

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**



THIS PUBLICATION IS COPYRIGHT PROTECTED

Copyright © 2020 IEC, Geneva, Switzerland

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

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

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

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

About the IEC

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

About IEC publications

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

IEC publications search - webstore.iec.ch/advsearchform

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 It also gives information on projects, replaced and withdrawn publications.

IEC Just Published - webstore.iec.ch/justpublished

Stay up to date on all new IEC publications. Just Published details all new publications released. Available online and once a month by email.

IEC Customer Service Centre - webstore.iec.ch/csc

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

Electropedia - www.electropedia.org

The world's leading online dictionary on electrotechnology, containing more than 22 000 terminological entries in English and French, with equivalent terms in 16 additional languages. Also known as the International Electrotechnical Vocabulary (IEV) online.

IEC Glossary - std.iec.ch/glossary

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

A propos de l'IEC

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

A propos des publications IEC

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

Recherche de publications IEC -

webstore.iec.ch/advsearchform

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

IEC Just Published - webstore.iec.ch/justpublished

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

Service Clients - webstore.iec.ch/csc

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

Electropedia - www.electropedia.org

Le premier dictionnaire d'électrotechnologie en ligne au monde, avec plus de 22 000 articles terminologiques en anglais et en français, ainsi que les termes équivalents dans 16 langues additionnelles. Egalement appelé Vocabulaire Electrotechnique International (IEV) en ligne.

Glossaire IEC - std.iec.ch/glossary

67 000 entrées terminologiques électrotechniques, en anglais et en français, extraites des articles Termes et Définitions des publications IEC parues depuis 2002. Plus certaines entrées antérieures extraites des publications des CE 37, 77, 86 et CISPR de l'IEC.

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

ICS 33.160.01

ISBN 978-2-8322-8862-7

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

CONTENTS

FOREWORD	7
INTRODUCTION	9
1 Scope	12
2 Normative references	12
3 Terms and definitions	12
4 Description of the STI model	18
4.1 Overview	18
4.2 Applicability of the STI model	19
4.3 Theoretical details	19
4.3.1 Envelope function and envelope spectrum	19
4.3.2 Reduction of modulation	20
4.3.3 Role of the octave-band noise carriers	20
4.3.4 Theoretical overview	20
4.4 Measurement of STI	22
4.4.1 Direct and indirect methods	22
4.4.2 Full STI	23
4.4.3 STIPA	23
4.4.4 Choice of method	23
5 Direct method of measuring STI – User guidance	25
5.1 Overview	25
5.2 STIPA	26
5.3 Application	26
5.4 Limitations	26
6 Indirect method of measuring STI (impulse response) – User guidance	27
6.1 Overview	27
6.2 Application	27
6.3 Limitations (non-linear distortion)	28
7 Limitations of the STI model	29
7.1 General	29
7.2 Frequency shifts	30
7.3 Centre clipping	30
7.4 Dropouts	30
7.5 Jitter	30
7.6 Digital voice compression systems	30
7.7 Overestimation of STI under low background noise conditions	31
7.8 Frequency response	31
7.9 Echoes	32
7.10 Fast amplitude compression and expansion	34
7.11 Non-linear distortion	35
7.12 Hearing impaired listeners	35
7.13 Impulsive and fluctuating noise	35
7.14 Conclusion	35
8 Measurement procedures, post-processing of data and applications	36
8.1 General	36
8.2 Acoustical input	36

8.3	Acoustical output	38
8.4	Electrical input	38
8.5	Electrical output	38
8.6	Examples of input/output combinations	38
8.6.1	Acoustical input – Acoustical output	38
8.6.2	Electrical input – Electrical output (e.g. assessment of wired and wireless) communication systems)	38
8.6.3	Acoustical input – Electrical output (e.g. assessment of microphones)	38
8.6.4	Electrical input – Acoustical output (e.g. assessment of PA systems)	39
8.7	Spatial averaging of STI measurements	39
8.8	Post-processing of measured MTF data	40
8.9	Issues concerning noise	40
8.9.1	General	40
8.9.2	Measurement of background noise	41
8.9.3	Fluctuating noise	41
8.10	Analysis and interpretation of the results	41
8.11	Binaural STI measurements	42
9	Use of the STI as a design prediction tool	42
9.1	Overview	42
9.2	Statistical predictions	43
9.3	Prediction from simulated impulse response	43
Annex A (informative)	The basis of the STI concept	44
A.1	Introduction to this annex	44
A.1.1	Purpose	44
A.1.2	Modulation transfer function (MTF)	44
A.1.3	STI model	45
A.1.4	STI modulation frequencies	46
A.2	Calculation of the STI	46
A.2.1	General equation for STI	46
A.2.2	Gender-specific octave band weighting and redundancy factors	47
A.2.3	Adjustment of the MTF for ambient noise	48
A.2.4	Adjustment of the MTF for auditory masking and threshold effects	48
A.3	Calculation of the modulation transfer ratio values	49
A.3.1	Direct method: Analysis of the STI test signal	49
A.3.2	Indirect method: Determination of the modulation transfer function (MTF)	50
A.4	Auditory effects on the STI	51
A.4.1	Overview	51
A.4.2	Level-dependent auditory masking	51
A.4.3	Absolute speech reception threshold	54
A.5	Generation of the STI test signal (direct method)	54
A.5.1	Pink noise source signal	54
A.5.2	Generating octave band carrier signals	54
A.5.3	Intensity modulation of the carrier signals	55
A.5.4	Applying the speech spectrum to the STI test signal	55
A.6	Spectrum of STI test signal	55
A.6.1	Standardized speech spectrum	55
A.6.2	Speech-shaped noise	55
Annex B (normative)	STIPA method	57

B.1	Overview.....	57
B.2	Test signal	57
Annex C	(normative) Verification of STI measuring devices	59
C.1	Specification of the measuring device	59
C.2	Signals for testing STI implementations	59
C.3	Testing the dynamic range in the modulation domain	59
C.3.1	General	59
C.3.2	Modulation depth testing for STIPA direct method	59
C.3.3	Modulation depth testing for STI indirect method	60
C.4	Testing of cross-talk between octave-band filters	61
C.4.1	Flank attenuation slopes	61
C.4.2	Octave band filter testing – STIPA direct method	61
C.4.3	Performance verification files.....	62
Annex D	(informative) Use of STI measuring devices	63
D.1	Overview.....	63
D.2	STIPA characterises only the speech transmission channel	63
D.3	Examples of test scenarios for STIPA tests	64
D.4	Equipment and resources needed for a STIPA test	67
D.4.1	Availability of the test signal	67
D.4.2	A source of the STIPA test signal	67
D.4.3	A STIPA analyser	67
D.5	Steps in the overall procedure	67
Annex E	(informative) Qualification of the STI and relationships with other speech intelligibility measures.....	68
E.1	Relationship between the STI and word/sentence scores	68
E.2	Relationship between STI and listening difficulty	68
Annex F	(informative) Nominal qualification bands for STI	70
Annex G	(informative) Examples of STI qualification bands and typical applications	71
Annex H	(informative) Non-native listeners	72
Annex I	(informative) Effect of age-related hearing loss and hearing impairment on speech intelligibility.....	73
Annex J	(normative) Setting and adjustment of STI test signal level.....	74
J.1	Overview.....	74
J.2	The concept of 'speech level' and the method of measurement	74
J.3	Real speech level	74
J.4	Corrected speech level derived from real speech level.....	75
J.5	Comparison of dynamic structures of speech and test signals.....	75
Annex K	(informative) Example test report sheet for STI measurements	77
Annex L	(normative) Prediction of the STI using statistical methods.....	79
Annex M	(informative) Adjustments to STI data to simulate alternative ambient noise spectra and different speech levels	81
Annex N	(informative) Other methods of determining speech intelligibility	91
N.1	Overview.....	91
N.2	Word tests	91
N.3	Modified rhyme tests	91
N.4	Speech intelligibility index (SII)	92
N.5	PESQ	92
Annex O	(informative) Alternative direct methods for measuring Full STI	

