

INTERNATIONAL STANDARD

NORME INTERNATIONALE

**Sound system equipment –
Part 16: Objective rating of speech intelligibility by speech transmission index**

**Equipements pour systèmes électroacoustiques –
Partie 16: Evaluation objective de l'intelligibilité de la parole au moyen de l'indice
de transmission de la parole**





THIS PUBLICATION IS COPYRIGHT PROTECTED

Copyright © 2011 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 la CEI ou du Comité national de la CEI du pays du demandeur.

Si vous avez des questions sur le copyright de la CEI ou si vous désirez obtenir des droits supplémentaires sur cette publication, utilisez les coordonnées ci-après ou contactez le Comité national de la CEI de votre pays de résidence.

IEC Central Office
3, rue de Varembe
CH-1211 Geneva 20
Switzerland
Email: inmail@iec.ch
Web: 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.

- Catalogue of IEC publications: www.iec.ch/searchpub

The IEC on-line Catalogue enables you to search by a variety of criteria (reference number, text, technical committee,...). It also gives information on projects, withdrawn and replaced publications.

- IEC Just Published: www.iec.ch/online_news/justpub

Stay up to date on all new IEC publications. Just Published details twice a month all new publications released. Available on-line and also by email.

- Electropedia: www.electropedia.org

The world's leading online dictionary of electronic and electrical terms containing more than 20 000 terms and definitions in English and French, with equivalent terms in additional languages. Also known as the International Electrotechnical Vocabulary online.

- Customer Service Centre: www.iec.ch/webstore/custserv

If you wish to give us your feedback on this publication or need further assistance, please visit the Customer Service Centre FAQ or contact us:

Email: csc@iec.ch
Tel.: +41 22 919 02 11
Fax: +41 22 919 03 00

A propos de la CEI

La Commission Electrotechnique Internationale (CEI) 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 CEI

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

- Catalogue des publications de la CEI: www.iec.ch/searchpub/cur_fut-f.htm

Le Catalogue en-ligne de la CEI vous permet d'effectuer des recherches en utilisant différents critères (numéro de référence, texte, comité d'études,...). Il donne aussi des informations sur les projets et les publications retirées ou remplacées.

- Just Published CEI: www.iec.ch/online_news/justpub

Restez informé sur les nouvelles publications de la CEI. Just Published détaille deux fois par mois les nouvelles publications parues. Disponible en-ligne et aussi par email.

- Electropedia: www.electropedia.org

Le premier dictionnaire en ligne au monde de termes électroniques et électriques. Il contient plus de 20 000 termes et définitions en anglais et en français, ainsi que les termes équivalents dans les langues additionnelles. Egalement appelé Vocabulaire Electrotechnique International en ligne.

- Service Clients: www.iec.ch/webstore/custserv/custserv_entry-f.htm

Si vous désirez nous donner des commentaires sur cette publication ou si vous avez des questions, visitez le FAQ du Service clients ou contactez-nous:

Email: csc@iec.ch
Tél.: +41 22 919 02 11
Fax: +41 22 919 03 00



INTERNATIONAL STANDARD

NORME INTERNATIONALE

**Sound system equipment –
Part 16: Objective rating of speech intelligibility by speech transmission index**

**Équipements pour systèmes électroacoustiques –
Partie 16: Évaluation objective de l'intelligibilité de la parole au moyen de l'indice
de transmission de la parole**

INTERNATIONAL
ELECTROTECHNICAL
COMMISSION

COMMISSION
ELECTROTECHNIQUE
INTERNATIONALE

PRICE CODE **XB**
CODE PRIX

ICS 33.160.01

ISBN 978-2-88912-672-9

CONTENTS

FOREWORD.....	5
INTRODUCTION.....	7
1 Scope.....	9
2 Normative references	9
3 Terms and definitions	9
4 Description of the STI method	15
4.1 General.....	15
4.1.1 Rationale for the STI method.....	15
4.1.2 Applicability of the STI method	15
4.2 Background of the STI method	16
4.2.1 General	16
4.2.2 Theoretical overview.....	17
4.2.3 Measurement of STI	18
4.3 Applicability of STI test methods	19
4.4 Use of direct and indirect methods	20
4.5 Limitations of the STI method.....	21
4.5.1 General	21
4.5.2 Frequency shifts	21
4.5.3 Centre clipping	22
4.5.4 Drop outs	22
4.5.5 Jitter.....	22
4.5.6 Vocoders	22
4.5.7 Overestimation of STI under low background noise conditions.....	22
4.5.8 Frequency response	22
4.5.9 Echoes	23
4.5.10 Fast amplitude compression and expansion.....	23
4.5.11 Non-linear distortion	24
4.5.12 Impulsive and fluctuating noise.....	24
4.5.13 Hearing impaired listeners	24
4.6 Conclusion	24
5 Direct method of measuring STI	24
5.1 Overview	24
5.2 STIPA	25
5.3 Application	26
5.4 Limitations.....	26
6 Indirect method of measuring STI using the impulse response.....	26
6.1 Overview	26
6.2 Application	27
6.3 Limitations (non-linear distortion)	28
7 Measurement procedures, post-processing of data and applications.....	28
7.1 General.....	28
7.2 Acoustical input.....	28
7.3 Acoustical output.....	30
7.4 Electrical input	30
7.5 Electrical output	30
7.6 Examples of input/output combinations.....	30

7.6.1	Acoustical input – Acoustical output	30
7.6.2	Electrical input – Electrical output (e.g. assessment of wired and wireless) communication systems).....	30
7.6.3	Acoustical input – Electrical output (e.g. assessment of microphones).....	31
7.6.4	Electrical input – Acoustical output (e.g. assessment of PA systems)	31
7.7	Post-processing of measured MTF data	31
7.8	Issues concerning noise	31
7.8.1	General	31
7.8.2	Measurement of background noise	32
7.8.3	Fluctuating noise	32
7.9	Analysis and interpretation of the results	32
7.10	Binaural STI measurements	33
8	Use of STI as a design prediction tool	33
8.1	Overview	33
8.2	Statistical predictions	33
8.3	Prediction from simulated impulse response.....	34
Annex A (normative)	Speech transmission index (STI) and revised STI methods	35
Annex B (normative)	STIPA method.....	48
Annex C (normative)	STITEL method.....	49
Annex D (informative)	RASTI method (obsolete)	50
Annex E (informative)	Qualification of the STI and relationships with other speech intelligibility measures.....	52
Annex F (informative)	Nominal qualification bands for STI	54
Annex G (informative)	Examples of STI qualification bands and typical applications	55
Annex H (informative)	Non-native listeners	56
Annex I (informative)	Effect of age-related hearing loss and hearing impairment on speech intelligibility.....	57
Annex J (normative)	Calibration of STI test signal level	58
Annex K (informative)	Example test report sheet for STI measurements	60
Annex L (normative)	Prediction of STI using statistical methods	62
Annex M (informative)	Adjustments to measured STI and STIPA results for simulation of occupancy noise and different speech levels	64
Annex N (informative)	Other methods of determining speech intelligibility.....	68
Bibliography.....		70
Figure 1 –	Concept of the reduction in modulation due to a transmission channel	16
Figure 2 –	Modulation transfer function – Input/output comparison.....	17
Figure A.1 –	Envelope function (panel A) of a 10 s speech signal for the 250 Hz octave band and corresponding envelope spectrum (panel B).....	36
Figure A.2 –	Theoretical expression of the MTF	37
Figure A.3 –	Measurement system and frequencies for the STI method	39
Figure A.4 –	Auditory masking of octave band ($k - 1$) on octave band (k).....	40
Figure D.1 –	Illustration of a practical RASTI test signal	51
Figure E.1 –	Relationships between some speech intelligibility measures	52
Figure E.2 –	Relationship between STI, speech intelligibility scores and listening difficulty ratings [34], [35].....	53
Figure F.1 –	STI qualification bands.....	54

Table 1 – Comparison of STI test methods for different types of distortion	19
Table 2 – Applicability of test	20
Table 3 – Choice of method	21
Table A.1 – Auditory masking as a function of the octave band level.....	41
Table A.2 – Absolute speech reception threshold level in octave bands	42
Table A.3 – MTI octave band weighting factors	42
Table A.4 – Octave band levels (dB) relative to the A-weighted speech level	43
Table B.1 – Modulation frequencies for the STIPA method.....	48
Table C.1 – Modulation frequencies for the STITEL method.....	49
Table D.1 – Modulation frequencies for the RASTI method	50
Table E.1 – Categories for listening difficulty	53
Table G.1 – Examples between STI qualification bands and typical applications	55
Table H.1 – Adjusted intelligibility qualification tables for non-native listeners.....	56
Table I.1 – Adjusted intelligibility qualification tables for normal listeners and people over 60 years old with hearing loss	57
Table M.1 – Example calculation	64

INTERNATIONAL ELECTROTECHNICAL COMMISSION

SOUND SYSTEM EQUIPMENT –

**Part 16: Objective rating of speech intelligibility
by speech transmission index**

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.

International Standard IEC 60268-16 has been prepared by IEC technical committee 100: Multimedia equipment and systems.

This fourth edition cancels and replaces the third edition, published in 2003, and constitutes a technical revision.

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

- development of more comprehensive, complete and unambiguous standardization of the STI methodology;
- the term STI_r is discontinued. A new function for the prediction of auditory masking effects is introduced;
- the concept of 'speech level' and the setting of the level of the test signal have been introduced;

- additional information has been included on prediction and measurement procedures.

NOTE See Introduction for a historical summary referring to the various changes from the first to the fourth edition (current edition).

This bilingual version corresponds to the monolingual English version, published in 2011-06.

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

FDIS	Report on voting
100/1812/FDIS	100/1849/RVD

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.

A list of all the parts in the IEC 60268 series, published under the general title *Sound system equipment* 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 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.

INTRODUCTION

The Speech Transmission Index (STI) is an objective measure to predict the intelligibility of speech transmitted from talker to listener by a transmission channel. The STI method has been the subject of ongoing development and refinement since its introduction in the 1970s. Major improvements of the STI have been consolidated by incorporating them in successive revisions of IEC 60268-16.

The history of revisions is as follows.

- Revision 1: 1988. In the first version of the STI standard, a gender-independent test signal spectrum was used.
- Revision 2: 1998. Gender specific test signals were introduced, for male and female talkers, each gender relating to a specific set of weighting factors. In addition, weightings were introduced for redundancy factors. The term STI_r was introduced to signify the use of these redundancy factors.
- Revision 3: 2003. Important differences between Revision 2 and Revision 3 are the introduction of
 - level dependent masking functions,
 - the STI derivative STIPA.

STIPA was specially developed as a fast measurement method that could deal with electro-acoustic and acoustic effects while determining the speech transmission quality of PA systems.

- Revision 4: 2010. The aim of Revision 4 (this revision) is to provide a more comprehensive, complete and unambiguous standardization of the STI methodology. The term STI_r is now discontinued. A new function for the prediction of auditory masking effects is introduced.

Speech is considered to be the major method of communication between humans. In many situations the speech signal is degraded by the signal path or the transmission channel between talker and listener, resulting in a reduction of the intelligibility of the speech at the listener's location.

To quantify the deterioration of the speech intelligibility induced by the transmission channel, a fast and objective measuring method was developed; the Speech Transmission Index (STI). The STI method applies a specific test signal to the transmission channel and by analysing the received test signal; the speech transmission quality of the channel is derived and expressed in a value between 0 and 1, as the Speech Transmission Index (STI). Using the obtained STI-value, the potential speech intelligibility can be determined.

Although there are limitations to the STI method, the use of STI has proved useful in many situations and has gained international acceptance.

Items that have changed in this revision

Specific changes that have been incorporated in this revision are:

- refinement of the STI model with respect to the level dependent masking function;
- Room Acoustic Speech Transmission Index (RASTI) has become obsolete and should not be used;
- calculations to add or remove the effects of background noise and to change the speech level and a worked example;
- notes regarding limitations of the STI method;
- methods to predict the STI performance of transmission channels based on the predicted (as distinct from measured) performance of parts or all of the transmission channel;

- introduction of STI corrections for non-native language listeners;
- introduction of STI corrections for listeners with some specific forms of hearing loss;
- relationships between STI and 'Listening Difficulty' scale.

Potential applications of STI

STI may be used to measure the potential intelligibility of a wide range of electronic systems and acoustic environments. Typical applications include:

- measurement of Public Address and Sound Reinforcement Systems;
- measurement and Certification of Voice Alarm and emergency sound systems;
- measurement of communication channels / systems such as intercoms and wireless communication;
- measurement of potential speech intelligibility and communication in rooms and auditoria;
- evaluation of direct speech communication (situations without electronic amplification) in rooms or acoustic spaces including vehicles;
- evaluation of the potential intelligibility of Assistive Hearing Systems;

NOTE The STI method is not validated for the measurement and evaluation of speech privacy or speech masking systems.

Potential users of STI

The range of users of STI measurements is diverse. Among the users who may apply this method are:

- certifiers of voice alarm and other types of emergency systems;
- certifiers of sound reinforcement and audio systems;
- audio and telecommunication equipment manufacturers;
- audio and communication engineers;
- acoustical and electro-acoustical engineers;
- sound system installers;
- researchers into STI methods and developers of instruments to measure STI.

To avoid misinterpretation of STI results, it is important that all users have an understanding of the basic principles, the application domain and its limitations.

SOUND SYSTEM EQUIPMENT –

Part 16: Objective rating of speech intelligibility by speech transmission index

1 Scope

This part of IEC 60268 specifies objective methods for rating the transmission quality of speech with respect to intelligibility.

The objective of this standard is to provide a comprehensive manual for all types of users of the STI method in the fields of audio, communications and acoustics.

This standard does not provide STI criteria for certification of transmission channels (e.g. criteria for a voice-alarm system).

Three methods are presented, which are closely related and are referred to as STI, STIPA, and STITEL. The first two methods are intended for rating speech transmission performance with or without sound systems. The STITEL method has more restricted uses.

NOTE None of the methods are suitable for the measurement and assessment of speech privacy and speech masking systems, as STI has not been validated for conditions that represent speech privacy applications [1]¹.

The following information is included:

- measurement techniques;
- prediction techniques.

2 Normative references

The following referenced documents are 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:

IEC 61260:1995, *Electroacoustics – Octave-band and fractional-octave-band filters*
Amendment 1 (2001)

ISO 18233:2006, *Acoustics – Application of new measurement methods in building and room acoustics*

3 Terms and definitions

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

3.1 speech intelligibility

rating of the proportion of speech that is understood

¹ Figures in square brackets refer to the Bibliography.

3.2

speech quality

rating of sound quality of a speech signal

3.3

speech transmission index

STI

metric ranging between 0 and 1 representing the transmission quality of speech with respect to intelligibility by a speech transmission channel

3.4

speech intelligibility index

SII

objective method for prediction of speech intelligibility based on the Articulation Index

3.5

STI method

FULL STI

objective method for prediction and measurement of the speech transmission index that uses 14 modulation frequencies over a range of 7 octave bands

3.6

distortion

any unintentional and generally undesired change of the form of a signal occurring in a speech transmission channel

NOTE Distortion can include both linear and non-linear effects in both frequency and time domain.

3.7

speech transmission index for public address systems

STIPA

method obtained by using a condensed version of the STI method but still responsive to distortions found in room acoustics and/or public address systems

NOTE STIPA is applied as a direct method.

3.8

speech transmission index for telecommunication systems

STITEL

method obtained by using a condensed version of the STI method but still responsive to distortions found in communication systems

NOTE STITEL is applied as a direct method.

3.9

room acoustical speech transmission index

RASTI

method obtained by using a condensed version of the STI method, to be used for screening purposes only and focused on direct communication between persons without making use of an electro-acoustic communication system

NOTE 1 RASTI accounts for noise interference and distortions in the time domain (echoes, reverberation).

NOTE 2 RASTI is now obsolete.

3.10

direct STI method

method using modulated (speech like) test signals to directly measure the modulation transfer function

3.11**indirect STI method**

method using the impulse response and forward energy integral (Schroeder integral) to derive the modulation transfer function

3.12**speech transmission channel**

acoustic or electro-acoustic signal path between a talker and a listener

3.13**public address system****PA**

electronic sound distribution system, employing microphones, amplifiers and loudspeakers, used to reinforce or amplify a given sound (such as an announcement or a pre-recorded message) and distributing the sound within a building or a space

3.14**voice alarm system****VAS**

sound distribution system that broadcasts speech messages or warning signals, or both, in an emergency

3.15**real speech level**

signal level of a speech signal in dB A where only the segments that contribute to the speech signal are taken into account; pauses and silences between words and sentences are ignored

NOTE See also Annex J.

3.16**reference speech level**

speech level equivalent to 60 dB A at 1 m distance in front of the talker's mouth

3.17**vocal effort**

exertion of the speaker, quantified objectively by the A-weighted speech level at 1 m distance in front of the mouth and qualified subjectively by a description

3.18**artificial mouth**

device consisting of a loudspeaker mounted in an enclosure and having a directivity and radiation pattern similar to those of the average human mouth

NOTE The degree of similarity required cannot be easily specified and depends on the particular application. See for example ITU-T P.50 [47].

3.19**non-native speaker**

person speaking a language which is different from the language that was learned as primary language during the childhood of the speaker

3.20**absolute speech reception threshold**

absolute threshold of hearing increased by the minimal required dynamic range for the correct recognition of speech

3.21

auditory masking

process by which the threshold of hearing (audibility) for one sound is raised by the presence of another (masking) sound

NOTE Within the STI method, auditory masking is also referred to as the upward spread of masking.

3.22

artificial ear

device with similar characteristics as the human ear for the reception of acoustic signals

NOTE See IEC 60318 [2].

3.23

intensity function

the squared amplitude signal as a function of time

3.24

envelope function

envelope of the intensity function

3.25

envelope spectrum

relative contribution of spectral components of the envelope function

3.26

modulation frequency

frequency of the sinusoidal variation of the envelope function

NOTE The modulation frequency f_m is expressed in Hertz (Hz).

3.27

modulation index

value between 0 and 1 that describes the depth of a sinusoidal modulation of the intensity function

3.28

modulation transfer ratio

ratio between the modulation depth of the received and the original (transmitted) modulation depth of the intensity function

3.29

modulation transfer function

MTF

modulation transfer ratios as a function of the modulation frequency

3.30

octave band weighting factor

α

relative contribution of each octave band to the speech transmission index

3.31

octave band redundancy factor

β

fraction of information overlap between two adjacent octave bands with respect to the speech intelligibility

3.32**background noise**

all sounds including noise remaining in the absence of the speech or test signal

3.33**fluctuating noise**

continuous sound or noise whose sound pressure level varies significantly, but not in an impulsive manner, during the observation period

3.34**impulsive noise**

sound or noise characterized by brief bursts of sound pressure

3.35**signal-to-noise ratio*****SNR***

difference between the sound pressure level of the speech or test signal and the sound pressure level of the background noise where the sound pressure levels are determined with a standardized frequency weighting

NOTE The signal-to-noise ratio *SNR* is expressed in decibels (dB).

3.36**effective signal-to-noise ratio** **SNR_{eff}**

difference between the level of the intensity modulation and the level of the intensity of all the distortions of a received STI test signal

NOTE 1 The effective signal-to-noise ratio is expressed in decibels (dB).

NOTE 2 Examples of distortions are reverberation field levels, ambient noise levels, non-linear distortion levels and masking levels.

3.37**crest factor**

difference between the peak and the RMS sound pressure levels during a given time-interval

NOTE The crest-factor is expressed in decibels (dB).

3.38**Lombard effect**

spontaneous increase of the vocal effort induced by the increase of the ambient noise level at the speaker's ear

NOTE Voice pitch shift at higher talking levels is not accounted for here.

3.39**fractional-octave-band filter**

bandpass filter for which the ratio of upper cut-off frequency f_2 to lower cut-off frequency f_1 is two raised to an exponent equal to the fraction of an octave band

NOTE 1 In symbols, the ratio of the cut-off frequencies is $f_2/f_1 = 2^{1/b}$, with $1/b$ denoting the fraction of an octave.

EXAMPLE 1 For half-octave band filters, the frequency ratio is $2^{1/2} = \sqrt{2}$.

EXAMPLE 2 For octave band filters, the frequency ratio is 2.

NOTE 2 For further information, refer to IEC 61260.

**3.40
reference sound pressure**

p_0
sound pressure, conventionally chosen to be equal to 20 µPa for airborne sound

**3.41
sound pressure level**

twenty times the logarithm to the base ten of the ratio of RMS sound pressure to the reference sound pressure

NOTE The sound pressure level is expressed in decibels (dB). The notation is L_p .

**3.42
equivalent continuous sound pressure level**

ten-fold logarithm to the base ten of the ratio of the squared RMS sound pressure level for a given time-interval to the squared reference sound pressure.

NOTE 1 The sound pressure level $L_{eq,T}$ is given by the following equation:

$$L_{eq,T} = 10 \lg \frac{\frac{1}{T} \int_{t_1}^{t_2} p^2(t) dt}{p_0^2}$$

where

$p(t)$ is the instantaneous sound pressure at time t ;

t is the integration variable for time;

$T = t_2 - t_1$, is the length of the time interval, for which the continuous sound pressure level is determined and

p_0 the reference sound pressure (20 µPa).

The numerator in the argument of the logarithm in the given equation is the RMS sound pressure for the averaging time T .

NOTE 2 As a matter of principle, no time-weighting is applied in the determination of the continuous sound pressure level.

NOTE 3 For further definitions, see IEC 61672 [3].

**3.43
percentile level**

ten-fold decimal logarithm of the ratio of the squared RMS sound pressure level being exceeded for a given part of the measurement time to the squared reference sound pressure where the RMS sound pressure is determined with a standardized time and frequency weighting, e.g., L_{10} or L_{A10}

NOTE 1 For application within the framework of this standard, the time-weighting “Fast” is to be applied for the determination of the percentile level.

NOTE 2 L_{A10} is the A and Fast-weighted sound pressure level being exceeded in 10 % of the measurement time.

**3.44
modulation transfer index
MTI**

unweighted mean of the scaled effective signal to noise ratios for a given octave band

**3.45
operational speech level**

sound pressure level of speech signal that will be used or is found in the applicable situation

3.46

operational background noise level

sound pressure level of background that will be present or is found in the applicable situation

NOTE This level is used for predictions and post-processing of measurements.

4 Description of the STI method

4.1 General

4.1.1 Rationale for the STI method

The STI method was developed as a fast and objective test method for determining the speech transmission quality of speech transmission channels. Using the speech transmission index, the speech intelligibility can be predicted for different types of word and sentence formats for a wide range of speech transmission systems.

In speech, the intensity of the signal varies with time producing a variation in the intensity envelope of the speech. Slow fluctuations of the intensity envelope correspond with word and sentence boundaries while fast fluctuations coincide with individual phonemes within words. Within the STI concept, preservation of the intensity envelope is considered to be of the utmost importance.

In contrast to the original approach of the articulation index [4], which is based on the signal to noise ratios in different speech spectral bands, the STI measurement determines the degree to which the intensity envelope of the speech signal is affected by a transmission channel. A Modulation Transfer Function (MTF) is determined which quantifies how the channel affects the intensity envelope of the speech signal.

The STI produces a metric on a scale of 0 to 1, based on weighted contributions from a range of frequency bands present in speech.

The STI method and its derivatives (see below) can be used to determine the potential intelligibility of a speech transmission channel at various locations and for various conditions. In particular, the effect of changes in the acoustic properties of spaces can be assessed.

4.1.2 Applicability of the STI method

The STI method is an objective and validated measure of speech transmission quality for a wide range of acoustic and electro-acoustical distortions that influence intelligibility. However, as it is a simplification of human speech, the STI model can be limited in its applicability. Users of the STI method that apply the method beyond its current limits may obtain inaccurate intelligibility predictions. An overview of the applications and limitations is therefore given that aims to help STI users decide on which STI method is most suitable for their application, so as to obtain the most meaningful and accurate results.

The STI method was validated for an acoustic output using a single omnidirectional microphone. The use of a directional microphone produces different and uncorrelatable results and is not normally advised. Further information is given in clause 7.10.

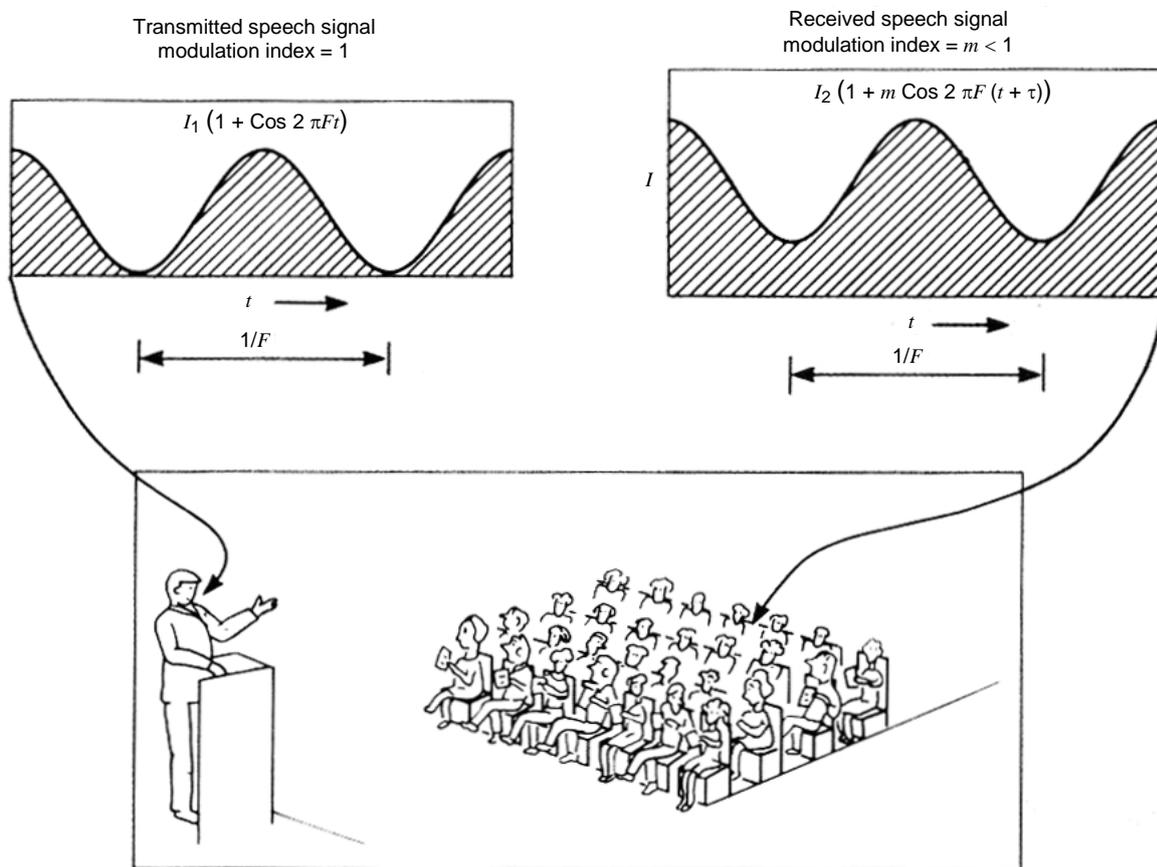
If the situation or the transmission channel does not allow the use of STI methods, alternative techniques for assessing intelligibility shall be used. Other methods exist to assess the quality of speech communication, each with their advantages and disadvantages and therefore have different users. Annex N describes a number of other measures of intelligibility.

4.2 Background of the STI method

4.2.1 General

The STI concept is based on the empirical finding that the fluctuations in speech signals carry the most relevant information relating to speech intelligibility, see [5], [6] and [7]. Fluctuations in speech result from the acoustic separation of sentences, words and phonemes, which are the fundamental elements of speech. The fluctuations, termed modulations, can be quantified as a function of modulation frequency F producing the modulation spectrum. For clear speech, the modulation frequencies typically extend from 0,5 Hz up to 16 Hz with maximum modulation at approximately 3 Hz.

Any deterioration of the modulation spectrum by the transmission channel is generally considered to result in a reduction of the speech intelligibility. This deterioration of the modulation spectrum corresponds to a reduction of the modulation depth at one or more modulation frequencies and is calculated as a modulation transmission value for each octave band over the speech spectral range. Figure 1 shows the concept of the reduction in modulation that can occur between a talker and listener.



IEC 1148/11

Figure 1 – Concept of the reduction in modulation due to a transmission channel

The STI method has been optimised and validated with subject-based intelligibility experiments using CVC (Dutch)-word scores for a large variety of distortions in transmission channels. Such distortions include noise, reverberation, echoes, non-linear distortion, and digital encoding techniques.

Using parameters derived from speech material, the STI test signal was developed. In general, the STI test signal comprises seven octave band noise signals corresponding with the octave bands from 125 Hz up to 8 kHz. Each noise carrier is modulated with one or more modulation frequencies at one-third octave intervals ranging from 0,63 Hz up to and including 12,5 Hz.

The STI method, described in Annex A, determines the modulation transfer function $m(F)$ of the transmission channel. A total of 98 results are obtained, corresponding to the 14 modulation frequencies and the seven octave bands (see Figure A.3). The RMS level of each octave-band carrier matches the relative level of the average, long term spectrum of speech material (see also 4.5 for further information). Each octave band has a contribution to speech intelligibility which is weighted according to that band. Using the weighted sum of these transmission index values, the overall STI value for the transmission channel is determined.

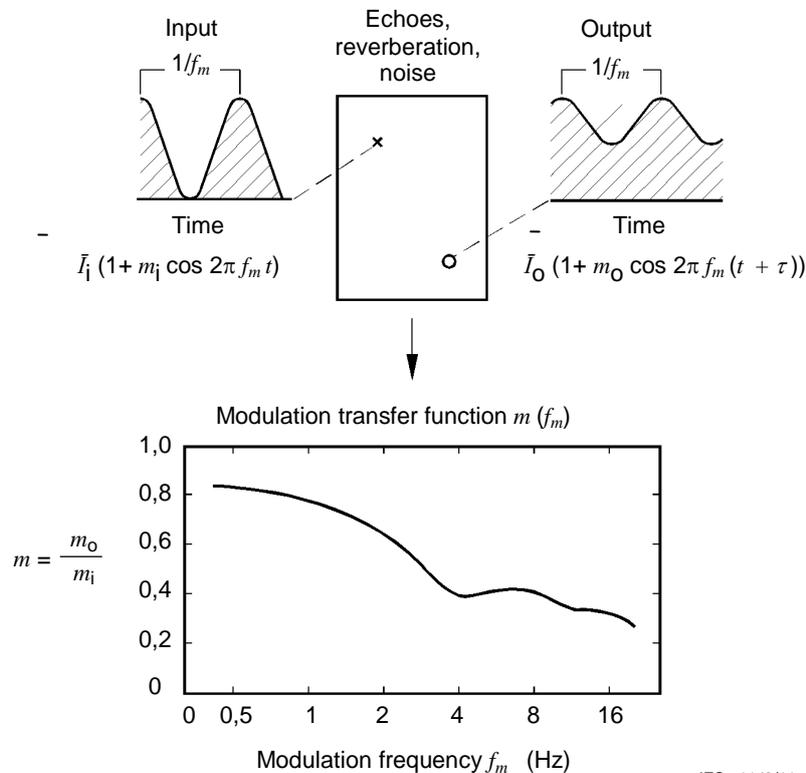
Research [4] has shown that adjacent octave bands contain redundant information with respect to speech intelligibility. If one octave band does not contribute to intelligibility (e.g. by masking from reverberation or background noise) then neighbouring octave bands can partly compensate for this missing contribution. This insight has led to the use of redundancy factors in the STI-methodology.

4.2.2 Theoretical overview

The modulation index m_i of a test signal is played into a room or through a communication channel and received at a listener position as the modulation index m_o . To measure, for example, the STI for the situation in Figure 1, the test signal would be transmitted by a sound source simulating a human talker situated at the talker's position with a receiving test microphone located at any listener position.

For the sound source, the important characteristics are physical size and directivity, position, sound pressure level and frequency response.

The typical test signal consists of a carrier with a speech-shaped frequency spectrum and a sinusoidal intensity modulation with modulation frequency f_m (see Figure 2).



IEC 1149/11

NOTE m_i and m_o are the modulation indices of the input and the output signals, respectively. \bar{I}_i and \bar{I}_o are the input and output intensities, the intensities being equal to the square of the sound pressure levels (p^2).

Figure 2 – Modulation transfer function – Input/output comparison

The reduction in the modulation depth at frequency f_m is quantified by the modulation transfer function $m(f_m)$ which is determined by

$$m(F) = \frac{m_o(F)}{m_i(F)}$$

and is interpreted in terms of an effective signal-to-noise ratio SNR_{eff} (irrespective of the cause of the reduction which can be reverberation, echoes, non-linear distortion components or interfering noise). It is determined by

$$SNR_{\text{eff}} = 10 \lg \left(\frac{m(F)}{1 - m(F)} \right)$$

The values of the effective signal-to-noise ratio are limited to the range of -15 dB to $+15$ dB. Values less than -15 dB are given the value of -15 dB and values greater than 15 dB are given the value of 15 dB.

The speech transmission index STI combines the modulation transfer index values from measurements in seven octave bands into one overall weighted value.

Annex A provides a more detailed description of the calculation of the speech transmission index.

4.2.3 Measurement of STI

The FULL STI is based on a complete set of 98 (7 x 14) modulation indices.

Two simplified forms of the STI, based on measurements using a lower number of modulation indices, are STIPA and STITEL (see Clause 5).

STIPA consists of a test signal with a predefined set of two modulations per octave band that are generated simultaneously giving a total of 14 modulation indices.

STITEL consists of a test signal with a predefined set of seven modulation frequencies, one per octave band, that are generated simultaneously giving a total of seven modulation indices.

There are two methods to measure STI:

- direct methods using modulated test signals;
- indirect methods based on the system's impulse response using the Schroeder equation.

The direct methods using STIPA and STITEL have substantially shorter measurement durations than the direct FULL STI. Note that the direct FULL STI is rarely used in practice.

Annex B and Annex C provide detailed descriptions of STIPA and STITEL, respectively. Annex D provides details about the now obsolete method RASTI.

The STI method, whether direct or indirect, has been proven to give valid results for a great number of linear distortions in both the time and frequency domains. The following distortions are accounted for by the STI method:

- temporal distortion, e.g. reverberation and echoes;
- noise;
- strong spectral distortion e.g. band-pass filtering.

NOTE Some types of spectral distortions may not be accounted for, see 4.5.8.

In addition, the direct STI methods account for non-linear distortion, e.g. clipping, whereas the indirect methods should only be used for linear systems. Additional information about the effects of non-linear distortion is given in Clause 6. Table 1 gives an overview of the STI test methods versus the types of linear and non-linear distortion for which they are appropriate.

Table 1 – Comparison of STI test methods for different types of distortion

Method	Type of distortion			
	Noise	Reverberation, echoes	Non-linear distortion	Spectral distortion ^a
Direct FULL STI	yes	yes	condition dependent	yes
Direct STIPA	yes	yes	condition dependent	yes
Direct STITEL	yes	condition dependent	condition dependent	yes
Indirect FULL STI using MLS ^e	yes ^b	yes	no	yes
Indirect FULL STI using swept sine signal ^c	no ^d	yes	no	yes

NOTE The term 'condition dependent' is used to indicate that the corresponding test signal type may or may not produce sufficiently accurate results, depending on the exact distortion type. For example:

- centre clipping is unlikely to have any effect on the modulation depth, whereas peak clipping reduces the modulation depth but generally has little effect on the intelligibility of speech, so the measured STI value may be pessimistic;
- STITEL can be used in reverberant environments, provided that the reverberation time is not largely dependent on frequency;
- similarly, STIPA can be used for PA systems that produce non-linear distortion components, unless the signal is severely clipped in various frequency bands.

^a The frequency response of the transmission channel may produce a perceived loss of intelligibility that is not adequately accounted for in the result, see 4.5.8.

^b Signal averaging of time domain data shall not be used and the excitation spectrum shall be speech-shaped.

^c This includes time delay spectrometry.

^d However, the effects of noise may be computed mathematically.

^e Theoretically, other mathematically deterministic pseudo-noise (random phase) signal could be employed.

4.3 Applicability of STI test methods

Table 2 provides an overview as to which forms of STI are recommended for various types of application. The + and – symbols are a general indication of the suitability of the method.

If significant parts of the listener population are non-native and/or older listeners, the STI should be interpreted as noted in Annex H.

Table 2 – Applicability of test

Application	Recommended test	FULL STI ^a	STIPA	STITEL	Limitations	Work-arounds
Assessing suitability of room acoustics for speech communication (no electronic amplification)	STIPA	++	++	+/-	Suitability of STITEL depends on reverberation	
Evaluating PA and VA systems	STIPA	+	++	+/-	Suitability of STITEL depends on reverberation and echoes	
Evaluating telecommunication channels (phone, radio)	STITEL	+	+	++	STITEL has more diagnostic power	
Channel features amplitude compression	STIPA	+	+	+		
Difference between male and female voices needs specific attention	FULL STI	++	-	+	STIPA not suitable for female (male spectrum only)	
Strong centre clipping	None	-	-	-		none
Strongly fluctuating noise	STIPA	+/-	+/-	+/-		Report several STI measurements
Speech and noise clearly spatially separated or a strong direct-field component exists in a highly reverberant environment	STI	+	+/-	+/-	To be used with caution. Currently standardised methods are inaccurate.	See 7.10
Channels that do not permit artificial test signals, such as vocoders	None	+/-	+/-	+/-	Currently standardised methods are inaccurate.	^b Use a speech-based STI test signal or listener tests
++ very well suited method, + well suited method, +/- suitable method, - not a suitable method ^a See Table 1 for suitability of measurement methods. ^b This is a direct method and may be included in a future addition of the standard.						

4.4 Use of direct and indirect methods

Table 3 below compares a number of practical issues relating to the use of direct and indirect measurement methods.

Table 3 – Choice of method

Subject	Direct method	Indirect method
Post processing	possible	mandatory
Handheld device	possible	possible
Amplitude nonlinearities	reduce the reliability of the result	reduce the reliability of the result
Frequency nonlinearities (Uneven spectrum) ^a	possible	possible
Frequency shift	not possible	not possible
Noise suppression	no	yes
Sample rate accuracy between the clock frequencies of the signal source and receiver	errors less than 20×10^{-6}	errors less than $0,5 \times 10^{-6}$
^a See 4.5.8 for further details.		

4.5 Limitations of the STI method

4.5.1 General

The STI method would ideally reflect all the changes in a transmission channel that are relevant to speech intelligibility. However, it is important to realise that the STI modelling approach is still a simplification of human processing. Also, the STI test signal differs from human speech in more subtle temporal and spectral aspects, such as:

- the dynamic range of speech, which depends on the integration time;
- the energy distribution in each time frame;
- the distribution of signal levels over the entire length of a speech segment or test signal (percentile exceedances);
- the lack of gaps in the test signal;
- the carriers in speech are not restricted to the fixed carrier bands and modulation frequencies;
- the spectral differences between individual words and the STI signal;
- the spectral differences between various talkers.

NOTE The speech spectrum specified for STI differs from the spectrum specified by ANSI [4].

As a consequence, for certain situations and possible (narrow-band) transmission channels, care shall be taken when using the STI. In some cases, intelligibility may suffer little from a distortion, whilst the STI shows a significant reduction. In other cases, in which the STI shows only minimal changes, the intelligibility can be considerably reduced. The following clauses discuss potential limitations in more detail.

4.5.2 Frequency shifts

This type of distortion may occur with

- playing a digital signal at the wrong sampling rate,
- devices for preventing acoustic feedback,
- single sideband radio transmissions.

Frequency shifts can have a large effect on STI with generally little effect on intelligibility, so the STI may underestimate intelligibility for systems with frequency shifts.

4.5.3 Centre clipping

This type of distortion may occur when low-level parts of a signal are not transmitted faithfully or are silenced. This could happen in amplifiers and corroded connectors. The STI overestimates the intelligibility for systems that show effects of severe centre clipping.

NOTE Centre clipping is also known as crossover distortion and origin distortion.

4.5.4 Drop outs

Signal drop-out at regular intervals can result from selective fading patterns in wireless transmissions and corruption of digital signals. The STI may not be reduced much, but intelligibility may be very poor. Analysis of the fine structure of the received modulated signal is recommended in order to flag drop-outs and where possible allow computation of the STI with the drop-outs removed.

4.5.5 Jitter

Time shifts of speech, as applied in digital signal transmission to compensate for variation in transmission rate, have no effect on intelligibility but can severely reduce the STI, so the STI may underestimate intelligibility for systems with jitter.

4.5.6 Vocoders

Although digital voice coders have little influence on intelligibility, depending on the type of codecs used, the STI tends to be increased. In situations with low intelligibility, the use of speech based test signals or subject-based measures is recommended.

STI should not be used to measure systems such as vocoders that encode speech segments. For example, linear predictive coding techniques which might use code-book related synthesis or the introduction of errors related to voiced/unvoiced speech fragments and pitch errors.

4.5.7 Overestimation of STI under low background noise conditions

It should be noted that the STI model inherently assumes a non-infinite signal to noise ratio in each octave band as the hearing reception threshold in the model operates as a source of background noise. If the background noise levels or the reception threshold values are set to zero during measurements or simulations, STI values may be too high.

As an example, this issue arises when mathematically investigating the behaviour of STI with changes in the spectrum of the input signal. If an MTF matrix with every value at unity (i.e. no contamination from reverberation or background noise) is used with an input signal that deviates from the specified speech spectrum, the STI result often shows little change, even with large changes in the input spectrum, see [8].

It is therefore essential that STI predictions and measurements should always incorporate a level of background noise that is realistic for the application. For example, measurements with an acoustic output should use a realistic background noise as well as the speech reception thresholds.

4.5.8 Frequency response

Research so far [8], [9], [10], [11], [12], [13] indicates that the frequency response of the transmission channel (which is manifest as the perceived tonal balance of speech) is much more important for perceived intelligibility than is indicated by STI measurements, especially in the presence of reverberation. If the frequency response is not reasonably flat, it is possible that the STI can indicate values that are too high compared to the perceived intelligibility.

Systems with measured STIs exceeding 0,5 have been reported where the perceived speech intelligibility has been found to be inadequate due to the poor frequency response / tonal

balance of the system. The application of equalisation to improve the frequency response improved the perceived intelligibility.

Acknowledging this limitation of the STI-method, a suitable solution for ensuring an even amplitude response is to perform a separate measurement of the amplitude versus frequency response of the system, preferably at a higher resolution than one octave bandwidths. Nonetheless, there are significant factors that may be not included in such measurements.

- The frequency response deduced from impulse response data is highly dependent on the length of time data used for the measurement and the time window that is applied to that data.
- There is no measure that is well-correlated to the perceived tonal balance for a variety of acoustical environments. For example, in low-reverberation situations, the influence of the direct field response on the tonal balance is typically much higher than in very reverberant environments, where the power response of the source becomes more dominant.
- The influence of varying talker position on the microphone's frequency response.

Some sound-system practitioners have indicated that small changes to the frequency response of sound systems that reduce the audible coloration of speech can reduce the degree of concentration that a listener needs to exert to achieve satisfactory intelligibility. This can be particularly important in long term listening situations or in the case of a non-native talker or listener. Examples of colorations include the presence of narrow band peaks or resonances where adjustments to the system of as little as 1 dB over a bandwidth as narrow as 1/3 octave have proved beneficial to the resultant perceived intelligibility [9], [10].

4.5.9 Echoes

Situations have been encountered in which audible echoes (late reflections) cause significant loss of perceived speech intelligibility whilst the corresponding measured STI values are significantly higher than the perceived intelligibility would indicate.

NOTE This issue is the subject of ongoing research, see e.g. [8].

In situations with audible echoes, other diagnostic acoustic methods should be used to measure and assess the severity of the echo.

4.5.10 Fast amplitude compression and expansion

Measured STI and STIPA values may be altered whenever compression or expansion is applied to the test signal. However, experience shows that only minor changes in perceived intelligibility occur with appropriate compression or expansion. It is also noted that compression schemes generally alter the perceived tonal balance of speech which in turn may adversely affect the perceived speech intelligibility.

When properly implemented, companders (complementary compression and expansion devices) are likely to have no overall effect on intelligibility.

