International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

J.248

(06/2008)

SERIES J: CABLE NETWORKS AND TRANSMISSION OF TELEVISION, SOUND PROGRAMME AND OTHER MULTIMEDIA SIGNALS

Measurement of the quality of service

Requirements for operational monitoring of video-to-audio delay in the distribution of television programs

Recommendation ITU-T J.248

Recommendation ITU-T J.248

Requirements for operational monitoring of video-to-audio delay in the distribution of television programs

Summary

Since the advent of digital television networks for program transmission, and the introduction of high-efficiency bit-rate reduction (BRR) devices and of other types of digital image processing devices, audiences sometimes complain that the television programs they receive are out of "lip-sync". Lip-sync errors are generally due to the fact that audio and video are separately processed in the television chain, and processing delays are typically different for video than for the accompanying audio signal.

Recommendation ITU-T J.248 analyses the problem and provides guidance on means to measure lip-sync errors in the context of operational monitoring in television programme transmission chains.

Source

Recommendation ITU-T J.248 was approved on 13 June 2008 by ITU-T Study Group 9 (2005-2008) under Recommendation ITU-T A.8 procedure.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at http://www.itu.int/ITU-T/ipr/.

© ITU 2009

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without the prior written permission of ITU.

CONTENTS

Page

Recommendation ITU-T J.248

Requirements for operational monitoring of video-to-audio delay in the distribution of television programs

1 Scope

This Recommendation specifies requirements for operational monitoring aimed to help minimizing video-to-audio delay, thus minimizing lip-sync errors in the transmission of television programs.

NOTE – The structure and content of this Recommendation have been organized for ease of use by those familiar with the original source material specifications; as such, the usual style of ITU-T recommendations has not been applied.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

[ITU-T J.243]	Recommendation ITU-T J.243 (2006), *Requirements for operational monitoring in television programme transmission chains*.
[ITU-R BT.1359]	Recommendation ITU-R BT.1359-1 (1998), *Relative timing of sound and vision for broadcasting*.
[ITU-R BT.1729]	Recommendation ITU-R BT.1729 (2005), *Common 16 x 9/4 x 3 aspect ratio digital television reference test pattern*.

3 Definitions

3.1 Terms defined elsewhere

None.

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 final edited master: The final edited master is the final instance of a television program as it is provided at the end of the program production chain, ready to be dispatched to the distributors and the end users.

3.2.2 frame synchronizer: A device that receives a video signal from a remote source, and synchronizes it to the local video synchronization pulses, in order that it may be seamlessly mixed with locally generated video signals.

3.2.3 interframe coding: Bit rate reduction video signal encoding that exploits the video signal redundancy over several pictures.

3.2.4 lip synchronization (lip-sync): Operation to provide the feeling that the speaking motion of the displayed person is synchronized with that person's voice, or other sounds are synchronized to their visually displayed source. Alternatively, the minimization of the relative delay between the visual display of a person speaking and the audio of the voice of the person speaking. The objective is to achieve a natural relationship between the visual image and the aural message for the viewer/listener.

3.2.5 primary distribution: Use of a transmission channel for transferring audio and/or video information to one or several destination points without a view to further post-processing on reception (e.g., from a continuity studio to a transmitter network).

3.2.6 secondary distribution: Use of a transmission channel for distribution of programs to viewers at large.

3.2.7 source coding (bit-rate reduction): The encoding of the original digital signal (video, audio or data) in bit-rate reduction (BRR) representation before protection is applied against bit errors in the channel.

4 Abbreviations and acronyms

None.

5 Conventions

None.

6 The reference television chain

For the purpose of this Recommendation, the reference television chain, from acquisition to presentation can be described as consisting of the four sections, listed below.

6.1 Television program production

This is the section of the television chain, in which program material is captured locally or acquired from remote sources. The production section starts from the camera and microphone, and it ends where the complete program material is presented on the finished edited master, ready to be dispatched to distributors and end users. Except for the simplest programs, the audio and video components of all television programs are acquired and processed in separate production chains, and they are generally only brought together on the finished edited master of the program, under the supervision of the program director and producer.

6.2 Primary distribution

This is the section of the television chain, in which programs are sent from the program provider to the program distributor (e.g., the input of the cable head-end). The program audio and video may need to be separately processed in this section, e.g., to be mixed with local program material or commentary.

