REPORT ITU-R BT.2044

Tolerable round-trip time delay for sound-programme and television broadcast programme inserts – Context and rationale

(Question ITU-R 35/6)

(2004)

CONTENTS

Page

1 Scope

This Report is intended to review the effects of delay and level of echo in an audio foldback loop in a broadcast production context. It also reviews the effect of audio-video delay. It does not attempt to take into account the effect of additional reverberation and noise in the listening environment and assumes that there is no significant loss in signal quality from the monitoring system.

2 References and Bibliography

[1] Technical Report ETR 250 [July 1996] Transmission and Multiplexing (TM); Speech communication quality from mouth to ear for 3.1 kHz handset telephony across networks. European Telecommunications Standards Institute. http://pda.etsi.org/pda/queryform.asp

[2] Technical Report ETR 262 [January 1996] Broadband integrated services digital network (B-ISDN); Asynchronous transfer mode (ATM); Video on demand (VOD) network aspects. European Telecommunications Standards Institute.

[3] Technical Report ETR 275 [April 1996] Transmission and Multiplexing (TM); Considerations on transmission delay and transmission delay values for components on connections supporting speech communication over evolving digital networks. European Telecommunications Standards Institute. http://pda.etsi.org/pda/queryform.asp

[4] Recommendation ITU-R BT.1359 Relative timing of sound and vision for broadcasting. International Telecommunication Union (November 1998).

[5] ETSI/GSM Recommendation 06.10 Version 3.2.0 [February 1992] GSM Full Rate Speech Transcoding. European Telecommunications Standards Institute. http://pda.etsi.org/pda/queryform.asp

[6] DAVIS, P. [1995] A tutorial on MPEG audio compression. *IEEE Multimedia*, p. 60-74.

[7] ONVURAL, R. [1994] *Asynchronous transfer mode networks: performance issues*. Boston, Artech House.

[8] Technical Report TR 100 815 V1.1.1 [February 1999] Digital video broadcasting; guidelines for the handling of Asynchronous transfer mode signals in DVB systems. European Telecommunications Standards Institute.

[9] C.S0014-0 Version 1.0 Enhanced Variable Rate Codec (EVRC) (CDMA2000 specification) Third Generation Partnership Project 2 (3GPP2) EIA/TIA (December 1999).

[10] SIU-WAH WONG [1991] An Evaluation of 6.4 kbit/s Speech Codecs for Inmarsat-M System. *IEEE*, p. 629-632.

[11] Recommendation ITU-R BT.1377 Labelling of video and audio apparatus throughput (processing) delay. International Telecommunication Union (1998).

[12] EVEREST, F. A. [1994] *The Master Handbook of Acoustics*. TAB Books/McGraw-Hill.

[13] CCITT Recommendation G.114 (1989) Mean one-way propagation time, Blue Book, Fasc. III.1.

[14] CCITT Recommendation G.131 (1989) Stability and echo, Blue Book, Fasc. III.1.

[15] ES 200 677 V1.2.1 Public Switched Telephone Network (PSTN) [March 1998] Requirements for handset telephony. European Telecommunications Standards Institute.

[16] LOCHNER, J. P. A. and BURGER, J. F. [1958] The subjective masking of short time delayed echoes by their primary sounds and their contribution to the intelligibility of speech. *Acustica,* Vol. 8, **1**, p. 1-10.

[17] MEYER, E. and SCHODDER, G. R. [1952] Über den Einfluß von Schallrückwürfen auf Richtungslokalisation und Lautstärke bei Sprache. *Nachr. Akad. Wiss.*, Göttingen, **6**.

[18] CHURCH, S. On beer & audio coding – why something called AAC is cooler than a pilsner & how it got to be that way. http://www.broadcastpapers.com/audio/TelosAAC08.htm

3 Terms and definitions

Latency: The delay in packet-switched systems due to packet routing and queuing.

Jitter: The variation in delay during a transmission in a packet-switched system.

