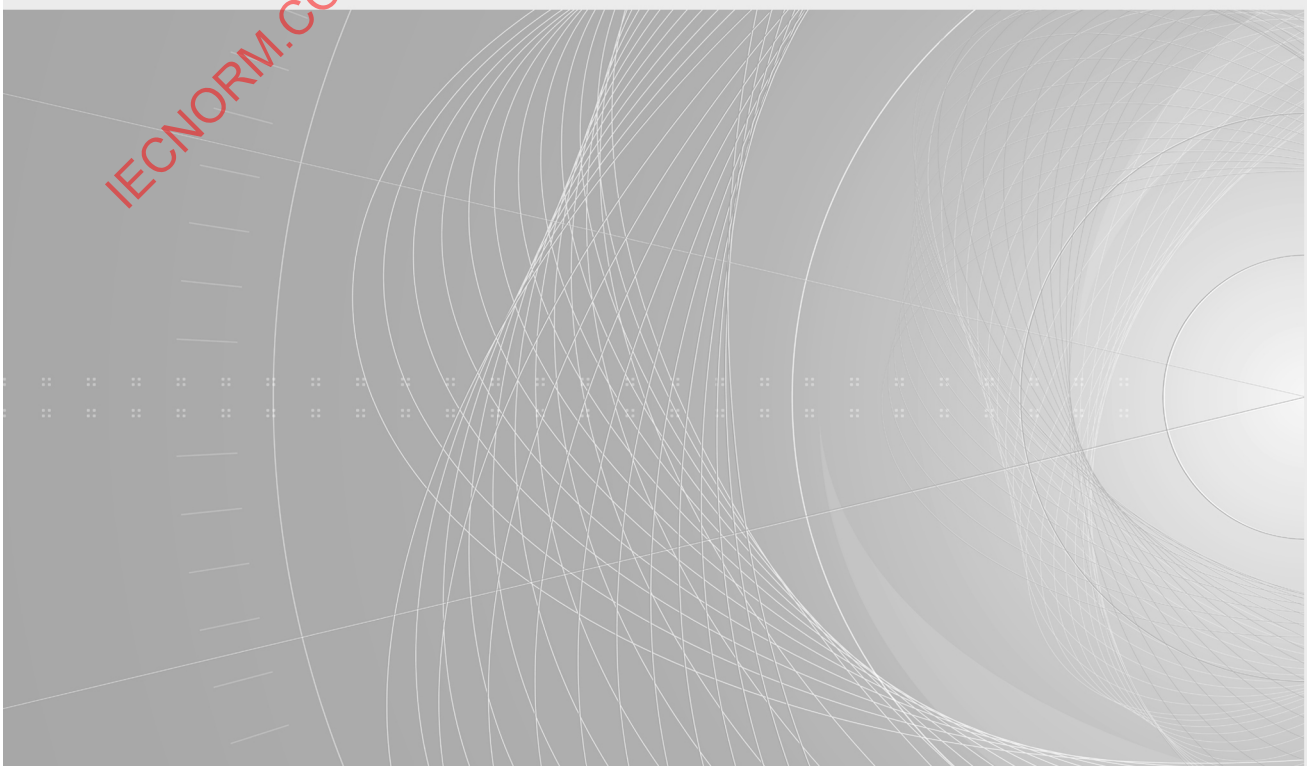
IEC TS 62312-2

Edition 2.0 2018-11
REDLINE VERSION

TECHNICAL SPECIFICATION

colour inside

Guideline for synchronization of audio and video –
Part 2: Methods for synchronization of audio and video systems



IEC TS 62312-2:2018-11 RLV(en)

IEC Central Office
3, rue de Varembé
CH-1211 Geneva 20
Switzerland

Tel.: +41 22 919 02 11
info@iec.ch
www.iec.ch

About the IEC
The International Electrotechnical Commission (IEC) is the leading global organization that prepares and publishes International Standards for all electrical, electronic and related technologies.

About IEC publications
The technical content of IEC publications is kept under constant review by the IEC. Please make sure that you have the latest edition, a corrigenda or an amendment might have been published.

IEC Catalogue - webstore.iec.ch/catalogue
The stand-alone application for consulting the entire bibliographical information on IEC International Standards, Technical Specifications, Technical Reports and other documents. Available for PC, Mac OS, Android Tablets and iPad.

IEC publications search - webstore.iec.ch/advsearchform
The advanced search enables to find IEC publications by a variety of criteria (reference number, text, technical committee,…). It also gives information on projects, replaced and withdrawn publications.

IEC Just Published - webstore.iec.ch/justpublished
Stay up to date on all new IEC publications. Just Published details all new publications released. Available online and also once a month by email.

Electropedia - www.electropedia.org
The world's leading online dictionary of electronic and electrical terms containing 21 000 terms and definitions in English and French, with equivalent terms in 16 additional languages. Also known as the International Electrotechnical Vocabulary (IEV) online.

IEC Glossary - std.iec.ch/glossary
67 000 electrotechnical terminology entries in English and French extracted from the Terms and Definitions clause of IEC publications issued since 2002. Some entries have been collected from earlier publications of IEC TC 37, 77, 86 and CISPR.

IEC Customer Service Centre - webstore.iec.ch/csc
If you wish to give us your feedback on this publication or need further assistance, please contact the Customer Service Centre: sales@iec.ch.

IEC TS 62312-2

Edition 2.0 2018-11
REDLINE VERSION

TECHNICAL SPECIFICATION

Guideline for synchronization of audio and video – Part 2: Methods for synchronization of audio and video systems

INTERNATIONAL
ELECTROTECHNICAL
COMMISSION

ICS 33.160.01

ISBN 978-2-8322-6219-1

Warning! Make sure that you obtained this publication from an authorized distributor.

CONTENTS

INTERNATIONAL ELECTROTECHNICAL COMMISSION

_____

GUIDELINE FOR SYNCHRONIZATION OF AUDIO AND VIDEO –

Part 2: Methods for synchronization of audio and video systems

FOREWORD

1) The International Electrotechnical Commission (IEC) is a worldwide organization for standardization comprising all national electrotechnical committees (IEC National Committees). The object of IEC is to promote international co-operation on all questions concerning standardization in the electrical and electronic fields. To this end and in addition to other activities, IEC publishes International Standards, Technical Specifications, Technical Reports, Publicly Available Specifications (PAS) and Guides (hereafter referred to as "IEC Publication(s)"). Their preparation is entrusted to technical committees; any IEC National Committee interested in the subject dealt with may participate in this preparatory work. International, governmental and non-governmental organizations liaising with the IEC also participate in this preparation. IEC collaborates closely with the International Organization for Standardization (ISO) in accordance with conditions determined by agreement between the two organizations.
2) The formal decisions or agreements of IEC on technical matters express, as nearly as possible, an international consensus of opinion on the relevant subjects since each technical committee has representation from all interested IEC National Committees.
3) IEC Publications have the form of recommendations for international use and are accepted by IEC National Committees in that sense. While all reasonable efforts are made to ensure that the technical content of IEC Publications is accurate, IEC cannot be held responsible for the way in which they are used or for any misinterpretation by any end user.
4) In order to promote international uniformity, IEC National Committees undertake to apply IEC Publications transparently to the maximum extent possible in their national and regional publications. Any divergence between any IEC Publication and the corresponding national or regional publication shall be clearly indicated in the latter.
5) IEC itself does not provide any attestation of conformity. Independent certification bodies provide conformity assessment services and, in some areas, access to IEC marks of conformity. IEC is not responsible for any services carried out by independent certification bodies.
6) All users should ensure that they have the latest edition of this publication.
7) No liability shall attach to IEC or its directors, employees, servants or agents including individual experts and members of its technical committees and IEC National Committees for any personal injury, property damage or other damage of any nature whatsoever, whether direct or indirect, or for costs (including legal fees) and expenses arising out of the publication, use of, or reliance upon, this IEC Publication or any other IEC Publications.
8) Attention is drawn to the Normative references cited in this publication. Use of the referenced publications is indispensable for the correct application of this publication.
9) Attention is drawn to the possibility that some of the elements of this IEC Publication may be the subject of patent rights. IEC shall not be held responsible for identifying any or all such patent rights.

