BS EN IEC 60268-16:2020

BSI Standards Publication

Sound system equipment

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

bsi.

National foreword

This British Standard is the UK implementation of EN IEC 60268-16:2020. It is identical to IEC 60268-16:2020. It supersedes BS EN 60268-16:2011, which is withdrawn.

The UK participation in its preparation was entrusted to Technical Committee EPL/100, Audio-visual equipment.

A list of organizations represented on this committee can be obtained on request to its committee manager.

This publication does not purport to include all the necessary provisions of a contract. Users are responsible for its correct application.

Published by BSI Standards Limited 2020

ISBN 978 0 580 97338 3

ICS 33.160.01

Compliance with a British Standard cannot confer immunity from legal obligations.

This British Standard was published under the authority of the Standards Policy and Strategy Committee on 30 November 2020.

Amendments/corrigenda issued since publication

Date	Text affected

EUROPEAN STANDARD

NORME EUROPÉENNE

EUROPÄISCHE NORM

EN IEC 60268-16

November 2020

ICS 33.160.01

Supersedes EN 60268-16:2011 and all of its amendments and corrigenda (if any)

English Version

Sound system equipment - Part 16: Objective rating of speech intelligibility by speech transmission index (IEC 60268-16:2020)

Équipements pour systèmes électroacoustiques - Partie 16: Évaluation objective de l'intelligibilité de la parole au moyen de l'indice de transmission de la parole (IEC 60268-16:2020)

Elektroakustische Geräte - Teil 16: Objektive Bewertung der Sprachverständlichkeit durch den Sprachübertragungsindex (IEC 60268-16:2020)

This European Standard was approved by CENELEC on 2020-10-30. CENELEC members are bound to comply with the CEN/CENELEC Internal Regulations which stipulate the conditions for giving this European Standard the status of a national standard without any alteration.

Up-to-date lists and bibliographical references concerning such national standards may be obtained on application to the CEN-CENELEC Management Centre or to any CENELEC member.

This European Standard exists in three official versions (English, French, German). A version in any other language made by translation under the responsibility of a CENELEC member into its own language and notified to the CEN-CENELEC Management Centre has the same status as the official versions.

CENELEC members are the national electrotechnical committees of Austria, Belgium, Bulgaria, Croatia, Cyprus, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, the Netherlands, Norway, Poland, Portugal, Republic of North Macedonia, Romania, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey and the United Kingdom.

European Committee for Electrotechnical Standardization
Comité Européen de Normalisation Electrotechnique
Europäisches Komitee für Elektrotechnische Normung

CEN-CENELEC Management Centre: Rue de la Science 23, B-1040 Brussels

European foreword

The text of document 100/3202/CDV, future edition 5 of IEC 60268-16, prepared by IEC/TC 100 "Audio, video and multimedia systems and equipment" was submitted to the IEC-CENELEC parallel vote and approved by CENELEC as EN IEC 60268-16:2020.

The following dates are fixed:

• latest date by which the document has to be implemented at national level by publication of an identical national standard or by endorsement	(dop)	2021-07-30
• latest date by which the national standards conflicting with the document have to be withdrawn	(dow)	2023-10-30

This document supersedes EN 60268-16:2011 and all of its amendments and corrigenda (if any).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. CENELEC shall not be held responsible for identifying any or all such patent rights.

Endorsement notice

The text of the International Standard IEC 60268-16:2020 was approved by CENELEC as a European Standard without any modification.

In the official version, for Bibliography, the following notes have to be added for the standards indicated:

IEC 60318-1:2009	NOTE	Harmonized as EN 60318-1:2009 (not modified)
IEC 61672-1	NOTE	Harmonized as EN 61672-1
IEC 60118-4	NOTE	Harmonized as EN 60118-4
ISO 9921:2003	NOTE	Harmonized as EN ISO 9921:2003 (not modified)
ISO/TR 22411:2008	NOTE	Harmonized as CEN ISO/TR 22411:2011 (not modified)
ISO 3382-1:2009	NOTE	Harmonized as EN ISO 3382-1:2009 (not modified)

Annex ZA
(normative)

Normative references to international publications with their corresponding European publications

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.

NOTE 1 Where an International Publication has been modified by common modifications, indicated by (mod), the relevant EN/HD applies.

NOTE 2 Up-to-date information on the latest versions of the European Standards listed in this annex is available here: www.cenelec.eu.

Publication	Year	Title	EN/HD	Year
IEC 61260-1	2014	Electroacoustics - Octave-band and fractional-octave-band filters - Part 1: Specifications	EN 61260-1	2014

This page deliberately left blank

CONTENTS

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

$L_{eq,T}$

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 $L_{eq,T}$ is given by the following equation:

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

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 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.

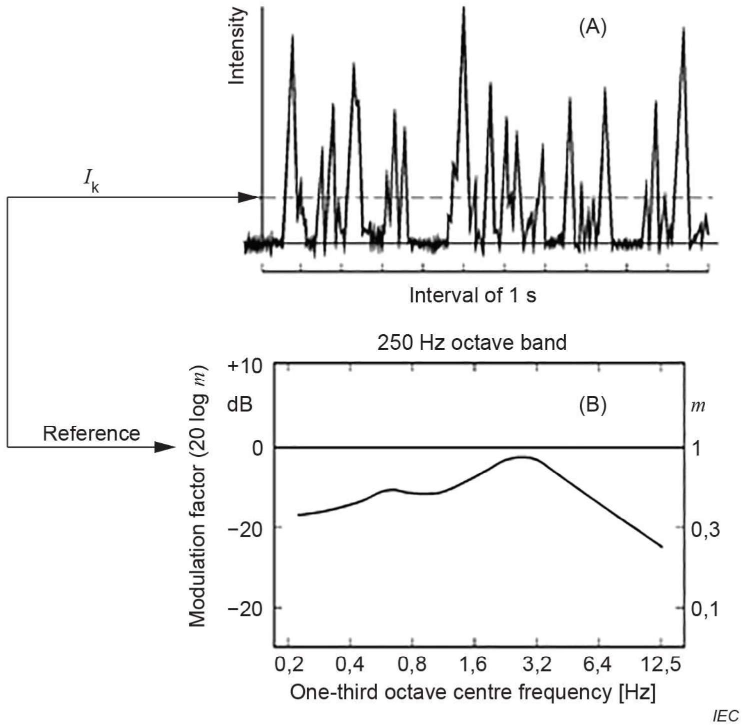


Figure 1 – Envelope function (panel A) of a 10 s speech signal for the 250 Hz octave band and corresponding envelope spectrum (panel B)

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

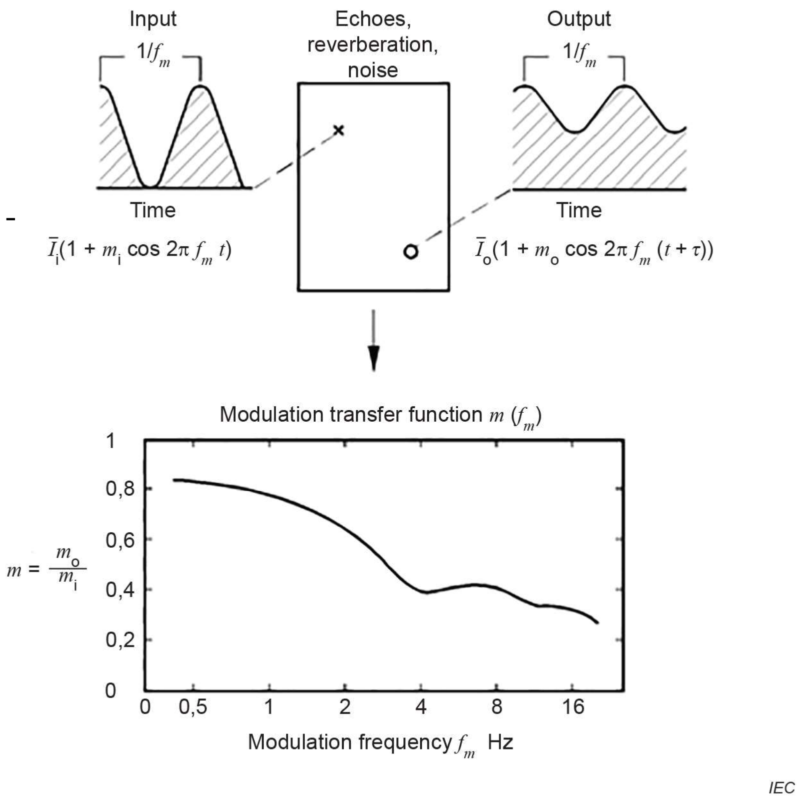
By proper choice of the form of test signal, the effective signal-to-noise ratio can include and allow for distortions in the time domain and non-linearities as well as background noise, etc.

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.



NOTE m_i and m_o are the modulation depths of the input and the output signals, respectively. $\overline{I}_\mathrm{i}$ and $\overline{I}_\mathrm{o}$ 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

$$m(f_m) = \frac{m_\mathrm{o}(f_m)}{m_\mathrm{i}(f_m)}$$

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.

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

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.

a) Measurements of the impulse response shall be conducted in accordance with ISO 18233.

b) The length of the acquired impulse response shall be at least 1,6 s and not less than half of the reverberation time of the room.

c) 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.

d) The use of excitation signals with a white frequency spectrum (e.g. as with time delay spectrometry (TDS), or maximum length sequences (MLS)) should be avoided under normal circumstances unless the background noise level is very low. A pink frequency spectrum (−3 dB/octave) produced with pink noise or logarithmic sine sweep (more rigorously, "exponential sweep") is generally more suitable. However, a speech shaped MLS signal can also be used without averaging to measure the effect of background noise on the STI directly.

e) Impulsive signals such as the Dirac function are not generally suitable when background noise, pass-band limiting and non-linear distortion are significant, since the average frequency spectrum and level distribution of typical speech are not represented in the test signal.

f) The impulse response method is only applicable to linear, time-invariant systems. If the transmission channel has functions with non-linear signal processing, these functions should be bypassed during the speech intelligibility measurement. If, for instance, the effective playback sound pressure level is increased by a nonlinear reduction of signal dynamics, this shall be 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.

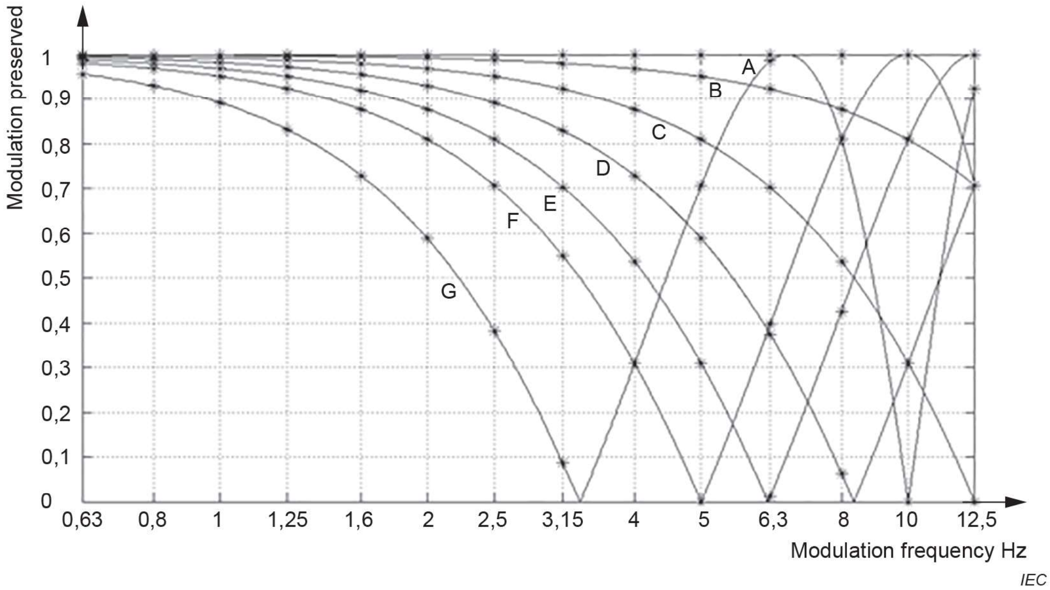
For a simple transmission system that consists only of the direct sound and a single reflection (or secondary arrival), the reflection always produces a reduction in modulation. With very short delay times, up to two or three milliseconds, this reduction appears at high modulation frequencies and is 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

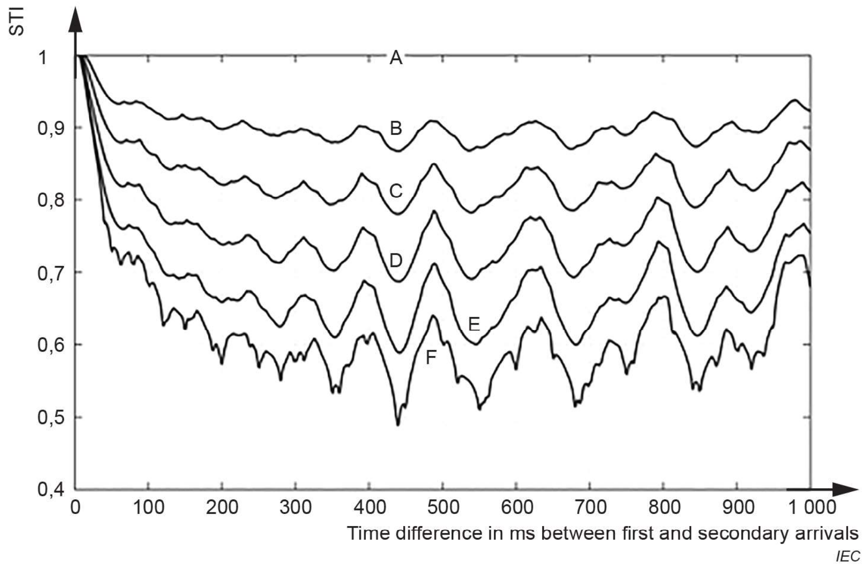
E 80 ms

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, companders (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 ±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 ±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.

e) Set the acoustic test source on the axis of the system microphone at the normal talker position and distance.

f) 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.

g) Run the STI or STIPA test sequence. Normally, and where available, the "with noise" option should be selected.

h) 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.

i) 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].

j) 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.

k) 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. $L_{\mathrm{A,eq}}$) 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).