I-frame: The initial frame in an MPEG group of pictures.

B-frame: A bidirectionally interpolated frame between an I-frame and a P-frame in an MPEG group of pictures.

P-frame: A forward-interpolated frame in an MPEG group of pictures.

Sidetone: An attenuated portion of the transmitted audio returned to the originator. This can be intentional in telephones. It is caused by an unbalanced 2-to-4 wire hybrid or an incorrectly terminated line.

2-wire system: A system in which signals are transmitted both ways over the same transmission line.

4-wire system: A system in which signals are transmitted on a separate transmission line in each direction.

Audio hybrid: A differential audio circuit used to convert between a 2-wire system and a 4-wire system. It generally uses an audio transformer with a split winding in a bridge configuration, or an active equivalent.

4 Abbreviations and acronyms

ms: Millisecond (10^{-3} s)

µs: Microsecond (10^{-6} s)

ns: Nanosecond (10^{-9} s)

codec: Encoder/decoder

ETSI: European Telecommunication Standards Institute

ITU: International Telecommunication Union

MPEG: Moving Picture Experts Group

fps: Frames per second.

5 Causes of delay and echo

5.1 Context

Looped communications systems are commonly encountered in broadcast production where multiple locations or remote locations are used, particularly in live programme production. The loop may be used to connect a studio interviewer and a remote guest or it may be used for foldback from the studio to a presenter at a remote location. Ideally, each separate stream of audio and video should only travel in one direction on this loop. In practice, it is quite common for audio to leak back around the loop to its source, particularly where telephones are part of the circuit. Because of distance and/or the need for encode/decode processes to compress the digitized signal for transmission and restore it again on reception, the signal can be delayed significantly. Delay and leakage around an audio loop combine to produce echo.

Because video has traditionally needed much more stringent synchronization than audio, it has not been a common practice to transmit video around long loops. With the advent of digital video, synchronization is simplified, equipment is shrinking and video transmission around long loops may become a more common practice. At the time of writing however, it is still uncommon for video to travel around a full loop back to its source. The need for synchronization of audio with video may however still make the video delay the limiting factor in an audio-video communication loop.

When using long loops, delay and echo can cause considerable difficulty, both in maintaining technical quality in the signal and from disturbance to participants. Both of these effects interfere with production operations.

Delay and echo are separate but related problems. Delay is due to several factors, principally encode/decode processing and path length from origin to destination[1]. Echo is due to leakage around a loop. As the delay increases, the associated echo becomes more noticeable and more disturbing.

5.2 Historical background

Annoyance from sound delay and echo is familiar in large auditoriums and sporting grounds and is hardly a new problem. In electronic communications, the problem first appeared in telephony, and much of our knowledge of causes and effects of delay in broadcasting has been learned from experience in this area. It is useful therefore to know a little about the causes and effects of delay in telephony and strategies that have been developed to manage the problem. Synchronization of sound and picture is also not a new problem, and experience in cinema has contributed much of our knowledge in this area.

When telephone systems first appeared, each telephone line required a separate pair of wires, not only to the local exchange, but also between exchanges. The cost of wire pairs was thus a large part of the total infrastructure cost. The use of an audio "hybrid" transformer (essentially a bridge system for combining and separating balanced circuits) allowed a single pair of wires (a 2-wire system) to carry signals in both directions, halving the transmission costs over a 4-wire system.

---

[1] The length of a signal path may be considerably longer than the direct or geodesic distance between two points. The difference between signal path length and geodesic distance is particularly large when geostationary satellites are involved.