This redline version of the official IEC Standard allows the user to identify the changes made to the previous edition. A vertical bar appears in the margin wherever a change has been made. Additions are in green text, deletions are in strikethrough red text.

The main task of IEC technical committees is to prepare International Standards. In exceptional circumstances, a technical committee may propose the publication of a Technical Specification when

- the required support cannot be obtained for the publication of an International Standard, despite repeated efforts, or
- the subject is still under technical development or where, for any other reason, there is the future but no immediate possibility of an agreement on an International Standard.

Technical Specifications are subject to review within three years of publication to decide whether they can be transformed into International Standards.

IEC 62312-2, which is a technical specification, has been prepared by technical area 11: Quality for audio, video and multimedia systems, of IEC technical committee 100: Audio, video and multimedia systems and equipment.

This second edition cancels and replaces the first edition published in 2007. This edition constitutes a technical revision.

This edition includes the following significant technical changes with respect to the previous edition:

a) fingerprint is newly introduced;
b) addition of the synchronization information from fingerprint (SMPTE spec.);
c) addition of the method for using the above information.

The text of this Technical Specification is based on the following documents:

Draft TS	Report on voting
100/3049/DTS	100/3106/RVDTS

Full information on the voting for the approval of this Technical Specification can be found in the report on voting indicated in the above table.

This document has been drafted in accordance with the ISO/IEC Directives, Part 2.

The list of all the parts of IEC 62312, published under the general title *Guideline for synchronization of audio and video*, can be found on the IEC website.

The committee has decided that the contents of this publication will remain unchanged until the stability date indicated on the IEC website under "http://webstore.iec.ch" in the data related to the specific publication. At this date, the publication will be

- transformed into an International standard,
- reconfirmed,
- withdrawn,
- replaced by a revised edition, or
- amended.

IMPORTANT – The "colour inside" logo on the cover page of this publication indicates that it contains colours which are considered to be useful for the correct understanding of its contents. Users should therefore print this publication using a colour printer.

INTRODUCTION

Audio, video and multimedia systems and equipment have begun to use digital technologies. Digital systems or equipment may cause delay of audio and video signals because digital signal processing causes latency and delay. The unevenness of the delay between audio and video ~~may~~ can cause synchronization problems.

For instance, a digital broadcasting system uses signal compression of audio and video. A receiver has signal decoders, and this ~~may~~ can cause unevenness of the delay between audio and video. Digital video signal processing of the display causes a significant delay in the reproduction time of the video image. Another example is an audio-video system consisting of a digital media player, an audio amplifier and a display. A digital media player outputs audio and video signals separately to the amplifier and display through digital interfaces. This may cause synchronization problems of audio and video when ~~each~~ the processing time of ~~the~~ each piece of equipment is different.

To solve synchronization problems of audio and video reproduction on the user side, this document gives guidelines for general methods for the synchronization of audio and video.

GUIDELINE FOR SYNCHRONIZATION OF AUDIO AND VIDEO –

Part 2: Methods for synchronization of audio and video systems

1 Scope

The IEC 62312 series gives guidelines for methods of synchronization of audio and video.

This part of IEC 62312 describes the system model and general methods for the synchronization of audio and video. The methods exclude the synchronization of the signal source and the spatial delay of audio reproduction.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

IEC 60958-1, *Digital audio interface – Part 1: General*

IEC 60958-3:2006, *Digital audio interface – Part 3: Consumer applications*
IEC 60958-3:2006/AMD2:2015

IEC 61883-6, *Consumer audio/video equipment – Digital interface – Part 6: Audio and music data transmission protocol*

SMPTE 12M, *Television, Audio and Film – Time and Control Code*

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

- IEC Electropedia: available at http://www.electropedia.org/
- ISO Online browsing platform: available at http://www.iso.org/obp

3.1
latency
inevitable delay of the signal that is caused by its principle or unavoidable signal processing

Note 1 to entry: Signal encoding and decoding, signal compression and de-compression, and signal transmission through interfaces cause latency.

3.2
delay
general meaning of delay other than latency, including delay that is caused by functional signal processing and delay that is set intentionally

Note 1 to entry: The term "delay" is sometimes used as to meaning latency. In this technical specification, "delay" is defined as delay other than latency. Functional signal processing causes delay or delay is set intentionally.

3.3
latency information
information of the sum value of latency and delay

3.4
time ~~stamp~~ address
quantized timing in which an event occurs on the basis of a reference clock

4 System model

4.1 Audio and video system

An audio and video system consists of audio and video devices and these devices ~~may~~ can include digital signal processing that causes latency and delay. Each device is connected to the other devices by analogue or digital signal interfaces. The final outputs from the audio and video system to the user are reproduction of audio with speakers and a visual image with a display device. These reproductions have no information of synchronization and the method for synchronization should be applied at the stage prior to the final reproduction.

To control synchronization of audio and video, synchronization information should be provided, and the controller should control the delay of the audio and video devices with the synchronization information. The synchronization information and the controller signal are transmitted through digital interfaces or a control line.

~~**4.2 Latency and delay**~~

~~Latency means an inevitable delay of signal in the audio and video device that is caused by principle or unavoidable signal processing. For instance, signal encoding and decoding, signal compression and de-compression, and signal transmission through interface cause latency.~~

~~The term "delay" is sometimes used as meaning latency. In this technical specification, "delay" is defined as delay other than latency. Functional signal processing causes delay or delay is set intentionally.~~

4.2 System model

Figure 1 shows the system model.

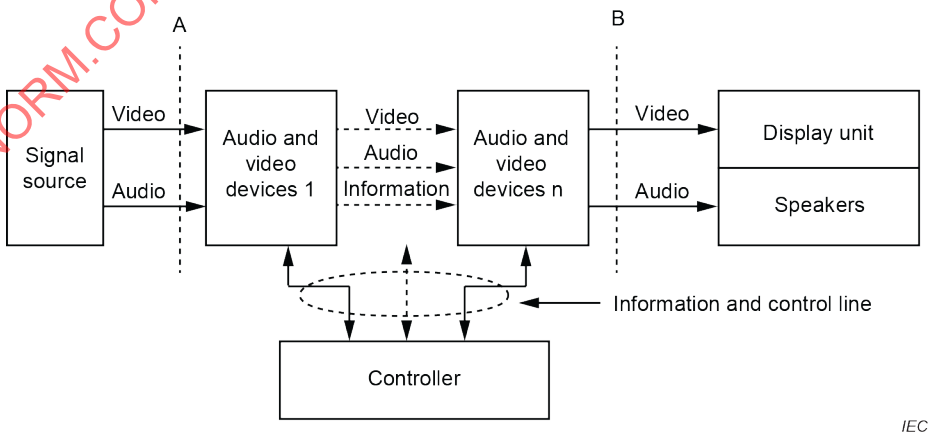


Figure 1 – System model

Audio and video synchronization of the signal source is out of the scope of this system model because this model cannot identify it or control it. The audio and video devices have latencies and delays because of their structure. The system consists of a number of audio and video

devices. The display unit is a visual reproducer, such as a CRT or a flat panel display device. The speakers are audio reproducers. These display units and speakers are defined as an ideal device that has no latency. A spatial delay of audio is outside the scope. The lines A and B in Figure 1 show the border of the control of synchronization; the controller controls latency and delay in the audio and video devices with the synchronization information.

4.3 Audio and video device

The audio and video device is defined as a device consisting of audio and video units. Each unit includes latency and delay.

Figure 2 shows a model for an audio and video device. This audio and video device consists of units that have latencies and delays and signal inputs and outputs.

