An Overview of Audio Technologies in Teleconferencing: From Source to Receiver and Back

Bradley Basnett
Teleconferencing Department
Bell-Northern Research
Ottawa, Ontario

INTRODUCTION

In this paper, a typical round-trip teleconference audio path will be described. Figure 1 shows the most common components along the teleconference voice path.

SOURCE AND ROOM INTERACTION

A teleconference participant's mouth is located approximately 0.5-1.0 m from a surface mounted microphone exhibiting a half-hemispherical cardioid polar pattern. The acoustic energy arriving at the microphone is comprised of direct and reflected energy. The reflections degrade the intelligibility and overall transmission quality of the talker's voice. The arrival time and amplitude of each reflection are dependent upon the proximity and composition of the room boundary surfaces. Fortunately, the directional nature of the microphone reduces the pick-up of off-axis reflections. To further reduce the impact of the reflections, the well designed teleconference suite will be outfitted with strategically located absorptive treatment. Suggested acoustic treatments for teleconference suites are given in [1,2].

Early /Late sound ratios are an indicator of acoustic clarity [3]. Early/Late ratios with an early time limit of 50 ms have been measured to be 25-30 dB for a cardioid microphone at 0.5 m from an on-axis talker in a teleconference suite with a average decay time of 350 ms [2].

NOISE

Room ambient noise is an unwanted traveler on the teleconference audio path. Teleconference suites should not have ambient noise in excess of NC-30. The largest contributor to ambient noise is usually the HVAC system. Meeting tools such as overhead projectors and computers can increase the ambient noise by 5-10 dB . Often, these meeting tools are located on the conference table in close proximity to the teleconference microphones. Every effort should be made to reduce HVAC noise and to select meeting tools with quiet cooling systems. The conference room partitions should have an STC rating of 50 dB to insure that outside noise does not interfere with the teleconference.

In addition to the audible impact on the near-end and far-end participants, room ambient noise degrades the convergence performance of acoustic echo cancellers [4]. Fast converging acoustic echo cancellers are essential for full-duplex conversations involving teleconference suites.

Recently available teleconferencing systems have incorporated DSP based noise suppression. Active noise control for rooms is also available [5].

ELECTROACOUSTIC SYSTEM INTEGRATION

Microphones are positioned in the teleconference suite so that each participant is within +/- 30 degrees of the most sensitive axis of the cardioid microphone. A typical suite may have 6-8 surface mounted cardioid microphones. The transmission of room noise and acoustic echo is reduced by passing each microphone signal through an automatic microphone mixer. The mixer selects the microphone that it will turn "on" based on acoustic level and arrival time of the talker's voice. The mixer will turn "on" additional microphones as required. Care must be taken with the amount of attenuation applied to the "off" microphones as it negatively impacts the adaptation performance of acoustic echo cancellers. "Off" attenuation of 8-10 dB has proven to be effective.

SIGNAL PROCESSING

Subsequent components in the audio path may contain automatic gain control (AGC). The AGC circuitry attempts to bring the level of audio signal from the microphone to a pre-defined target value. This target value is chosen to ensure that a quiet talker and a loud talker are passed to the transmission circuitry at a nominal level.

The majority of teleconference systems operate partially or completely in the digital domain past this point in the audio chain. The audio signal from the microphone mixer is digitized and processed. Currently, the amount of processing that occurs in this portion of the audio path varies widely. One of the more elegant methods [6] will be examined.

In aid of later echo-cancellation, echo-suppression, double-talk detection, and 7 kHz transmission bandwidth, the microphone mixer signal is sampled at 16kHz and passed to a bandpass filter bank. The bandpass filters are numerous (29) and narrow (250 Hz). These bands are then subsampled at 1kHz. The subsampling enables efficient DSP to be performed.

The subsampled microphone signal then serves as one of the inputs to the acoustic echo canceller. The echo canceller is in the form of an adaptive filter. The adaptive filter subtracts that portion of the microphone signal that contains energy related to the receive signal as radiated by the loudspeaker. The other input to the adaptive filter is the subsampled version of the receive signal that is eventually radiated by the loudspeaker.

The loudspeaker(s) is typically positioned to provide good coverage of the seated teleconference participants while being off-axis to the teleconference microphones.

The size (number of taps) required of the adaptive filter is related to the decay time of the room in each band and any known non-linearity in the system [7]. The echo canceller described above can reduce the resultant acoustic echo by approx. 20 dB. In the case of audio for low-bit rate videoconferencing, transmission delays of up to 500 ms are inserted at the transmit codec to achieve lip-synch. 20 dB of echo reduction is not sufficient under this condition. Non-linear echo suppression is required to further reduce the acoustic