The combining and separating functions in this system worked imperfectly as they required a good impedance match with the line at the sending end and a well-terminated line at the receiving end. The electrical characteristics of wire transmission lines vary with frequency, distance and atmospheric conditions (early telephones, like the telegraph, used aerial lines) and early hybrids could only be matched for one fixed line characteristic. Leakage through the hybrid at the handset, known as sidetone, is therefore unavoidable, both from local mismatch and from line echo. The amplitude and phase of the sidetone is not uniform with frequency, and the sound reaching the earpiece from a mildly mismatched hybrid is neither pleasant nor confidence-inspiring. Allowing a small amount of local broadband leakage to swamp the non-uniform matching leakage gives a more uniform response of the talker's voice at the earpiece and also gives the talker some confidence that the apparatus is working correctly, at least at the near end. It is a long-established practice therefore to deliberately allow some leakage between the halves of an audio loop in 2-wire systems.

The echo from the far end of a mismatched line was a relatively minor concern in early telephone systems, as line lengths were limited to about 30 km. When amplification became available and long-distance telephony became possible, the echo problem increased, as not only were there more points at which the line could be mismatched[2], but the delay of the echo became significantly longer, increasing its audibility. Echo cancelling became a major concern for telephone companies and continues to be important in maintaining the quality of long-distance lines.

With the advent of multiplexing and digitization, modern telephone systems are effectively 4-wire systems between exchanges and in many cases[3], this form is maintained right up to the handset. Eventually, it is likely that digitization and 4-wire circuits (or their optical equivalent) will extend to all handsets. Line echo will then be a problem only on loudspeaking telephones.

In broadcasting, the use of 2-wire communications systems has been limited to order wires (intercoms) and phone-in talkback programmes. Outside broadcasts have generally used separate lines for foldback and for programme contribution. Echo is relatively easy to control in these situations[4] and it is delay rather than echo that is the main problem for system designers.

5.3 System design and configuration factors versus operating practices

Delay and echo can be due to system design, as in telephone systems, due to system configuration, such as the Group-of-Pictures (GoP) setting in an MPEG video encoder, or due to poor operating or technical practice, such as foldback loudspeakers or headphones being allowed to leak into open microphones. Adequate system design, suitable system configuration and good operating practice are all needed to control delay and echo in loops.

---

[2] A long-distance analogue telephone line requires repeaters about every 30 km. Each of these points has the potential for line mismatch and resulting echo.

[3] For example, digital PABX systems linked digitally to the local exchange.

[4] This is no guarantee however that it is controlled.

5.4 System design and configuration factors which cause delay and echo

There are two principal causes of delay in audio/video systems: encode/decode processes and signal propagation.

5.4.1 Encode/decode processes

Encode/decode processes in digital systems consist of two stages. The first stage of encoding is analogue-to-digital conversion and the second stage is data compression. In decoding, the process is reversed: first the signal is decompressed, then it is converted from digital to analogue.

Analogue to digital conversion and digital to analogue conversion are relatively quick processes – they are generally performed in the interval between samples. For broadcast audio this interval is around 1/48 kHz, or around 21 µs. For telephone audio this interval is around 1/8 kHz, or around 125 µs. For standard definition broadcast video it is around 1/13 MHz or around 77 ns.

Data compression is used to reduce bandwidth needs, and is almost universally used in broadcasting for transmission over long distances (between buildings, localities, cities or countries). Audio and video in the broadcast chain are usually transmitted in compressed digital form during programme production, exchange, distribution and emission. Data compression is also almost universally used for recording, although some of the compression standards for recording differ from those for transmission.

The delays due to data compression are somewhat longer than those for analogue-to-digital and digital-to-analogue conversion. To achieve high compression factors, redundancy between video frames is reduced by comparing frames to a key frame and encoding only the differences. Video encoding therefore is based on a group of pictures rather than frame by frame. The frame comparison process at the encode stage can use blocks of up to 15 frames with subgroups of up to four frames. A picture group size of 15 pictures results in a minimum delay of four frames. Total encoding delay including processing time may be considerably longer than the minimum number of buffer frames and is usually at least 7 or 8 frames for this configuration. Similar delays are incurred at the decode stage.

The delay due to data compression of digitized audio is of a similar order, although the compression techniques are somewhat different from those used for video.