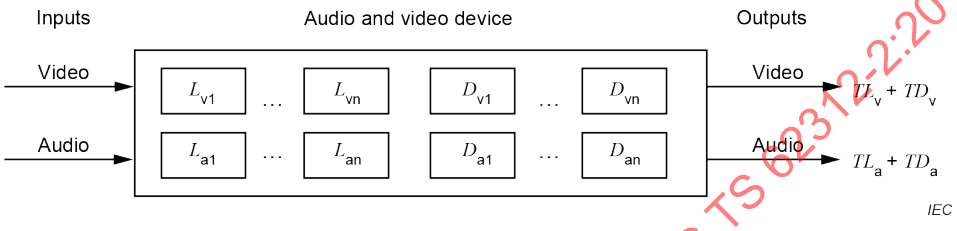


Figure 2 – Audio and video device

The latency of the unit is defined as L_{vn} for video and L_{an} for audio, and the delay of the unit is defined as D_{vn} for video and D_{an} for audio

where

n is a numeric number of units;
L_{vn} is the video latency of the video unit number n;
D_{vn} is the video delay of the video unit number n;
TL_v is the total latency of video;
TD_v is the total delay of video;
L_{an} is the audio latency of the audio unit number n;
D_{an} is the audio delay of the audio unit number n;
TL_a is the total latency of audio;
TD_a is the total delay of audio.

The total value of latency and delay is the sum of each value.

$$TL_v = L_{v1} + L_{v2} + \ldots\ldots + L_{vn}$$
$$TD_v = D_{v1} + D_{v2} + \ldots\ldots + D_{vn}$$
$$TL_a = L_{a1} + L_{a2} + \ldots\ldots + L_{an}$$
$$TD_a = D_{a1} + D_{a2} + \ldots\ldots + D_{an}$$

To control synchronization of audio and video, the audio and video devices should have the information and control function that processes the synchronization information and control delay.

The information and control function of the audio and video device is described in Figure 3. This is a general model that is applied for both the audio part and the video part independently.

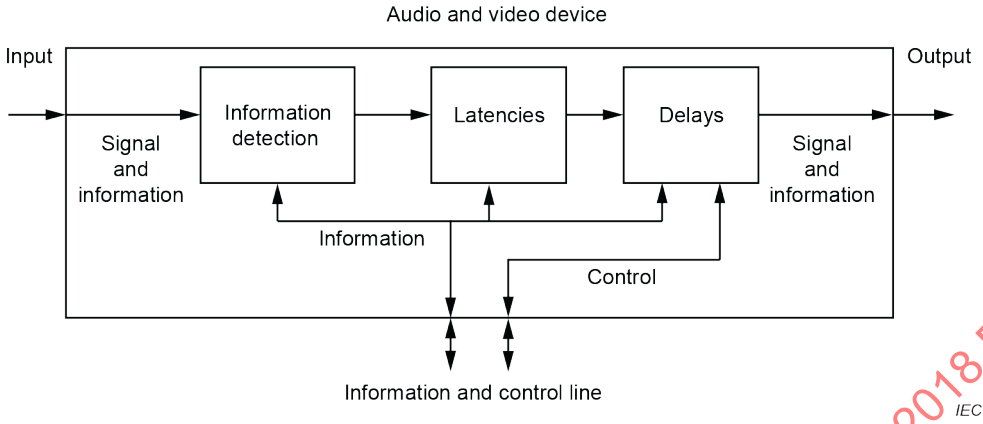


Figure 3 – Information and control in audio and video device

The audio and video signal and the synchronization information from the previous audio and video device are input into the next audio and video device. The information detection unit detects the information of synchronization from the input signal, and the controller accesses the information through the information line.

Signal processing causes delays and latencies, which are TD_v and TL_v for video, and TD_a and TL_a for audio. The controller accesses the information of latencies and delays through the information line.

To control the delay of the audio and video signals of the audio and video device, at least one controllable delay unit is required. The controller controls the controllable delay unit through the control line.

4.4 Controller

The controller controls the delay of both audio and video or either audio or video signals with the controllable delay unit. The controller calculates the value of delay that is necessary to synchronize the audio and video signals. The method is described in Clause 5.

The controller may be included in the audio and video device, as described in 4.6.

4.5 Source device and display device

Figure 4 shows the source device. The source device is defined as the device that consists of the signal source and the audio and video device. The latency and delay of the signal source is outside the scope of this device. The latencies and delays of the source device are defined as being the same as those of the audio and video device. For instance, a DVD player consists of the signal source and the audio and video device. In a DVD player, the source signal is the content~~s~~ of the DVD ~~disc~~, and it is assumed to have synchronized audio and video data. A digital TV tuner consists of the signal source and the audio and video device, and the signal source is a broadcast~~ing~~ content that is assumed to have synchronized audio and video data. The signal is carried by an RF signal or data stream with an encoded format, which causes latencies at the decoding process.

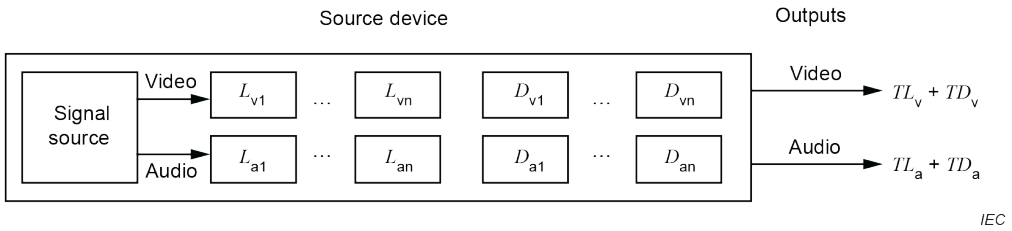


Figure 4 – Source device

Figure 5 shows the display device, such as a video monitor. The display device consists of an audio and video device and a display unit and speakers. The latencies and delays of the source device are defined as being the same as those of the audio and video device. The display device may have only a video part without an audio part.

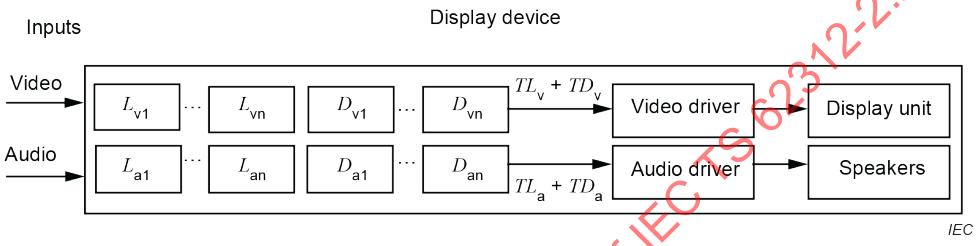


Figure 5 – Display device

Figure 6 shows the audio amplifier. This consists of only an audio part and ~~drive~~ speakers.

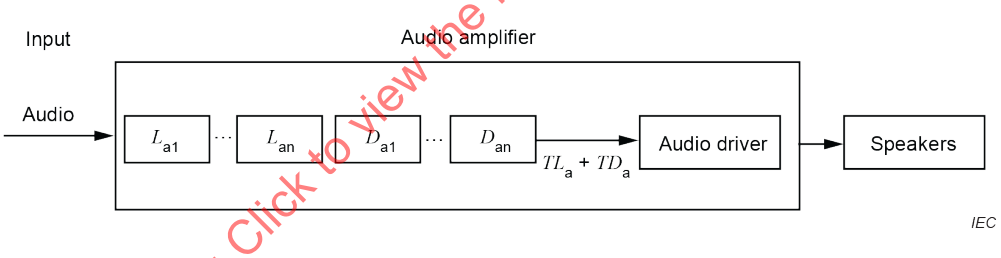


Figure 6 – Audio amplifier

4.6 Controller function of audio and video device

The controller can be equipped as a function of the audio and video device. This case is described in Figure 7. The information, or synchronization information, and control are transmitted between devices. The controller is located in the audio and video device that controls the synchronization.

The information and control line is provided by the digital interface or the network. For instance, IEC 60958 is a digital audio interface, which carries digital audio data and information including synchronization information, and IEEE 1394 is a network interface that carries similar information.