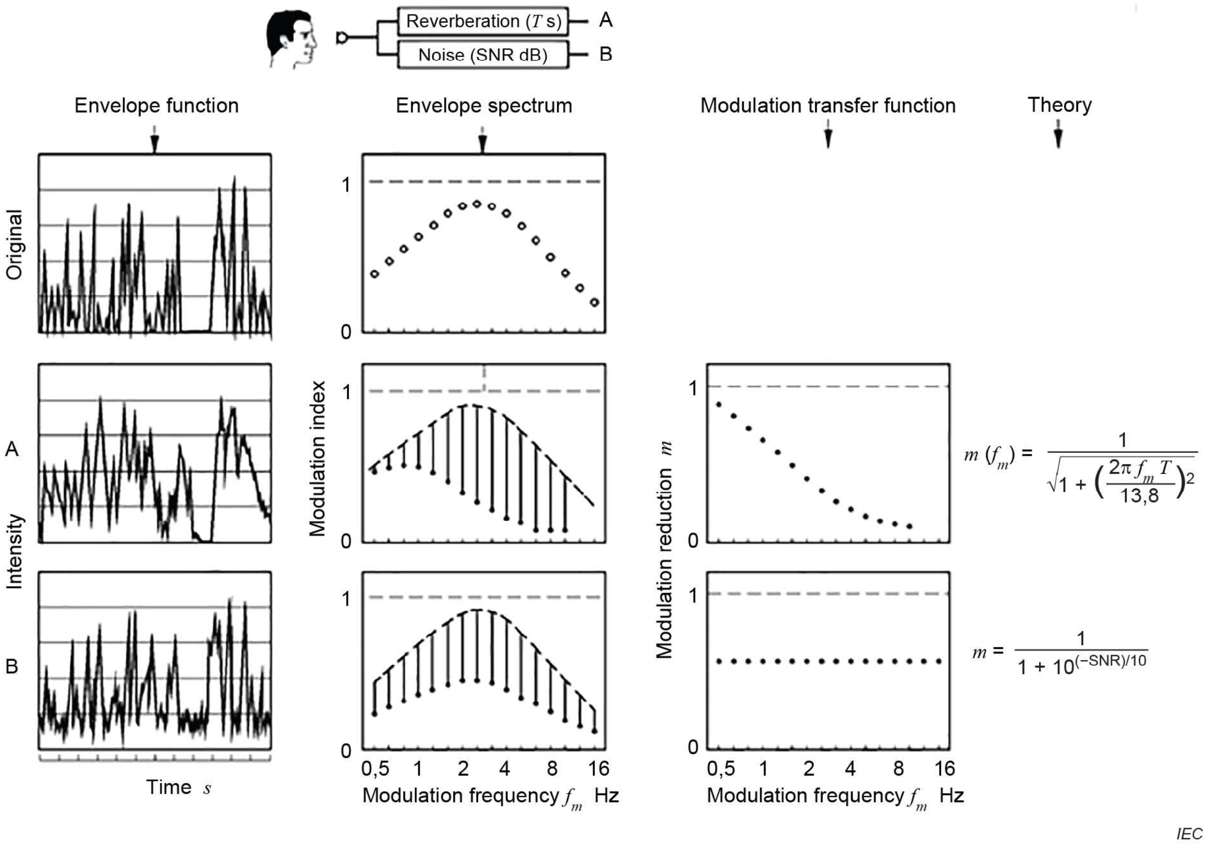


Figure A.1 – Theoretical expression of the MTF

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

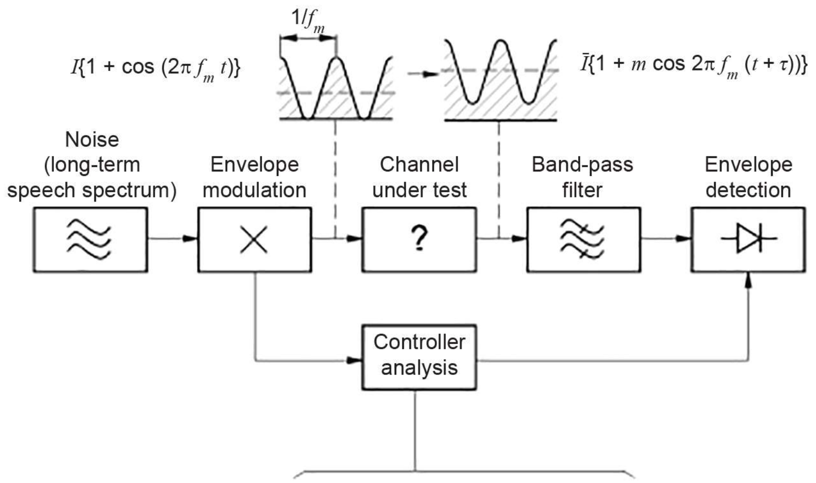
The STI model uses a discretized version of the MTF, with the modulation transfer ratios (also known as m values) being computed at specific modulation frequencies.

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 unpractical 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.



Octave-band Hz	125	250	500	1k	2k	4k	8k
f_1 = 0,63 Hz	m						
f_2 = 0,8 Hz							
f_3 = 1 Hz							
f_4 = 1,25 Hz							
f_5 = 1,6 Hz							
f_6 = 2 Hz							
f_7 = 2,5 Hz							
f_8 = 3,15 Hz							
f_9 = 4 Hz							
f_{10} = 5 Hz							
f_{11} = 6,3 Hz							
f_{12} = 8 Hz							
f_{13} = 10 Hz							
f_{14} = 12,5 Hz							
$\bar{I}_k$							

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:

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

where

$SNR_{\mathrm{eff}\,k,f_m}$ 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:

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

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.

The injected test signal shall have the specified speech spectrum shape.

Modulation transfer ratio values higher than 1,0 shall be truncated to 1,0. An m-value higher than 1,3 is very unlikely and is likely to be a result non-sinusoidal fluctuations or impulsive noises.

NOTE The auditory masking intensity ($I_{\mathrm{am,k}}$) is determined by logarithmically summing the level of the ambient noise and speech signal.

A.3 Calculation of the modulation transfer ratio values

A.3.1 Direct method: Analysis of the STI test signal

A.3.1.1 Basic processing steps

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

- filter the input signal with the seven octave band filters;
- determine the intensities I_k in each octave band k;
- determine the intensity modulation I at each modulation frequency f_m;
- determine the m_{k} values in each octave band k;
- calculate the STI in accordance with 4.3 and A.2.

A.3.1.2 Filtering and determination of intensities

The received modulated-noise test signal shall be band-pass filtered into seven octave-wide signals. The filters shall be one octave band wide with centre frequencies ranging from 125 Hz up to 8 000 Hz.

The shape and tolerance limits of the band-pass filters shall comply with IEC 61260-1, class 1. The input signal shall be split by the filter bank into output signals without power loss. Additionally, to minimize cross-talk between adjacent octave bands, filter slopes shall comply with the requirements of Annex C and provide 42 dB minimum attenuation at the centre frequency of each adjacent band. Filters can be implemented as IIR (infinite impulse response) or FIR (finite impulse response) types.

The phase characteristics of the band-pass filters should be as linear as possible to avoid distortion of the phase relationship of the amplitude modulations by the settling behaviour of the filters. The phase characteristics of the filters shall not give rise to a systematic error higher than 0,01 STI for the range between 0,1 and 0,9 STI (between −12 dB and 12 dB ρ_{eff}).

The intensity envelope shall be calculated by squaring the outputs of the bandpass filters and applying a low pass filter at a cut-off frequency of approximately 100 Hz to the intensity signal.

During one measurement, all intensities shall be calculated using the same time segment of the input signal; no implicit time weighting is allowed.

A.3.1.3 Derivation of the modulation transfer function (MTF) using the direct method

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

Depending on the test signal, the MTF for each octave band is derived by correlating the intensity envelope with sine and cosine signals with specific duration and modulation frequencies. The modulation transfer at each modulation frequency f_m is calculated by first deriving the modulation depth of the received signal (m_o) for the output of each octave band k (see Figure 2).

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

where

f_m is the modulation frequency in Hz;

t is the time in seconds;

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

The summation shall be made over the measurement duration using a whole number of periods for each specific modulation frequency.

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

NOTE 2 The use of window functions with non-integer periods leads to inaccuracies caused by data leakage.

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

$$m_{k,f_m} = m_\mathrm{o}(k, f_m) / m_\mathrm{i}(k, f_m)$$

where

$m_\mathrm{o}\left(k, f_m\right)$ is the modulation depth of the received test signal for octave band k and modulation frequency f_m;

$m_\mathrm{i}\left(k, f_m\right)$ is the modulation depth of the transmitted test signal for octave band k and modulation frequency f_m.

NOTE The derivation of the modulation transfer function (MTF) using the indirect method is described in 6.1.

A.3.2 Indirect method: Determination of the modulation transfer function (MTF)

A.3.2.1 Basic processing steps

The following sections set out the procedure for analysing the received signal and calculating the resultant STI using the indirect method. The procedure is broken down into the following primary steps:

- filter the impulse response with the seven octave band filters;
- determine the m_k values in each octave band k using equation in A.3.1.3;
- calculate the STI in accordance with 4.3 and Clause A.2.

A.3.2.2 Filtering and determination of intensities

The received impulse response shall be band-pass filtered into seven octave-wide signals. The filters shall be one octave band wide with centre frequencies ranging from 125 Hz up to 8 kHz.

The shape and tolerance limits of the band-pass filters shall comply with IEC 61260-1, class 1. The input signal shall be split by the filter bank into output signals without power loss, and therefore the roll-offs of adjacent filters shall be complementary with frequency and intersect at the −3 dB attenuation points. Additionally, to minimize cross-talk between adjacent octave bands, filter slopes shall comply with the requirements of Annex C and provide 42 dB minimum attenuation at the centre frequency of each adjacent band.

Backward filtering of the impulse response shall be used to minimise transient distortion of the modulation components of the impulse response due to settling of the filters. The phase characteristics of the analysis filters shall not give rise to a systematic error higher than 0,01 STI for the range between 0,1 and 0,9 STI (between −12 dB and 12 dB SNR).

Filters can be implemented as IIR or FIR types.

A.4 Auditory effects on the STI

A.4.1 Overview

The STI model models two specific hearing-related aspects by applying appropriate noise terms. These two aspects are upward auditory masking (the reduction in aural sensitivity by a stronger, lower frequency sound) [36] and the absolute reception threshold. These auditory effects reduce the effective signal-to-noise ratio in the various octave bands and can be expressed as a reduction of the modulation transfer function, which result in lower STI values.

The only parameter in the STI model that interlinks adjacent frequency bands with respect to the effective SNR is the auditory masking function. This masking function comes into play where strong aberrations in the amplitude frequency response are present (see also 7.8) and/or when sound pressure levels are high.

Auditory effects shall only be considered when test signals are obtained acoustically (in dB SPL), which is often the case in practice. If test signals are obtained electrically, without reference to a sound pressure level, this shall be noted, and auditory effects disabled in the measurement.

A.4.2 Level-dependent auditory masking

Auditory masking is an inherent effect of the human hearing process. When a loud, low-frequency sound is presented at the ear, it always masks higher frequencies, possibly rendering them inaudible if the difference between their relative levels exceeds a given threshold. This phenomenon is referred to as upward spread of masking [36]. The auditory masking effect also depends on the absolute sound pressure level of both frequency components. A low-frequency sound at low sound pressure level causes less masking of a high-frequency sound than if it were at a higher sound pressure level, i.e. the masking slope at low sound pressure levels is steeper than at high sound pressure levels [37].

