

INTERNATIONAL STANDARD

NORME INTERNATIONALE



Multimedia quality – Method of assessment of synchronization of audio and video

Qualité pour les systèmes multimédia – Méthode d'évaluation de la synchronisation de l'audio et de la vidéo

IECNORM.COM : Click to view the full PDF of IEC 62503:2008



THIS PUBLICATION IS COPYRIGHT PROTECTED

Copyright © 2008 IEC, Geneva, Switzerland

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from either IEC or IEC's member National Committee in the country of the requester. If you have any questions about IEC copyright or have an enquiry about obtaining additional rights to this publication, please contact the address below or your local IEC member National Committee for further information.

Droits de reproduction réservés. Sauf indication contraire, aucune partie de cette publication ne peut être reproduite ni utilisée sous quelque forme que ce soit et par aucun procédé, électronique ou mécanique, y compris la photocopie et les microfilms, sans l'accord écrit de l'IEC ou du Comité national de l'IEC du pays du demandeur. Si vous avez des questions sur le copyright de l'IEC ou si vous désirez obtenir des droits supplémentaires sur cette publication, utilisez les coordonnées ci-après ou contactez le Comité national de l'IEC de votre pays de résidence.

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

Tel.: +41 22 919 02 11
Fax: +41 22 919 03 00
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 - www.iec.ch/searchpub

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 more than 30 000 terms and definitions in English and French, with equivalent terms in 14 additional languages. Also known as the International Electrotechnical Vocabulary (IEV) online.

IEC Glossary - std.iec.ch/glossary

More than 55 000 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: csc@iec.ch.

A propos de l'IEC

La Commission Electrotechnique Internationale (IEC) est la première organisation mondiale qui élabore et publie des Normes internationales pour tout ce qui a trait à l'électricité, à l'électronique et aux technologies apparentées.

A propos des publications IEC

Le contenu technique des publications IEC est constamment revu. Veuillez vous assurer que vous possédez l'édition la plus récente, un corrigendum ou amendement peut avoir été publié.

Catalogue IEC - webstore.iec.ch/catalogue

Application autonome pour consulter tous les renseignements bibliographiques sur les Normes internationales, Spécifications techniques, Rapports techniques et autres documents de l'IEC. Disponible pour PC, Mac OS, tablettes Android et iPad.

Recherche de publications IEC - www.iec.ch/searchpub

La recherche avancée permet de trouver des publications IEC en utilisant différents critères (numéro de référence, texte, comité d'études,...). Elle donne aussi des informations sur les projets et les publications remplacées ou retirées.

IEC Just Published - webstore.iec.ch/justpublished

Restez informé sur les nouvelles publications IEC. Just Published détaille les nouvelles publications parues. Disponible en ligne et aussi une fois par mois par email.

Electropedia - www.electropedia.org

Le premier dictionnaire en ligne de termes électroniques et électriques. Il contient plus de 30 000 termes et définitions en anglais et en français, ainsi que les termes équivalents dans 14 langues additionnelles. Egalement appelé Vocabulaire Electrotechnique International (IEV) en ligne.

Glossaire IEC - std.iec.ch/glossary

Plus de 55 000 entrées terminologiques électrotechniques, en anglais et en français, extraites des articles Termes et Définitions des publications IEC parues depuis 2002. Plus certaines entrées antérieures extraites des publications des CE 37, 77, 86 et CISPR de l'IEC.

Service Clients - webstore.iec.ch/csc

Si vous désirez nous donner des commentaires sur cette publication ou si vous avez des questions contactez-nous: csc@iec.ch.

INTERNATIONAL STANDARD

NORME INTERNATIONALE



Multimedia quality – Method of assessment of synchronization of audio and video

Qualité pour les systèmes multimédia – Méthode d'évaluation de la synchronisation de l'audio et de la vidéo

INTERNATIONAL
ELECTROTECHNICAL
COMMISSION

COMMISSION
ELECTROTECHNIQUE
INTERNATIONALE

PRICE CODE
CODE PRIX

M

ICS 33.160.01

ISBN 978-2-8322-1359-9

**Warning! Make sure that you obtained this publication from an authorized distributor.
Attention! Veuillez vous assurer que vous avez obtenu cette publication via un distributeur agréé.**

CONTENTS

FOREWORD.....	3
INTRODUCTION.....	5
1 Scope.....	6
2 Normative reference	6
3 Terms and definitions	6
4 Overview of methods of assessment.....	7
5 Subjective assessment of lip sync	7
5.1 Items to be assessed	7
5.2 Preparation of test video clips and test video sequence.....	8
5.2.1 Selection of content of a test video clip.....	8
5.2.2 Creation of a test video sequence.....	8
5.3 Procedures and condition for assessment of lip sync at the section 3-3'	9
5.4 Reporting of the result of assessment.....	9
6 Data processing	10
6.1 Items to be assessed	10
6.2 Method of assessment.....	10
6.3 Reporting of the result of estimation	11
Bibliography.....	12
Figure 1 – Overview.....	7
Figure 2 – Preparation of test video clips with time shifted audio	8
Figure 3 – An example of subjective assessment of lip sync	10
Figure 4 – Normalized response for grading impairment caused by lip sync mismatch	11
Table 1 – Five-grade impairment scale and explanation of subjective opinion score.....	9

INTERNATIONAL ELECTROTECHNICAL COMMISSION

**MULTIMEDIA QUALITY –
METHOD OF ASSESSMENT OF
SYNCHRONIZATION OF AUDIO AND VIDEO**

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 provides no marking procedure to indicate its approval and cannot be rendered responsible for any equipment declared to be in conformity with an IEC Publication.
- 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.

International Standard IEC 62503 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 bilingual version (2014-04) corresponds to the monolingual English version, published in 2008-09.

The text of this standard is based on the following documents:

CDV	Report on voting
100/1277/CDV	100/1358/RVC

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

The French version of this standard has not been voted upon.

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

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

- 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 document using a colour printer.

IECNORM.COM : Click to view the full PDF of IEC 62503:2008

INTRODUCTION

Contemporary multimedia systems are realized by digital technology. Depending on what digital processing is being applied, time delays differ among medium channels for reproduction as perceptible stimulus for a human audience. An example is video delay against audio, which is identified by such terms as lip sync or AV-sync. Video delay against audio will be inevitable for large sized displays, because necessary time for rendering and visualization will be proportional to the number of picture elements.