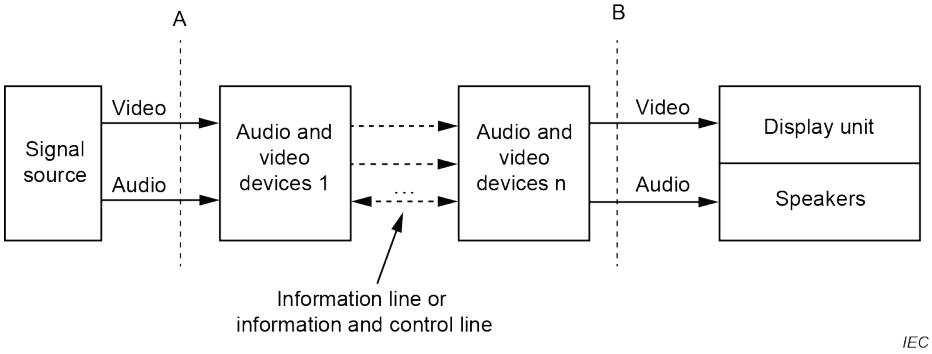


Figure 7 – Controller function of audio and video device

5 Methods for synchronization

5.1 Information for synchronization

5.1.1 Time-code information

Time-code information is given to the original signal of audio and video when they are recorded as a recording time ~~stamp~~ address. Through the process of editing, post-production and authoring, the final time code information is given to the signal source of the source device in order to ascertain the time when the audio and video signals should be reproduced. This time-code information is used to synchronize audio and video signals in the audio and video system. The source device reproduces the time-code information from the signal source and provides this information to the audio and video device with the information line.

The SMPTE time code is used for the original recording time ~~stamp~~ address and may also be used for the signal source of the source device. In addition, the secondary time code is used for synchronization. This time code is produced from the original time code, such as the SMPTE time code. Further details of time-code information are explained in Annex A.

5.1.2 Latency information

"Latency information" is defined as meaning the information of the sum value of latency and delay. The source device, the audio and video device, the display device or the audio amplifier causes latency and delay. Latency information is used to know the time difference between the audio and video signals in each device. The audio and video device provides this information to the other audio and video devices or the controller with the information line. The value of latency information is defined in 4.3 as TL_a, TD_a and TL_v, TD_v. Further details of latency information are explained in Annex A.

5.1.3 Synchronization information

Synchronization information shows the timing when the audio and video signals are captured as a snapshot at a point where the audio and video are known to be synchronized. This information is generated from audio and video signal natures at periods in the flow of audio and video signals, and the information is embedded in or carried with the audio and video signals. The audio, video and synchronization information is transmitted through a media and, at the reproduction side, the relative position of video and audio frames with matching fingerprints are analysed to obtain delay or synchronization information.

One method of synchronization information is the fingerprint information. Fingerprints calculated using a simple algorithm are made from each audio and video signal. Video fingerprints and audio fingerprints are generated periodically at the same point in the system and they are transmitted with the audio and video signals in an established data format which depends upon the transport mechanism. At the reproduction of these audio and video signals,

these fingerprints are analysed to obtain video delay, audio delay and synchronization information. Further details of fingerprint information are explained in Annex A.

5.2 Methods for synchronization

5.2.1 General

To synchronize audio and video signals, an additional delay is added to either the audio or the video signal ahead of another signal of audio or video as compensation for the difference.

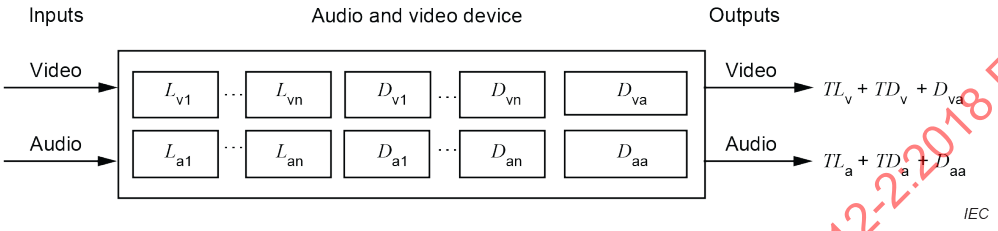


Figure 8 – Additional delay

Figure 8 shows the audio and video device with the additional delay unit for audio and video. This additional delay unit makes the additional delay D_{va} and D_{aa}.

D_{va} is the additional delay for video by the additional delay unit.

D_{aa} is the additional delay for audio by the additional delay unit.

The same model is applied for the source device, the display device and the audio amplifier. In the audio and video system, the overall value of latency and delay is the sum of all latency and delay of the audio and video device and the additional delay for audio and video. To synchronize audio and video, each value should be equal to:

$$\sum(TL_a + TD_a + D_{aa}) = \sum(TL_v + TD_v + Dv_a)$$

Generally, one additional delay to each audio and video is adequate, because one additional delay can compensate for the difference in the audio and video system that consists of serial connection of the audio and video devices. In this case, the condition of overall delay and latency is:

$$(\sum(TL_a + TD_a))+ D_{aa} = (\sum(TL_v + TD_v)) + D_{va}$$

One additional delay of either audio or video can achieve synchronization. In this case, at least one additional delay unit is required in any device of the audio and video system. This is shown as:

$$(\sum(TL_a + TD_a)) + D_{aa} = (\sum(TL_v + TD_v))$$

$$(\sum(TL_a + TD_a)) = (\sum(TL_v + TD_v)) + D_{av}$$

In general, the video signal tends to delay more than the audio signal because of its signal processing in the audio and video device. In this case, the additional delay is added to the audio signal.

Figure 9 shows the case of multiple reproductions. If the audio and video system has multiple reproduction devices in multiple branches, the additional delay should be added to each branch.

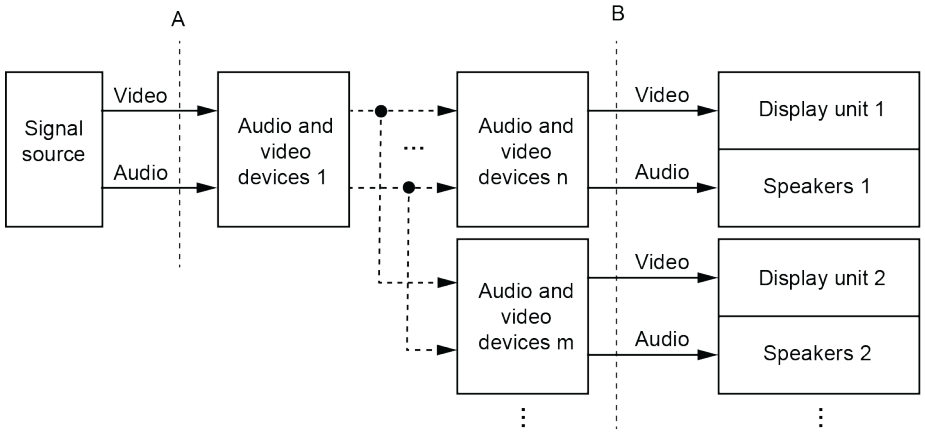


Figure 9 – Multiple reproductions

To calculate the value of the additional delay, the synchronization information is used. There are two types of synchronization information: one is time-code and the other is latency information.

The controller gathers the synchronization information from the source device, the audio and video device, the display device and the audio amplifier, and calculates the necessary value of the additional delay. Then the controller controls the delay value of the additional delay unit to synchronize audio and video signals. The timing of the execution of this process depends on the audio and video system. When some event in the audio and video system changes the overall value of the latency and delay, the controller may execute this control process.

It depends on the audio and video system which device's synchronization information should be used and which device is adequate for the addition of the additional delay.

5.2.2 Method with time-code information

To calculate the value of the additional delay, the difference in the time code of the audio and the video signal is used. The source device reproduces the audio and video time code from the signal source. This information is transferred through the information line to the audio and video devices, the display device and the audio amplifier.