The main principle of the auditory masking as modelled in the STI is shown in Figure A.3. It shows that a lower octave band ($k - 1$) has a masking effect on the next (higher) octave band (k). The slope of the masking function in turn depends on the total sound pressure level present in octave band ($k - 1$). Note that downward masking is not included in the model.

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

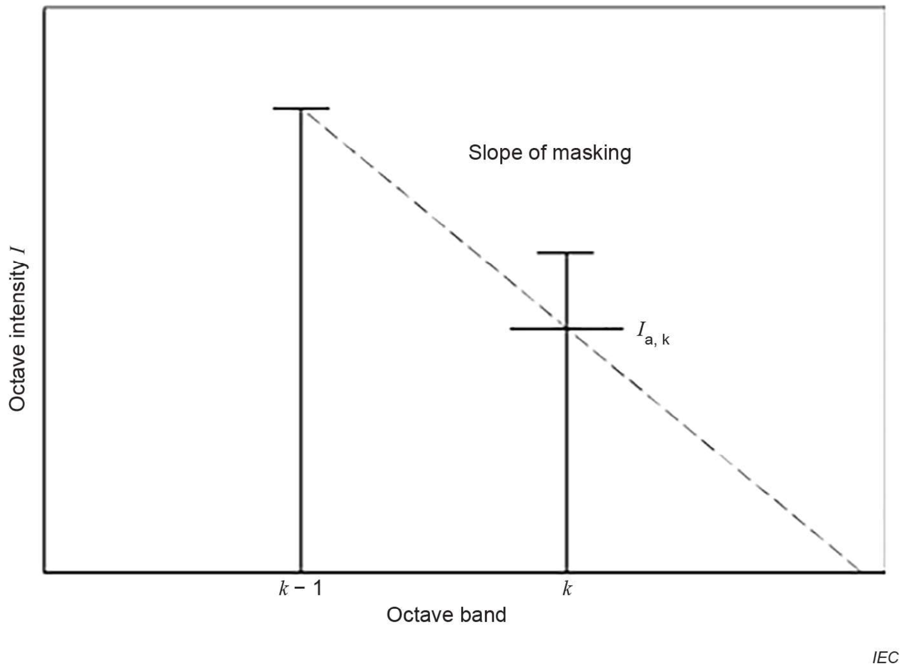


Figure A.3 – Auditory masking of octave band (k – 1) on octave band (k)

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

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

where

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

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

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

$$I_{k-1} = 10^{\left(\frac{L_{k-1}}{10}\right)}$$

where

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

The auditory masking factor a for octave band k therefore depends on the intensity of octave band k – 1.

In Table A.2, the level dependent auditory masking is given in dB for octave band k as a function of the sound pressure level L (dB) in octave band $k - 1$. It is noted that the auditory masking level is a function of the total sound pressure level in that octave band consisting of both the test signal level and the ambient noise level.

Table A.2 – Auditory masking as a function of the octave band level

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

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

$$a = 10^{\left(L_a/10\right)}$$

where

L_a is the octave band level dependent auditory masking value in dB as derived from Table A.2.

NOTE 1 If a mathematical adjustment is made to the STI for a different ambient noise situation, the auditory masking factor depends on both the measured signal intensity and the added ambient noise intensity in a specific octave band. Both intensities are added to obtain the overall acoustic intensity for an octave band to enable derivation of the appropriate auditory masking factor for that octave band.

NOTE 2 The masking scheme introduced with edition 3 was discrete and resulted in non-continuous STI results as a function of the overall sound pressure level. Since edition 4, the auditory masking scheme presented in Table A.2 yields continuous STI results as a function of the sound pressure level.

The effect of level dependent masking is shown in Figure A.4. This figure shows the STI values for a range of A-weighted speech levels from 0 dB to 120 dB with an MTF matrix produced by seven different reverberation times with equal reverberation times in each octave band. The effect of the reception threshold is also included in this graph.

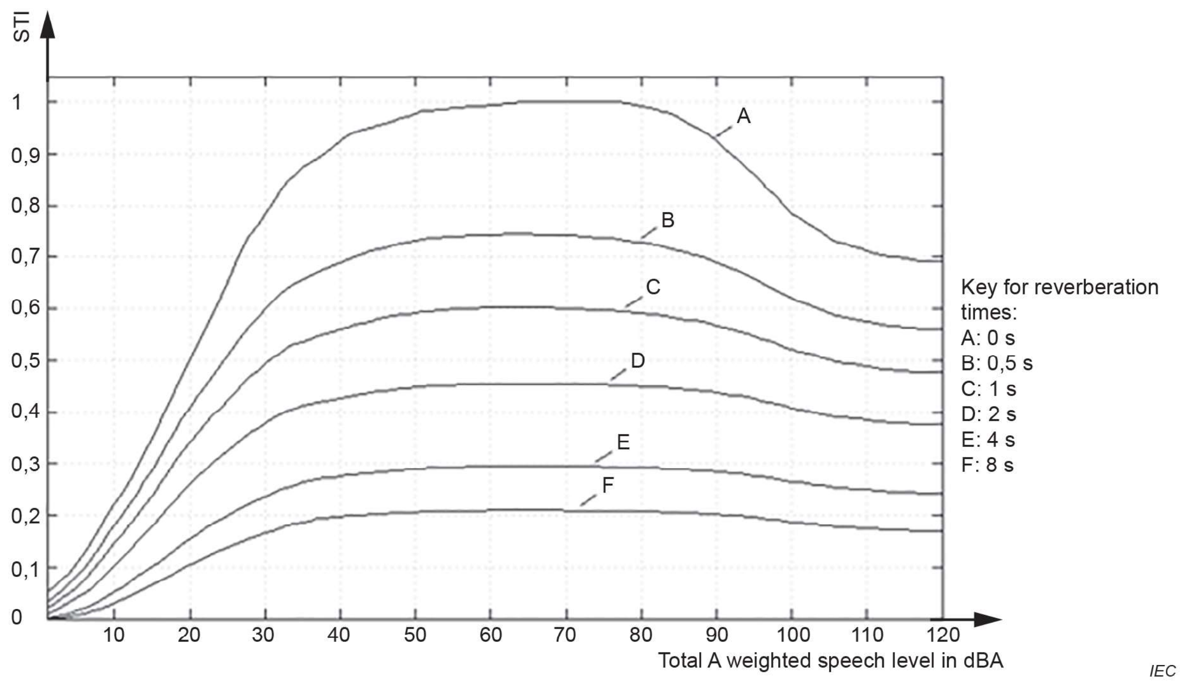


Figure A.4 – Relationship between STI and speech level for different reverberation times.

A.4.3 Absolute speech reception threshold

The absolute speech reception threshold is defined by the absolute threshold of hearing and the minimal required dynamic range for the correct recognition of speech. The absolute speech reception threshold intensity (I_{rt}) is modelled in the STI as an intrinsic noise floor which reduces the effective signal to noise ratio when speech levels are low. The speech reception threshold is given in Table A.3 (in dB SPL) as a level in each octave band.

Table A.3 – Absolute speech reception threshold level in octave bands

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

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

$$I_{\text{rt},k} = 10^{(A_k/E)}$$

where

A_k is the absolute speech reception threshold in dB for octave band k (see Table A.3).

A.5 Generation of the STI test signal (direct method)

A.5.1 Pink noise source signal

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

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

Pink noise can be produced from white noise using a low pass filter with a −3 dB per octave slope commencing at 63 Hz. The crest factor of the signal should typically lie between 12 dB and 14 dB, with the L_1 percentile exceedance (1 %) value typically lying between 8 dB and 11 dB.

A.5.2 Generating octave band carrier signals

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

EXAMPLE 12th order 0,2 dB ripple type I Chebyshev filters can be used for generating the half-octave wide carrier signals.

A.5.3 Intensity modulation of the carrier signals

Each carrier is amplitude modulated with the square root of a raised sinusoidal modulation at the maximum modulation depth (m = 1), as described by the following equation. Applying this function in the amplitude domain will result in a sinusoidal modulation in the intensity domain.

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

where

f_m is the modulation frequency in Hz;

t is the time in seconds;

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

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

For the generation of the STIPA test signal, see Annex B.

A.5.4 Applying the speech spectrum to the STI test signal

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

A.6 Spectrum of STI test signal

A.6.1 Standardized speech spectrum

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

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

Octave band Hz	125	250	500	1 000	2 000	4 000	8 000	A-weighted
Males dB	−2,5	0,5	0	−6	−12	−18	−24	0,0
NOTE 1 For guidance in determining the speech level, see Clause J.4.								

NOTE The male spectrum has changed in this edition of the standard. Prior to making this change, a theoretical investigation was conducted into the extent of changes to STI values with the above spectrum compared to the previous spectrum. Approximately 1,5 million scenarios were investigated; refer to [38].

A.6.2 Speech-shaped noise

To shape a noise signal to the speech spectrum specified in Table A.4, a combination of IIR (infinite impulse response) filters with gain can be applied to a pink noise signal. The required spectrum can be produced using a combination of high and low pass filters, a biquadratic peaking filter and gain settings. An example of filter parameters that produce that spectrum is given in Table A.5, and includes two types of common biquadratic peaking filters, which have different s-plane polynomials.

The spectrum produced by this combination is based on the test signal being integrated into octave bands and is not directly equivalent to the frequency response of the filter combination.

The levels in the 31,5 Hz, 63 Hz and 16 kHz octave bands shall be at least 20 dB lower than the 125 Hz and 8 kHz bands, respectively.

Table A.5 – Filter parameters and s-plane polynomials that produce speech-shaped pink noise.

Filter	Type	Parameter	s-plane polynomial
Filter 1:	Second order high-pass filter	resonant frequency f_h = 142 Hz Q = 0,58	$H_1(s) = \dfrac{s_n^2}{s_n^2 + \dfrac{s_n}{Q} + 1}$ where $s_n = \dfrac{s}{2\pi f_h}$
Filter 2 Alternative A	Biquadratic peaking filter	Centre frequency f_c = 500 Hz Q = 1,78 Gain g = 2,7 dB	$H_2(s) = \dfrac{s^2 + Bs + \omega_0^2}{s^2 + As + \omega_0^2}$ where $\omega_0 = 2\pi f_\mathrm{c}$ $G = 10^{\left(\frac{g}{20}\right)}$ $\Delta_W = \omega_0\left(2^{\frac{W}{2}} - 2^{-\frac{W}{2}}\right)$ where $W = \dfrac{2\sinh^{-1}\left(\dfrac{1}{2Q}\right)}{\ln(2)}$ $A = \Delta_W\sqrt{\dfrac{1}{G}}$ $B = GA$
Filter 2 Alternative B	Biquadratic peaking filter	Centre frequency f_c = 500 Hz Q = 2,04 Gain g = 2,7 dB	$H_2(s) = \dfrac{s_n^2 + Bs_n + 1}{s_n^2 + As_n + 1}$ where $s_n = \dfrac{s}{2\pi f_\mathrm{c}}$ $G = 10^{\left(\frac{g}{20}\right)}$ $B = \dfrac{G}{f_\mathrm{c}Q}$ for $g \geq 0$ $B = \dfrac{1}{f_\mathrm{c}Q}$ for $g < 0$ $A = \dfrac{1}{f_\mathrm{c}Q}$ for $g \geq 0$ $A = \dfrac{1}{f_\mathrm{c}QG}$ for $g < 0$
Filter 3:	First order low-pass filter	Turnover frequency f_l = 315 Hz	$H_3(s) = \dfrac{1}{s_n + 1}$ where $s_n = \dfrac{s}{2\pi f_l}$
Broadband gain		4,0 dB	

Annex B
(normative)

STIPA method

B.1 Overview

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

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

The STIPA method is only validated for the male speech spectrum. To obtain stability in the modulation domain with the noise carrier, the measurement duration shall be approximately 18 s, with a recommend range of 15 s to 25 s.

In theory, a further increase in measurement accuracy can be achieved by increasing the measurement time beyond 25 s. However, in practice there is also the risk that the accuracy of the measurement decreases with longer measurement times, owing to a possible (slight) mismatch in sampling frequency between the STIPA source and the STIPA analyser. If a higher accuracy is needed, it is recommended to compute the mean STI across multiple 15 s to 25 s measurements, rather than lengthening the measurement time.

Within a measurement range, for each individual modulation frequency, the maximum whole numbers of periods shall be analysed to minimise leakage by the time windowing. Accordingly, the analysis time is different for each per modulation frequency.

Table B.1 – Modulation frequencies for the STIPA method

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

If the indirect method is used to derive a STIPA result, it shall be referred to as STIPA(IR). STIPA(IR) is only useful to predict the deviation between STIPA and Full STI measurement results.