Annex P (normative) Information to be provided by manufacturers.....	94
P.1 Purpose of this annex	94
P.2 Form in which the information is to be provided	94
P.3 Required information.....	94
P.4 Declaration	94
Annex Q (informative) Effect of uncertainties of selected parameters on STI uncertainty.....	95
Q.1 STI calculation framework.....	95
Q.1.1 Overview	95
Q.1.2 Statistical MTF.....	95
Q.1.3 Corrections.....	95
Q.1.4 Effective SNR	96
Q.1.5 Modulation transfer index (MTI)	96
Q.1.6 Speech transmission index (STI)	96
Q.2 The effect of RT uncertainty on STI uncertainty	97
Q.2.1 General	97
Q.2.2 Modulation transfer function.....	97
Q.2.3 Uncertainty in the STI	97
Q.2.4 Conclusions:.....	99
Q.3 The effect of S/N uncertainty on STI uncertainty	99
Q.3.1 General	99
Q.3.2 Ideal transfer function	99
Q.3.3 Reverberation.....	100
Q.3.4 Conclusions:.....	101
Q.4 The effect of signal level uncertainty on STI uncertainty.....	101
Q.4.1 Overview	101
Q.4.2 Auditory masking	101
Q.4.3 Conclusions.....	103
Bibliography.....	104
Figure 1 – Envelope function (panel A) of a 10 s speech signal for the 250 Hz octave band and corresponding envelope spectrum (panel B).....	20
Figure 2 – Modulation transfer function – Input/output comparison.....	21
Figure 3 – Effect of a single delayed arrival on the MTF (idealised conditions).....	33
Figure 4 – Idealised STI (Male speech Spectrum) versus delay and level of secondary arrival	34
Figure A.1 – Theoretical expression of the MTF	44
Figure A.2 – Measurement system and frequencies for the STI method	46
Figure A.3 – Auditory masking of octave band ($k-1$) on octave band (k).....	52
Figure A.4 – Relationship between STI and speech level for different reverberation times.....	53
Figure D.1 – Schematic representation of the definition of a speech transmission channel.....	64
Figure E.1 – Relationships between some speech intelligibility measures	68
Figure E.2 – Relationship between STI, speech intelligibility scores and listening difficulty ratings [43], [44].....	69
Figure F.1 – STI qualification bands.....	70

Figure Q.1 – Uncertainty in absolute value of STI vs reverberation time RT with various degrees of uncertainty in RT	99
Figure Q.2 – Uncertainty in absolute value of STI vs reverberation time RT with 1 dB uncertainty in SNR at various SNRs	101
Figure Q.3 – Uncertainty in absolute value of STI versus reverberation time RT with various degrees of masking.	103
Table 1 – How to use this document	10
Table 2 – Comparison of direct and indirect methods	22
Table 3 – Suitability of STI test methods for different types of distortion	24
Table 4 – Test-method suitability	24
Table 5 – Measurement applications	25
Table A.1 – MTI octave band weighting factors	48
Table A.2 – Auditory masking as a function of the octave band level.....	53
Table A.3 – Absolute speech reception threshold level in octave bands	54
Table A.4 – Octave band levels (dB) relative to the A-weighted speech level	55
Table A.5 – Filter parameters and s-plane polynomials that produce speech-shaped pink noise.	56
Table B.1 – Modulation frequencies for the STIPA method.....	57
Table C.1 – Specification of an STI measuring device.....	59
Table D.1 – Scenario 1, PA with "live" announcer	65
Table D.2 – Scenario 2, PA with pre-recorded announcements	65
Table D.3 – Scenario 3, "live" meetings and conversations	66
Table D.4 – Scenario 4, lecture.....	66
Table E.1 – Categories for listening difficulty	69
Table G.1 – Examples between STI qualification bands and typical applications	71
Table H.1 – Adjusted intelligibility qualification tables for non-native listeners.....	72
Table I.1 – Adjusted intelligibility qualification tables for normal listeners and people over 60 years old with hearing loss	73
Table J.1 – Typical speech and test signal dynamics	75
Table J.2 – Comparison of speech and the test signal	76
Table K.1 – Example test report sheet	77
Table K.2 – Measurement data record sheet.....	78
Table M.1 – Flow chart of post-processing adjustment steps.....	82
Table M.2 – Example calculation.....	87

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: Audio, video and multimedia equipment and systems.

This fifth edition cancels and replaces the fourth edition published in 2011. This edition constitutes a technical revision.

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

- a) the spectrum of the male speech test signal has been changed, with significant reductions in the 125 Hz and 250 Hz bands being implemented;
- b) some corrections to formulae have been made;
- c) additional information has been included on prediction and measurement procedures;
- d) spectrum and weighting factors for female speech have been removed;
- e) verification information for STI measurement devices added;
- f) the relationships between STI and number of other speech intelligibility measures have been updated in Annex E;

- g) greater information is given in Annex M about adjustments to the measured STI results to simulate effects of alternative ambient noise and speech levels.

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

The text of this International Standard is based on the following documents:

CDV	Report on voting
100/3202/CDV	100/3422/RVC

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

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

A list of all 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 document will remain unchanged until the stability date indicated on the IEC website under "<http://webstore.iec.ch>" in the data related to the specific document. At this date, the document will be

- reconfirmed,
- withdrawn,
- replaced by a revised edition, or
- amended.

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

INTRODUCTION

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.

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.

To avoid misinterpretation of STI results, it is important that all users of the STI understand the basic principles behind the operation of the STI, the application domain and the limitations. This document provides substantial information to assist users.

Potential applications of the STI

The STI can 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 emergency sound and communication systems;
- measurement of communication channels and 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 was not designed for the measurement and evaluation of speech privacy or speech masking systems and, therefore, has not been validated for these situations. It is not recommended to use the STI below 0,3, but if this is to be undertaken, specialist expertise and techniques beyond the scope of this standard are required.

Potential users of STI

The range of users of STI measurements is diverse. Among the users who might 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;
- acoustic and electroacoustic consultants;
- sound system installers;
- researchers into STI methods and developers of instruments to measure the STI.

Table 1 summarises which sections of the document may apply to different users and applications.

Table 1 – How to use this document

Purpose	Topic	Clauses
All users	Introduction to the STI method	
Routine check of voice-alarm or sound system with STIPA	Direct method of measuring STI	4
In-depth check of or to certify sound system with STIPA and/or impulse response methods	Description of the STI method	5
	Direct method of measuring STI	4 and 5
	Indirect method of measuring STI using the impulse response	4 and 6
	Measurement procedures, and applications	8
	Post-processing of measured MTF data	8.8
	Limitations of the measurement methods	5.4, 6.3
	Optional: Theory and equations governing STI methods	Annex A and Annex B
	Optional: Relationship between subjective and objective measures of intelligibility	Annex F
	Optional: Measurement uncertainties	Annex Q
Measure telecommunication equipment	Direct method only	8.6.2
Manufacturer of STIPA device	Theory and equations governing STI methods	Annex A and Annex B
	Verification of STI measurement device performance	Annex C
	Information to be provided	Annex D
Manufacturer of acoustical analyser and simulation software	Theory and equations governing STI methods	Annex A
	Calibration of STI instruments	Annex C
	Information to be provided	Annex P
Research into intelligibility	Theory and equations governing STI methods	Annex A and Annex B
Using simulation software	Prediction methods	Annex M
Post processing of STI and STIPA measurement	Post processing measurement results	Annex M
	Optional – As per in-depth measurements of STI listed above	
	Optional -Worked calculation example	Annex M
Evaluation of the potential intelligibility of Assistive Listening Systems	As per in-depth measurements of STI listed above	
	Special process for Assistive Listening Systems	8.6.3

Revision history

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: 2011.
 - The terms STI_r and Room Acoustic Speech Transmission Index (RASTI) were discontinued.
 - A new function for the prediction of auditory masking effects was introduced.
 - STI corrections for non-native language listeners and some forms of hearing loss were introduced.

SOUND SYSTEM EQUIPMENT –

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

1 Scope

This part of IEC 60268 defines the STI model, test signals, measurement and prediction methods.

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

This document does not provide STI criteria for certification of transmission channels (e.g. criteria for a voice-alarm system), but some typical application values are provided in Annex G.

Every measurement method has limitations, and the reader is referred to clauses relating to limitations such as speech privacy, echo and systems using digital voice compression (vocoders).

This document does not cover the case of fluctuating noise on the STI, although some general comments on dealing with this complex issue are provided in 7.13 and 8.9.3.

2 Normative references

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

IEC 61260-1:2014, *Electroacoustics – Octave-band and fractional-octave-band filters – Part 1: Specifications*

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply. ISO and IEC maintain terminological databases for use in standardization at the following addresses:

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

3.1

speech intelligibility

rating of the proportion of speech that is understood

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

Note 1 to entry: This note applies to the French language only.

3.4**speech intelligibility index****SII**

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

Note 1 to entry: This note applies to the French language only.

3.5**STI model**

framework for quantifying the potential effect that a transmission path between a talker and listener has on speech intelligibility

Note 1 to entry: The model predicts the speech intelligibility based on the degree to which the intensity modulations of speech are preserved during transmission.

3.6**Full STI**

model for prediction and measurement of the speech transmission index that uses 14 modulation frequencies in each of the 7 octave bands

3.7**distortion**

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

Note 1 to entry: Distortion can include both linear and non-linear effects in both the frequency and time domains.

3.8**speech transmission index for public address systems****STIPA**

model using a condensed version of the Full STI that uses only 2 modulation frequencies in each of 7 octave bands

Note 1 to entry: This note applies to the French language only.

3.9**direct method**

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

3.10**indirect method**

method using the impulse response to derive the modulation transfer function by applying Schroeder's equation

3.11**speech transmission channel**

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

3.12

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

Note 1 to entry: This note applies to the French language only.