5.4.2 Propagation delay

Propagation delay can be divided into two factors: distance delay and switching delay or latency.

Propagation delays due to distance are of the order of 5 µs/km of path length in a waveguide, cable or optical fibre and of the order of 3.3 µs/km for radiated signals. In both cases, the path length may be similar to the direct point-to-point (geodesic) distance or it may be considerably longer, as with satellite transmission. For long distances, this delay may be comparable with encode/decode delays. A signal travelling terrestrially to the opposite side of the earth has a distance delay of around 100 ms. By satellite the distance delay for the same destination is around 280 ms, assuming only a single geostationary satellite is involved in the relay (based on an altitude of 36 000 km overhead or 42 000 km from the horizon).

Switching delays are generally not perceptually significant in analogue carrier systems or in synchronous digital systems. Packet-switched digital systems such as asynchronous transmission mode (ATM) incur delays at each switching node as each packet is held and its header is read before retransmitting on the next available space on a suitable line. There is delay therefore from the reading/path allocating process and from queuing. This delay is known as latency. Because the transmission path is allocated dynamically for each packet in packet-switched systems, there is also some variability in transmission delay, known as jitter. Packet-switching delays are small however (~ a few ms) compared with distance delay for long distances. Packet-switching delays are also fairly consistent for a constant path: most of the jitter generally comes from variation in path length.

5.5 Magnitude of delay in various transmission media

There are two principal causes of delay: transmission delay and encode/decode delay. Transmission delay varies with the mode of transmission, depending on whether the transmission is terrestrial or via satellite and continuous or packetized.

– Transmission delay

  – Satellite path: for geostationary satellites at an altitude of about 36 000 km above the surface, such as broadcast satellites, the path delay due to distance varies from about 239 ms when the satellite is directly over the source and receiver to about 281 ms when the satellite is at the horizon for both source and receiver.

  – For low-orbit satellites such as those used in the Iridium mobile telephone system, the altitude is around 780 km from the surface, and the delay is proportionally less. As low-orbit satellites are not geostationary, the delay from these satellites varies with their position. Allowing for an 1 800 km radius of coverage for each satellite (there are 66 satellites in the Iridium system) the maximum distance to the satellite is 2 000 km, giving a one-way propagation delay of about 6.6 ms. Additional delays due to inter-satellite links and on-board processing would increase the overall delay in the Iridium system.

  – Analogue and ISDN: transmission delay on both terrestrial analogue and terrestrial ISDN carrier systems is around 5 µs/km, based on an optical or copper waveguide propagation speed of around 2/3 speed of light. This gives up to 100 ms delay for a trip to the opposite side of the globe.

  – ATM: as well as the distance factor of 5 µs/km, ATM systems have a switching latency due to the method of routing packetized data. Each node has a latency of up to 150 µs and there may be up to 40 nodes in a long transmission path. In practice, latency is usually closer to 20-40 µs per node.

5.6 Magnitude of delay in video codecs

Broadcast video encoding almost universally uses MPEG compression for long-distance transmission. The video encoding delay in MPEG depends on the GoP parameter which on most codecs can be set between 1 and 15 frames. The video encoding delay depends not on the overall length of the group of pictures but on the length of the subgroup containing the first frame (the I-frame) and the first P-frame. The GoP is typically set to 15 to maximize the data compression.

This gives an initial subgroup length of 4 frames, with a minimum encoding delay of 133 ms at 30 fps or 160 ms for 25 fps and typical encoding delay of approximately double these figures. Shorter GoP settings can reduce the minimum delay down to 33 ms and 40 ms respectively, at the cost of less data compression (higher bit rate) or poorer picture quality at constant bit rate.

5.7 Magnitude of delay in audio codecs

Audio codecs vary widely in their characteristics, depending on the intended application. In broadcasting, fixed and mobile telephones are frequently used for programme contribution, especially for news and talkback programmes. Internet audio streaming is also becoming an increasingly common part of broadcasting services but is not yet widely used for programme contribution. High-quality programme audio may also be encoded for transmission with a broadcast-quality codec.