B.2 Test signal

The STIPA test signal is defined by a summation of seven noise octave-band carriers multiplied by their accompanying amplitude modulator, as given by:

$$\sum_{k=1}^{7} G_k N_k(t) A_k(t) \tag{B.1}$$

where

G_k is the octave band weighting factor;

$N_k(t)$ is the bandwidth-limited noise-carrier signal;

$A_k(t)$ is the amplitude modulator;

K is the octave band number.

The generation of noise band carriers for STI test signals is described in Clause A.5.

The octave band weighting factor G_k is given by:

$$G_k = 10^{\frac{L_k}{20}} \tag{B.2}$$

where

L_k is the level in dB in the octave band k.

The octave band levels are based on the male spectrum from Table A.4.

The modulator $A_k(t)$ for each octave band is described by:

$$A_k(t) = \sqrt{0{,}5\left(1 + 0{,}55\left(\sin\left(2\pi f_{1k} t\right) - \sin\left(2\pi f_{2k} t\right)\right)\right)} \tag{B.3}$$

where

f_{1k} is the first modulation frequency in Hz in the k^{th} octave band;

f_{2k} is the second modulation frequency in Hz in the k^{th} octave band.

NOTE With STIPA, both sinusoidal oscillations are added in opposite phase so that the modulation signal's crest factor is minimized. This allows for the maximum modulation factor of 0,55.

Annex C
(normative)

Verification of STI measuring devices

C.1 Specification of the measuring device

In order to verify the measuring device, an explicit specification is required, against which to verify the performance. Table C.1 gives the details of the specification and their relation to the clauses of Annex C. Manufacturers shall specify all the characteristics listed in Table C.1 and are free to add information about other characteristics that is consistent with, and does not obscure, the listed characteristics.

Table C.1 – Specification of an STI measuring device

Characteristic	Sub-clause reference	Requirements
Modulation depth for STIPA direct method	C.3.2	The absolute value of the error between the derived and the theoretical m-value shall not exceed 0,05. Overall m-value errors shall not yield a systematic absolute error (offset) in the STI results greater than 0,01. m-values shall be reported for the highest and the lowest possible measuring levels of the measuring device meeting the criteria.
Modulation depth for STIPA indirect method	C.3.3	
Crosstalk between octave-band filters	C.4	Crosstalk shall not increase the effective noise level representing the auditory masking by more than 3 dB. If, in testing, an m-value lower than 0,5 ± 0,05 is obtained, the level of the observed modulated carrier shall be increased in 1 dB steps to a level where an m-value of 0,5 ± 0,05 results. With the measured relative level, the corresponding sound pressure level as per Table A.1 is determined and reported as the sound pressure level below which the STI is underestimated.

C.2 Signals for testing STI implementations

Test signals are preferably stored as uncompressed wave files with a sample rate of at least 48 kHz and at least 16 bits (signed) per sample. The duration of the signal should be greater than the duration of the analysis period. Test signals shall either be injected electrically or inserted at an algorithm level.

C.3 Testing the dynamic range in the modulation domain

C.3.1 General

The modulation depth of the envelope function is the key factor for STI calculations. Any error or deviations within the modulation domain directly reflects as an error in the STI value. Although the determined modulation transfer ratios (m-values) are not affected by masking features, it is recommended that auditory masking features are disabled during testing.

C.3.2 Modulation depth testing for STIPA direct method

Since the direct method uses modulated noise band carrier signals, it is relatively easy to replace the noise carriers with sine wave carriers for low-noise testing applications. The sine carrier can then subsequently be amplitude modulated with different modulation depths to test the capabilities of measuring devices or algorithms.

Equation (C.1) yields a modulated carrier $A(t)$ that allows the testing of the dynamic range of STIPA implementations in the modulation domain for different octave bands.

$$A_k(t) = \sin(2\pi c_k t)\sqrt{0,5\left(1+0,55\,m\left(\sin(2\pi f_{1k} t) - \sin(2\pi f_{2k} t)\right)\right)} \qquad \text{(C.1)}$$

where

k is the octave band index k=1…7;

fc_k is the carrier band centre frequency, in Hz;

t is the time variable, in s;

m is the specified m-value;

f_{1k} is the lower modulation frequency, in Hz, as per Table B.1;

f_{2k} is the higher modulation frequency, in Hz, as per Table B.1.

Manufacturers of STI implementations shall report the specified m-value and the derived m-values for the range of 0,0 to 1,0 in 0,1 steps for each octave band. The absolute value of the error between the derived and the theoretical m-value shall not exceed 0,05. Overall m-value errors shall not yield a systematic absolute error (offset) in the STI results greater than 0,01.

m-values shall be reported for the highest and the lowest possible measuring levels of the measuring device meeting the criteria.

NOTE The levels of each octave band are identical.

C.3.3 Modulation depth testing for STI indirect method

The indirect method derives the m-values from the impulse response. Equation (C.2) yields an exponentially decayed sine wave carrier that is used as a substitute impulse response for testing the dynamic range of the modulation domain. The decay of the function (C.2) is defined by the −60 dB reverberation time.

$$A(t) = \sin(2\pi\, fc_k\, t)\,1\,000^{-\frac{t}{RT_{60}}} \qquad \text{(C.2)}$$

where

k is the octave band index k = 1...7;

fc_k is the octave band centre frequency, in Hz;

t is the time variable, in s;

RT_{60} is the reverberation time, in s.

For a given reverberation time, the m-values derived from the impulse response generated using Equation (C.2) should match the theoretical m-values given by Equation (C.3).

$$m(f_m RT_{60}) = \frac{1}{\sqrt{1+\left[\dfrac{2\pi f_m RT_{60}}{\log\left(10^6\right)}\right]^2}} \qquad \text{(C.3)}$$

where

m is the m-value

f_m is the modulation frequency in Hz

RT_{60} is the reverberation time to −60 dB in s

log is the natural logarithm, $\log(10^6) \approx 13{,}8$.

Manufacturers of STI implementations shall report the specified m-value and the derived m-values for RT_{60} values of 0,125 s, 0,25 s, 0,5 s, 1 s, 2 s, 4 s and 8 s in each octave band. The absolute value of the error between the derived and the theoretical m-value shall not exceed 0,05. Overall m-value errors shall not yield a systematic absolute error (offset) in the STI results greater than 0,01.

C.4 Testing of cross-talk between octave-band filters

C.4.1 Flank attenuation slopes

Crosstalk that is present between the octave bands can influence the derived m-values in each band. Signals leaking from adjacent octave bands will be manifest as noise, which reduces the m-values in the observed octave band.

For crosstalk between bands to not cause undue corruption of the m-values, leakage should be lower than the amount of auditory masking produced by the octave band below the observed band. Leakage at that level would effectively increase the effective noise level representing the auditory masking by 3 dB.

Examination of the auditory masking functions shows that the steepest level-dependent masking slope of relevance is approximately 41 dB/octave. This maximum slope is derived by finding the lowest level in the 500 Hz band (48 dB), which produces a level approximately equal to the lowest reception threshold (6,5 dB at 1 000 Hz).

To fully accommodate the auditory masking functions, crosstalk between octave-band filters should be minimised by using filter slopes of at least 42 dB/octave.

C.4.2 Octave band filter testing – STIPA direct method

Octave-band filter slopes shall be checked using a 100 % modulated sine carrier in the observed band at a relative level of −41 dB compared to a non-modulated sine carrier in one of the adjacent octave bands. If the slope of the band-pass filters is exactly 41 dB/octave, a m-value of 0,5 will be obtained, corresponding to an SNR of 0 dB.

Since filters are likely to be asymmetric, both sides of the band-pass filter shall be investigated. The appropriate test signals for each observed octave band k are generated using Equations C.4 and C.5. Note that the level of the non-modulated sine signal is adjusted by 3 dB to compensate for the fact that it is a non-modulated signal.

$$A_k(t) = 10^{\left(-41/20\right)} \sin\left(2\pi fc_k t\right) \sqrt{0{,}5\left(1 + 0{,}55\left(\sin\left(2\pi f1_k t\right) - \sin\left(2\pi f2_k t\right)\right)\right)} + 10^{\left(-3/20\right)} \sin\left(2\pi fc_{k-1} t\right) \tag{C.4}$$

$$A_k(t) = 10^{\left(-41/20\right)} \sin\left(2\pi fc_k t\right) \sqrt{0{,}5\left(1 + 0{,}55\left(\sin\left(2\pi f1_k t\right) - \sin\left(2\pi f2_k t\right)\right)\right)} + 10^{\left(-3/20\right)} \sin\left(2\pi fc_{k+1} t\right) \tag{C.5}$$

where

k is the observed octave band index k = 1...7;

fc_{k} is the observed carrier band centre frequency, in Hz;

t is the time variable, in s;

$f1_{\mathrm{k}}$ is the lower modulation frequency, in Hz, as per Table B.1;

$f2_{\mathrm{k}}$ is the higher modulation frequency, in Hz, as per Table B.1;

$fc_{\mathrm{k-1}}$ is the lower octave non-modulated carrier frequency, in Hz;

$fc_{\mathrm{k+1}}$ is the higher octave non-modulated carrier frequency, in Hz.

The m-values in the observed octave band k shall be 0,5 ± 0,05 or higher, with a non-modulated sine carrier in the lower adjacent octave band or in the higher adjacent octave band with a relative level of 41 dB.

If an m-value lower than 0,5 ± 0,05 is obtained, the level of the observed modulated carrier shall be increased in 1 dB steps to a level where an m-value of 0,5 ±0,05 results. With the measured relative level, the corresponding sound pressure level as per Table A.1 is determined and reported as the sound pressure level below which the STI is underestimated.

NOTE Testing both sides of each octave-band filter requires non-modulated sine carriers in the 63 Hz and 16 kHz octave bands.

C.4.3 Performance verification files

A number of reference files to verify the performance of STI measuring equipment can be downloaded from a number of websites including the following URLs: www.aes.org/standards and www.stipa.info. [39] [40]

The reference files enable manufacturers and users to confirm STI performance under a range of simulated conditions.

Annex D
(informative)

Use of STI measuring devices

D.1 Overview

This annex illustrates how STI measuring devices are used in various test scenarios. Focus is on the practical use of the test instruments instead of the various theoretical aspects of STI testing, which are addressed elsewhere in this document. For the purposes of this annex, it is assumed that all measurements take place with the direct method, using the STIPA test signal.

D.2 STIPA characterises only the speech transmission channel

Although STIPA measurements are often referred to as "speech intelligibility measurements," this is not literally correct. The speech transmission index reflects how a transmission path affects speech intelligibility; it is a physical measure that does not take listeners and talkers into account, but simply characterises the transmission path or channel. This means that factors such as hearing loss, poor articulation and other (human) limitations are not considered.

In practice, this is often beneficial for users. For example, the supplier of a PA system that is being certified using STIPA does not have to worry about a poor speaking style or hearing impairments of the evaluators (or other factors out of your control) that could affect the outcome of the tests.

Standards for STI performance usually set performance limits that are based on the (often implicit) assumption that all talkers and listeners are "normal." However, this has a potential drawback: it means that expectations based on STIPA measurements can be optimistic if, for example, large populations of hearing-impaired people need to be addressed, or if announcements are made using an accented voice or with a poorly-articulated speaking style. In those cases, performance limits should be set to higher STI values to ensure sufficient subjective intelligibility is provided.

In other words, STIPA measurements only indicate what the *speech transmission channel* does to the speech in terms of intelligibility. Before commencing, the user should consider the structure of the speech transmission channel that is to be tested.

In the context of STIPA testing, the term "speech transmission channel" is used in a broader sense than (for instance) in telecommunications engineering. The term "channel" suggests to some that electronic equipment (e.g. for radio transmission) is used, which is not necessarily the case in STIPA situations.

Figure D.1 shows the definition of the speech transmission channel: essentially everything that influences intelligibility, except for the talkers and listeners themselves.

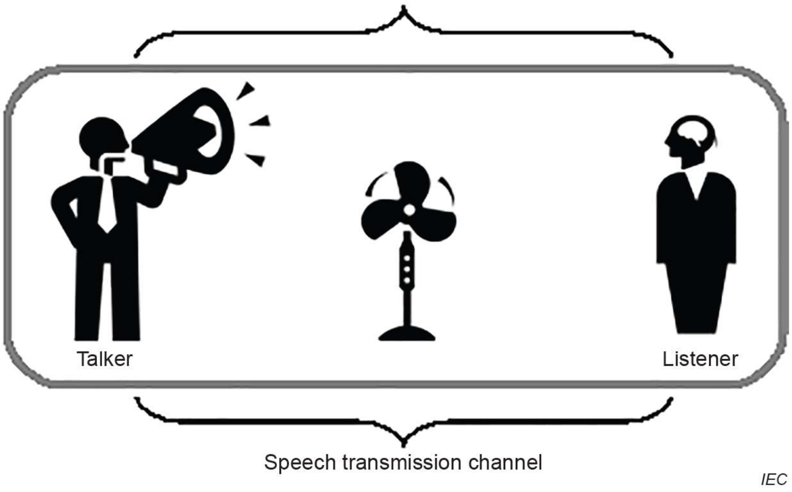


Figure D.1 – Schematic representation of the definition of a speech transmission channel

NOTE The channel comprises everything between the talker and listener that influences intelligibility, including noise sources and the acoustics of the environment, except for the talker and listener themselves.