3.13

emergency sound and communication system

sound distribution or communication system that can broadcast speech in an emergency

3.14

real speech level

sound pressure level of the broadcast speech signal that is measured and is used to derive the corrected speech level, typically by adding 3 dB

Note 1 to entry: The difference between the real speech level and the corrected speech level represents the reduction in the long-term level produced by the pauses and silences between words.

3.15

corrected speech level

long-term speech level in dBA where only the segments that contribute to the speech signal are taken into account and pauses and silences between words and sentences are ignored, as defined in Annex J

3.16

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

3.17

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 1 to entry: The degree of similarity required cannot be easily specified and depends on the particular application. See for example ITU-T P.51 [1].

3.18

talkbox

loudspeaker mounted in an enclosure designed to exhibit directivity and radiation patterns similar to those of the average human head and produce a calibrated frequency response for reproduced test signals

3.19

non-native speaker

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

3.20

absolute speech reception threshold

threshold of hearing increased by the minimum required dynamic range to enable 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 1 to entry: Within the STI model, 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 1 to entry: See IEC 60318-1 [2].

3.23

intensity function

squared amplitude signal as a function of time

3.24

envelope function

temporal fluctuations of the intensity of a speech signal within a certain frequency band that has been low-pass filtered at approximately 50 Hz to remove the fine structure of the carrier waveform

3.25

envelope spectrum

spectral components of the envelope function

3.26

modulation frequency

F_e

frequency of the sinusoidal variation of the envelope function generally lying in the range 0,1 Hz to 30 Hz

3.27

specific modulation frequency

f_m

specific frequency of the sinusoidal variation of the envelope function which lies in the range of 0,63 Hz to 12,5 Hz

Note 1 to entry: The subscript variable m , the modulation frequency index, is not the same variable as in IEC 61260-1, the phrase ' m -value' (see 3.29) nor the modulation depth symbols m_i and m_o used in 4.3.4.

Note 2 to entry: m takes the values 1 to 14. See Figure A.1 and A.2.1.

3.28

modulation depth

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

3.29

modulation transfer ratio

m -value

ratio between the modulation depths of the intensity functions of the received and original (transmitted) signals

3.30

modulation transfer function

MTF

modulation transfer ratio as a function of the modulation frequency

Note 1 to entry: This note applies to the French language only.

3.31
modulation transfer index

MTI

$m(f_m)$

unweighted mean of the transmission indices over all modulation frequencies within a given octave band

Note 1 to entry: This note applies to the French language only.

3.32
transmission index

TI

T

effective signal-to-noise ratios scaled to a value between 0 and 1

Note 1 to entry: This note applies to the French language only.

3.33
octave band weighting factor

α

relative contribution in each octave band to the speech transmission index

3.34
octave band redundancy factor

β

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

3.35
background noise

sounds comprising stationary, fluctuating and impulsive noise remaining in the absence of the speech or test signals

3.36
stationary noise

continuous noise with an approximately constant level

Note 1 to entry: This level is used for predictions and post-processing of measurements.

3.37
impulsive noise

sound or noise characterized by short individual bursts of sound pressure

3.38
fluctuating noise

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

3.39
signal-to-noise ratio
SNR

ρ

difference in dB 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 1 to entry: This note applies to the French language only.

3.40

effective signal-to-noise ratio
SNR_{eff}

ρ_{eff}

modulation transfer function transformed into the signal-to-noise ratio domain, expressed in dB

3.41

crest factor

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

3.42

Lombard effect

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

Note 1 to entry: Voice pitch shift at higher talking levels is not accounted for here.

3.43

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

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 1 to entry: The ratio of the cut-off frequencies is $f_2 / f_1 = 2^{1/b}$, with $1/b$ denoting the fraction of an octave.

Note 2 to entry: Filters derived using the more commonly-employed base 10 can also be used.

Note 3 to entry: For further information, refer to IEC 61260-1.

3.44

reference sound pressure

p_0

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

3.45

sound pressure level

L_p

twenty times the logarithm to the base ten of the ratio of RMS sound pressure to the reference sound pressure, expressed in dB

3.46

equivalent continuous sound pressure level

$L_{\text{eq},T}$

ten times 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 to entry: The sound pressure level $L_{\text{eq},T}$ is given by the following equation:

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

with

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

t the integration variable for time;

$T = t_2 - t_1$, 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 to entry: As a matter of principle, no time-weighting is applied in the determination of the continuous sound pressure level.

Note 3 to entry: For further definitions, see IEC 61672 [3].

4 Description of the STI model

4.1 Overview

The STI model is an objective and validated framework for evaluating speech transmission quality for communication channels that may be affected by a wide range of acoustic and electro-acoustical distortions that affect speech intelligibility.

The model was developed as a fast and objective test method for determining the quality of speech transmission provided by a speech transmission channel or system. Using the speech transmission index, the potential speech intelligibility can be predicted for different types of word and sentence formats for a wide range of conditions within speech transmission systems. Such conditions include reverberation and ambient noise.

The STI model represents an idealised situation in which a talker with the standardised male speech spectrum is speaking with good articulation (clear speech) at a nominal word rate of 3 to 4 syllables per second and assumes listeners have normal hearing. Corrections may be applied for non-native speakers/listeners and for listeners with hearing loss, as indicated in Annex H and Annex I respectively.

A speech signal level varies rapidly with time producing variations (or fluctuations) in the intensity envelope of the sound. Slower fluctuations of this intensity envelope correspond with word and sentence boundaries, while faster fluctuations coincide with individual phonemes within words. Phonemes are the fundamental elements of speech and connected discourse can be considered as a sequence of phonemes.

The STI concept is based on the empirical finding that these fluctuations carry the most relevant information relating to speech intelligibility, and preservation of the intensity envelope is considered to be of the utmost importance see [4], [5] and [6]. Time-domain distortions within a transmission channel (such as reverberation, echoes and automatic gain control) along with noise can degrade the fluctuating speech-signal and reduce the intelligibility. The extent of degradation in the fluctuations determines the potential speech intelligibility and the STI model measures the degree to which the fluctuations are preserved.

The STI model has been optimised and validated with subject-based intelligibility experiments using CVC (consonant-vowel-consonant) (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.

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

Research [8] 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 modulations in neighbouring octave bands can partly compensate for this missing contribution. This insight led to the use of redundancy factors. Equations used in the STI model and more technical details are presented in Annex A.

However, as the STI is a simplification of the human speech communication process, the STI model can be limited in its applicability. Users that apply the model beyond its current limits might obtain inaccurate intelligibility predictions. Accordingly, an overview of the applications and limitations is given to help users decide which method is most suitable for their application, so that meaningful and accurate results can be obtained.

In contrast to the approach of the Articulation Index [7], 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 or fluctuations of the speech signal.

NOTE A comparison of the STI methods with other methods of assessing speech intelligibility is given in Annex E.

4.2 Applicability of the STI model

The STI model is monaural and was validated using acoustic measurements made in the acoustic free-field with an omnidirectional microphone. The use of a directional microphone for measurement produces different and uncorrelatable results and is not normally advised. Further information is given in 7.10.

If the situation or the transmission channel does not allow the use of STI models, alternative techniques for assessing intelligibility shall be used. Other methods exist to assess the quality of speech communication, and as each has advantages and disadvantages. Annex N describes other measures of intelligibility.

4.3 Theoretical details

4.3.1 Envelope function and envelope spectrum

The fluctuations in speech intensity are termed modulations and can be quantified as a function of modulation frequency F producing the modulation spectrum. For well-articulated (clear) speech, the modulation frequencies typically extend from 0,5 Hz up to 16 Hz, with maximum modulation occurring at approximately 3 Hz.

Each phoneme is characterized by a specific frequency spectrum in the intensity envelope, and the shape of the envelope is unique for a specific sequence of phonemes. To achieve speech clarity, these spectral differences of the phonemes shall be preserved. Degradation of the speech envelope, such as by noise or reverberation, results in a reduction in the degree of fluctuations of the envelope and this is reflected by a reduction of the spectral differences between phonemes.

Figure 1, panel A, shows an example envelope for the 250 Hz octave frequency band with fluctuating intensity being clearly visible. The spectrum of the envelope provides a description of the envelope fluctuations and is obtained from a spectral analysis of the envelope in one-third octave-bands. Typically, a speech excerpt of 1 min in length is analysed to give the spectral distribution of the envelope fluctuations. This allows the formation of the modulation ratio as a function of modulation frequency as shown in Figure 1, panel B, where the spectrum is normalized with respect to the mean intensity I_k .

NOTE k is the octave-band number. See 6.1.

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.