In each device, the audio and video time code will differ if that device has different delay and latency in each audio and video unit.

There are two types of audio and video devices. One is the device that transfers the time-code information to the next device. The other is the device that has the additional delay unit. Generally, the former type is the source device that is located in the middle of the audio and video system. The latter type is the display device and the audio amplifier. However, the type of device depends on the audio and video system.

In the case of the source device and the middle-located device, this does not reproduce sound and image but outputs the audio and video signal to the next audio and video device. The middle device may not have the compensation method and it should transfer the time code to the next audio and video device.

In the case of the final reproduction device, such as the display device and the audio amplifier, the additional delay controls the presentation time of audio and video to be synchronized.

5.2.3 Method with latency information

The latency information is used to ascertain how much time is the difference between the audio and the video signal in the audio and video device. To reproduce audio and video in the same presentation time, the controller should know each value of the latency information of the audio and video devices through the information line. In one audio and video device, the latency information of audio and video is:

$$TL_{a} + TD_{a} \quad \text{and} \quad TL_{v} + TD_{v}$$

In the system having multiple audio and video devices, the overall value of the latency is the sum of all the latency information of the audio and video devices as:

$$\sum(TL_{a} + TD_{a}) \quad \text{and} \quad \sum(TL_{v} + TD_{v})$$

Each audio and video device has its latency information, and the controller asks for the latency information of each audio and video device. The controller calculates the overall value of the latency information of audio and video devices. The value of the additional delays, D_{aa} and D_{va}, is calculated from the value of that overall value of latency information.

5.2.4 Method with synchronization information

Synchronization information is typically generated at the time when the original content is made. This information shows the exact timing of audio and video frames at the time of audio and video signals are generated, and this information is provided periodically. After transmission and signal processing through a media, this information allows the video and audio frames to be rematched. The reproduction system detects the timing of audio and video signals at that reproduction time; from that, the value of delay between audio and video signal is obtained.

5.3 Method with IEC 60958-3

Subclause 6.3 of IEC 60958-3:2006/AMD2:2015 specifies information for synchronization in user data. These are SMPTE time-code information and latency information. Annex T of IEC 60958-3:2006 describes application cases of synchronization with the SMPTE time-code and latency information.

IEC 60958-3 is a mono-directional interface and carries audio data and latency information and has no functionality of control. The latency information is only transmitted to the next audio and video device. To calculate the overall value of the latency information, each device adds its latency information to the latency information that is detected from the input signal. This bucket brigade gives the sum of the latency information from the source device to the current audio and video device. This information is input to the next audio and video device.

5.4 Method with IEC 61883-6

IEC 61883-6 is based on the 1394 TA specification. This is one of the application specifications of IEEE 1394. As IEEE 1394 is a bi-directional interface and has a packet transmission structure, various protocols are transmitted at the same time in packets including IEC 61883-6. 1394 TA-specified application specifications include the AV/C command that is used for controlling systems and devices.

To synchronize audio and video with time-code information, the SMPTE Time Code and Sample Count Transmission Protocol Ver.1.0 ~~(document number 1999024)~~ specifies the transmission of the SMPTE time code in packets. This specification may be used with IEC 61883-6.

To synchronize audio and video with latency information, the AV/C Command for AV Synchronization Version 1.0 ~~(document number 2005005)~~ specifies a method for getting latency information and controlling latency and delay. This specification may be used with IEC 61883-6.

In addition, the IEC 60958-conformant data format is defined in IEC 61883-6, and this format carries the IEC 60958 frame format as it is through IEC 61883-6. The IEC 60958 frame format is defined in IEC 60958-1. IEC 61883-6 carries IEC 60958 data and its SMPTE time-code and latency information.

The IEC 60958-conformant data format can be applied to the other interface specification in the same way as IEC 61883-6 applies.

5.5 Method with other interface specifications

For other interface and control specifications, the system model and the general method for synchronization described in this document can be applied.

Annex A
(informative)

Information for synchronization

A.1 Time-code information

SMPTE time code is a common time-code information. Other time-code information is employed for each system, for instance MPEG employs various kinds of time-code information for their various MPEG specifications.

This information addresses a time address of an audio and video signal that should be presented; the information is carried with audio and video signals and the information is addressed at the content production stage.

A.2 Latency information

Latency is caused by signal processing in each device; latency information is recognized by the audio and video device.

The information is generated by the signal processing stage, the information is carried from the signal processing stage to the other audio and video devices.

A.3 Synchronization information

This information is different from time-code information. Synchronization information is synthesized from characteristics of the audio and video signals and embedded in or carried with the signals, so this information is captured at an originally synchronized state of audio and video. SMPTE ST-2064 defines the synchronization information called "fingerprint" and the transport mechanisms for it.

Bibliography

IEEE 1394:1995, *Standard for a High Performance Serial Bus*

1394TA 2001024, *Audio and Music Data Transmission Protocol 2.1*

1394TA 1999024, *SMPTE Time Code and Sample Count Transmission Protocol Ver.1.0*

1394TA 2005005, *AV/C Command for AV Synchronization Version 1.0*

~~SMPTE 12M, *Television, Audio and Film – Time and Control Code*~~

ST 2064-1:2015, *Audio to Video Synchronization Measurement – Fingerprint Generation*

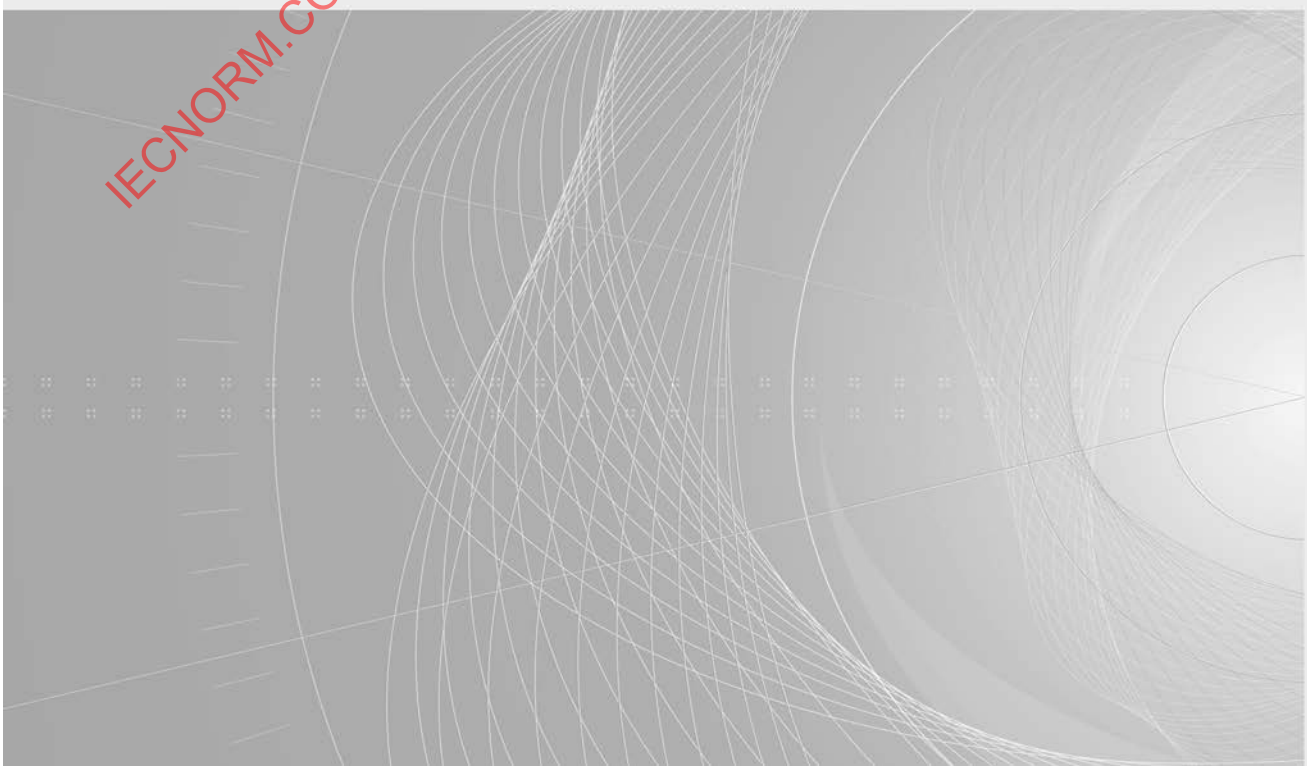
ST 2064-2:2015, *Audio to Video Synchronization Measurement – Fingerprint Transport*