In Figure D.1, the fan symbolizes a noise source interfering with speech from the talker. The talker and listener occupy the same space for which the acoustic properties (determined by wall materials, ceilings, etc.) will affect intelligibility. This is also considered by the STI. The horn used by the talker represents the use of electro-acoustic devices. Such devices, if present in the transmission channel, can introduce non-linear distortion components which are also considered by the (direct) STI method.

When performing STIPA tests:

- the talker is replaced by a source of the test signal;
- the listener is replaced by a STIPA analyser.

D.3 Examples of test scenarios for STIPA tests

Perhaps the most common application of the STIPA test method is to evaluate PA systems – hence the "PA" in STIPA. A few common STIPA testing scenarios are presented in Table D.1 for which the transmission channel and talkers and listeners are identified, together with the factors that would be expected to influence the STI. Table D.2 considers the case with pre-recorded announcements, whilst Table D.3 and Table D.4 consider scenarios for live meetings and lectures respectively.

Keep in mind that during STIPA tests, all talkers are replaced by a source of the STIPA test signal and all listeners (and listener locations) are measuring positions where the STIPA analyser is used.

Table D.1 – Scenario 1, PA with "live" announcer

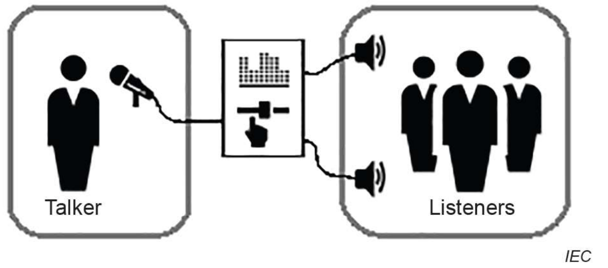
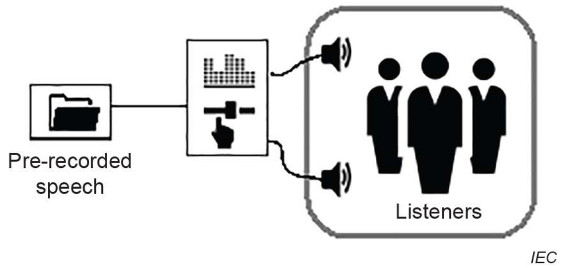
	
Talker	The talker is usually a single person making announcements, who might (or might not) have been trained for this purpose. The talker is usually out of range of the PA system and unable to directly hear the announcements.
Listeners	Listeners are all the people present in the venue for whom the announcements are intended. Listener locations to be considered are all spaces where the public is allowed.
Speech transmission channel to be tested	Everything from the paging microphone (in its acoustic environment) to all relevant listener locations and includes sound-system electronics.
Factors influencing the STI	• Noise and reverberation at the talker location, • Paging microphone characteristics and speaking distance • Frequency response and distortion of the sound system • Noise and reverberation at the listener location • Overall sound pressure level produced by the sound system

Table D.2 – Scenario 2, PA with pre-recorded announcements

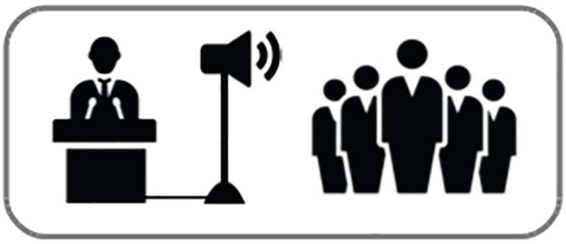
	
Talker	Instead of a live talker, recorded speech is used.
Listeners	Listeners are all the people present in the venue for whom the announcements are intended. Listener locations to be considered are all spaces where the public is allowed.
Speech transmission channel to be tested	Everything from (digital) audio storage and playback electronics up to all relevant listener locations and includes sound-system electronics.
Factors influencing the STI	• Frequency response and distortion of the sound system • Noise and reverberation at the listener location • Overall sound pressure level due to PA

Scenario 1 and Scenario 2 represent the most common public address and voice evacuation scenarios. Other common scenarios are shown below. Scenario 3 is the "classic" application of the speech transmission index to pure room acoustics, without the involvement of electronics for sound reproduction. STIPA evaluations can be very useful in identifying the impact of factors relating to room acoustics (e.g. lack of acoustic absorption materials) and ambient noise (e.g. due to air-conditioning systems) on speech intelligibility. Scenario 4 is a typical lecture-type situation, where one lecturer speaks to a larger number of listeners in the same room.

Table D.3 – Scenario 3, "live" meetings and conversations

	 IEC
Talker/listeners	In meetings and conversations, the same people take turns acting as talkers and listeners. All positions around a meeting table are therefore to be considered as talker positions as well as listener positions.
Speech transmission channel to be tested	Each individual talker and listener position combines into a transmission channel.
Factors influencing the STI	• Distance between talker and listener • Reverberation in the meeting room • Ambient noise in the meeting room; interfering speech from adjacent rooms • Vocal effort; speaking levels (relaxed vs. raised voice)

Table D.4 – Scenario 4, lecture

	 IEC
Talker	A single lecturer usually addresses a room full of people. The talker position is at the lectern, using a fixed microphone, or a somewhat larger presentation area if a wireless microphone is used.
Listeners	All seats in the audience are regarded as listening positions. Generally, there are more seats than can realistically be covered by STIPA measurements. A selection of representative seats (which shall always cover the expected worst-case seats) shall be selected.
Speech transmission channel to be tested	Everything from the microphone up to all listener positions in the room.
Factors influencing the STI	• Noise and reverberation in the lecture hall • Microphone characteristics and speaking distance • Frequency response and distortion of the sound system and possible the influence of acoustic feedback • Overall sound pressure level generated by the sound system, which will differ from seat to seat.

D.4 Equipment and resources needed for a STIPA test

D.4.1 Availability of the test signal

Access to the test signal is indispensable, but not always easily obtained. For instance, the operator of a PA system at a shopping mall might not permit the playback of test signals during opening hours. Or, even worse, the complete blocking of ground-to-air radio communications used for air traffic control by playing 18 s of test signal is completely unacceptable. Therefore, the channel shall be available in a configuration that matches "normal operation."

D.4.2 A source of the STIPA test signal

In Scenario 2 (pre-recorded speech), the source of the test signal may be an audio file of the STIPA signal. In other cases, such as Scenarios 1 and 4 (Table D.1 to Table D.4), a talkbox or calibrated test loudspeaker is most likely the best option as a test signal source. It shall match, as closely as possible, the directivity of a human talker (see 8.2).

D.4.3 A STIPA analyser

A STIPA analyser is basically a combination of a microphone, pre-amplifier, analogue-to-digital converter and a combination of hardware and software to provide the processing needed to compute the STI. All of this can be integrated into a single device, or a combination of discrete hardware and software components can be used.

D.5 Steps in the overall procedure

Generally, the overall procedure of most STI measuring sessions comprises the following steps.

a) Planning: study technical documentation, blueprints and all other documents related to the project and the transmission channel
b) Measurement plan: draft a detailed measurement plan, in which all aspects of the measurement session are planned (equipment to use, calibration procedure, numbers and locations of measurements, etc.).
c) Calibration: make sure that all equipment is properly calibrated and tested prior to use
d) Set up the signal source: depending on what type of scenario is involved, choose a suitable signal source (e.g. file, audio player, talkbox). Configure the signal source for the correct sound pressure level (nominally 60 dB A-weighted at 1 m distance). Start the test signal at least 2 s before the start of the measurement and keep it playing.
e) Collect measurement data: following the measuring positions and grids laid out in the planning phase, carry out all measurements and record all data. It is usual to carry out multiple measurements for each location
f) Post-processing: use software tools (such as worksheets made available by manufacturers of measuring equipment) (or the process in Annex M) to carry out any necessary post-hoc operations on the STI data. For instance, if measurements were made in noise-free (unoccupied) conditions, add the ambient noise spectrum corresponding to the operational state into the MTF matrix by computation.
g) If the indirect method is used, the real speech level and spectrum shall be entered along with the real noise level.
h) Annex M provides an example of the method to process the data.
i) Report: describe the setup, the measurement results, and the conclusions.

Annex E
(informative)

Qualification of the STI and relationships with other speech intelligibility measures

E.1 Relationship between the STI and word/sentence scores

The relationships between the STI and various speech intelligibility measures are given in Figure E.1. The nonsense word score for equally balanced CVC words is obtained from [41]. The relation with PB words in the so-called "Harvard list" with binaural listening is taken from [42]. The relation with sentence intelligibility is based on SRT (speech reception threshold) results.

NOTE The STI-PB words relationship, presented in Figure E.1, has been updated and differs from previous editions of this document.