Fast compression reacts on the instantaneous amplitude envelopes of a range of frequency bands. With this compression, signal level variations above the compression threshold level (knee point) are reduced according to the compression ratio. As compression reduces the dynamic range of the signal, the modulation depth may also be reduced.

On the other hand, automatic gain control (AGC) has a fast reaction time, but a very slow recovery time and does not reduce the short-term dynamic range.

Compression and AGC techniques are often applied to improve speech intelligibility (e.g. for the hearing impaired who suffer from a limited dynamic range) and can also be applied in public address systems.

Sentence intelligibility as measured by the speech reception threshold (SRT) has been found to increase by up to an equivalent of a 4 dB change in effective SNR, but this is dependent on the amount and type of compression.

The effect of compression on intelligibility at high signal and noise levels, such as in public address systems, awaits the outcome of further research.

4.5.11 Non-linear distortion

Although the STI is sensitive to distortion, the result is highly dependent on the measurement method adopted. (This is discussed further in 6.3.)

4.5.12 Impulsive and fluctuating noise

Two types of background noise should be distinguished in STI measurements:

- impulsive;
- fluctuating.

Impulsive noise and undesired short events, such as a hammer dropping, result in inaccurate STI results, especially with narrow band transmission, as well as in the incorrect diagnosis of the contribution of frequency bands.

Fluctuating noise, such as babbling voices or machinery that is repeatedly turned on/off or is cyclical, can lead to variations in the STI value obtained for repeated measurements and may also lead to considerable underestimation or overestimation of intelligibility measurements.

Subjectively, the intelligibility of sentences in fluctuating noise is known to be higher than in stationary noise with the same time-averaged RMS output [14].

If STI measurements are conducted in the presence of impulsive or fluctuating noise, then the indirect method (described in Clause 6) should be used. Signal averaging with MLS or slow sine-sweeps should be used to reduce the noise in the measurement. The degrading effects of the noise can then be added back to the MTF by post-processing the 'noise-free' MTF data.

When using sine-sweeps to determine the STI, a noise-free measurement is required. For practical purposes, a noise free-measurement is obtained if the SNR in each octave band is at least 20 dB. 7.8.3 provides further information.

4.5.13 Hearing impaired listeners

Without specific corrections, the STI method is not a reliable predictor of the intelligibility of speech for hearing-impaired listeners [15]. The measurement of hearing assistive systems or channels is possible, though specific corrections may be also required [16].

4.6 Conclusion

In general, the STI method is a conservative approach and may underestimate intelligibility in some applications, but there are exceptions such as given in 4.5.3.

5 Direct method of measuring STI

5.1 Overview

STI may be measured either directly using a suitably modulated signal or indirectly by means of mathematical manipulation of a system impulse response using a relationship proposed by Schroeder [17].

The research described in [4], [5], [18], [19] and [20] developed the basis and method for the FULL STI. From subsequent research came the two current simplified forms, STIPA and STITEL, which require less measuring time. RASTI was also developed but is now obsolete.

FULL STI – consists of 98 separate test signals using 14 different modulation frequencies for seven octave bands. Each test signal contains only one modulation frequency for only one octave band noise carrier; the other octave bands contain no signal. The test signals are generated sequentially. With an average of 10 s per test signal, a FULL STI measurement requires approximately 15 min. An alternative version of the Full STI signal contains random modulations in the other octave bands in addition to the modulation frequency and octave band under test.

STIPA – consists of only one test signal with a predefined set of two modulations in each of the seven octave bands. The 14 modulations are generated simultaneously. One measurement takes between 10 s and 15 s.

STITEL – consists of only one test signal with a predefined set of seven modulation frequencies, one per octave band, that are generated simultaneously. One measurement takes approximately 12 s. STITEL may be used for its higher sensitivity (see Annex C) but great care needs to be taken in its use.

RASTI – consists of only one test signal with a predefined set of nine modulation frequencies that are generated simultaneously, five for the 2 000 Hz octave band and four for the 500 Hz octave band. One measurement takes approximately 30 s. RASTI is now obsolete, but for completeness, details are given in Annex D.

Table 2 compares the accuracy of the two simplified test signals with that of the FULL STI for various test conditions.

For an STI to take account of the operational signal-to-noise ratios and the absolute speech level, the mean intensity of the test signal should be equivalent to the normal speech level at the test position. Applying the method described in Annex J, the L_{Aeq} of the test signal is adjusted to be 3 dB A greater than the typical L_{Aeq} of the measured continuous speech at the test position (i.e. a 3 dB correction factor needs to be added).

5.2 STIPA

The STI test signal can be simplified if the uncorrelated (or speech-like) modulations that are required for the accurate interpretation of non-linear distortions are omitted [21]. This allows simultaneous modulation and parallel processing of all frequency bands, thus reducing measurement time, but this reduces the ability to account for some forms of non-linear distortion, as noted in Table 1. For each octave frequency band the modulation transfer function is determined for two modulation frequencies.

The STIPA method, described in Annex B, employs this simplification and has a measurement time of between 10 s and 15 s. The STIPA method is suitable for the measurement of natural speech (room acoustic transmission) as well as sound systems.

The designation STIPA refers specifically to a modulated, speech shaped signal (as described in Annex B). If STIPA is derived from an impulse response, for example by prediction, this shall be clearly stated and the designation STIPA(IR) shall be used to avoid confusion. It should be noted that the standard STIPA signal is based on a male speech spectrum.

Without specific corrections, the STIPA method is not a reliable predictor of the intelligibility of speech for hearing-impaired listeners [15]. The measurement of hearing assistive systems or channels is possible, though specific corrections may be also required [16].

5.3 Application

The direct STI method can be applied to almost any digital, analogue, electro-acoustic and acoustic speech transmission channel. With the determined STI-value, the intelligibility of different types of speech material can be predicted for many types of transmission systems.

For all tests in which reference is made to this standard, the relevant parameters and results should be stated in a measurement report sheet. A sample report sheet is given in Annex K.

5.4 Limitations

In addition to the limitations of the STI method described in Clause 4, there are a number of other limitations to the direct method of measuring the STI.

As the test signal is band-limited random or pseudo-random noise, repetition of measurements does not normally produce identical results, even under conditions of steady interference. The results centre on a mean with a certain deviation. This depends, amongst other factors, on the number of discrete measurements of the modulation transfer function (usually 98 for the STI method or 14 for STIPA) and the measuring time involved.

Typically, with FULL STI, the maximum deviation is about 0,02 STI for a measuring time of 10 s for each modulation index $m(f_m)$ and with stationary noise interference. With STIPA and a measurement time of 15 s, the maximum deviation is approximately 0,03 STI for repeated measurements.

With fluctuating noise (for example, a babble of voices), higher deviations may be found, possibly with a systematic error (bias). This can be checked by carrying out a measurement in the absence of the test signal, which should result in a residual STI value less than 0,20. An estimate of the deviation should be made by repeating measurements for at least a restricted set of conditions.

It is therefore good practice to average the STI results over two or three measurements for a specific condition.

6 Indirect method of measuring STI using the impulse response

6.1 Overview

The modulation transfer function MTF, as the basis of the STI, can also be computed from the impulse response of a transmission channel, using the process known as the Schroeder method [14]. The impulse response is acquired (usually by computer-based equipment) and the MTF derived from which the STI is subsequently calculated.

The following equation (of which the first factor is the Schroeder equation), should be used to calculate the modulation transfer function $m_{f,k}$ at modulation frequency f_m in octave band k .

$$m_k(f_m) = \frac{\left| \int_0^{\infty} h_k(t) e^{-j2\pi f_m t} dt \right|}{\int_0^{\infty} h_k(t)^2 dt} \cdot \left[1 + 10^{-SNR_k/10} \right]^{-1}$$

where

$h_k(t)$ is impulse response of octave band k ;

f_m is the modulation frequency;
 SNR_k is the signal-to-noise ratio in dB.

The indirect method is only applicable to linear, time-invariant systems.

Considerable experience is required to use this method as the measurement systems allow a variety of parameters to be adjusted, which may affect the final result.

This method is also applicable to the simplified forms of STI. As the processing time of this technique is quite short, it is recommended to calculate the FULL STI. However, calculation of the shorter derivatives of STI can be useful.

STIPA values derived from impulse response measurements shall be termed STIPA(IR).

6.2 Application

When deriving STI values from impulse response measurements, it is usual to make a noise free measurement and then correct this for the effects of background noise and speech level. However, techniques are available that enable the effects of background noise to be directly accounted for within the measurement, for example, through the use of a speech shaped MLS signal without averaging. Measurement procedures used for determining the impulse response shall meet the following requirements among others, with further information provided in ISO 18233.

- a) Measurements of the impulse response shall be conducted in accordance with ISO 18233.
- b) The length of the acquired impulse response shall be at least 1,6 s and not less than half of the reverberation time of the room.
- c) In order to produce a “noise-free” impulse response, a SNR of at least 20 dB should be obtained in all seven octave bands. If necessary, signal averaging can be used to achieve this.
- d) The use of excitation signals with a white frequency spectrum (e.g. as with Time Delay Spectrometry, TDS or Maximum Length Sequences, MLS) should be avoided under normal circumstances unless the background noise level is very low. A pink frequency spectrum (–3 dB/octave) produced with pink noise or logarithmic sine sweep (more rigorously, “exponential sweep”) is generally more suitable. However, a speech shaped MLS signal can also be used without averaging to measure the effect of background noise on the STI directly.
- e) Impulsive signals such as the Dirac function are not generally suitable when background noise, pass-band limiting and non-linear distortion are significant, since the average frequency spectrum and level distribution of typical speech are not represented in the test signal.
- f) The impulse response method is only applicable to linear, time-invariant systems. If the transmission channel has functions with non-linear signal processing, these functions should be bypassed during the speech intelligibility measurement. If, for instance, the effective playback sound pressure level is increased by a nonlinear reduction of signal dynamics, this shall be taken into account by separately measuring the maximum sound pressure level and applying an appropriate correction.
- g) Time variances due to movements of the air (wind) or climatic changes during the measurement process shall be avoided (they also invalidate averaging over longer periods of time). The average wind speed during MLS measurements, for example, should not exceed 4 m/s. Measurements using maximum length sequences (MLS) are more vulnerable in this respect than measurements performed with sine-sweeps.
- h) It should be ensured that the components involved in the transmission of sound (loudspeakers, room surfaces, reflectors, measurement microphone, people) do not move during the measurement cycle.

- i) Under critical conditions, the repeatability of the measurement results shall be proven by repeated measurements.
- j) The impact of background noise (L_n) and operational speech level (L_s) in each octave band k shall be incorporated into the result by post-processing (see Annex M).

6.3 Limitations (non-linear distortion)

In addition to the limitations of the STI method described in Clause 4, there are a number of other limitations to the impulse response method of measuring the STI of which non-linear distortions are of special importance.

Distortions of the measurement signal should be avoided as the indirect method does not correctly account for the effects of distortion. When this method is used, the sensitivity to distortion strongly depends on the measurement procedure applied [13], [22]. Fourier transform based methods, for example, are only error-free for linear systems.

Critical analysis is therefore required of how the impulse response is obtained and potentially influenced by non-linearities in the transmission system, particularly as in practice, system components can be operated at the limits of their performance range. When using sine sweep techniques, the non linear distortion components appear at the beginning or end of the recovered impulse response and so can be evaluated. However, errors may arise if the reverberation time is long, as the reverberant tail of the distortion components may smear into the main impulse response.

When using an MLS signal, distortion components tend to appear as noise and are not so readily discernable. DC components and time aliasing artefacts occur as pre-arrivals (pre-echoes) before the arrival of the signal.

When using a sine sweep technique, the distortion components that are inherent within the method shall be edited out or removed from the IR before calculation of the STI can be undertaken.

7 Measurement procedures, post-processing of data and applications

7.1 General

Although STI measurements are normally performed acoustically, in certain situations it is not always possible or necessary to use acoustic excitation or perform acoustic measurements. For example, in situations when different systems are rated with respect to their speech transmission quality or more diagnostic information is needed, the test signal may be injected and/or received electrically.

It is essential that in any post-processing of the MTF matrix, a realistic level of background noise is used [8]. If the output of the transmission channel is acoustic, the hearing reception threshold shall be used as a minimum.

All relevant parameters should be stated in a measurement report. A sample report is given in Annex K.

7.2 Acoustical input

Applying the test signal via a special loudspeaker (see below) to the microphone of the system under test ensures that factors at the microphone location that could reduce intelligibility (such as ambient noise or feedback, for example) are taken into account. In addition, some electro-acoustic systems do not have any alternative way of injecting the test signal. As this procedure requires the test signal to be reproduced acoustically, it is necessary to use a specific loudspeaker (e.g. an artificial mouth) that emulates a natural talker.

Correct adjustment of the test signal spectrum to match the standard speech spectrum is also required for electrical injection of the test signal. When using the direct method, the standardized test signal shall be used for this purpose.

The following procedure shall be used.

- a) Verify the integrity of the test signal (e.g. via means of a loop back measurement). This is particularly important if the test signal is generated from a CD player, although PCM (e.g. .wav file) generators should also be checked. (Digitally compressed signal formats e.g. MP3 should not normally be used, though compression schemes employing at least 128 kbit/s have been shown to work without apparent error). Further information is available in [9].
- b) Verify that the 1/3 octave frequency response of the test signal source (artificial mouth or suitable test loudspeaker) is within ± 1 dB over the required frequency range when measured in a free field (free of reflections) for either of the following measurement techniques:
 - over the range 88 Hz to 11,3 kHz using a FULL STI or MLS or other impulse response measurement signal (the limits of the 125 Hz and 8 kHz octave bands) or
 - individual octave band levels over the range 125 Hz to 8 kHz when using a STIPA or other speech shaped test signal.

NOTE 1 For indirect measurements, the frequency response derived from an MLS or other impulse response measurement can be processed to calculate an octave-band spectrum.

If necessary, adjust the equalisation (if any) of the artificial mouth or test loudspeaker to satisfy this requirement.

- c) Set the source on the axis of the appropriate microphone at the appropriate talker position / distance and direct it in the normal speaking direction.

In the absence of an artificial mouth, a suitable transducer, such as a small, single-source, high-quality loudspeaker (cone diameter not exceeding 100 mm), may be used and shall be described with the results.

NOTE 2 Generally, in a listening space, speech intelligibility depends upon the directivity of the source; therefore, a mouth simulator having similar directivity characteristics to those of the human head/mouth (see ITU-T Recommendation P.51 [48]) should be used when assessing the intelligibility of unamplified talkers. However, the directional characteristics of the test source (talker simulator loudspeaker or mouth simulator) can be of significant importance when making measurements in large or reverberant spaces, or when the pick-up microphone is located at some distance from the talker [16], [23].

When speech is relayed through a sound system, a simulator may not be required. However, where the source microphone is situated in either a reverberant or noisy location or if a close talking or noise cancelling microphone is involved, then either a talker or mouth simulator should be employed.

- d) Set the test signal level at the microphone position to the operational speech level that will be used in the system. The speech and test signal levels shall be matched according to the method described in Annex J.

NOTE 3 This test is likely to stress the amplifier. See 14.9 of IEC 60268-3 [24]. It may be convenient to apply the test signal for 1 min, for example, followed by several minutes of zero signal to allow cooling to take place.

In the absence of a correct match between the test signal level and the operational speech level, a default equivalent level of 60 dB A at 1 m in front of the artificial mouth or test loudspeaker should be used for the source.

Smaller talker distances typically result in speech levels of approximately 86 dB A to 94 dB A for handheld microphones (distances of 5 cm to 2 cm), while speech levels of approximately 80 dB A to 86 dB A result for gooseneck microphones (distances of 10 cm to 5 cm).

NOTE 4 The above levels are subject to wide variations in practice.

- e) Run the STI, STIPA (or STITEL) test sequence. Normally, and where available, the “with noise” option should be selected.
- f) If an MLS signal is used to measure the impulse response and if it is required to take account of the background noise, the excitation spectrum should be adjusted to the

standardised speech spectrum by appropriate filtering. Signal averaging should be disabled or a single sequence should be employed [25].

- g) If sine-sweeps, MLS or TDS are used to determine the noise-free impulse response, appropriate adjustments to speech and noise levels at both the microphone and receiver locations shall be applied to the noise-free MTF by post processing.
- h) The test signal shall be fed into the system in such a way as to ensure that all signal processing components relevant for speech reproduction (equalisers, signal delays, etc.) are correctly taken into account during the measurement process.

7.3 Acoustical output

The measurement device (microphone/artificial ear/head simulator) shall be acoustically calibrated with respect to sensitivity and frequency response. Measurements shall be performed at the listener's normal location and listening height (or at a specified listening height). If a single microphone is used, it shall be Omni-directional and of diffuse field type.

7.4 Electrical input

Follow the above procedure in 7.2, replacing step d) by the step below, and selecting the injection point for the signal to be as close as possible to the normal signal input, so as to include as much of the system as possible in the test.

The STI test signal at the point of injection shall be adjusted to be equivalent to the level of speech at that point. The speech level is determined using the speech level measurement method as described in Annex J.

7.5 Electrical output

Since no acoustic conditions are involved at the electrical output, hearing-related effects, such as masking and the reception threshold, shall be disabled on the measurement device. If this is not possible, the electrical input to the measurement device shall be adjusted to simulate a sound pressure level well above the reception threshold but below a point where level-dependent masking becomes noticeable in the STI results. Broad band output levels should be A-weighted and then reported as A-weighted voltage levels in dB relative to a stated reference, e.g. 1 V.

7.6 Examples of input/output combinations

7.6.1 Acoustical input – Acoustical output

In the normal STI measurement set-up for PA systems and in auditoria, a sound source is used to acoustically generate the STI test signal. The test signal level is calibrated and corresponds to the nominal speech level. A situation-dependent and representative talking distance should be employed as described in 7.2. A calibrated STI measuring device is used at the receiver location to determine the STI of the transmission channel.

7.6.2 Electrical input – Electrical output (e.g. assessment of wired and wireless communication systems)

Purely electrical STI measurements are generally performed to rate different communication systems with respect to their speech transmission quality rather than to obtain an absolute value for the speech intelligibility. It is advisable to perform these measurements at different input signal levels (e.g. from –10 dB to +10 dB relative to the reference operational level) to gather information of the influence of the dynamic range, noise floor and signal processing capabilities on the intelligibility of speech. These types of measurements are likely to be conducted on wired or wireless speech transmission systems such as telephone lines and radio communication systems.

7.6.3 Acoustical input – Electrical output (e.g. assessment of microphones)

To compare microphones with respect to their effect on the intelligibility, the STI test signal level at the microphone should be calibrated as given in 7.2. Measurements are performed in combination with the appropriate ambient noise spectrum and as a function of the noise level to determine the microphone's noise rejection behaviour. Preferably, measurements should be made at different speech levels to examine the effect of a lowered or raised voice on the intelligibility.

NOTE Special methods may be required when measuring the STI of assistive hearing systems and Audio Frequency Induction Loop Systems (AFILS), in particular [16]. However, much of the general guidance given in this clause is applicable.

7.6.4 Electrical input – Acoustical output (e.g. assessment of PA systems)

To compare different transducers (loudspeakers, headsets), the STI test signal can be electrically injected. The test signal shall be reproduced at the listeners' location at a sound pressure level that is representative of normal operation.

In the case of a public address or similar sound distribution system, the measurements should be performed for a representative number of locations. Taking a simple mean value of the results can be misleading. A better method, that takes account of the spatial variation in the results is the value obtained by computing the mean of the measured data minus one standard deviation. This is also sometimes known as the rating of the space and indicates that a given location will statistically have an 84 % probability or level of confidence in achieving a given target value (assuming a Gaussian distribution). A more precise method is to plot the complete statistical distribution of the results.

When assessing headsets, an in-ear microphone (MIRE) or an artificial ear should be used.

7.7 Post-processing of measured MTF data

There are a number of corrections that can be made to measured MTF data:

- elimination of noise from (de-noising) a measured MTF;
- addition of an occupancy noise level and spectrum;
- consideration of the hearing reception threshold;
- adjustment of the speech level and spectrum;
- correction for different reverberation times.

The effect of occupancy noise can be determined either

- a) by manually entering noise data into the noise data table used by the measuring equipment or
- b) by mixing an artificial or recorded noise signal of the correct spectral content and level with either the direct signal input to the analyser or a recorded signal.

Annex M gives an example of removing the noise from a measured MTF matrix and adding operational background noise and desired speech levels. The equations listed in Annex A are used for this process.

7.8 Issues concerning noise

7.8.1 General

As with all linear systems, the influence of distortions such as reverberation is independent of the amplitude response. Consequently, the variables that are dependent on the signal level are the signal to noise ratio in each octave band and the associated upward masking. Therefore, the STI method can be relatively insensitive to changes in the amplitude frequency response of the transmission channel, especially when the background noise is low.

When low levels of background noise are added to the MTF matrix, representing the noise levels that would occur in practice when using an electro-acoustic system, the overall STI shows more sensitivity to changes in the input spectrum.

An essentially noiseless situation, where only the auditory hearing threshold acts as a residual noise source, is usually not a realistic assumption for most practical cases. Even in quiet environments, such as libraries or court rooms, a residual noise level of 25 dB to 35 dB SPL is not uncommon and should be taken into account. This can be achieved by applying a suitable criterion, such as NCB, RC or NR curves (see [26]).

Undesired short events (for example impulsive noise) can be detected by analysing the statistics of the signal. However, it is easier in practice to repeat the STI measurement with the noise source physically eliminated or use the indirect method and averaging techniques.

Fluctuating noise is detected by measuring the direct STI in the absence of the test signal. If the STI is too high (e.g. $STI > 0,2$), the measurement results are likely to be erroneous. Preferably, the STI measurement should be carried out without the noise being present. The noise should then be separately measured (see 7.8.2), and the STI computed mathematically.

7.8.2 Measurement of background noise

To correct an STI measurement with regard to background noise, it is necessary to accurately characterize the background noise. The equivalent continuous sound pressure level (L_{eq}) of the background noise in each of the seven octave bands (125 Hz to 8 kHz) should be measured over a sufficient period of time. The positions, durations and times of the measurement shall be recorded together with the notes on unusual circumstances that may affect the validity of the measurements.

It should be noted that for the corrective calculation described here, it is not sufficient to determine a single broadband value for the background noise (e.g. $L_{A,eq}$) and to use a single broadband value for the speech signal (i.e. the operational speech level). Also see 7.8.3.

7.8.3 Fluctuating noise

If fluctuating noise cannot be eliminated, its influence should be minimised by amplifying the signal until it is at least 15 dB above the noise level in each octave band. From the modulation indices, calculate the STI based on the original signal levels before amplification. This method requires some computational skills.

If the influence of fluctuating noise cannot be reduced, measurements should be repeated at least three times before taking the average STI. If the spread is lower than 0,03 STI over the three repetitions, further repetition of the measurement is not necessary.

Interpretation of the speech intelligibility in the presence of fluctuating noise is extremely difficult and is currently beyond the scope of this standard. However, it has been found that listeners listen to speech in the gaps between the fluctuating noise and perceive a higher intelligibility than the STI would predict, based simply on the L_{eq} of the fluctuating noise.

NOTE If the fluctuation is great (e.g. 15 dB or more), it may be necessary to use the L_{10} in each octave band.

7.9 Analysis and interpretation of the results

It is important to examine the MTF data in each octave band to determine the reliability of the results, as follows:

- constant or slightly reducing modulation transfer ratio values as a function of modulation frequency indicate that noise is the dominant mechanism;

- modulation transfer ratio values monotonically decreasing with modulation frequency indicate that reverberation is the main mechanism;
- values that initially reduce and then increase with modulation frequency indicate the presence of strong reflections arriving later than 50 ms, which may produce an over-optimistic conclusion with regard to intelligibility. It is recommended that if this effect is detected, it should be reported with the result.

7.10 Binaural STI measurements

Although the STI is a well-accepted and standardized method for prediction of intelligibility, the STI model is essentially based on monaural listening. The advantages of binaural listening to speech intelligibility are disregarded by the model.

Subjectively, the binaural advantage might be significant. However, no clear measurement methods are available. The current STI method may produce an underestimation of intelligibility, especially if speech and noise arrive at the receiver from different directions. This issue is currently being researched.

When performing binaural STI measurements using an artificial head, the recommended approach is to use the STI results for the best ear. For further information, see [27].

8 Use of STI as a design prediction tool

8.1 Overview

During the design stage of a sound system, it is useful to predict the STI performance from the predicted room acoustic parameters. Two methods are available:

- calculation based on a predicted direct field, combined with an exponential reverberant decay and simple electro-acoustic parameters. Statistically calculated reverberation times may be used here;
- prediction based on a simulated impulse response of the system in the acoustic space.

Predictions based on simulated impulse response offer a higher degree of precision. This method is also preferred in cases where statistically-calculated reverberation times (Sabine/Eyring) are known to be in error, e.g. in coupled spaces, or spaces with uneven distribution of absorption.

It is critical that the operational speech level be used for prediction of the STI as this affects both the effective *SNR* and masking effects. A broadband speech signal shall be used for this prediction and shall ensure that the transmission channel is capable of producing the operational sound pressure level.

8.2 Statistical predictions

Prediction of the STI performance of a sound system shall be based on the MTF matrix that is calculated from the predicted room acoustic and electro-acoustic parameters and the measured or estimated background noise levels for each octave band contributing to the STI version chosen. Calculations shall use the method of Houtgast et al. [28] which is given in Annex L.

Access shall be available to the MTI values in each octave band and the octave band levels of the output speech signal.

If the prediction is made using commercially-available software, the results shall state:

- that a statistical estimate has been made using the method of Houtgast et al. [28];
- that the STI has been computed using the appropriate male or female weightings;

note that:

- RASTI shall not be used as an indication of the predicted STI;
- the STI shall not be estimated by converting a %Alcons value;
- the method of statistical prediction is even less sensitive than direct STI to the effects of strong discrete early and late arrivals and the possible loss of intelligibility due to poor frequency response.

8.3 Prediction from simulated impulse response

Prediction of the STI from a simulated impulse response shall be undertaken as follows:

- a) the MTF matrix shall be calculated using the Schroeder equation (see 6.1). The duration of the impulse response shall not be less than half the reverberation time and at least 1,6 s to ensure a reliable calculation of the modulation indices for the lowest modulation frequency of 0,63 Hz;
- b) both the hearing reception thresholds and the measured or estimated background noise sound pressure levels for each octave band shall then be introduced into the MTF matrix;
- c) the speech spectrum and operational speech level shall be selected and the auditory masking corrections listed in Table A.1 applied to the MTF matrix;
- d) the octave band specific male and female weighting factors given in Table A.3 shall be applied to the MTI values.

For each prediction location, access shall be available to the MTI values in each octave band and the octave band levels of the output speech signal along with the frequency response.

For predictions with multiple listener positions, the statistical properties and distribution of the results over the listening area shall be stated.

The results shall also state:

- that the STI has been calculated from an MTF derived from a predicted impulse response with the appropriate male or female weighting applied;
- the background noise levels which have been applied to the prediction.

Annex A (normative)

Speech transmission index (STI) and revised STI methods

A.1 Background

A.1.1 Envelope function and envelope spectrum

Connected discourse can be considered as a sequence of the smallest speech fragments, called phonemes. Each phoneme is characterized by a specific frequency spectrum. Clarity requires that the spectral differences of the phonemes are preserved. These spectral differences can be typified by the envelope function within a number of frequency bands. The envelope function describes the temporal fluctuations of the intensity of a speech signal within a certain frequency band. The shape of the envelope function is unique for a specific sequence of phonemes. Distortion of the speech envelope, such as by noise or reverberation, results in a reduction of the spectral differences between phonemes and this is reflected by a reduction in the degree of fluctuations of the envelope function.

Figure A.1, panel A shows an exemplary envelope function for the 250 Hz octave frequency band. The envelope spectrum gives a more general description of the fluctuations of the envelope function and results from a one-third octave-band spectral analysis of the envelope function. Typically a speech excerpt of 1 min length is analysed to give the spectral distribution of the envelope fluctuations about the mean intensity. This allows the formation of the modulation index as a function of modulation frequency as shown in Figure A.1, panel B, where the spectrum is normalized with respect to the mean intensity I_k .

A comparison of the envelope spectra obtained directly from the talker with the corresponding spectra obtained via a transmission path gives the reduction in fluctuations due to the transmission path. This reduction leads to the modulation transfer function or MTF, which represents the reduction of the modulation depth as a function of modulation frequency.

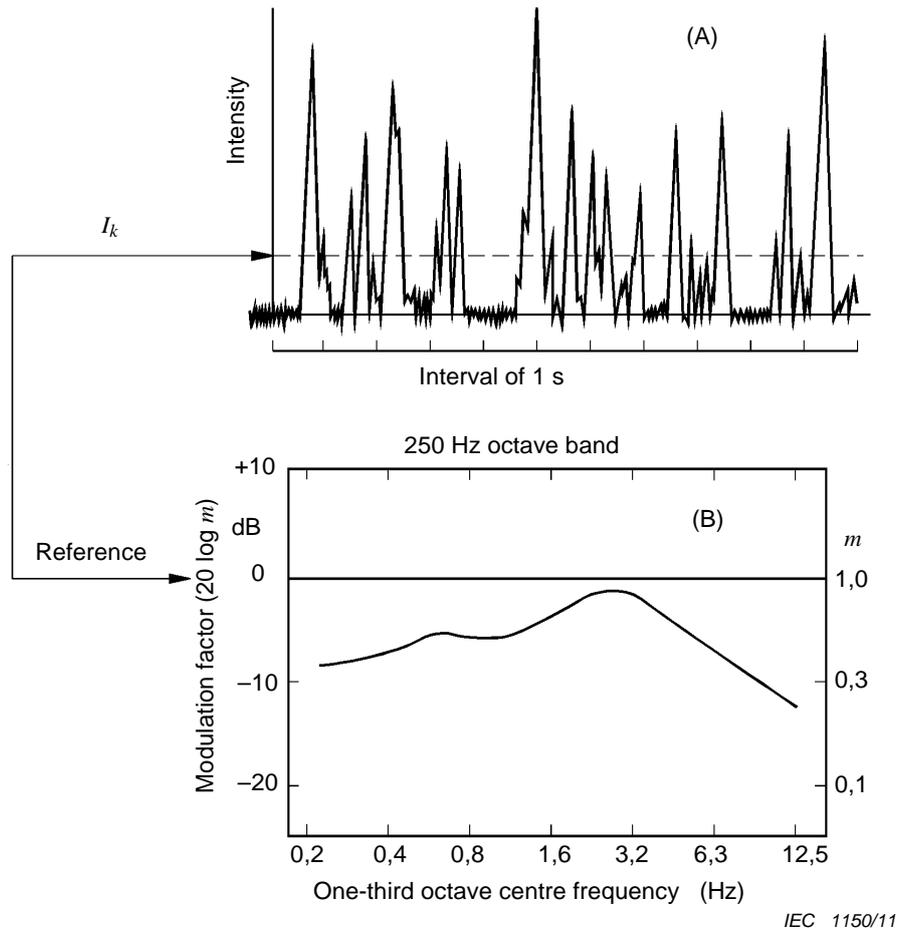


Figure A.1 – Envelope function (panel A) of a 10 s speech signal for the 250 Hz octave band and corresponding envelope spectrum (panel B)

A.1.2 Modulation transfer function (MTF)

The rationale underlying the application of the MTF concept to studies of room acoustics has been described elsewhere [5], [6], [19], [20]. The MTF quantifies the extent of the reductions in the modulations of the original material as a function of the modulation frequency. The modulations are defined by the intensity envelope of the signal, as it is in the intensity domain that interfering noise or reverberation will affect only the depth of modulation of a sinusoidal modulation without changing its shape. Figure A.2 illustrates this for the octave-band centred on 250 Hz for two simple transmission systems, one with exponential reverberation only (case A: $T = 2,5$ s) and the other with only interfering noise (case B; signal-to-noise ratio $SNR = 0$ dB).

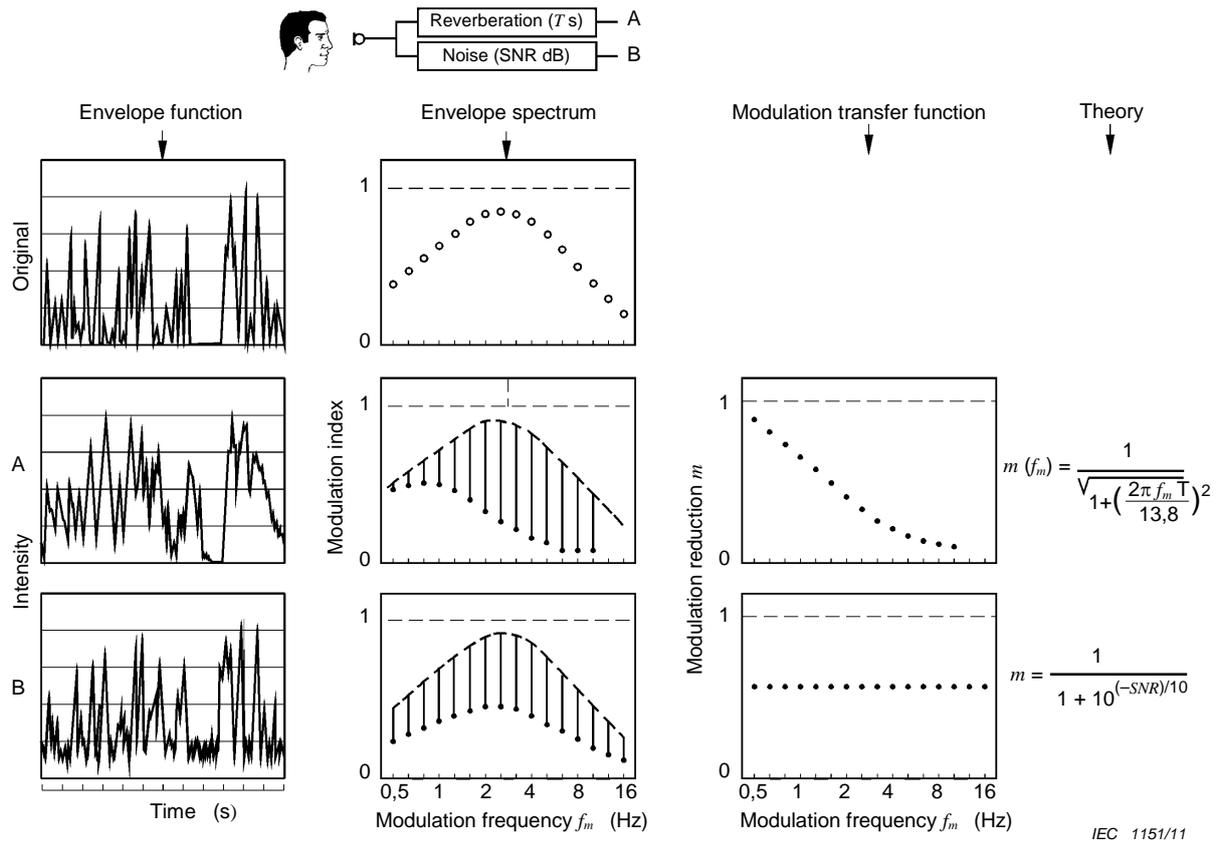


Figure A.2 – Theoretical expression of the MTF

With reverberation, the MTF has the shape of a low-pass filter: the faster fluctuations being relatively more affected than slower fluctuations. In the theoretical case of purely exponential reverberant decay, the MTF can be derived mathematically (see Figure A.2, case A) and the product of f_m and T determines the roll-off as given by:

$$m(f_m) = \frac{1}{\sqrt{1 + \left(\frac{2\pi f_m T}{13,8}\right)^2}}$$

where

f_m is the modulation frequency;

T is the reverberation time in seconds.

For noise, the MTF is defined by the signal-to-noise ratio and is independent of the modulation frequency (see Figure A.2, case B). The noise, by increasing the mean intensity, reduces the modulation depth for all modulation frequencies as given by:

$$m = \frac{1}{1 + 10^{(-SNR/10)}}$$

where

SNR is the signal to noise ratio in dB.

With strong echoes (pronounced reflections) the MTF shows the shape of a notch filter, rolling off first and rising again with higher modulation frequencies.

A.2 STI technique

A.2.1 General

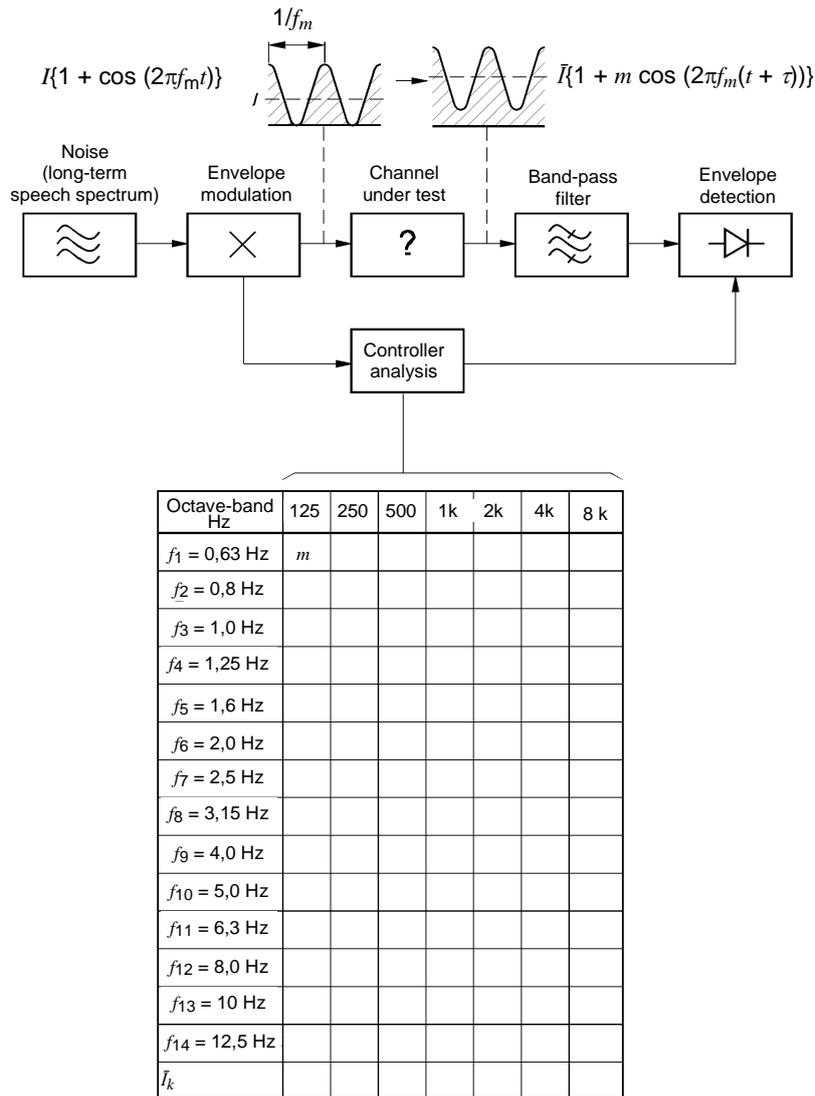
The speech transmission index (STI) is an objective measure, based on the weighted contribution of a number of frequency bands within the frequency range of speech signals, the contributions being set by the effective signal-to-noise ratio. Its description and the octave-band weighting factors and redundancy factors are given in [29]. By proper choice of the form of test signal, this effective signal-to-noise ratio can include and allow for distortions in the time domain and non-linearities as well as background noise, etc.

Distortions in the time domain (such as reverberation, echoes and automatic gain control) may degrade the fluctuating speech signal and reduce the intelligibility. This is modelled in the STI procedure by determining the modulation transfer function for the range of relevant frequencies present in the envelope of natural speech signals. The relevant range for these modulation frequencies extends from 0,63 Hz to 12,5 Hz in 14 one-third octave bands. Figure A.3 illustrates a measuring arrangement in which the modulation transfer function, $m(f_m)$, is determined separately for each modulation frequency in each octave band.

The most comprehensive measurement of STI is the FULL STI. The direct FULL STI method uses only one modulation frequency for one octave band per test signal with each test signal being approximately 10 s long. To obtain a single STI value, the FULL STI method uses 98 independent test signals (14×7).

The STI method was originally developed using direct FULL STI signals (one modulation frequency for one octave band) with random modulations for the octave bands that do not contain modulations. The random modulations were based on energy distributions as found in natural speech and had an instantaneous level which was approximately 3 dB higher than the overall speech level for the particular octave band under evaluation.

Since measuring the direct FULL STI is unpractical in many situations, faster methods (STIPA, STITEL) have been developed. The direct FULL STI method is generally now only used for background STI research.



IEC 1152/11

NOTE The value of the modulation transfer function (m) is determined for all cells of the matrix of seven octave bands and 14 modulation frequencies. Also, the octave intensity levels (\bar{I}_k , equal to the square of the sound pressure levels) are obtained for use in calculating auditory masking effects.

Figure A.3 – Measurement system and frequencies for the STI method

A.2.2 STI modulation frequencies

The 14 STI modulation frequencies are at nominal one-third octave intervals and consist of 0,63; 0,80; 1,00; 1,25; 1,60; 2,00; 2,50; 3,15; 4,00; 5,00; 6,3; 8,00; 10,0; 12,5 Hz.

NOTE For frequencies used by STIPA, see Table B.1.

A.3 Auditory effects on the STI

A.3.1 Overview

Hearing related aspects such as auditory masking (the reduction in aural sensitivity by a stronger, lower frequency sound) [30] as well as the absolute reception threshold are modelled in the STI calculation by applying appropriate noise terms. The auditory effects will reduce the effective signal-to-noise ratio in the various octave bands and can be expressed as a reduction of the modulation transfer function resulting in generally lower STI values.

STI evaluates the modulation loss and calculates the effective signal-to-noise-ratio in each octave band independently of the other octave bands. The only parameter that interlinks adjacent frequency bands with respect to the effective signal to noise ratio is the auditory masking function, in which aberrations in the amplitude frequency response are reflected when sound pressure levels are high and the slopes of the masking functions are low, see also 4.5.8.

NOTE Auditory effects can only be taken into account when test signals are obtained acoustically (in dB SPL). In general, this is often the case. However, if test signals are obtained electrically, without reference to a sound pressure level (for example when rating different type of microphones), this should be noted and auditory effects should be disabled in the measurement.

A.3.2 Level dependent auditory masking

Auditory masking is an inherent effect of the human hearing process. When a loud, low frequency sound is presented at the ear, it will always mask higher frequencies, possibly rendering them inaudible if the difference between their relative levels exceeds a given threshold. This phenomenon is referred to as upward spread of masking [30]. The auditory masking effect also depends on the absolute sound pressure level of both frequency components. A low frequency sound at low sound pressure level causes less masking of a high frequency sound than if it were at a higher sound pressure level, i.e. the masking slope at low sound pressure levels is steeper than at high sound pressure levels [31].

The main principle of the auditory masking as modelled in the STI is shown in Figure A.4. It shows that a lower octave band ($k - 1$) has a masking effect on the next (higher) octave band (k). The slope of the masking function in turn depends on the overall sound pressure level for octave band ($k - 1$).

The auditory masking as modelled in the STI has an effect on the 250 Hz, 500 Hz, 1 000 Hz, 2 000 Hz, 4 000 Hz and 8 000 Hz octave bands and extends one octave band upwards. Accordingly, the 125 Hz octave band masks the 250 Hz octave band, the 250 Hz octave band masks the 500 Hz octave band, etc. The 125 Hz octave band is not masked at all.