Typical transcoding delays for a range of systems are:

– GSM mobile phone [5] – recommended transcoding delay: minimum 20 ms and maximum 30 ms.

– CDMA mobile phone [10] – recommended maximum encode delay 20 ms; recommended maximum decode delay: 3 ms.

– Satellite phone systems – there are a number of competing systems including ICO, Iridium, AceS, AMSC-TMI and Inmarsat. These systems can be classed as low-Earth orbit (LEO) (ICO, Globalstar, Iridium and Teledesic),) or geostationary Earth orbit (GEO) (Inmarsat, Satphone, ASC, Thuraya, APMT, EAST). A typical GEO system, Inmarsat Mini-M, uses advanced multi-band excitation (AMBE) encoding at 4.8 kbit/s, with encode/decode delays comparable with those of GSM and CDMA terrestrial mobile telephone systems. In the GEO case however, the distance delay tends to make the encode/decode delays insignificant.

– Studio MPEG codec.

Audio for broadcast may be embedded with video using MPEG-2 encoding, or it may be sent on separate ISDN/ATM lines. If embedded with video, the audio will be synchronized to the video. If sent separately, the audio may be uncompressed, giving a negligible encode/decode delay, or it may be compressed. Compression algorithms vary between manufacturers, so minimum and typical delays are cited.

Due to the nature of audio signals a minimum bit rate for a given audio quality would deviate by a factor of 10 or more over time. In general time-slots with a large demand on local bit rate are very short and surrounded by time-slots with very low bit rate. Most state of the art audio coding schemes therefore do some averaging of the bit rate over time to provide a constant bit rate. Depending whether the buffer for this averaging is in the encoder or decoder side, the encoding or the decoding delay is larger.

There are several options for MPEG-1 audio compression: Layer I, Layer II and Layer III. Layer III is also known as MP3. MPEG-2 in addition to Layer I, II and III, offers advanced audio coding (AAC). In MPEG-4 AAC was chosen as the baseline codec for natural audio and extended by several new tools and functionalities. In the context of round trip delay the low delay (LD) version of AAC coding is the most prominent extension.

Minimum delays for MPEG-1 Layer I, II, III, from [6], [18]

Layer	Target bit rate (kbit/s)	Compression ratio	Theoretical minimum delay (ms)
AAC-LD	64	12:1	20
Layer I	192	4:1	19
Layer II	128	6:1	35
Layer II	64	12:1	59

Coding	Bit rate (kbit/s)	Sampling rate (kHz)	Typical delay (ms)
AAC-LD stereo	128	48	60
AAC-LD mono	64	48	50
AAC stereo	128	48	172
Layer III stereo	128	48	326
Layer II stereo	128	48	224
Layer II stereo	128	24 (half mode)	398
G.722	64	48	10

End-to-end delay (including ISDN channel) for a commercial available audio codec [18].

Codec	Bit rate (kbit/s)	Algorithmic delay with bit reservoir set to zero (ms)	100% workload, burst transmission (ms)	100% workload, continuous transmission (ms)	30% workload, burst transmission (ms)	30% workload, continuous transmission (ms)
Layer 2	192	34	Not available	Not available	Not available	Not available
Layer 3	128	54	118	142	107	131
MPEG-4 AAC	96	55	82	211	63	192
MPEG-4 HE AAC	56	129	184	361	145	322
MPEG-4 AAC-LD	128	20	33	44	24	35

End-to-end delay of current audio coding schemes.

MPEG audio encoded with MPEG video will generally have the same delay as the video.

5.8 Circuit symmetry

It should not be automatically assumed that the two half-loops in a programme contribution loop will be identical in processing delay, path delay, overall delay or leakage. It is common practice in broadcasting to use a high-quality circuit on the outside broadcast programme line only, with a lower quality circuit on the foldback line. It is also common practice to use audio-only foldback, even on video contribution loops. Both of these factors will affect the overall loop delay and the perceptual disturbance from delay and echo.