IEC TS 62312-2

Edition 2.0 2018-11

TECHNICAL SPECIFICATION

Guideline for synchronization of audio and video –
Part 2: Methods for synchronization of audio and video systems



IEC TS 62312-2:2018-11(en)

CONTENTS

INTERNATIONAL ELECTROTECHNICAL COMMISSION

_____

GUIDELINE FOR SYNCHRONIZATION OF AUDIO AND VIDEO –

Part 2: Methods for synchronization of audio and video systems

FOREWORD

1) The International Electrotechnical Commission (IEC) is a worldwide organization for standardization comprising all national electrotechnical committees (IEC National Committees). The object of IEC is to promote international co-operation on all questions concerning standardization in the electrical and electronic fields. To this end and in addition to other activities, IEC publishes International Standards, Technical Specifications, Technical Reports, Publicly Available Specifications (PAS) and Guides (hereafter referred to as "IEC Publication(s)"). Their preparation is entrusted to technical committees; any IEC National Committee interested in the subject dealt with may participate in this preparatory work. International, governmental and non-governmental organizations liaising with the IEC also participate in this preparation. IEC collaborates closely with the International Organization for Standardization (ISO) in accordance with conditions determined by agreement between the two organizations.
2) The formal decisions or agreements of IEC on technical matters express, as nearly as possible, an international consensus of opinion on the relevant subjects since each technical committee has representation from all interested IEC National Committees.
3) IEC Publications have the form of recommendations for international use and are accepted by IEC National Committees in that sense. While all reasonable efforts are made to ensure that the technical content of IEC Publications is accurate, IEC cannot be held responsible for the way in which they are used or for any misinterpretation by any end user.
4) In order to promote international uniformity, IEC National Committees undertake to apply IEC Publications transparently to the maximum extent possible in their national and regional publications. Any divergence between any IEC Publication and the corresponding national or regional publication shall be clearly indicated in the latter.
5) IEC itself does not provide any attestation of conformity. Independent certification bodies provide conformity assessment services and, in some areas, access to IEC marks of conformity. IEC is not responsible for any services carried out by independent certification bodies.
6) All users should ensure that they have the latest edition of this publication.
7) No liability shall attach to IEC or its directors, employees, servants or agents including individual experts and members of its technical committees and IEC National Committees for any personal injury, property damage or other damage of any nature whatsoever, whether direct or indirect, or for costs (including legal fees) and expenses arising out of the publication, use of, or reliance upon, this IEC Publication or any other IEC Publications.
8) Attention is drawn to the Normative references cited in this publication. Use of the referenced publications is indispensable for the correct application of this publication.
9) Attention is drawn to the possibility that some of the elements of this IEC Publication may be the subject of patent rights. IEC shall not be held responsible for identifying any or all such patent rights.

The main task of IEC technical committees is to prepare International Standards. In exceptional circumstances, a technical committee may propose the publication of a Technical Specification when

- the required support cannot be obtained for the publication of an International Standard, despite repeated efforts, or
- the subject is still under technical development or where, for any other reason, there is the future but no immediate possibility of an agreement on an International Standard.

Technical Specifications are subject to review within three years of publication to decide whether they can be transformed into International Standards.

IEC 62312-2, which is a technical specification, has been prepared by technical area 11: Quality for audio, video and multimedia systems, of IEC technical committee 100: Audio, video and multimedia systems and equipment.

This second edition cancels and replaces the first edition published in 2007. This edition constitutes a technical revision.

This edition includes the following significant technical changes with respect to the previous edition:

a) fingerprint is newly introduced;
b) addition of the synchronization information from fingerprint (SMPTE spec.);
c) addition of the method for using the above information.

The text of this Technical Specification is based on the following documents:

Draft TS	Report on voting
100/3049/DTS	100/3106/RVDTS

Full information on the voting for the approval of this Technical Specification can be found in the report on voting indicated in the above table.

This document has been drafted in accordance with the ISO/IEC Directives, Part 2.

The list of all the parts of IEC 62312, published under the general title *Guideline for synchronization of audio and video*, can be found on the IEC website.

The committee has decided that the contents of this publication will remain unchanged until the stability date indicated on the IEC website under "http://webstore.iec.ch" in the data related to the specific publication. At this date, the publication will be

- transformed into an International standard,
- reconfirmed,
- withdrawn,
- replaced by a revised edition, or
- amended.

A bilingual version of this publication may be issued at a later date.

INTRODUCTION

Audio, video and multimedia systems and equipment have begun to use digital technologies. Digital systems or equipment may cause delay of audio and video signals because digital signal processing causes latency and delay. The unevenness of the delay between audio and video can cause synchronization problems.

For instance, a digital broadcasting system uses signal compression of audio and video. A receiver has signal decoders, and this can cause unevenness of the delay between audio and video. Digital video signal processing of the display causes a significant delay in the reproduction time of the video image. Another example is an audio-video system consisting of a digital media player, an audio amplifier and a display. A digital media player outputs audio and video signals separately to the amplifier and display through digital interfaces. This may cause synchronization problems of audio and video when the processing time of each piece of equipment is different.

To solve synchronization problems of audio and video reproduction on the user side, this document gives guidelines for general methods for the synchronization of audio and video.

GUIDELINE FOR SYNCHRONIZATION OF AUDIO AND VIDEO –

Part 2: Methods for synchronization of audio and video systems

1 Scope

The IEC 62312 series gives guidelines for methods of synchronization of audio and video.

This part of IEC 62312 describes the system model and general methods for the synchronization of audio and video. The methods exclude the synchronization of the signal source and the spatial delay of audio reproduction.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

IEC 60958-1, *Digital audio interface – Part 1: General*

IEC 60958-3:2006, *Digital audio interface – Part 3: Consumer applications*
IEC 60958-3:2006/AMD2:2015

IEC 61883-6, *Consumer audio/video equipment – Digital interface – Part 6: Audio and music data transmission protocol*

SMPTE 12M, *Television, Audio and Film – Time and Control Code*

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

- IEC Electropedia: available at http://www.electropedia.org/
- ISO Online browsing platform: available at http://www.iso.org/obp

3.1
latency
inevitable delay of the signal that is caused by its principle or unavoidable signal processing

Note 1 to entry: Signal encoding and decoding, signal compression and de-compression, and signal transmission through interfaces cause latency.

3.2
delay
general meaning of delay other than latency, including delay that is caused by functional signal processing and delay that is set intentionally

Note 1 to entry: The term "delay" is sometimes used as to meaning latency. In this technical specification, "delay" is defined as delay other than latency. Functional signal processing causes delay or delay is set intentionally.

3.3
latency information
information of the sum value of latency and delay

3.4
time address
quantized timing in which an event occurs on the basis of a reference clock

4 System model

4.1 Audio and video system

An audio and video system consists of audio and video devices and these devices can include digital signal processing that causes latency and delay. Each device is connected to the other devices by analogue or digital signal interfaces. The final outputs from the audio and video system to the user are reproduction of audio with speakers and a visual image with a display device. These reproductions have no information of synchronization and the method for synchronization should be applied at the stage prior to the final reproduction.

To control synchronization of audio and video, synchronization information should be provided, and the controller should control the delay of the audio and video devices with the synchronization information. The synchronization information and the controller signal are transmitted through digital interfaces or a control line.

4.2 System model

Figure 1 shows the system model.