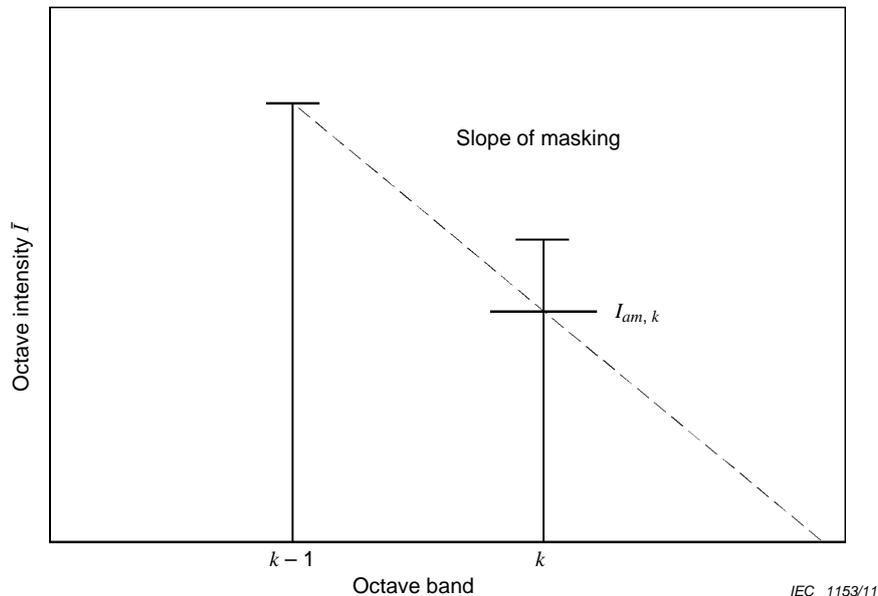


Figure A.4 – Auditory masking of octave band ($k - 1$) on octave band (k)

The masking intensity ($I_{am,k}$) for octave band k is given by:

$$I_{am,k} = I_{k-1} \times amf$$

where

I_{k-1} is the intensity of the adjacent lower octave band (octave band $k - 1$);

amf is the level dependent auditory masking factor which is a function that is dependent on the intensity (I_{k-1}) of the adjacent lower octave band.

The intensity (I_{k-1}) for an octave band $k-1$ is given by:

$$I_{k-1} = 10^{(L_{k-1}/10)}$$

where

L_{k-1} is the overall sound pressure level for octave band $k - 1$ in dB.

The auditory masking factor amf for octave band k therefore depends on the intensity of octave band $k - 1$.

In Table A.1, the level dependent auditory masking is given in dB ($amdB$) for octave band k as a function of the sound pressure level L (dB) in octave band $k - 1$. It is noted that the auditory masking level is a function of the total sound pressure level in that octave band consisting of both the test signal level and the ambient noise level.

Table A.1 – Auditory masking as a function of the octave band level

Sound pressure level L of octave band $k-1$ dB	<63	≥63 and <67	≥67 and <100	≥100
Auditory masking $amdB$ dB	$0,5 \times L - 65$	$1,8 \times L - 146,9$	$0,5 \times L - 59,8$	-10

The auditory masking factor amf for an octave band is given by:

$$amf = 10^{(amdB/10)}$$

where

$amdB$ is the octave band level dependent auditory masking value in dB as derived from Table A.1.

NOTE 1 In the case that a mathematical correction on the STI is made for a different ambient noise situation, the auditory masking factor depends on both the measured signal intensity and the added ambient noise intensity in a certain octave band. Both intensities are added to obtain the overall acoustic intensity for an octave band to be able to derive the appropriate auditory masking factor for that octave band.

NOTE 2 The previous masking scheme was discrete and resulted in non continuous STI results as a function of the overall sound pressure level. The auditory masking scheme as presented in Table A.1 yields continuous STI results as a function of the sound pressure level.

A.3.3 Absolute speech reception threshold

The absolute speech reception threshold is defined by the absolute threshold of hearing and the minimal required dynamic range for the correct recognition of speech. The absolute speech reception threshold intensity (I_{rt}) is modelled in the STI as an intrinsic noise floor which reduces the effective signal to noise ratio when speech levels are low. The absolute speech reception threshold is given in Table A.2 (in dB SPL) as a level in each octave band.

Table A.2 – Absolute speech reception threshold level in octave bands

Octave band Hz	125	250	500	1 000	2 000	4 000	8 000
Absolute speech reception threshold ART dB SPL	46	27	12	6,5	7,5	8	12

The reception threshold intensity ($I_{rt,k}$) for octave band k is given by:

$$I_{rt,k} = 10^{(ART_k / 10)}$$

where

ART_k is the absolute speech reception threshold for octave band k in dB (see Table A.2).

A.3.4 Gender-specific octave band weighting and redundancy factors

The STI method can discriminate between male and female speech signals. Gender related factors are expressed in different test signal spectra and different weighting factors. Since female speech is generally considered to be more intelligible than male speech, male speech is generally used to assess speech transmission channels.

The STI weighting factors (α) and redundancy factors (β) for male and female speech are shown in Table A.3 as a function of the octave bands.

Table A.3 – MTI octave band weighting factors

Octave band Hz		125	250	500	1 000	2 000	4 000	8 000
Males	α	0,085	0,127	0,230	0,233	0,309	0,224	0,173
	β	0,085	0,078	0,065	0,011	0,047	0,095	–
Females	α	–	0,117	0,223	0,216	0,328	0,250	0,194
	β	–	0,099	0,066	0,062	0,025	0,076	–

NOTE The weighting factors for male STI contain an artefact which can occasionally appear when performing predictive calculations. If the modulation transfer ratio of the 250 Hz octave band is equal to or less than 0,08 (equivalent to a transmission index TI of 0,15) with the other octave bands at maximum transmission index of (1,0), the STI value will be larger than 1,0. If the contribution of the 250 Hz band is completely removed, the STI will become 1,03. For this situation, it is recommended to truncate the STI at 1,0. In practical STI measurement situations, it is unlikely that this artefact will appear, as noise will prevent this from occurring.

A.3.5 Gender-specific spectra of STI test signals

The spectra of the STI test signals are specified by the octave band mean levels given in Table A.4. The octave band levels are normalized to an A-weighted level of 0 dB for easy scaling to an overall A-weighted sound pressure level. This spectrum may instantaneously be exceeded by 3 dB by a modulated test signal.

Table A.4 – Octave band levels (dB) relative to the A-weighted speech level

Octave band Hz	125	250	500	1 000	2 000	4 000	8 000	A-weighted
Males dB	2,9	2,9	–0,8	–6,8	–12,8	–18,8	–24,8	0,0
Females dB	–	5,3	–1,9	–9,1	–15,8	–16,7	–18,0	0,0

NOTE 1 The STIPA test signal is based on the male speech spectrum and only this signal should be generally employed.

NOTE 2 For guidance in determining the speech level, see Annex J.4.

A.4 STI test signal generation

A.4.1 Pink noise source signal

The direct STI method uses test signals that have similar spectral and temporal properties to those found in natural speech. Consequently, STI test signals consist of a number of frequency bands of noise whose intensity is sinusoidally modulated.

The STI test signal is initially generated from a pink noise source, which provides a flat frequency spectrum when measured with fractional octave-band filtering.

Pink noise can be produced from white noise using a low pass filter with a –3 dB per octave slope commencing at 63 Hz. The crest factor of the signal should typically lie between 12 dB and 14 dB, with the L_1 exceedance value lying between 8 dB and 11 dB.

A.4.2 Generating octave band carrier signals

To generate the seven STI noise carrier signals, a pink noise signal is fed into seven half-octave band wide filters at centre frequencies 125 Hz up to 8 000 Hz. Between each noise carrier signal there is a half-octave band wide gap which should not contain any significant signal. The half-octave band filters shall have a stop-band attenuation rate of at least 60 dB per octave to minimise the overlap between neighbouring carriers. The roll-off should be continuous and contain no ripple in the stop-band as exhibited e.g. by type II Chebyshev or elliptic filters. The ripple in the pass-band should not exceed 0,3 dB.

EXAMPLE 12th order 0,2 dB ripple type I Chebyshev filters can be used for generating the half-octave wide carrier signals. Other viable methods are frequency-domain related.

A.4.3 Intensity modulation of the carrier signals

Each carrier signal is sinusoidally intensity-modulated with one or more modulation frequencies, at the maximum possible modulation depth (i.e. $m = 1,0$). To obtain a sinusoidally intensity modulated carrier, the carrier signal is amplitude-modulated with the square root of a raised sinusoidal modulation as given by:

$$m_{f_m}(t) = \sqrt{0,5 \cdot (1 + m \cdot \cos(2\pi f_m t))}$$

where

f_m is the modulation frequency in Hz;

t is the time in seconds;

m is the depth of the modulation (maximally equal to 1).

If more than one modulation frequency per carrier band is impressed on a band at the same time, modulation factors shall be equal and the phase relationships and the modulation depth shall be selected to prevent over-modulation of the carrier.

A.4.4 Applying the speech spectrum to the STI signal

Each carrier signal is assigned a relative octave band level according to the speech spectrum as given in Table A.4. Since the carrier signals originate from a pink noise signal, the octave band levels are similar and only need to be derived according to Table A.4. All modulated carrier signals are summed to obtain the STI test signal.

A.5 Analysis of the STI test signal and calculation of the STI

A.5.1 Signal processing

A.5.1.1 Basic processing steps

The following subclauses set out the procedure for analysing the received signal and calculating the resultant STI. The procedure is broken down into the following basic steps:

- filter the input signal with the seven octave band filters;
- determine the intensities I_k in each octave band k ;
- determine the intensity modulation I at each modulation frequency f_m ;
- calculate the STI according to 4.3 and Clause A.1.

A.5.1.2 Filtering and determination of intensities

When filtering the received signal into the seven octave bands, the tolerance limits of the filters shall comply with IEC 61260, class 0 or class 1. The power of the input signal has to be split by the filter bank into output signals without power loss, so roll-offs of adjacent filters shall be complementary over frequency and intersect at –3 dB attenuation.

When applying the direct method, the received modulated noise test signal is filtered into octave band signals. The octave band filters are one octave band wide with centre frequencies ranging from 125 Hz up to 8 000 Hz.

The phase characteristics of the analysis filters should be as linear as possible to avoid distortion of the phase relationship of any of the amplitude modulations. During one measurement, all intensities should be calculated referring to the same time segment of the input signal; no implicit time weighting is allowed.

The phase characteristics of the analysis filters should not give rise to a systematic error higher than 0,01 STI in the end result for the range between 0,1 and 0,9 STI (between –12 dB and 12 dB *SNR*).

The intensities are calculated by squaring the outputs of the bandpass filters.

A.5.2 Derivation of the modulation transfer function (MTF) using the direct method

The calculation of the STI is based on the modulation transfer function. The MTF of a transmission path can be determined in various ways, the principal being the derivation of the modulation reduction factor from the comparison of the intensity modulations at the output and at the input to the path.

The filter outputs are transformed into intensity time signals by squaring the output signals of the filter. By applying a low pass filter at a cut-off frequency of approximately 100 Hz, the intensity envelope for each octave band is obtained.

Depending on the test signal, the MTF for each octave band is derived by correlating the intensity envelope with sine and cosine signals of a specific time length and specific modulation frequencies. The modulation transfer at each modulation frequency is calculated by first deriving the modulation depth of the received signal (mdr) for each octave band output (k).

$$mdr_{k, f_m} = 2 \times \frac{\sqrt{[\sum I_k(t) \cdot \sin(2\pi f_m t)]^2 + [\sum I_k(t) \cdot \cos(2\pi f_m t)]^2}}{\sum I_k(t)}$$

where

f_m is the modulation frequency in Hz;

t is the time in seconds;

$I_k(t)$ is the intensity envelope as a function of time for octave band k ;

NOTE 1 $I_k(t) \cdot \sin(2\pi f_m t)$ is the inner product of the intensity envelope for octave band k and a specific sinusoidal modulation f_m .

NOTE 2 The summation is made over the measurement duration with a whole number of periods for each specific modulation frequency.

Using the modulation indices of the received signal and the transmitted signal, the modulation transfer ratio can be calculated. This value is often referred to as the m -value. All derived modulation transfer ratio values $m(k, f_m)$ form the so called MTF matrix. The modulation transfer ratio is given by:

$$m_{k, f_m} = mdr_{k, f_m} / mdt_{k, f_m}$$

where

mdr_{k, f_m} is the modulation depth of the received test signal for octave band k and modulation frequency f_m ;

mdt_{k, f_m} is the modulation depth of the transmitted test signal for octave band k and modulation frequency f_m .

NOTE The derivation of the modulation transfer function (MTF) using the indirect method is described in Clause 6.

A.5.3 Correction of the MTF using auditory masking

The derived modulation transfer ratio values (m -values) are corrected using auditory masking effects by applying the following formula:

$$m'_{k,f_m} = m_{k,f_m} \times \frac{I_k}{I_k + I_{am,k} + I_{rt,k}}$$

where

m_{k,f_m} is the derived modulation transfer ratio value for octave band k and modulation frequency f_m ;

I_k is the acoustic intensity level for octave band k ;

$I_{am,k}$ is the acoustic intensity level for the level dependent auditory masking effect on octave band k ;

$I_{rt,k}$ is the acoustic intensity level of the reception threshold for octave band k .

NOTE 1 Modulation transfer ratio values higher than 1,0 should be truncated to 1,0. An m -value higher than 1,3 is very unlikely and may be a result of other than sinusoidal fluctuations or impulsive noises.

NOTE 2 When mathematically applying an additional ambient noise level, the term in the denominator should be extended with $I_{n,k}$ which represents the acoustic intensity level of the ambient noise for octave band k . Note that also the auditory masking intensity ($I_{am,k}$) is affected by mathematically applying additional ambient noise.

A.5.4 Calculation of the effective signal to noise ratio

The corrected modulation transfer ratio values are transformed into an effective signal to noise ratio SNR_{eff} (dB) as given by:

$$SNR_{\text{eff } k, f_m} = 10 \times \log \frac{m'_{k,f_m}}{1 - m'_{k,f_m}}$$

where

m'_{k,f_m} is the corrected modulation transfer ratio value for octave band k and modulation frequency f_m .

Since the outcome of the signal-to-noise ratio calculation may become infinite, values shall be limited to the range of -15 dB to +15 dB.

A.5.5 Calculation of the Transmission Index (TI)

The transmission index (TI) for each octave band and modulation frequency is calculated using:

$$TI_{k,f_m} = \frac{SNR_{\text{eff } k, f_m} + 15}{30}$$

where

$SNR_{\text{eff } k, f_m}$ is the effective signal to noise ratio for each octave band k and modulation frequency f_m expressed in dB.

A.5.6 Calculation of the STI

The derived transmission indices (TI) are averaged over modulation frequencies to obtain the modulation transfer index (MTI_k) per octave band k using:

$$MTI_k = \frac{1}{n} \sum_{m=1}^n TI_{k,f_m}$$

where

TI_{k,f_m} is the transmission index for each octave band k and modulation frequency f_m ;

m is the index of the modulation frequency.

n is the number of modulation frequencies per octave band.

With the modulation transfer indices (MTI_k) for each octave band k , the STI is calculated using:

$$STI = \sum_{k=1}^7 \alpha_k \times MTI_k - \sum_{k=1}^6 \beta_k \times \sqrt{MTI_k \times MTI_{k+1}}$$

where

MTI_k is the modulation transfer index for octave band k ;

α_k is the weight factor for octave band k ;

β_k is the redundancy factor between octave band k and octave band $k+1$.

NOTE In the event that STI values higher than 1,0 are obtained, they should be set at 1,0.

Annex B
(normative)

STIPA method

Instead of the 14 modulation frequencies being applied successively to all seven octave bands as per the procedure for the FULL STI, the STIPA method applies two unique modulation frequencies simultaneously to each of the seven frequency bands as shown in Table B.1. A total of $2 \times 7 = 14$ modulation frequencies are therefore used.

As each octave band is modulated by two modulation frequencies simultaneously at a frequency ratio of 5, the modulation index (depth) for each modulation frequency shall be 0,55 for a sinusoidal addition of the two components with a phase difference of 180° between the components.

The STIPA method is only validated for the male speech spectrum and its measurement time is approximately 15 s to 20 s.

Table B.1 – Modulation frequencies for the STIPA method

Octave band Hz	125	250	500	1 000	2 000	4 000	8 000
First modulation frequency Hz	1,60	1,00	0,63	2,00	1,25	0,80	2,50
Second modulation frequency Hz	8,00	5,00	3,15	10,0	6,25	4,00	12,5

Annex C (normative)

STITEL method

C.1 General

A simplification can be applied to the test signal if the uncorrelated (speech-like) modulations required for the correct interpretation of non-linear distortions, are omitted. This allows modulation and processing of all seven frequency bands simultaneously, thus reducing measuring time. The STITEL method employs this simplification and requires between 10 s and 15 s for one measurement.

STITEL uses one modulation frequency per octave band carrier which allows 100 % modulation of the test signal to be employed and thus increases the SNR by 3 dB.

In place of the 14 modulation frequencies that are applied successively to all seven octave bands for the FULL STI, the STITEL method simultaneously applies a unique modulation frequency to each of the seven octave bands as shown in Table C.1. The test signal includes all seven modulated octave bands, and these are all analysed simultaneously.

Table C.1 – Modulation frequencies for the STITEL method

Octave band Hz	125	250	500	1 000	2 000	4 000	8 000
Modulation frequency Hz	1,12	11,33	0,71	2,83	6,97	1,78	4,53

Other than the above modulation frequencies, the normal STI calculation scheme is applied for STITEL.

C.2 Limitations

The STITEL method should not be used for transmission channels:

- a) which introduce frequency shifts or frequency multiplication;
- b) which include vocoders (i.e. LPC, CELP, RELP, etc.);
- c) which introduce strong non-linear distortion components;
- d) for which reverberation time is strongly frequency-dependent;
- e) having echoes stronger than –10 dB referred to the primary signal;
- f) if the background noise has audible tones and/or marked peaks or troughs in the octave-band spectrum;
- g) if the background noise is impulsive and/or the space is not substantially free of discrete echoes, particularly flutter echoes whose repetition frequency is an integral multiple of one or more of the modulation frequencies.

If c), d), or e) or all three apply, or possibly apply, the STI method should be used instead, or used to verify the results obtained by the STITEL method.

Annex D
(informative)

RASTI method (obsolete)

Although now obsolete, details of the RASTI method are shown for the sake of completeness. The technique was created to provide a faster method of obtaining an STI measurement, there being an order of magnitude reduction in the data and corresponding computational effort. The reliance of the technique operating on just two octave carriers was found to be a serious limitation of the method when testing electroacoustic systems.

As shown in Table D.1, a total number of nine modulation frequencies applied to two octave bands is used.

Table D.1 – Modulation frequencies for the RASTI method

Modulation frequency Hz	f_1	f_2	f_3	f_4	f_5
500 Hz octave band	1,0	2,0	4,0	8,0	–
2 000 Hz octave band	0,7	1,4	2,8	5,6	11,2

The frequency weighting is applied indirectly by using four contributions for the 500 Hz octave band and five for the 2 kHz octave band, i.e. the weightings are 4/9 (0,45) and 5/9 (0,55), respectively.

Figure D.1 illustrates a practical RASTI test signal.

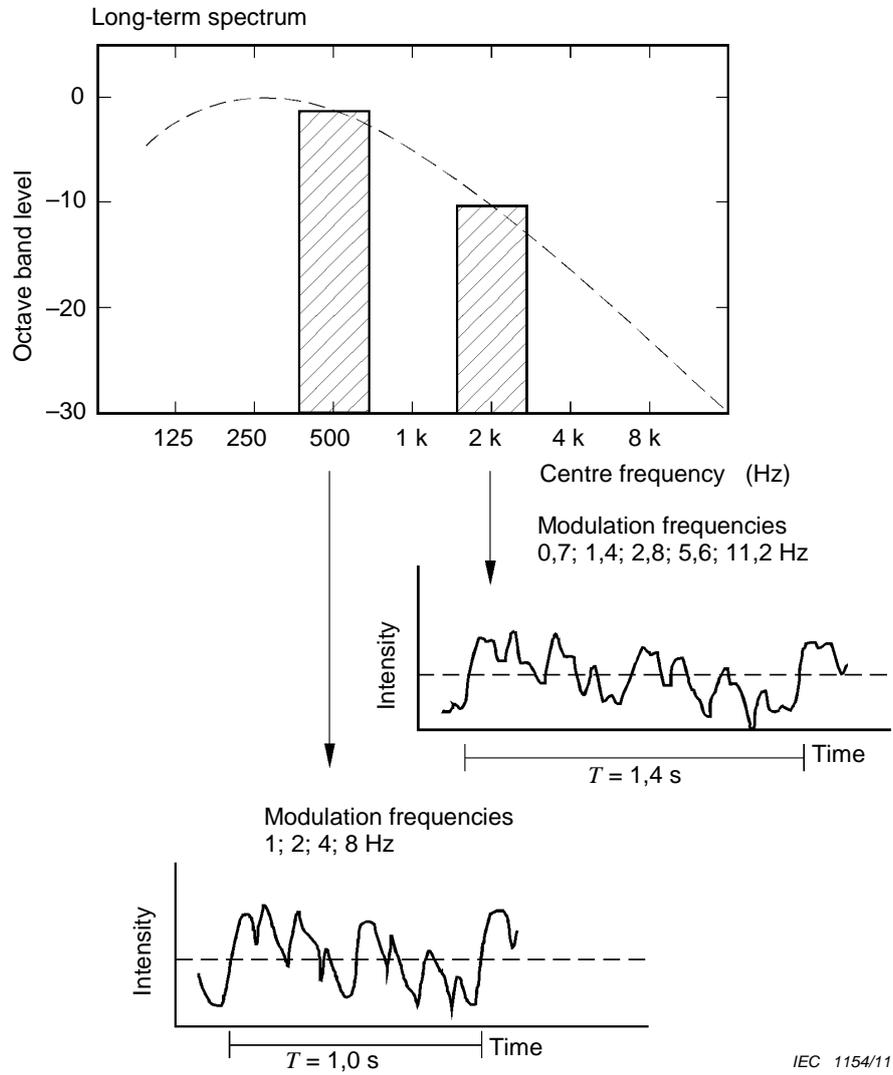


Figure D.1 – Illustration of a practical RASTI test signal

Annex E (informative)

Qualification of the STI and relationships with other speech intelligibility measures

E.1 Relationship between STI and word/sentence scores

The relationships between the STI and various speech intelligibility measures are given in Figure E.1. The nonsense word score for equally balanced CVC words is obtained from [32]. The relation with PB words of the so-called “Harvard list” is according to [33]. The relation with sentence intelligibility is based on SRT (Speech Reception Threshold) results.

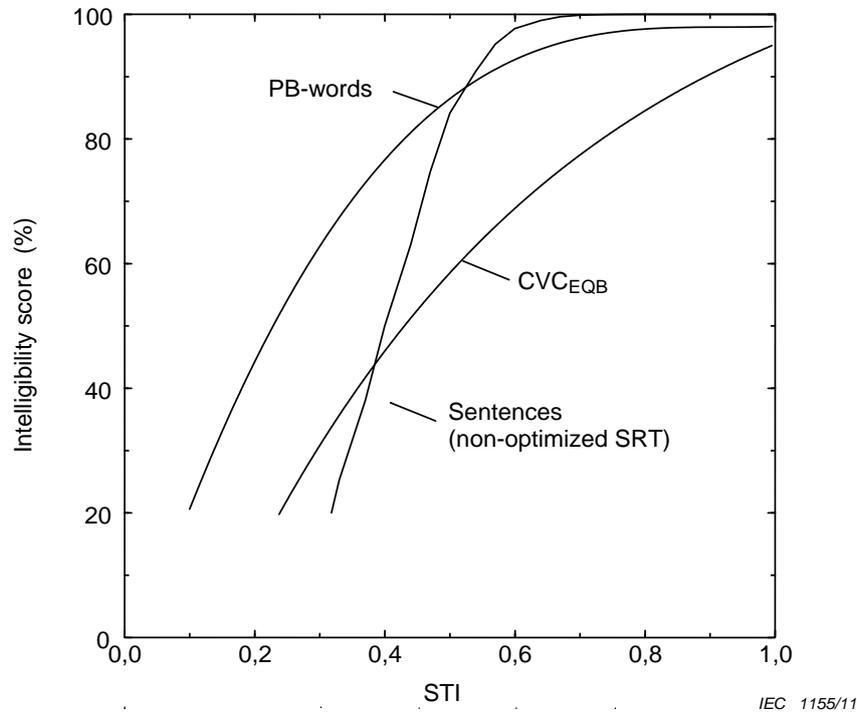


Figure E.1 – Relationships between some
speech intelligibility measures

E.2 Relationship between STI and listening difficulty

In some circumstances, listening difficulty may be a more suitable alternative for the evaluation of speech intelligibility performance than word or sentence scores, as the listening difficulty metric is more sensitive to transmission channels that exhibit a high performance level of speech transmission.

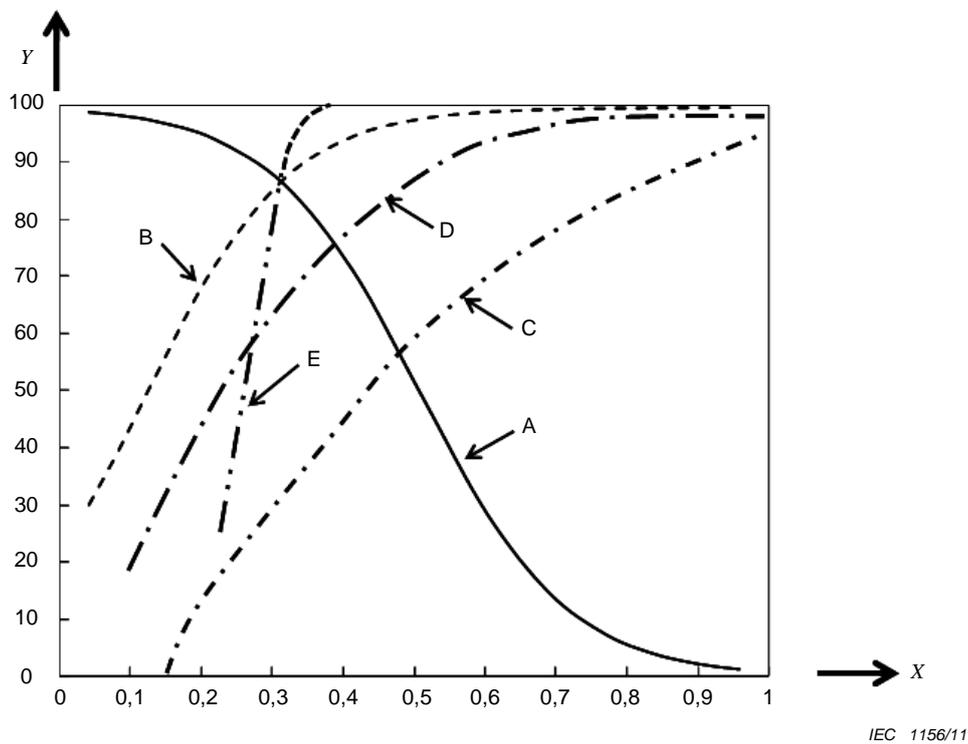
The listening difficulty rating is the percentage of responses indicating a certain degree of difficulty when listening to speech through a transmission channel. Note that listening difficulty ratings decrease for conditions with improved speech transmission, contrary to conventional intelligibility scores [31].

Table E.1 lists the categories which are typically used to describe the listening difficulty.

Table E.1 – Categories for listening difficulty

Category	Description
0	Not difficult: no effort is required, equivalent to a relaxed listening condition
1	Slightly difficult: slight attention is required
2	Moderately difficult: moderate attention is required
3	Very difficult: considerable attention is required

Figure E.2 shows the relationship between listening difficulty ratings, intelligibility scores in Figure E.1 and the STI.

**Key**

X axis is STI

Y axis is percentage of correct responses and listening difficulty

Curve A is listening difficulty

Curve B is word recognition

Curve C is CVC_{EQB}

Curve D is PB-Words

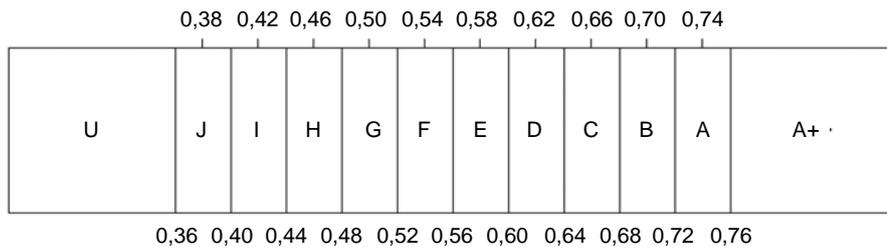
Curve E is Sentences (SRT)

Figure E.2 – Relationship between STI, speech intelligibility scores and listening difficulty ratings [34], [35]

Annex F (informative)

Nominal qualification bands for STI

In order to provide both flexibility for different applications and an inbuilt measurement and/or prediction tolerance, the qualification scale is divided into a number of bands. The STI value required for a given application or situation may then be obtained from an associated performance table (for an example, see Annex G). Figure F.1 shows the qualification bands:



IEC 1157/11

Key

Upper row of numbers: STI values at the centre of the bands;

Row of letters: band designations;

Lower row of numbers: STI values at the edges of the bands.

Figure F.1 – STI qualification bands

The spacing of the intervals in Figure F.1 is based on the typical uncertainty of direct STI measurements.

When STI measurements are made over a specific area, the spread of results should be analysed by examining the statistical distribution.

Annex G (informative)

Examples of STI qualification bands and typical applications

The information in the Table G.1 is presented as an example of usage.

Table G.1 – Examples between STI qualification bands and typical applications

Category	Nominal STI value	Type of message information	Examples of typical uses (for natural or reproduced voice)	Comment
A+	>0,76		Recording studios	Excellent intelligibility but rarely achievable in most environments
A	0,74	Complex messages, unfamiliar words	Theatres, speech auditoria, parliaments, courts, Assistive Hearing Systems (AHS)	High speech intelligibility
B	0,7	Complex messages, unfamiliar words		
C	0,66	Complex messages, unfamiliar words	Theatres, speech auditoria, teleconferencing, parliaments, courts	High speech intelligibility
D	0,62	Complex messages, familiar words	Lecture theatres, classrooms, concert halls	Good speech intelligibility
E	0,58	Complex messages, familiar context	Concert halls, modern churches	High quality PA systems
F	0,54	Complex messages, familiar context	PA systems in shopping malls, public buildings' offices, VA systems, cathedrals	Good quality PA systems
G	0,5	Complex messages, familiar context	Shopping malls, public buildings' offices, VA systems	Target value for VA systems
H	0,46	Simple messages, familiar words	VA and PA systems in difficult acoustic environments	Normal lower limit for VA systems
I	0,42	Simple messages, familiar context	VA and PA systems in very difficult spaces	
J	0,38		Not suitable for PA systems	
U	<0,36		Not suitable for PA systems	
NOTE 1 These values should be regarded as minimum target values.				
NOTE 2 Perceived intelligibility relating to each category will also depend on the frequency response at each listening position.				
NOTE 3 The STI values refer to measured values in sample listening positions or as required by specific application standards.				

Annex H (informative)

Non-native listeners

Generally, compared to the intelligibility obtained with native listeners, non-native listeners require a 4 dB to 5 dB improvement of the signal-to-noise ratio for similar intelligibility (see ISO 9921:2003 [49]).

Adjusted intelligibility qualification tables for three groups of non-natives can be defined, based on experimental data [36]. For this purpose, the non-native proficiency of the listener should be classified, based on language experience, age of learning and frequency of use of the second language.

Table H.1 is indicative only. For details refer to ergonomics standards (ISO 9921 [49]). For low-proficiency non-native listeners, good or excellent intelligibility cannot be achieved.

NOTE As a guide, an increase in STI of 0,1 corresponds to 3 dB change in effective signal-to-noise ratio.

Table H.1 – Adjusted intelligibility qualification tables for non-native listeners

STI label range	Standard STI	Non-native category I experienced, daily second language use	Non-native category II intermediate experience and level of second language use	Non-native category III new learner, infrequent second language use
bad – poor	0,30	0,33	0,38	0,44
poor – fair	0,45	0,50	0,60	0,74
fair – good	0,60	0,68	0,86	impossible
good – excellent	0,75	0,86	impossible	impossible

NOTE 1 For details on STI label categories, refer to ISO 9921 [49].

EXAMPLE For a non-native listener of category II and to achieve an intelligibility equivalent to an STI of 0,45, the transmission system needs to achieve a performance of 0,60.

NOTE 2 For intermediate values between the stated standard STI, interpolation should be used to estimate the adjusted STI.

Annex I (informative)

Effect of age-related hearing loss and hearing impairment on speech intelligibility

For hearing-impaired persons, deriving adjusted intelligibility qualification tables is more complex than for non-native listeners and requires defining the type of hearing impairment.

The STI method cannot give reliable results for all types of hearing impairment and, in general, it is recommended that subject-based listening tests or other prediction methods such as the speech intelligibility index (SII) are used.

For listeners beyond 50 years old, hearing levels and the spread between individuals increases considerably [37]. Nevertheless, age-related hearing impairment shows good correlation between intelligibility and hearing loss.

As a rule of thumb, to reach intelligibility at the critical point of 50 % sentence intelligibility (where redundant sentences typically have to be repeated twice), hearing impaired listeners require 4,5 dB higher *SNR* for 20 dBHL [38]. Here, dBHL refers to the hearing loss (HL) in dB, defined as the pure-tone average hearing level (PTA) of 0,5 kHz, 2,0 kHz and 4,0 kHz, relative to 18-year old normal listeners.

Table I.1 provides an indication of the adjusted intelligibility qualifications. It should be noted that the maximum intelligibility that can be reached depends on the degree of hearing loss [15], [38]. For more details, refer to other standards, such as [39].

**Table I.1 – Adjusted intelligibility qualification tables for normal listeners
and people over 60 years old with hearing loss**

STI label category	Normal listeners (Standard STI)	Older listeners PTA = 15 dB	Older listeners PTA = 20 dB	Older listeners PTA = 30 dB
bad – poor	0,30	0,42	0,47	0,51
poor – fair	0,45	0,57	0,62	0,66
fair – good	0,60	0,72	cannot be achieved	cannot be achieved
good – excellent	0,75	cannot be achieved	cannot be achieved	cannot be achieved

NOTE 1 For details on categories, refer to ISO 9921 [49].

NOTE 2 Typical normal listeners have a PTA of between 0 and 5 dB.

EXAMPLE For an older listener with PTA = 15 dB and to achieve an intelligibility equivalent to an STI of 0,45, the transmission system needs to achieve an STI of 0,57.

Annex J (normative)

Calibration of STI test signal level

J.1 Overview

For performing reproducible STI measurements, speech levels and noise levels should be carefully calibrated. However, speech signals and noise signals may have different temporal and spectral characteristics, which make it difficult to easily and accurately compare them. For determining the speech-to-noise ratio, a level measurement method should be equally suitable for various types of speech (male versus female, connected discourse versus isolated words), measurement conditions (background noise, bandwidth) and should also be applicable to noise-like signals.

The basic methods of measuring the real speech level are based on removing the silent parts of the speech signal, e.g. the gaps between words. An accurate method (the speech level meter procedure) is described in Clause J.2. Other methods may be employed as well, if less accuracy is required. A simple method, available also with standardized sound pressure level meters is described in Clause J.3 and Clause J.4.

J.2 The concept of 'speech level' and the method of measurement

The speech level measurement method, was developed by Houtgast and Steeneken (1978, 1986). In order to overcome spectral differences between signals and to have the signal levels closely match a perceived level, the measurement method is based on the A-weighted filtering of the signal. In general, speech signals are not continuous and contain numerous pauses. When specifying the signal-to-noise ratio of speech and noise signals based on the measurement of $L_{A,eq}$ values, it is evident that the number and the duration of the silences between utterances will affect the result. For example, speech at exactly the same nominal level but with a different pattern of pauses will lead to a different $L_{A,eq}$ value being measured and therefore to differences in the resultant signal-to-noise ratio.

The speech level measurement method deals with this phenomenon by removing all silences before calculating the level and in a manner such that only the parts of the speech signal which contain relevant signal information are taken into account. The A-weighted signal is therefore divided into frames of 10 ms to 20 ms in length and the energy per frame is calculated. Next, all the energy per frame values are accumulated in a level distribution histogram. Using the level distribution histogram, the RMS level of the speech is iteratively determined by cumulating all histogram data starting at a threshold that lays 14 dB below the calculated RMS level of the speech signal. Following this procedure, the relatively silent parts of the signal are left out and as a result, the signal level will become independent of the temporal distribution of the signal.

Extensive research [40] carried out on different speech level measures shows that the difference in RMS level of connected discourse and CVC words embedded in carrier phrases is minimal (<0,5 dB) when applying the speech level measurement method.

J.3 Test speech level

For measurements on a PA system, the test speech level is the level actually obtained from the system when working correctly at a specified reference position. If the signal input is from a talker or other acoustic source, the paragraph below applies when setting the input level to the system.

For measurements with a talker or other acoustic source, in the absence of a PA system, the test speech level shall be set to 60 dB A measured at 1 m distance, on the axis of main radiation of the artificial mouth or talker. If it is required to simulate a condition with a raised vocal effort (Lombard effect), the test speech level shall be set to 70 dB A.

J.4 Corrected speech level

An approximate corrected speech level measurement procedure based on a simple measurement of an A-weighted equivalent sound pressure level ($L_{A,eq}$) is as follows. The measured level is corrected by an empirically derived factor in order to obtain an estimate of the real speech level as determined with the method described in Clause J.2. In order to obtain the approximate corrected speech level, proceed as follows.

- Determine the $L_{A,eq}$ of a voice signal, with a length of at least 40 s unless the signal is a recorded announcement of shorter duration.
- Add 3 dB to the result.

NOTE The 3 dB correction factor may vary according to the speech rate and characteristics of a specific talker.

Annex K
(informative)

Example test report sheet for STI measurements

This is a general set of guidelines and prompts to assist the measurement of STI over a range of applications. Not all categories are applicable in every case.

GENERAL INFORMATION	
Measurement method	
Project / location	
Occupancy / configuration	
Date of measurement	
Method: Indirect (IR) or Direct (STIPA)	
Source	
Signal type: MLS; swept sine; STIPA; other	
Source type: Signal generator, CD player; .wav or .mp3 player	
Test signal spectrum: Adjusted to the standardized speech spectrum?	
Method of signal insertion: Electronic input or broadcast from an acoustic source?	
Details of test loudspeaker / artificial mouth / type	
Distance of source to system microphone (m)	
Directional pattern of sound system microphone	
Distance of microphone to nearest reflecting surface (m)	
System signal processing	
Status of system signal processing, e.g. compression, limiters, equalisation	
Was any part of the signal chain clipping during the measurement?	
Measurement hardware	
Brand/Type – serial no / version	
STI or STIPA loop back-to-back test result	
Measurement microphone	
Brand, model and type (free field, random incidence)	
Monaural or binaural?	
Height above floor/ground (m)	
Aiming point of microphone	
Subjective impression with speech transmission	
Tonal characteristics: natural, muffled, boomy, resonant, harsh?	
Are there audible distortions or echoes?	
For acoustical signal insertion: Is there audible ringing or regeneration (feedback)?	

MEASUREMENT DATA PER POSITION (use a different sheet for each position)								
Position / location reference								
Time of measurement								
Ambient noise levels (without test signal)								
	$L_{A,eq}$ (dB)	Octave band (Hz)						
		125	250	500	1 000	2 000	4 000	8 000
L_{eq} measured over 15 s (dB)								
Description of the ambient noise: e.g. steady, transient, impulsive								
Climatic conditions: wind, rain, temperature, humidity								
STIPA method								
STIPA signal level	$L_{A,eq}$ (dB)	Octave band (Hz)						
		125	250	500	1 000	2 000	4 000	8 000
L_{eq} measured over 15 s (dB)								
STIPA measurement number	Average	1	2	3	4	5	6	
STIPA results for each measurement: take 3 measurements; if variation >0,03, take 3 more								
Maximum variation of results								
Impulse response (IR) method								
Length of acquired impulse response Minimum for STI calculation: 1,6 s								
Test signal level	$L_{A,eq}$ (dB)	Octave band (Hz)						
		125	250	500	1 000	2 000	4 000	8 000
L_{eq} with test signal (dB)								
Signal to noise ratio SNR (dB)								
STI results	STI	Transmission index MTI						
Transmission indices and STI								
Deviation of frequency response from 250 Hz to 12 kHz relative to 1 kHz.								
Frequency response measurement: Time window applied, smoothing.								
Does IR show arrivals likely to cause audible echoes?								
Results of processing measured data for different signal and noise levels								
Signal levels	$L_{A,eq}$ (dB)	Octave band (Hz)						
		125	250	500	1 000	2 000	4 000	8 000
Proposed speech level (dB)								
Proposed ambient noise level (dB)								
STI results	STI	Transmission index MTI						
Calculated STI and MTIs								

Annex L (normative)

Prediction of STI using statistical methods

According to [28], the overall modulation transfer function including temporal distortion and noise can be written as:

$$m(f_m) = \frac{\left| \int_0^{\infty} h(t)^2 e^{-j2\pi f_m t} dt \right|}{\int_0^{\infty} h(t)^2 dt} \cdot \left[1 + 10^{-SNR/10} \right]^{-1}$$

where

$m(f_m)$ is the modulation transfer function of the transmission channel;

f_m is the modulation frequency;

$h(t)$ is the impulse response of the transmission channel;

t is the integration variable for time;

SNR is the signal to noise ratio in dB.

Assuming a diffuse reverberant field, the impulse response for both the direct and reverberant field components for a single source can be written as:

$$h(t) = \frac{Q}{r^2} \cdot \delta(t) + \frac{13,8 Q}{r_c^2 T} e^{-\frac{13,8t}{T}}$$

where

Q is the directivity factor for the sound source (loudspeaker or talker);

$\delta(t)$ is the Dirac (or delta) function;

r is the talker to listener distance;

r_c is the critical distance in the room or space (also known as critical radius);

T is the reverberation time of the room or space.

The modulation transfer function including temporal distortion and noise can then be written as:

$$m(f_m) = \frac{\sqrt{A^2 + B^2}}{C}$$

with

$$A = \frac{Q}{r^2} + \frac{1}{r_c^2} \left[1 + \left(\frac{2\pi f_m T}{13,8} \right)^2 \right]^{-1};$$

$$B = \frac{2\pi f_m T}{13,8 r_c^2} \left[1 + \left(\frac{2\pi f_m T}{13,8} \right)^2 \right]^{-1};$$

$$C = \frac{Q}{r^2} + \frac{1}{r_c^2} + Q \cdot 10^{-SNR/10};$$

NOTE The prediction described above is only accurate for receiver locations within the main radiation direction of the talker or loudspeaker. Thus, the direct field component of the above equations should be adjusted to account for any off-axis loss of the loudspeakers due to directivity and the number of loudspeakers contributing to the direct field.

It is critical that the operational speech level be used for prediction of the STI as this affects both the effective *SNR* and masking effects. A broadband speech signal shall be used for this prediction and shall ensure that the transmission channel is capable of producing the operational sound pressure level.

The above method does not account for the arrival-time difference of multiple direct-field signals, nor can it account for echoes.

If the space exists, the measured reverberation times should be used in preference to the predicted reverberation times.

Annex M (informative)

Adjustments to measured STI and STIPA results for simulation of occupancy noise and different speech levels

An example of a suitable method to adjust the measured STI and STIPA values to simulate occupancy noise and different speech levels is given in Table M.1.