There should also be additional factors to be considered. They include synchronization problem during medium gathering, production, post-production, processing in various aspects to combine these multiple media and to be sent out or recorded as “multimedia”.

There is a need for international standards to provide the following three related methodologies:

- a) an objective method of measurement for difference of delays between reproduced audio and video (lip sync) by multimedia systems and equipment,
- b) a subjective (or perceptible) and statistical method of assessment of overall difference of delays between a real world and a reproduced scene and sound,
- c) a method of estimation of implied difference of delays inherent in multimedia received, recorded or under reproduction.

This International Standard addresses the item b) using typical multimedia content such as bust shots of news casters because of easiness in defining synchronization of audio and video. Since a range of perceptibly allowable miss-synchronization and sensitivity of human audience for lead and delayed audio against accompanied video depends on human sensation and the conditions for assessment, a clearly defined method of assessment of such characteristics should be standardised.

This International Standard is intended to supplement Recommendation ITU-R BT.1359-1, [1]¹, as well as partly answer the request of ITU-R to IEC that has been stated in Recommendation ITU-R BT.1377 [2].

Technical contents are based on a study in Faculty of Engineering, Chiba University in Japan, conducted in April 2006.

¹ Numbers in square brackets refer to the Bibliography.

MULTIMEDIA QUALITY – METHOD OF ASSESSMENT OF SYNCHRONIZATION OF AUDIO AND VIDEO

1 Scope

This International Standard provides a subjective (or perceptible) and statistical method of assessment of overall, or end-to-end, difference of delays between real world and reproduced scenes in terms of video and accompanying audio recoded in a medium.

This International Standard does not specify limiting values for those results obtained by the application of the provisions in this standard. It excludes applications to professional broadcast systems.

2 Normative reference

The following referenced document is indispensable for the application 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.

ITU-R BT.500-11:2002, *Methodology for the subjective assessment of the quality of television pictures*

3 Terms and definitions

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

3.1

lip sync

video delay against accompanying audio

3.2

outlier

subjective opinion score outside $m \pm s$, where m is a sample mean of the original scores of a set of subjects for the same video delay and s is a standard deviation of the scores

3.3

subject

ordinary untrained human audience of audio and video reproduction; random sample of individual members of general public

3.4

test video clip

short duration of video frames with accompanying audio to be used as original

3.5

test video sequence

random series of test video clips where the audio channels are shifted in time compared to the original

4 Overview of methods of assessment

Figure 1 depicts overview of possible objective methods of measurement and subjective method of assessment to acquire necessary parameters corresponding to lip sync.

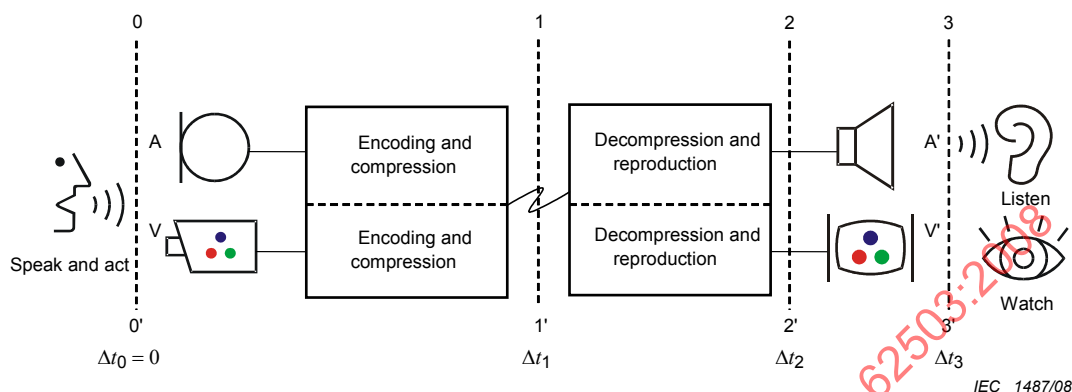


Figure 1 – Overview

The leftmost in Figure 1 is a real world and the rightmost in Figure 1 is a reproduced world.

Lip sync at the section 0-0', i.e. scene delay in reference to accompanying sound, is normally zero; in other words null video delay against accompanying audio is expected; $\Delta t_0 = 0$. Where $\Delta t_0 \neq 0$ is foreseen, it shall also be taken into account.

Lip sync at the section 1-1' is supposed to be introduced by separate acquisition of physical phenomena by microphones and video cameras followed by yet further separate digital processing for audio and video data. It will cause lip sync of $\Delta t_1 \neq 0$.

NOTE In case of MPEG-2 encoding, there is the scheme of synchronization using Decoding Time Stamp (DTS) as well as Presentation Time Stamp (PTS) embedded in the header of Packetized Elementary Stream (PES). See ISO/IEC 13818-1 [11].

Lip sync at the section 2-2' is supposed to be introduced by reproduction process for audio and video channels separately such as decompression, rendering and reproduction. It will cause lip sync of $\Delta t_2 \neq 0$, which can be measured using a reference test multimedia material with $\Delta t_1 = 0$.

Lip sync at the section 3-3' is in the reproduced multimedia world and assessed by human subjects. Subjective opinion scores on lip sync are statistically analyzed to find estimated value for $\Delta t_3 \neq 0$; corresponding to the amount of compensation for just-synchronized reproduction.

5 Subjective assessment of lip sync

5.1 Items to be assessed

Subjective grading level of miss-synchronization of video and audio.

5.2 Preparation of test video clips and test video sequence

5.2.1 Selection of content of a test video clip

Since lip sync is a kind of human perception, it may depend on the contents of the video and accompanying audio. Especially when it is related to movement of lips of a human speaker, a match between a spoken language and a mother tongue may affect the result.

NOTE In this International Standard, in order to provide worked examples, speech in Japanese language uttered by a well trained professional news reader is watched and listened to by the subjects with the same mother tongue.

A bust shot of a news reader shall be extracted, duration of which should be around 10 s to 20 s. Data of audio channel of the video clip shall be taken as the timing reference.

Possible amount of time caused by miss-synchronization in this original video clip, Δt_1 at the section 1-1', is unknown. However, this international standard provides the method to estimate overall lip sync Δt_3 including Δt_0 and Δt_1 . Namely, $\Delta t_3 = \Delta t_0 + \Delta t_1 + \Delta t_2$.

