurred during the encoding and decoding stage. Generally, the higher the coding gain the greater the latency, the lower the coding gain the shorter the latency. An ideal system would have maximum coding gain with no latency – currently this is not feasible.

Perceptual encoding is not a perfect solution for professional audio simply because it removes elements of the audio signal. Frequencies determined to be irrelevant, or not perceived are removed form the spectrum. This could have a detrimental affect upon the timbre of particular sounds or might result in destroying spatial cues used in sound localization. Amplitude levels are altered, which can lead to fatigue in the listener after prolonged listening. The type of windows used in the filters and transforms can generate transient effects within the signal. Careful consideration of the type of codec is very important as each codec has strengths in different areas. For example, a codec that was designed for compressing pop-music could be disastrous for classical music. However, there are not many options available when the network infrastructure cannot support the real-time transfer of high quality audio data.

NETWORK CONSIDERATIONS

When designing the network infrastructure there are a number of tools available to ensure a particular level of service and data reliability, of which Quality of Service (QoS) and Error Detection/Correction are the most important. A more detailed review of networking technologies can be found in another paper presented at this conference by Bailey [8].

QoS

In the field of networking, user and application requirements are mapped into a QoS, which guarantees a minimum level of service for data transfer. QoS is essentially concerned with resource management. Where a resource is a system component used for the manipulation of data. Resources can be shared or used exclusively, and are normally divided into (1) passive resources (e.g. bandwidth) and (2) active resources (processor time). Steinmetz and Nahrstedt [9] have divided QoS into four distinct areas:

1. Applications QoS. These are the requirements for the application services; they are usually specified in terms of (1) media quality, e.g. minimum acceptable delay, and (2) media relations, e.g. synchronization between separate audio channels.
2. System QoS. These are the requirements of the OS and the communication services as a result of the Applications QoS. These can be categorized as quantitative criteria (e.g. bits per second, BER, processing time, etc.) and qualitative criteria (scheduling, error handling, etc.).
3. Network QoS. This may be specified in terms of (1) network load and (2) network performance.
4. Device QoS parameters specify timing and throughput demands for media data units.

Most of the current network protocols are based on 'best effort services'– these are services based on either no guarantees (i.e. no QoS specification required), or on limited guarantees (i.e. some level of QoS is specified). When specifying a QoS for transmitting audio certain resources need to be reserved, this is known as 'resource reservation'. If no resource reservation is made at either the end-system or in the routers and switches then the likelihood of dropped or delayed packets is high. This will lead to errors when the audio data is played.

Depending upon the QoS specified the system will alter some elements of the data. This is known as 'scaling' in network parlance. Scaling involves the subsampling of a data stream and results in only a portion of the original content being transferred. In general, scaling can be done either at the source or at the receiver. There are two types of scaling available: (1) transparent (undertaken by the transport system – dropping portions of the data) and (2) non-transparent (interaction between the network layers – modification of stream before it is presented to the transport layer, e.g. modifying a coding algorithm's compression ratio). In relation to audio, transparent scaling is difficult. Any modification of the audio data will be perceived by the listener, for instance, dropping a channel of stereo. Therefore non-transparent scaling must be used at the source.

QoS can also specify the transfer times for a network using Rate Control. A high rate control will result in synchronous data handling which equates to an imposed fixed data rate. Rate Control has to calculate time constraints (including delays), space constraints (system buffers), device constraints (frame grabber allocations), frequency constraints (network bandwidth and system bandwidth), and finally reliability constraints (error detection/correction).

Error Detection and Correction
A high degree of data reliability in audio applications is necessary for a number of reasons. The majority of audio compression schemes cannot tolerate loss. Lost packets will affect synchronization within the audio data or will be realized as audible clicks or distortion. Another important aspect of data loss is that the error might be irreparable. For instance, in real-time recording, lost packets could be disastrous, as there is no opportunity for a re-take.

For these reasons audio data needs to be protected to some degree and the mechanism for doing this incorporates the detection and correction of errors. Error correction can be done at the source level (by the codec for instance) or by the network layers (error checking in transfer protocols for example). There are several error correction schemes available, some of which are optimized for speed while others focus on data reliability.

In a shared network protection from errors is advisable, as the likelihood of data loss is high. On a dedicated network the need for data protection diminishes, primarily because there is no other traffic to interfere with the transfer of audio data. As with the coding process, error protection incurs a heavy penalty in terms of latency. This is especially true if there is more than one protection scheme in place (e.g. where both the codec and network protocol are using error protection).

Latency
Latency is one of the biggest issues in the transmission of digital audio data over networks. System latencies include delays introduced by the coding process, the transmission of the data, the decoding of the data, and any duplex handshake operation such as receipt acknowledgements. Where possible latencies should be avoided at every stage of the coding-transfer-decoding process.

Interface compatibility
Complex networks can involve inter-networking (networking between different network types) and requires the re-framing of packets to ensure compatibility between the networks and devices. When planning a network the designer should be conscious of the number of transformations the audio data has to undergo in order to pass through different interfaces. Some codecs place the coded audio data into a framed stream, which is also packetized by the network layers. If possible a uniform approach should be adopted across the network. This also reduces any potential latency that might occur during data conversion.