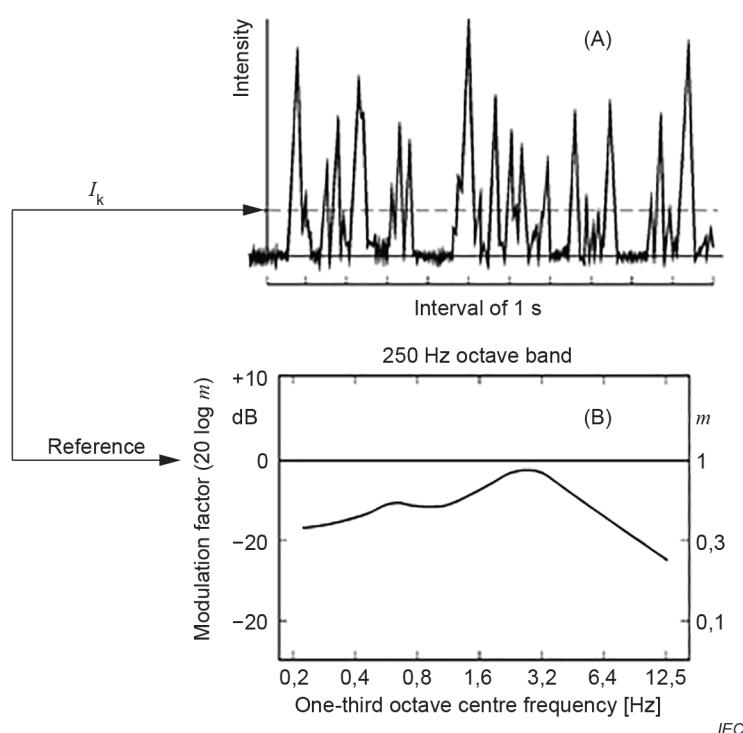


Figure 1 – Envelope function (panel A) of a 10 s speech signal for the 250 Hz octave band and corresponding envelope spectrum (panel B)

4.3.2 Reduction of modulation

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 2 shows the concept of the reduction in modulation that can occur between a talker (input) and listener (output).

4.3.3 Role of the octave-band noise carriers

The STI test signal was developed from parameters derived from speech material. In general, the STI test signal comprises noise signals in the seven octave-bands ranging from 125 Hz to 8 kHz. As the noise signals in these octave-bands carries modulation signals, they are termed "noise carriers". 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 model determines the modulation transfer function $m(F)$ of the transmission channel. In the Full STI method, 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 5.4 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. Its description and the octave-band weighting factors and redundancy factors are given in [8].

4.3.4 Theoretical overview

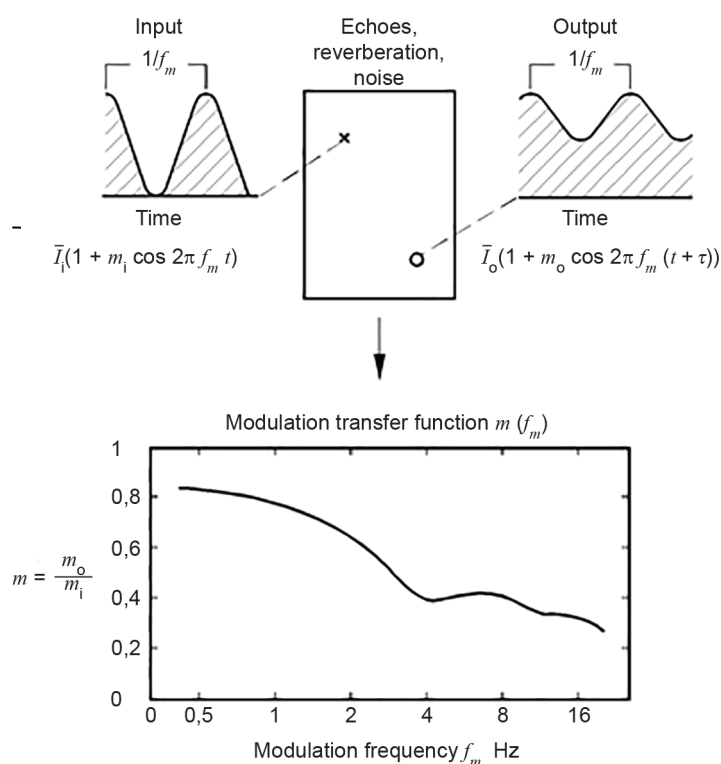
By proper choice of the form of test signal, the effective signal-to-noise ratio can include and allow for distortions in the time domain and non-linearities as well as background noise, etc.

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.2 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 modulation depth m_i of a test signal is played into a room or through a communication channel and received at a listener position with degraded modulation depth m_o . 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 as illustrated in Figure 2.



IEC

NOTE m_i and m_o are the modulation depths 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

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

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

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

The values of the effective signal-to-noise ratio are then 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 effective signal-to-noise ratios are used to calculate the modulation transfer index (MTI) in each octave band. The speech transmission index STI combines the MTI 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.

4.4 Measurement of STI

NOTE Clause 8, Annex D, Annex P and Annex Q give details on practical measurements, the specifications of measuring systems and the uncertainty of results.

4.4.1 Direct and indirect methods

There are two methods to measure STI:

- direct method using modulated test signals
- indirect method based on the system's impulse response

Each method has advantages and disadvantages, some of which are shown in Table 2.

It should be noted that the direct and indirect methods may not always give identical results. This is generally owing to the noise-based carrier used in the direct method in comparison to the more-repeatable nature of the test signal used to derive the impulse response.

Table 2 – Comparison of direct and indirect methods

Subject	Direct method	Indirect method ^b
Post processing	possible	mandatory
Handheld device	possible	possible
Amplitude nonlinearities	reduce the reliability of the result	reduce the reliability of the result
Frequency response 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 during the measurement period	errors less than 20×10^{-6}	errors less than $0,5 \times 10^{-6}$
^a See 7.8 for further details		
^b Different methods of deriving the impulse response may produce small differences		

4.4.2 Full STI

The research described in [4], [5], [6], [8],[9], [10], and [11] developed the basis and method for the Full STI.

Originally, the full STI measurement consisted of 98 separate test signals using 14 different modulation frequencies in 7 octave bands. Each test signal contained only one modulation frequency for only one octave band noise carrier; the other octave bands contained no signal. The test signals were generated sequentially. With an average of 10 s per modulation signal, a Full STI measurement required approximately 15 min to execute and, therefore, it is now rarely used.

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.

Simultaneous use of a number of modulation frequencies enables the Full STI to be measured in a much shorter time. Further information about these techniques is given in Annex O.

The indirect method is also widely employed for Full STI.

4.4.3 STIPA

STIPA is a simplified form of the Full STI and is based on measurements using a lower number of modulation indices (see Clause 5). The STIPA test signal uses a predefined set of two modulations per octave band that are generated simultaneously, giving a total of 14 modulation indices. STIPA has a substantially shorter measurement duration than the Full STI and is the primary use of the direct method. Annex B provides a detailed description of STIPA.

The STIPA test signal 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 15 s and 25 s.

STIPA can also be derived using the indirect method and shall be referred to as STIPA(IR).

4.4.4 Choice of method

The STI model, 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 model:

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

NOTE Some types of spectral distortions might not be accounted for, see 7.8.

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 3 and Table 4 give an overview of the STI test methods versus the types of linear and non-linear distortion for which they are appropriate. The + and – symbols are a general indication of the suitability of the method.

Table 3 – Suitability 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	yes	condition dependent	yes
Direct STIPA	yes	yes	limited	condition dependent	yes
Indirect Full STI ^c	no ^b	yes	yes	no	yes
Indirect STIPA ^c	no ^b	yes	limited	no	yes

NOTE The term 'condition dependent' is used to indicate that the corresponding test signal type might or might 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 might be pessimistic;
- 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 might produce a perceived loss of intelligibility that is not adequately accounted for in the result, see 7.8.

^b Yes, if a MLS test signal is used, however signal averaging of time domain data shall not be employed, and the excitation spectrum shall be speech-shaped.

^c This includes time delay spectrometry.

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

The effects of noise should be computed mathematically.

Table 4 – Test-method suitability

Type of Distortion	Full STI Direct	STIPA	Full STI Indirect	Limitations	Work-arounds
Non-linear	++	++	--		
Reverb	++	++	++		
Echo delay	++	-	++		
Noise	++	++	+/-	Depends on test signal	Post addition of noise to MTF matrix
AGC	++	++	+/-	Depends on test signal	
Reverb + noise	++	++	+/-		Post addition of noise to MTF matrix
Analog phase or frequency shifting	++	++	--	All methods unsuitable with changes to the digital sample rate of the test signal	

The + and – symbols are a general indication of the suitability of the method.

Table 5 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 5 – Measurement applications

Application	Full STI Direct	STIPA	Full STI Indirect	Comment	Work-arounds/ section in text
Assessing suitability of room acoustics for speech communication (no electronic amplification)	++	++	++		
Evaluating PA and VA systems	++	++	++		
Evaluating telecommunication channels (phone, radio)	+	+	--		
Channel features amplitude compression	+	+	--		
Measurements of industrial noise situations with fluctuating noise	+/-	+/-	+/-	Caution required	See 8.9 Measure levels and post process
Speech and noise clearly spatially separated, or a strong direct-field component exists in a highly reverberant environment	+	+/-		To be used with caution. Currently standardised methods are inaccurate.	See 8.11
Channels that do not permit artificial test signals, such as vocoders	+/-	+/-		Currently standardised methods are inaccurate.	Use a speech-based STI test signal or listener tests
The + and – symbols are a general indication of the suitability of the method.					