Table M.1 – Example calculation

1 Acquire measurement data with signal and noise levels present during measurement							
	125 Hz	250 Hz	500 Hz	1 000 Hz	2 000 Hz	4 000 Hz	8 000 Hz
Signal level L_{eq}	77,9	77,9	74,2	68,2	62,2	56,2	50,2
Background noise levels L_{eq}	48,0	40,0	34,0	30,0	27,0	25,0	23,0
MTF matrix with noise, temporal effects, masking and threshold factors							
0,63 Hz	0,982	0,952	0,960	0,969	0,979	0,983	0,994
0,80 Hz	0,966	0,928	0,941	0,954	0,969	0,976	0,992
1,00 Hz	0,945	0,897	0,914	0,933	0,955	0,965	0,989
1,25 Hz	0,919	0,862	0,881	0,908	0,939	0,952	0,984
1,60 Hz	0,884	0,819	0,836	0,873	0,915	0,932	0,978
2,00 Hz	0,850	0,784	0,793	0,838	0,890	0,911	0,971
2,50 Hz	0,815	0,750	0,749	0,799	0,862	0,888	0,961
3,15 Hz	0,772	0,715	0,716	0,760	0,832	0,863	0,950
4,00 Hz	0,740	0,678	0,691	0,730	0,800	0,836	0,938
5,00 Hz	0,724	0,623	0,665	0,721	0,772	0,811	0,926
6,30 Hz	0,713	0,553	0,643	0,708	0,745	0,785	0,913
8,00 Hz	0,669	0,515	0,611	0,664	0,720	0,764	0,901
10,00 Hz	0,590	0,479	0,545	0,603	0,693	0,748	0,890
12,50 Hz	0,553	0,442	0,513	0,602	0,678	0,736	0,881
2 Remove background noise, masking and threshold factors							
	125 Hz	250 Hz	500 Hz	1 000 Hz	2 000 Hz	4 000 Hz	8 000 Hz
Signal to noise ratio during measurement dB	29,90	37,90	40,20	38,20	35,20	31,20	27,20
$mk(f)$ for noise only	0,999	1,000	1,000	1,000	1,000	0,999	0,998
Adjustment to remove background noise	1,001	1,000	1,000	1,000	1,000	1,001	1,002
Combined speech and noise level dB	77,90	77,90	74,20	68,20	62,20	56,20	50,21
Auditory masking factor amf dB	not applicable	-20,8	-20,8	-22,7	-25,7	-33,9	-36,9
Combined squared sound pressure I_k , MPa ²	61,7	61,7	26,3	6,61	1,66	0,417	0,105
Auditory masking factor $amf \times 1\ 000$	not applicable	8,22	8,22	5,37	2,69	0,407	0,204
$I_{am,k}$	0	508 000	507 000	141 000	17 800	676	85,2
Absolute reception threshold ART dB	46	27	12	6,5	7,5	8	12
Intensity - absolute reception threshold $I_{rt,k}$	40 000	501	15,8	4,5	5,6	6,3	15,8

Adjustment to remove masking and threshold	1,001	1,008	1,019	1,021	1,011	1,002	1,001
Combined adjustments	1,002	1,008	1,019	1,022	1,011	1,002	1,003
Adjusted MTF matrix without noise, masking and threshold							
0,63 Hz	0,983	0,960	0,978	0,990	0,990	0,986	0,997
0,80 Hz	0,968	0,936	0,959	0,974	0,980	0,979	0,995
1,00 Hz	0,947	0,904	0,931	0,953	0,966	0,968	0,992
1,25 Hz	0,920	0,869	0,898	0,927	0,949	0,955	0,987
1,60 Hz	0,886	0,826	0,852	0,892	0,925	0,935	0,981
2,00 Hz	0,851	0,791	0,808	0,856	0,900	0,914	0,974
2,50 Hz	0,816	0,756	0,764	0,816	0,871	0,891	0,964
3,15 Hz	0,773	0,721	0,730	0,776	0,841	0,866	0,953
4,00 Hz	0,741	0,684	0,705	0,745	0,809	0,838	0,941
5,00 Hz	0,726	0,628	0,678	0,736	0,780	0,812	0,929
6,30 Hz	0,714	0,557	0,656	0,723	0,753	0,786	0,916
8,00 Hz	0,670	0,520	0,623	0,678	0,728	0,765	0,904
10,00 Hz	0,591	0,483	0,556	0,615	0,701	0,749	0,893
12,50 Hz	0,554	0,446	0,523	0,614	0,685	0,737	0,884
3 Adjust MTF matrix for operational levels and masking and threshold effects							
Operational speech L_{eq}	82,9	82,9	79,2	73,2	67,2	61,2	55,2
Operational background noise L_{eq}	55,5	47,5	41,5	37,5	34,5	32,5	30,5
Signal to Noise ratio	27,40	35,40	37,70	35,70	32,70	28,70	24,70
$mk(f)$ for noise only	0,998	1,000	1,000	1,000	0,999	0,999	0,997
Combined speech and noise level dB	82,9	82,9	79,2	73,2	67,2	61,2	55,2
Auditory masking factor amf dB	not applicable	-18,3	-18,3	-20,2	-23,2	-26,2	-34,4
Combined squared sound pressure I_k , MPa ²	195	195	83,2	20,9	5,25	1,32	0,332
Auditory masking factor $amf \times 1\ 000$	not applicable	14,6	14,6	9,55	4,79	2,40	0,363
$I_{am,k}$	0	2 850 000	2 850 000	795 000	100 000	12 600	480
Absolute reception threshold ART, dB	46	27	12	6,5	7,5	8	12
Intensity - absolute reception threshold $I_{rt,k}$	40 000	500	15,8	4,5	5,6	6,3	15,8
Correction for masking and threshold	1,000	0,986	0,967	0,963	0,981	0,991	0,999
Combined adjustments	0,998	0,985	0,967	0,963	0,981	0,989	0,995
Adjusted MTF matrix for operational levels, masking and threshold							
0,63 Hz	0,981	0,946	0,946	0,953	0,971	0,975	0,992
0,80 Hz	0,966	0,922	0,927	0,938	0,961	0,968	0,990
1,00 Hz	0,945	0,891	0,900	0,918	0,947	0,957	0,987
1,25 Hz	0,919	0,856	0,868	0,893	0,931	0,944	0,982
1,60 Hz	0,884	0,814	0,823	0,859	0,907	0,925	0,976
2,00 Hz	0,850	0,779	0,781	0,824	0,882	0,904	0,969
2,50 Hz	0,814	0,745	0,738	0,786	0,855	0,881	0,959
3,15 Hz	0,772	0,710	0,706	0,747	0,825	0,856	0,948
4,00 Hz	0,739	0,674	0,681	0,718	0,793	0,829	0,936
5,00 Hz	0,724	0,619	0,656	0,709	0,765	0,804	0,924

6,30 Hz	0,713	0,549	0,634	0,696	0,739	0,778	0,911
8,00 Hz	0,668	0,512	0,602	0,653	0,714	0,757	0,900
10,00 Hz	0,589	0,476	0,537	0,593	0,687	0,741	0,889
12,50 Hz	0,553	0,439	0,505	0,592	0,672	0,729	0,880
4 Process MTF matrix to yield STI							
	125 Hz	250 Hz	500 Hz	1 000 Hz	2 000 Hz	4 000 Hz	8 000 Hz
4a Convert into effective SNRs							
0,63 Hz	17,21	12,44	12,42	13,09	15,21	15,93	21,01
0,80 Hz	14,55	10,73	11,04	11,83	13,90	14,83	20,02
1,00 Hz	12,34	9,13	9,56	10,47	12,52	13,50	18,86
1,25 Hz	10,52	7,74	8,17	9,22	11,31	12,30	17,41
1,60 Hz	8,82	6,41	6,69	7,84	9,91	10,88	16,13
2,00 Hz	7,52	5,47	5,52	6,71	8,76	9,73	14,98
2,50 Hz	6,42	4,66	4,51	5,64	7,70	8,69	13,72
3,15 Hz	5,29	3,89	3,80	4,71	6,73	7,75	12,64
4,00 Hz	4,53	3,16	3,30	4,06	5,84	6,87	11,68
5,00 Hz	4,19	2,11	2,79	3,87	5,14	6,12	10,87
6,30 Hz	3,95	0,85	2,38	3,60	4,51	5,44	10,13
8,00 Hz	3,04	0,21	1,80	2,74	3,97	4,94	9,52
10,00 Hz	1,57	-0,42	0,65	1,63	3,42	4,57	9,02
12,50 Hz	0,92	-1,06	0,10	1,61	3,12	4,31	8,64
4b Truncate $SNR_{k,f}$							
0,63 Hz	15,0	12,4	12,4	13,1	15,0	15,0	15,0
0,80 Hz	14,6	10,7	11,0	11,8	13,9	14,8	15,0
1,00 Hz	12,3	9,1	9,6	10,5	12,5	13,5	15,0
1,25 Hz	10,5	7,7	8,2	9,2	11,3	12,3	15,0
1,60 Hz	8,8	6,4	6,7	7,8	9,9	10,9	15,0
2,00 Hz	7,5	5,5	5,5	6,7	8,8	9,7	15,0
2,50 Hz	6,4	4,7	4,5	5,6	7,7	8,7	13,7
3,15 Hz	5,3	3,9	3,8	4,7	6,7	7,7	12,6
4,00 Hz	4,5	3,2	3,3	4,1	5,8	6,9	11,7
5,00 Hz	4,2	2,1	2,8	3,9	5,1	6,1	10,9
6,30 Hz	4,0	0,9	2,4	3,6	4,5	5,4	10,1
8,00 Hz	3,0	0,2	1,8	2,7	4,0	4,9	9,5
10,00 Hz	1,6	-0,4	0,6	1,6	3,4	4,6	9,0
12,50 Hz	0,9	-1,1	0,1	1,6	3,1	4,3	8,6
4c Convert to Transmission Indices $MTI_{k,f}$							
0,63 Hz	1,00	0,91	0,91	0,94	1,00	1,00	1,00
0,80 Hz	0,99	0,86	0,87	0,89	0,96	0,99	1,00
1,00 Hz	0,91	0,80	0,82	0,85	0,92	0,95	1,00
1,25 Hz	0,85	0,76	0,77	0,81	0,88	0,91	1,00
1,60 Hz	0,79	0,71	0,72	0,76	0,83	0,86	1,00
2,00 Hz	0,75	0,68	0,68	0,72	0,79	0,82	1,00
2,50 Hz	0,71	0,66	0,65	0,69	0,76	0,79	0,96
3,15 Hz	0,68	0,63	0,63	0,66	0,72	0,76	0,92

4,00 Hz	0,65	0,61	0,61	0,64	0,69	0,73	0,89
5,00 Hz	0,64	0,57	0,59	0,63	0,67	0,70	0,86
6,30 Hz	0,63	0,53	0,58	0,62	0,65	0,68	0,84
8,00 Hz	0,60	0,51	0,56	0,59	0,63	0,66	0,82
10,00 Hz	0,55	0,49	0,52	0,55	0,61	0,65	0,80
12,50 Hz	0,53	0,46	0,50	0,55	0,60	0,64	0,79
MTI_k	0,73	0,66	0,67	0,71	0,77	0,80	0,92
Octave band weighting alpha, males	0,085	0,127	0,230	0,233	0,309	0,224	0,173
Sum of alpha* $MTI = MTI_k \times$ alpha weighting	0,062	0,083	0,155	0,165	0,237	0,179	0,159
Octave band weighting beta, males	0,085	0,078	0,065	0,011	0,047	0,095	none
Sum of beta* $MTI = MTI_k \times$ beta weighting	0,059	0,052	0,045	0,008	0,037	0,081	0,000
Σ alpha* MTI	1,040						
Σ beta* MTI	0,282						
STI	0,76						

Annex N (informative)

Other methods of determining speech intelligibility

N.1 Overview

Intelligibility prediction metrics can be broadly divided into two categories: relatively complex predictors including explicit and sophisticated perceptual and cognitive modelling, and simpler metrics that are easier to measure and understand and are therefore accessible to greater populations of acousticians. The STI and SII both fall into the first category, although the STI leans towards the ease-of-use which is the benefit of the second category, while the SII more dominantly possesses the flexibility and scientific rigor that is the benefit of the first category.

Another example of the first category (complex perceptual models) is the Speech Recognition Sensitivity model [41], which elegantly works around shortcomings of other models, but has not seen much “field experience” or independent evaluation. Complex models have also been developed to predict speech quality and intelligibility specifically for telecommunication channels (for example, the PESQ model [42], [43]). The added value of the STI, in relation to these models, is the wider applicability (room acoustics and telecommunications), combined with its widespread use and third-party evaluations. The fact that various vendors have implemented the STI method in their measuring devices helps in this respect.

The category of simpler metrics includes the Speech Interference Level (SIL) as described in ISO 9921 [49], a metric that predicts intelligibility of speech in noise by averaging the speech-to-noise ratio in three octave bands. This second category also includes various measures based on early-to-late energy ratios derived from impulse responses, such as clarity and definition [44]. These are specifically of interest when investigating reverberant environments. Under the conditions and for the type of applications that these measures are intended for, their level of accuracy may approach that of the STI. In more complex situations, the accuracy of the STI outperforms all simpler metrics.

N.2 Word tests

The limitations of word tests are given in ISO/TR 4870 [45]. It should be noted that, because the method is based on the perception of words by listeners, there are no limitations with respect to the characteristics of the sound system or those of the environment. It is essential that the words are embedded in a carrier phrase in case of use in combination with temporal distortions (reverberation, echoes, automatic gain control).

N.3 Modified rhyme tests

The limitations are similar to those given in ISO/TR 4870 [45]. It should be noted that, because the method is based on the perception of words by listeners, there are no limitations in respect of the characteristics of the sound system or those of the environment.

N.4 Speech Intelligibility Index (SII)

The SII is also often preferred by those who are interested in comparing effects of different speech materials rather than different channels. However, in contrast to STI, SII cannot be measured directly, but shall be calculated. It is commonly used by experimental audiologists, because of its higher frequency resolution and its sensitivity to the intelligibility decrease at high vocal efforts.

NOTE SII appears not to be significantly more sensitive than STI to the effects of frequency response [12].

N.5 Articulation loss of consonants (%ALcons)

The limitations are similar to those given in ISO/TR 4870 [45]. It should be noted that the measurement procedure does not include vowels. This may cause a systematic error with respect to word tests [46]. As the test is based on the reception of words by listeners, there are no limitations in respect to the characteristics of the sound system or those of the environment. It should be noted that %ALcons cannot normally be accurately measured acoustically. It should be noted that there is no accurate way of electroacoustically measuring %ALcons.

N.6 PESQ

The Perceptual Evaluation of Speech Quality (PESQ) is especially useful for situations of high intelligibility and is based on mean opinion scores and especially suitable for measuring the high quality transmission where speech intelligibility is less of an issue [42], [43].

Bibliography

- [1] Mapp, P, *New Techniques for Measuring Speech Privacy & Efficiency of Sound Masking Systems. AES 122nd Convention, Vienna, May 2007*
- [2] IEC 60318-1:2009, *Electroacoustics – Simulators of human head and ear – Part 1: Ear simulator for the measurement of supra-aural and circumaural earphones*
- [3] IEC 61672 (all parts), *Electroacoustics – Sound level meters*
- [4] ANSI (1969). ANSI S3.5-1969, *American National Standard, "Methods for Calculation of the Articulation Index" (American National Standards Institute New York.)*
- [5] Steeneken, H.J.M. and Houtgast, T., "Some applications of the Speech Transmission Index (STI) in auditoria", *Acustica 51, 1982, p.229-234*
- [6] Steeneken, H.J.M. and Houtgast, T., "A physical method for measuring speech transmission quality", *J. Acoust. Soc. Amer. 67, 1980, 31, p.318-326*
- [7] Houtgast, T. and Steeneken, H.J.M., "The modulation transfer function in room acoustics as a predictor of speech intelligibility", *Acustica 28, 1973, p.66-73*
- [8] Steinbrecher, T, "Speech Transmission Index: Too weak in time and frequency?", *Proc IOA Vol 30 Part 6 2008*
- [9] Mapp, P, "Is STIPA a robust measure of speech intelligibility performance," *AES 118th Convention, Barcelona, 2005.*
- [10] Mapp, P, *Some Effects of Equalisation on Sound System Intelligibility and Measurement AES 115th Convention 2003 Preprint*
- [11] Leembruggen, G and A Stacey, "Should the Matrix be reloaded Proc IOA Vol 28 Part 6 2006
- [12] Leembruggen, G., "Is SII better than STI at recognising the effects of poor tonal balance on intelligibility?" *Proc IOA Vol 28 Part 6 2006*
- [13] Mapp P, *Practical application of STI to assessing Public Address and Emergency Sound Systems. Past, present and future of the Speech Transmission Index, TNO 2002*
- [14] Festen, JM & Plomp, R., *Effects of fluctuating noise and interfering speech on the speech reception threshold for impaired and normal hearing, J. Acoust. Soc. Amer., 88(4) 1990*
- [15] Duquesnoy, A.J.H/M and Plomp, R., (1980) "Effect of reverberation and noise on the intelligibility of sentences in case of presbycusis". *J. Acoust. Soc. Am. 68, p.537-544*
- [16] Mapp, P, "Assessing the Potential Intelligibility of Assistive Audio Systems for the Hard of Hearing and Other Users", *AES 124th convention, Amsterdam, 2008*
- [17] Schroeder, M, *Modulation Transfer Functions: Definition and Measurement, Acustica, 49, 1981*
- [18] Houtgast, T. and Steeneken, H.J.M., "The modulation transfer function in room acoustics as a predictor of speech intelligibility", *Acustica 28, 1973, p.66-73*
- [19] Houtgast, T. and Steeneken, H.J.M., "A Multi-lingual evaluation of the Rasti-method for estimating speech intelligibility in auditoria", *Acustica 54, 1984, p.185-199*

- [20] Steeneken, H.J.M. and Houtgast, T., *“The temporal envelope spectrum of speech and its significance in room acoustics”*, Proc. 11th International Congress on Acoustics, Paris, 1983, Vol. 7, p.85-88
- [21] Steeneken, H.J.M., Verhave, J.A., McManus, S., and Jacob, K.D., (2001) *“Development of an Accurate, Handheld, Simple-to-use Meter for the Prediction of Speech Intelligibility”*, Proceedings IoA 2001, Reproduced sound (17). Stratford-upon-Avon, UK
- [22] Mapp, P. *Further thoughts on Speech Transmission Index (STI)*. IOA Reproduced Sound 18.Proc. IOA Vol 24 Pt 8. 2002
- [23] Bozzoli, F, Farina, A, *“Influence of artificial mouth’s directivity in determining Speech Transmission Index”*., AES 119th Convention, New York, 2005
- [24] IEC 60268-3:2000, *Sound system equipment – Part 3: Amplifiers*
- [25] Rife, D., *“Modulation Transfer Function Measurements with Maximum-Length Sequences”*, J. Audio Eng. Soc., Oct. 1992, Vol. 40, No. 10
- [26] ANSI standard S12.2 1995, *Criteria for evaluation of room noise*
- [27] Wijngaarden, S, Drullman, R, *“Binaural intelligibility prediction based on the speech transmission index”* J. Acoust. Soc. Amer. 123, 2008, p.4514-4523
- [28] Houtgast, T, Steeneken, H and Plomp, R., *“Predicting Speech Intelligibility in Rooms from the Modulation Transfer Function. i. General Room Acoustics”* Acustica Vol 46 1980
- [29] Steeneken, H.J.M. and Houtgast, T., *“Mutual dependency of the octave-band weights in predicting speech intelligibility,”* Speech Communication 28, 1999, p.109-123
- [30] Zwicker, E. and Feltkeller, *“The ear as an information receiver”*, (ASA)
- [31] Wijngaarden, S.J. van and Steeneken, H.J.M. (1999). *“Objective prediction of speech intelligibility at high ambient noise levels using the speech transmission index.”* In Eurospeech99 – Proceedings of the 6th European Conference on Speech Communication and Technology, Budapest, Vol 6, p.2639-2642
- [32] Steeneken, H.J.M. and Houtgast, T., *“Validation of the STIr method with the revised model,”* Speech Communication 38, 2002, p.413-425
- [33] Anderson, B.W., and Kalb, J.T. 1987. *“English verification of the STI method for estimating speech intelligibility of a communications channel,”* J. Acoust. Soc. Am. 81, 1982-1985
- [34] Hiroshi Sato, John S. Bradley, and Masayuki Morimoto, *Using listening difficulty ratings of conditions for speech communication in rooms* , J. Acoust. Soc. Am. 117, 1157 (2005)
- [35] Hayato Sato, Masayuki Morimoto, Hiroshi Sato, and Megumi Wada, *Relationship between listening difficulty and acoustical objective measures in reverberant sound fields,* J. Acoust. Soc. Am. 123, 2087-2093 (2008)
- [36] Wijngaarden, S.J., Steeneken, H.J.M. and Houtgast, T. (2002) *“Quantifying the intelligibility of speech in noise for non-native listeners,”* J. Acoust. Soc Am. 112, p.3004-3013

- [37] ISO 7029:2000, *Acoustics – Statistical distribution of hearing thresholds as a function of age*
- [38] Sato,H, Kurakata,K, Mizunami,T, “Accessible speech messages for the elderly in rooms” *9th Western Pacific Acoustics Conference Seoul, Korea, 2006*
- [39] ISO/TR 22411:2008, *Ergonomics data and guidelines for the application of ISO/IEC Guide 71 to products and services to address the needs of older persons and persons with disabilities*
- [40] Steeneken, H.J.M. and Houtgast, T (1986), *Comparison of some methods for measuring speech levels Report IZF1986-20 TNO Institute for Perception Soesterberg Netherlands*
- [41] H. Müsch and S. Buus. ‘Using statistical decision theory to predict speech intelligibility. I. Model structure’. *J.Acoust.Soc.Am.* 109, 2896–2909. (2001)
- [42] ITU-T Rec P.862, *Perceptual evaluation of speech quality (PESQ)” An objective method of end-to-end speech quality assessment of narrow band telephone networks and speech codecs. (International Telecommunication Union, Geneva Switzerland 2001 Feb)*
- [43] J. G. Beerends, A. P. Hekstra, A. W. Rix and M. P. Hollier. ‘PESQ, the new ITU standard for objective measurement of perceived speech quality, Part II – Perceptual model’. *J. Audio Eng. Soc.*, vol. 50, pp. 765-778. (2002 Oct.)
- [44] ISO 3382-1:2009, *Acoustics – Measurement of room acoustic parameters – Part 1: Performance spaces*
- [45] ISO/TR 4870:1991, *Acoustics – The construction and calibration of speech intelligibility tests (Withdrawn)*
- [46] Steeneken, H.J.M. and Houtgast, T., “Phoneme-group specific octave-band weights in predicting speech intelligibility,” *Speech Communication* 38, 2002, p.399-411
- [47] ITU-T P.50, *Telephone transmission quality, telephone installations, local line networks – Objective measuring apparatus – Artificial voices*
- [48] ITU-T P.51, *Telephone transmission quality – Objective measuring apparatus – Artificial mouth*
- [49] ISO 9921:2003, *Ergonomics – Assessment of speech communication*
-

SOMMAIRE

AVANT-PROPOS.....	77
INTRODUCTION.....	79
1 Domaine d'application	81
2 Références normatives.....	81
3 Termes et définitions	81
4 Description de la méthode STI.....	87
4.1 Généralités.....	87
4.1.1 Justification de la méthode STI.....	87
4.1.2 Applicabilité de la méthode STI	87
4.2 Notions de base pour la méthode STI.....	88
4.2.1 Généralités.....	88
4.2.2 Vue d'ensemble de la théorie.....	90
4.2.3 Mesure du STI.....	91
4.3 Applicabilité des méthodes d'essai STI.....	92
4.4 Utilisation des méthodes directes et indirectes	94
4.5 Limitation de la méthode STI.....	94
4.5.1 Généralités.....	94
4.5.2 Glissements de fréquence	94
4.5.3 Mise à zéro des petites valeurs (" <i>centre clipping</i> ").....	95
4.5.4 Pertes de niveau (" <i>drop outs</i> ").....	95
4.5.5 Gigue	95
4.5.6 Vocodeurs	95
4.5.7 Surestimation de l'indice STI dans des conditions de faible bruit de fond.....	95
4.5.8 Réponse en fréquence.....	96
4.5.9 Echos	96
4.5.10 Compression et expansion rapides de l'amplitude.....	97
4.5.11 Distorsion non linéaire	97
4.5.12 Bruit impulsif et fluctuant	97
4.5.13 Auditeurs malentendants	98
4.6 Conclusion	98
5 Méthode directe de mesure du STI	98
5.1 Vue d'ensemble.....	98
5.2 STIPA	99
5.3 Application	99
5.4 Limitations.....	99
6 Méthode indirecte de mesure du STI utilisant la réponse impulsionnelle	100
6.1 Vue d'ensemble.....	100
6.2 Application	101
6.3 Limitations (distorsion non linéaire)	101
7 Modes opératoires de mesure, post-traitement des données et applications	102
7.1 Généralités.....	102
7.2 Entrée acoustique	102
7.3 Sortie acoustique	104
7.4 Entrée électrique.....	104
7.5 Sortie électrique	104

7.6	Exemples de combinaisons entrée/sortie.....	104
7.6.1	Entrée acoustique – Sortie acoustique.....	104
7.6.2	Entrée électrique – Sortie électrique (par exemple, évaluation de systèmes de communication filaire et sans fil)	104
7.6.3	Entrée acoustique – Sortie électrique (par exemple, évaluation des microphones).....	105
7.6.4	Entrée électrique – Sortie acoustique (par exemple, évaluation des systèmes PA)	105
7.7	Post-traitement des données de MTF (fonction de transfert de modulation) mesurées	105
7.8	Questions concernant le bruit.....	106
7.8.1	Généralités.....	106
7.8.2	Mesure du bruit de fond.....	106
7.8.3	Bruit fluctuant.....	106
7.9	Analyse et interprétation des résultats.....	107
7.10	Mesures de STI binaural	107
8	Utilisation du STI comme moyen de prédiction en conception de systèmes acoustiques	107
8.1	Vue d'ensemble.....	107
8.2	Prédictions statistiques	108
8.3	Prédiction à partir d'une réponse impulsionnelle simulée.....	108
	Annexe A (normative) Méthode de l'indice de transmission de la parole (STI) et méthode STI révisée	110
	Annexe B (normative) Méthode STIPA	123
	Annexe C (normative) Méthode STITEL.....	124
	Annexe D (informative) Méthode RASTI (désuète).....	125
	Annexe E (informative) Qualification du STI et relations avec d'autres mesures de l'intelligibilité de la parole.....	127
	Annexe F (informative) Bandes nominales pour qualifier un indice STI	129
	Annexe G (informative) Exemples d'intervalles ou "bandes" pour qualifier un indice STI et applications typiques	130
	Annexe H (informative) Auditeurs de langue étrangère	131
	Annexe I (informative) Effet sur l'intelligibilité de la parole de la perte d'audition et de troubles de l'audition liés à l'âge	132
	Annexe J (normative) Etalonnage du niveau du signal d'essai STI	133
	Annexe K (informative) Exemple de fiche de rapport d'essai pour les mesures de STI	135
	Annexe L (normative) Prédiction du STI à l'aide de méthodes statistiques	138
	Annexe M (informative) Ajustements des résultats de STI et de STIPA mesurés pour la simulation du bruit d'occupation et des différents niveaux de parole.....	140
	Annexe N (informative) Autres méthodes de détermination de l'intelligibilité de la parole	144
	Bibliographie.....	146
	Figure 1 – Concept de réduction de la modulation due au canal de transmission	89
	Figure 2 – Fonction de transfert de modulation – comparaison entrée/sortie	90
	Figure A.1 – Fonction enveloppe (schéma A) d'un signal vocal de 10 s pour une bande d'une octave centrée sur 250 Hz et spectre de l'enveloppe correspondant (schéma B)	111
	Figure A.2 – Expressions théoriques de la MTF	112
	Figure A.3 – Système de mesure et fréquences pour la méthode STI	114

Figure A.4 – Masquage auditif de la bande d'une octave ($k - 1$) sur la bande d'une octave (k).....	115
Figure D.1 – Illustration d'un signal d'essai RASTI en pratique	126
Figure E.1 – Relations entre quelques mesures de l'intelligibilité de la parole	127
Figure E.2 – Relation entre STI, notes (<i>scores</i>) d'intelligibilité de la parole et évaluations de la difficulté d'écoute [34], [35].....	128
Figure F.1 – "Bandes" ou intervalles servant à qualifier un STI	129
Tableau 1 – Comparaison des méthodes d'essai STI pour différents types de distorsion	92
Tableau 2 – Applicabilité de l'essai	93
Tableau 3 – Choix de la méthode.....	94
Tableau A.1 – Masquage auditif en fonction du niveau de pression sonore de la bande d'une octave	116
Tableau A.2 – Niveau de seuil absolu de réception de la parole dans les bandes d'une octave.....	117
Tableau A.3 – Facteurs de pondération de bandes d'une octave STI	117
Tableau A.4 – Niveaux de bande d'une octave (dB) par rapport au niveau de parole pondéré A.....	118
Tableau B.1 – Fréquences de modulation pour la méthode STIPA	123
Tableau C.1 – Fréquences de modulation pour la méthode STITEL	124
Tableau D.1 – Fréquences de modulation pour la méthode RASTI.....	125
Tableau E.1 – Catégories pour apprécier la difficulté d'écoute	128
Tableau G.1 – Exemples de correspondances entre les "bandes" qualifiant un indice STI et des applications typiques	130
Tableau H.1 – Indice d'intelligibilité ajusté pour des auditeurs de langue maternelle étrangère selon leurs compétences linguistiques	131
Tableau I.1 – Tableaux de qualification de l'intelligibilité ajustée pour les auditeurs normaux et âgés de plus de 60 ans présentant une perte d'audition <i>PTA = pure tone average hearing level</i>	132
Tableau M.1 – Exemple de calcul	140

COMMISSION ÉLECTROTECHNIQUE INTERNATIONALE

ÉQUIPEMENTS POUR SYSTÈMES ÉLECTROACOUSTIQUES –

**Partie 16: Evaluation objective de l'intelligibilité de la parole
au moyen de l'indice de transmission de la parole**

AVANT-PROPOS

- 1) La Commission Electrotechnique 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. A 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 elle-même ne fournit aucune attestation de conformité. Des organismes de certification indépendants fournissent des services d'évaluation de conformité et, dans certains secteurs, accèdent aux marques de conformité de la CEI. La CEI n'est responsable d'aucun des services effectués par les organismes de certification indépendants.
- 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 60268-16 a été établie par le comité d'études 100 de la CEI: Systèmes et appareils audio, vidéo et multimédia.

Cette quatrième édition annule et remplace la troisième édition parue en 2003, et constitue une révision technique.

Cette édition inclut les modifications techniques importantes suivantes par rapport à l'édition précédente:

- Le développement d'une normalisation plus aboutie, plus complète et moins ambiguë de la méthodologie STI.
- Le terme STI_r n'est plus utilisé. Une nouvelle fonction pour la prédiction des effets du masquage auditif est introduite.

- Le concept de "niveau de parole" et le réglage du niveau du signal d'essai ont été introduits.
- Des informations supplémentaires relatives aux modes opératoires de prédiction et de mesures ont été incluses.

NOTE Voir l'Introduction donnant un résumé de l'historique des différentes modifications effectuées de la première à la quatrième édition (la présente édition).

Le texte anglais de cette norme est issu des documents 100/1812/FDIS et 100/1849/RVD. Le rapport de vote 100/1849/RVD 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.

Une liste de toutes les parties de la série de normes CEI 60268, publiées sous le titre général *Equipements pour systèmes électroacoustiques*, peut être consultée sur le site web de la CEI.

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

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

INTRODUCTION

L'indice de transmission de la parole (STI) est une mesure objective pour prédire l'intelligibilité de la parole transmise de locuteur à auditeur par un canal de transmission. La méthode STI a fait l'objet de développements et de perfectionnements permanents depuis son introduction dans les années 1970. Les améliorations majeures de l'indice STI ont été consolidées en les incorporant dans les révisions successives de la CEI 60268-16.

L'historique des révisions est la suivante:

- Révision 1: 1988. Dans la première version de la norme STI, il avait été utilisé un spectre de signal d'essai indépendant du sexe du locuteur.
- Révision 2: 1998. Des signaux d'essai spécifiques au sexe ont été introduits, pour les locuteurs masculins et féminins, chaque genre se rapportant à un jeu spécifique de facteurs de pondération. En outre, des pondérations ont été introduites pour les facteurs de redondance. Le terme STI_r a été introduit pour signifier l'utilisation de ces facteurs de redondance.
- Révision 3: 2003. Des différences importantes entre la Révision 2 et la Révision 3 consistent en l'introduction de
 - fonctions de masquage dépendant du niveau,
 - la méthode STIPA dérivée de la méthode STI.

La méthode STIPA avait été spécialement développée comme une méthode de mesure rapide qui pouvait traiter des effets électroacoustiques et acoustiques lors de la détermination de la qualité de transmission de la parole des systèmes de sonorisation PA (Public Address system).

- Révision 4: 2010. Le but de la Révision 4 (la présente révision) est d'assurer une normalisation plus complète et aboutie et moins ambiguë de la méthodologie STI. Le terme STI_r n'est désormais plus utilisé. Une nouvelle fonction est introduite pour la prédiction des effets du masquage auditif.

La parole est considérée comme la méthode principale de communication entre les êtres humains. Dans un grand nombre de situations, le signal vocal est dégradé par le trajet du signal ou par le canal de transmission entre locuteur et auditeur, ce qui réduit l'intelligibilité de la parole à l'emplacement où se trouve l'auditeur.

Pour quantifier la détérioration de l'intelligibilité de la parole apportée par le canal de transmission, une méthode de mesure rapide et objective a été développée: l'indice de transmission de la parole (STI, Speech Transmission Index). La méthode STI applique un signal d'essai spécifique au canal de transmission et analyse le signal d'essai reçu; la qualité de transmission de la parole par le canal s'en déduit et est exprimée par une valeur comprise entre 0 et 1, appelée indice de transmission de la parole (STI). Avec la valeur STI ainsi obtenue on peut déterminer l'intelligibilité potentielle de la parole.

Bien que la méthode STI ait ses limites, elle s'est avérée utile dans un grand nombre de situations et est devenue une technique internationalement reconnue et pratiquée.

Éléments qui ont changé dans la présente révision

Les modifications spécifiques qui ont été incorporées dans la présente révision sont:

- perfectionnement du modèle STI quant à la fonction de masquage dépendant du niveau;
- l'indice de transmission de la parole relatif à l'acoustique d'une salle (Room Acoustic Speech Transmission Index, RASTI) est devenu caduc et il convient donc de ne pas l'utiliser;
- calculs pour ajouter ou éliminer les effets d'un bruit de fond et pour modifier le niveau de la parole et un exemple d'application;

- notes relatives aux limitations de la méthode STI;
- méthodes de prévision de la performance en indice STI de canaux de transmission, employant la performance prédite (par opposition à mesurée) de tout ou partie du canal de transmission;
- introduction des corrections STI pour les auditeurs de langue maternelle étrangère;
- introduction des corrections STI pour les auditeurs présentant certaines formes spécifiques de perte d'audition;
- relations entre STI et échelle de "difficulté d'écoute".

Applications potentiel du STI

L'indice STI peut être utilisé pour mesurer l'intelligibilité potentielle d'une grande gamme de systèmes électroniques et d'environnements acoustiques. Des applications types sont notamment:

- mesure des systèmes de sonorisation et de renforcement du son;
- mesure et certification des systèmes électroacoustiques d'alarme vocale et d'urgence;
- mesure des canaux / systèmes de communication, tels que des interphones ou une communication sans fil;
- mesure de l'intelligibilité potentielle de la parole et de la communication dans des salles et des auditoriums;
- évaluation de la communication parlée directe (situations sans amplification électronique) dans des salles ou espaces acoustiques, dont les véhicules;
- évaluation de l'intelligibilité potentielle de systèmes d'aide à l'audition.

NOTE La méthode STI n'a pas été validée pour la mesure et l'évaluation de la confidentialité des conversations ou des systèmes de masquage de la parole.

Utilisateurs potentiels du STI

L'étendue des utilisateurs de mesures STI est diverse. Les utilisateurs pouvant appliquer cette méthode sont, entre autres:

- certificateurs des systèmes d'alarme vocale et autres types de systèmes d'urgence;
- certificateurs de systèmes de renforcement du son et de systèmes audio;
- fabricants de matériel audio et de télécommunication;
- techniciens de l'audio et de la communication;
- techniciens acousticiens et électroacousticiens;
- installateurs de systèmes électroacoustiques;
- chercheurs en matière de méthodes STI et développeurs d'instruments pour mesurer le STI.

Afin d'éviter toute erreur d'interprétation des résultats de la méthode STI, il est important que tous les utilisateurs en comprennent les principes fondamentaux, le domaine d'application et les limitations.

ÉQUIPEMENTS POUR SYSTÈMES ÉLECTROACOUSTIQUES –

Partie 16: Evaluation objective de l'intelligibilité de la parole au moyen de l'indice de transmission de la parole

1 Domaine d'application

La présente partie de la CEI 60268 spécifie des méthodes objectives d'évaluation de la qualité de transmission de la parole sous l'angle de l'intelligibilité.

L'objectif de la présente Norme est de mettre un manuel complet à la disposition de tous les types d'utilisateurs de la méthode STI dans les domaines de l'audio, des communications et de l'acoustique.

La présente Norme ne donne pas de critères STI destinés à la certification des canaux de transmission (par exemple, critères pour un système d'alarme vocale).

Trois méthodes, étroitement liées, sont présentées sous le nom de STI, STIPA, et STITEL. Les deux premières méthodes sont destinées à l'évaluation des performances de la transmission de la parole avec ou sans systèmes électroacoustiques. La méthode STITEL a des utilisations plus restreintes.

NOTE Aucune de ces méthodes n'est adaptée à la mesure et à l'évaluation des systèmes de confidentialité des conversations ou de masquage de la parole, car le STI n'a pas été validé pour des conditions représentatives des applications assurant la confidentialité des conversations [1]¹.

Les informations suivantes sont incluses:

- les techniques de mesure;
- les techniques de prédiction.

2 Références normatives

Les documents de référence suivants sont indispensables 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 tous ses éventuels amendements).

CEI 61260:1995, *Electroacoustique – Filtrés de bande d'octave et de bande d'une fraction d'octave*
Amendement 1 (2001)

ISO 18233:2006, *Acoustique – Application de nouvelles méthodes de mesurage dans l'acoustique des bâtiments et des salles*

3 Termes et définitions

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

¹ Les chiffres entre crochets renvoient à la Bibliographie.

3.1

intelligibilité de la parole (*Speech Intelligibility*)

évaluation de la proportion de la parole qui est comprise

3.2

qualité de la parole (*Speech quality*)

évaluation de la qualité sonore d'un signal de parole

3.3

indice de transmission de la parole (*Speech Transmission Index*)

STI

métrique, comprise entre 0 et 1, représentant la qualité de transmission de la parole par rapport à l'intelligibilité par un canal de transmission de la parole

3.4

indice d'intelligibilité de la parole (*Speech Intelligibility Index*)

SII

méthode objective pour la prédiction de l'intelligibilité de la parole basée sur l'indice d'articulation

3.5

méthode STI

FULL STI (MÉTHODE STI COMPLÈTE)

méthode objective de prédiction et de mesure de l'indice de transmission de la parole utilisant 14 fréquences de modulation sur une gamme de 7 bandes d'une octave

3.6

distorsion

toute déformation involontaire et généralement non désirée d'un signal se produisant dans un canal de transmission de la parole

NOTE La distorsion peut comprendre des effets linéaires et non linéaires, tant dans le domaine fréquentiel que dans le domaine temporel.

3.7

indice de transmission de la parole pour systèmes de sonorisation (*Speech Transmission Index for Public Address systems*)

STIPA

méthode obtenue par une version simplifiée de la méthode STI mais encore sensible aux distorsions apparaissant dans l'acoustique des salles et/ou dans les systèmes de sonorisation

NOTE La méthode STIPA est appliquée comme une méthode directe.

3.8

indice de transmission de la parole pour systèmes de télécommunications (*Speech Transmission Index for Telecommunication systems*)

STITEL

méthode obtenue par une version simplifiée de la méthode STI, mais encore sensible aux distorsions apparaissant dans les systèmes de communication

NOTE La méthode STITEL est appliquée comme une méthode directe.

3.9

indice de transmission de la parole relatif à l'acoustique de la salle (*Room Acoustical Speech Transmission Index*)

RASTI

méthode obtenue par une version simplifiée de la méthode STI, qui est uniquement utilisée à des fins de filtrage et concentrée sur la communication directe entre des personnes, sans utilisation d'un système de communication électroacoustique

NOTE 1 La méthode RASTI tient compte des perturbations et distorsions acoustiques dans le domaine temporel (échos, réverbération).

NOTE 2 La méthode RASTI est désormais désuète.

3.10

méthode STI directe

méthode utilisant des signaux d'essai modulés (similaires à la parole) pour mesurer directement la fonction de transfert de modulation

3.11

méthode STI indirecte

méthode utilisant la réponse impulsionnelle et l'intégrale de l'énergie directe (l'intégrale de Schroeder) pour obtenir la fonction de transfert de modulation

3.12

canal de transmission de la parole

chemin du signal acoustique ou électroacoustique entre un locuteur et un auditeur

3.13

système de sonorisation (*Public Address system*)

PA

système de distribution électronique du son, utilisant des microphones, amplificateurs et haut-parleurs pour renforcer ou amplifier un son donné (tel qu'une annonce ou un message préenregistré) et distribuant le son dans un bâtiment ou un espace

3.14

système d'alarme vocale (*Voice Alarm System*)

VAS

système de distribution du son diffusant des messages et/ou des signaux d'avertissement, en cas d'urgence

3.15

niveau réel de la parole (*real speech level*)

niveau de signal pour un signal vocal, exprimé en dB A, où seuls sont pris en compte les segments contribuant au signal vocal; les pauses et silences entre les mots et les phrases sont ignorés

NOTE Voir également l'Annexe J.

3.16

niveau de référence de la parole (*reference speech level*)

niveau de parole équivalant à 60 dB A à une distance d'un mètre en face de la bouche du locuteur

3.17

effort vocal

effort exercé par le locuteur, quantifié objectivement par le niveau vocal pondéré A à une distance de 1 m de la bouche et qualifié subjectivement par une description

3.18

bouche artificielle

dispositif constitué d'un haut-parleur monté sur une enceinte et ayant un diagramme directionnel et un diagramme de rayonnement semblables à ceux de la bouche humaine moyenne

NOTE Le degré de ressemblance requis ne peut pas être facilement spécifié et dépend de l'application particulière. Voir par exemple l'UIT-T P.50 [47].

3.19

locuteur de langue maternelle étrangère

personne parlant une langue différente de celle apprise comme langue principale dans l'enfance du locuteur

3.20

seuil absolu de réception de la parole

seuil absolu d'audition accru de la gamme dynamique minimale requise pour une reconnaissance correcte de la parole

3.21

masquage auditif (*auditory masking*)

processus par lequel le seuil d'audition (d'audibilité) pour un son est augmenté par la présence d'un autre son (de masquage)

NOTE Dans le cadre de la méthode STI, le masquage auditif est également appelé " dispersion du masquage".

3.22

oreille artificielle

dispositif présentant des caractéristiques semblables à celles de l'oreille humaine pour la réception de signaux acoustiques

NOTE Voir la CEI 60318 [2].

3.23

fonction intensité

carré du signal d'amplitude en fonction du temps

3.24

fonction enveloppe

enveloppe de la fonction intensité

3.25

spectre enveloppe

contribution relative des composantes spectrales de la fonction enveloppe

3.26

fréquence de modulation

fréquence de la variation sinusoïdale de la fonction enveloppe

NOTE La fréquence de modulation f_m est exprimée en Hertz (Hz).

3.27

indice de modulation

valeur, comprise entre 0 et 1, décrivant la profondeur d'une modulation sinusoïdale de la fonction intensité

3.28

rapport de transfert de la modulation (*modulation transfer ratio*)

rapport entre la profondeur de modulation du signal reçu et la profondeur de modulation du signal d'origine (émis), la profondeur de modulation étant celle de la fonction intensité

3.29

fonction de transfert de modulation (*Modulation Transfer Function*)

MTF

rapports de transfert de modulation pour les différentes fréquences de la fréquence de modulation

3.30**facteur de pondération de bandes d'une octave** α

contribution relative de chaque bande d'une octave à l'indice de transmission de la parole

3.31**facteur de redondance de bandes d'une octave** β

fraction de chevauchement d'informations entre deux bandes d'une octave adjacentes sous l'angle de l'intelligibilité de la parole

3.32**bruit de fond**

ensemble de tous les sons, bruit inclus, restant en l'absence de signal de parole ou d'essai

3.33**bruit fluctuant**

son ou bruit continu dont le niveau de pression sonore varie sensiblement, mais pas de manière impulsive, pendant la période d'observation

3.34**bruit impulsif**

son ou bruit caractérisé par de brèves remontées de la pression sonore

3.35**rapport signal sur bruit** (*Signal-to-Noise Ratio*) SNR

différence entre le niveau de pression sonore de la parole ou du signal d'essai et le niveau de pression sonore du bruit de fond, les niveaux de pression sonore étant déterminés avec une pondération normalisée des fréquences

NOTE Le rapport signal sur bruit SNR est exprimé en décibels (dB).