5.9 Operating practices which cause echo

While some delay is unavoidable in an audio circuit, echo is generally avoidable. Echo can be caused both by system design and by poor operating practices. Although good operating practices will not automatically ensure an echo-free loop, they will minimize the echo within a given system design.

The effectiveness of good operating practice in controlling echo depends to a large extent on the inherent echo in a loop. In a 2-wire circuit, the inherent echo is determined by system design. In a 4-wire circuit, echo caused by audio leakage from foldback lines to programme lines can be minimized by:

– using a “mix-minus” or “clean feed” to foldback, and/or
– using appropriate muting of foldback when microphones are open, either by system interlock if a manual talk button is used, or by using a voice-operated mute circuit (a “ducker”), and/or
– by using closed headphone monitoring or in-ear monitoring to minimize leakage into open microphones.

5.10 Sound to picture delay

Delay between sound and picture has also been well studied and standards [4] have been in place for some time. The main perceptual problem with this type of delay is loss of lip-sync on speech. The ITU standard currently specifies an acceptable delay between sound and picture of +25 ms (sound leads picture) to –100 ms (picture leads sound) as measured at the final programme source selection element (usually the video master control room).

Resynchronizing separate sound and picture with varying delay is conceptually a simple task, although the equipment can be expensive. If the sound leads the picture, a simple inexpensive audio delay unit may be employed to correct the sync. If picture leads sound, the picture must be delayed: this equipment is generally more expensive, although such a facility may be available in an existing video server.

6 Effects of delay and echo

6.1 Audio delay and video delay

Video communications loops have existed for some years in videoconferencing and in broadcasting but are a relatively new technology compared with audio communications loops. For this reason, the subjective effect of long delay in a video communications loop (> 1 frame) is not as well-studied as the subjective effect of long delay in an audio loop. Because of this there is a relative scarcity of information on the subjective effects of delay in video loops compared with that for audio loops. This contribution is therefore based mainly on information available on the subjective effect of delay and echo in an audio loop. Until more detailed information becomes available on subjective effects of delay in video loops, audio delay and echo is assumed to be the primary cause of disturbance, regardless of whether it is the limiting factor in overall audio-video loop delay.

6.2 Variables affecting disturbance

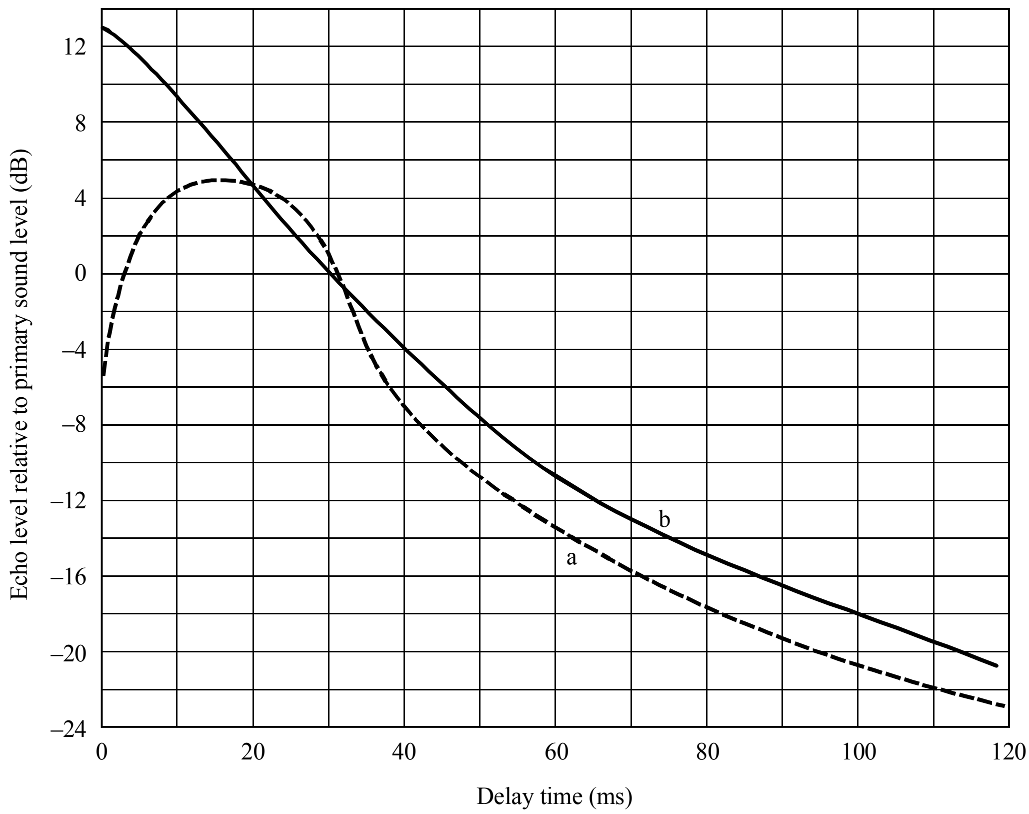
Audio delays have a number of disruptive effects on speech communications systems. The effect of the delay depends primarily on two factors: the length of the delay and the return loss in the loop [1, 3].

