

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

	<i>Page</i>
1 Scope	2
2 References and Bibliography	2
3 Terms and definitions	3
4 Abbreviations and acronyms	3
5 Causes of delay and echo	4
5.1 Context	4
5.2 Historical background	4
5.3 System design and configuration factors versus operating practices	5
5.4 System design and configuration factors which cause delay and echo	6
5.4.1 Encode/decode processes	6
5.4.2 Propagation delay	6
5.5 Magnitude of delay in various transmission media	7
5.6 Magnitude of delay in video codecs	7
5.7 Magnitude of delay in audio codecs	8
5.8 Circuit symmetry	10
5.9 Operating practices which cause echo	10
5.10 Sound to picture delay	10
6 Effects of delay and echo	10
6.1 Audio delay and video delay	10
6.2 Variables affecting disturbance	11
6.3 Echoes and fusion	11
6.4 Types of disturbance and their effects	12
6.5 Adaptation	12
7 Summary	12

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 destination¹. 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 mismatched², 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 cases³, 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 situations⁴ 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.