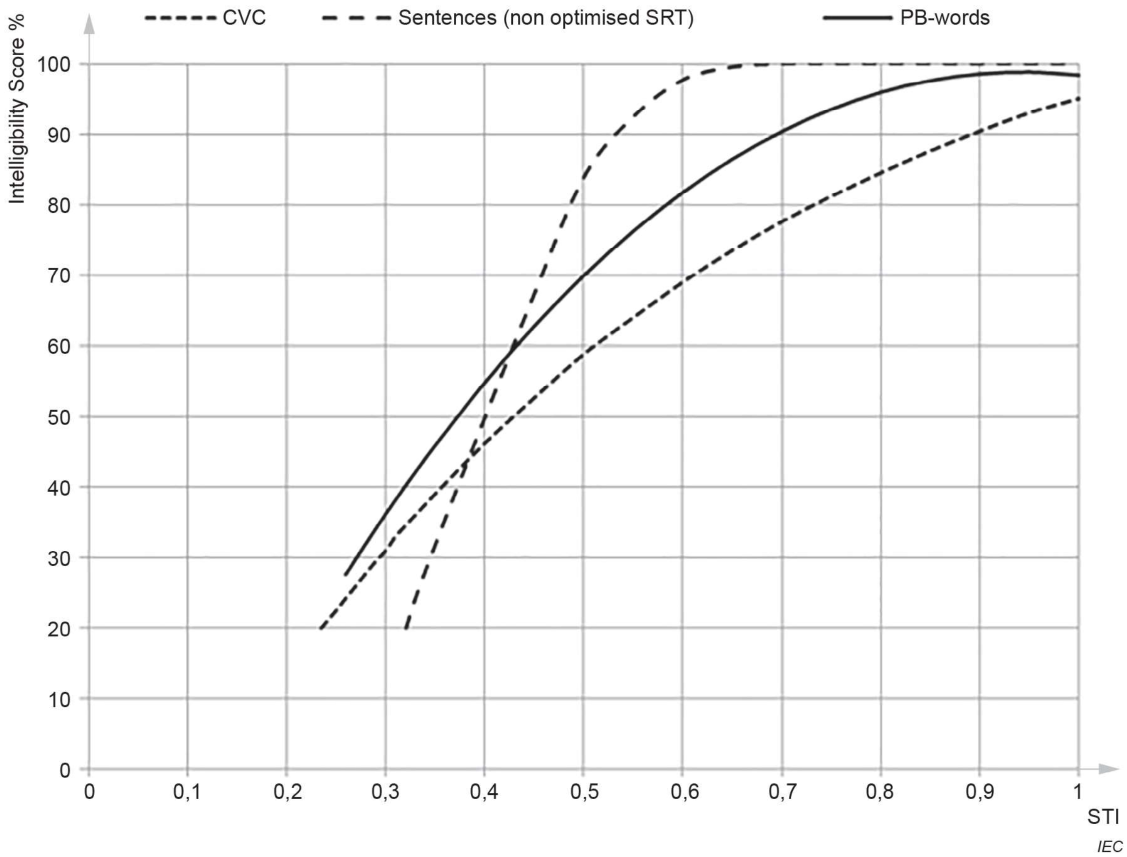


Figure E.1 – Relationships between some speech intelligibility measures

E.2 Relationship between STI and listening difficulty

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

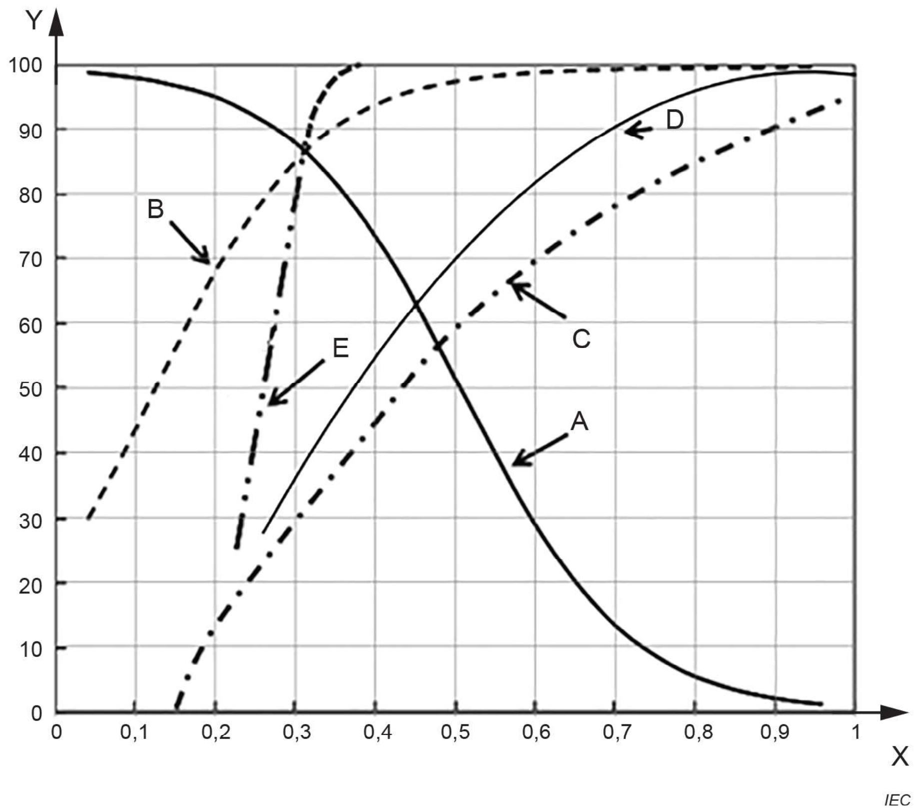
The listening difficulty rating is the percentage of responses indicating a certain degree of difficulty when listening to speech through a transmission channel. Note that listening difficulty ratings decrease for conditions with improved speech transmission, contrary to conventional intelligibility scores [43].

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

Table E.1 – Categories for listening difficulty

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

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



Key:

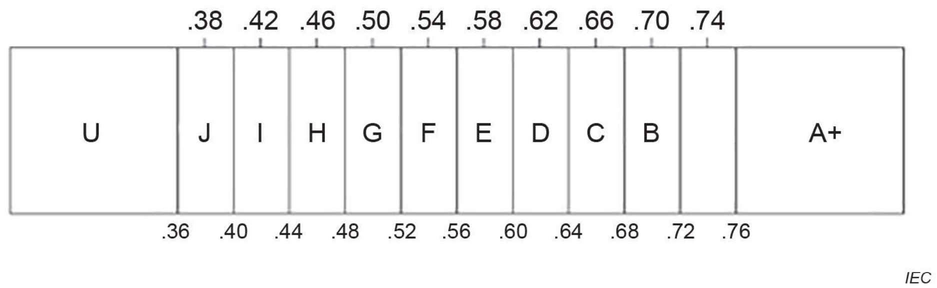
X axis is STI
Y axis is percentage of correct responses and listening difficulty
Curve A is listening difficulty
Curve B is word recognition
Curve C is CVC (EQB)
Curve D is PB-Words
Curve E is Sentences (SRT)

Figure E.2 – Relationship between STI, speech intelligibility scores and listening difficulty ratings [43], [44]

Annex F
(informative)

Nominal qualification bands for STI

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



Key

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

Row of letters: band designations;

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

Figure F.1 – STI qualification bands

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

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

Annex G
(informative)

Examples of STI qualification bands and typical applications

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

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

Category	**Nominal STI value**	**Type of message information**	**Examples of typical uses** (for natural or reproduced voice)	**Comment**
A+	> 0,76		Recording studios	Excellent intelligibility but rarely achievable in most environments
A	0,74	Complex messages, unfamiliar words	Theatres, speech auditoria, parliaments, courts, Assistive Hearing Systems (AHS)	High speech intelligibility
B	0,7	Complex messages, unfamiliar words		
C	0,66	Complex messages, unfamiliar words	Theatres, speech auditoria, teleconferencing, parliaments, courts	High speech intelligibility
D	0,62	Complex messages, familiar words	Lecture theatres, classrooms, concert halls	Good speech intelligibility
E	0,58	Complex messages, familiar context	Concert halls, modern churches	High quality PA systems
F	0,54	Complex messages, familiar context	PA systems in shopping malls, public buildings offices, VA systems, cathedrals	Good quality PA systems
G	0,5	Complex messages, familiar context	Shopping malls, public buildings offices, VA systems	Target value for VA systems
H	0,46	Simple messages, familiar words	VA and PA systems in difficult acoustic environments	Normal lower limit for VA systems
I	0,42	Simple messages, familiar context	VA and PA systems in very difficult spaces	Limited intelligibility
J	0,38		Not suitable for PA systems	
U	< 0,36		Not suitable for PA systems	

These values should be regarded as minimum target values.

NOTE 1 Perceived intelligibility relating to each category also depend on the frequency response at each listening position.

NOTE 2 The STI values refer to measured values in sample listening positions or as required by specific application standards.

Annex H
(informative)

Non-native listeners

Generally, compared to the intelligibility obtained with native listeners, non-native listeners require a 4 dB to 5 dB improvement of the signal-to-noise ratio for similar intelligibility (ISO 9921:2003 [45]).

Adjusted intelligibility qualification tables for three groups of non-natives can be defined, based on experimental data [46]. For this purpose, the non-native proficiency of the listener should be classified, based on language experience, age of learning and frequency of use of the second language.

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

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

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

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

For a non-native listener of category II and to achieve an intelligibility equivalent to an STI of 0,45, the transmission system needs to achieve a performance of 0,60. For intermediate values between the stated standard STI, interpolation should be used to estimate the adjusted STI.

NOTE For details on STI label categories, refer to ISO 9921.

Annex I
(informative)

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

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

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

NOTE 1 The speed of delivery of the speech has been found to have a large influence, for hearing-impaired persons, on intelligibility, and that is outside the scope of this document.

For listeners beyond 50 years old, hearing levels and the spread between individuals increases considerably [47]. Nevertheless, age-related hearing impairment shows good correlation between intelligibility and hearing loss.

As a rule of thumb, to reach intelligibility at the critical point of 50 % sentence intelligibility (where redundant sentences typically are be repeated twice), hearing impaired listeners require 4,5 dB higher SNR for 20 dBHL [48]. Here, dBHL refers to the hearing loss (HL) in dB, defined as the pure-tone average hearing level (PTA) of 0,5 kHz, 2,0 kHz and 4,0 kHz, relative to 18-year normal listeners.

NOTE 2 A 20 dB HL is a mild loss; hearing aids are often not required for HL values less than 35 dB. However, this does not apply to certain types of hearing loss, for which PTA is an inappropriate metric.

Table I.1 provides an indication of the adjusted intelligibility qualifications. It should be noted that the maximum intelligibility that can be reached depends on the degree of hearing loss [13], 48]. For more details, refer to other standards, such as [49].

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

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

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

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

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

Annex J
(normative)

Setting and adjustment of STI test signal level

J.1 Overview

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

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

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

The speech level measurement method was developed by Houtgast and Steeneken ([50]). In order to overcome spectral differences between signals and to have the signal levels closely match a perceived level, the measurement method is based on the A-weighted filtering of the signal. In general, speech signals are not continuous and contain numerous pauses. When specifying the signal-to-noise ratio of speech and noise signals based on the measurement of L_{Aeq} values, it is evident that the number and the duration of the silences between utterances affect the result. For example, speech at the identical nominal level but with a different pattern of pauses lead to a different L_{Aeq} value being measured and therefore to differences in the resultant signal-to-noise ratio.

The speech level measurement method deals with this phenomenon by removing all silences before calculating the level, such that only those parts of the speech signal which contain relevant signal information are taken into account.

The A-weighted signal is divided into frames of 10 ms to 20 ms in length and the energy per frame is calculated. Next, all the energy per frame values are accumulated in a level distribution histogram. Using the level distribution histogram, the RMS level of the speech is iteratively determined by cumulating all histogram data starting at a threshold that lies 14 dB below the calculated RMS level of the speech signal. Following this procedure, the relatively silent parts of the signal are removed allowing the signal level to become independent of the temporal distribution of the signal.

Extensive research [51] carried out on different speech level measures shows that the difference in RMS level of connected discourse and CVC words embedded in carrier phrases is minimal (< 0,5 dB) when applying the speech level measurement method.

J.3 Real speech level

For measurements on a PA system, the corrected (real) speech level is the level actually obtained from the system when working correctly at a specified reference position.

For measurements with a talker or other acoustic source, in the absence of a PA system, the A-weighted level of the modulated STIPA signal shall be set to 60 dB, measured at 1 m distance, on the axis of main radiation of the artificial mouth or talker. In practice, a STIPA A-weighted signal level of 66 dB measured at 500 mm helps to minimise the contribution of reverberation to the measured level.

If it is required to simulate a condition with a raised vocal effort (Lombard effect), the real speech A-weighted level shall be set to 70 dB.

If the test signal level needs to be adjusted to match the level of actual announcement, then the process described in Clause J.4 shall be used.

J.4 Corrected speech level derived from real speech level

Real speech levels can be approximately adjusted to provide the corrected speech level using a simple measurement of an A-weighted equivalent sound pressure level (L_{Aeq}). The measured level is adjusted by an empirically derived factor to obtain an estimate of the corrected speech level as described in Clause J.2. To obtain the approximate corrected speech level, use the following method:

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

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

J.5 Comparison of dynamic structures of speech and test signals

The dynamic characteristics of real speech and test signals can be very different and therefore the way in which they excite an electroacoustic system varies. Table J.1 compares the dynamic characteristics of a typical speech signal with that of a STIPA signal set to have the same equivalent L_{Aeq} value. As the table shows, speech has a much wider dynamic range.

Table J.1 – Typical speech and test signal dynamics

Signal	L_{Aeq}	L_{Apk}	L_{A1}	L_{A10}
Typical Speech (dB)	60,0	79,9	67,1	63,3
STIPA (dB)	60,0	72,5	61,8	61,0
Difference	0,0	7,4	5,3	2,3
NOTE L_{A1}, and L_{A10} are 1 %, 10 % percentile exceedance levels over the speech sample and are measured with a FAST time constant.				

Table J.2 shows these data in a different format that again illustrates the differences between typical speech and test signals. Further information can be found in [52].

Table J.2 – Comparison of speech and the test signal

Signal	Typical crest factor, dB		$L_{A1} - L_{Aeq}$, dB	$L_{A10} - L_{Aeq}$, dB
	A-weighted	C-weighted		
Typical speech	20,0	16,7	7,1	3,3
Pink noise	12,0	11,2	1,8	0,1
STIPA	12,4	11,6	1,8	1,0

Annex K
(informative)

Example test report sheet for STI measurements

Table K.1 and Table K.2 give a general set of guidelines and prompts to assist the measurement of STI over a range of applications. Not all categories are applicable in every case.

Table K.1 – Example test report sheet

GENERAL INFORMATION	
Measurement method	
Project / location	
Occupancy / configuration	
Date of measurement	
Edition of IEC 60268-16 used	
Method: Indirect (IR) or Direct (STIPA)	
Source	
Signal type: MLS; swept sine; STIPA; other	
Source type: Signal generator, CD player; .wav or .mp3 player	
Test signal spectrum: Adjusted to the standardized speech spectrum?	
Method of signal insertion: Electronic input or broadcast from an acoustic source?	
Details of test loudspeaker / artificial mouth / type	
Distance of source to system microphone (m)	
Directional pattern of sound system microphone	
Distance of microphone to nearest reflecting surface (m)	
System signal processing	
Status of system signal processing, e.g. compression, limiters, equalisation	
Was any part of the signal chain clipping during the measurement?	
Measurement hardware	
Brand/Type – serial no / version	
STI or STIPA loop back / direct connection test result	
Measurement microphone	
Brand, model and type (free field, random incidence)	
Monaural or binaural?	
Height above floor/ground (m)	
Aiming point of microphone	
Subjective impression with speech transmission	
Tonal characteristics: natural, muffled, boomy, resonant, harsh?	
Are there audible distortions or echoes?	
For acoustical signal insertion: Is there audible ringing or regeneration (feedback)?	

Table K.2 – Measurement data record sheet

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

Annex L
(normative)

Prediction of the STI using statistical methods

According to [35], the complete modulation transfer function, at modulation frequency f_m in octave band k including temporal distortion and noise can be written as:

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

where

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

$h_k(t)$ is impulse response of the transmission channel in octave band k;

f_m is the modulation frequency;

t is the integration variable for time;

ρ_k is the signal-to-noise ratio in dB.