3.36**rapport signal sur bruit réel** (*effective Signal-to-Noise Ratio*) SNR_{eff}

différence entre le niveau de la modulation d'intensité et le niveau de l'intensité de toutes les distorsions d'un signal d'essai STI reçu

NOTE 1 Le rapport signal sur bruit réel est exprimé en décibels (dB).

NOTE 2 Des exemples de distorsions sont les niveaux des champs de réverbération, les niveaux des bruits ambiants, les niveaux des distorsions non linéaires et les niveaux des masquage.

3.37**facteur de crête**

différence entre les niveaux de pression sonore de crête et quadratique moyen pendant un intervalle de temps donné

NOTE Le facteur de crête est exprimé en décibels (dB).

3.38**effet Lombard**

augmentation spontanée de l'effort vocal induite par l'augmentation du niveau du bruit ambiant à l'oreille du locuteur

NOTE Le décalage de la tonie vocale à des niveaux de locution plus élevés n'est pas pris en compte ici.

3.39

filtre d'octave fractionnaire

filtre passe-bande pour lequel le rapport de la fréquence de coupure supérieure f_2 à la fréquence de coupure inférieure f_1 est de deux à un exposant égal à la fraction d'une bande d'une octave

NOTE 1 En symboles, le rapport des fréquences de coupure est $f_2/f_1 = 2^{1/b}$, où $1/b$ désigne la fraction d'une octave.

EXEMPLE 1: Pour les filtres demi-octaves, le rapport des fréquences est de $2^{1/2} = \sqrt{2}$.

EXEMPLE 2: Pour les filtres d'octaves, le rapport des fréquences est de 2.

NOTE 2 Pour davantage d'informations, on se référera à la CEI 61260.

3.40

pression sonore de référence

p_0

pression sonore, choisie par convention, comme étant égale à 20 μPa pour un son aérien

3.41

niveau de pression sonore

vingt fois le logarithme décimal du rapport de la pression sonore quadratique moyenne sur la pression sonore de référence

NOTE Le niveau de pression sonore est exprimé en décibels (dB). La notation est L_p .

3.42

niveau de pression sonore équivalent continu (*equivalent continuous sound pressure level*)

dix fois le logarithme décimal du rapport du niveau de pression sonore quadratique moyen, pour un intervalle de temps donné, sur le carré de la pression sonore de référence

NOTE 1 Le niveau de pression sonore, $L_{eq,T}$, est donné par l'équation suivante:

$$L_{eq,T} = 10 \lg \frac{\frac{1}{T} \int_{t_1}^{t_2} p^2(t) dt}{p_0^2}$$

où

$p(t)$ est la pression sonore instantanée au temps t ;

t est la variable d'intégration pour le temps;

$T = t_2 - t_1$ est longueur de l'intervalle de temps, pour laquelle le niveau de pression sonore continu est déterminé; et

p_0 est la pression sonore de référence (20 μPa).

Dans l'équation donnée, le numérateur dans l'argument du logarithme est la pression sonore quadratique moyenne pour le temps d'intégration T .

NOTE 2 Par principe, il n'est appliqué aucune pondération temporelle lors de la détermination du niveau de pression sonore continu.

NOTE 3 Pour d'autres définitions, voir la CEI 61672 [3].

3.43

niveau de centile

dix fois le logarithme décimal du rapport du niveau de pression sonore quadratique moyen, dépassé pendant une partie donnée du temps de mesure, au carré de la pression sonore de référence; la pression sonore quadratique moyenne est déterminée avec une pondération normalisée temporelle et fréquentielle, par exemple, L_{10} ou L_{A10}

NOTE 1 Dans le cadre d'application de la présente Norme, la pondération temporelle "Fast" est à appliquer pour la détermination du niveau de centile.

NOTE 2 L_{A10} est le niveau de pression sonore pondérée A et Fast dépassé dans 10 % du temps de mesure.

3.44

indice de transfert de modulation (*Modulation Transfer Index*)

MTI

moyenne non pondérée des rapports signal sur bruit réel mis à l'échelle, pour une bande d'octave donnée

3.45

niveau de parole opérationnel

niveau de pression sonore d'un signal vocal qui sera utilisé ou qui est trouvé dans la situation ou le contexte de la mesure

3.46

niveau de bruit de fond opérationnel

niveau de pression sonore de bruit de fond qui sera présent ou qui est trouvé dans la situation ou le contexte de la mesure

NOTE Ce niveau est utilisé pour les prédictions et le post-traitement des mesures.

4 Description de la méthode STI

4.1 Généralités

4.1.1 Justification de la méthode STI

La méthode STI a été développée comme une méthode d'essai rapide et objective pour déterminer la qualité de transmission de la parole de canaux de transmission de la parole. Avec l'indice de transmission de la parole, l'intelligibilité de la parole peut être prédite pour différents types de formats de mots et de phrases pour une vaste gamme de systèmes de transmission de la parole.

Dans la parole, l'intensité du signal varie en fonction du temps en produisant une variation de l'enveloppe de l'intensité de la parole. Les fluctuations lentes de l'enveloppe de l'intensité correspondent aux limites de mots et des phrases alors que les fluctuations rapides coïncident avec les phonèmes individuels au sein des mots. Dans le concept STI, la préservation de l'enveloppe de l'intensité est considérée comme étant de la plus haute importance.

Contrairement à l'approche d'origine faisant appel à l'indice d'articulation [4], qui se fonde sur les rapports signal sur bruit dans différentes bandes spectrales de la parole, la mesure STI détermine comment l'enveloppe de l'intensité du signal de la parole est affectée par un canal de transmission. Une fonction de transfert de modulation (MTF) est déterminée, qui quantifie la manière dont le canal affecte l'enveloppe de l'intensité du signal de la parole.

La méthode STI produit une métrique sur une échelle de 0 à 1, fondée sur les contributions pondérées d'une gamme de bandes de fréquences présentes dans la parole.

La méthode STI et ses dérivées (voir ci-dessous) peuvent être utilisées pour déterminer l'intelligibilité potentielle d'un canal de transmission de la parole en divers emplacements et dans diverses conditions. En particulier, l'effet des modifications des propriétés acoustiques des espaces peut être évalué.

4.1.2 Applicabilité de la méthode STI

La méthode STI est une méthode objective et validée de mesure de la qualité de transmission de la parole pour une large gamme de distorsions acoustiques et électroacoustiques qui ont un effet sur l'intelligibilité. Cependant, comme il s'agit d'une simplification de la parole humaine, le modèle STI peut ne pas s'appliquer partout. Des utilisateurs de la méthode STI

qui l'emploieraient au-delà de ses limites actuelles pourraient obtenir des prédictions inexactes de l'intelligibilité. Une vue d'ensemble des applications et des limitations est donc donnée, pour d'aider les utilisateurs du STI à choisir la méthode STI la mieux adaptée à leur application et obtenir ainsi les résultats les plus significatifs et les plus exacts.

La méthode STI a été validée pour une sortie acoustique utilisant un seul microphone omnidirectionnel. L'utilisation d'un microphone directionnel donne des résultats différents et qu'il n'est pas possible de corrélérer; elle n'est habituellement pas à conseiller. Des informations supplémentaires sont données en 7.10.

Si la situation ou le canal de transmission ne permet pas l'utilisation des méthodes STI, d'autres méthodes d'évaluation de l'intelligibilité doivent être utilisées. Il existe d'autres méthodes pour évaluer la qualité de la communication parlée, chacune de ces méthodes ayant ses avantages et inconvénients et donc des utilisateurs différents. L'Annexe N décrit un certain nombre d'autres mesures de l'intelligibilité.

4.2 Notions de base pour la méthode STI

4.2.1 Généralités

Le concept d'indice STI est fondé sur la constatation empirique que ce sont les fluctuations des signaux de parole qui transportent les informations les plus pertinentes pour l'intelligibilité de la parole, voir [5], [6] et [7]. Les fluctuations de la parole résultent de la séparation acoustique des phrases, mots et phonèmes, qui sont les éléments fondamentaux de la parole. Les fluctuations, appelées modulations, peuvent être quantifiées en fonction de la fréquence de modulation F produisant le spectre de modulation. Pour la parole en clair, les fréquences de modulation s'étendent typiquement de 0,5 Hz jusqu'à 16 Hz avec une modulation maximale à 3 Hz approximativement.

Toute détérioration du spectre de modulation par le canal de transmission est considérée comme aboutissant à une réduction de l'intelligibilité de la parole. Cette détérioration du spectre de modulation correspond à une réduction de la profondeur de modulation à une ou plusieurs fréquences de modulation; elle est calculée comme valeur de la transmission de la modulation pour chaque bande d'une octave, sur la bande spectrale de la parole. La Figure 1 montre le concept de réduction de la modulation qui peut se produire entre un locuteur et un auditeur.

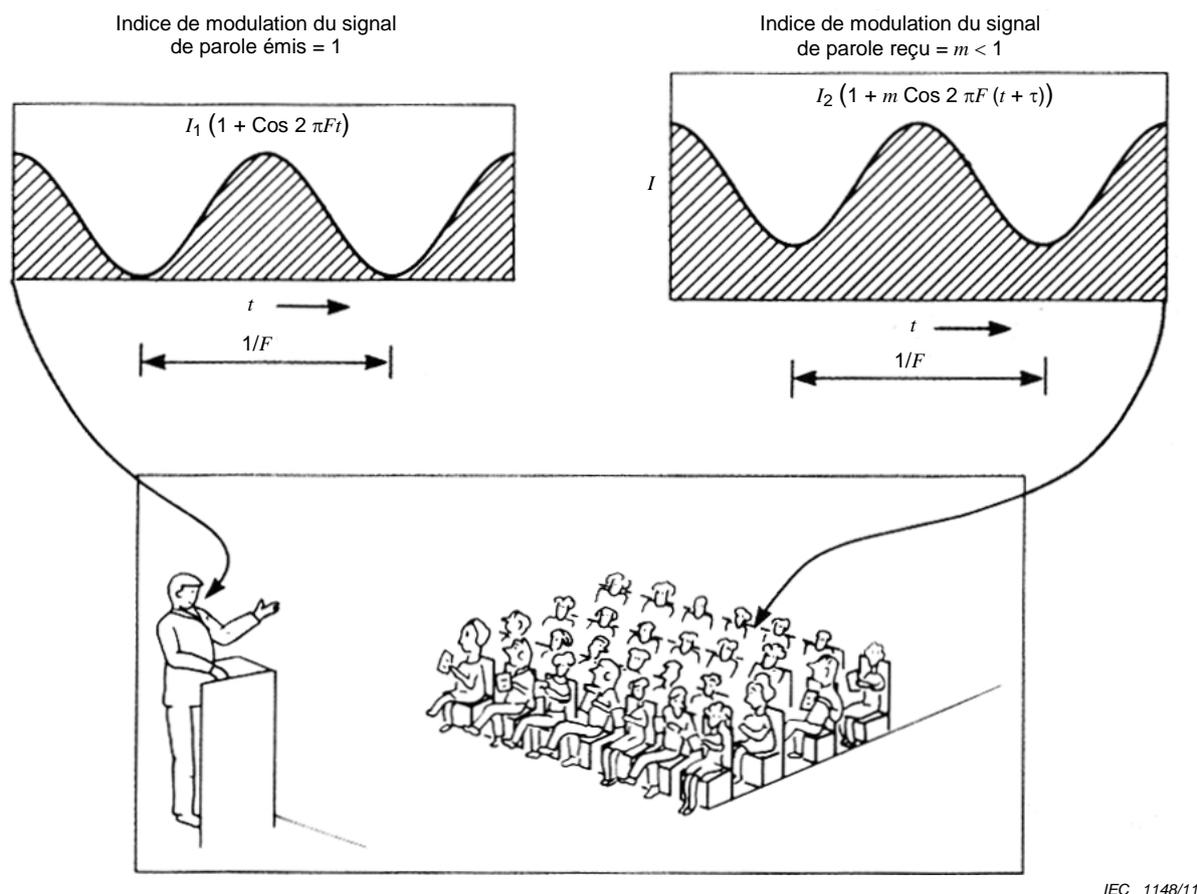


Figure 1 – Concept de réduction de la modulation due au canal de transmission

La méthode STI a été optimisée et validée avec des expérimentations de l'intelligibilité, notée par des sujets, de mots CVC (consonne-voyelle-consonne) (en Hollandais) pour une grande diversité de distorsions de canaux de transmission. Ces distorsions sont notamment le bruit, la réverbération, les échos, la distorsion non linéaire et les techniques de codage numérique.

Le signal d'essai STI a été mis au point en utilisant des paramètres dérivés du matériau vocal. En général, le signal d'essai comprend sept signaux de bruit chacun d'une octave, couvrant les octaves centrées sur 125 Hz jusqu'à 8 kHz. Chaque porteuse de bruit est modulée avec une ou plusieurs fréquences de modulation à des intervalles d'un tiers d'octave dans la gamme de 0,63 Hz jusqu'à 12,5 Hz compris.

La méthode STI, décrite dans l'Annexe A, détermine la fonction de transfert de modulation $m(F)$ du canal de transmission. Il est obtenu un total de 98 résultats correspondant aux 14 fréquences de modulation et aux sept bandes d'une octave (voir Figure A.3). Le niveau quadratique moyen de chaque porteuse d'une octave est en proportion du niveau relatif du spectre moyen de long terme du matériau vocal (voir aussi 4.5 pour de plus amples informations). Chaque bande d'une octave apporte une contribution à l'intelligibilité de la parole, pondérée selon la bande en question. La somme pondérée de ces valeurs d'indice de transmission détermine la valeur de l'indice STI global pour le canal de transmission.

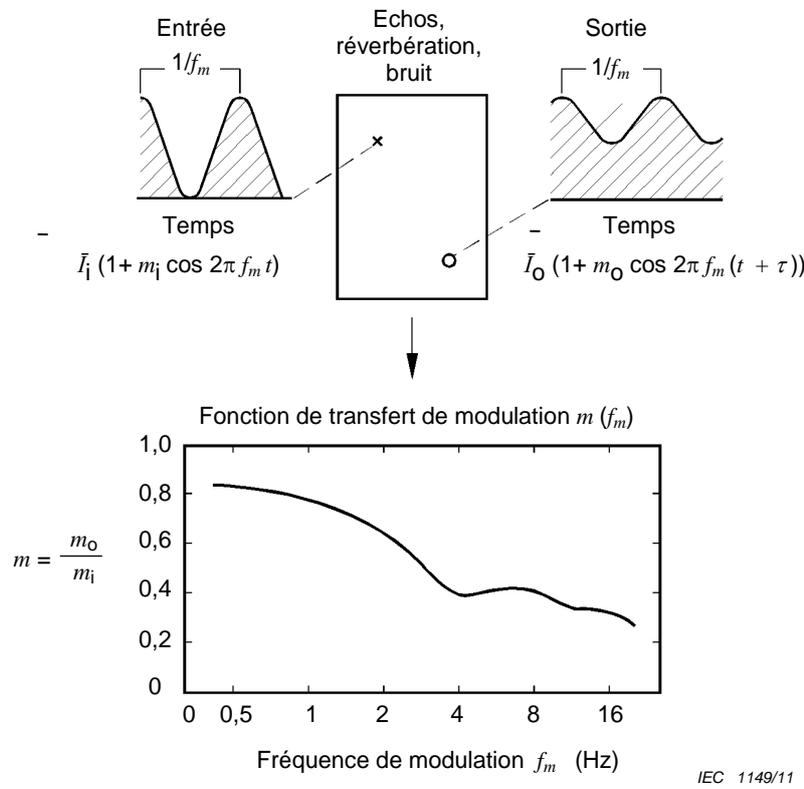
Des travaux de recherche [4] montrent que des bandes d'une octave adjacentes contiennent des informations redondantes sous l'angle de l'intelligibilité de la parole. Si une bande d'une octave ne contribue pas à l'intelligibilité (par exemple, par masquage par rapport à la réverbération ou au bruit de fond), les bandes d'une octave voisines peuvent compenser partiellement cette contribution manquante. Cette connaissance a conduit à l'utilisation de facteurs de redondance dans la méthodologie STI.

4.2.2 Vue d'ensemble de la théorie

L'indice de modulation m_i d'un signal d'essai est joué dans une salle ou sur un canal de communication et est reçu à l'emplacement d'un auditeur comme indice de modulation m_o . Pour mesurer, par exemple, le STI dans la situation illustrée à la Figure 1, le signal d'essai serait émis par une source sonore simulant un locuteur humaine située à l'emplacement du locuteur avec un microphone récepteur d'essai placé à n'importe quel emplacement d'auditeur.

Les caractéristiques essentielles de la source sonore sont les dimensions physiques, la directivité, la position, le niveau de pression sonore et la réponse en fréquence.

Le signal d'essai type est une porteuse dont le spectre dans le domaine fréquentiel a la forme d'un spectre de parole, modulé en intensité par un signal sinusoïdal de fréquence f_m (voir Figure 2).



NOTE m_i et m_o sont respectivement les indices de modulation des signaux d'entrée et de sortie. \bar{I}_i et \bar{I}_o sont les intensités d'entrée et de sortie, ces intensités étant égales au carré des niveaux de pression sonore (p^2).

Figure 2 – Fonction de transfert de modulation – comparaison entrée/sortie

La baisse de la profondeur de modulation à la fréquence f_m est quantifiée par la fonction de transfert de la modulation $m(f_m)$, déterminée par

$$m(F) = \frac{m_o}{m_i}$$

Cette baisse est interprétée en termes de rapport signal sur bruit réel, SNR_{eff} , (indépendamment de la cause de la baisse qui peut être dû à la réverbération, aux échos, aux composantes de distorsion non linéaire ou au bruit d'interférences). Elle est déterminée par

$$SNR_{\text{eff}} = 10 \lg \left(\frac{m(F)}{1 - m(F)} \right)$$

Les valeurs du rapport réel signal sur bruit sont limitées à la gamme de -15 dB à +15 dB. Les valeurs inférieures à 15 dB sont notées -15 dB et les valeurs supérieures à 15 dB sont notées 15 dB.

L'indice de transmission de la parole STI combine les valeurs de l'indice de transfert de modulation tirées de mesures dans sept bandes d'une octave en une seule valeur pondérée globale.

L'Annexe A fournit une description plus détaillée du calcul de l'indice de transmission de la parole.

4.2.3 Mesure du STI

Le FULL STI (MÉTHODE STI COMPLÈTE) est fondé sur un ensemble complet de 98 (7 x 14) indices de modulation.

Deux formes simplifiées du STI, fondées sur un nombre plus petit d'indices de modulation, sont STIPA et STITEL (voir Article 5).

Le STIPA consiste en un signal d'essai avec un ensemble prédéfini de deux modulations par bande d'une octave qui sont générées simultanément pour donner un total de 14 indices de modulation.

Le STITEL consiste en un signal d'essai avec un ensemble prédéfini de sept fréquences de modulation, une par bande d'une octave, qui sont générées simultanément pour donner un total de sept indices de modulation.

Il existe deux types de méthodes pour mesurer le STI:

- les méthodes directes utilisant des signaux d'essai modulés;
- les méthodes indirectes basées sur la réponse impulsionnelle du système utilisant l'équation de Schroeder.

Les méthodes directes utilisant le STIPA et le STITEL ont des durées de mesure sensiblement plus courtes que la méthode directe FULL STI. Noter que la méthode directe FULL STI est rarement utilisée en pratique.

L'Annexe B et l'Annexe C donnent respectivement des descriptions détaillées de la méthode STIPA et de la méthode STITEL. L'Annexe D fournit des détails relatifs à la méthode RASTI, désormais désuète.

La méthode STI, qu'elle soit directe ou indirecte, s'est avérée donner des résultats valides pour un grand nombre de distorsions linéaires, tant dans le domaine temporel que dans le domaine fréquentiel. Les distorsions suivantes sont prises en compte par la méthode STI:

- distorsion temporelle, par exemple, réverbération et échos;
- bruit;
- forte distorsion spectrale, par exemple, filtrage passe-bande.

NOTE Certains types de distorsions spectrales peuvent ne pas être pris en compte, voir 4.5.8.

De plus, les méthodes STI directes prennent en compte des distorsions non linéaires, par exemple l'écrêtage, alors que les méthodes indirectes ne valent que pour des systèmes linéaires. Des informations supplémentaires sur les effets de distorsions non linéaires sont

données à l'Article 6. Le Tableau 1 donne une vue d'ensemble des méthodes d'essai STI en fonction des types de distorsion linéaire et non linéaire pour lesquels elles sont appropriées.

Tableau 1 – Comparaison des méthodes d'essai STI pour différents types de distorsion

Méthode	Type de distorsion			
	Bruit	Réverbération, échos	Distorsion non linéaire	Distorsion spectrale ^a
FULL STI directe	oui	oui	dépendant des conditions	oui
STIPA directe	oui	oui	dépendant des conditions	oui
STITEL directe	oui	dépendant des conditions	dépendant des conditions	oui
FULL STI indirecte utilisant la séquence de type longueur maximale (MLS, Maximum Length Sequence) ^e	oui ^b	oui	non	oui
FULL STI indirecte utilisant le signal sinusoïdal de balayage ^c	non ^d	oui	non	oui
<p>NOTE Le terme "dépendant des conditions" sert à indiquer que le type de signal d'essai correspondant peut ou peut ne pas produire des résultats suffisamment exacts, en fonction du type exact de distorsion. Par exemple:</p> <ul style="list-style-type: none"> • La mise à zéro des petites valeurs (<i>centre clipping</i>) est peu susceptible d'avoir le moindre effet sur la profondeur de modulation, alors que l'écrtage par coupure des crêtes réduit la profondeur de modulation, mais a en général peu d'effet sur l'intelligibilité de la parole et, donc, la valeur STI mesurée peut être pessimiste. • La méthode STITEL peut être utilisée dans des environnements réverbérants, à condition que le temps de réverbération ne dépende pas fortement de la fréquence. • De même, la méthode STIPA peut être utilisée pour les systèmes PA produisant des composantes de distorsion non linéaire, à moins que le signal ne soit fortement écrté dans ses diverses bandes de fréquences. 				
<p>^a La réponse en fréquence du canal de transmission peut produire une perte d'intelligibilité perçue, dont il n'est pas correctement tenu compte dans le résultat, voir 4.5.8.</p> <p>^b Le moyennage de signal des données temporelles ne doit pas être utilisé et le spectre d'excitation doit ressembler à celui de la parole.</p> <p>^c Cela inclut la spectrométrie par retard temporel.</p> <p>^d Les effets du bruit peuvent toutefois être calculés mathématiquement.</p> <p>^e Théoriquement, un autre signal de pseudo-bruit (à phase aléatoire) mathématiquement déterministe pourrait être employé.</p>				

4.3 Applicabilité des méthodes d'essai STI

Le Tableau 2 donne une vue d'ensemble des formes de STI recommandées pour divers types d'application. Les symboles + et – représentent une indication générale de l'adéquation de la méthode.

Si des parties significatives de la population d'auditeurs sont de langue maternelle étrangère et/ou sont âgés, il convient d'interpréter le STI comme indiqué à l'Annexe H.

Tableau 2 – Applicabilité de l'essai

Application	Essai recommandé	FULL STI ^a	STIPA	STITEL	Limitations	Solutions de rechange
Évaluation de l'adéquation de l'acoustique de la salle pour la communication parlée (sans amplification électronique)	STIPA	++	++	+/-	L'adéquation de la méthode STITEL dépend de la réverbération.	
Évaluation des systèmes PA et VA (Voice Alarm, Alarme vocale)	STIPA	+	++	+/-	L'adéquation de la méthode STITEL dépend de la réverbération et des échos.	
Évaluation des canaux de télécommunications (téléphone, radio)	STITEL	+	+	++	La méthode STITEL présente une puissance de diagnostic supérieure.	
Le canal présente une compression d'amplitude	STIPA	+	+	+		
La différence entre voix masculine et voix féminine nécessite une attention spécifique	FULL STI	++	-	+	La méthode STIPA n'est pas adaptée au spectre féminin (spectre masculin uniquement).	
Mise à zéro des petites valeurs "fortes" (<i>strong centre clipping</i>)	Aucun	-	-	-		aucune
Bruit fortement fluctuant	STIPA	+/-	+/-	+/-		Communiquer plusieurs mesures de STI.
Parole et bruit clairement séparés dans l'espace ou une forte composante de champ direct existe dans un environnement hautement réverbérant	STI	+	+/-	+/-	A utiliser avec précaution. Les méthodes actuellement normalisées sont inexactes.	Voir 7.10
Canaux ne permettant pas les signaux d'essai artificiels, tels que les vocodeurs	Aucun	+/-	+/-	+/-	Les méthodes actuellement normalisées sont inexactes.	^b Utiliser un signal d'essai STI vocal ou des essais sur des auditeurs.
++ méthode très bien adaptée, + méthode bien adaptée, +/- méthode adaptée, - méthode inadaptée						
a Voir le Tableau 1 concernant l'adéquation des méthodes de mesure.						
b Il s'agit d'une méthode directe qui peut être incluse dans une future édition de la norme.						

4.4 Utilisation des méthodes directes et indirectes

Le Tableau 3 ci-dessous compare un certain nombre de questions pratiques ayant trait à l'utilisation des méthodes de mesure directes et indirectes.

Tableau 3 – Choix de la méthode

Objet	Méthode directe	Méthode indirecte
Post-traitement	possible	obligatoire
Dispositif tenu à la main	possible	possible
Non linéarités de l'amplitude	réduisent la fiabilité du résultat	réduisent la fiabilité du résultat
Non linéarités de la fréquence (spectre irrégulier) ^a	possible	possible
Glissement de fréquence	pas possible	pas possible
Suppression du bruit	non	oui
Précision de la fréquence d'échantillonnage entre les fréquences d'horloge de la source du signal et du récepteur	erreurs inférieures à 20×10^{-6}	erreurs inférieures à $0,5 \times 10^{-6}$
^a Voir 4.5.8 pour de plus amples détails.		

4.5 Limitation de la méthode STI

4.5.1 Généralités

La méthode STI refléterait idéalement tous les changements dans un canal de transmission qui sont pertinents pour l'intelligibilité de la parole. Cependant, il est essentiel de réaliser que l'approche de la modélisation STI reste une simplification du traitement humain. De plus, le signal d'essai STI diffère de la parole réelle par des aspects temporels et spectraux plus subtils, tels que:

- la gamme dynamique de la parole, qui dépend du temps d'intégration;
- la distribution de l'énergie dans chaque tranche de temps;
- la distribution des niveaux de signal sur toute la longueur d'un segment de parole ou d'un signal d'essai (percentiles de dépassements);
- l'absence de trous dans le signal d'essai;
- les porteuses présentes dans la parole ne sont pas limitées aux bandes de porteuses et aux fréquences de modulation fixées;
- les différences spectrales entre les mots pris un par un et le signal STI;
- les différences spectrales entre divers locuteurs.

NOTE Le spectre de parole spécifié pour la méthode STI diffère du spectre spécifié par l'ANSI [4].

Par conséquent, il faut faire preuve de prudence pour utiliser la méthode STI dans certaines situations et pour certains canaux de transmission (bande étroite). Dans un certain nombre de cas, l'intelligibilité peut ne souffrir que peu de la distorsion, alors que la méthode STI montre une réduction considérable d'intelligibilité. Dans d'autres cas, où la méthode STI ne montre que de petits changements, l'intelligibilité peut être considérablement réduite. Les articles ci-après débattent plus en détail des limitations potentielles.

4.5.2 Glissements de fréquence

Ce type de distorsion peut se produire avec

- une lecture d'un signal numérique à la mauvaise fréquence d'échantillonnage,

- des dispositifs évitant l'accrochage acoustique (anti-Larsen),
- des émissions radio à bande latérale unique.

Les glissements de fréquence peuvent avoir un effet important sur la méthode STI mais ont généralement un effet faible sur l'intelligibilité, si bien que cette méthode peut sous-estimer l'intelligibilité de systèmes présentant des glissements de fréquence.

4.5.3 Mise à zéro des petites valeurs ("*centre clipping*")

Ce type de distorsion peut se produire lorsque des parties de bas niveau d'un signal ne sont pas transmises fidèlement ou sont rendues muettes. Cela peut se produire dans des amplificateurs ou sur des connecteurs corrodés. La méthode STI surestime l'intelligibilité des systèmes qui présentent de forts effets de mise à zéro des petites valeurs.

NOTE La mise à zéro des petites valeurs est également appelée "distorsion au passage par zéro" et "distorsion à l'origine" ("*crossover distortion*" ou "*origin distortion*")

4.5.4 Pertes de niveau ("*drop outs*")

Une perte de niveau d'un signal, à des intervalles réguliers, peut résulter de phénomènes d'évanouissement sélectifs dans les transmissions sans fil et de la corruption des signaux numériques. Le STI peut ne pas être trop réduit, mais l'intelligibilité peut être très médiocre. L'analyse de la structure fine du signal modulé reçu est recommandée, afin de détecter et marquer ces pertes de niveau et, dans la mesure du possible, permettre le calcul du STI après élimination de ces pertes de niveau.

4.5.5 Gigue

Les décalages temporels de la parole, tels qu'appliqués en transmission numérique pour compenser les variations de la vitesse de transmission, n'ont aucun effet sur l'intelligibilité, mais peuvent réduire fortement le STI, si bien que ce dernier peut sous-estimer l'intelligibilité pour des systèmes présentant de la gigue.

4.5.6 Vocodeurs

Bien que les codeurs vocaux numériques aient peu d'influence sur l'intelligibilité, en fonction du type des codecs utilisés, le STI tend à augmenter. En cas de faible intelligibilité, il est recommandé d'utiliser des signaux d'essai de parole ou des mesures de l'intelligibilité perçue par des sujets.

STI ne doit pas être employé pour mesurer des systèmes tels que des vocodeurs qui codent des segments de parole, avec par exemple les techniques de codage linéaire prédictif avec synthèse par livre de code (code book) ou l'introduction d'erreurs sur les fragments de parole voisés/ non-voisés et d'erreurs de tonie.

4.5.7 Surestimation de l'indice STI dans des conditions de faible bruit de fond

Il convient de noter que le modèle STI suppose intrinsèquement un rapport signal sur bruit non infini dans chaque bande d'une octave, car le seuil de réception d'audition dans le modèle fonctionne comme une source de bruit de fond. Si les niveaux de bruit de fond ou les valeurs de seuil de réception sont fixés à zéro au cours des mesures ou des simulations, les valeurs du STI peuvent être trop élevées.

A titre d'exemple, cette question se pose lors de l'investigation mathématique du comportement du STI en fonction des changements du spectre du signal d'entrée. Si une matrice MTF, où chaque valeur est égale à l'unité (à savoir, absence de contamination par la réverbération ou le bruit de fond), est utilisée avec un signal d'entrée qui s'écarte du spectre de parole spécifié, le résultat STI montre souvent peu de variation, même en cas de fortes variations du spectre d'entrée, voir [8].

Pour cette raison il convient que les mesures de l'indice de transmission de la parole soient faites avec un bruit de fond réaliste pour ce type d'application. Par exemple, il convient que les mesures de la sortie acoustique utilisent un bruit de fond réaliste de même des seuils de réception de la parole (*speech reception thresholds*).

4.5.8 Réponse en fréquence

Jusqu'à présent, les travaux de recherche [8], [9], [10], [11], [12], [13] indiquent que la réponse en fréquence du canal de transmission, (qui se manifeste comme l'équilibre tonal perçu de la parole) est beaucoup plus importante pour l'intelligibilité perçue que ne l'indiquent les mesures du STI, particulièrement en présence de réverbération. Si la réponse en fréquence n'est pas sensiblement plate, il est possible que le STI puisse indiquer des valeurs trop élevées en comparaison de l'intelligibilité perçue.

Il a été rapporté le cas de systèmes avec des valeurs mesurées du STI supérieures à 0,5, dans lesquels l'intelligibilité perçue de la parole s'est avérée inadéquate, en raison d'une médiocre réponse en fréquence / d'un médiocre équilibre tonal du système. L'application d'une égalisation, afin de corriger la réponse en fréquence, a amélioré l'intelligibilité perçue.

Vu cette limitation de la méthode STI, une solution appropriée pour assurer une réponse uniforme en amplitude est de faire une mesure séparée de la réponse amplitude / fréquence du système avec de préférence une résolution supérieure à celle des largeurs de bandes d'une octave. Néanmoins, il existe des facteurs importants qui peuvent ne pas être inclus dans ces mesures:

- La réponse en fréquence déduite des mesures de la réponse impulsionnelle dépend fortement de la longueur des données temporelles utilisée pour la mesure et de la fenêtre temporelle de filtrage appliquée à ces données.
- Pour nombre d'environnements acoustiques, il n'existe pas de mesure qui soit bien corrélée à l'équilibre tonal perçu. Par exemple, dans des situations avec une faible réverbération, l'influence de la réponse en champ direct sur l'équilibre tonal est typiquement beaucoup plus forte que dans les environnements très réverbérants, où la réponse en puissance de la source devient prépondérante.
- L'influence des changements de position du locuteur sur la réponse en fréquence du microphone.

Certains praticiens des systèmes électroacoustiques ont indiqué que de petites modifications de la réponse en fréquence des systèmes électroacoustiques, qui réduisent la coloration audible de la parole, peuvent réduire le degré de concentration qu'un auditeur doit avoir pour arriver à une intelligibilité satisfaisante. Cela peut être particulièrement important dans des situations d'écoute longue ou dans le cas d'un locuteur ou auditeur de langue maternelle étrangère. Des exemples de colorations sont la présence de crêtes de bande étroite, ou de résonances: des ajustements apportés au système aussi faibles que 1 dB sur une largeur de bande aussi étroite que 1/3 octave se sont avérés bénéfiques pour l'intelligibilité perçue [9], [10].

4.5.9 Echos

Il a été rencontré des situations où des échos audibles (réflexions tardives) provoquent une perte sensible d'intelligibilité perçue de la parole alors que les valeurs mesurées du STI sont sensiblement plus élevées que ne l'indiquerait l'intelligibilité perçue.

NOTE Cette question fait l'objet de recherches en cours (voir, par exemple, [8]).

En présence d'échos audibles, il convient d'utiliser d'autres méthodes acoustiques de diagnostic pour mesurer et évaluer la gravité de l'écho.

4.5.10 Compression et expansion rapides de l'amplitude

Les valeurs mesurées de STI et de STIPA peuvent être altérées chaque fois qu'une compression ou une expansion est appliquée au signal d'essai. Cependant, l'expérience montre qu'avec la compression ou l'expansion appropriée il n'y a que des changements mineurs de l'intelligibilité perçue. Il est également noté que les plans de compression altèrent généralement l'équilibre tonal perçu de la parole qui, à son tour, peut avoir un effet défavorable sur l'intelligibilité perçue de la parole.

Lorsqu'ils sont correctement mis en application, les compresseurs-extenseurs (dispositifs complémentaires de compression et d'expansion) n'ont vraisemblablement aucun effet global sur l'intelligibilité.

La compression rapide réagit sur les enveloppes instantanées d'amplitude d'une gamme de bandes de fréquences. Avec cette compression, les variations de niveau du signal au-dessus du niveau de seuil de compression (le coude) sont réduites du taux de compression. La compression réduisant la plage dynamique du signal, la profondeur de modulation peut aussi être réduite.

D'autre part, la commande automatique de gain (AGC, Automatic Gain Control) a un temps de réaction rapide mais un temps de récupération très long et ne réduit pas la plage dynamique à court terme.

Les techniques de compression et de commande automatique de gain sont souvent appliquées pour améliorer l'intelligibilité de la parole (par exemple pour les malentendants souffrant d'une plage dynamique limitée) et peuvent aussi être appliquées aux systèmes de sonorisation.

L'intelligibilité de phrases telle que mesurée par le seuil de réception de la parole (SRT, Speech Reception Threshold) a été améliorée de jusqu'à 4 dB en équivalent de SNR effectif mais cela dépend de la valeur et du type de compression.

L'effet de la compression sur l'intelligibilité à des niveaux élevés de signal et de bruit, tel que dans les systèmes de sonorisation, attend les résultats de futurs travaux de recherche.

4.5.11 Distorsion non linéaire

Bien que le STI soit sensible à la distorsion, le résultat dépend fortement de la méthode de mesure adoptée. (Cela fait l'objet d'un débat plus approfondi en 6.3).

4.5.12 Bruit impulsif et fluctuant

Dans les mesures du STI, il convient de distinguer deux types de bruit de fond:

- impulsif;
- fluctuant.

Le bruit impulsif et les événements brefs non désirés, tels qu'une chute de marteau, conduisent à des résultats inexacts du STI, notamment dans le cas de la transmission à bande étroite, ainsi qu'au diagnostic incorrect de la contribution des bandes de fréquences.

Des bruits fluctuants, tels que des brouhahas de voix ou des machines mises sous tension/hors tension de façon répétitive ou des machines à fonctionnement cyclique, peuvent, lors de mesures répétées, faire varier l'indice STI et conduire aussi à une sous-estimation ou à une surestimation de l'intelligibilité.

De façon subjective, l'intelligibilité de phrases dans un bruit fluctuant est réputée être plus élevée que dans un bruit stationnaire de même valeur quadratique moyennée sur le temps [14].

Si les mesures du STI sont réalisées en présence de bruit impulsif ou fluctuant, il convient d'utiliser la méthode indirecte (décrite à l'Article 6). Il convient d'utiliser un moyennage du signal avec des séquences de longueur maximale (MLS, Maximum Length Sequences) ou des balayages sinusoïdaux lents pour dans cette mesure réduire le bruit. Les effets de la dégradation par le bruit peuvent alors être ajoutés à la fonction de transfert de modulation MTF par post-traitement des données de la MTF "exemptes de bruit".

L'utilisation de balayages sinusoïdaux pour déterminer le STI impose une mesure sans bruit. Pour des raisons pratiques, une mesure sans bruit est obtenue si le *SNR* dans chaque bande d'une octave est d'au moins 20 dB. 7.8.3 donne des informations complémentaires.

4.5.13 Auditeurs malentendants

Sans corrections spécifiques, la méthode STI n'est pas une méthode de prédiction fiable de l'intelligibilité de la parole par des auditeurs malentendants [15]. La mesure des systèmes ou canaux d'aide à l'audition est possible, même si des corrections spécifiques peuvent s'avérer indispensables [16].

4.6 Conclusion

En général, la méthode STI est une approche "conservative" et peut sous-estimer l'intelligibilité dans certaines applications, mais il existe des exceptions, telles que celles données en 4.5.3.

5 Méthode directe de mesure du STI

5.1 Vue d'ensemble

Le STI peut être mesuré, soit directement en utilisant un signal convenablement modulé, soit indirectement au moyen d'une manipulation mathématique d'une réponse impulsionnelle du système, en utilisant une relation proposée par Schroeder [17].

Les travaux de recherche décrits dans [4], [5], [18], [19] et [20] ont développé la base et la méthode pour le FULL STI. Des travaux de recherche ultérieurs ont donné les deux formes simplifiées actuelles, à savoir STIPA et STITEL, qui exigent un moindre temps de mesure. La méthode RASTI a aussi été mise au point, mais elle est désormais désuète.

FULL STI – consiste en 98 signaux d'essai distincts, utilisant 14 fréquences de modulation différentes pour sept bandes d'une octave. Chaque signal d'essai contient une seule fréquence de modulation pour une seule porteuse de bruit de bande d'une octave; les autres bandes d'une octave ne contiennent pas de signal. Les signaux d'essai sont générés séquentiellement. Avec une moyenne de 10 s par signal d'essai, une mesure de FULL STI prend environ 15 min. Une autre version du signal de Full STI contient des modulations aléatoires dans les autres bandes d'une octave en plus de la fréquence de modulation et de la bande d'une octave soumise à essai.

STIPA – consiste en un seul signal d'essai avec un ensemble prédéfini de deux modulations dans chacune des sept bandes d'une octave. Les 14 modulations sont générées simultanément. Une mesure prend entre 10 s et 15 s.

STITEL – consiste en un seul signal d'essai avec un ensemble prédéfini de sept fréquences de modulation, une par bande d'une octave, qui sont générées simultanément. Une mesure prend approximativement 12 s. Le STITEL peut être utilisé pour sa meilleure sensibilité (voir Annexe C), mais il convient de prendre beaucoup de précautions lors de son utilisation.

RASTI – consiste en un seul signal d'essai avec un ensemble prédéfini de neuf fréquences de modulation, qui sont générées simultanément, cinq pour la bande d'une octave de 2 000 Hz et quatre pour la bande d'une octave de 500 Hz. Une mesure prend approximativement 30 s. La

méthode RASTI est désormais désuète, mais, pour une information exhaustive, des détails sont donnés dans l'Annexe D.

Le Tableau 2 compare l'exactitude des deux signaux d'essai simplifiés à celle du FULL STI, pour diverses conditions d'essai.

Pour que le STI tienne compte des rapports signal sur bruit opérationnels et du niveau absolu de la parole, il convient que l'intensité moyenne du signal d'essai soit équivalente au niveau normal de la parole, à la position d'essai. En appliquant la méthode décrite dans l'Annexe J, le $L_{A,eq}$ (*A-weighted equivalent sound pressure level*, niveau de pression sonore équivalent pondéré A) du signal d'essai est réajusté pour être de 3 dB A supérieur au $L_{A,eq}$ typique de la parole continue mesurée à la position d'essai (c'est-à-dire qu'il est nécessaire d'ajouter un facteur de correction de 3 dB).

5.2 STIPA

Le signal d'essai STI peut être simplifié, si sont omises [21] ces modulations non corrélées (ou de type parole) requises pour l'interprétation précise des distorsions non linéaires. Ceci permet la modulation simultanée et le traitement parallèle de toutes les bandes de fréquences, réduisant de ce fait le temps de mesure, mais réduisant aussi sa capacité à prendre en compte certaines formes de distorsion non linéaire, comme indiqué au Tableau 1. Pour chaque bande de fréquence d'une octave, la fonction de transfert de modulation est déterminée pour deux fréquences de modulation.

La méthode STIPA, décrite en Annexe B, utilise cette simplification et a un temps de mesure compris entre 10 s et 15 s. La méthode STIPA est adaptée à la mesure en parole naturelle (transmission acoustique des salles), ainsi qu'aux systèmes électroacoustiques.

La dénomination STIPA renvoie spécifiquement à un signal vocal modulé (comme décrit en Annexe B). Si le STIPA est dérivé d'une réponse impulsionnelle, par exemple, par prédiction, cela doit être clairement énoncé et pour éviter toute confusion, la dénomination STIPA(IR) doit être utilisée. Il convient de noter que le signal STIPA normalisé est fondé sur le spectre de la parole masculine.

Sans corrections spécifiques, la méthode STIPA n'est pas une méthode de prédiction fiable de l'intelligibilité de la parole pour les auditeurs malentendants [15]. La mesure des systèmes ou canaux d'aide à l'audition est possible, même si des corrections spécifiques peuvent s'avérer indispensables [16].

5.3 Application

La méthode STI directe peut être appliquée à pratiquement n'importe quel canal de transmission de la parole, qu'il soit numérique, analogique, électroacoustique ou acoustique. Avec la valeur de STI déterminée, l'intelligibilité des différents types de matériau vocal peut être prédite pour un grand nombre de types de systèmes de transmission.

Pour tous les essais faisant référence à la présente Norme, il convient d'énoncer les paramètres et résultats correspondants dans une fiche de rapport de mesure. Un exemple de fiche de rapport est donné à l'Annexe K.

5.4 Limitations

En plus des limitations de la méthode STI décrites à l'Article 4, la méthode directe de mesure du STI a un certain nombre d'autres limitations.

Comme le signal d'essai est un bruit aléatoire ou pseudo-aléatoire de bande limitée, des mesures successives ne donnent normalement pas des résultats identiques, même en présence de perturbations constantes. Les résultats sont centrés sur une moyenne avec un certain écart-type. Ce dernier dépend, entre autres facteurs, du nombre de mesures discrètes

de la fonction de transfert de modulation (généralement 98 pour la méthode STI ou 14 pour la méthode STIPA) et du temps de mesure employé.