6.3 Echoes and fusion

For delays below the threshold in Fig. 1, the human hearing system cannot distinguish between the two arrivals on most types of speech and music. The two sounds therefore appear to "fuse" and become one louder sound. This has a positive effect on both perceived loudness and on speech intelligibility. The fusion effect becomes progressively weaker as delay increases, particularly for delays over 30 ms. Where fusion does not occur, the echo becomes disturbing.

FIGURE 1

Audibility of delayed signal in the presence of direct signal, as a function of delay and relative level (from [16])



Comparison of threshold of perception curves.

Curves a Meyer and Schodder (primary sound level 55 Phons)
b Lochner and Burger (primary sound level 50 Phons)

Rap 2044-01

6.4 Types of disturbance and their effects

There are principally three types of disruptive effect from delay and echo:

– *Long delay without significant return echo.* This type of delay can be classified as a half-loop delay. This can cause difficulty in normal conversation, causing parties to talk over each other inadvertently and to break off sentences once they realize this has happened. This causes the conversation to be prolonged and broken and it is also distracting. Delays in satellite circuits (~ 240-280 ms for each satellite link) are common causes of this type of disturbance.

– *Echoes returning to the speaker.* This type of delay can be classified as an integral multiple (1, 2, …*n*) of a full loop delay. These interfere with the normal speech feedback mechanism from brain to mouth to ear to brain, causing stuttering and hesitation. Both the length of the delay and the level of the echo have a significant effect on the disturbance, the disturbance increasing with delay at a constant level and decreasing with level at a constant delay. For long delays (> 240 ms) return losses as low as –50 dB can still cause some disturbance.

– *Echoes returning to the listener.* This can be classified as a 1 1/2, 2 1/2, …etc. loop delay. These disturb the listener by reducing speech intelligibility. As with full loop delay, the disturbance increases with delay length. It is a common problem in public address systems where it is known as "slap echo" for single reflections and "flutter echo" for multiple reflections. The audibility of this type of delay is shown in Fig. 1, from [16]. Echoes should be below this curve to be inaudible as separate sounds.

The effects of echo delay and echo level on speech intelligibility and speech difficulty have been extensively studied by telecommunications authorities [1], [3], [14] and are relatively well understood. The effects on speech intelligibility are probably better understood however than the effects on speech difficulty.