6.3 Secondary distribution

This is the section of the television chain, in which programs are sent from the primary distribution point (e.g., the output of the cable head-end) to the end user of the programs (e.g., the cable television subscriber).

6.4 Presentation

This is the section of the television chain, in which the audio and video of program are presented on the audiovisual display of the end user.

7 Main causes for loss of lip-sync

As long as audio and video signals are multiplexed in a single bit stream, they preserve the lip-sync they had at the input of the multiplexer.

However, lip-sync may be lost whenever the audio and video signals are separately processed. In this event, the amount of lip-sync error depends on the type of processing and on the number of cascaded processes.

The reference television chain contains a large number of devices that the video or respectively the audio signals must go through. Every device along the chain introduces some delay in the audio or the video signal that goes through it. Audio delays introduced along the chain are generally small enough not to appreciably affect lip-sync. Video delays are also generally small, except for some specific devices. The devices that introduce the most significant video delays, often large enough to visibly affect lip-sync are frame synchronizers, bit-rate-reduction video encoders and decoders, and complex image processing devices found in production and in presentation, such as image correctors, interlacers/de-interlacers etc., which may be built into consumer displays[1].

a) Frame synchronizers are devices that receive a video signal from a remote source and synchronize it to the local video synchronization pulses, in order that it may be seamlessly mixed with locally generated video signals. They are found in program production, but also in primary and secondary distribution. The video delay that they introduce is inherently variable, since it depends on the phase of remote sync with respect to the local sync, and it may be the order of one video picture.

b) Source {bit-rate-reduction} encoders and decoders can introduce very large but essentially fixed video delays, whose amount depends on the GoP (temporal interpolation) mode to which they are set to operate, and on the computation delay in the encoder, which can be quite high in advanced encoders such as MPEG-4 ones. The delay introduced in the encoder is then cumulated with the one introduced in the decoder. The total delay can be quite large, of the order of many television images, and in extreme case of the order of some seconds.

c) Complex image processing devices are present both in production and in presentation. They can introduce appreciable video delays depending on the operating mode to which they are set.

d) In some cases, the video signal travels on a path different from the audio signal. In these cases, a delay can be introduced, due to the different transit time of the video and the audio signals over their respective transmission path. This is the case, for instance, of a sports program in which the video is sent via satellite and the comment, mixed with the international sound, is sent via a land line.

For the purpose of this Recommendation, which addresses the primary and secondary distribution of programs, it is assumed that lip-sync is perfect on the final edited master at the end of television program production, and any lip-sync error is introduced downstream from it, namely in the contribution section of the chain, in the distribution section of the chain and in program

[1] Lip-sync errors can also be introduced by some program production tools.

As an example, a video-wall behind the anchor-man of a news bulleting will generally introduce some delay in the displayed video signal, due the image processing circuits it contains. If the video-wall displays the image of an interviewee that is present in the news studio, the delay between the interviewee's direct image and the one on the video-wall may be objectionable.

As another example, radiocameras are often used in live broadcasts. The radiocamera signal generally has to go through a frame synchronizer and perhaps a color corrector, thus suffering some delay. When the radiocamera signal is mixed with a signal coming directly from other cameras that shoot the same event, the lip-sync error may be noticeable.

presentation, i.e. in the consumer display. This is a credible assumption for most programs, notably recorded ones, since the creative staff can be expected to certify that the final edited master is correct before the program is released.

8 User requirements for operational monitoring

The following requirements are derived from [ITU-T J.243] in the light of specific issues in this Recommendation.

8.1 Operational aspect

1) Capability of in-service monitoring;
2) Applicability to the video formats in use such as SDTV and HDTV;
3) Applicability to the numbers of audio channels in use;
4) Applicability to the coding bit rates in use, irrespective of variable bit rate (VBR) or constant bit rate (CBR);
5) Applicability to the transmission bit rates in use;
6) Applicability to the coding parameters and tools (e.g., profile/level, picture structure, range of motion vectors) in use;
7) Applicability to different signal processing such as compression coding, standards conversion, aspect ratio conversion and switching between remote program and local program;
8) Applicability to different sources of degradation (e.g., compression ratio and transmission error rate);
9) Applicability to different program contents;
10) Applicability to the system configurations in use;
11) Traceability of the causes of malfunction, failure and degradation;
12) Availability of precise information for delay adjustment tool.