Synchronization
Synchronization within an audio system is crucial, if the synchronization between two devices is wrong then this will have a noticeable affect upon the audio. Synchronization problems also occur when multiple channels of audio are transmitted. Not alone must the timing within each channel (**intra**) be correct, the synchronization between channels (**inter**) must be exact in order to reproduce the required sound. This is crucial in multi-channel environments where poor synchronization can lead to a multitude of problems including phasing, and a breakdown of the spatial image within the sound field.

Network Technologies
Real-time audio transfers necessitate the use of a high capacity network that can facilitate reliable and timely transfers. Network technologies such as Ethernet are not suitable for real-time transportation of audio data. The main reasons for this are the high latencies associated with Ethernet and the increased potential for dropped packets. Ethernet was designed for the reliable transfer of computer data and not continuous media data. Hence, when network usage increases the performance hit is spread across equally to all traffic on the network. Transmitting audio data rely requires a dedicated channel or path with a fixed bandwidth.

To ensure reliability of data transfer Ethernet employs a system for checking the success of the transfer and for reissuing any errant packets again. While these are important factors for transferring computer data they incur a heavy time penalty that makes this infrastructure inadequate for the transmission of continuous media data. That is why protocols like TCP are not suitable for the delivery of real-time audio.

A more suitable network environment for professional audio applications might include technologies such as ATM over Fiber Optic cable. While this may be considered excessive the advantages soon outweigh what may be considered extravagance. Fiber optic technology offers higher capacity than Ethernet and doesn't suffer from the physical limitations of copper cable.

ATM's use of virtual connections and paths guarantees a fixed bandwidth and facilitates multiple connections (one for data and one for control information maybe). ATM doesn't employ the same packet reliability as Ethernet, however because there is a point-to-point connection, albeit logical, the potential for lost packets is greatly reduced.

CONCULSION

Audio Data
Professional audio data must be of the highest quality and this level of quality must be maintained throughout the recording/production process. Therefore careful consideration is required when implementing a network infrastructure or transferring the audio data over an existing network. The network is merely a vehicle for transporting data from point A to point B and should not have a negative impact upon the quality of the audio data.

Selection of the codec

In professional audio applications, one would expect that the audio material would not be degraded by means of some perceptual codec. Therefore careful attention is required when selecting the appropriate codec. Recent research by Soulodre et al. indicates that there is a large discrepancy in terms of subjective quality amongst a range of professional audio codecs when compressing two-channel material [10].

Currently, the ITU-R are assessing a new objective measurement tool, called PEAQ, for the assessment of perceived audio quality [11], [12]. A tool of this type combined with subjective evaluation would be an invaluable instrument in determining the suitability of a perceptual codec for professional audio applications.

Selecting the Network

Real-time transfer of continuous media, such as audio data, necessitates a large capacity network and protocols that can manage the data transfer in a timely and reliable manner. Common data networks, such as TCP/IP over Ethernet, are not capable of transferring high quality audio data in real-time. Hence, more sophisticated configurations are more appropriate.

The network infrastructure should incorporate a level of QoS that guarantees timely and reliable delivery of audio data. A minimum level of error detection/correction is also advisable. Mixed networks should be avoided where possible and interface compatibility should be taken into account.

When designing the network, a capacity for future expansion should be built into the design. This could incorporate a mixed media network (video and audio), an element of computer data networking, and a facility to cope with increased capacity demands.

REFERENCES

[1] H. Fowler et al., *The Concise Oxford Dictionary*, 9th Ed. (OUP, Oxford, 1995).

[2] F. Rumsey, *The Audio Workstation Handbook* (Focal Press, Oxford, 1996).

[3] B. Moore, *An Introduction to the Psychology of Hearing*, 4th Ed. (Academic Press, London, 1997).

[4] H. Fletcher, "Auditory Patterns," *Rev. Mod. Phys.*, vol. 12, pp. 47 – 65 (1940)

[5] E. Zwicker, "Subdivision of the audible frequency range into critical bands," *J. Acoustical Soc. of America*, vol. 33, p248 (1961).

[6] J. Watkinson, *The Art of Digital Audio*, 2nd Ed. (Focal Press, Oxford, 1994).

[7] K. Brandenburg et al., *MPEG-4 Natural Audio Coding*, http://www.tnt.uni-hannover.de/project/mpeg/audio/

[8] A. Bailey, "Tutorial: introduction to network principles for audio engineers", Moving Audio: pro-audio networking and transfer, *AES UK Conference*, (2000 May).

[9] R. Steinmetz and K. Nahrstedt, *Multimedia: Computing, Communications & Applications*, (Prentice Hall, New Jersey, 1995).

[10] G. A. Souldore, T. Grusec, M. Lavoie, and L. Thibault, "Subjective Evaluation of State-of-the Art Two-Channel Audio Codecs," *J. Audio Eng. Soc.*, vol. 46, pp. 164 – 177 (1998 March).

[11] T. Thiede et al, "PEAQ – The ITU Standard for Objective Measurement of Perceived Audio Quality," *J. Audio Eng. Soc.*, vol. 48, pp. 3 – 29 (2000 January/February).

[12] W. C. Treurniet, and G. A. Souldore, "Evauluation of the ITU-R Objective Audio Quality Measurement Method," *J. Audio Eng. Soc.*, vol. 48, pp. 164 – 173 (2000 March).