5.2.2 Creation of a test video sequence

The test video sequence shall be a randomised series of the video clip selected in 5.2.1, in which each of the audio channels shall be replaced by time-shifted audio data with necessary duration of padding as a leader or a trailer depending on the direction of the time shift. Preparation of such video clips is shown in Figure 2 as in the image frames with delayed audio and with led audio. The amount of time shifts T_l and T_d is subject to be adjusted.

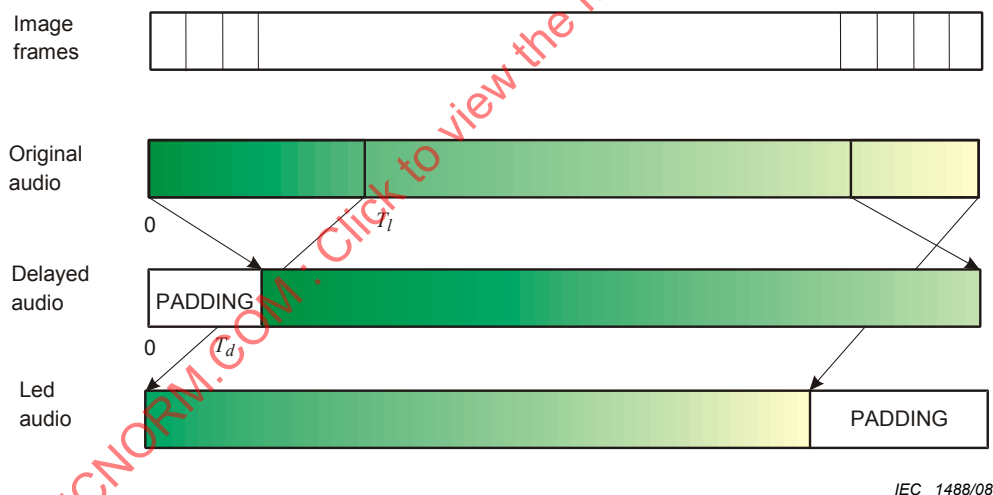


Figure 2 – Preparation of test video clips with time shifted audio

To allow for each of the video clips with the time-shifted audio composing the test video sequence to be visually identified by a subject, each video clip prepared in accordance with Figure 2 should be preceded by a necessary number of title frames which include a sequence number.

The amount of time shifts for audio data, T_l and T_d , shall be determined taking into account the sum of the lip sync in reproduction system, Δt_2 , and possible lip sync in the original video clip, Δt_1 . The amount of increment and decrement of the time shift ΔT for T_l or T_d shall be decided in accordance with precision of assessment.

In this standard, $\Delta T = 10$ ms is recommended.

The test video sequence should be stored in a medium such as CD-ROM for use in 5.3 without losing audio-video synchronization.

5.3 Procedures and condition for assessment of lip sync at the section 3-3'

The procedures described below shall be followed.

- The test video sequence being composed of randomized order of the same short video clips of different time-shifted audio (plus and minus) in reference to video shall be reproduced to subjects.
- The number of the subjects shall be at least 15. Each of them shall be asked to report its subjective opinion score for each of the video clips in the test video sequence under a fixed viewing and audible condition.
- An advance instruction to the subjects shall be provided on the five-grade impairment scale recommended by ITU-R BT.500-11 for subjective judgement, as shown in Table 1.

Table 1 – Five-grade impairment scale and explanation of subjective opinion score

Grade	Explanation
5	Imperceptible (Complete agreement of audio and video)
4	Perceptible but not annoying (miss-synchronization is within the acceptable degree)
3	Slightly annoying (slight miss-synchronization is recognized)
2	Annoying (apparent miss-synchronization is recognized and beyond acceptable degree)
1	Very annoying (not synchronized at all)

- Viewing distance of the reproduced test video sequence L shall be four times of reproduction height H of video frames: $L = 4H$.
- Audio shall be reproduced by a headphone or a loudspeaker. Use of the headphone is recommended to prevent any multi-path echo in a test room.

NOTE Taking into account the travel speed of sound waves, the time for the sound pressure of a speaker to reach to ears of the observer should be regarded as an additional audio delay.

- Illumination of a test room where the audio-video reproduction system is installed should be around 300 lx with correlated colour temperature of about 6 500 K.

5.4 Reporting of the result of assessment

Taking into account outliers from the original five-grade opinion scores, averaged subjective opinion scores shall be plotted against predetermined audio shift, T_l or $-T_d$, as exemplified in Figure 3, with error bars of $m \pm c$, where m is a sample mean of the opinion scores of a set of subjects for the same video clip and c is a 95 % confidence interval of each of the respective mean opinion scores.

The horizontal axis shall be the amount of time shifts of the audio channel or the value of T_l in Figure 2.