Typiquement, pour la FULL STI, l'écart-type maximal est d'environ 0,02 STI pour un temps de mesure de 10 s, pour chaque indice de modulation $m(f_m)$, et avec des perturbations stationnaires. Avec le STIPA et un temps de mesure de 15 s, l'écart maximal est approximativement de 0,03 STI, pour des mesures répétées.

Avec un bruit fluctuant (par exemple, un brouhaha de voix), des écarts-types supérieurs peuvent être constatés, et éventuellement être accompagnés d'une erreur systématique (biais). Cela peut être vérifié en effectuant une mesure en l'absence du signal d'essai, dont il convient que le résultat ne soit pas une valeur STI résiduelle inférieure à 0,20. Il convient de réaliser une estimation de l'écart-type en répétant les mesures, au moins pour un nombre limité de conditions.

Il est donc de bonne pratique de moyenner les résultats du STI sur deux ou trois mesures, pour une condition spécifique.

6 Méthode indirecte de mesure du STI utilisant la réponse impulsionnelle

6.1 Vue d'ensemble

La fonction de transfert de modulation MTF, en tant que base du STI, peut aussi être calculée à partir de la réponse impulsionnelle d'un canal de transmission, en utilisant le processus appelé "méthode de Schroeder" [14]. La réponse impulsionnelle est acquise (généralement par un matériel informatique), la MTF est déduite et le STI est ensuite calculé à partir de celle-ci.

Il convient d'utiliser l'équation suivante (dont le premier facteur est l'équation de Schroeder), pour calculer la fonction de transfert de modulation $m_{f,k}$, à la fréquence de modulation f_m , dans la k^e bande d'une octave:

$$m_k(f_m) = \frac{\left| \int_0^{\infty} h_k(t) e^{-j2\pi f_m t} dt \right|}{\int_0^{\infty} h_k(t)^2 dt} \cdot \left[1 + 10^{-SNR_k / 10} \right]^{-1}$$

où

- $h_k(t)$ est la réponse impulsionnelle de la k^e bande d'une octave;
- f_m est la fréquence de modulation;
- SNR_k est le rapport signal sur bruit en dB.

La méthode indirecte s'applique uniquement aux systèmes linéaires, invariants dans le temps.

Une expérience considérable est requise pour utiliser cette méthode, car les systèmes de mesure permettent d'ajuster beaucoup de paramètres divers, ce qui peut affecter le résultat final.

Cette méthode est également applicable aux formes simplifiées de la méthode STI. Comme pour cette technique le temps de traitement est très court, il est recommandé de calculer le FULL STI. Néanmoins, le calcul des méthodes plus rapides dérivées du STI peut être utile.

Les valeurs de STIPA dérivées des mesures de réponse impulsionnelle doivent être appelées STIPA(IR).

6.2 Application

Quand on tire l'indice STI de mesures de la réponse impulsionnelle, il est habituel de réaliser une mesure "exempte de bruit" puis de la corriger des effets du bruit de fond et du niveau de la parole. Cependant, des techniques disponibles permettent de tenir directement compte des effets du bruit de fond dans la mesure, par exemple, en utilisant un signal fait de séquences de longueur maximale (*maximum length sequences*) MLS de type vocal sans moyennage. Les modes opératoires de mesure pour déterminer la réponse impulsionnelle doivent satisfaire aux exigences suivantes, parmi d'autres, avec des informations complémentaires fournies par l'ISO 18233:

- a) Les mesures de la réponse impulsionnelle doivent être réalisées conformément à l'ISO 18233.
- b) La longueur de la réponse impulsionnelle acquise doit être d'au moins 1,6 s et ne doit pas être inférieure à la moitié du temps de réverbération de la salle.
- c) Afin de produire une réponse impulsionnelle "exempte de bruit", il convient d'obtenir un *SNR* de 20 dB au moins, dans toutes les sept bandes d'une octave. S'il y a lieu, le moyennage de signal peut être utilisé pour obtenir ce résultat;
- d) Il convient d'éviter d'utiliser des signaux d'excitation avec un spectre de fréquence blanc (par exemple, tel qu'utilisé dans la spectrométrie temporisée (TDS, Time Delay Spectrometry) ou dans les séquences de longueur maximale (MLS)) dans les conditions normales, à moins que le niveau du bruit de fond soit très bas. Un spectre de fréquences rose (-3 dB/octave) produit avec un bruit rose ou un balayage sinusoïdal logarithmique (plus rigoureusement "un balayage exponentiel") est généralement mieux adapté. Toutefois, un signal MLS vocal peut être utilisé également sans moyennage pour mesurer directement l'effet du bruit de fond sur le STI.
- e) Les signaux impulsionnels, tels que la fonction de Dirac, ne sont généralement pas adaptés, lorsque le bruit de fond, la limitation de la bande passante ou la distorsion non linéaire sont importants, car le spectre moyen des fréquences et la répartition des niveaux pour une parole typique ne sont pas représentés dans le signal d'essai.
- f) La méthode de la réponse impulsionnelle s'applique uniquement aux systèmes linéaires constants dans le temps. Si le canal de transmission a des fonctions avec un traitement non linéaire du signal, il convient que ces fonctions soient contournées au cours de la mesure de l'intelligibilité de la parole. Si, par exemple, le niveau effectif de pression sonore de la restitution est augmenté par une réduction non linéaire de la dynamique du signal, cela doit être pris en compte en mesurant séparément le niveau maximal de pression sonore et en appliquant une correction appropriée.
- g) Les variations de temps dues aux mouvements de l'air (vent) ou aux changements climatiques au cours du processus de mesure doivent être évitées (elles invalident également le moyennage sur des durées plus longues). Il convient que la vitesse moyenne du vent au cours des mesures de MLS, par exemple, ne dépasse pas 4 m/s. Les mesures utilisant les séquences de longueur maximale (MLS) sont plus vulnérables à cet égard que les mesures réalisées avec des balayages sinusoïdaux.
- h) Il convient de s'assurer que les composants impliqués dans la transmission du son (haut-parleurs, surfaces des salles, réflecteurs, microphone de mesure, personnes) ne bougeront pas au cours du cycle de mesure.
- i) Dans des conditions critiques, la répétabilité des résultats de mesure doit être prouvée par des mesures répétées.
- j) L'impact du bruit de fond (L_n) et du niveau de parole opérationnel (L_s) dans chaque bande d'une octave k doit être incorporé dans le résultat par un post-traitement (voir Annexe M).

6.3 Limitations (distorsion non linéaire)

La méthode de mesure de l'indice STI a des limitations décrites à l'Article 4; la méthode de mesure par la réponse impulsionnelle en a bien d'autres, dont surtout l'effet des distorsions non-linéaires.

Il convient d'éviter les distorsions du signal de mesure, car la méthode indirecte ne prend pas correctement en compte les effets de distorsion. Lors de l'utilisation de cette méthode, la sensibilité à la distorsion dépend fortement du mode opératoire de mesure [13], [22]. Par exemple, les méthodes employant les transformations de Fourier, ne sont exemptes d'erreur que pour les systèmes linéaires.

S'impose une analyse critique de la mesure de la réponse impulsionnelle et de l'influence possible des non-linéarités du canal de transmission, car les composants du système peuvent dans la pratique se trouver fonctionner à la limite de leur plage de bon fonctionnement. Lors de l'utilisation des techniques de balayage sinusoïdal, les composantes de distorsion non linéaire apparaissent au début ou à la fin de la réponse impulsionnelle récupérée et peuvent ainsi être évaluées. Cependant, des erreurs peuvent apparaître si le temps de réverbération est long, car la trainée de réverbération des composantes de distorsion peut polluer la réponse impulsionnelle principale.

Lors de l'utilisation d'un signal MLS (suites de longueur maximale), les composantes de distorsion tendent à apparaître comme du bruit et ne sont donc pas directement discernables. Les composantes continues (C.C) et les artéfacts de repliement temporels apparaissent comme des précurseurs (pré-échos) avant l'arrivée du signal.

Lors de l'utilisation d'une technique de balayage sinusoïdal, les composantes de la distorsion, inhérentes à la méthode, doivent être supprimées ou éliminées de l'IR (Impulse Response, réponse impulsionnelle), avant que le calcul du STI puisse être entrepris.

7 Modes opératoires de mesure, post-traitement des données et applications

7.1 Généralités

Bien que les mesures de STI soient normalement réalisées par des moyens acoustiques, il n'est dans certaines situations, pas toujours possible ou nécessaire, d'utiliser l'excitation acoustique ou de réaliser des mesures acoustiques. Par exemple, dans les situations où les différents systèmes sont évalués selon leur qualité de transmission de la parole ou si de plus amples informations de diagnostic sont nécessaires, le signal d'essai peut être injecté et/ou reçu électriquement.

Dans tout post-traitement de la matrice MTF (modulation transfer function), il est essentiel d'utiliser un niveau réaliste de bruit de fond [8]. Si la sortie du canal de transmission est acoustique, le seuil d'audition en réception doit être utilisé comme un minimum.

Il convient d'énoncer les paramètres pertinents dans un rapport de mesure. Un exemple (extrait) de rapport est donné à l'Annexe K.

7.2 Entrée acoustique

Le fait d'appliquer le signal d'essai, via un haut-parleur spécial (voir ci-dessous), au microphone du système en essai assure la prise en compte de ces facteurs qui à l'emplacement du microphone sont susceptibles de réduire l'intelligibilité, par exemple le bruit ambiant ou la rétroaction acoustique (feedback). En outre, certains systèmes électroacoustiques n'ont pas d'autre moyen d'injecter le signal d'essai. Comme ce mode opératoire exige que le signal d'essai soit reproduit acoustiquement, il est nécessaire d'utiliser un haut-parleur spécifique (par exemple une bouche artificielle) qui émule un locuteur naturel.

Un ajustement correct du spectre du signal d'essai pour l'accorder au spectre normalisé de la parole est également requis pour l'injection électrique du signal d'essai. Lors de l'utilisation de la méthode directe, le signal d'essai normalisé doit être utilisé à cet effet.

Le mode opératoire suivant doit être utilisé:

- a) Vérifier l'intégrité du signal d'essai (par exemple par une mesure en boucle fermée (*loop back*)). Ceci est particulièrement important si le signal d'essai est généré par un lecteur de CD, quoique les générateurs de modulation par impulsion et codage MIC (PCM) (par exemple les fichiers .wav) soient aussi à vérifier. (Normalement, il convient de ne pas utiliser les formats à compression du signal par exemple le MP3, bien que des systèmes de compression utilisant au moins 128 kbit/s se soient avérés fonctionner sans erreur apparente). Des informations supplémentaires sont disponibles dans la référence [9].
- b) Vérifier que la réponse en fréquence par 1/3 d'octave de la source de signal d'essai (bouche artificielle ou haut-parleur d'essai adéquat) est, lorsqu'elle est mesurée en champ libre (exempt de réflexions), à ± 1 dB sur la gamme de fréquences requise, pour l'une ou l'autre des techniques de mesure suivantes:
- sur la gamme de 88 Hz à 11,3 kHz en utilisant la méthode FULL STI ou la MLS (maximum length sequence) ou tout autre signal de mesure de la réponse impulsionnelle (88 Hz et 11,3 kHz sont les limites des bandes d'une octave centrées sur 125 Hz et 8 kHz) ou
 - les niveaux individuels des bande d'une octave centrées de 125 Hz à 8 kHz lors de l'utilisation d'un STIPA ou tout autre signal d'essai vocal.

NOTE 1 Pour les mesures indirectes, la réponse en fréquence dérivée d'une MLS ou d'une autre mesure de la réponse impulsionnelle peut être traitée pour calculer le spectre par bandes d'une octave.

Si cela s'avère nécessaire, régler l'égalisation (si il y en a une) de la bouche artificielle ou du haut-parleur d'essai pour satisfaire à cette exigence.

- c) Placer la source sur l'axe du microphone approprié, à l'emplacement approprié ou à la distance adéquate du locuteur et la diriger dans la direction normale de locution.

En l'absence de bouche artificielle, il est permis d'utiliser un transducteur approprié, tel qu'une petite enceinte mono-source équipée d'un haut-parleur unique de haute qualité (diamètre du cône ne dépassant pas 100 mm) et ce transducteur doit être décrit avec les résultats.

NOTE 2 Généralement, dans un espace d'écoute, l'intelligibilité de la parole dépend de la directivité de la source; par conséquent, il convient lors de l'évaluation de l'intelligibilité de locuteurs sans amplification d'utiliser une bouche artificielle présentant des caractéristiques de directivité similaires à celles de la tête/bouche humaine (voir Recommandation UIT-T P.51 [48]). Cependant, les caractéristiques directionnelles de la source d'essai (haut-parleur simulant le locuteur ou bouche artificielle) peuvent être d'une importance considérable lors de mesures dans des espaces vastes ou réverbérants ou lorsque le microphone capteur est placé à bonne distance du locuteur [16], [23].

Lorsque la parole est relayée par un système électroacoustique, un simulateur peut ne pas être indispensable. Par contre, lorsque le microphone source est placé dans un emplacement réverbérant ou bruyant ou si un microphone de proximité ou un microphone antibruit est impliqué, il convient d'utiliser soit un simulateur de locuteur, soit une bouche artificielle.

- d) Régler le niveau du signal d'essai à l'emplacement du microphone au niveau opérationnel de parole qui sera utilisé dans le système. Les niveaux de parole et de signal d'essai doivent être accordés entre eux conformément à la méthode décrite dans l'Annexe J.

NOTE 3 Cet essai est susceptible de mettre l'amplificateur sous contrainte. Voir le 14.9 de la CEI 60268-3 [24]. Il peut être commode d'appliquer le signal d'essai pendant par exemple, 1 min, suivie de plusieurs minutes de signal zéro pour permettre le refroidissement.

En l'absence d'un accord correct entre le niveau du signal d'essai et le niveau opérationnel de parole, il convient de donner par défaut, à la source, un niveau équivalent, de 60 dB A, à 1 m en face de la bouche artificielle ou du haut-parleur d'essai.

Des distances au locuteur plus courtes donnent typiquement des niveaux de parole d'environ 86 dB A à 94 dB A pour des microphones tenus à la main (distances de 5 cm à 2 cm), alors que des niveaux de parole d'environ 80 dB A à 86 dB A sont obtenus pour des microphones à col de cygne (distances de 10 cm à 5 cm).

NOTE 4 Les niveaux donnés ci-dessus sont dans la pratique sujets à de grandes variations.

- e) Lancer la suite d'essai STI, STIPA (ou STITEL). Normalement, et quand elle est disponible, il convient de choisir l'option "avec bruit".
- f) Si un signal MLS (suites de longueur maximale) est utilisé pour mesurer la réponse impulsionnelle et s'il faut tenir compte du bruit de fond, il convient, par un filtrage

approprié, d'ajuster le spectre d'excitation au spectre de parole normalisé. Il convient de désactiver le moyennage de signal ou d'utiliser une seule séquence [25].

- g) Si des balayages sinusoïdaux, la MLS ou la TDS sont utilisés pour déterminer la réponse impulsionnelle sans bruit, des ajustements appropriés des niveaux de parole et de bruit aux emplacements du microphone et du récepteur doivent par post-traitement être appliqués à la MTF (fonction de transfert de modulation) sans bruit.
- h) Le signal d'essai doit alimenter le système de façon que tous les traitements de signal (égaliseurs, lignes à retards, etc.) qui jouent sur la qualité de la reproduction de la parole soient bien pris en compte lors de la mesure.

7.3 Sortie acoustique

Le dispositif de mesure (microphone/oreille artificielle/simulateur de la tête) doit être étalonné acoustiquement, en sensibilité et en réponse en fréquence. Les mesures doivent être réalisées à l'emplacement normal de l'auditeur et à la hauteur normale d'écoute (ou à une hauteur d'écoute spécifiée). Si un seul microphone est utilisé, il doit être omnidirectionnel et du type pour champ diffus.

7.4 Entrée électrique

Suivre le mode opératoire du 7.2 ci-dessus, en remplaçant l'étape d) par l'étape ci-dessous et en choisissant le point d'injection du signal pour qu'il soit le plus près possible de l'entrée normale du signal, afin de mettre dans l'essai le plus possible du système..

Le signal d'essai STI, au point d'injection, doit être ajusté pour qu'il soit équivalent au niveau de parole à cet endroit. Ce niveau de parole est déterminé par la méthode de mesure de niveau de parole décrite dans l'Annexe J.

7.5 Sortie électrique

Comme aucune condition acoustique n'est impliquée à la sortie électrique, les effets liés à l'audition, tels que le masquage et le seuil de réception, doivent sur le dispositif de mesure être rendus inopérants. Si ce n'est pas possible, l'entrée électrique du dispositif de mesure doit être ajustée pour simuler un niveau de pression sonore situé bien au-dessus du seuil de réception, mais en dessous d'un point où le masquage dépendant du niveau devient décelable dans les résultats du STI. Il convient d'appliquer une pondération A aux niveaux de sortie large bande, puis de les consigner comme niveaux de tension avec pondération A, en dB par rapport à une référence qui sera notée, par exemple, 1 V.

7.6 Exemples de combinaisons entrée/sortie

7.6.1 Entrée acoustique – Sortie acoustique

Le montage de mesure normal du STI, pour les systèmes de sonorisation PA (*public address*) et dans les auditoriums, comporte une source sonore pour générer acoustiquement le signal d'essai STI. Le niveau du signal d'essai est étalonné et correspond au niveau de parole nominal. Il convient d'utiliser une distance de locution représentative, selon la situation, comme décrit en 7.2. Un dispositif de mesure de STI étalonné est utilisé à l'emplacement du récepteur, pour déterminer le STI du canal de transmission.

7.6.2 Entrée électrique – Sortie électrique (par exemple, évaluation de systèmes de communication filaire et sans fil)

Des mesures purement électriques du STI sont en général réalisées pour évaluer les différents systèmes de communication sous l'angle de leur qualité de transmission de la parole, plutôt que pour obtenir une valeur absolue de l'intelligibilité de la parole. Il est conseillé de réaliser ces mesures à différents niveaux de signal d'entrée (par exemple, de -10 dB à +10 dB par rapport au niveau opérationnel de référence) pour recueillir des informations relatives à l'influence sur l'intelligibilité de la parole de la plage dynamique, du plancher de bruit et des traitements de signal. Ces types de mesures sont susceptibles d'être

faits sur des systèmes de transmission de la parole, filaires ou sans fil, tels que des lignes téléphoniques et des systèmes de radiocommunications.

7.6.3 Entrée acoustique – Sortie électrique (par exemple, évaluation des microphones)

Pour comparer les microphones par leur effet sur l'intelligibilité, il convient d'étalonner le niveau du signal d'essai STI au microphone, comme indiqué en 7.2. Les mesures sont réalisées en combinaison avec le spectre de bruit ambiant approprié et en fonction du niveau de bruit, pour déterminer le comportement du microphone en rejection de bruit. Il convient, de préférence, d'effectuer les mesures à différents niveaux de parole, pour étudier l'effet d'une baisse ou d'une hausse de la voix sur l'intelligibilité.

NOTE Des méthodes spéciales peuvent être requises pour mesurer le STI de systèmes d'aide à l'audition et des systèmes de boucle d'induction de fréquence audio (AFILS, Audio Frequency Induction Loop Systems), en particulier [16]. Cependant, une grande partie des recommandations générales données dans cet article s'applique.

7.6.4 Entrée électrique – Sortie acoustique (par exemple, évaluation des systèmes PA)

Pour comparer différents transducteurs (haut-parleurs, casques d'écoute), le signal d'essai STI peut être injecté électriquement. Le signal d'essai doit être reproduit, à l'emplacement des auditeurs, à un niveau de pression sonore représentatif du fonctionnement normal.

Dans le cas d'un système de sonorisation ou système de distribution du son de ce genre, il convient de réaliser les mesures pour un nombre représentatif d'emplacements. Prendre une simple valeur moyenne des résultats peut être trompeur. Une méthode meilleure, qui prend en compte la variation de la mesure selon les emplacements, est de calculer la moyenne des mesures moins un écart type. Cette quantité est parfois appelée "qualité de l'espace": un emplacement aura, sous l'hypothèse d'une répartition gaussienne, une probabilité – ou un "niveau de confiance"- de 84% d'avoir au moins cette qualité. Une méthode plus précise est de tracer la distribution statistique complète des résultats.

Lors de l'évaluation des casques d'écoute, il convient d'utiliser un microphone intra-auriculaire (MIRE) ou une oreille artificielle.

7.7 Post-traitement des données de MTF (fonction de transfert de modulation) mesurées

Un certain nombre de corrections peuvent être apportées aux valeurs mesurées de la MTF:

- élimination du bruit (débruitage) d'une MTF mesurée;
- ajout d'un niveau et d'un spectre de bruit d'occupation;
- prise en considération du seuil de réception en audition;
- ajustement du niveau de parole et du spectre;
- correction pour différents temps de réverbération.

L'effet du bruit d'occupation peut être déterminé soit

- a) en entrant manuellement des données relatives au bruit dans le tableau des données de bruit utilisé par l'équipement de mesure, soit
- b) en mélangeant un signal de bruit artificiel ou enregistré, de contenu et de niveau spectral adéquat, avec l'entrée du signal directe dans l'analyseur ou avec un signal enregistré.

L'Annexe M donne un exemple d'élimination de bruit d'une matrice MTF mesurée et d'ajout de bruit de fond opérationnel et de niveaux de parole utile. Les équations énumérées dans l'Annexe A sont utilisées pour ce processus.

7.8 Questions concernant le bruit

7.8.1 Généralités

Comme dans le cas de tous les systèmes linéaires, l'influence des distorsions, telles que la réverbération, est indépendante de la réponse en amplitude. En conséquence, les variables qui dépendent du niveau du signal sont le rapport signal sur bruit dans chaque bande d'une octave et le masquage croissant associé. Par conséquent, la méthode STI peut être relativement insensible aux variations de la réponse amplitude/fréquence du canal de transmission, surtout lorsque le bruit de fond est faible.

Lorsque des bruits de fond de faibles niveaux sont ajoutés à la matrice MTF, bruits représentatifs des niveaux de bruit rencontrés dans la pratique lors de l'utilisation d'un système électroacoustique, le STI global montre plus de sensibilité aux variations du spectre d'entrée.

Une situation essentiellement sans bruit, dans laquelle seul le seuil d'audition auditive agit comme source de bruit résiduel, n'est, dans la plupart des cas pratiques, en général pas une hypothèse réaliste. Même dans des environnements silencieux, tels que les bibliothèques ou les salles de tribunaux, un niveau de bruit résiduel de 25 dB SPL à 35 dB SPL n'est pas rare et il convient d'en tenir compte. Cela peut se faire en appliquant un critère approprié, tel que les courbes NCB, RC ou NR (voir [26]).

Des brefs événements non désirés (par exemple, du bruit impulsif) peuvent être détectés par une analyse des statistiques du signal. Cependant, il est en pratique facile de répéter la mesure du STI avec la source de bruit physiquement éliminée ou d'utiliser la méthode indirecte et des techniques de moyennage.

Le bruit fluctuant est détecté en mesurant le STI direct en l'absence du signal d'essai. Si le STI est trop élevé (par exemple, $STI > 0,2$), les résultats des mesures seront probablement erronés. La méthode recommandée est de mesurer le STI en l'absence du bruit, mesurer ensuite séparément le bruit (voir 7.8.2) et calculer mathématiquement le STI.

7.8.2 Mesure du bruit de fond

Pour corriger une mesure de STI quant au bruit de fond, il est nécessaire de caractériser exactement ce bruit de fond. Il est recommandé de mesurer le niveau de pression sonore équivalent continu (L_{eq}) du bruit de fond dans chacune des sept bandes d'une octave (125 Hz à 8 kHz), sur une durée suffisante. Les emplacements, durées et heures de mesure doivent être consignés en même temps que les remarques sur des circonstances inhabituelles qui pourraient affecter la validité des mesures.

Pour ce calcul correctif ici décrit, il ne suffit pas de déterminer une seule valeur large bande pour le bruit de fond (par exemple $L_{A,eq}$) et d'utiliser une seule valeur large bande pour le signal de la parole (c'est-à-dire le niveau de parole opérationnel). Voir aussi 7.8.3.

7.8.3 Bruit fluctuant

Il est recommandé, si le bruit fluctuant ne peut être éliminé, de réduire au minimum son influence en amplifiant le signal au moins à 15 dB au-dessus du niveau de bruit dans chacune des bandes d'une octave. Calculer le STI, avec les indices de modulation, pour les niveaux originaux avant amplification. Cette méthode exige certaines compétences en calcul.

Si l'influence du bruit fluctuant ne peut être réduite, il convient de répéter les mesures au moins trois fois avant de prendre la moyenne des STI. Si la dispersion de ces trois mesures est moindre que 0,03 STI, d'autres mesures ne sont pas nécessaires.

L'interprétation de l'intelligibilité de la parole en présence de bruit fluctuant est extrêmement difficile et ne relève pas actuellement du domaine d'application de la présente Norme.

Toutefois, il s'est avéré que les auditeurs écoutent la parole dans les trous entre le bruit fluctuant et perçoivent une intelligibilité plus élevée que ne le prédirait le STI basé simplement sur le L_{eq} du bruit fluctuant.

NOTE Si la fluctuation est forte (par exemple 15 dB ou plus), il peut être nécessaire d'utiliser le L_{10} dans chaque bande d'une octave.

7.9 Analyse et interprétation des résultats

Il est important d'examiner les données MTF dans chacune des bandes d'une octave pour déterminer la fiabilité des résultats, comme suit:

- si le rapport de transfert de modulation en fonction de la fréquence de modulation, présente en fonction de la fréquence des valeurs constantes ou en légère diminution, le bruit est le mécanisme dominant;
- si le rapport de transfert de modulation décroît de façon monotone avec la fréquence de modulation, la réverbération est le mécanisme principal;
- si les valeurs baissent puis augmentent avec la fréquence de modulation, il y a de fortes réflexions avec des retards supérieurs à 50 ms, ce qui peut conduire à une évaluation de l'intelligibilité excessivement optimiste. Si ce phénomène est détecté, il sera consigné avec le résultat.

7.10 Mesures de STI binaural

Même si le STI est une méthode bien acceptée et normalisée pour la prédiction de l'intelligibilité, le modèle STI est essentiellement fondé sur l'écoute monaurale. Les avantages de l'écoute binaurale pour l'intelligibilité de la parole sont ignorés par le modèle.

Subjectivement, l'avantage binaural peut être important. Toutefois, il n'existe aucune méthode claire de mesure. La méthode STI actuelle peut produire une sous-estimation de l'intelligibilité, notamment si la parole et le bruit arrivent au récepteur à partir de directions différentes. Cette question fait actuellement l'objet de travaux de recherche.

Lors des mesures binaurales du STI utilisant une tête artificielle, l'approche recommandée est d'utiliser les résultats du STI pour la meilleure oreille. Pour plus d'informations, voir [27].

8 Utilisation du STI comme moyen de prédiction en conception de systèmes acoustiques

8.1 Vue d'ensemble

Pendant la phase de conception d'un système électroacoustique, il est utile de prédire la performance en STI à partir des paramètres acoustiques prédits pour la salle. Deux méthodes sont disponibles:

- un calcul fondé sur une prédiction du champ direct, combiné à une réverbération à décroissance exponentielle dans le temps et des paramètres électroacoustique simples. Des temps de réverbération calculés statistiquement peuvent être employés ici;
- une prédiction fondée sur une réponse impulsionnelle simulée du système dans l'espace acoustique.

Les prédictions fondées sur une réponse impulsionnelle simulée offrent un degré de précision plus élevé. Cette méthode est également préférable dans les cas où les temps de réverbération calculés statistiquement (Sabine/Eyring) sont connus pour être entachés d'erreur, par exemple, dans des espaces couplés, ou dans des espaces avec une répartition non uniforme de l'absorption.

Il est essentiel d'utiliser le niveau de parole opérationnel pour la prédiction du STI, car cela affecte à la fois le SNR réel et les effets du masquage. Pour cette prédiction un signal de

parole large bande doit être employé, et le canal de transmission doit pouvoir produire le niveau opérationnel de pression sonore

8.2 Prédiction statistique

La prédiction de la performance en STI d'un système électroacoustique doit être fondée sur la matrice MTF (fonction de transfert de modulation), qui est calculée à partir des paramètres acoustiques et électroacoustiques prédits de la salle et des niveaux de bruit de fond mesurés ou estimés pour chaque bande d'une octave contribuant à la version choisie du STI. Les calculs doivent utiliser la méthode de Houtgast et al. [28], qui est donnée à l'Annexe L.

Les valeurs de MTI (indice de transfert de modulation) dans chaque bande d'une octave et les niveaux du signal de parole de sortie par bande d'une octave doivent être accessibles.

Si la prédiction est faite en utilisant un logiciel du commerce, les résultats doivent énoncer:

- qu'une estimation statistique a été effectuée en utilisant la méthode de Houtgast et al.; [28];
- que le STI a été calculé en utilisant les pondérations masculines ou féminines appropriées;

Noter que:

- le RASTI ne doit pas être utilisé comme une indication du STI prédit;
- le STI ne doit pas être estimé en convertissant une valeur de %Alcons (Articulation Loss of consonants, perte d'articulation des consonnes);
- la méthode de prédiction statistique est même moins sensible que la méthode STI directe aux effets d'arrivées discrètes fortes en avance ou en retard et à la perte d'intelligibilité possible qui peut venir d'une réponse en fréquence médiocre.

8.3 Prédiction à partir d'une réponse impulsionnelle simulée

La prédiction du STI à partir d'une réponse impulsionnelle simulée doit être effectuée de la manière suivante:

- a) la matrice MTF doit être calculée en utilisant l'équation de Schroeder (voir 6.1). La durée de la réponse impulsionnelle ne doit pas être inférieure à la moitié du temps de réverbération mais doit être d'au moins 1,6 s pour assurer un calcul fiable des indices de modulation pour la modulation de fréquence la plus basse de 0,63 Hz;
- b) tant les seuils de réception en écoute que les niveaux de pression sonore mesurés ou estimés du bruit de fond pour chaque bande d'une octave doivent alors être introduits dans la matrice MTF;
- c) le spectre de la parole et le niveau opérationnel de la parole doivent être choisis, et les corrections de masquage auditif énumérées au Tableau A.1 appliquées à la matrice MTF;
- d) les facteurs de pondération masculins et féminins spécifiques à chaque bande d'une octave donnés au Tableau A.3 doivent être appliqués aux valeurs de MTI.

Pour chaque emplacement objet d'une prévision, les valeurs du MTI et les niveaux de sortie du signal de parole doivent pour chaque bande d'une octave être disponibles ainsi que la réponse en fréquence.

Pour les prévisions avec de multiples emplacements d'écoute (d'auditeurs) les propriétés statistiques et la distribution des résultats sur toute la zone d'écoute doivent être consignées.

Les résultats doivent également énoncer:

- que le STI a été calculé à partir d'une MTF tirée d'une réponse impulsionnelle prédite, avec l'application de la pondération masculine ou féminine appropriée;
- les niveaux de bruit de fond qui ont été appliqués à la prévision.

Annexe A (normative)

Méthode de l'indice de transmission de la parole (STI) et méthode STI révisée

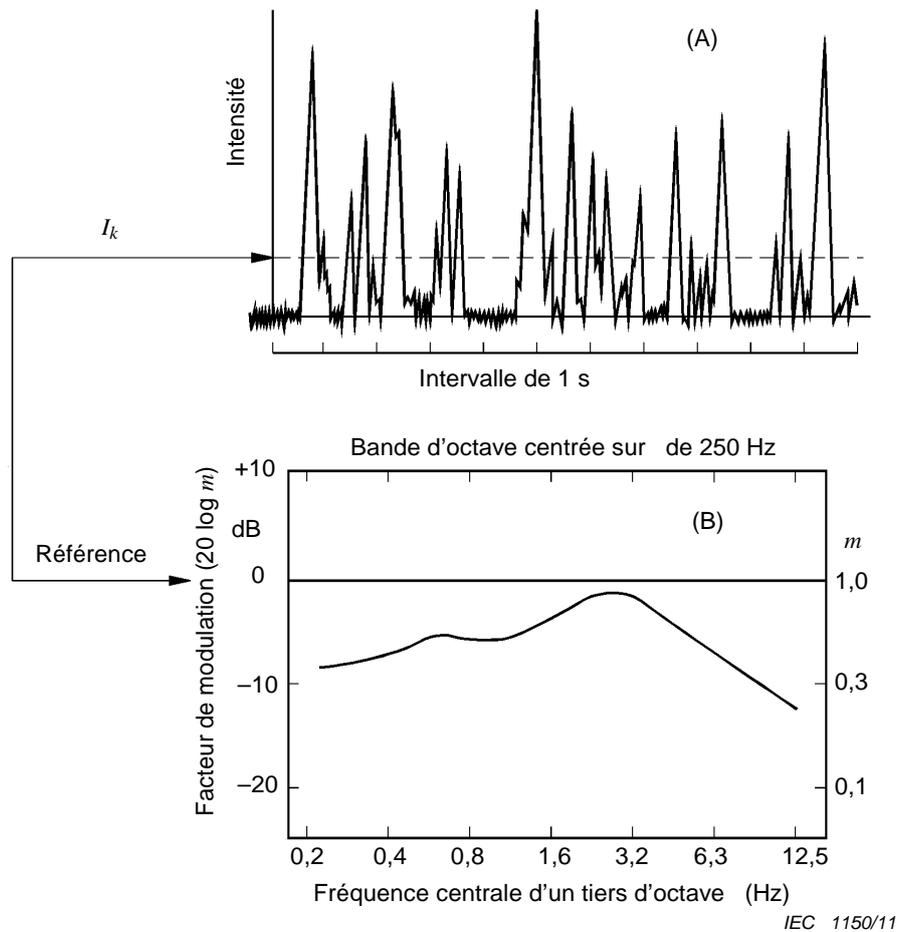
A.1 Contexte

A.1.1 Fonction enveloppe et spectre de l'enveloppe

Un discours continu peut être considéré comme une suite de fragments élémentaires de parole, appelés phonèmes. Chaque phonème se caractérise par un spectre particulier en fréquence. La clarté exige que soient préservées les différences spectrales entre phonèmes. Ces différences spectrales peuvent être caractérisées par la fonction enveloppe prise dans un certain nombre de bandes de fréquences. La fonction enveloppe décrit les fluctuations temporelles de l'intensité d'un signal de parole à l'intérieur d'une certaine bande de fréquences. La forme de la fonction enveloppe est unique pour une séquence spécifique de phonèmes. Une distorsion de l'enveloppe de la parole, par exemple du fait d'un bruit ou d'une réverbération, réduit les différences spectrales entre phonèmes, ce qui se reflète par une diminution du degré des fluctuations de la fonction enveloppe.

La Figure A.1, schéma A, montre une fonction enveloppe exemplaire pour une bande de fréquences d'octave centrée sur 250 Hz. Le spectre de l'enveloppe fournit une description plus générale des fluctuations de la fonction enveloppe et résulte d'une analyse spectrale par bande de tiers d'octave de la fonction enveloppe. Habituellement, un extrait de discours d'une minute est analysé pour donner la distribution spectrale des fluctuations de l'enveloppe autour de l'intensité moyenne. Cela permet la formation de l'indice de modulation en fonction de la fréquence de modulation comme montré à la Figure A.1, schéma B, où le spectre est normalisé par rapport à l'intensité moyenne I_k .

Une comparaison des spectres de l'enveloppe mesurés directement au niveau du locuteur avec les spectres correspondants trouvés au bout d'une voie de transmission indique la perte des fluctuations dues à cette voie de transmission. Cette perte conduit à la fonction de transfert de modulation ou MTF, qui représente la diminution de la profondeur de modulation en fonction de la fréquence de modulation.



IEC 1150/11

Figure A.1 – Fonction enveloppe (schéma A) d'un signal vocal de 10 s pour une bande d'une octave centrée sur 250 Hz et spectre de l'enveloppe correspondant (schéma B)

A.1.2 Fonction de transfert de modulation (MTF)

La raison sous-jacente à l'application du concept de MTF aux études d'acoustique des salles a été décrite ailleurs [5], [6], [19], [20]. La MTF quantifie la diminution des modulations du signal original en fonction de la fréquence de modulation. Les modulations sont définies par l'enveloppe de l'intensité du signal: en effet c'est en intensité que le bruit ou la réverbération n'affectent que la profondeur de modulation d'une modulation sinusoïdale sans en modifier la forme. La Figure A.2 en fournit une illustration pour la bande d'une octave de fréquence médiane 250 Hz dans le cas de deux systèmes de transmission simples, l'un avec seulement une réverbération exponentielle (cas A; $T = 2,5$ s) et l'autre avec seulement un brouillage par un bruit (cas B; rapport signal sur bruit $SNR = 0$ dB).

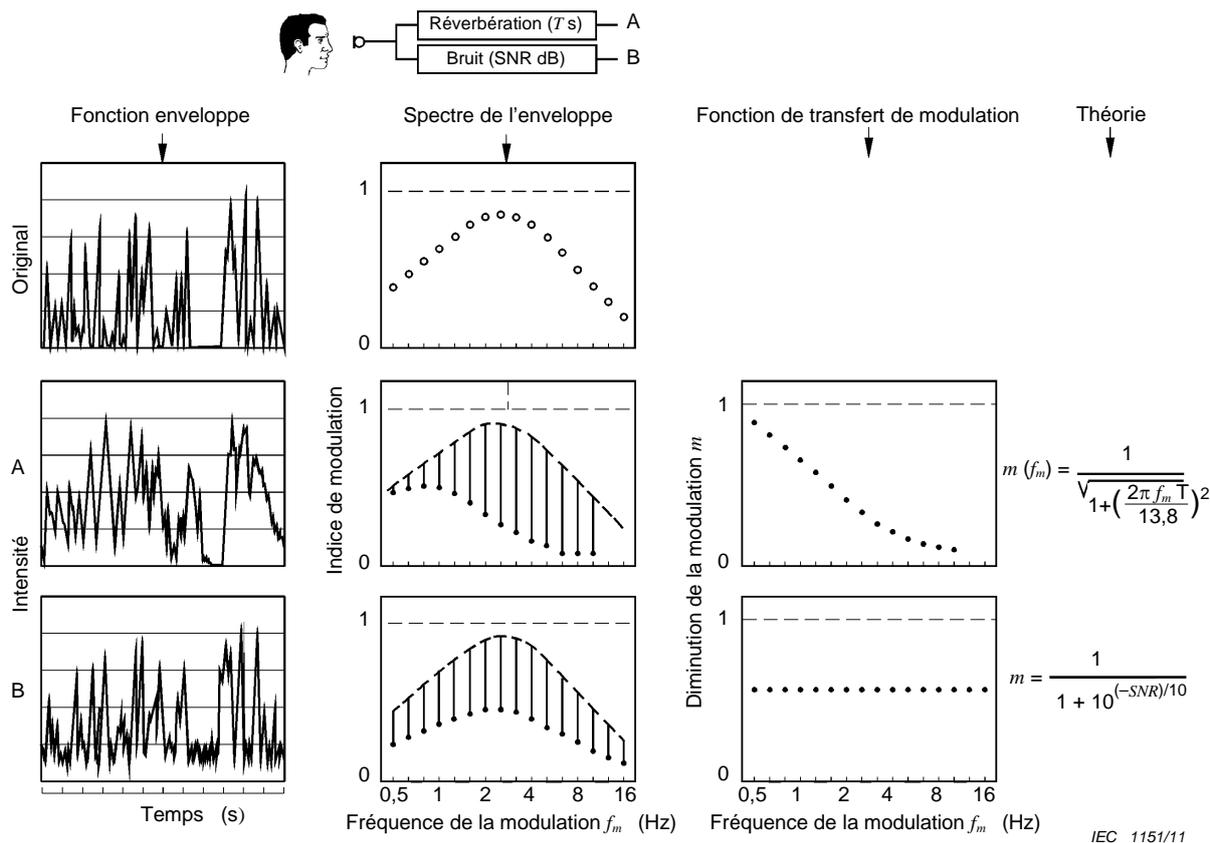


Figure A.2 – Expressions théoriques de la MTF

En présence de réverbération, la MTF a la forme d'un filtre passe-bas: les fluctuations les plus rapides étant relativement plus affectées que les fluctuations plus lentes. Dans le cas théorique d'une réverbération à décroissance purement exponentielle, la MTF peut être mathématiquement explicitée (voir Figure A.2, cas A) et le produit de f_m et de T détermine le roll-off comme suit.

$$m(f_m) = \frac{1}{\sqrt{1 + \left(\frac{2\pi f_m T}{13,8}\right)^2}}$$

où

- f_m est la fréquence de modulation;
- T est le temps de réverbération en secondes.

Dans le cas du bruit, la MTF est définie par le rapport signal sur bruit et elle est indépendante de la fréquence de modulation (voir Figure A.2, cas B). Le bruit, en augmentant l'intensité moyenne, réduit la profondeur de modulation pour toutes les fréquences de modulation selon:

$$m = \frac{1}{1 + 10^{(-SNR/10)}}$$

où

- SNR est le rapport signal sur bruit en dB.

Avec des forts échos (réflexions prononcées), la MTF montre la forme d'un filtre coupe-bande, d'abord décroissant progressivement puis augmentant de nouveau à des fréquences de modulation plus élevées.

A.2 La technique du STI

A.2.1 Généralités

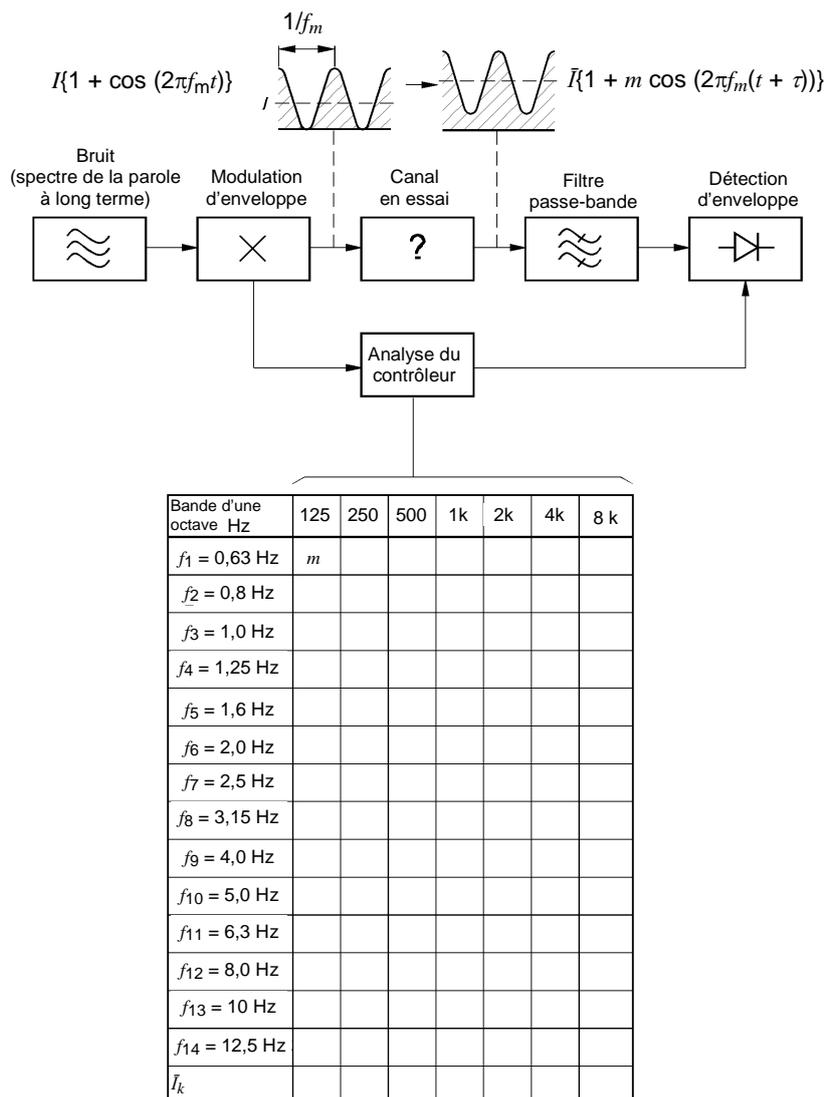
L'indice de transmission de la parole (STI) est une mesure objective, fondée sur la contribution pondérée d'un certain nombre de bandes de fréquences prises dans la gamme de fréquences des signaux vocaux, ces contributions étant déterminées par le rapport signal sur bruit réel ou effectif. Sa description ainsi que les facteurs de pondération de bandes d'une octave et les facteurs de redondance sont donnés en [29]. Pour un choix correct de la forme du signal d'essai, ce rapport signal sur bruit effectif peut inclure et permettre la prise en compte des distorsions dans le domaine temporel des non-linéarités, du bruit de fond, etc.