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

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

where

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

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

r is the talker to listener distance;

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

T is the reverberation time of the room or space.

The above equation for the impulse response can be re-written as:

$$h(t) = I_\mathrm{D} \cdot \delta(t) + I_\mathrm{R}\, e^{-\frac{13{,}8t}{T}}$$

where

I_D is the intensity of the direct sound

I_R is the intensity of the reverberant sound

The modulation transfer function including temporal distortion and noise for a single source can then be written as:

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

with

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

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

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

where

$$SNR = 10\log \cdot \left(\frac{I_{\mathrm{s1m}}}{I_{\mathrm{N}}}\right)$$

and

I_{N} is the intensity of the noise;

I_{S1m} is the intensity of the source at 1 m.

The equations for terms A, B and C can be rewritten as:

$$A = I_{\mathrm{D}} + I_{\mathrm{R}}\left[1+\left(\frac{2\pi f_m T}{13{,}8}\right)^2\right]^{-1}$$

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

$$C = I_{\mathrm{D}} + I_{\mathrm{R}} + \left[1+\left(\frac{I_N}{I_{\mathrm{D}}+I_{\mathrm{R}}}\right)\right]^{-1}$$

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

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

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

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

Annex M
(informative)

Adjustments to STI data to simulate alternative ambient noise spectra and different speech levels

Annex M illustrates adjustments that can be made to measured and calculated STI data to simulate the effects of alternative levels of background noise and speech. The process is based on making corrections to the modulation transfer function (MTF).

In essence, the MTF or m-values are intensity ratios. Accordingly, signal and noise levels L that are obtained in dB are converted into intensities by applying $I = 10^{(L/10)}$.

The generalized form of adjustments to MTF data is implemented by multiplying each individual m-value by an appropriate intensity-ratio correction factor C. For example, to convert m_1, which equals the ratio $\frac{I_s}{I_s + I_{n1}}$, into $m_2 = \frac{I_s}{I_s + I_{n2}}$, m_1 is multiplied by the intensity-ratio correction-factor $C = \frac{I_s + I_{n1}}{I_s + I_{n2}}$.

The midpoint of the overall process is the derivation of an MTF matrix which is free of ambient noise and auditory-masking effects. In this state, the new MTF matrix provides the starting point for further processing. There is a range of starting points in the overall process, which depend on the assessment situation and data. The steps are described in Table M.1.

Table M.1 – Flow chart of post-processing adjustment steps

Step	Sub	Situation and action	Comments and further actions
1	a)	**Collect all available data**: – TI-values or the modulation transfer function (MTF) per octave band – The (acoustic) levels per octave band The amount of TI data or MTF data depends on the selected STI method or implementation. Full STI provides 14 values per octave band (in total 98 values), while STIPA yields 2 values per octave band (in total 14 values). There are generally seven octave band levels, one level for each octave band.	The MTF can contain contributions from many different sources such as linear and non-linear distortion, digital (codec) effects, reverberation, echo, ambient noise and auditory masking contributions etc. An STI value or an MTI (modulation transmission index per octave band) value should not be used to construct MTF data.
	b)	**Only the TI values are known** If only the TI values are known (and not the MTF), then convert TI values into m-values using the processing of Step 2.	Confirm the appropriate number of TI-values per octave band.
	c)	**The MTF data is fully processed MTF and contains auditory masking contributions.** If the MTF data is fully processed and MTF contains auditory masking contributions, then remove the auditory contributions from the MTF data using processing Step 3.	When performing this step, the acoustic levels per octave band should also be measured or predicted. If simulations or measurements were done with no acoustic reference at all (e.g. electrical measurements), then skip this step.
	d)	**The MTF data is free from auditory contributions but not free from ambient noise contributions**. If the MTF data is free from auditory contributions but not free from ambient noise contributions. then remove the ambient noise contributions from the MTF using processing Step 4.	This is likely to be the starting point when m-values are acquired from a measuring device and measured in the *presence* of ambient noise. To perform Step 4, the signal-to-ambient noise ratios shall be known or predicted.
	e)	**The MTF data is free from auditory contributions and free from ambient noise** If the MTF data is free from auditory contributions and free from ambient noise contributions, then use processing Step 5 to add contributions from an alternative ambient noise spectrum and new speech levels into to MTF.	This is likely to be the starting point when m-values are acquired from a measuring device in the *absence* of ambient noise. To perform Step 5, the signal-to-ambient-noise ratios shall be calculated.
	f)	**Compute the STI** Compute the STI using processing Step 6	Predicted or measured levels per octave band are needed to calculate auditory masking contributions.
2	**Convert TI values into an MTF**		Go to Step 3 if the MTF is already available
	a)	Calculate the effective signal-to-noise ratio from the T_{k,f_m} values for octave band k and modulation frequency f_m using the formula: $\rho_{\text{eff}\,k,f_m} = 30\ T_{k,f_m} - 15$	This step is used in situations where only transmission indices T_{k,f_m} are available.
	b)	Calculate the MTF_k from the effective signal-to-noise ratio $\rho_{\text{eff}\,k,f_m}$ (from Step 2a) using the formula: $m_{k,f_m} = \dfrac{1}{10^{\left(\frac{-\rho_{\text{eff}\,k,f_m}}{10}\right)} - 1}$	The MTF_k is the series of m_{k,f_m} values corresponding with octave band k. All MTF_k series from all octave bands form the so-called MTF matrix.

Step	Sub	Situation and action	Comments and further actions
3		**Removal of auditory contributions from e MTF**	If the m-values were obtained *before* auditory processing or were obtained electrically (and therefore do not contain any auditory contributions), then skip this step and go to Step 4.
	a)	Determine the acoustic octave band levels L_k in dB SPL for octave band k and convert them into intensities I_k. L_k is a combination of the received (deteriorated) test signal (including reflections) *and* the ambient noise level.	The octave band levels L_k is the sum of all acoustic sources for octave band k at the listener's ear or the measurement microphone. Keep the octave band intensities I_k for further processing.
	b)	By using Table A.1, compute the auditory masking slope La_k for octave band k using the acoustic level L_{k-1} from the lower octave band (k – 1).	Within the STI concept, octave band 125 Hz has no lower neighbouring masking octave band. Results for the 125 Hz octave band are therefore calculated *without* auditory masking contributions.
	c)	Compute the auditory masking factor a_k for octave band k based on the computed La_k using the formula: $$a_k = 10^{\left(\frac{La_k}{10}\right)}$$	Note that $amfdB_k$ always is a negative value.
	d)	Compute the auditory masking intensity Iam_k for octave band k using the auditory masking factor a_k from octave band k and the intensity I_{k-1} from the lower octave band k – 1 by using the formula: $$Iam_k = a_k \times I_{k-1}$$	Keep the auditory masking intensity Iam_k for octave band k for further processing.
	e)	Compute the intensities of the absolute reception thresholds Irt_k for octave band k using Table A.2	Keep the intensities of the absolute reception thresholds Irt_k in octave band k for further processing.
	f)	Compute the intensity ratio correction factor C_k, to be able to remove all auditory contributions from the MTF, using the formula: $$C_k = \frac{I_k}{I_k + Iam_k + Irt_k}$$	Use the stored intensities from Steps 3a, 3d and 3e.
	g)	Compute the MTF free of auditory contributions by multiplying the m-values with $\frac{1}{C_k}$ (from Step 3f) using the formula: $$m'_{k,fm} = m_{k,f_m} \times \frac{1}{C_k}$$	This processing step is explained as: $$m_{k,f_m} \times \frac{1}{C_k} =$$ $$\frac{Io_k}{I_k + Iam_k + Irt_k} \times \frac{I_k + Iam_k + Irt_k}{I_k}$$ $$= \frac{Io_k}{I_k} = m'_{k,fm}$$ where Io_k would be the received test signal intensity when free of any electro-acoustic contamination.
	h)	Keep the $m'_{k,fm}$ values for further processing steps	Continue with Step 4

Step	Sub	Situation and action	Comments and further actions
4	**Removal of the ambient noise contributions from the MTF**		Ensure that the MTF does not contain any auditory contributions, otherwise process the data using Step 3 first. If no ambient noise contributions are included in the MTF, skip this step and go to Step 5.
	a)	Determine the received STI signal levels Ls_k for octave band k and compute the corresponding STI signal intensities Is_k.	The received STI signal levels Ls_k represent the speech signal levels near the listener and can still contain contributions due to electro-acoustic effects such as non-linearities and reflections but not from the ambient noise. The speech signal levels might have been measured, predicted or determined based on signal-to-noise ratio information.
	b)	Determine the ambient noise levels Ln_k for octave band k and compute the ambient noise intensities In_k.	Ambient noise levels Ln_k might have been measured in the absence of the STI test signal or predicted.
	c)	Compute the intensity-ratio correction factor C_k to enable removal of the ambient noise contributions from the MTF using the formula: $$C_k = \frac{Is_k}{Is_k + In_k}$$	
	d)	Compute the MTF free of ambient noise contributions by multiplying the MTF (obtained in Step 3) with $\frac{1}{C_k}$ (from Step 4c) using formula: $$m''_{k,fm} = m'_{k,fm} \times \frac{1}{C_k}$$	
	e)	*Keep* the $m''_{k,fm}$ values for further processing steps	Continue with Step 5
5	**Add the new noise spectrum, new speech levels and auditory masking contributions to the MTF**		Ensure that the MTF does not contain any ambient noise or auditory contributions; if it does, process the MTF data with Steps 3 and 4 first.
	a)	Define the new STI signal levels Ls_k for octave band k and compute the corresponding STI test signal intensities Is_k.	The new received STI signal levels Ls_k represent the speech signal levels at the listener. These values can be based on an equalized version of the original STI signal levels.
	b)	Determine the new ambient noise levels Ln_k for octave band k and compute the ambient noise intensities In_k.	Ambient noise levels Ln_k might have been measured in the absence of the STI test signal or predicted.
	c)	Compute the intensity ratio correction factor C_k to add the new ambient noise level and new signal level contributions to the MTF using the formula: $$C_k = \frac{Is_k}{Is_k + In_k}$$	The intensity ratio correction factor C_k is also equivalent to: $$\frac{1}{1+10^{\left(-\frac{\rho_k}{10}\right)}}$$ where ρ_k is the signal-to-ambient noise ratio in dB in octave band k.

Step	Sub	Situation and action	Comments and further actions
	d)	Compute the new MTF including new ambient noise contributions by multiplying the MTF (obtained in Step 4) with C_k (from Step 5c) using the formula: $$m^{*}_{k,fm} = m''_{k,fm} \times C_k$$	NOTE If the octave band levels have no acoustic reference at all, for example when doing electrical STI measurements, then skip Steps 5e to 5k and go to processing Step 6.
5 cont.	e)	Compute the total acoustic level L_k and the acoustic intensity I_k for octave band k using the formulas: $$I_k = Is_k + In_k$$ $$L_k = 10 \times \lg 10\left(I_k\right)$$	The total acoustic octave-band level is needed for calculating auditory masking features. Keep the acoustic intensity for later use.
	f)	By using Table A.1, compute the auditory masking slope La_k for octave band k using the acoustic level L_{k-1} from the lower octave band (k – 1).	Within the STI concept, octave band 125 Hz has no lower neighbouring masking octave band. Results for the 125 Hz octave band are therefore calculated *without* masking contributions.
	g)	Compute the auditory masking factor a_k for octave band k based on the computed La_k using the formula: $$a_k = 10^{\left(\frac{La_k}{10}\right)}$$	Note that La_k is always a negative value.
	h)	Compute the auditory masking intensity Iam_k for octave band k using the auditory masking factor a_k from octave band k and the intensity I_{k-1} from the lower octave band k – 1 by using the formula: $$Iam_k = a_k \times I_{k-1}$$	*Keep* the auditory masking intensity Iam_k for octave band k for further processing.
	i)	Compute the intensities of the absolute reception thresholds Irt_k for octave band k using Table A.2	*Keep* the intensities of the absolute reception thresholds Irt_k for octave band k
	j)	Compute the intensity ratio correction factor C_k to add all auditory contributions to the MTF using the formula: $$C_k = \frac{I_k}{I_k + Iam_k + Irt_k}$$	Combine the results from Step 5e, 5h and 5i.
	k)	Compute the new MTF including auditory contributions by multiplying the MTF (obtained in Step 5d) with C_k (from Step 5j) using the formula: $$m^{**}_{k,fm} = m^{*}_{k,fm} \times C_k$$	Continue with Step 6

Step	Sub	Situation and action	Comments and further actions
6	**Compute STI using the adjusted MTF**		Get the adjusted MTF from Step 5
	a)	Convert adjusted MTF values into effective signal to noise ratio's $\rho\mathrm{eff}_{k,f_m}$ using the formula: $$\rho\mathrm{eff}_{k,fm} = 10 \times \log 10\left(\frac{m_{k,f_m}}{1-m_{k,f_m}}\right)$$	The adjusted MTF can be obtained from Step 5d (if electrical measurements of STI are to be used) or Step 5k (if acoustical measurements are to be used).
	b)	Truncate the $\rho\mathrm{eff}_{k,f_m}$ values to the range of −15 dB and +15 dB	
	c)	Convert the truncated $\rho\mathrm{eff}_{k,f_m}$ values into transmission indices TI_{k,f_m} using the formula: $$TI_{k,f_m} = \frac{\rho\mathrm{eff}_{k,fm} + 15}{30}$$	
	d)	Compute mean transmission indices MTI_k for octave band k using: $$MTI_k = \frac{1}{n}\sum_{m=1}^{n} TI_{k,f_m}$$	The value of n depends on the chosen STI method or implementation. For Full ST,I n equals 14, for STIPA, n equals 2.
	e)	Compute the STI value using the weight factors of Table A.1: $$STI = \sum_{k=1}^{7}\alpha_k \times MTI_k - \sum_{k=1}^{6}\beta_k \times \sqrt{MTI_k \times MTI_{k+1}}$$	Eventually, clip STI values higher than 1,00 to 1,00.