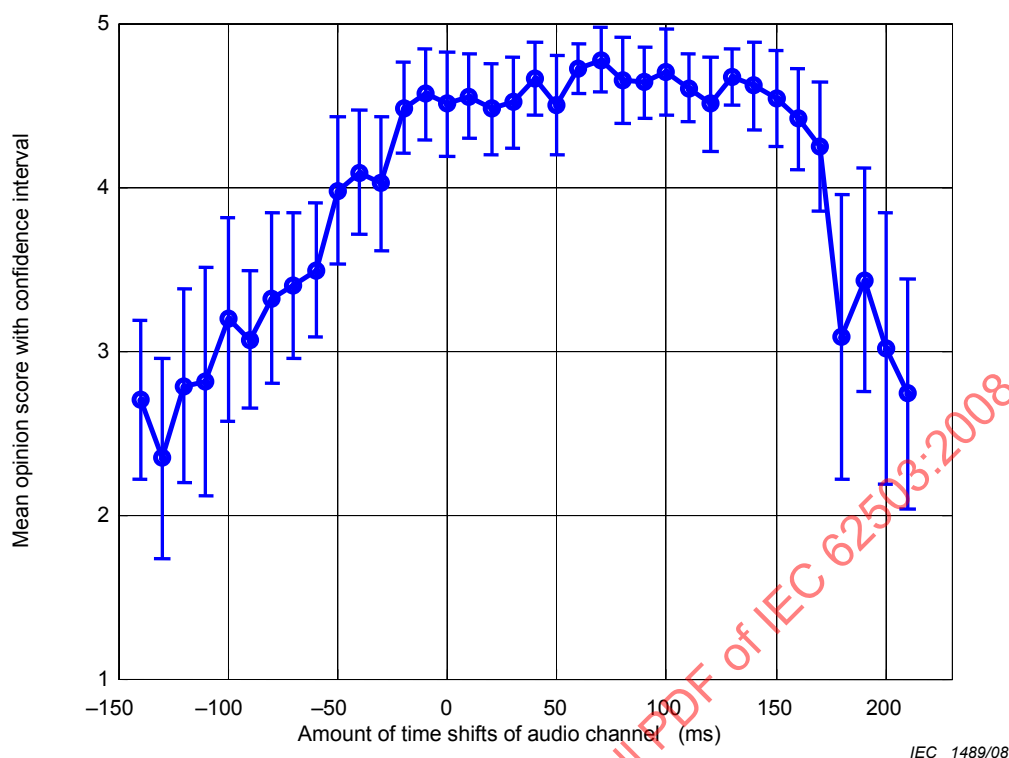


Figure 3 – An example of subjective assessment of lip sync

The report shall also include the following information:

- description of a kind of audio and video source used in the assessment, such as spoken language in case of a speech and duration of the test video clip;
- the number of the subjects;
- illumination level and correlated colour temperature of lighting used; and
- the position and distance from the display and speakers to the observer, in case of headphone not being used.

6 Data processing

6.1 Items to be assessed

Subjective time difference in milliseconds between reproduced video and associated audio.

6.2 Method of assessment

The mean opinion scores in the five-grade impairment scale reported in figure 3 shall be regressed by the stepwise linear functions as shown by the set of equations (1) to (3).

$$g = a_1 (t - t_1) + g_0, \text{ for } t < t_1 \quad (1)$$

$$g = g_0, \text{ for } t_1 \leq t \leq t_2 \quad (2)$$

$$g = a_2 (t - t_2) + g_0, \text{ for } t_2 < t \quad (3)$$

A horizontal coordinate value, or an estimate for Δt_3 in terms of Figure 1, should be calculated as an intersection of the two lines defined by equation (1) and equation (3). Namely, Δt_3 is expressed by equation (4).

$$\Delta t_3 = \frac{a_1 t_1 - a_2 t_2}{a_1 - a_2} \quad (4)$$

NOTE An amount of the lip sync, i.e. video delay against audio needed to be compensated falls between t_1 and t_2 in equation (2). The value of Δt_3 calculated by equation (4) is one of possible estimates.

Figure 4 explains the scheme of the method of estimation by an example, in which the vertical line at the intersection coordinate of the two lines corresponds to an estimated lip sync compensation Δt_3 .

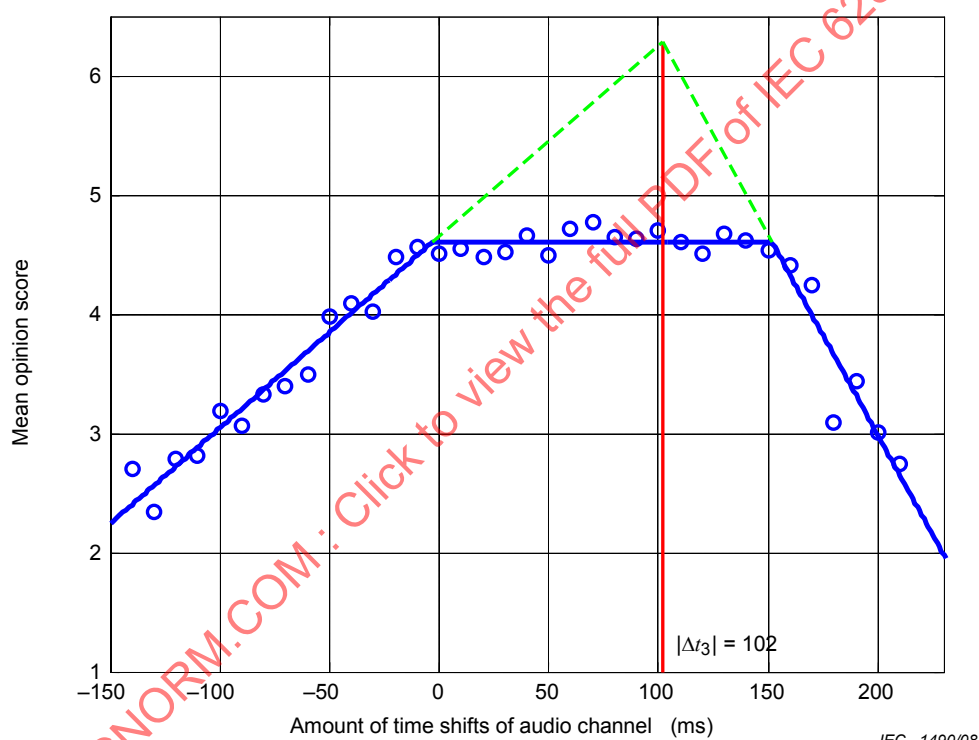


Figure 4 – Normalized response for grading impairment caused by lip sync mismatch

6.3 Reporting of the result of estimation

The amount of overall audio time shift to compensate lip sync obtained by the regression and the extrapolation shall be reported in milliseconds together with the time parameters t_1 and t_2 as shown below.

$$\begin{aligned} t_1 &= -3 \text{ ms} \\ t_2 &= 152 \text{ ms} \\ \Delta t_3 &= 102 \text{ ms} \end{aligned}$$

Bibliography

- [1] Recommendation ITU-R BT.1359-1: Relative timing of sound and vision for broadcasting (1998)
 - [2] Recommendation ITU-R BT.1377: Labelling of video and audio apparatus throughput (processing) delay (1998)
 - [3] ATSC Implementation Subcommittee Finding: Relative timing of sound and vision for broadcast operations, Advanced Television Systems Committee Doc. IS-191 (26 June 2003)
 - [4] EBU Technical Recommendation R37-2007: The relative timing of the sound and vision components of a television signal, the European Broadcasting Union
 - [5] IEC 60417, Graphical symbols for use on equipment,
<http://www.graphical-symbols.info/equipment>
 - [6] Hiroaki IKEDA, Reiko IWAI and Junichi YOSHIO: Kansei on the time difference between audio and video, *Hoso Gigyutsu (Broadcasting Technologies)*, vol.59, No.10, pp.151-155 (2006-10)
 - [7] Hiroaki IKEDA and Junichi YOSHIO: Synchronization in multimedia – Technical review, *Journal of IEEJ*, Vol. 126, No.5, pp.288-291 (May 2006)
 - [8] Question ITU-R 68/6: Synchronization necessary for the satisfactory reception of sound and picture signals (2003-10) – still under study as of 2006
 - [9] Question ITU-T 11/9: Requirements and methods for sound and television transmission over IP networks “webcasting” (Study period: 2005-2008)
 - [10] ISO/IEC 13818-1 Ed.2:2000, Amendments 1, 2, 3, 4, 5, Information technology – Generic coding of moving pictures and associated audio information – Part 1: systems
 - [11] IEC/TS 62312-2, Guideline for synchronization of audio and video – Part 2: Methods for synchronization of audio and video systems
-