5 Direct method of measuring STI – User guidance

5.1 Overview

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 typically takes between 15 s and 25 s.

For the STI to take account of the real signal-to-noise ratios and the corrected speech level, the mean intensity of the test signal should be equivalent to the corrected (real) speech level at the test position. This is obtained using the method described in Annex J, in which the L_{Aeq} of the test signal is adjusted to be 3 dB greater than the typical L_{Aeq} of the measured real speech level at the measurement location (i.e. a 3 dB correction factor is added).

5.2 STIPA

The STI test signal can be simplified if the related modulations in other octave bands that are required for the accurate interpretation of non-linear distortions are omitted [12]. 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 5. 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 15 s and 25 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 [13]. The measurement of hearing assistive systems or channels is possible, though specific corrections can be also required [14].

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 model described in Clause 4, there are other limitations to the direct method of measuring the STI.

Because 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 $m(f_m)$ 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; see [12], [15], [16].

With fluctuating noise (for example, a babble of voices), higher deviations can 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 0,30. An estimate of the deviation should be made by repeating measurements for at least a restricted set of conditions. It is unwise to try and measure STI in the presence of significant impulsive noise, whose effects are complex and highly variable.

It is therefore good practice to average the STI results over two or three measurements for a specific condition. A number of standards require assessment of the variations and subsequent averaging.

6 Indirect method of measuring STI (impulse response) – User guidance

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" [17]. 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 $m_k(f_m)$ at modulation frequency f_m in octave band k .

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

where

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

f_m is the modulation frequency;

ρ_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 can affect the 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, using a speech-shaped maximum length sequence (MLS) signal without averaging. Measurement procedures used for determining the impulse response shall meet the following requirements, with further information provided in ISO 18233.

- Measurements of the impulse response shall be conducted in accordance with ISO 18233.
- 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.
- To produce a "noise-free" impulse response, an SNR of at least 20 dB should be obtained in all seven octave bands. If necessary, signal averaging can be used to achieve this.
- 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 considered 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 real 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 model described in Clause 4, there are other limitations to the impulse response method of measuring the STI of which non-linear distortions are of special importance.

Non-linear distortions of the measurement signal should be avoided as the indirect method does not correctly account for the effects of this distortion. When this method is used, the sensitivity to distortion strongly depends on the measurement procedure applied [16], [17]. For example, Fourier transform based methods 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 test signals, the non-linear distortion components appear at the beginning or end of the recovered impulse response and so can be evaluated. However, errors can arise if the reverberation time is long, as the reverberant tail of the distortion components can smear into the main impulse response.

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

When using a sine sweep test signal, any distortion components detected shall be edited out or removed from the IR before calculation of the STI can be undertaken.

7 Limitations of the STI model

7.1 General

It is important to realise that the STI itself is not a complete measure of speech intelligibility as it does not include the intelligibility of the talker. By definition, speech intelligibility can only be measured through listening tests with human participants. Instead, the STI is a measure which predicts the influence that a transmission channel has on speech intelligibility. The STI has proven itself as an accurate prediction tool, as long it is applied within its intended scope. Inherently, the method also has its limitations.

The STI is based on the general observation that the loss of intelligibility, as speech is transmitted through a specific channel, is related to a reduction in the intensity modulations in the speech signal.

To simplify the measurement of the degree to which intensity modulations are reduced, speech is replaced in STI tests by an artificial test signal. This means that the applicability of the STI has limits in at least three aspects:

- Some types of channels have a measurable impact on speech intelligibility, yet leave the modulation spectrum unaffected. This is the case with certain specific types of signal distortion (listed later), where the fine structure of the signal is severely degraded while the envelope remains unaffected.
- Some transmission channels are designed to specifically adapt to speech; the response of such channels to the artificial STI test signals might not be representative.
- Even if the STI accurately corresponds with speech intelligibility in theory, the technical method of measuring the STI can introduce errors. In other words, the measuring tools might be incompatible with the channel under test.

It should be noted that the STI test signal differs a little from human speech in temporal and spectral aspects. These differences can produce differences between STI and perceived intelligibility, and include factors such as:

- the dynamic range of speech, the measured value of which depends on the integration time;
- the energy distribution of speech in each time frame;
- the distribution of signal levels over the entire length of a speech segment or test signal (percentile exceedances);
- the absence of gaps in the test signal;
- the carriers in speech not being 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 [7].

Consequently, for certain situations and possible (narrow-band) transmission channels, care shall be taken when using the STI. In some cases, intelligibility can 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 subclauses (7.2 to 7.14) discuss specific situations in which the applicability of the STI is limited, in more detail.

7.2 Frequency shifts

This type of distortion can occur with:

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

Frequency shifts interfere with the correlation process through which modulations in the received STI test signal are matched with the modulations in the source signal by the analyser. Small frequency shifts can have a profound impact on the measured STI, whilst generally having little effect on intelligibility. Consequently, the measured STI might underestimate intelligibility for systems with frequency shifts.

Although frequency shifts rarely occur with real-time channels (such as PA systems), they do occur quite often with recorded speech (in particular, when speech is replayed from CD players).

7.3 Centre clipping

This type of distortion can 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, since the effect on intelligibility is a result of the degraded fine structure of the signal is undetected by the STI model.

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

7.4 Dropouts

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

7.5 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 can underestimate intelligibility for systems with jitter.

7.6 Digital voice compression systems

Digital voice compression systems are often based on models of human speech. STI test signals, on the other hand, are based on modulated noise, which these systems tend to suppress rather than reproduce. STI test signals are therefore fundamentally not suited for digital voice compression systems.

This issue is a limitation inherent to STI measuring model. In addition, voice coders also tend to affect the fine structure of speech to such a degree that intelligibility is affected.

Generally, STI measurements should not be made with channels that include digital voice compression systems. Exceptions can be made for those cases where it is demonstrated that the test signal (the fine structure as well as intensity envelope) remains unaffected by the voice coder. This can be the case with higher bit-rate compression systems.

7.7 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 might be too high.

As an example, this issue can arise when investigating the behaviour of STI with changes to the form of the test signal spectrum. If an MTF matrix having every m value at 1 (i.e. no degradation 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 [18].

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.

7.8 Frequency response

Research so far [16], [18], [19], [20], [21], [22] 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 under conditions of low noise have been reported where the perceived speech intelligibility has been found to be inadequate owing to the poor frequency response or tonal balance of the system. The application of equalisation to improve the frequency response substantially 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. (For example, 1/3 octave bandwidth or 1/3 octave smoothing might be employed). Nonetheless, there are significant factors that might not be 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 effective frequency response of the microphone.

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 [19], [20].

7.9 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. The effect is also dependent on the rate of the transmitted speech. This issue is the subject of ongoing research, see e.g. [21], [23]

This subclause describes in detail the influence of a single secondary reflection or a delayed arrival on the MTF.

Whereas reverberation produces an effect on the variation with frequency of the MTF to that of a low-pass filter, a secondary arrival (or echo) has a similar effect to a notch-filter on the MTF frequency response.

NOTE In the example of Figure 2, the frequency response of the MTF extends from 0,5 Hz to below 20 Hz. For STI measurements, frequencies above 12,5 Hz are not taken into account.

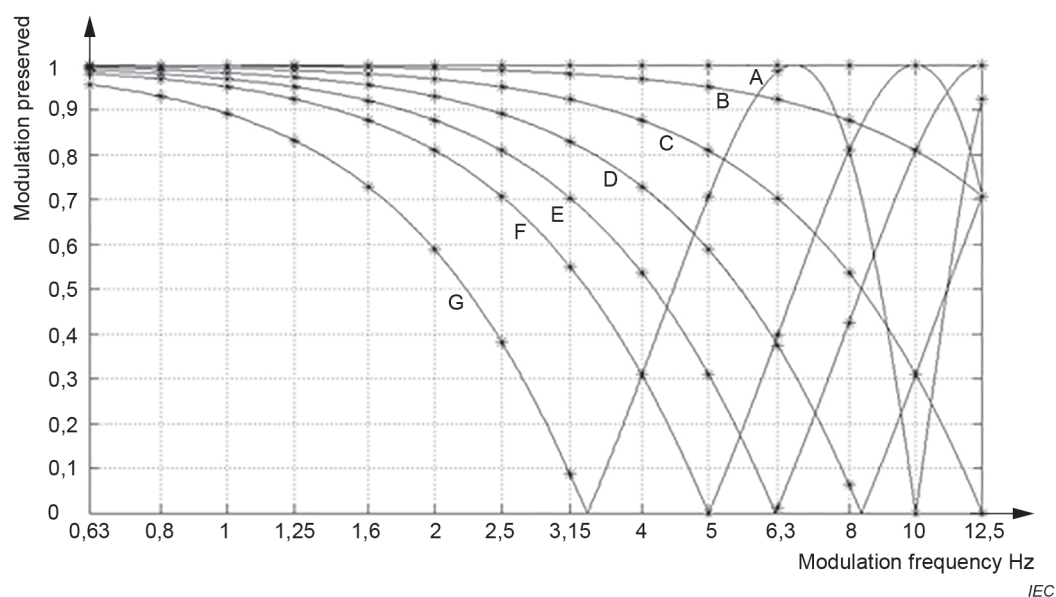
For a simple transmission system that consists only of the direct sound and a single reflection (or secondary arrival), the reflection always produces a reduction in modulation. With very short delay times, up to two or three milliseconds, this reduction appears at high modulation frequencies and is therefore mostly outside the 12,5 Hz upper limit range of the STI method. As delay times further increase, the notch moves towards lower modulation frequencies and since the notch repeats at multiple modulation frequencies, eventually multiple notches appear in the MTF.

