Proceedings of the Institute of Acoustics

INTERNET AUDIO QUALITY

R. C. CROSS.

Audio Engineer.

1. INTRODUCTION.

It is said that one year in the computer industry is equivalent to three years in 'normal' life. Internet Protocol Local Area Networks are becoming more powerful. Ethernet bandwidth is increasing from 10 Mbps to 100 Mbps over the next couple of years and Gigabit Ethernet is just around the corner [1]. The rate of development of computer and telephony convergence and the decrease in the cost of ownership of hardware and access to the internet means that streaming audio will increasingly become an everyday delivery system for music listening. An obvious example would be real time Juke Boxes rather than overnight non-real time equivalents.

This paper discusses the differences between the subjective assessment of voice and music delivered over IP networks. A range of samples will be played during the presentation together with the bit rate information. Finally an ad-hoc experiment will be conducted with the conference audience. The audience will be invited to listen to randomised identical music samples with the objective of guessing the bit rates used. Although it's always useful to understand the Technology, - properly scoped technology should be transparent - leaving the listener to answer the question, what does it actually sound like ?

2. VOIP.

VOIP or Voice Over Internet Protocol is being used more and more at the present moment as a cheap alternative to the telephony provided by conventional operators. The companies providing the service are sometimes known as 3G (Third Generation) Telco's as the service is similar to a piggy back or parasitic operation, using the Internet backbone as the trans country link. This will probably lead to call billing which is independent of distance. Use of VOIP technology is already available to provide PABX telephone facilities within a Corporate Intranet.

Each manufacturer uses their own proprietary coding / decoding algorithm but they are all similar to the GSM (Global Systeme Mobile) scheme which is optimised for speech. As VOIP has become more popular and usable, the coding schemes are being standardised. The ITU (International Telecommunications Union) have recently produced the H.323 V2 from Study Group 16. H.323 is a recommendation that sets standards for multimedia communications over Local Area Networks (LANs) that do not provide a guaranteed Quality of Service (QoS). Recommendations include; Codec Standards, Interoperability, common call set-up and control protocols, Network Independence, Platform and Application Independence, Multipoint Support, Bandwidth Management and many associated with multipoint conferencing.

The introduction of H.323 demonstrates the advances made in this area and highlights by omission two elements which are essential to telephony conversation Subjective Opinion Scores; namely Quality of Service and Latency. Quality of Service is essentially the speed of call set up and continuing availability of the link. Latency is the conversational delay caused by the fact that the speech is intrinsically packetized, multirouted and stored by the network. Rather than use the spare PC memory as a buffer, newer network nodes now incorporate their own cache or buffer to reassemble the data, thus reducing the latency.

Naturally, Latency or delay is much more of a subjective disturbance during a two way conversation than simply listening. For this reason, listeners will concentrate on the bandwidth / noise / distortion aspects and people engaged in conversation will concentrate less on these and more on the impediments to ease of conversation. To further complicate the issue, it is suspected that the measure of QoS and Latency and its effect on the Conversational Mean Opinion Score will not be a simple linear interaction with the instantaneous speech quality.

3. SUBJECTIVE ASSESSMENT OF MUSIC QUALITY - LISTENING TEST.

Six different clips of a selection of music are available, recorded at different bit-rates. The audience will be presented with these in a randomised order. Those who wish to take part will be asked to rank the samples in reproduction quality order, with further rows for bit rate and Opinion Score. A separate formsheet will be available for this purpose at the conference presentation. The table is shown below.

Clip	1	2	3	4	5	6
Rank						
Opinion Score						
Bit rate ?						

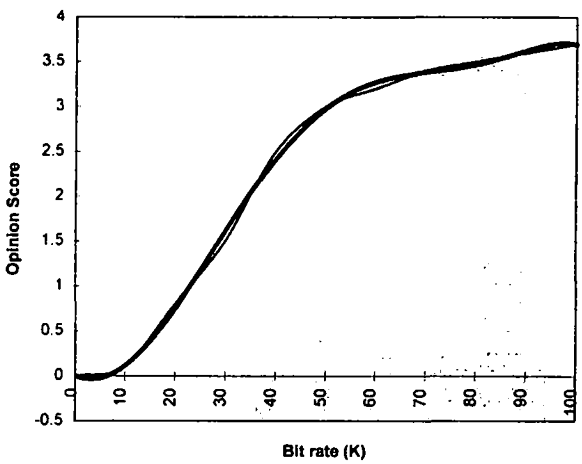
Note: For this purpose the Opinion Score will use the following Scale:

4 = Excellent, 3 = Good, 2 = Fair, 1 = Poor, 0 = Bad.

Interpolation between these integers is allowed.

This is by no means a controlled experiment since the music content may not be to everyone's liking and the upper limit will be set by the ambient noise level and the replay system level. Additionally, there is no reference condition used for subject training. The clips will be played twice and it will probably be easier to mark on the second play. The information will be useful to the author and whoever takes part will be encouraged to leave the filled form behind for further processing.

The data should be a curve similar to the following:



Opinion Score Vs Bit rate for a range of Clips.

4. CONCLUSION.

The topic area, research and uses of Internet Audio are expanding at a dizzy rate. Presently, we can only take snapshots of the progress that is being made. Rather than the technology, the Author is interested in the Subjective performance of these delivery systems, particularly with reference to the discerning listener. It's intended to run a small ad-hoc experiment using Reproduced Sound 14's audience to gain insight into the perception of the sound quality encoded in a widely used proprietary system.

5. REFERENCES.

[1]: Databeam H323 primer. http://www.databeam.com/h323/h323primer.html