IECNORM.COM : Click to view the full PDF of IEC 62503:2008

SOMMAIRE

AVANT-PROPOS	15
INTRODUCTION.....	17
1 Domaine d'application	18
2 Références normatives	18
3 Termes et définitions	18
4 Vue d'ensemble des méthodes d'évaluation.....	19
5 Évaluation subjective de la synchronisation labiale.....	20
5.1 Éléments devant être évalués	20
5.2 Préparation des clips vidéo d'essai et de la séquence vidéo d'essai.....	20
5.2.1 Choix du contenu d'un clip vidéo d'essai.....	20
5.2.2 Création d'une séquence vidéo d'essai	20
5.3 Procédures et condition pour l'évaluation de la synchronisation labiale à la section 3-3'	21
5.4 Rapport relatif au résultat d'évaluation	22
6 Traitement des données	23
6.1 Éléments devant être évalués	23
6.2 Méthode d'évaluation	23
6.3 Rapport relatif au résultat d'estimation	24
Bibliographie.....	25
Figure 1 – Vue d'ensemble	19
Figure 2 – Préparation de clips vidéo avec de l'audio décalé dans le temps.....	21
Figure 3 – Exemple d'évaluation subjective de la synchronisation labiale.....	23
Figure 4 – Réponse normalisée pour la dégradation de notes due à une discordance de synchronisation labiale.....	24
Tableau 1 – Échelle de dégradation à cinq notes et explication d'une note d'opinion subjective	21

COMMISSION ÉLECTROTECHNIQUE INTERNATIONALE

**QUALITÉ POUR LES SYSTÈMES MULTIMÉDIA –
MÉTHODE D'ÉVALUATION DE LA SYNCHRONISATION
DE L'AUDIO ET DE LA VIDÉO**

AVANT-PROPOS

- 1) La Commission Électrotechnique Internationale (CEI) est une organisation mondiale de normalisation composée de l'ensemble des comités électrotechniques nationaux (Comités nationaux de la CEI). La CEI a pour objet de favoriser la coopération internationale pour toutes les questions de normalisation dans les domaines de l'électricité et de l'électronique. À cet effet, la CEI – entre autres activités – publie des Normes internationales, des Spécifications techniques, des Rapports techniques, des Spécifications accessibles au public (PAS) et des Guides (ci-après dénommés "Publication(s) de la CEI"). Leur élaboration est confiée à des comités d'études, aux travaux desquels tout Comité national intéressé par le sujet traité peut participer. Les organisations internationales, gouvernementales et non gouvernementales, en liaison avec la CEI, participent également aux travaux. La CEI collabore étroitement avec l'Organisation Internationale de Normalisation (ISO), selon des conditions fixées par accord entre les deux organisations.
- 2) Les décisions ou accords officiels de la CEI concernant les questions techniques représentent, dans la mesure du possible, un accord international sur les sujets étudiés, étant donné que les Comités nationaux de la CEI intéressés sont représentés dans chaque comité d'études.
- 3) Les Publications de la CEI se présentent sous la forme de recommandations internationales et sont agréées comme telles par les Comités nationaux de la CEI. Tous les efforts raisonnables sont entrepris afin que la CEI s'assure de l'exactitude du contenu technique de ses publications; la CEI ne peut pas être tenue responsable de l'éventuelle mauvaise utilisation ou interprétation qui en est faite par un quelconque utilisateur final.
- 4) Dans le but d'encourager l'uniformité internationale, les Comités nationaux de la CEI s'engagent, dans toute la mesure possible, à appliquer de façon transparente les Publications de la CEI dans leurs publications nationales et régionales. Toutes divergences entre toutes Publications de la CEI et toutes publications nationales ou régionales correspondantes doivent être indiquées en termes clairs dans ces dernières.
- 5) La CEI n'a fixé aucune procédure concernant le marquage comme indication d'approbation et sa responsabilité n'est pas engagée quand un matériel est déclaré conforme à l'une de ses Publications.
- 6) Tous les utilisateurs doivent s'assurer qu'ils sont en possession de la dernière édition de cette publication.
- 7) Aucune responsabilité ne doit être imputée à la CEI, à ses administrateurs, employés, auxiliaires ou mandataires, y compris ses experts particuliers et les membres de ses comités d'études et des Comités nationaux de la CEI, pour tout préjudice causé en cas de dommages corporels et matériels, ou de tout autre dommage de quelque nature que ce soit, directe ou indirecte, ou pour supporter les coûts (y compris les frais de justice) et les dépenses découlant de la publication ou de l'utilisation de cette Publication de la CEI ou de toute autre Publication de la CEI, ou au crédit qui lui est accordé.
- 8) L'attention est attirée sur les références normatives citées dans cette publication. L'utilisation de publications référencées est obligatoire pour une application correcte de la présente publication.
- 9) L'attention est attirée sur le fait que certains des éléments de la présente Publication de la CEI peuvent faire l'objet de droits de brevet. La CEI ne saurait être tenue pour responsable de ne pas avoir identifié de tels droits de brevets et de ne pas avoir signalé leur existence.

La Norme internationale CEI 62503 a été établie par le domaine technique 11: Qualité pour les systèmes audio, vidéo et multimédia, du comité d'études 100 de la CEI: Systèmes et équipements audio, vidéo et multimédia.

La présente version bilingue (2014-04) correspond à la version anglaise monolingue publiée en 2008-09.

Le texte anglais de cette norme est issu des documents 100/1277/CDV et 100/1358/RVC.