If the intensities of the two arrivals are equal, the modulation value at the notch frequency reduces to zero. The larger the level difference between the two intensities, the smaller is the reduction in modulation.

The frequency at which a notch occurs does not necessarily coincide with any of the third-octave band frequencies at which the MTF is sampled, so for many delay conditions, the modulation in a given band is not reduced to zero at the MTF sampling frequency.

Figure 3 shows the effect of secondary arrivals on the modulation transfer function and resultant MTF values for a range of differential arrival times. The intensities of the first and second arrival are assumed to be identical, resulting in m -values ranging from 0 to 1. The graphs have been computed with a continuous frequency input, with the heavy dots in each graph showing the values that would be recorded in the associated MTF matrix.

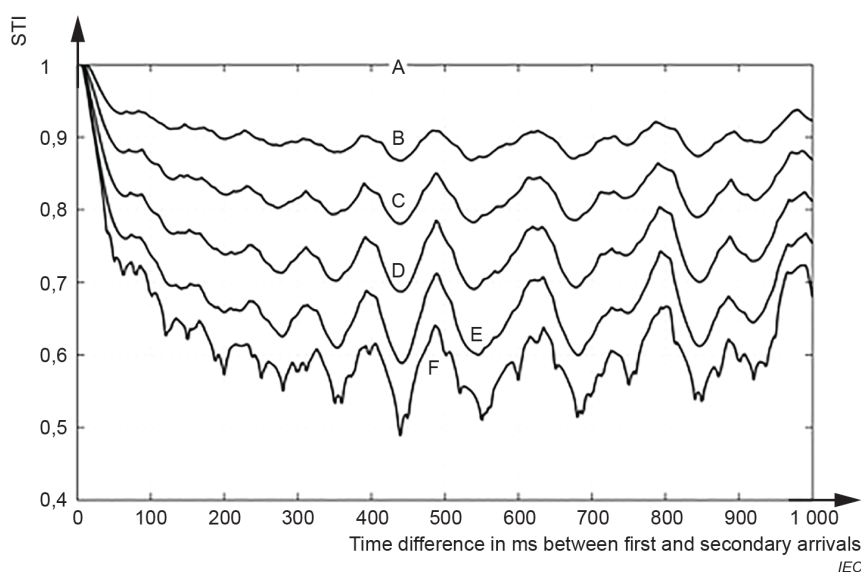
Notches in the MTF are clearly evident. For the six conditions from Figure 1, the STI values resulting from the respective MTFs are 1, 0,92, 0,77, 0,72, 0,71 and 0,64, respectively.

**Key**

- A 0 ms
- B 20 ms
- C 40 ms
- D 60 ms
- E 80 ms
- F 100 ms
- G 150 ms

Figure 3 – Effect of a single delayed arrival on the MTF (idealised conditions)

Figure 4 shows the resulting STI values when MTFs become notch-filtered by secondary arrivals with delay times between 0 ms and 1 000 ms, and various relative levels between the two arrivals.



Key

- A ±15 dB
- B ±12 dB
- C ±9 dB
- D ±6 dB
- E ±3 dB
- F 0 dB

NOTE The secondary arrival can be either higher or lower in level than the primary signal – either situation has the same effect on the STI.

**Figure 4 – Idealised STI (Male speech Spectrum)
versus delay and level of secondary arrival**

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

7.10 Fast amplitude compression and expansion

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

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

Fast compression reacts on the near-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 is also likely to be reduced [22].

On the other hand, automatic gain control (AGC) has slow reaction recovery times and generally 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.

7.11 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.)

7.12 Hearing impaired listeners

Without specific corrections, the STI model is not a reliable predictor of the intelligibility of speech for hearing-impaired listeners [13]. The measurement of hearing assistive systems or channels is possible, though specific corrections might be required [14]. In particular, the reception thresholds and masking function need to be disabled and the bandwidth considered. Further information is given in Annex I.

7.13 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.

Interpretation of the speech intelligibility in the presence of fluctuating noise is extremely difficult and cannot be addressed in this edition of this document. 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.

Fluctuating noise, such as babbling voices or machinery that is repeatedly turned on or off or is cyclical, can lead to variations in the STI value obtained for repeated measurements and can also produce 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 [24].

If the impulsive or fluctuating background noise cannot be removed, then an STI measurement without the test signal should have a value less than 0,3 to ensure that the temporal variations in the noise will not seriously degrade the STI measurement with the test signal.

It is preferable that STIPA meters indicate errors resulting from fluctuating and impulsive noise.

7.14 Conclusion

In general, the STI model is a conservative approach and can underestimate intelligibility in some applications, but there are a number of important exceptions.

8 Measurement procedures, post-processing of data and applications

8.1 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 [21]. If the output of the transmission channel is acoustic, the hearing reception threshold (SPL) shall be used as a minimum.

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

Measurements or predictions of STI should state which edition of this standard has been used.

8.2 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 considered. 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 by directly connecting the output of the test signal source to the analyser input). 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 [15].
- 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 frequency range 88 Hz to 11,6 kHz (the effective limits of the 125 Hz and 8 kHz octave bands) when measured in a free field (free of reflections).
- c) Verify that the individual octave band L_{eq} levels over the range 125 Hz to 8 kHz are within ± 1 dB and preferably $\pm 0,5$ dB of the values for the male spectrum signal given in Table A.4 when using a STIPA or other speech-shaped test signal conforming to the STI spectrum.

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.

- d) In the absence of an artificial mouth, a suitable transducer such as a small, single-source, high-quality loudspeaker with cone diameter or aperture not exceeding 65 mm, may be used, and shall be described with the results. The following parameters shall be provided by the source:

- directionality to match human talker;

- the shape of the test spectrum at 50 mm from the source shall not deviate from the defined STI spectrum shape (see Table A.4) by more than $\pm 2,5$ dB when measured at the specified reference point of 250 mm or 500 mm (as nominated by the manufacturer);
- the distortion characteristics associated with the system (e.g. driver excursion, amplifier power capacity, enclosure vibrational modes) shall be sufficiently low that the m values (in the MTF matrix) are unity when measured under anechoic conditions at the reference position with the maximum corrected speech level.

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 should be used when assessing the intelligibility of unamplified talkers or the acoustic pickup from microphones. The directional characteristics of the acoustic test source can have a significant effect on the results when making STI/STIPA measurements particularly in reverberant spaces, or when the pick-up microphone is located at some distance from the talker.

Apertures (cone diameters) not exceeding 65 mm are generally more representative of the directivity of a human talker. If larger diameter drivers are used to simulate live talkers, the high-frequency directivity might be too high for accurate STI measurements especially when using distant microphones. For further information, see [14],[25],[26],[27].

When the system (source) microphone is situated in either a reverberant or noisy location or if a close-talking or noise-cancelling microphone is involved, a mouth/talker simulator should be employed as the test-signal source. Under low noise/low reverberation conditions, direct injection of the test signal may be suitable.

- Set the acoustic test source on the axis of the system microphone at the normal talker position and distance.
- Set the test signal level at the microphone position to the corrected speech level that is used in the system. The speech and test signal levels shall be matched according to the method described in Annex J.

If the corrected speech level is unknown, a default equivalent level of 60 dBA at 1 m in front of the artificial mouth or test loudspeaker should be used.

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

NOTE 2 This test can stress the amplifier driving the source. See 14.9 of IEC 60268-3[28]. It can 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.

NOTE 3 The above levels are subject to wide variations in practice.

- Run the STI or STIPA test sequence. Normally, and where available, the "with noise" option should be selected.
- The sound field of the test signal should be allowed to develop and stabilise in the space for a minimum of 2 s before commencing a measurement. In highly reverberant spaces, e.g. road tunnels, this minimum period may need to be extended to between 5 s and 10 s. An insufficient stabilisation period can lead to over-estimation of the STI.
- 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 [29].
- 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.
- 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 (equalizers, signal delays, etc.) are correctly considered during the measurement process.

8.3 Acoustical output

The STI model is based on the use of a single omni-directional measurement microphone that 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). Alternatively, a measurement can be made with an artificial binaural ear/head simulator with appropriate adjustments as described in 8.11.