Les distorsions dans le domaine temporel (réverbération, échos et commande automatique de gain) peuvent dégrader le signal vocal fluctuant et en diminuer l'intelligibilité. Cela est dans le mode opératoire STI, modélisé par la détermination de la fonction de transfert de modulation sur la gamme des fréquences présentes dans l'enveloppe des signaux vocaux naturels. La plage pertinente de ces fréquences de modulation est de 0,63 Hz à 12,5 Hz sur quatorze bandes d'un tiers d'octave. La Figure A.3 illustre un dispositif de mesure où la fonction de transfert de modulation, $m(f_m)$, est déterminée séparément pour chaque fréquence de modulation dans chaque bande d'une octave.

La mesure la plus complète du STI est le FULL STI. La méthode FULL STI directe utilise une seule fréquence de modulation pour une bande d'une octave par signal d'essai, chaque signal d'essai étant d'une durée approximative de 10 s. Pour obtenir une seule valeur de STI, la méthode FULL STI utilise 98 (14×7) signaux d'essai indépendants.

La méthode STI a été mise au point à l'origine en utilisant des signaux FULL STI directs (une fréquence de modulation pour une bande d'une octave) avec des modulations aléatoires pour les bandes d'une octave sans modulation. Les modulations aléatoires étaient fondées sur des distributions d'énergie semblables à celles de la parole naturelle et avaient un niveau instantané qui était approximativement de 3 dB plus élevé que le niveau global de la parole pour cette bande d'une octave en cours d'évaluation.

La mesure directe du FULL STI étant, dans un grand nombre de situations, peu pratique, des méthodes plus rapides (STIPA, STITEL) ont été développées. La méthode FULL STI directe n'est actuellement utilisée que pour les travaux de recherche sur le STI.



IEC 1152/11

NOTE La valeur de la fonction de transfert de modulation (m) est déterminée pour toutes les cellules de la matrice des sept bandes d'une octave et des quatorze fréquences de modulation. De même, les niveaux d'intensité par octave, (\bar{I}_k , égal au carré des niveaux de pression sonore) sont obtenus pour calculer les effets du masquage auditif.

Figure A.3 – Système de mesure et fréquences pour la méthode STI

A.2.2 Fréquences de modulation STI

Les quatorze fréquences de modulation STI sont à des intervalles nominaux d'un tiers d'octave 0,63 Hz; 0,80 Hz; 1,00 Hz; 1,25 Hz; 1,60 Hz; 2,00 Hz; 2,50 Hz; 3,15 Hz; 4,00 Hz; 5,00 Hz; 6,3 Hz; 8,00 Hz; 10,0 Hz; 12,5 Hz.

NOTE En ce qui concerne les fréquences utilisées par la méthode STIPA, voir le Tableau B.1.

A.3 Effets auditifs sur le STI

A.3.1 Vue d'ensemble

Les aspects liés à l'audition, tels que le masquage auditif (perte de sensibilité auditive due à un son plus fort et de fréquence plus basse) [30] et le seuil absolu de réception, sont modélisés dans le calcul du STI par application de termes de bruit appropriés. Les effets auditifs réduisent le rapport signal sur bruit effectif dans les diverses bandes d'une octave et

peuvent être exprimés comme une réduction de la fonction de transfert de modulation conduisant à des valeurs de STI généralement plus faibles.

Le STI évalue la perte de modulation et calcule le rapport signal sur bruit réel dans chaque bande d'une octave indépendamment des autres bandes d'une octave. Le seul paramètre reliant entre elles des bandes de fréquences adjacentes quant au rapport signal sur bruit effectif est la fonction de masquage auditif, qui reflète les aberrations de la réponse amplitude/fréquence lorsque les niveaux de pression sonore sont élevés et que les fonctions de masquage ont des pentes faibles (voir aussi 4.5.8).

NOTE Les effets auditifs ne peuvent être pris en compte que lorsque les signaux d'essai sont obtenus acoustiquement (en dB SPL). En général, tel est souvent le cas. Cependant, si les signaux d'essai sont obtenus électriquement, sans référence à une pression sonore (par exemple lors de l'évaluation de différents types de microphones), il convient de le noter et de rendre les effets auditifs inopérants dans la mesure.

A.3.2 Masquage auditif dépendant du niveau

Le masquage auditif est un effet inhérent au processus d'audition humaine. Lorsqu'un son fort et à basse fréquence est présenté à l'oreille, il masquera toujours des fréquences plus élevées, les rendant éventuellement inaudibles, si la différence entre leurs niveaux relatifs dépasse un seuil donné. Ce phénomène est appelé masquage des hautes fréquences par les basses [30]. L'effet du masquage auditif dépend aussi de la pression sonore absolue des composantes des deux fréquences. Un son basse fréquence, à un niveau de pression sonore faible, provoque moins de masquage d'un son haute fréquence que s'il était à un niveau de pression sonore plus élevé, c'est-à-dire que la pente de masquage aux niveaux de pression sonore faibles est plus abrupte qu'aux niveaux de pression sonore élevés [31].

Le principe premier du masquage auditif, tel que modélisé dans le STI, est montré à la Figure A.4. Une bande d'une octave inférieure ($k - 1$) a un effet de masquage sur la bande suivante d'une octave immédiatement supérieure (k). La pente de la fonction de masquage dépend du niveau de pression sonore global sur la bande d'une octave ($k - 1$).

Le masquage auditif, tel que modélisé dans le STI, a un effet sur les bandes d'une octave de fréquence médiane 250 Hz, 500 Hz, 1 000 Hz, 2 000 Hz, 4 000 Hz et 8 000 Hz et s'étend sur la bande d'une octave plus haut. En conséquence, la bande d'une octave à 125 Hz masque la bande d'une octave à 250 Hz, la bande d'une octave à 250 Hz masque la bande d'une octave à 500 Hz, etc. La bande d'une octave à 125 Hz n'est pas masquée du tout.

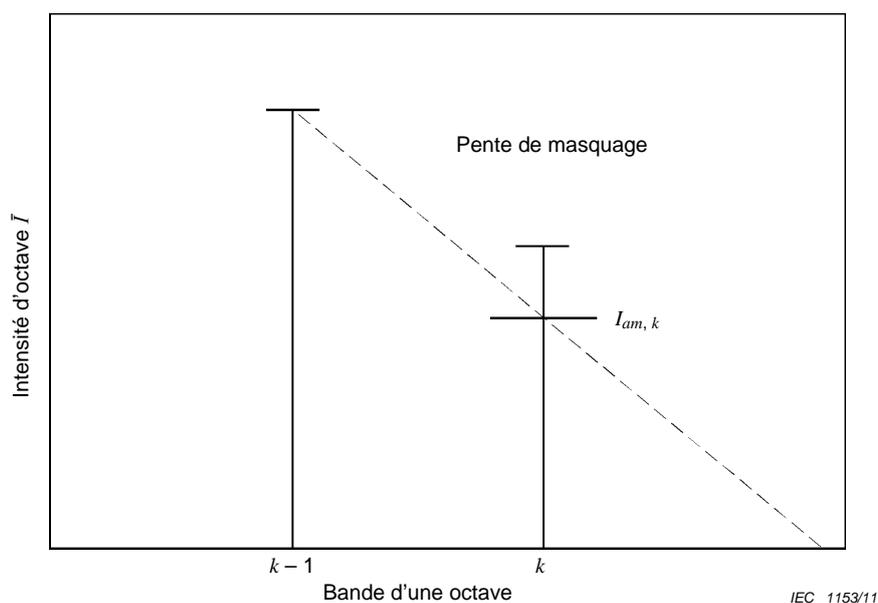


Figure A.4 – Masquage auditif de la bande d'une octave ($k - 1$) sur la bande d'une octave (k)

L'intensité du masquage ($I_{am,k}$) pour la bande d'une octave k est donnée par:

$$I_{am,k} = I_{k-1} \times amf$$

où

I_{k-1} est l'intensité de la bande d'une octave inférieure adjacente (bande d'une octave $k - 1$);

amf est le facteur de masquage auditif dépendant du niveau qui est une fonction dépendant de l'intensité (I_{k-1}) de la bande d'une octave adjacente inférieure.

L'intensité (I_{k-1}) pour la bande d'une octave $k-1$ est donnée par:

$$I_{k-1} = 10^{(L_{k-1}/10)}$$

où

L_{k-1} est le niveau de pression sonore global pour la bande d'une octave $k - 1$, en dB.

Le facteur de masquage auditif amf pour la bande d'une octave k dépend donc de l'intensité de la bande d'une octave $k - 1$.

Dans le Tableau A.1, le masquage auditif dépendant du niveau est donné en dB ($amdB$) pour la bande d'une octave k , en fonction du niveau de pression sonore L (dB) de la bande d'une octave $k - 1$. On notera que le niveau de masquage auditif est une fonction du niveau de pression sonore total dans cette bande d'une octave, constitué à la fois du niveau du signal d'essai et du niveau du bruit ambiant.

Tableau A.1 – Masquage auditif en fonction du niveau de pression sonore de la bande d'une octave

Niveau de pression sonore L de la bande d'une octave $k-1$ dB	<63	≥63 et <67	≥67 et <100	≥100
Masquage auditif $amdB$ dB	$0,5 \times L - 65$	$1,8 \times L - 146,9$	$0,5 \times L - 59,8$	-10

Le facteur de masquage auditif amf pour une bande d'une octave est donné par:

$$amf = 10^{(amdB/10)}$$

où

$amdB$ est la valeur du masquage auditif dépendant du niveau pour la bande d'une octave, en dB, telle que déduit du Tableau A.1.

NOTE 1 Si une correction mathématique du STI est faite pour une situation différente du bruit ambiant, le facteur de masquage auditif dépend à la fois de l'intensité de signal mesurée et de l'intensité additionnelle du bruit ambiant dans une certaine bande d'une octave. Ces deux intensités sont ajoutées pour obtenir une intensité acoustique globale pour une bande d'une octave, pour en déduire le facteur de masquage auditif approprié pour cette bande d'une octave en question.

NOTE 2 Le plan de masquage antérieur était discret et aboutissait à des résultats de STI non continus en fonction du niveau de pression sonore global. Le plan de masquage auditif, tel que présenté dans le Tableau A.1, fournit des résultats de STI continus en fonction du niveau de pression sonore.

A.3.3 Seuil absolu de réception de la parole

Le seuil de réception de la parole absolu est défini par le seuil absolu d'audition et la gamme dynamique minimale requise pour la reconnaissance correcte de la parole. L'intensité du seuil de réception de la parole absolu (I_{rt}) est modélisée dans le STI comme un bruit de fond intrinsèque réduisant le rapport signal sur bruit réel lorsque les niveaux de parole sont bas. Le seuil absolu de réception de la parole est donné dans le Tableau A.2 (en dB SPL) comme niveau dans chaque bande d'une octave.

Tableau A.2 – Niveau de seuil absolu de réception de la parole dans les bandes d'une octave

Bande d'une octave Hz	125	250	500	1 000	2 000	4 000	8 000
Seuil de réception de la parole absolu (ART) dB SPL	46	27	12	6,5	7,5	8	12

L'intensité du seuil de réception ($I_{rt,k}$) pour la bande d'une octave k est donnée par:

$$I_{rt,k} = 10^{(ART_k / 10)}$$

où

ART_k est le seuil de réception de la parole absolu pour la bande d'une octave k en dB (voir Tableau A.2).

A.3.4 Facteurs de pondération et facteurs de redondance de bandes d'une octave spécifiques au genre

La méthode STI peut distinguer entre les signaux de parole masculine et féminine. Les facteurs liés au genre sont exprimés dans des spectres de signal d'essai différents et dans des facteurs de pondération différents. La parole féminine étant généralement considérée comme étant plus intelligible que la parole masculine, c'est la parole masculine qui est utilisée en général pour évaluer les canaux de transmission de la parole.

Les facteurs de pondération (α) et les facteurs de redondance (β) du STI pour la parole masculine et la parole féminine sont donnés dans le Tableau A.3 en fonction des bandes d'une octave.

Tableau A.3 – Facteurs de pondération de bandes d'une octave STI

Bande d'une octave Hz		125	250	500	1 000	2 000	4 000	8 000
Hommes	α	0,085	0,127	0,230	0,233	0,309	0,224	0,173
	β	0,085	0,078	0,065	0,011	0,047	0,095	–
Femmes	α	–	0,117	0,223	0,216	0,328	0,250	0,194
	β	–	0,099	0,066	0,062	0,025	0,076	–

NOTE Les facteurs de pondération pour le STI masculin contiennent un artéfact qui peut parfois apparaître lors de calculs prédictifs. Si le rapport de transfert de modulation de la bande d'une octave de 250 Hz est inférieur ou égal à 0,08 (équivalent à un indice de transmission TI de 0,15), les autres bandes d'une octave étant à l'indice de transmission maximal de (1,0), la valeur STI sera supérieure à 1,0. Si la contribution de la bande de 250 Hz est entièrement éliminée, le STI deviendra 1,03. Pour cette situation, il est recommandé de tronquer le STI à 1,0. Dans des situations pratiques de mesure, il est peu vraisemblable que cet artéfact apparaisse, car le bruit l'empêchera de se produire.

A.3.5 Spectres des signaux d'essai STI spécifiques au genre

Les spectres des signaux d'essai STI sont spécifiés par les niveaux moyens des bandes d'une octave donnés dans le Tableau A.4. Les niveaux des bande d'une octave sont normalisés au niveau de 0 dB en pondération A pour faciliter le passage à un niveau global de pression sonore en pondération A. Ce spectre peut pendant de courts moments être dépassé de 3 dB par un signal d'essai modulé.

Tableau A.4 – Niveaux de bande d'une octave (dB) par rapport au niveau de parole pondéré A

Bande d'une octave Hz	125	250	500	1 000	2 000	4 000	8 000	Pondéré A
Hommes dB	2,9	2,9	-0,8	-6,8	-12,8	-18,8	-24,8	0,0
Femmes dB	-	5,3	-1,9	-9,1	-15,8	-16,7	-18,0	0,0

NOTE 1 Le signal d'essai STIPA vient du spectre de la parole masculine et, en général, il convient d'utiliser seulement ce signal.
NOTE 2 Voir l'Article 4 de l'Annexe J pour des directives sur la détermination du niveau de parole.

A.4 Génération du signal d'essai STI

A.4.1 Signal source de bruit rose

La méthode directe STI utilise des signaux d'essai de propriétés spectrales et temporelles similaires à celles trouvées dans la parole naturelle. En conséquence, les signaux d'essai STI se composent d'un certain nombre de bandes de fréquences de bruit, dont l'intensité est modulée par des sinusoïdes.

Le signal d'essai STI est initialement fait avec une source de bruit rose, donnant un spectre de fréquences plat, lorsqu'il est mesuré avec un filtrage d'octave fractionnaire.

Le bruit rose peut être produit à partir d'un bruit blanc en utilisant un filtre passe-bas, avec une pente de -3 dB par octave, commençant à 63 Hz. Il convient que le facteur de crête du signal se situe typiquement entre 12 dB et 14 dB, la valeur de dépassement L_1 se situant entre 8 dB et 11 dB.

A.4.2 Génération des "porteuses" des bandes d'une octave

Pour générer les sept signaux des "porteuses" de bruit STI, un signal de bruit rose est introduit dans sept filtres d'une largeur d'une demi-octave à des fréquences centrales comprises entre 125 Hz et 8 000 Hz. Après chaque "porteuse" de bruit il y a un trou large d'une demi-octave où il ne doit pas y avoir de signal à un niveau perceptible. Les filtres demi-octaves doivent, en bande coupée, atténuer d'au moins 60 dB par octave, afin de minimiser le recouvrement entre porteuses voisines. Il est recommandé que l'affaiblissement soit continu, qu'il n'ait aucune ondulation en bande coupée, comme cela serait le cas pour, par exemple, des filtres Tchebychev de type II ou elliptiques, et que l'ondulation dans la bande passante ne dépasse pas 0,3 dB.

EXEMPLE: Des filtres Tchebychev de type I, d'ordre 12, à ondulation de 0,2 dB, peuvent être utilisés pour générer les signaux des "porteuses" d'une demi-octave de largeur. D'autres méthodes valables relèvent du domaine fréquentiel.

A.4.3 Modulation d'intensité des "porteuses"

Chaque "porteuse" est modulée sinusoidalement en intensité avec une ou plusieurs fréquences de modulation, à la profondeur maximale de modulation possible (c'est-à-dire $m = 1,0$). Pour obtenir une porteuse modulée sinusoidalement en intensité, le signal de porteuse est modulé en amplitude avec la racine carrée d'une modulation sinusoidale, telle que donnée par:

$$m_{f_m}(t) = \sqrt{0,5 \cdot (1 + m \cdot \cos(2\pi f_m t))}$$

où

f_m est la fréquence de modulation en Hz;

t est le temps en secondes;

m est la profondeur de la modulation (égale au maximum à 1).

Si une bande porteuse est en même temps modulée par plus d'une fréquence de modulation, les facteurs de modulation doivent être égaux et les relations de phase et la profondeur de modulation doivent être choisis pour éviter une surmodulation de la porteuse.

A.4.4 Application du spectre de la parole au signal STI

Chaque "porteuse" se voit attribuer un niveau relatif, par bande d'une octave, selon le spectre de la parole donné au Tableau A.4. Comme ces "porteuses" proviennent d'un signal de bruit rose, les niveaux des bandes d'une octave sont similaires et déduits simplement selon le Tableau A.4. Tous les "signaux de porteuse" modulés sont sommés pour donner le signal d'essai STI.

A.5 Analyse du signal d'essai STI et calcul du STI

A.5.1 Traitement du signal

A.5.1.1 Étapes de base du traitement

Les paragraphes suivants exposent le mode opératoire pour l'analyse du signal reçu et le calcul du STI résultant. Le mode opératoire est décomposé en différentes étapes de base, comme suit:

- filtrer le signal d'entrée avec les sept filtres d'octaves;
- déterminer les intensités I_k dans chaque bande d'une octave k ;
- déterminer la modulation de l'intensité I à chaque fréquence de modulation f_m ;
- calculer le STI en suivant 4.3 et l'Article A.1.

A.5.1.2 Filtrage et détermination des intensités

Lors du filtrage du signal reçu dans les sept bandes d'une octave, les limites des tolérances des filtres doivent être conformes à la classe 0 ou à la classe 1 de la CEI 61260. La puissance du signal d'entrée doit être divisée par la batterie de filtres en signaux de sortie sans perte de puissance et, donc, les affaiblissements des filtres adjacents doivent être complémentaires en fonction de la fréquence et se couper à une atténuation de -3 dB.

Lors de l'application de la méthode directe, le signal d'essai de bruit modulé reçu est filtré en signaux de bandes d'une octave. Les filtres d'octaves ont une largeur d'une bande d'une octave avec des fréquences centrales allant de 125 Hz jusqu'à 8 000 Hz.

Il convient que les caractéristiques de phases soient aussi linéaires que possible de façon à éviter une distorsion de la relation de phase des modulations d'amplitude. Au cours de la

même mesure, il convient de calculer toutes les intensités en se référant au même segment de temps du signal d'entrée; aucune pondération temporelle implicite n'est permise.

Il convient que les caractéristiques de phase des filtres d'analyse ne produisent pas une erreur systématique supérieure à 0,01 STI dans le résultat final, pour la plage des valeurs comprises entre 0,1 et 0,9 STI (entre -12 dB et 12 dB de *SNR*).

Les intensités sont calculées en prenant le carré des signaux des sorties des filtres passe-bande.

A.5.2 Calcul de la fonction de transfert de modulation (MTF) à l'aide de la méthode directe

Le calcul du STI emploie la fonction de transfert de modulation. La MTF d'une voie de transmission peut être déterminée de multiples façons, la principale étant le calcul du facteur de perte de modulation par comparaison des modulations d'intensité à la sortie et à l'entrée de la voie de transmission.

Les sorties des filtres sont transformées en signaux temporels d'intensité par élévation au carré des signaux de sortie du filtre. L'application d'un filtre passe-bas à une fréquence de coupure de 100 Hz environ permet d'obtenir l'enveloppe de l'intensité pour chaque bande d'une octave.

En fonction du signal d'essai, la MTF pour chaque bande d'une octave est calculée par corrélation de l'enveloppe de l'intensité avec des signaux en sinus et cosinus d'une durée spécifique et de fréquences de modulation spécifiques. Le transfert de modulation à chaque fréquence de modulation est calculé en évaluant d'abord la profondeur de modulation du signal reçu (*mdr*) pour chaque sortie de bande d'une octave (*k*).

$$mdr_{k,f_m} = 2 \times \frac{\sqrt{[\sum I_k(t) \cdot \sin(2\pi f_m t)]^2 + [\sum I_k(t) \cdot \cos(2\pi f_m t)]^2}}{\sum I_k(t)}$$

où

f_m est la fréquence de modulation en Hz;

t est le temps en secondes;

$I_k(t)$ est l'enveloppe de l'intensité en fonction du temps, pour la bande d'une octave k .

NOTE 1 $I_k(t) \cdot \sin(2\pi f_m t)$ est le produit intérieur de l'enveloppe de l'intensité pour la bande d'une octave k et d'une modulation sinusoïdale spécifique f_m .

NOTE 2 La sommation est faite sur la durée de mesure, avec un nombre entier de périodes pour chaque fréquence spécifique de modulation.

Le rapport de transfert de modulation peut être calculé en utilisant les indices de modulation du signal reçu et du signal émis. Cette valeur est souvent appelée "valeur m ". L'ensemble des valeurs ainsi obtenues $m(k, f_m)$ des rapports de transfert de modulation est ce que l'on appelle la matrice MTF (fonction de transfert de modulation). Le rapport de transfert de modulation est donné par:

$$m_{k,f_m} = mdr_{k,f_m} / mdt_{k,f_m}$$

où

mdr_{k,f_m} est la profondeur de modulation du signal d'essai reçu pour la bande d'une octave k et la fréquence de modulation f_m ;

mdt_{k,f_m} est la profondeur de modulation du signal d'essai émis pour la bande d'une octave k et la fréquence de modulation f_m .

NOTE L'Article 6 décrit le calcul de la fonction de transfert de modulation (MTF) par la méthode indirecte.

A.5.3 Correction de la MTF en utilisant le masquage auditif

Les valeurs (valeurs m) ainsi calculées des rapports de transfert de modulation sont corrigées en utilisant les effets du masquage auditif, en appliquant la formule suivante:

$$m'_{k,f_m} = m_{k,f_m} \times \frac{I_k}{I_k + I_{am,k} + I_{rt,k}}$$

où

m_{k,f_m} est la valeur du rapport de transfert de modulation calculée pour la bande d'une octave k et la fréquence de modulation f_m ;

I_k est le niveau d'intensité acoustique pour la bande d'une octave k ;

$I_{am,k}$ est le niveau d'intensité acoustique pour l'effet de masquage auditif dépendant du niveau, sur la bande d'une octave k ;

$I_{rt,k}$ est le niveau d'intensité acoustique du seuil de réception, pour la bande d'une octave k .

NOTE 1 Il convient de tronquer à 1,0 les valeurs du rapport de transfert de modulation supérieures à 1,0. Une valeur m supérieure à 1,3 est très peu probable et peut être le résultat d'un phénomène autre que les fluctuations sinusoïdales ou les bruits impulsifs.

NOTE 2 Lors de l'application mathématique d'un niveau du bruit ambiant supplémentaire, il convient d'étendre le terme du dénominateur avec $I_{n,k}$ qui représente le niveau d'intensité acoustique du bruit ambiant pour la bande d'une octave k . Noter que l'intensité du masquage auditif ($I_{am,k}$) est aussi affectée par l'application mathématique du bruit ambiant supplémentaire.

A.5.4 Calcul du rapport "signal à bruit" effectif

Les valeurs corrigées du rapport de transfert de modulation sont transformées en un rapport signal sur bruit effectif SNR_{eff} (dB) donné par:

$$SNR_{\text{eff } k, f_m} = 10 \times \log \frac{m'_{k, f_m}}{1 - m'_{k, f_m}}$$

où

m'_{k, f_m} est la valeur du rapport de transfert de modulation corrigée pour la bande d'une octave k et la fréquence de modulation f_m .

Le résultat du calcul du rapport signal sur bruit pouvant être infini, les valeurs doivent être limitées à la gamme de -15 dB à +15 dB.

A.5.5 Calcul de l'indice de transmission (TI)

L'indice de transmission (TI) pour chaque bande d'une octave et chaque fréquence de modulation est calculé par:

$$TI_{k,f_m} = \frac{SNR_{\text{eff } k, f_m} + 15}{30}$$

où

$SNR_{\text{eff } k, f_m}$ est le rapport signal sur bruit effectif, exprimé en dB, pour chaque bande d'une octave k et chaque fréquence de modulation f_m .

A.5.6 Calcul du STI

Les indices de transmission (TI) calculés sont moyennés sur les fréquences de modulation pour obtenir l'indice de transfert de modulation (MTI_k) par bande d'une octave k en utilisant:

$$MTI_k = \frac{1}{n} \sum_{m=1}^n TI_{k,f_m}$$

où

TI_{k,f_m} est l'indice de transmission pour chaque bande d'une octave k et chaque fréquence de modulation f_m ;

m est l'indice de la fréquence de modulation;

n est le nombre de fréquences de modulation par bande d'une octave.

Avec les indices de transfert de modulation (MTI_k) pour chaque bande d'une octave k , le STI est calculé par:

$$STI = \sum_{k=1}^7 \alpha_k \times MTI_k - \sum_{k=1}^6 \beta_k \times \sqrt{MTI_k \times MTI_{k+1}}$$

où

MTI_k est l'indice de transfert de modulation pour la bande d'une octave k ;

α_k est le facteur de pondération pour la bande d'une octave k ;

β_k est le facteur de redondance entre la bande d'une octave k et la bande d'une octave $k+1$.

NOTE Si les valeurs de STI obtenues sont supérieures à 1,0, il convient de les mettre à 1,0.

Annexe B (normative)

Méthode STIPA

Au lieu des quatorze fréquences de modulation appliquées successivement à l'ensemble des sept bandes d'une octave du mode opératoire de la méthode FULL STI, la méthode STIPA applique simultanément deux fréquences de modulation uniques à chacune des sept bandes de fréquences, comme indiqué dans le Tableau B.1. Il est donc utilisé un total de $2 \times 7 = 14$ fréquences de modulation.

Chaque bande d'une octave est modulée simultanément par deux fréquences de modulation dont le rapport est 5; l'indice de modulation (profondeur) pour chaque fréquence de modulation doit être de 0,55 pour une addition sinusoïdale des deux composantes avec une différence de phase de 180° entre celles-ci.

La méthode STIPA n'est validée que pour le spectre de la parole masculine et son temps de mesure est approximativement de 15 s à 20 s.

Tableau B.1 – Fréquences de modulation pour la méthode STIPA

Bande d'une octave Hz	125	250	500	1 000	2 000	4 000	8 000
Première fréquence de modulation Hz	1,60	1,00	0,63	2,00	1,25	0,80	2,50
Seconde fréquence de modulation Hz	8,00	5,00	3,15	10,0	6,25	4,00	12,5

Annexe C (normative)

Méthode STITEL

C.1 Généralités

Il est possible de simplifier le signal d'essai si sont omises ces modulations non corrélées (similaires à de la parole) nécessaires à une interprétation correcte des distorsions non linéaires. Cela permet une modulation et un traitement simultanés de toutes les sept bandes de fréquences, ce qui réduit le temps de mesure. La méthode STITEL utilise cette simplification et prend de 10 s à 15 s par mesure.

La méthode STITEL utilise une fréquence de modulation par porteuse de bande d'une octave, ce qui permet d'appliquer une modulation à 100 % du signal d'essai et augmente ainsi le *SNR* de 3 dB.

Au lieu des quatorze fréquences de modulation appliquées successivement à l'ensemble des sept bandes d'une octave pour la méthode FULL STI, la méthode STITEL applique simultanément une unique fréquence de modulation à chacune des sept bandes d'une octave, comme indiqué dans le Tableau C.1. Le signal d'essai comprend toutes les sept bandes modulées d'une octave qui sont analysées simultanément.

Tableau C.1 – Fréquences de modulation pour la méthode STITEL

Bande d'une octave Hz	125	250	500	1 000	2 000	4 000	8 000
Fréquence de modulation Hz	1,12	11,33	0,71	2,83	6,97	1,78	4,53

Mis à part les fréquences de modulation ci-dessus, le schéma normal de calcul du STI est appliqué pour le STITEL.

C.2 Limitations

Il convient de ne pas utiliser la méthode STITEL pour des canaux de transmission:

- qui introduisent des glissements de fréquence ou une multiplication de fréquences;
- qui comprennent des vocodeurs (c'est-à-dire LPC, CELP, RELP, etc.);
- qui introduisent des composantes de distorsion fortement non linéaire;
- pour lesquels le temps de réverbération est fortement dépendant de la fréquence;
- qui présentent des échos à plus de -10 dB par rapport au signal primaire;
- si le bruit de fond contient des sons purs perceptibles, et/ou des crêtes ou des creux marqués dans le spectre de bande d'une octave;
- si le bruit de fond est de caractère impulsif et/ou que l'endroit n'est pas réellement dépourvu d'échos discrets, plus particulièrement d'échos flottants, dont la fréquence de répétition est un multiple entier d'une ou de plusieurs des fréquences de modulation.

Si c), d), e) ou tous les trois s'appliquent, ou sont susceptibles de s'appliquer, il convient d'utiliser plutôt la méthode STI ou de l'utiliser pour vérifier les résultats obtenus par la méthode STITEL.

Annexe D (informative)

Méthode RASTI (désuète)

Bien qu'elle soit désormais désuète, la méthode RASTI est présentée en détail par souci d'exhaustivité. Cette technique avait été créée pour être une façon plus rapide d'obtenir une mesure du STI, les données et l'effort de calcul y étant réduits d'un ordre de grandeur. Que la méthode n'emploie que deux octaves a été trouvé être une insuffisance sérieuse pour l'essai de systèmes électroacoustiques. L'utilisation de deux octaves seulement a été considérée insuffisante pour l'essai de systèmes électroacoustiques.

Comme montré au Tableau D.1, un nombre total de neuf fréquences de modulation est appliqué à deux bandes d'une octave.

Tableau D.1 – Fréquences de modulation pour la méthode RASTI

Fréquence de modulation Hz	f_1	f_2	f_3	f_4	f_5
Bande d'une octave de 500 Hz	1,0	2,0	4,0	8,0	-
Bande d'une octave de 2 000 Hz	0,7	1,4	2,8	5,6	11,2

La pondération en fréquence est appliquée indirectement en utilisant quatre contributions pour la bande d'une octave à 500 Hz et cinq pour la bande d'une octave à 2 kHz, c'est-à-dire que les pondérations sont respectivement de 4/9 (0,45) et de 5/9 (0,55).

La Figure D.1 montre un signal d'essai RASTI en pratique.

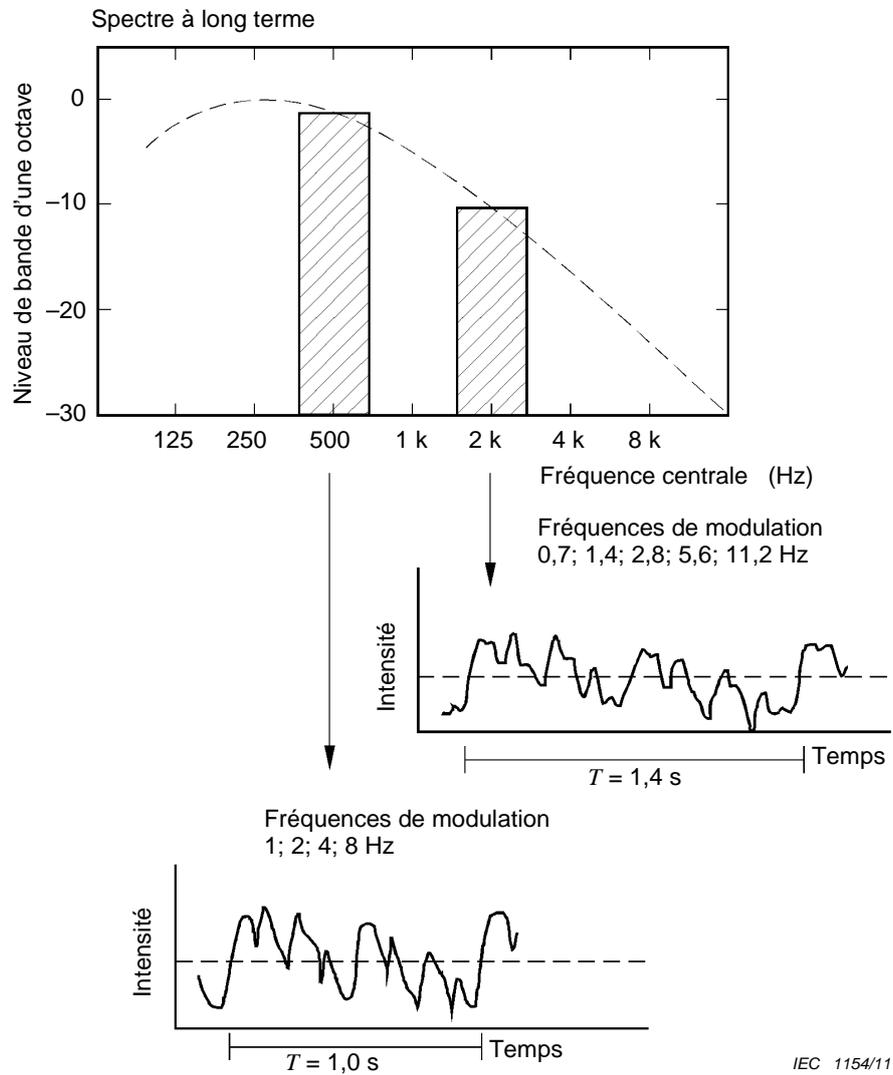


Figure D.1 – Illustration d'un signal d'essai RASTI en pratique

Annexe E (informative)

Qualification du STI et relations avec d'autres mesures de l'intelligibilité de la parole

E.1 Relation entre STI et "note d'intelligibilité" de mots ou de phrases

Les relations entre le STI et diverses mesures de l'intelligibilité de la parole sont données à la Figure E.1. L'indicateur d'intelligibilité de mots dépourvus de sens pour des mots de type consonne-voyelle-consonne (CVC) équilibrés est tiré de [32]. Le rapport du STI avec des mots anglais comprenant les deux lettres consécutives PB de la liste dite de "Harvard" est tiré de [33]. Le rapport avec l'intelligibilité de phrases est tiré de résultats sur le seuil de réception de la parole (*Speech Reception Threshold, SRT*).

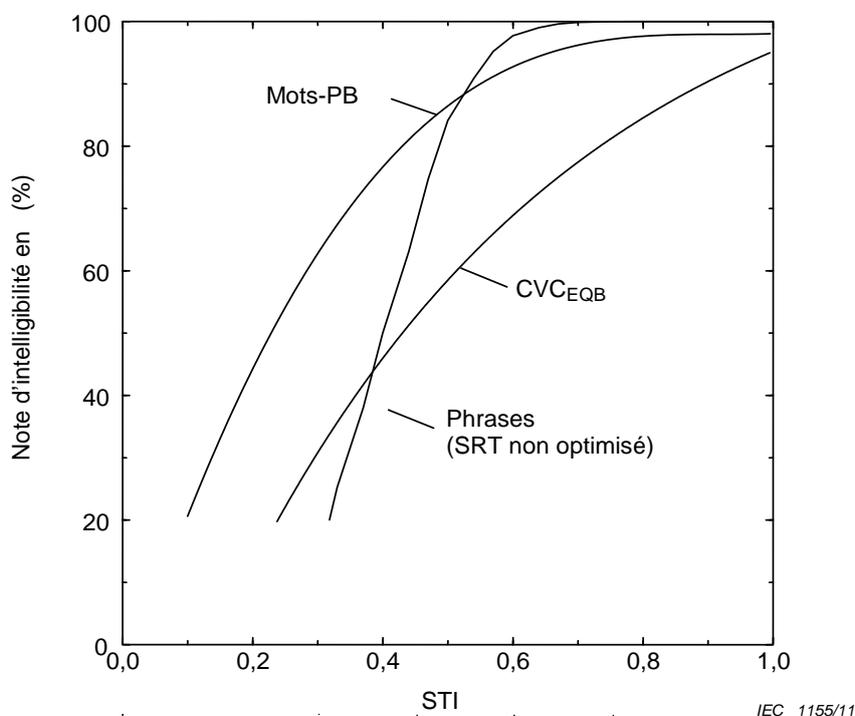


Figure E.1 – Relations entre quelques mesures de l'intelligibilité de la parole

E.2 Relation entre STI et "difficulté d'écoute"

La "difficulté d'écoute" peut, dans certains cas, être une meilleure évaluation de l'intelligibilité que les "notes d'intelligibilité" de mots ou de phrases. La métrique "difficulté d'écoute" est mieux adaptée aux canaux de transmission de la parole de haute qualité.

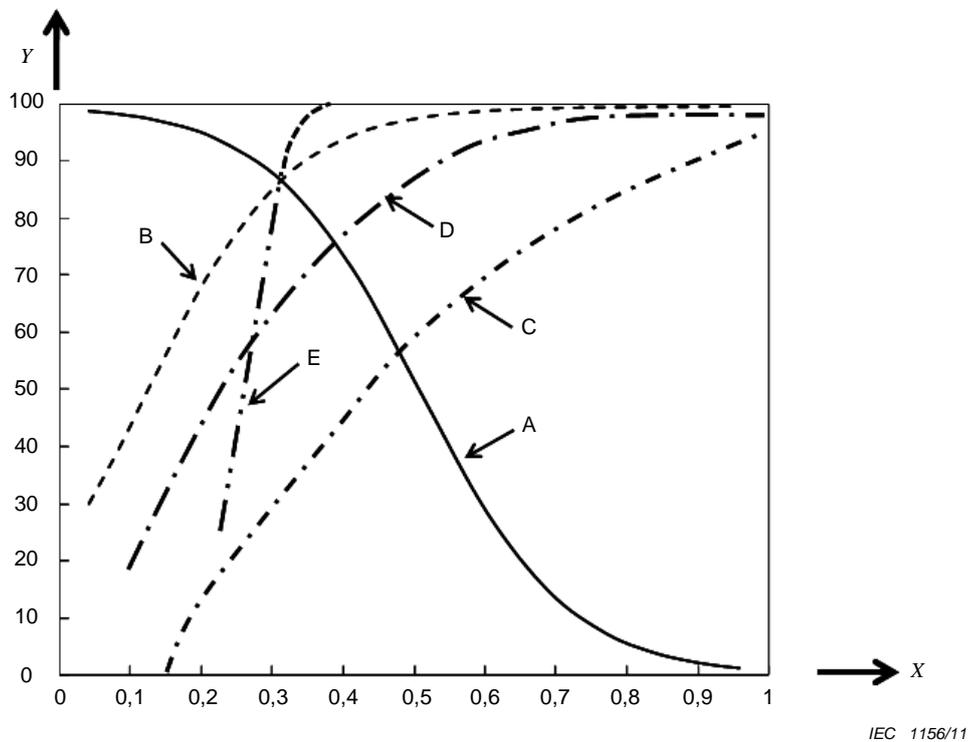
La "difficulté d'écoute" est le pourcentage de réponses indiquant un certain degré de difficulté pendant l'écoute d'un discours à travers un canal de transmission. Noter que les évaluations de la "difficulté d'écoute" sont des chiffres décroissant quand les conditions de transmission sont meilleures, contrairement aux "notes d'intelligibilité" conventionnelles [31].

Le Tableau E.1 énumère les catégories typiquement utilisées pour décrire la difficulté d'écoute.

Tableau E.1 – Catégories pour apprécier la difficulté d'écoute

Catégorie	Description
0	Pas difficile: aucun effort n'est requis, cela équivaut à une condition d'écoute relâchée.
1	Légèrement difficile: une légère attention est requise.
2	Modérément difficile: une attention modérée est requise.
3	Très difficile: une attention considérable est requise.

La Figure E.2 montre la relation entre les évaluations de la difficulté d'écoute, les notes d'intelligibilité de la Figure E.1 et le STI.



Légende

L'axe X correspond aux valeurs de STI.

L'axe Y correspond aux pourcentages de réponses correctes et de difficulté d'écoute.

La courbe A est celle de la difficulté d'écoute.

La courbe B est celle de la reconnaissance de mots.

La courbe C est celle de CVC_{EQB} (consonne-voyelle-consonne, "équilibré")

La courbe D est celle des mots PB (mots anglais avec p et b consécutifs)

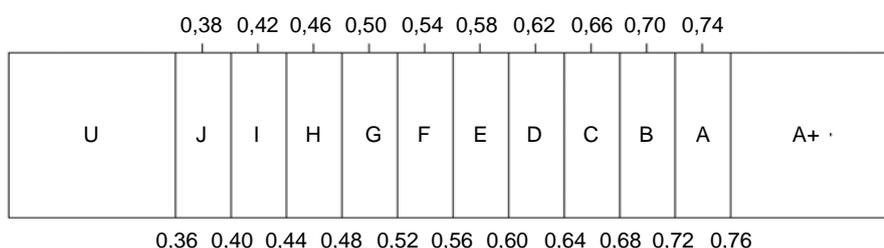
La courbe E est celle des phrases (SRT, *Speech Reception Threshold*).

Figure E.2 – Relation entre STI, notes (scores) d'intelligibilité de la parole et évaluations de la difficulté d'écoute [34], [35]

Annexe F (informative)

Bandes nominales pour qualifier un indice STI

L'échelle de qualification est divisée en plusieurs intervalles ou "bandes" pour fournir de la souplesse pour des applications multiples et pour exprimer les tolérances ou marges d'erreur applicables aux mesures et aux prédictions. La valeur de l'indice STI nécessaire pour une application ou une situation s'en déduit par la table des performances associées à ces intervalles (voir un exemple en Annexe G). La Figure F.1 montre ces intervalles de l'échelle de qualification.



IEC 1157/11

Légende

Rangée supérieure de nombres: valeurs du STI au centre des intervalles ou "bandes".

Rangée de lettres: désignations des intervalles ou "bandes".

Rangée inférieure de nombres: valeurs du STI aux bords des intervalles ou "bandes".

Figure F.1 – "Bandes" ou intervalles servant à qualifier un STI

L'espacement des intervalles de la Figure F.1 se fonde sur l'incertitude typique des mesures directes du STI.

Pour des mesures faites sur toute une zone ou surface, il est recommandé d'analyser la dispersion des résultats par leur distribution statistique.

Annexe G (informative)

Exemples d'intervalles ou "bandes" pour qualifier un indice STI et applications typiques

Les informations données dans le Tableau G.1 sont présentées comme un exemple d'utilisation.