Le rapport de vote 100/1358/RVC donne toute information sur le vote ayant abouti à l'approbation de cette norme.

La version française de cette norme n'a pas été soumise au vote.

Cette publication a été rédigée selon les Directives ISO/CEI, Partie 2.

Le comité a décidé que le contenu de cette publication ne sera pas modifié avant la date de maintenance indiquée sur le site web de la CEI sous "<http://webstore.iec.ch>" dans les données relatives à la publication recherchée. À cette date, la publication sera

- reconduite,
- supprimée,
- remplacée par une édition révisée, ou
- amendée.

IMPORTANT – Le logo "*colour inside*" qui se trouve sur la page de couverture de cette publication indique qu'elle contient des couleurs qui sont considérées comme utiles à une bonne compréhension de son contenu. Les utilisateurs devraient, par conséquent, imprimer cette publication en utilisant une imprimante couleur.

IECNORM.COM : Click to view the full PDF of IEC 62503:2008

INTRODUCTION

Les systèmes multimédia contemporains sont réalisés par technologie numérique. Selon le traitement numérique appliqué, les délais diffèrent parmi les canaux de support pour la reproduction sous forme de stimulus perceptible pour un auditoire humain. Un exemple est le retard de la vidéo par rapport à l'audio, qui est identifié par des termes tels que «synchronisation labiale» ("lip sync") ou «synchronisation AV» ("AV-sync"). Le retard de la vidéo par rapport à l'audio est inévitable pour les affichages de grandes dimensions, car le temps nécessaire pour le rendu et la visualisation est proportionnel au nombre d'éléments d'image.

Il convient également qu'il y ait d'autres facteurs à prendre en compte. Ils comprennent le problème de synchronisation pendant la collecte de supports, la production, la postproduction, le traitement en divers aspects pour combiner ces multiples supports et de les envoyer ou enregistrer comme "multimédia".

Il existe un besoin de normes internationales pour fournir les trois méthodologies connexes suivantes:

- a) une méthode objective de mesure de la différence des retards entre l'audio et la vidéo reproduits (synchronisation labiale) par les systèmes et équipements multimédia,
- b) une méthode subjective (ou perceptible) et statistique d'évaluation de la différence globale des retards entre le monde réel et une scène et un son reproduits,
- c) une méthode d'estimation de la différence impliquée des retards inhérents aux multimédia reçus, enregistrés ou reproduits.

La présente Norme internationale traite de l'élément b) en utilisant un contenu multimédia type tel que les plans rapprochés des journalistes de télévision en raison de la facilité à définir la synchronisation de l'audio et de la vidéo. Sachant qu'une plage de mauvaise synchronisation admissible du point de vue de la perception et la sensibilité de l'auditoire humain pour l'audio en avance et en retard par rapport à la vidéo accompagnée dépendent de la sensation humaine et des conditions d'évaluation, il convient de normaliser une méthode clairement définie d'évaluation de ces caractéristiques.

La présente Norme internationale vise à compléter la Recommandation UIT-R BT.1359-1, [1]¹ et aussi à répondre en partie à la demande de l'UIT-R à la CEI qui a été énoncée dans la Recommandation UIT-R BT.1377 [2].

Le contenu technique repose sur une étude réalisée en avril 2006 à la Faculté d'ingénierie de l'Université de Chiba au Japon.

¹ Les chiffres entre crochets renvoient à la Bibliographie.

QUALITÉ POUR LES SYSTÈMES MULTIMÉDIA – MÉTHODE D'ÉVALUATION DE LA SYNCHRONISATION DE L'AUDIO ET DE LA VIDÉO

1 Domaine d'application

La présente Norme internationale fournit une méthode subjective (ou perceptible) et statistique d'évaluation de la différence globale, ou de bout en bout, des retards entre le monde réel et les scènes reproduites en termes de vidéo et d'audio d'accompagnement enregistrés sur un support.

La présente Norme internationale ne spécifie pas de valeurs limites pour ces résultats obtenus par l'application des dispositions qu'elle contient. Elle exclut les applications aux systèmes professionnels de diffusion.

2 Références normatives

Le document de référence suivant est indispensable pour l'application du présent document. Pour les références datées, seule l'édition citée s'applique. Pour les références non datées, la dernière édition du document de référence s'applique (y compris les éventuels amendements).

UIT-R BT.500-11:2002, *Méthodologie d'évaluation subjective de la qualité des images de télévision*

3 Termes et définitions

Pour les besoins du présent document, les termes et définitions suivants s'appliquent.

3.1

synchronisation labiale (lip sync)

retard de la vidéo par rapport à l'audio d'accompagnement

3.2

valeur aberrante

note d'opinion subjective tombant à l'extérieur de la plage $m \pm s$, où m est une moyenne d'échantillons des notes d'origine d'un ensemble de sujets pour le même retard de vidéo et s est l'écart type des notes

3.3

sujet

auditoire humain ordinaire sans formation de la reproduction de l'audio et de la vidéo; échantillon aléatoire de membres individuels du grand public

3.4

clip vidéo d'essai

courte durée d'images vidéo avec l'audio d'accompagnement à utiliser comme original

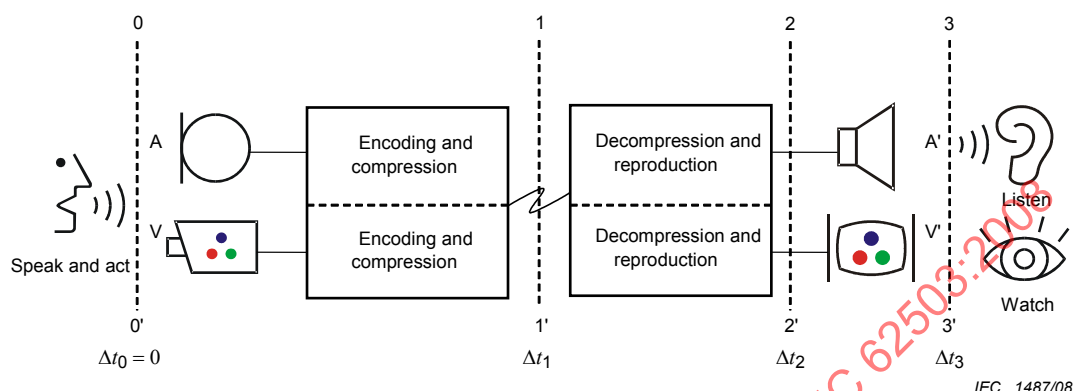
3.5

séquence vidéo d'essai

série aléatoire de clips vidéo d'essai où les canaux audio sont décalés dans le temps en comparaison à l'original

4 Vue d'ensemble des méthodes d'évaluation

La Figure 1 illustre une vue d'ensemble de possibles méthodes objectives de mesure et méthodes subjectives d'évaluation pour acquérir les nécessaires paramètres correspondant à la synchronisation labiale.