6.5 Adaptation

A learned or acquired tolerance to the effects of a stimulus or irritant is known as adaptation. Although some adaptation to echo-affected signals can be acquired, the amount of adaptation which is possible is generally quite limited. It can also take some time to adapt. The amount of adaptation which can be learned and the time taken to learn it is likely to vary considerably between individuals.

Since much broadcast content, especially on live shows, is provided by guests rather than presenters, guidelines should be formulated on the assumption that no adaptation will occur in the available time.

7 Summary

Looped communications systems are commonly encountered in broadcast production where multiple locations or remote locations are used, particularly in live programme production. The loop may be used to connect a studio interviewer and a remote guest or it may be used for foldback from the studio to a presenter at a remote location. Ideally, each separate stream of audio and video should only travel in one direction on this loop. In practice, it is quite common for audio to leak back around the loop to its source, particularly where telephones are part of the circuit. Because of distance and/or the need for encode/decode processes to compress the digitized signal for transmission and restore it again on reception, the signal can be delayed significantly.

Delay and leakage combine to create echo. Both the delay and the echo can cause disturbance. This disturbance adversely affects both the programme participants and the programme audience. There are principally three types of disruptive effect from delay and echo:

– long delay without significant return echo;

– echoes returning to the speaker;

– echoes returning to the listener.

The effect of echo on speech intelligibility is relatively well understood. The effect of delay on the speaker and on the listener is less well understood. Guidelines are therefore suggested here based on audibility and effect on speech intelligibility.

Annoyance from sound delay and echo is familiar in large auditoriums and sporting grounds and is hardly a new problem. In electronic communications, the problem first appeared in 2-wire telephony, and much of our knowledge of causes and effects of delay in broadcasting has been learned from experience in this area. It is difficult to control leakage between the halves of an audio loop in 2-wire systems, and it is a long-established practice to deliberately allow some controlled leakage in these systems to make the uncontrollable leakage less noticeable.

There are two principal causes of delay: propagation delay and encode/decode delay. Propagation delay can be divided into two factors: distance delay and switching delay or latency.

Propagation delays due to distance are of the order of 5 μs/km of path length in a waveguide, cable or optical fibre and of the order of 3.3 μs/km for radiated signals. A signal travelling terrestrially to the opposite side of the Earth has a distance delay of around 100 ms. By satellite the distance delay for the same destination is around 280 ms, assuming only a single geostationary satellite is involved in the relay.

The need for audio-video synchronization may make the video delay the limiting factor in an audio-video communication loop. Video encoding is based on a group of pictures rather than individual frames. A GoP size of 15 pictures is the largest commonly used size and gives the highest data compression. This results in a minimum delay of four frames. Total encoding delay including processing time may be considerably longer than the minimum number of buffer frames and is usually at least 7 or 8 frames (230-320 ms) for this configuration.

Audio transcoding delays in most mobile phone systems such as GSM, CDMA and satellite phone systems are low (3-30 ms). For higher quality broadcast audio encoding such as MPEG AAC they may be of the same order as the video encoding delays (170-400 ms). More recent systems such as MPEG AAC-LD have reduced this to around 60 ms.

ATM systems have a switching latency due to the method of routing packetized data. Each node has a latency of up to 150 μs and there may be up to 40 nodes in a long transmission path. In practice, latency is usually closer to 20-40 μs per node resulting in a relatively small delay of around 1.6 ms.

Echo can be minimized both by good system design and by good operating practices.

Three general principles which should be observed in designing and configuring facilities for programme contribution loops are:

– minimizing the number of encode/decode stages;

– keeping differential audio-video delay within the limit of +25 ms to –100 ms;

– keeping the foldback leakage limit below the audibility threshold.

Operating practices which will minimize echo include:

– using a "mix-minus" or "clean feed" to foldback; and/or

– using appropriate muting of foldback when microphones are open, either by system interlock if a manual talk button is used, or by using a voice-operated mute circuit (a "ducker"); and/or

– by using closed headphone monitoring or in-ear monitoring to minimize leakage into open microphones.