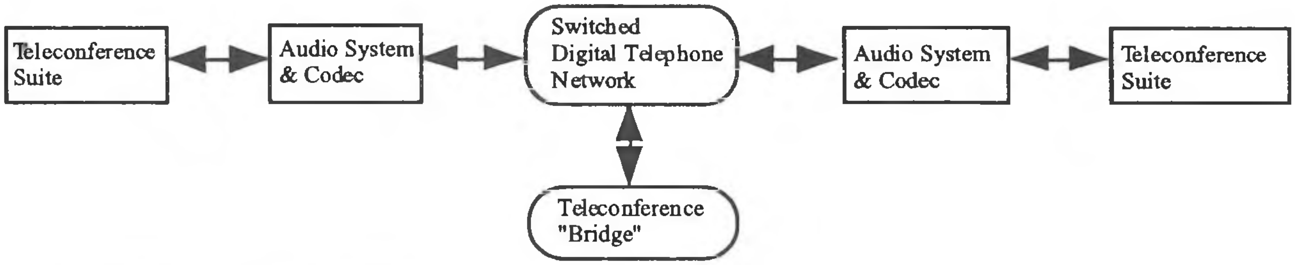

Figure 1. Typical Teleconference Audio Path

echo to an acceptable level. The suppression is in the form of subband transmit gain reduction. When the algorithm determines that the signal at the microphone is largely comprised of signals from the loudspeaker, the gain of that particular transmit subband is reduced.

TRANSMISSION

The processed microphone signal is then prepared for transmission. In some cases, this means that it will be converted back to an analog audio signal and enter the public telephone network. The increasingly more common scenario, whereby the signal enters and exits the telephone network in digital form, will be described here.

In order to transmit digital audio over digital telephone lines, some form of data compression is required to reduce the number of bits per second to a manageable and cost-effective number. Data compression is achieved through redundancy and irrelevancy removal techniques. Most commonly available digital telephone lines are 56kbit/s (switched 56) or 64 kbit/s (ISDN BRI). There are several compression (coding) techniques for transmitting audio over the digital telephone network. Standardized algorithms are now seeing broader use but several proprietary methods are also in active service. Proprietary techniques usually offer lower bit rates via greater compression. Standards offer compatibility while proprietary methods offer lower portions of the data stream being used for audio thus freeing up more bandwidth for video in the case of low-bit rate videoconferencing. G.711 and G.728 are ITU standards for narrow band (300-3.4kHz) transmission, and require 64 kbit/s and 16 kbit/s respectively. G.722 is the international standard for wide band (50-7.0 kHz) coding and it requires 48 or 64 kbit/s. Wide band transmission has become commonplace due to the ever expanding digital telephone network. It should be noted that there are devices available that can combine multiple switched 56 kbit/s or ISDN BRI lines to create a virtual higher bit-rate channel for audio bandwidth up to 22 kHz [8]. These devices are known as inverse multiplexers.

Now that the digital audio is coded and compressed it can make its way into the digital network. One scenario has the audio transmission arriving at a similar teleconference system at the far-end. The other is the case where the teleconference has more than two participants and the audio passes from one site to another via a "bridge". This is known as a "multipoint" teleconference. The "bridge" must broadcast audio from the active talker's site to all the other sites. The "bridge" also operates in the digital domain and usually incorporates automatic gain control (AGC), automatic talker recognition (gating), line echo-cancellation, and automatic mixing.

Audio from the active talker's site arrives at the far-end(s) via the bridge. It is then decoded and used as one of the inputs to the echo canceller and to the input of a D/A converter for amplification and reproduction by a loudspeaker(s) in the teleconference suite. The loudspeaker in a video teleconference suite is usually co-located with the video monitors. The loudspeaker output will reach the listener directly and indirectly via the perimeter surfaces of the room . The Early/late ratio (50ms) is 10-12 dB for most seats in a 12 seat video conference suite.

Depending of the acoustic distance between the loudspeaker(s) and microphone(s), the room treatment and directional characteristics of the microphone and loudspeaker, some of the speaker's acoustic radiation may reach the microphone(s). The acoustic echo cancellation and suppression process described above must also be in place at the far-end to facilitate full-duplex communication.

CONCLUSION

Current teleconferences employ many audio components and processes to achieve high quality, wide band, digital transmission from one acoustic space to another.

REFERENCES

[1] D'Antonio, Peter, "Teleconferencing Facilities Acoustical Design", Sound and Communications, pp 59-62, February 1991

[2] Basnett, Bradley, "Acoustical Aspects Related to the Performance of Teleconference Facilities" 126th Meeting of the Acoustical Society America Denver CO, October 1993

[3] Bradley, J.S. and Halliwell, R.E., "Making Auditorium Acoustics More Quantitative" Sound and Vibration, Feb .1989.

[4] Mapes-Riordan, Dan and Zhao, John, "Echo Control in Teleconferencing Using Adaptive Filters", AES 95th Convention Preprint #3755, October 1993

[5] Elliott, S.J. and Nelson P.A., "Active Noise Control", IEEE Signal Processing Magazine, October 1993

[6] Chu, Peter, "Weaver SSB Subband Acoustic Echo Canceller", IEEE ASP Workshop Proceedings, October 1993

[7] Knappe. M, and Goubran.R, "Steady-State Performance of Full-Band Acoustic Echo Cancellers", ICASSP Proceedings, April 1994.

[8] Audio Processing Technology Ltd, "ISDN for Audio", Revision E, May 1994.