Légende

Anglais	Français
Speak and act	Parler et agir
Encoding and compression	Codage et compression
Decompression and reproduction	Décompression et reproduction
Listen	Écouter
Watch	Regarder

Figure 1 – Vue d'ensemble

La partie la plus à gauche de la Figure 1 est un monde réel alors que la partie la plus à droite de la Figure 1 est un monde reproduit.

La synchronisation labiale à la section 0-0', à savoir le retard scénique en référence au son d'accompagnement, est normalement de zéro; autrement dit, on s'attend à un retard vidéo nul par rapport à l'audio d'accompagnement; $\Delta t_0 = 0$. Lorsque $\Delta t_0 \neq 0$ est prévisible, il doit aussi être pris en compte.

La synchronisation labiale à la section 1-1' est supposée être introduite par une acquisition séparée de phénomènes physiques par des microphones et des caméras vidéo, suivie d'un traitement numérique ultérieur pour les données audio et vidéo. Cela donne lieu à une synchronisation labiale de $\Delta t_1 \neq 0$.

NOTE En cas de codage MPEG-2, il existe un plan de synchronisation utilisant le champ horodatage de décodage (Decoding Time Stamp (DTS)) ainsi que le champ horodatage de présentation (Presentation Time Stamp (PTS)) intégrés dans l'en-tête de la suite élémentaire groupée en paquets (Packetized Elementary Stream (PES)). Voir l'ISO/CEI 13818-1 [11].

La synchronisation labiale à la section 2-2' est supposée être introduite séparément par un processus de reproduction pour les canaux audio et vidéo tel que la décompression, le rendu et la reproduction. Cela donne lieu à une synchronisation labiale de $\Delta t_2 \neq 0$, qui peut être mesurée en utilisant un support multimédia d'essai de référence avec $\Delta t_1 = 0$.

La synchronisation labiale à la section 3-3' se situe dans le monde multimédia reproduit et elle est évaluée par des sujets humains. Les notes d'opinion subjective relatives à la

synchronisation labiale sont analysées statistiquement pour trouver la valeur estimée pour $\Delta t_3 \neq 0$; ce qui correspondant à la quantité de compensation pour la reproduction juste synchronisée.

5 Évaluation subjective de la synchronisation labiale

5.1 Éléments devant être évalués

Niveau de notation subjective de la mauvaise synchronisation de la vidéo et de l'audio.

5.2 Préparation des clips vidéo d'essai et de la séquence vidéo d'essai

5.2.1 Choix du contenu d'un clip vidéo d'essai

La synchronisation labiale étant une perception humaine, elle peut dépendre du contenu de la vidéo et de l'audio d'accompagnement. Notamment lorsqu'elle est relative au mouvement des lèvres d'un orateur humain, une correspondance entre une langue parlée et une langue maternelle peut affecter le résultat.

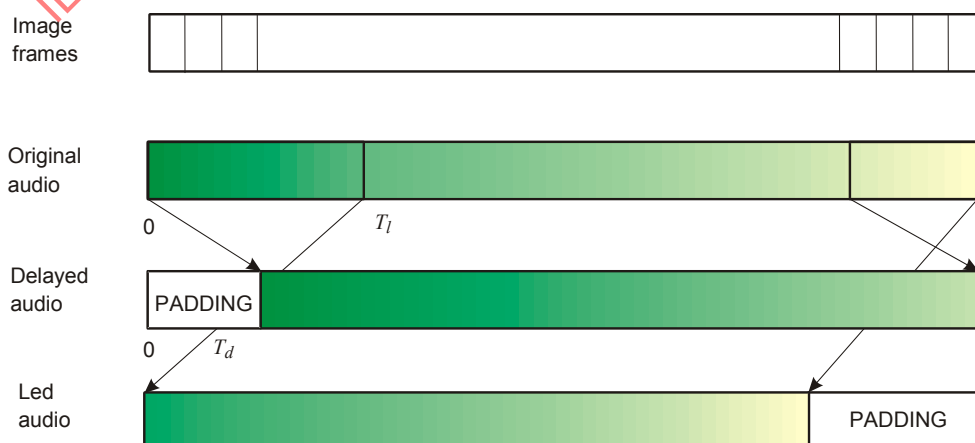
NOTE Dans la présente Norme internationale, afin de fournir des exemples pratiques, un discours en langue japonaise prononcé de façon processionnelle par un lecteur de nouvelles bien entraîné est regardé et écouté par des sujets ayant la même langue maternelle.

Un plan rapproché d'un lecteur de nouvelles doit être extrait et il convient que sa durée soit de l'ordre de 10 s à 20 s. Des données de canal audio du clip vidéo doivent être prélevées comme référence de temporisation.

La durée possible due à la mauvaise synchronisation dans le clip vidéo d'origine, Δt_1 à la section 1-1', est inconnue. Cependant, la présente Norme internationale fournit la méthode pour estimer la synchronisation labiale globale Δt_3 contenant Δt_0 et Δt_1 . À savoir, $\Delta t_3 = \Delta t_0 + \Delta t_1 + \Delta t_2$.

5.2.2 Création d'une séquence vidéo d'essai

La séquence vidéo d'essai doit être une série aléatoire du clip vidéo choisie en 5.2.1, dans laquelle chacun des canaux audio doit être remplacé par les données audio décalées dans le temps avec la durée nécessaire du bourrage comme amorce ou fin selon le sens du décalage temporel. La préparation de ces clips vidéo est montrée à la Figure 2 comme dans les trames d'image avec l'audio retardé et avec l'audio mené. La quantité de décalages temporels T_l et T_d est susceptible d'être réajustée.