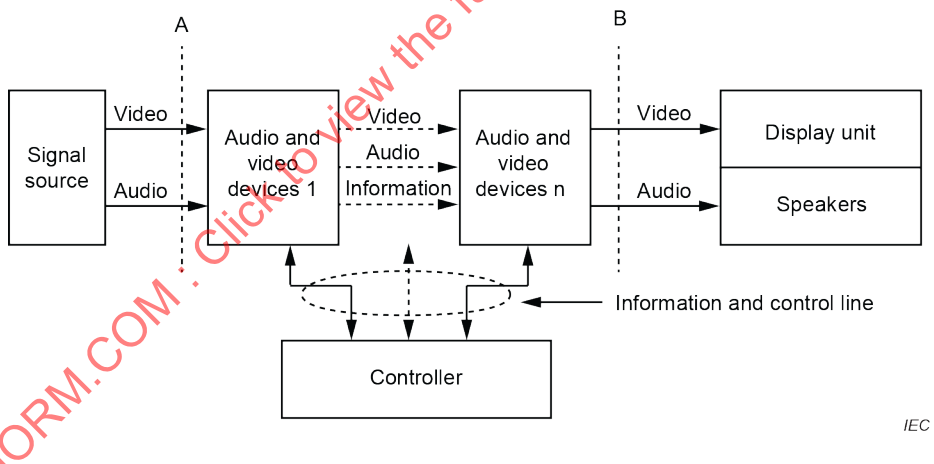


Figure 1 – System model

Audio and video synchronization of the signal source is out of the scope of this system model because this model cannot identify it or control it. The audio and video devices have latencies and delays because of their structure. The system consists of a number of audio and video devices. The display unit is a visual reproducer, such as a CRT or a flat panel display device. The speakers are audio reproducers. These display units and speakers are defined as an ideal device that has no latency. A spatial delay of audio is outside the scope. The lines A and B in Figure 1 show the border of the control of synchronization; the controller controls latency and delay in the audio and video devices with the synchronization information.

4.3 Audio and video device

The audio and video device is defined as a device consisting of audio and video units. Each unit includes latency and delay.

Figure 2 shows a model for an audio and video device. This audio and video device consists of units that have latencies and delays and signal inputs and outputs.

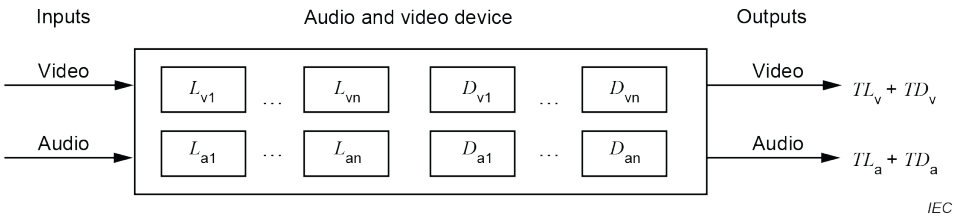


Figure 2 – Audio and video device

The latency of the unit is defined as L_{vn} for video and L_{an} for audio, and the delay of the unit is defined as D_{vn} for video and D_{an} for audio

where

n is a numeric number of units;
L_{vn} is the video latency of the video unit number n;
D_{vn} is the video delay of the video unit number n;
TL_v is the total latency of video;
TD_v is the total delay of video;
L_{an} is the audio latency of the audio unit number n;
D_{an} is the audio delay of the audio unit number n;
TL_a is the total latency of audio;
TD_a is the total delay of audio.

The total value of latency and delay is the sum of each value.

$$TL_v = L_{v1} + L_{v2} + \ldots\ldots + L_{vn}$$
$$TD_v = D_{v1} + D_{v2} + \ldots\ldots + D_{vn}$$
$$TL_a = L_{a1} + L_{a2} + \ldots\ldots + L_{an}$$
$$TD_a = D_{a1} + D_{a2} + \ldots\ldots + D_{an}$$

To control synchronization of audio and video, the audio and video devices should have the information and control function that processes the synchronization information and control delay.

The information and control function of the audio and video device is described in Figure 3. This is a general model that is applied for both the audio part and the video part independently.

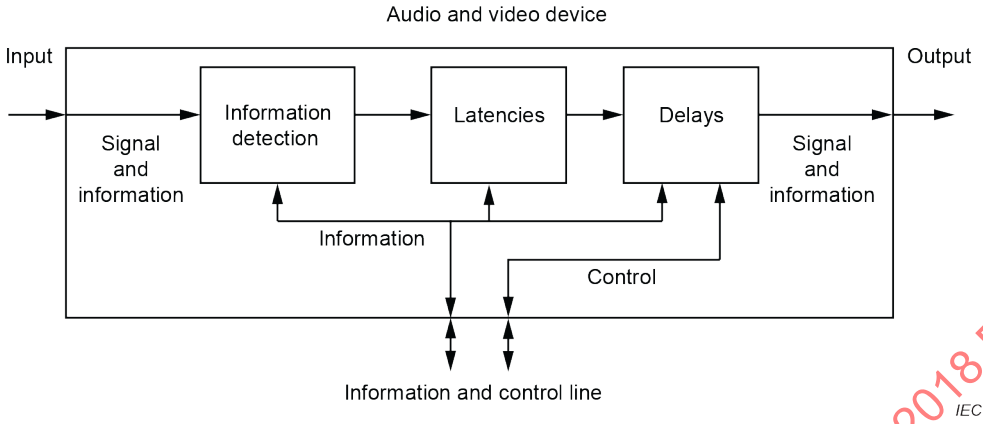


Figure 3 – Information and control in audio and video device

The audio and video signal and the synchronization information from the previous audio and video device are input into the next audio and video device. The information detection unit detects the information of synchronization from the input signal, and the controller accesses the information through the information line.

Signal processing causes delays and latencies, which are TD_v and TL_v for video, and TD_a and TL_a for audio. The controller accesses the information of latencies and delays through the information line.

To control the delay of the audio and video signals of the audio and video device, at least one controllable delay unit is required. The controller controls the controllable delay unit through the control line.

4.4 Controller

The controller controls the delay of both audio and video or either audio or video signals with the controllable delay unit. The controller calculates the value of delay that is necessary to synchronize the audio and video signals. The method is described in Clause 5.

The controller may be included in the audio and video device, as described in 4.6.

4.5 Source device and display device

Figure 4 shows the source device. The source device is defined as the device that consists of the signal source and the audio and video device. The latency and delay of the signal source is outside the scope of this device. The latencies and delays of the source device are defined as being the same as those of the audio and video device. For instance, a DVD player consists of the signal source and the audio and video device. In a DVD player, the source signal is the content of the DVD, and it is assumed to have synchronized audio and video data. A digital TV tuner consists of the signal source and the audio and video device, and the signal source is a broadcast content that is assumed to have synchronized audio and video data. The signal is carried by an RF signal or data stream with an encoded format, which causes latencies at the decoding process.

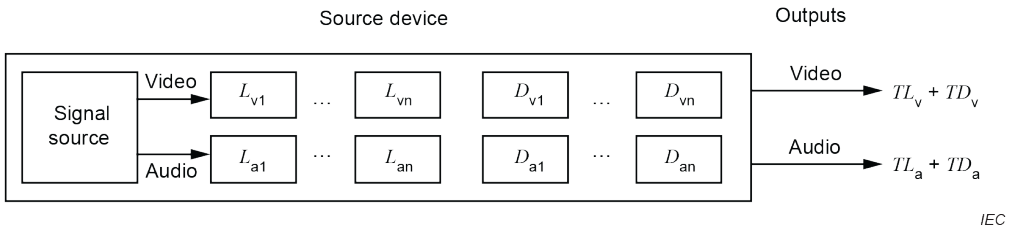


Figure 4 – Source device

Figure 5 shows the display device, such as a video monitor. The display device consists of an audio and video device and a display unit and speakers. The latencies and delays of the source device are defined as being the same as those of the audio and video device. The display device may have only a video part without an audio part.