Tableau G.1 – Exemples de correspondances entre les "bandes" qualifiant un indice STI et des applications typiques

Catégorie	Valeur nominale du STI	Informations sur le type de message	Exemples typiques d'utilisations (pour la voix naturelle ou reproduite)	Commentaire
A+	>0,76		Studios d'enregistrement	Intelligibilité excellente mais rarement obtenue dans la plupart des environnements
A	0,74	Messages complexes, mots inconnus	Théâtres, auditoriums, parlements, tribunaux, systèmes d'aide à l'audition (AHS, Assistive Hearing Systems)	Intelligibilité élevée de la parole
B	0,7	Messages complexes, mots inconnus		
C	0,66	Messages complexes, mots inconnus	Théâtres, auditoriums, téléconférences, parlements, tribunaux	Intelligibilité élevée de la parole
D	0,62	Messages complexes, mots familiers	Salles de conférence, salles de classe, salles de concert	Bonne intelligibilité de la parole
E	0,58	Messages complexes, contexte familier	Salles de concert, églises modernes	Systèmes PA (<i>public address</i> , sonorisation) de haute qualité
F	0,54	Messages complexes, contexte familier	Systèmes PA (sonorisation) dans les galeries marchandes, bureaux dans les bâtiments publics, systèmes VA (voice alarm), cathédrales	Systèmes PA de bonne qualité
G	0,5	Messages complexes, contexte familier	Galeries marchandes, bureaux dans les bâtiments publics, systèmes VA	Valeurs cibles pour les systèmes VA (voice alarm)
H	0,46	Messages simples, mots familiers	Systèmes VA et PA dans des environnements acoustiques difficiles	Limite inférieure normale pour les systèmes VA
I	0,42	Messages simples, contexte familier	Systèmes VA et PA dans des endroits très difficiles	
J	0,38		Ne convient pas aux systèmes PA	
U	<0,36		Ne convient pas aux systèmes PA	

NOTE 1 Ces valeurs sont à voir comme des objectifs minimaux.

NOTE 2 L'intelligibilité perçue pour chaque catégorie dépend aussi de la réponse en fréquence à chaque position d'écoute.

NOTE 3 Les valeurs de l'indice STI se réfèrent aux valeurs mesurées sur un échantillon de positions d'écoute ou selon les exigences de normes particulières d'application.

Annexe H (informative)

Auditeurs de langue étrangère

Les auditeurs de langue maternelle étrangère ont en général besoin d'une amélioration de 4 dB à 5 dB du rapport signal sur bruit pour une intelligibilité similaire (ISO 9921:2003) à celle obtenue avec des auditeurs écoutant dans leur langue maternelle.

Des tableaux de l'indice d'intelligibilité, ajustée pour trois classes de personnes de langue étrangère, peuvent être définis à partir de données expérimentales [36]. La compétence linguistique de la personne de langue étrangère sera classée selon l'expérience de la langue, l'âge de l'apprentissage et la fréquence de l'utilisation de cette seconde langue.

Le Tableau H.1 n'est donné qu'à titre indicatif. Pour plus de détails, se référer aux normes ergonomiques (ISO 9921). Une intelligibilité bonne ou excellente ne peut être atteinte par des auditeurs de langue maternelle étrangère et de faible compétence linguistique.

NOTE A titre de guide, une augmentation de 0,1 du STI correspond à une modification de 3 dB du rapport signal sur bruit effectif.

Tableau H.1 – Indice d'intelligibilité ajusté pour des auditeurs de langue maternelle étrangère selon leurs compétences linguistiques

Plage d'évaluation correspondant à l'indice STI standard ou "étiquette STI"	STI standard	Catégorie I de personnes de langue étrangère expérimentées, utilisant quotidiennement la seconde langue	Catégorie II de personnes de langue étrangère expérience et niveau intermédiaires d'utilisation de la seconde langue	Catégorie III de personnes de langue étrangère nouvel apprenant, utilisation rare de la seconde langue
mauvais à médiocre	0,30	0,33	0,38	0,44
médiocre à passable	0,45	0,50	0,60	0,74
passable à bon	0,60	0,68	0,86	impossible
bon à excellent	0,75	0,86	impossible	impossible
<p>NOTE 1 Pour plus de détails sur les catégories d'étiquettes STI, se référer à l'ISO 9921 [49].</p> <p>EXEMPLE: Pour qu'un auditeur de langue étrangère de catégorie II arrive à une intelligibilité équivalant à un STI de 0,45, le système de transmission doit assurer une performance de 0,60.</p> <p>NOTE 2 Pour les valeurs intermédiaires entre les indices STI standard, il convient d'utiliser l'interpolation pour déduire l'indice STI ajusté.</p>				

Annexe I
(informative)

Effet sur l'intelligibilité de la parole de la perte d'audition et de troubles de l'audition liés à l'âge

Pour les malentendants, la dérivation des tableaux de qualification de l'intelligibilité ajustée est plus complexe que pour les auditeurs de langue étrangère et exige de définir le type de trouble de l'audition.

La méthode STI ne peut pas donner de résultats fiables pour tous les types de trouble de l'audition et, en général des tests d'écoute par des sujets ou d'autres méthodes de prédiction telles que l'indice d'intelligibilité de la parole (*speech intelligibility index, SII*) sont recommandées.

Pour les auditeurs de plus de 50 ans, les niveaux d'audition et la dispersion entre individus augmentent considérablement [37]. Néanmoins, les troubles de l'audition liés à l'âge présentent une bonne corrélation entre l'intelligibilité et la perte d'audition.

Une règle empirique est que, pour atteindre l'intelligibilité au point critique d'intelligibilité des phrases de 50 % (où typiquement les phrases redondantes doivent être répétées deux fois), les auditeurs malentendants ont besoin d'un *SNR* plus élevé de 4,5 dB pour 20 dBHL [38]. Ici, l'unité dBHL se rapporte à la perte d'audition (HL, Hearing Loss) en dB, définie comme étant le niveau moyen d'audition d'un son pur (PTA, Pure-Tone Average Hearing Level) de 0,5 kHz, 2,0 kHz et 4,0 kHz, par rapport à des auditeurs normaux âgés de 18 ans.

Le Tableau I.1 donne une indication des qualifications d'intelligibilité ajustée. Noter que l'intelligibilité maximale qui peut être atteinte dépend du degré de la perte d'audition [15], [38]. Pour plus de détails, se référer à d'autres normes, telles que la référence [39].

Tableau I.1 – Tableaux de qualification de l'intelligibilité ajustée pour les auditeurs normaux et âgés de plus de 60 ans présentant une perte d'audition
PTA = pure tone average hearing level

Catégorie d'étiquette STI	Auditeurs normaux (STI standard)	Auditeurs âgés PTA = 15 dB	Auditeurs âgés PTA = 20 dB	Auditeurs âgés PTA = 30 dB
mauvais à médiocre	0,30	0,42	0,47	0,51
médiocre à passable	0,45	0,57	0,62	0,66
passable à bon	0,60	0,72	Ne peut être obtenue	Ne peut être obtenue
bon à excellent	0,75	Ne peut être obtenue	Ne peut être obtenue	Ne peut être obtenue

NOTE 1 Pour plus de détails sur les catégories, se référer à l'ISO 9921 [49].
 NOTE 2 Les auditeurs normaux typiques ont un PTA compris entre 0 dB et 5 dB.
 EXEMPLE: Pour qu'un auditeur âgé de PTA = 15 dB ait une intelligibilité équivalant à un STI de 0,45, le système de transmission doit assurer un STI de 0,57.

Annexe J (normative)

Etalonnage du niveau du signal d'essai STI

J.1 Vue d'ensemble

Pour réaliser des mesures reproductibles du STI, il convient d'étalonner soigneusement les niveaux de parole et les niveaux de bruit. Cependant, les signaux de parole et les signaux de bruit peuvent avoir des caractéristiques temporelles et spectrales différentes, ce qui rend difficile une comparaison aisée et précise. Il convient donc pour déterminer les rapports parole-à-bruit que la méthode de mesure de niveau soit adaptée à des types de parole divers (masculine / féminine, discours continu / mots isolés) à des conditions diverses (bruit de fond, largeur de bande) et qu'elle soit applicable aussi à des signaux semblables à du bruit.

Les méthodes de base de la mesure du niveau d'une parole réelle sont fondées sur l'élimination des parties silencieuses du signal de parole, par exemple des trous entre mots. Une méthode précise (le mode opératoire au volumètre) est décrite à l'Article J.2. D'autres méthodes peuvent encore être employées si la précision requise est moindre. Une méthode simple, disponible aussi sur des instruments courants de mesure du niveau de pression sonore est décrite aux Articles J.3 et J.4.

J.2 Le concept de "niveau de la parole" et sa méthode de mesure

La méthode de mesure du niveau de la parole a été mise au point par Houtgast et Steeneken (1978, 1986). Pour surmonter les différences spectrales entre les signaux et pour que les niveaux de signal soient en correspondance étroite avec le niveau perçu, la méthode de mesure se fonde sur le filtrage pondéré A du signal. En général, les signaux de parole ne sont pas continus et contiennent de nombreuses pauses. Spécifier le rapport signal à bruit des signaux parole et bruit à partir de la mesure de $L_{A,eq}$ rend évidemment le résultat dépendant du nombre et de la durée des silences entre les émissions de sons. Par exemple, une parole avec exactement le même niveau nominal, mais un modèle différent de pauses, aura une valeur $L_{A,eq}$ mesurée différente, d'où des différences entre les rapports signal à bruit résultants.

La méthode de mesure du niveau de la parole résout cette difficulté en éliminant tous les silences avant de calculer le niveau, de façon à ne prendre en compte que les seules parties du signal de parole contenant des informations pertinentes. Le signal pondéré A est donc divisé en trames d'une longueur de 10 ms à 20 ms et l'énergie par trame calculée. Ensuite, toutes les valeurs d'énergie par trame sont cumulées dans un histogramme de distribution des niveaux. Le niveau quadratique moyen (RMS) de la parole est déterminé à partir de cet histogramme de distribution des niveaux, par itérations, en cumulant toutes les données de l'histogramme en commençant à un seuil qui se situe 14 dB en dessous du niveau quadratique moyen du signal de parole calculé à l'itération précédente. Ce mode opératoire fait que les parties relativement silencieuses du signal sont laissées de côté; le niveau de signal devient donc indépendant de la distribution temporelle du signal.

Des travaux de recherche approfondie [40] sur différentes mesures du niveau de la parole montrent que la différence de niveau quadratique moyen (RMS) entre du discours continu et des mots CVC (consonne-voyelle-consonne) insérés dans des phrases porteuses est minimale (< 0,5 dB) quand on emploie cette méthode de mesure du niveau de la parole.

J.3 Niveau de parole d'essai

Le niveau de parole lors de mesures sur un système de sonorisation (*PA, public address*) est le niveau effectif trouvé à un point d'écoute spécifié, le système fonctionnant correctement. Si l'entrée du signal provient d'un locuteur ou d'une autre source acoustique, l'alinéa ci-dessous s'applique pour régler le niveau d'entrée au système.

Pour les mesures avec un locuteur ou une autre source acoustique, en l'absence de système de sonorisation (*PA*), le niveau de parole d'essai doit être réglé à 60 dB A mesuré à une distance de 1 m, sur l'axe de rayonnement principal de la bouche artificielle ou du locuteur. S'il est nécessaire de simuler une condition d'effort vocal accru (effet Lombard), le niveau de parole d'essai doit être réglé à 70 dB A.

J.4 Niveau de parole corrigé

Une mesure du "niveau de parole corrigé approché", avec une simple mesure d'un niveau de pression sonore équivalent pondéré A ($L_{A,eq}$), a le mode opératoire suivant. Le niveau mesuré est corrigé d'un facteur obtenu empiriquement, afin d'obtenir une estimation du niveau réel de la parole déterminé avec la méthode décrite à l'Article J.2. Afin d'obtenir le niveau de parole corrigé approché, procéder comme suit.

- Déterminer le $L_{A,eq}$ d'un signal vocal, avec une longueur d'au moins 40 s, à moins que le signal ne soit une annonce enregistrée de plus courte durée.
- Ajouter 3 dB au résultat.

NOTE Le facteur de correction de 3 dB peut varier en fonction du débit de parole et des caractéristiques d'un locuteur particulier.

Annexe K (informative)

Exemple de fiche de rapport d'essai pour les mesures de STI

Ceci est un ensemble général de directives et permet de faciliter la mesure du STI pour toute une gamme d'applications. Toutes les catégories ne sont pas applicables dans tous les cas.

INFORMATIONS GÉNÉRALES	
Méthode de mesure	
Projet / emplacement	
Occupation / configuration	
Date de mesure	
Méthode: Indirecte (IR) ou Directe (STIPA)	
Source	
Type de signal: MLS; balayage sinusoïdal; STIPA; autre	
Type de source: Générateur de signal, lecteur de CD; lecteur de .wav ou de .mp3	
Spectre du signal d'essai: Ajusté au spectre de parole normalisé?	
Méthode d'insertion du signal: Entrée électronique ou diffusion à partir d'une source acoustique?	
Détails du haut-parleur d'essai / de la bouche artificielle / type	
Distance de la source au microphone du système (m)	
Diagramme directionnel du microphone du système électroacoustique	
Distance du microphone à la surface réfléchissante la plus proche (m)	
Traitement de signal du système	
État du traitement de signal du système, par exemple, compression, limiteurs, égalisation	
Une partie de la chaîne de signaux est-elle écrêtée pendant la mesure?	
Équipement matériel de mesure	
Marque /Type - n° de série / version	
Résultat d'essai en opposition en boucle STI ou STIPA	
Microphone de mesure	
Marque, modèle et type (champ libre, incidence aléatoire)	
Monaural ou binaural?	
Hauteur au-dessus du plancher/sol (m)	
Point cible du microphone	

INFORMATIONS GÉNÉRALES	
Impression subjective avec transmission de la parole	
Caractéristiques du son: naturel, étouffé, caverneux, résonant, rauque?	
Y a-t-il des distorsions audibles ou des échos?	
Pour l'insertion d'un signal acoustique: Y a-t-il une sonnerie ou une régénération audible (rétroaction)?	

DONNÉES DE MESURE PAR POSITION (utiliser une fiche différente pour chaque position)								
Référence de position / emplacement								
Temps de mesure								
Niveaux du bruit ambiant (sans signal d'essai)								
	$L_{A,eq}$ (dB)	Bande d'une octave (Hz)						
		125	250	500	1 000	2 000	4 000	8 000
L_{eq} mesuré sur 15 s (dB)								
Description du bruit ambiant: par exemple, constant, transitoire, impulsif								
Conditions climatiques: vent, pluie, température, humidité								
Méthode STIPA								
	$L_{A,eq}$ (dB)	Bande d'une octave (Hz)						
		125	250	500	1 000	2 000	4 000	8 000
L_{eq} mesuré sur 15 s (dB)								
Numéro de mesure de STIPA	Moyenne	1	2	3	4	5	6	
Résultats du STIPA pour chaque mesure: prendre trois mesures; si la variation >0,03, en prendre trois de plus.								
Variation maximale des résultats								
Méthode de la réponse impulsionnelle (IR)								
Longueur de la réponse impulsionnelle acquise. Minimum pour le calcul de STI: 1,6 s								
	$L_{A,eq}$ (dB)	Bande d'une octave (Hz)						
		125	250	500	1 000	2 000	4 000	8 000
L_{eq} avec signal d'essai (dB)								
Rapport signal à bruit SNR (dB)								

Résultats de STI	STI	Indice de transmission MTI						
Indices de transmission et STI								
Écart de la réponse en fréquence de 250 Hz à 12 kHz par rapport à 1 kHz.								
Mesure de la réponse en fréquence: fenêtre de temps appliquée, lissage.								
La réponse IR montre-t-elle des arrivées susceptibles de provoquer des échos audibles?								
Résultats du traitement des données mesurées pour différents niveaux de signal et de bruit								
Niveaux de signal	$L_{A,eq}$ (dB)	Bande d'une octave (Hz)						
		125	250	500	1 000	2 000	4 000	8 000
Niveau de parole proposé (dB)								
Niveau du bruit ambiant proposé (dB)								
Résultats de STI	STI	Indice de transmission MTI						
STI et MTI calculés								

Annexe L (normative)

Prédiction du STI à l'aide de méthodes statistiques

Selon la référence [28], la fonction de transfert de modulation globale avec la distorsion temporelle et le bruit peut s'écrire:

$$m(f_m) = \frac{\left| \int_0^{\infty} h(t)^2 e^{-j2\pi f_m t} dt \right|}{\int_0^{\infty} h(t)^2 dt} \cdot \left[1 + 10^{-SNR/10} \right]^{-1}$$

où

$m(f_m)$ est la fonction de transfert de modulation du canal de transmission;

f_m est la fréquence de modulation;

$h(t)$ est la réponse impulsionnelle du canal de transmission;

t est la variable d'intégration pour le temps;

SNR est le rapport signal à bruit en dB.

En supposant un champ de réverbération diffus, la réponse impulsionnelle pour les composantes de champ direct et réverbérant pour une source unique peut s'écrire:

$$h(t) = \frac{Q}{r^2} \cdot \delta(t) + \frac{13,8 Q}{r_c^2 T} e^{-\frac{13,8t}{T}}$$

où

Q est le facteur de directivité pour la source sonore (haut-parleur ou locuteur);

$\delta(t)$ est la fonction de Dirac (ou delta);

r est la distance du locuteur à l'auditeur;

r_c est la distance critique dans la salle ou dans l'espace acoustique (également connu sous le nom de rayon critique);

T est le temps de réverbération de la salle ou de l'espace acoustique.

La fonction de transfert de modulation avec la distorsion temporelle et le bruit peut alors s'écrire:

$$m(f_m) = \frac{\sqrt{A^2 + B^2}}{C}$$

avec

$$A = \frac{Q}{r^2} + \frac{1}{r_c^2} \left[1 + \left(\frac{2\pi f_m T}{13,8} \right)^2 \right]^{-1};$$

$$B = \frac{2\pi f_m T}{13,8 r_c^2} \left[1 + \left(\frac{2\pi f_m T}{13,8} \right)^2 \right]^{-1};$$

$$C = \frac{Q}{r^2} + \frac{1}{r_c^2} + Q \cdot 10^{-SNR/10};$$

NOTE La prédiction décrite ci-dessus n'est exacte que pour les emplacements du récepteur situés dans la direction principale de rayonnement du locuteur ou du haut-parleur. Ainsi, il convient d'ajuster la composante de champ direct dans les équations ci-dessus pour prendre en compte toute perte hors axe des haut-parleurs, en raison de leur directivité et le nombre de haut-parleurs contribuant au champ direct.

Il est essentiel d'utiliser le niveau de parole opérationnel pour la prédiction du STI, car cela affecte à la fois les effets du *SNR* réel et ceux du masquage. Un signal de parole à large bande doit être utilisé pour cette prédiction et doit permettre de s'assurer que le canal de transmission a la capacité de produire le niveau opérationnel de pression sonore.

La méthode ci-dessus ne tient pas compte de la différence des temps d'arrivée des signaux multiples de champ direct; elle ne peut pas non plus tenir compte des échos.

Si l'espace acoustique existe, il convient que les temps de réverbération mesurés soient utilisés de préférence aux temps de réverbération prédits.

Annexe M (informative)

Ajustements des résultats de STI et de STIPA mesurés pour la simulation du bruit d'occupation et des différents niveaux de parole

Un exemple d'une méthode adaptée pour ajuster les valeurs mesurées de STI et de STIPA, afin de simuler le bruit d'occupation et différents niveaux de parole, est donné par le Tableau M.1.

Tableau M.1 – Exemple de calcul

1 Acquérir les données de mesure en présence de niveaux de signal et de bruit pendant la mesure.							
	125 Hz	250 Hz	500 Hz	1 000 Hz	2 000 Hz	4 000 Hz	8 000 Hz
Niveau de signal L_{eq}	77,9	77,9	74,2	68,2	62,2	56,2	50,2
Niveau de bruit de fond L_{eq}	48,0	40,0	34,0	30,0	27,0	25,0	23,0
Matrice MTF avec bruit, effets temporels, masquage et facteurs de seuil							
0,63 Hz	0,982	0,952	0,960	0,969	0,979	0,983	0,994
0,80 Hz	0,966	0,928	0,941	0,954	0,969	0,976	0,992
1,00 Hz	0,945	0,897	0,914	0,933	0,955	0,965	0,989
1,25 Hz	0,919	0,862	0,881	0,908	0,939	0,952	0,984
1,60 Hz	0,884	0,819	0,836	0,873	0,915	0,932	0,978
2,00 Hz	0,850	0,784	0,793	0,838	0,890	0,911	0,971
2,50 Hz	0,815	0,750	0,749	0,799	0,862	0,888	0,961
3,15 Hz	0,772	0,715	0,716	0,760	0,832	0,863	0,950
4,00 Hz	0,740	0,678	0,691	0,730	0,800	0,836	0,938
5,00 Hz	0,724	0,623	0,665	0,721	0,772	0,811	0,926
6,30 Hz	0,713	0,553	0,643	0,708	0,745	0,785	0,913
8,00 Hz	0,669	0,515	0,611	0,664	0,720	0,764	0,901
10,00 Hz	0,590	0,479	0,545	0,603	0,693	0,748	0,890
12,50 Hz	0,553	0,442	0,513	0,602	0,678	0,736	0,881
2 Éliminer le bruit de fond, le masquage et les facteurs de seuil							
	125 Hz	250 Hz	500 Hz	1 000 Hz	2 000 Hz	4 000 Hz	8 000 Hz
Rapport signal sur bruit pendant la mesure, dB	29,90	37,90	40,20	38,20	35,20	31,20	27,20
$mk(f)$ pour le bruit uniquement	0,999	1,000	1,000	1,000	1,000	0,999	0,998
Ajustement pour éliminer le bruit de fond	1,001	1,000	1,000	1,000	1,000	1,001	1,002
Niveau combiné de parole et de bruit, dB	77,90	77,90	74,20	68,20	62,20	56,20	50,21
Facteur de masquage auditif amf , dB	non applicable	-20,8	-20,8	-22,7	-25,7	-33,9	-36,9
Pression sonore au carré combinée I_k , MPa ²	61,7	61,7	26,3	6,61	1,66	0,417	0,105
Facteur de masquage auditif $amf \times 1\ 000$	non applicable	8,22	8,22	5,37	2,69	0,407	0,204
$I_{am,k}$	0	508 000	507 000	141 000	17 800	676	85,2
Seuil de réception absolu ART, dB	46	27	12	6,5	7,5	8	12
Intensité – seuil de réception absolu $I_{rt,k}$	40 000	501	15,8	4,5	5,6	6,3	15,8
Ajustement pour éliminer le masquage et le seuil	1,001	1,008	1,019	1,021	1,011	1,002	1,001

Ajustements combinés	1,002	1,008	1,019	1,022	1,011	1,002	1,003
Matrice MTF ajustée sans bruit, masquage et seuil							
0,63 Hz	0,983	0,960	0,978	0,990	0,990	0,986	0,997
0,80 Hz	0,968	0,936	0,959	0,974	0,980	0,979	0,995
1,00 Hz	0,947	0,904	0,931	0,953	0,966	0,968	0,992
1,25 Hz	0,920	0,869	0,898	0,927	0,949	0,955	0,987
1,60 Hz	0,886	0,826	0,852	0,892	0,925	0,935	0,981
2,00 Hz	0,851	0,791	0,808	0,856	0,900	0,914	0,974
2,50 Hz	0,816	0,756	0,764	0,816	0,871	0,891	0,964
3,15 Hz	0,773	0,721	0,730	0,776	0,841	0,866	0,953
4,00 Hz	0,741	0,684	0,705	0,745	0,809	0,838	0,941
5,00 Hz	0,726	0,628	0,678	0,736	0,780	0,812	0,929
6,30 Hz	0,714	0,557	0,656	0,723	0,753	0,786	0,916
8,00 Hz	0,670	0,520	0,623	0,678	0,728	0,765	0,904
10,00 Hz	0,591	0,483	0,556	0,615	0,701	0,749	0,893
12,50 Hz	0,554	0,446	0,523	0,614	0,685	0,737	0,884
3 Ajuster la matrice MTF pour les niveaux opérationnels et les effets de masquage et de seuil							
Niveau de parole opérationnel L_{eq}	82,9	82,9	79,2	73,2	67,2	61,2	55,2
Niveau de bruit de fond opérationnel L_{eq}	55,5	47,5	41,5	37,5	34,5	32,5	30,5
Rapport signal sur bruit	27,40	35,40	37,70	35,70	32,70	28,70	24,70
$mk(f)$ pour le bruit uniquement	0,998	1,000	1,000	1,000	0,999	0,999	0,997
Niveau combiné de parole et de bruit, dB	82,9	82,9	79,2	73,2	67,2	61,2	55,2
Facteur de masquage auditif amf , dB	non applicable	-18,3	-18,3	-20,2	-23,2	-26,2	-34,4
Pression sonore au carré combinée I_k , MPa^2	195	195	83,2	20,9	5,25	1,32	0,332
Facteur de masquage auditif $amf \times 1\ 000$	non applicable	14,6	14,6	9,55	4,79	2,40	0,363
$I_{am,k}$	0	2 850 000	2 850 000	795 000	100 000	12 600	480
Seuil de réception absolu ART, dB	46	27	12	6,5	7,5	8	12
Intensité - seuil de réception absolu $I_{rt,k}$	40 000	500	15,8	4,5	5,6	6,3	15,8
Correction pour le masquage et le seuil	1,000	0,986	0,967	0,963	0,981	0,991	0,999
Ajustements combinés	0,998	0,985	0,967	0,963	0,981	0,989	0,995
Matrice MTF ajustée pour les niveaux opérationnels, le masque et le seuil							
0,63 Hz	0,981	0,946	0,946	0,953	0,971	0,975	0,992
0,80 Hz	0,966	0,922	0,927	0,938	0,961	0,968	0,990
1,00 Hz	0,945	0,891	0,900	0,918	0,947	0,957	0,987
1,25 Hz	0,919	0,856	0,868	0,893	0,931	0,944	0,982
1,60 Hz	0,884	0,814	0,823	0,859	0,907	0,925	0,976
2,00 Hz	0,850	0,779	0,781	0,824	0,882	0,904	0,969
2,50 Hz	0,814	0,745	0,738	0,786	0,855	0,881	0,959
3,15 Hz	0,772	0,710	0,706	0,747	0,825	0,856	0,948
4,00 Hz	0,739	0,674	0,681	0,718	0,793	0,829	0,936
5,00 Hz	0,724	0,619	0,656	0,709	0,765	0,804	0,924
6,30 Hz	0,713	0,549	0,634	0,696	0,739	0,778	0,911
8,00 Hz	0,668	0,512	0,602	0,653	0,714	0,757	0,900

10,00 Hz	0,589	0,476	0,537	0,593	0,687	0,741	0,889
12,50 Hz	0,553	0,439	0,505	0,592	0,672	0,729	0,880
4 Traiter la matrice MTF pour obtenir le STI							
	125 Hz	250 Hz	500 Hz	1 000 Hz	2 000 Hz	4 000 Hz	8 000 Hz
4a Convertir en SNR réels							
0,63 Hz	17,21	12,44	12,42	13,09	15,21	15,93	21,01
0,80 Hz	14,55	10,73	11,04	11,83	13,90	14,83	20,02
1,00 Hz	12,34	9,13	9,56	10,47	12,52	13,50	18,86
1,25 Hz	10,52	7,74	8,17	9,22	11,31	12,30	17,41
1,60 Hz	8,82	6,41	6,69	7,84	9,91	10,88	16,13
2,00 Hz	7,52	5,47	5,52	6,71	8,76	9,73	14,98
2,50 Hz	6,42	4,66	4,51	5,64	7,70	8,69	13,72
3,15 Hz	5,29	3,89	3,80	4,71	6,73	7,75	12,64
4,00 Hz	4,53	3,16	3,30	4,06	5,84	6,87	11,68
5,00 Hz	4,19	2,11	2,79	3,87	5,14	6,12	10,87
6,30 Hz	3,95	0,85	2,38	3,60	4,51	5,44	10,13
8,00 Hz	3,04	0,21	1,80	2,74	3,97	4,94	9,52
10,00 Hz	1,57	-0,42	0,65	1,63	3,42	4,57	9,02
12,50 Hz	0,92	-1,06	0,10	1,61	3,12	4,31	8,64
4b Tronquer $SNR_{k,f}$							
0,63 Hz	15,0	12,4	12,4	13,1	15,0	15,0	15,0
0,80 Hz	14,6	10,7	11,0	11,8	13,9	14,8	15,0
1,00 Hz	12,3	9,1	9,6	10,5	12,5	13,5	15,0
1,25 Hz	10,5	7,7	8,2	9,2	11,3	12,3	15,0
1,60 Hz	8,8	6,4	6,7	7,8	9,9	10,9	15,0
2,00 Hz	7,5	5,5	5,5	6,7	8,8	9,7	15,0
2,50 Hz	6,4	4,7	4,5	5,6	7,7	8,7	13,7
3,15 Hz	5,3	3,9	3,8	4,7	6,7	7,7	12,6
4,00 Hz	4,5	3,2	3,3	4,1	5,8	6,9	11,7
5,00 Hz	4,2	2,1	2,8	3,9	5,1	6,1	10,9
6,30 Hz	4,0	0,9	2,4	3,6	4,5	5,4	10,1
8,00 Hz	3,0	0,2	1,8	2,7	4,0	4,9	9,5
10,00 Hz	1,6	-0,4	0,6	1,6	3,4	4,6	9,0
12,50 Hz	0,9	-1,1	0,1	1,6	3,1	4,3	8,6
4c Convertir en indices de transmission $MTI_{k,f}$							
0,63 Hz	1,00	0,91	0,91	0,94	1,00	1,00	1,00
0,80 Hz	0,99	0,86	0,87	0,89	0,96	0,99	1,00
1,00 Hz	0,91	0,80	0,82	0,85	0,92	0,95	1,00
1,25 Hz	0,85	0,76	0,77	0,81	0,88	0,91	1,00
1,60 Hz	0,79	0,71	0,72	0,76	0,83	0,86	1,00
2,00 Hz	0,75	0,68	0,68	0,72	0,79	0,82	1,00
2,50 Hz	0,71	0,66	0,65	0,69	0,76	0,79	0,96
3,15 Hz	0,68	0,63	0,63	0,66	0,72	0,76	0,92
4,00 Hz	0,65	0,61	0,61	0,64	0,69	0,73	0,89
5,00 Hz	0,64	0,57	0,59	0,63	0,67	0,70	0,86

6,30 Hz	0,63	0,53	0,58	0,62	0,65	0,68	0,84
8,00 Hz	0,60	0,51	0,56	0,59	0,63	0,66	0,82
10,00 Hz	0,55	0,49	0,52	0,55	0,61	0,65	0,80
12,50 Hz	0,53	0,46	0,50	0,55	0,60	0,64	0,79
MTI_k	0,73	0,66	0,67	0,71	0,77	0,80	0,92
Pondération de bandes d'une octave, alpha, hommes	0,085	0,127	0,230	0,233	0,309	0,224	0,173
Somme de alpha* $MTI = MTI_k \times$ pondération alpha	0,062	0,083	0,155	0,165	0,237	0,179	0,159
Pondération de bandes d'une octave, bêta, hommes	0,085	0,078	0,065	0,011	0,047	0,095	none
Somme de bêta* $MTI = MTI_k \times$ pondération bêta	0,059	0,052	0,045	0,008	0,037	0,081	0,000
\sum alpha* MTI	1,040						
\sum bêta* MTI	0,282						
STI	0,76						

Annexe N (informative)

Autres méthodes de détermination de l'intelligibilité de la parole

N.1 Vue d'ensemble

Les métriques de prédiction de l'intelligibilité peuvent être grossièrement divisées en deux catégories: les prédicteurs relativement complexes incluant une modélisation cognitive et perceptive sophistiquée et explicite, et les métriques plus simples qui sont plus faciles à mesurer et à comprendre et qui sont donc accessibles à une plus grande population d'acousticiens. Le STI et le SII s'inscrivent tous deux dans la première catégorie, même si le STI penche vers la facilité d'utilisation, qui est l'avantage de la seconde catégorie, alors que le SII possède plutôt la flexibilité et la rigueur scientifique, qui sont l'avantage de la première catégorie.

Un autre exemple de la première catégorie (modèles perceptifs complexes) est le modèle de sensibilité de la reconnaissance de la parole (Speech Recognition Sensitivity) [41], qui travaille élégamment autour des insuffisances d'autres modèles, mais n'a pas eu beaucoup "d'expérience de terrain" ou d'évaluation indépendante. Des modèles complexes ont aussi été mis au point pour prédire spécifiquement l'intelligibilité et la qualité de la parole pour les canaux de télécommunication (par exemple, le modèle PESQ (Perceptual Evaluation of Speech Quality, évaluation perceptive de la qualité de la parole) [42], [43]). La valeur ajoutée du STI, par rapport à ces modèles, est un domaine de préconisation plus vaste (acoustique de salles et télécommunications), une pratique généralisée et des évaluations par tierces parties. Le fait que divers fournisseurs aient mis en œuvre la méthode STI dans leurs appareils de mesure est, à cet égard, d'un grand secours.

La catégorie des métriques plus simples comprend le "niveau d'interférence sur la parole" (SIL, Speech Interference Level) décrit dans l'ISO 9921; cette métrique prédit l'intelligibilité de la parole dans le bruit en moyennant le rapport parole à bruit dans trois bandes d'une octave. Cette seconde catégorie comprend aussi diverses mesures employant des rapports d'énergie "précoce sur tardive" (échos précurseurs sur échos tardifs) dérivés de réponses impulsionnelles, comme la clarté et la définition [44]. Celles-ci présentent un intérêt spécifique lors des études sur les environnements réverbérants. Pour les conditions et les types d'applications que visent ces mesures, leur niveau d'exactitude peut approcher celui du STI. Dans des situations plus complexes, l'exactitude du STI surclasse toutes les métriques plus simples.

N.2 Essais avec des mots

Les limitations des essais avec des mots sont données dans l'ISO/TR 4870 [45]. Comme la méthode utilise la reconnaissance de mots par des auditeurs, il n'existe pas de contrainte quant aux caractéristiques du système électroacoustique ou de l'environnement. En cas d'utilisation en présence de distorsions temporelles (réverbération, échos, commande automatique du gain), les mots doivent être insérés dans une phrase porteuse.

N.3 Essais de rimes modifiées

Les limitations sont similaires à celles données dans l'ISO/TR 4870 [45]. Comme la méthode utilise la reconnaissance de mots par des auditeurs, il n'existe pas de contrainte quant aux caractéristiques du système électroacoustique ou de l'environnement.

N.4 Indice d'intelligibilité de la parole (SII)

Le SII est souvent préféré par ceux qu'intéresse la comparaison des effets de différents signaux de parole plus que la comparaison de différents canaux de transmission. Cependant, contrairement au STI, le SII ne peut pas être mesuré directement, mais doit être calculé. Il est couramment utilisé par les audiologistes expérimentateurs en raison de sa meilleure résolution en fréquence et de sa sensibilité à la baisse d'intelligibilité aux forts efforts vocaux.

NOTE Le SII apparaît ne pas être considérablement plus sensible que le STI aux effets de la réponse en fréquence [12].

N.5 Perte d'articulation des consonnes (%ALcons)

Les limitations sont similaires à celles données dans l'ISO/TR 4870 [45]. Il convient que le mode opératoire de mesure n'inclue pas les voyelles. Il peut en résulter une erreur systématique sur les essais de mots [46]. Comme la méthode utilise la reconnaissance de mots par des auditeurs, il n'existe pas de contrainte quant aux caractéristiques du système électroacoustique ou de l'environnement. Il convient de noter que le %ALcons ne peut normalement pas être mesuré exactement par des moyens acoustiques. Il n'existe aucune manière exacte de mesurer le %ALcons par des moyens électroacoustiques.

N.6 PESQ

L'évaluation perceptive de la qualité de la parole (PESQ) est particulièrement utile dans les situations d'intelligibilité élevée. Elle emploie des scores (ou notes) moyens d'opinion et est spécialement adaptée à la mesure d'une transmission de haute qualité, lorsque l'intelligibilité de la parole est un problème moindre [42], [43].

Bibliographie

- [1] Mapp, P, *New Techniques for Measuring Speech Privacy & Efficiency of Sound Masking Systems. AES 122nd Convention, Vienna, May 2007*
- [2] CEI 60318-1:2009, *Electroacoustique – Simulateurs de tête et d'oreille humaine – Partie 1: Simulateur d'oreille pour la mesure des écouteurs supra-auraux et circumauraux*
- [3] CEI 61672 (toutes les parties), *Electroacoustique – Sonomètres*
- [4] ANSI (1969). ANSI S3.5-1969, *American National Standard, "Methods for Calculation of the Articulation Index"* (American National Standards Institute New York.)
- [5] Steeneken, H.J.M. and Houtgast, T., "Some applications of the Speech Transmission Index (STI) in auditoria", *Acustica* 51, 1982, p.229-234
- [6] Steeneken, H.J.M. and Houtgast, T., "A physical method for measuring speech transmission quality", *J. Acoust. Soc. Amer.* 67, 1980, 31, p.318-326
- [7] Houtgast, T. and Steeneken, H.J.M., "The modulation transfer function in room acoustics as a predictor of speech intelligibility", *Acustica* 28, 1973, p.66-73
- [8] Steinbrecher, T, "Speech Transmission Index: Too weak in time and frequency?", *Proc IOA Vol 30 Part 6 2008*
- [9] Mapp, P, "Is STIPA a robust measure of speech intelligibility performance," *AES 118th Convention, Barcelona, 2005*
- [10] Mapp, P, *Some Effects of Equalisation on Sound System Intelligibility and Measurement AES 115th Convention 2003 Preprint*
- [11] Leembruggen, G and A Stacey "Should the Matrix be reloaded *Proc IOA Vol 28 Part 6 2006*
- [12] Leembruggen, G. "Is SII better than STI at recognising the effects of poor tonal balance on intelligibility?" *Proc IOA Vol 28 Part 6 2006*
- [13] Mapp P, *Practical application of STI to assessing Public Address and Emergency Sound Systems. Past, present and future of the Speech Transmission Index, TNO 2002*
- [14] Festen, JM & Plomp, R, *Effects of fluctuating noise and interfering speech on the speech reception threshold for impaired and normal hearing, J. Acoust. Soc. Amer., 88(4) 1990*
- [15] Duquesnoy, A.J.H/M and Plomp, R. (1980) "Effect of reverberation and noise on the intelligibility of sentences in case of presbycusis". *J. Acoust. Soc. Am.* 68, p.537-544
- [16] Mapp, P, "Assessing the Potential Intelligibility of Assistive Audio Systems for the Hard of Hearing and Other Users", *AES 124th convention, Amsterdam, 2008*
- [17] Schroeder, M, *Modulation Transfer Functions: Definition and Measurement, Acustica, 49, 1981*
- [18] Houtgast, T. and Steeneken, H.J.M., "The modulation transfer function in room acoustics as a predictor of speech intelligibility", *Acustica* 28, 1973, p.66-73

- [19] Houtgast, T. and Steeneken, H.J.M., "A Multi-lingual evaluation of the Rasti-method for estimating speech intelligibility in auditoria", *Acustica* 54, 1984, p.185-199
- [20] Steeneken, H.J.M. and Houtgast, T., "The temporal envelope spectrum of speech and its significance in room acoustics", *Proc. 11th International Congress on Acoustics, Paris, 1983, Vol. 7, p.85-88*
- [21] Steeneken, H.J.M., Verhave, J.A., McManus, S., and Jacob, K.D., (2001) "Development of an Accurate, Handheld, Simple-to-use Meter for the Prediction of Speech Intelligibility", *Proceedings IoA 2001, Reproduced sound (17)*. Stratford-upon-Avon, UK
- [22] Mapp, P. *Further thoughts on Speech Transmission Index (STI)*. IOA Reproduced Sound 18. Proc. IOA Vol 24 Pt 8. 2002
- [23] Bozzoli, F, Farina, A, "Influence of artificial mouth's directivity in determining Speech Transmission Index", *AES 119th Convention, New York, 2005*
- [24] CEI 60268-3:2000, *Equipements pour systèmes électroacoustiques – Partie 3: Amplificateurs*
- [25] Rife, D., "Modulation Transfer Function Measurements with Maximum-Length Sequences", *J. Audio Eng. Soc.*, Oct. 1992, Vol. 40, No. 10
- [26] ANSI standard S12.2 1995, *Criteria for evaluation of room noise*
- [27] Wijngaarden, S, Drullman, R, "Binaural intelligibility prediction based on the speech transmission index" *J. Acoust. Soc. Amer.* 123, 2008, p.4514-4523
- [28] Houtgast, T, Steeneken, H and Plomp, R. "Predicting Speech Intelligibility in Rooms from the Modulation Transfer Function. i. General Room Acoustics" *Acustica Vol 46 1980*
- [29] Steeneken, H.J.M. and Houtgast, T., "Mutual dependency of the octave-band weights in predicting speech intelligibility," *Speech Communication* 28, 1999, p.109-123
- [30] Zwicker, E. and Feltkeller, "The ear as an information receiver", (ASA)
- [31] Wijngaarden, S.J. van and Steeneken, H.J.M. (1999). "Objective prediction of speech intelligibility at high ambient noise levels using the speech transmission index." *In Eurospeech99 – Proceedings of the 6th European Conference on Speech Communication and Technology, Budapest, Vol 6, p.2639-2642*
- [32] Steeneken, H.J.M. and Houtgast, T., "Validation of the STIr method with the revised model," *Speech Communication* 38, 2002, p.413-425
- [33] Anderson, B.W., and Kalb, J.T. 1987. "English verification of the STI method for estimating speech intelligibility of a communications channel," *J. Acoust. Soc. Am.* 81, 1982-1985
- [34] Hiroshi Sato, John S. Bradley, and Masayuki Morimoto, *Using listening difficulty ratings of conditions for speech communication in rooms* , *J. Acoust. Soc. Am.* 117, 1157 (2005)
- [35] Hayato Sato, Masayuki Morimoto, Hiroshi Sato, and Megumi Wada, *Relationship between listening difficulty and acoustical objective measures in reverberant sound fields*, *J. Acoust. Soc. Am.* 123, 2087-2093 (2008)

- [36] Wijngaarden, S.J., Steeneken, H.J.M. and Houtgast, T. (2002) "Quantifying the intelligibility of speech in noise for non-native listeners," *J. Acoust. Soc Am.* 112, p.3004-3013
- [37] ISO 7029: 2000, *Acoustique – Distribution statistique des seuils d'audition en fonction de l'âge*
- [38] Sato,H, Kurakata,K, Mizunami,T, "Accessible speech messages for the elderly in rooms" *9th Western Pacific Acoustics Conference Seoul, Korea, 2006*
- [39] ISO/TR 22411:2008, *Données d'ergonomie et lignes directrices pour l'application du Guide ISO/CEI 71 aux produits et services, afin de répondre aux besoins des personnes âgées et de celles ayant des incapacités*
- [40] Steeneken, H.J.M. and Houtgast, T (1986), *Comparison of some methods for measuring speech levels Report IZF1986-20 TNO Institute for Perception Soesterberg Netherlands*
- [41] H. Müsch and S. Buus. 'Using statistical decision theory to predict speech intelligibility. I. Model structure'. *J.Acoust.Soc.Am.* 109, 2896–2909. (2001)
- [42] Recommandation UIT-T P.862, "Évaluation de la qualité vocale perçue (PESQ)": *méthode objective d'évaluation de la qualité vocale de bout en bout des codecs vocaux et des réseaux téléphoniques à bande étroite. (Union Internationale des Télécommunications, Genève Suisse, février 2001)*
- [43] J. G. Beerends, A. P. Hekstra, A. W. Rix and M. P. Hollier. 'PESQ, the new ITU standard for objective measurement of perceived speech quality, Part II - Perceptual model'. *J. Audio Eng. Soc.*, vol. 50, pp. 765-778. (2002 Oct.)
- [44] ISO 3382-1:2009, *Acoustique – Mesurage des paramètres acoustiques des salles – Partie 1: Salles de spectacles*
- [45] ISO/TR 4870:1991, *Acoustique – Élaboration et étalonnage des tests d'intelligibilité de parole (Annulée)*
- [46] Steeneken, H.J.M. and Houtgast, T., "Phoneme-group specific octave-band weights in predicting speech intelligibility," *Speech Communication* 38, 2002, p.399-411
- [47] ITU-T P.50, *Qualité de transmission téléphonique, installations téléphoniques et réseaux locaux – Appareils de mesures objectives – Voix artificielles*
- [48] ITU-T P.51, *Qualité de transmission téléphonique, installations téléphoniques et réseaux locaux – Appareils de mesures objectives – Bouche artificielle*
- [49] ISO 9921:2003, *Ergonomie – Evaluation de la communication parlée*
-

INTERNATIONAL
ELECTROTECHNICAL
COMMISSION

3, rue de Varembé
PO Box 131
CH-1211 Geneva 20
Switzerland

Tel: + 41 22 919 02 11
Fax: + 41 22 919 03 00
info@iec.ch
www.iec.ch