Légende

Anglais	Français
Image frames	Trames d'image
Original audio	Audio d'origine
Delayed audio	Audio en retard
PADDING	BOURRAGE
Led audio	Audio en avance

Figure 2 – Préparation de clips vidéo avec de l'audio décalé dans le temps

Pour permettre que chacun des clips vidéo avec audio décalé dans le temps composant la séquence vidéo d'essai soit identifié visuellement par un sujet, il convient que chaque clip vidéo préparé conformément à la Figure 2 soit précédé du nombre nécessaire de trames de titre qui comprennent un numéro de séquence.

La quantité de décalages temporels pour les données audio, T_l et T_d , doit être déterminée en tenant compte de la somme de la synchronisation labiale dans le système de reproduction, Δt_2 , et de la possible synchronisation labiale dans le clip vidéo d'origine, Δt_1 . La quantité d'incrément et de décrément du décalage temporel ΔT pour T_l ou T_d doit être décidée en fonction de la précision de l'évaluation.

Dans la présente norme, la valeur recommandée est $\Delta T = 10$ ms.

Il convient que la séquence vidéo d'essai soit stockée sur un support tel qu'un CD-ROM pour une utilisation en 5.3 sans perdre la synchronisation audio-vidéo.

5.3 Procédures et condition pour l'évaluation de la synchronisation labiale à la section 3-3'

Les procédures décrites ci-dessous doivent être suivies.

- La séquence vidéo d'essai étant composée d'un ordre aléatoire des mêmes clips vidéo courts de différents décalages temporels de l'audio (plus et moins) par rapport à la vidéo doit être reproduite pour les sujets.
- Le nombre de sujets doit être au moins de 15. Il doit être demandé à chacun d'eux d'indiquer sa note d'opinion subjective pour chacun des clips vidéo dans la séquence vidéo d'essai dans des conditions fixes d'observation et d'audition.
- Une instruction d'avance aux sujets doit être fournie sur l'échelle de dégradation à cinq notes recommandée par l'UIT-R BT.500-11 pour le jugement subjectif, conformément au Tableau 1.

Tableau 1 – Échelle de dégradation à cinq notes et explication d'une note d'opinion subjective

Note	Explication
5	Imperceptible (Accord total de l'audio et de la vidéo)
4	Perceptible, mais non gênant (la mauvaise synchronisation se situe dans les limites du degré acceptable)
3	Légèrement gênant (une légère mauvaise synchronisation est reconnue)
2	Gênant (une mauvaise synchronisation évidente est reconnue et se situe au-delà du degré acceptable)
1	Très gênante (pas de synchronisation du tout)

- d) La distance d'observation de la séquence vidéo d'essai reproduite L doit être égale à quatre fois la hauteur de reproduction H des images vidéo: $L = 4H$.
- e) L'audio doit être reproduit par un casque d'écoute ou un haut-parleur. L'utilisation du casque d'écoute est recommandée pour prévenir tout écho par trajets multiples dans une salle d'essais.

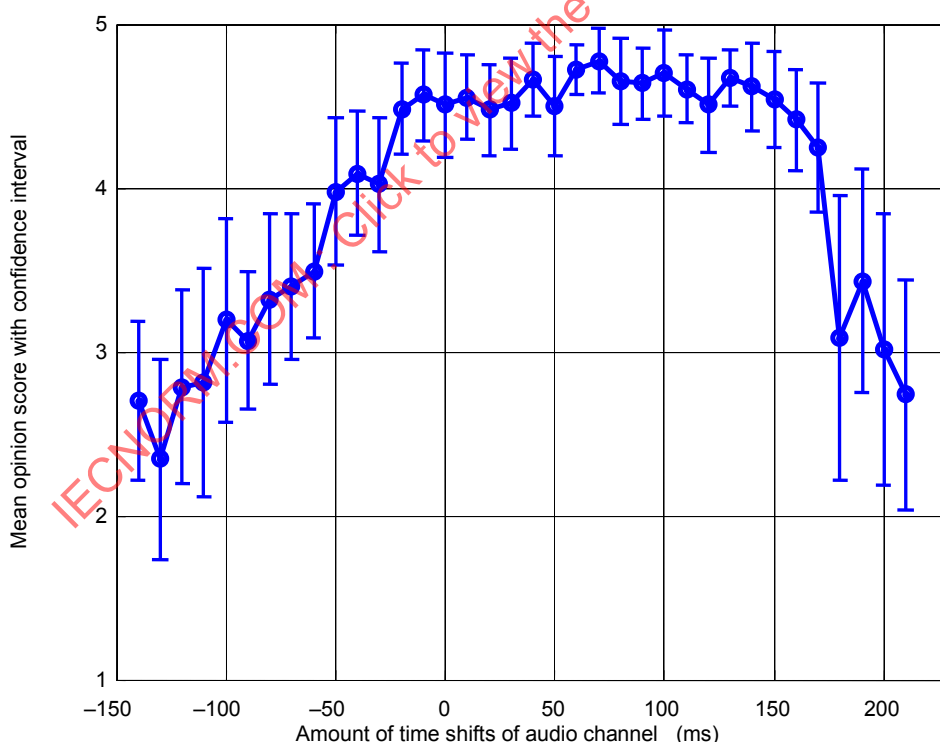
NOTE Compte tenu de la vitesse de déplacement des ondes acoustiques, il convient de considérer comme un retard audio supplémentaire le temps mis par la pression sonore d'un haut-parleur pour atteindre les oreilles de l'observateur.

- f) Il convient que l'éclairage d'une salle d'essais où est installé le système de reproduction audio-vidéo soit d'environ 300 lx avec une température de couleur proximale d'environ 6 500 K.

5.4 Rapport relatif au résultat d'évaluation

Compte tenu des valeurs aberrantes dans les notations d'opinion à cinq notes d'origine, les notes d'opinion objective moyennées doivent être pointées sur un graphique en fonction du décalage audio prédéterminé, T_l ou $-T_d$, conformément aux exemples montrés à la Figure 3, avec des barres d'erreur de $m \pm c$, où m est une moyenne d'échantillons de notes d'opinion d'un ensemble de sujets pour le même clip vidéo et c est un intervalle de confiance à 95 % de chacune des notes moyennées d'opinion respectives.

L'axe horizontal doit être la quantité de décalages temporels du canal audio ou la valeur de T_l à la Figure 2.



IEC 1489/08

Légende

Anglais	Français
Mean opinion score with confidence interval	Note d'opinion moyenne avec intervalle de confiance
Amount of time shifts of audio channel (ms)	Quantité de décalages temporels du canal audio