8.2 Measurement aspect

1) Ability to evaluate quantitatively the lip-sync;
2) Ability to perform lip-sync assessment using only bit streams (e.g., Transport Stream) if possible;
3) Ability to perform lip-sync assessment using only the signals concerned (i.e., non-reference methods) if possible;
4) Ability to detect the occurrence point of lip-sync loss;
5) Ability to perform lip-sync assessment using only baseband signals;
6) Repeatability (i.e., evaluation result should not be affected by the successive signals);
7) Ability to evaluate lip-sync in a short time or instantaneously.

Appendix I

Perceptual limits for lip-sync errors

(This appendix does not form an integral part of this Recommendation)

[ITU-R BT.1359] provides a thorough discussion of lip-sync, and it indicates the undetectable plateau, the detectability threshold and the acceptability threshold for lip-sync errors.

The threshold of detectability is about 45 ms to –125 ms (audio leading video is indicated as a positive value; audio lagging behind video is indicated as a negative value).

The threshold of acceptability is about 90 ms to –185 ms on the average, with respect to perfect lip-sync.

These thresholds are shown in Figure I.1. They can be taken to generally apply to the spoken word; narrower limits may apply to the case of impulsive sounds.

Each threshold is set to the lead or lag value at which 50% of the assessors cast the higher integer score and 50% cast the lower integer score in the 5 grades assessment scale.

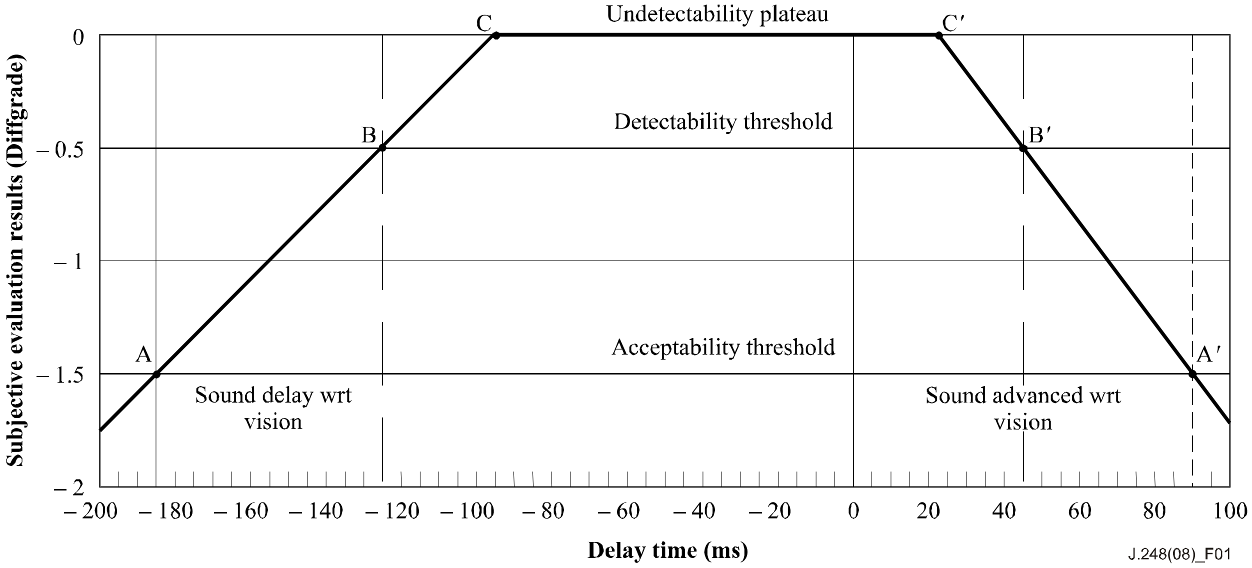

Figure I.1 – Detectability and acceptability thresholds

This Recommendation provides some insight in the causes for loss of lip-sync along the television chain, and it describes how lip-sync errors can be corrected or at least reduced to an acceptable limit, as indicated in Figure I.1.