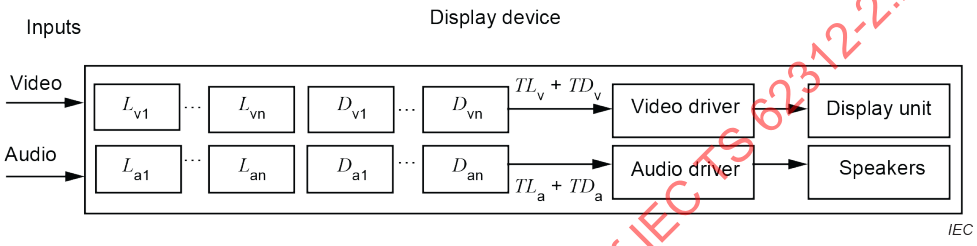


Figure 5 – Display device

Figure 6 shows the audio amplifier. This consists of only an audio part and speakers.

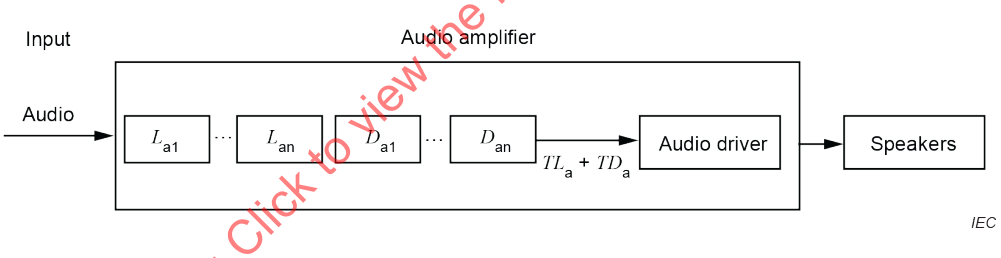


Figure 6 – Audio amplifier

4.6 Controller function of audio and video device

The controller can be equipped as a function of the audio and video device. This case is described in Figure 7. The information, or synchronization information, and control are transmitted between devices. The controller is located in the audio and video device that controls the synchronization.

The information and control line is provided by the digital interface or the network. For instance, IEC 60958 is a digital audio interface, which carries digital audio data and information including synchronization information, and IEEE 1394 is a network interface that carries similar information.

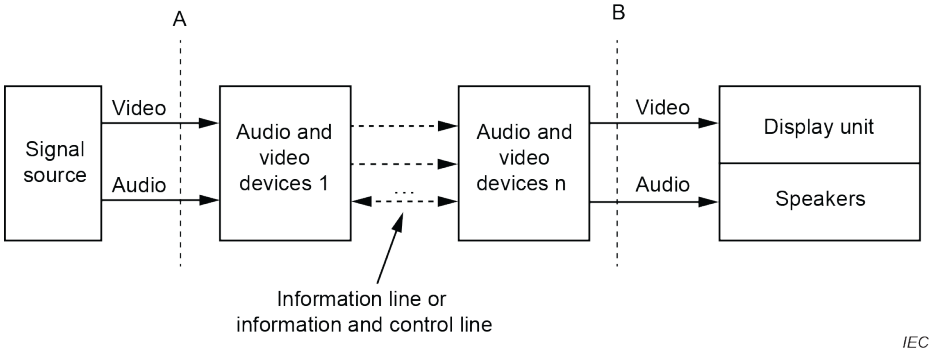


Figure 7 – Controller function of audio and video device

5 Methods for synchronization

5.1 Information for synchronization

5.1.1 Time-code information

Time-code information is given to the original signal of audio and video when they are recorded as a recording time address. Through the process of editing, post-production and authoring, the final time code information is given to the signal source of the source device in order to ascertain the time when the audio and video signals should be reproduced. This time-code information is used to synchronize audio and video signals in the audio and video system. The source device reproduces the time-code information from the signal source and provides this information to the audio and video device with the information line.

The SMPTE time code is used for the original recording time address and may also be used for the signal source of the source device. In addition, the secondary time code is used for synchronization. This time code is produced from the original time code, such as the SMPTE time code. Further details of time-code information are explained in Annex A.

5.1.2 Latency information

"Latency information" is defined as meaning the information of the sum value of latency and delay. The source device, the audio and video device, the display device or the audio amplifier causes latency and delay. Latency information is used to know the time difference between the audio and video signals in each device. The audio and video device provides this information to the other audio and video devices or the controller with the information line. The value of latency information is defined in 4.3 as TL_a, TD_a and TL_v, TD_v. Further details of latency information are explained in Annex A.

5.1.3 Synchronization information

Synchronization information shows the timing when the audio and video signals are captured as a snapshot at a point where the audio and video are known to be synchronized. This information is generated from audio and video signal natures at periods in the flow of audio and video signals, and the information is embedded in or carried with the audio and video signals. The audio, video and synchronization information is transmitted through a media and, at the reproduction side, the relative position of video and audio frames with matching fingerprints are analysed to obtain delay or synchronization information.

One method of synchronization information is the fingerprint information. Fingerprints calculated using a simple algorithm are made from each audio and video signal. Video fingerprints and audio fingerprints are generated periodically at the same point in the system and they are transmitted with the audio and video signals in an established data format which depends upon the transport mechanism. At the reproduction of these audio and video signals,

these fingerprints are analysed to obtain video delay, audio delay and synchronization information. Further details of fingerprint information are explained in Annex A.

5.2 Methods for synchronization

5.2.1 General

To synchronize audio and video signals, an additional delay is added to either the audio or the video signal ahead of another signal of audio or video as compensation for the difference.

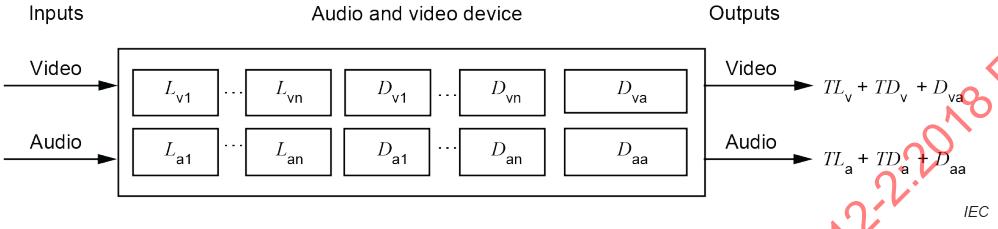


Figure 8 – Additional delay

Figure 8 shows the audio and video device with the additional delay unit for audio and video. This additional delay unit makes the additional delay D_{va} and D_{aa}.

D_{va} is the additional delay for video by the additional delay unit.

D_{aa} is the additional delay for audio by the additional delay unit.

The same model is applied for the source device, the display device and the audio amplifier. In the audio and video system, the overall value of latency and delay is the sum of all latency and delay of the audio and video device and the additional delay for audio and video. To synchronize audio and video, each value should be equal to:

$$\sum(TL_a + TD_a + D_{aa}) = \sum(TL_v + TD_v + Dv_a)$$

Generally, one additional delay to each audio and video is adequate, because one additional delay can compensate for the difference in the audio and video system that consists of serial connection of the audio and video devices. In this case, the condition of overall delay and latency is:

$$(\sum(TL_a + TD_a))+ D_{aa} = (\sum(TL_v + TD_v)) + D_{va}$$

One additional delay of either audio or video can achieve synchronization. In this case, at least one additional delay unit is required in any device of the audio and video system. This is shown as:

$$(\sum(TL_a + TD_a)) + D_{aa} = (\sum(TL_v + TD_v))$$

$$(\sum(TL_a + TD_a)) = (\sum(TL_v + TD_v)) + D_{av}$$

In general, the video signal tends to delay more than the audio signal because of its signal processing in the audio and video device. In this case, the additional delay is added to the audio signal.

Figure 9 shows the case of multiple reproductions. If the audio and video system has multiple reproduction devices in multiple branches, the additional delay should be added to each branch.