8.4 Electrical input

Follow the procedure in 8.2, 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, to include as much of the system as possible in the test.

The STI test signal level, 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.

8.5 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 (between approximately 55 dBA and 80 dBA). Broadband 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.

8.6 Examples of input/output combinations

8.6.1 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 8.2. A calibrated STI measuring device is used at the receiver location to determine the STI of the transmission channel.

8.6.2 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 real 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.

8.6.3 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 8.2. 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.

The STI may be used to measure the potential intelligibility performance of assistive listening systems (ALS) and in particular, their associated acoustic paths. In most ALS [e.g. hearing loop systems (HLS), also known as "audio frequency induction loop systems" (AFILS), or infra-red systems], it is the path between the pick-up microphone and transmission system that is of critical importance. Further information can be found in IEC 60118-4 [30]. In so-called "soundfield" or voice reinforcement systems, the path between the loudspeaker and listener is more relevant.

Special methods can be required when measuring the STI of assistive hearing systems and hearing loop systems, e.g. [14]. In particular, the reception thresholds and masking function need to be disabled. However, much of the general guidance given in Clause 7 is applicable.

8.6.4 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 at 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 statistically has 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 or an artificial ear (see IEC 60318 [2]) should be used.

8.7 Spatial averaging of STI measurements

In some applications, e.g. evacuation using speech messages, applicable standards give information about how STI measurements are to be executed in detail and how a combined result is to be calculated from multiple measurements representing an average over an area. Often, such standards require a space to be subdivided into areas that are characterised by a common scenario, e.g. reverberation time and/or background noise. Such areas are typically termed "acoustically distinguishable area" or ADA. Most applicable standards also specify the number of measurements to be taken or alternatively, they specify a typical grid on which measurements shall be performed (see for example ISO 7240-19 [31]).

For the calculation of a single-number result, averaging of individual measurements and some statistical post-processing is required. The applicable standards specify this process, but two commonly used methods are described below.

- Perform STI measurements at each specified location within an ADA in accordance with the applicable standard. To reduce statistical fluctuations, apply averaging at each individual location.
- One method then takes all positional results and from that dataset, calculates the arithmetic mean value as well as the standard deviation and then subtracts the standard deviation from the mean value to produce the result.
- Other methods require more complex post-processing of the individual result, e.g. generating a cumulative distribution function, discarding a certain percentage of samples and post-processing the remaining values for mean and minimum values.

Some basic rules should be observed when selecting analysis positions across a space or building:

- It is generally required to perform separate measurements for each room.

- Performing measurements of example rooms is permitted for time-saving purposes, provided that they are identical in terms of their room acoustics (including furniture and equipment), dimensions, sound system implementation and ambient noise.
- The selection of the measurement positions shall be representative. Positions immediately next to loudspeakers should not be included to avoid exceedingly good results. Grid size should be selected based on the application and the applicable standards, but should include as a minimum, two positions per room. For further information, refer to ISO 7240-19.
- No more than one-third of the measurements should be made on the axis of the loudspeakers.
- Symmetries in terms of room geometry, surface material and sound system design can be utilized for reducing the number of necessary measurement positions.
- The selection shall consider changing spatial conditions within a room (e.g. variation of the ceiling height, the acoustical absorption of the surface or shadowing effects).
- The microphone position during the measurements should be at assumed ear height. The height to be assumed is 1 m to 1,2 m for seated persons and 1,5 m to 1,7 m for standing persons.

8.8 Post-processing of measured MTF data

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

- elimination of noise (de-noising) from a measured MTF;
- addition of an ambient 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 the stationary background noise and desired real speech levels. The equations listed in Annex A are used for this process.

8.9 Issues concerning noise

8.9.1 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 model 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 in which 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 considered. This can be achieved by applying a suitable criterion, such as NCB, RC or NR curves (see [32]).

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 and impulsive noise is detected by measuring the direct STI in the absence of the test signal. If the STI is too high (e.g. $STI > 0,3$), 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 8.9.2), and the STI computed mathematically.

8.9.2 Measurement of background noise

To correct an STI measurement for the effect of 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 shall be measured over a sufficient period of time in order to accurately characterise it. The positions, durations and times of the measurement shall be recorded together with the notes on unusual circumstances that can 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 A-weighted broadband sound pressure level value for the speech signal in dB. Also see 8.9.3.

8.9.3 Fluctuating noise

If fluctuating noise cannot be eliminated, its influence should be minimised by amplifying the signal until it is ideally at least 20 dB above the noise level in each octave band. The time-averaged level (L_{eq}) of the real ambient noise is then determined. Using the measured modulation indices, the STI is computed based on the original signal levels before amplification and the time-averaged ambient noise. This process 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.

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 into 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.

8.10 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. Examples of items that should be examined are as follows:

- constant or slightly reducing modulation transfer ratio values as a function of modulation frequency indicate that noise is the dominant mechanism; refer to Figure A.1.
- modulation transfer ratio values monotonically decreasing with modulation frequency indicate that reverberation is the main mechanism; refer to Figure A.1.
- values that initially reduce and then increase with modulation frequency indicate the presence of strong reflections arriving later than 50 ms, which can produce an over-optimistic conclusion about intelligibility. It is recommended that if this effect is detected, it should be reported with the result.

The accuracy of the STI itself is higher than the accuracy of any single m -value, as the STI is based on a mean value over all m -values in each octave band.

In real life, spurious modulations commonly occur, and it is quite common to observe individual m -values greater than 1,00; m -values greater than 1,00 do not necessarily indicate that the measurement is invalid. For example: if under certain conditions the statistical error associated with a measured m -value is 0,15, and the true m -value which is being estimated has a value of 1,00, then estimated m -values up to 1,15 are to be expected. There can also be values as low as 0,85.

However, real-life testing environments do not lead to individual m -values greater than 1,3. If m -values great than 1,3 are measured, this is a clear indication that the measurement is invalid, most likely due to the influence of impulsive or fluctuating noise.

8.11 Binaural STI measurements

Although the STI is a well-accepted and standardized model 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 model might 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 [33].

ITU-T P.58 [34] gives the diffuse-field correction factors to enable conversion of levels measured in a free field to measurements using an artificial head. (This is important for the masking algorithm to properly account for the levels at the ear reference point.)

9 Use of the STI as a design prediction tool

9.1 Overview

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

- calculation based on a predicted direct field, possibly combined with some ray-tracing for early reflections and a simulated exponential reverberant decay and simple electro-acoustic parameters.
- calculations based on a predicted direct field and statistically calculated reverberation times.
- prediction based on a computed impulse response of the system in the acoustic space.

Predictions based on computed impulse response offer greater 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.

Where the real speech level and spectrum is known or estimated, this level shall be used for prediction of the STI as it 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 corrected speech sound pressure level.

9.2 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. [35], 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. [35];
- 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 owing to poor frequency response.

9.3 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 evaluated part 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 real 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 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 that have been applied to the prediction.

Annex A (informative)