Appendix II

Detection and adjustment of lip-sync errors

(This appendix does not form an integral part of this Recommendation)

II.1 Perceptual verification of lip-sync along television chains

It is possible to check that correct lip-sync is preserved along a television chain, through the use of test patterns, based on the principle of the clap-board used in the production of motion pictures.

For example, the test patterns specified in [ITU-R BT.1729] are suitable for such an application. This is a monoscope-type test pattern which also contains two horizontal counter-moving white bars. The instant when the two bars touch each other is marked by a short bip. This arrangement provides an audiovisual stimulus of remarkably high sensitivity.

Of course, such test patterns cannot be used on-line, during actual program transmission. They can only be used off-line or at best at station breaks which are designed to include some audiovisual stimuli similar to the ones described. Nevertheless, their use should be encouraged, whenever it is possible.

II.2 Maintaining lip-sync along the television chain – The ideal solution

The ideal solution to the problem of maintaining lip-sync along the television chain would be to add identical time-stamps to the video and audio components of the program at the end of the production section, and automatically correct any lip-sync loss by re-aligning the time-stamps whenever necessary along the chain.

This solution presents some difficulties.

First of all, only some specific, and sometimes proprietary video/audio equipment allows to add identical time-stamps to the audio and video component of a program. If such equipment is not used throughout the television distribution chain, the video and audio components of the program will not carry identical time-stamps, or some program components will not carry time-stamps at all.

Secondly, the television distribution chain is often complex, it can change configuration from moment to moment, and various parts of it are often under the responsibility of different operators. In this situation it is difficult to be sure that the original time-stamps will travel through it unscathed. If they do not, then an automatic lip-sync corrector will not work, or it will mis-work.

Thirdly, lip-sync may change when programs are switched in a program schedule or due to "zapping", if the switching is performed between programs that use different image formats or different compression parameters. In this case the lip-sync corrector at the reception point may need to adjust its audio/video delay to the appropriate value, and this requires timely detection e.g., by means of metadata or by a continuous check of time-stamps, if available.

II.3 Maintaining lip-sync along the television chain – A practical approximate solution

A practical, albeit approximate approach, is to identify those devices in the television chain that introduce large video delays, and pre-correct those delays at the input of each such device, by introducing an equivalent delay in the associated sound signal. In other terms, at every place in the television transmission chain where a lip-sync error is introduced, the error should be immediately (locally) corrected.

For instance, it is known that a bit-rate reduction encoder introduces a delay in the video signal. The amount of the delay is a function of the bit-rate reduction profile used, and it can be determined in advance.

The video bit-rate reduction decoder downstream from the encoder also introduces some video delay. This can also be determined in advance. The action suggested here is to delay the sound signal by the same amount as the sum of the delays introduced by the encoder and the decoder. This action should result in retaining lip-sync among video and audio signals that were in lip-sync at the input of the encoder.

Similarly, the video delay introduced by complex image processing in the consumer receiver varies depending on the setting of the receiver controls. The delay introduced by the setting of each control is known in advance and it can be compensated by introducing a corresponding delay in the audio signal within the consumer receiver.

Video frame synchronizer are generally only found in the television distribution chain at the hinge between primary and secondary distribution, where there is sometimes the need to switch between remotely-generated programs and locally-generated programs such as local news or commercials.

Video frame synchronizers introduce a delay in the video signal that can vary from zero to one video frame moment by moment. However, they often contain the circuitry needed to introduce the same delay in the audio signal. When this is not the case, the suggested approach is to introduce a delay in the sound signal that is equal to about half the total range of delay of the frame synchronizer. In this way, the audio will lead the video or it will lag behind it by a rather small amount, probably still within the undetectability plateau of Figure I.1.

The case described in clause 7 subitem d) above does not generally pose a problem, since it is normally possible to check and if necessary restore lip-sync during the circuit preparatory period, when the test pattern referred to in clause 7, or even a simple cinema-style clap-board can be used before the start of the contribution.

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	General tariff principles
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	**Cable networks and transmission of television, sound programme and other multimedia signals**
Series K	Protection against interference
Series L	Construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Telephone transmission quality, telephone installations, local line networks
Series Q	Switching and signalling
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks, open system communications and security
Series Y	Global information infrastructure, Internet protocol aspects and next-generation networks
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2009