The basis of the STI concept

A.1 Introduction to this annex

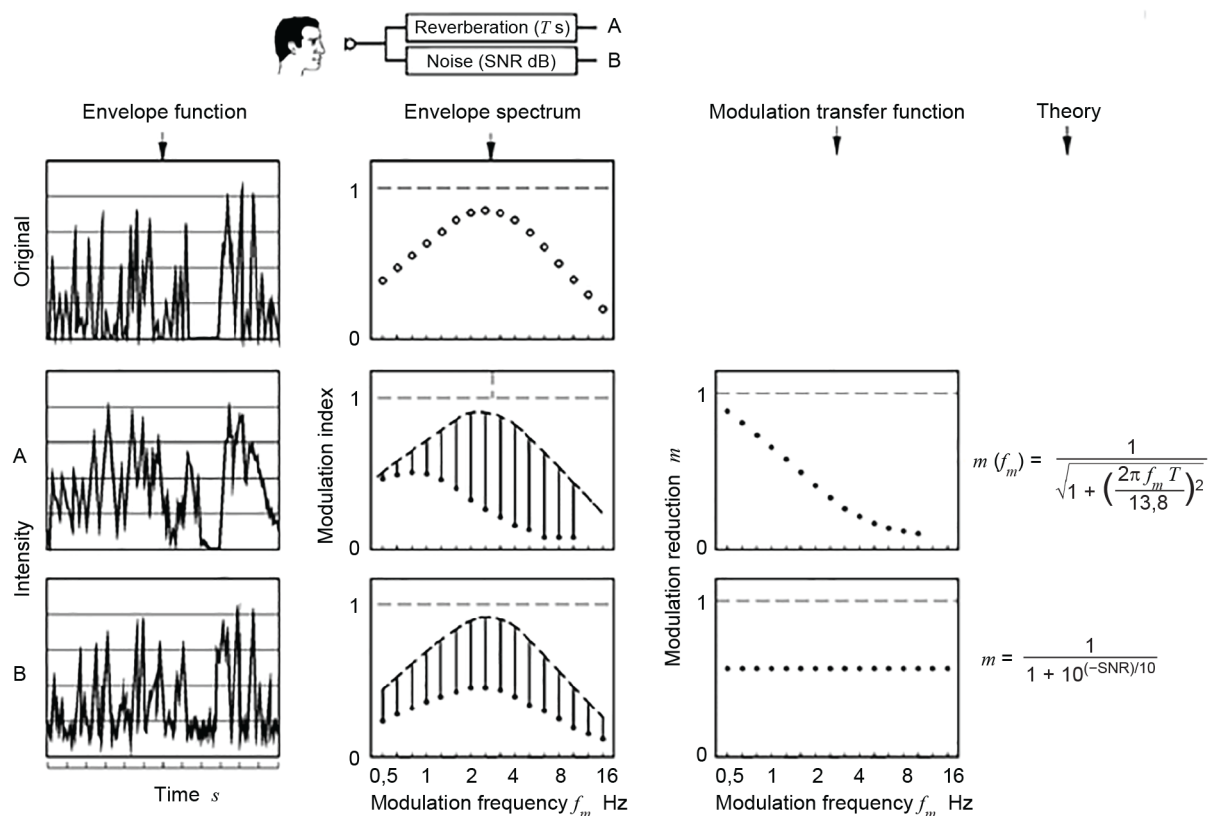
A.1.1 Purpose

This annex details the basis of the STI and its calculation steps, to assist manufacturers of measurement tools and prediction software. It is also of value for general users of the STI who wish to understand the theoretical basis and the post processing of MTF data.

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 [4], [5], [10], [11]. 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 normally affects only the depth of modulation of a sinusoidal modulation without changing its shape.

Figure A.1 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) (the vertical lines in the envelope spectrum indicate the reduction in modulation index at each modulation frequency).



IEC

Figure A.1 – Theoretical expression of the MTF

With reverberation, the MTF frequency response has the shape of a low-pass filter: the faster fluctuations being relatively more affected than slower fluctuations. In the theoretical case of a diffuse sound field with a purely exponential reverberant decay, the MTF can be derived mathematically (see Figure A.1, 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.1, case B). The noise, by increasing the mean intensity, reduces the modulation depth for all modulation frequencies as given by:

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

where

ρ is the signal to noise ratio in dB.

With strong echoes (pronounced reflections) the MTF frequency response shows the shape of a notch filter, rolling off first and then rising again with higher modulation frequencies. This is illustrated in Figure 3, in 7.9.

A.1.3 STI model

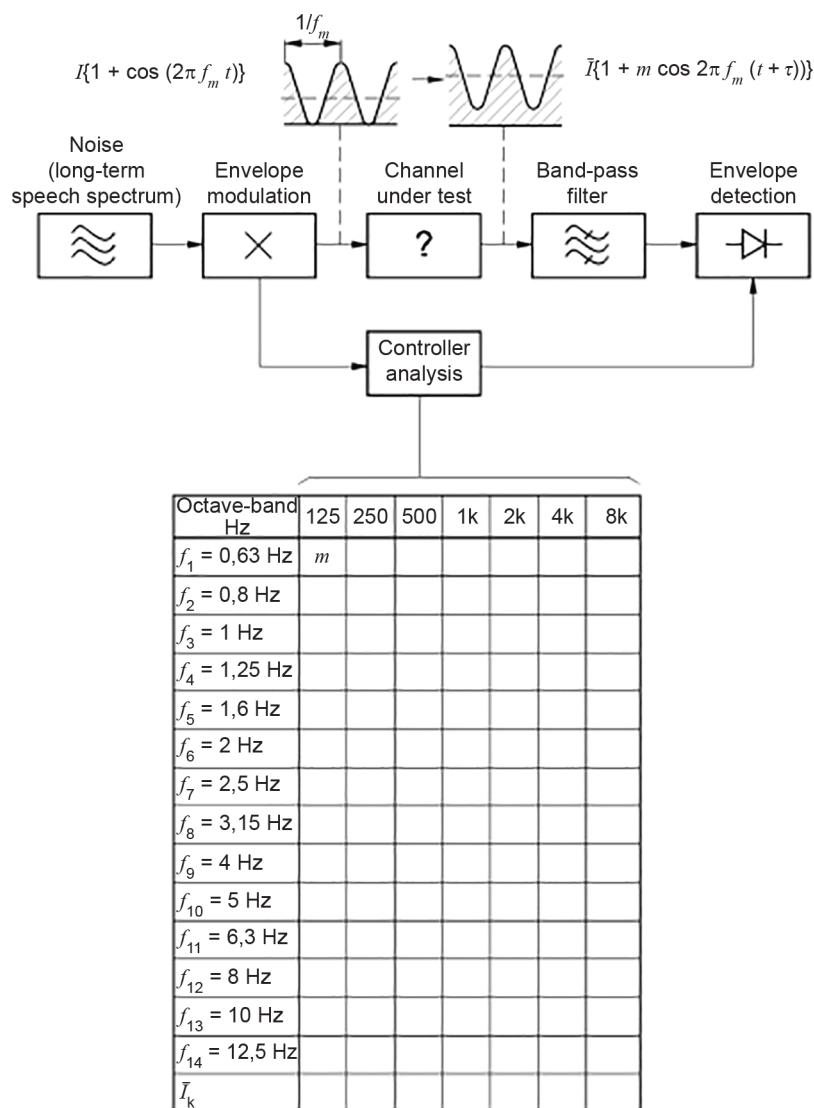
The STI model uses a discretized version of the MTF, with the modulation transfer ratios (also known as m values) being computed at specific modulation frequencies.

The most comprehensive measurement of the STI is the Full STI. To obtain a single STI value, the direct version of the Full STI model uses 98 independent test signals (14 modulation frequencies × 7 octave carrier bands).

The STI model was originally developed using a test signal that applied each modulation frequency sequentially to each octave band. Random modulations were used for all other the octave bands not under test. 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 specific octave band under evaluation. This model is referred to as the Full STI method.

Because each test signal is approximately 10 s long, the direct Full STI measurement requires 980 s to complete. As this is impractical in many situations, the faster method STIPA is now used. The direct Full STI method is only used for background STI research.

The concept is shown diagrammatically in Figure A.2.



IEC

NOTE The value of the modulation transfer function ($m(f_m)$) is determined for all cells of the matrix of 7 octave carrier 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.2 – Measurement system and frequencies for the STI method

A.1.4 STI modulation frequencies

The 14 STI modulation frequencies are at nominal one-third octave intervals and consist of 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 For frequencies used by STIPA, see Table B.1.

A.2 Calculation of the STI

A.2.1 General equation for STI

The STI is calculated using:

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

where

M_k is the modulation transfer index for octave band k ;

α_k is the gender-specific weight factor for octave band k ;

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

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

The modulation transfer index (MTI_k) per octave band k is obtained by averaging the transmission indices (TI) over the modulation frequencies:

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

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.

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.

It is calculated using the corrected modulation transfer ratio values:

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

where

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

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

A.2.2 Gender-specific octave band weighting and redundancy factors

The STI method can discriminate between male and female speech signals. However, in practice and to simplify the prediction and measurement process, only male speech shall be used.

The STI weighting factors (α) and redundancy factors (β) for male speech are shown in Table A.1 as a function of the octave bands.

Table A.1 – 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	–

The weighing factors for male STI contain an artefact that can occasionally appear when performing predictive calculations. For example, 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 is larger than 1,0. If the contribution of the 250 Hz band is completely removed, the STI becomes 1,03. For this situation, it is recommended to truncate the STI at 1,0. In practical STI measurements situations, it is unlikely that this artefact will appear, as noise will prevent this from occurring.

A.2.3 Adjustment of the MTF for ambient noise

$$m'_{k,f_m} = m_{k,f_m} \times \frac{I_{s,k}}{I_{s,k} + I_{n,k}}$$

where

m'_{k,f_m} is the derived modulation transfer ratio value for octave band k and modulation frequency f_m with ambient noise;

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

$I_{s,k}$ is the acoustic intensity level of the test signal in octave band k ;

$I_{n,k}$ is the acoustic intensity level of the background noise in octave band k .

In practice, m'_{k,f_m} is often measured directly; however, under very-high SNR conditions or simulations, the noise term $I_{n,k}$ might need to be added by post-processing.

NOTE $I_k = I_{s,k} + I_{n,k}$ equals the total acoustic intensity level, which is used in A.2.4 below.

A.2.4 Adjustment of the MTF for auditory masking and threshold effects

The derived modulation transfer ratio values (m -values) are corrected using auditory masking and threshold 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 total acoustic intensity level for octave band k and includes the received test signal level and noise ($I_k = I_{s,k} + I_{n,k}$);

$I_{am,k}$ is the total acoustic intensity level for the level dependent auditory masking effect on octave band k , as described in A.4.2. N.B.: this is the total level of signal and noise;

$I_{rt,k}$ is the acoustic intensity level of the reception threshold for octave band k , as described in A.4.3.