Table M.2 gives an example of how to adjust the measured STI values to simulate occupancy noise and different speech levels. Sections in which data is inserted are shaded in grey. Reference to some of the steps in Table M.1 is made. It should be noted that the calculations can vary slightly depending on the software used and rounding errors, but ultimately should be within 0,01 STI.

Table M.2 – Example calculation

1 Acquire raw MTF data with signal and noise levels							
	125 Hz	**250 Hz**	**500 Hz**	**1000 Hz**	**2000 Hz**	**4000 Hz**	**8000 Hz**
Signal level L_{eq}	73,0	70,0	72,0	68,2	62,2	56,2	52,0
Background noise levels L_{eq}	62,0	56,0	50,0	44,0	39,5	35,0	29,0
MTF matrix with noise, temporal effects, masking and threshold factors							
0,63 Hz	0,981 68	0,952 24	0,959 82	0,968 82	0,978 97	0,983 40	0,994 14
0,80 Hz	0,966 49	0,928 08	0,940 78	0,953 81	0,968 97	0,976 41	0,992 14
1,00 Hz	0,945 21	0,896 88	0,913 73	0,932 79	0,954 97	0,965 41	0,989 15
1,25 Hz	0,918 87	0,861 65	0,880 67	0,907 77	0,938 98	0,952 42	0,984 16
1,60 Hz	0,884 43	0,819 37	0,835 58	0,872 74	0,914 98	0,932 43	0,978 17
2,00 Hz	0,849 98	0,784 14	0,792 50	0,837 71	0,889 98	0,911 45	0,971 18
2,50 Hz	0,814 52	0,749 92	0,749 42	0,798 68	0,861 98	0,888 46	0,961 20
3,15 Hz	0,771 97	0,714 68	0,716 36	0,759 64	0,831 98	0,863 48	0,950 22
4,00 Hz	0,739 55	0,678 45	0,691 31	0,729 62	0,799 98	0,836 49	0,938 24
5,00 Hz	0,724 36	0,623 08	0,665 26	0,720 61	0,771 98	0,810 51	0,926 26
6,30 Hz	0,713 21	0,552 62	0,643 22	0,707 60	0,744 98	0,784 52	0,913 29
8,00 Hz	0,668 64	0,515 38	0,611 16	0,663 56	0,719 98	0,763 54	0,901 31
10,00 Hz	0,589 62	0,479 14	0,545 03	0,602 51	0,692 98	0,747 55	0,890 33
12,50 Hz	0,553 15	0,441 90	0,512 97	0,601 51	0,677 98	0,735 55	0,881 35
2 Remove background noise, auditory masking and reception threshold factors							
	125 Hz	**250 Hz**	**500 Hz**	**1 000 Hz**	**2 000 Hz**	**4 000 Hz**	**8 000 Hz**
$Is_{k(f)}$ – test signal only	19 952 623	10 000 000	15 848 932	6 606 934	1 659 587	416 869	158 489
$In_{k(f)}$ – noise only	1 584 893	398 107	100 000	25 119	8 913	3 162	794
Correction factor for ambient noise: $C1_k = Is_k / (Is_k + In_k)$ (see Step 4c)	0,926 41	0,961 71	0,993 73	0,996 21	0,994 66	0,992 47	0,995 01
Combined speech and noise level dB	73,332	70,170	72,027	68,216	62,223	56,233	52,022
Intensity of speech and noise I_k	21 537 516	10 398 107	15 948 932	6 632 053	1 668 499	420 032	159 284
Auditory masking factor La		-23,1340	-24,7152	-23,7863	-25,6918	-33,8884	-36,8836
Auditory masking factor a		0,00486	0,00338	0,00418	0,00270	0,00041	0,00020
$Iam,_k$	0,000	104 663,1	35 110,05	66695,63	17884,30	681,536	86,084
Reception Threshold $Lrt_{,k}$	46,0	27,0	12,0	6,5	7,5	8,0	12,0
$Irt,_k$	39 810,72	501,187	15,848 9	4,466 8	5,623 4	6,309 6	15,848 9
Correction factor for auditory masking & reception threshold: $C2_k = I_k / (I_k + Iam_k + Irt_k)$ (see Step 3f)	0,998 15	0,989 99	0,997 80	0,990 04	0,989 39	0,998 37	0,999 36
Inverse product of correction factors $C1_k$ & $C2_k$	1,081 43	1,050 33	1,008 53	1,013 90	1,016 15	1,009 24	1,005 66

3 Adjusted MTF matrix without noise, auditory masking and reception threshold							
0,63 Hz	1,061 62	1,000 16	0,968 00	0,982 29	0,994 78	0,992 49	0,999 76
0,80 Hz	1,045 19	0,974 79	0,948 80	0,967 06	0,984 62	0,985 43	0,997 75
1,00 Hz	1,022 18	0,942 02	0,921 52	0,945 75	0,970 40	0,974 33	0,994 74
1,25 Hz	0,993 69	0,905 01	0,888 18	0,920 39	0,954 14	0,961 22	0,989 72
1,60 Hz	0,956 44	0,860 61	0,842 71	0,884 87	0,929 75	0,941 05	0,983 70
2,00 Hz	0,919 19	0,823 60	0,799 26	0,849 35	0,904 35	0,919 86	0,976 67
2,50 Hz	0,880 85	0,787 66	0,755 81	0,809 78	0,875 90	0,896 67	0,966 63
3,15 Hz	0,834 83	0,750 65	0,722 47	0,770 20	0,845 41	0,871 45	0,955 59
4,00 Hz	0,799 78	0,712 59	0,697 20	0,739 76	0,812 90	0,844 22	0,943 55
5,00 Hz	0,783 34	0,654 44	0,670 93	0,730 63	0,784 45	0,817 99	0,931 50
6,30 Hz	0,771 29	0,580 43	0,648 70	0,717 43	0,757 01	0,791 77	0,918 45
8,00 Hz	0,723 08	0,541 32	0,616 37	0,672 78	0,731 61	0,770 59	0,906 41
10,00 Hz	0,637 63	0,503 25	0,549 68	0,610 88	0,704 17	0,754 45	0,895 37
12,50 Hz	0,598 19	0,464 14	0,517 35	0,609 87	0,688 93	0,742 35	0,886 33
4 Adjustment factors for operational speech and noise levels, auditory masking and reception threshold							
Operational speech L_{eq}	79,0	82,9	79,2	73,2	67,2	61,2	55,2
Operational background noise L_{eq}	67,0	70,0	72,0	68,0	65,0	63,0	60,0
$Is_{k(f)}$ – test signal only	79 432 823	194 984 460	83 176 377	20 892 961	5 248 075	1 318 257	331 131
$In_{k(f)}$ – noise only	5 011 872	10 000 000	15 848 932	6 309 573	3 162 278	1 995 262	1 000 000
Signal to noise ratio	12,00	12,90	7,20	5,20	2,20	-1,80	-4,80
Correction factor for Ambient Noise: $C3_k = Is_k / (Is_k + In_k)$ (see Step 5c)	0,940 65	0,951 22	0,839 95	0,768 05	0,624 00	0,397 84	0,248 76
Combined speech and noise level dB	79,27	83,12	79,96	74,35	69,25	65,20	61,24
Combined intensity I_k	84 444 696	204 984 46 0	99 025 309	27 202 535	8 410 352	3 313 519	1 331 131
Auditory masking factor La		-20,167	-18,241	-19,821	-22,627	-25,176	-29,535
Auditory masking factor a		0,009 622	0,014 992	0,010 420	0,005 461	0,003 037	0,001 113
$Iam_{,k}$	0,0	812 565,8	3 073 133, 4	1 031 856, 5	148 564,2	25 540,0	3 688,2
Reception threshold $Lrt_{,k}$	46	27	12	6,5	7,5	8	12
$Irt_{,k}$	39810,72	501,187	15,849	4,467	5,623	6,310	15,849
Correction factor for Auditory masking & Reception Threshold: $C4_k = I_k / (I_k + Iam_k + Irt_k)$ (see Step 5j)	0,999 53	0,996 05	0,969 90	0,963 45	0,982 64	0,992 35	0,997 23
Product of Correction Factors $C3_k$ & $C4_k$	0,940 21	0,947 46	0,814 67	0,739 98	0,613 17	0,394 80	0,248 07

5 Adjusted MTF matrix for operational speech and noise levels, auditory masking and reception threshold							
0,63 Hz	0,998 14	0,947 61	0,788 60	0,726 87	0,609 97	0,391 83	0,248 01
0,80 Hz	0,982 69	0,923 57	0,772 96	0,715 61	0,603 74	0,389 04	0,247 51
1,00 Hz	0,961 06	0,892 52	0,750 74	0,699 84	0,595 02	0,384 66	0,246 76
1,25 Hz	0,934 28	0,857 46	0,723 57	0,681 07	0,585 05	0,379 49	0,245 52
1,60 Hz	0,899 25	0,815 39	0,686 53	0,654 79	0,570 10	0,371 52	0,244 02
2,00 Hz	0,864 23	0,780 33	0,651 13	0,628 51	0,554 52	0,363 16	0,242 28
2,50 Hz	0,828 18	0,746 27	0,615 73	0,599 22	0,537 07	0,354 00	0,239 79
3,15 Hz	0,784 92	0,711 21	0,588 57	0,569 94	0,518 38	0,344 05	0,237 05
4,00 Hz	0,751 95	0,675 15	0,567 99	0,547 41	0,498 44	0,333 30	0,234 06
5,00 Hz	0,736 50	0,620 06	0,546 59	0,540 65	0,481 00	0,322 94	0,231 08
6,30 Hz	0,725 17	0,549 94	0,528 48	0,530 89	0,464 18	0,312 59	0,227 84
8,00 Hz	0,679 85	0,512 87	0,502 14	0,497 85	0,448 60	0,304 23	0,224 85
10,00 Hz	0,599 50	0,476 81	0,447 81	0,452 04	0,431 78	0,297 86	0,222 11
12,50 Hz	0,562 42	0,439 75	0,421 47	0,451 29	0,422 43	0,293 08	0,219 87
6 Process MTF matrix to yield STI							
6a, Convert m values into effective SNRs	**125 Hz**	**250 Hz**	**500 Hz**	**1000 Hz**	**2000 Hz**	**4000 Hz**	**8000 Hz**
0,63 Hz	27,298 70	12,574 16	5,717 56	4,250 96	1,942 14	-1,909 25	-4,817 45
0,80 Hz	17,541 12	10,822 25	5,320 57	4,007 63	1,828 72	-1,960 11	-4,829 06
1,00 Hz	13,923 34	9,192 91	4,788 26	3,676 49	1,670 96	-2,040 31	-4,846 49
1,25 Hz	11,527 53	7,792 79	4,178 97	3,294 94	1,491 97	-2,135 54	-4,875 63
1,60 Hz	9,506 54	6,451 07	3,404 59	2,780 14	1,225 77	-2,283 04	-4,910 73
2,00 Hz	8,038 30	5,505 05	2,710 06	2,283 57	0,950 89	-2,439 31	-4,951 87
2,50 Hz	6,830 47	4,685 27	2,047 63	1,746 81	0,645 23	-2,612 25	-5,010 98
3,15 Hz	5,622 14	3,914 19	1,555 02	1,222 92	0,319 49	-2,802 55	-5,076 49
4,00 Hz	4,816 58	3,177 17	1,188 48	0,826 05	-0,027 01	-3,011 04	-5,148 56
5,00 Hz	4,463 97	2,127 11	0,811 67	0,707 73	-0,330 23	-3,215 01	-5,221 27
6,30 Hz	4,213 79	0,870 39	0,495 25	0,537 27	-0,623 38	-3,422 43	-5,300 78
8,00 Hz	3,270 56	0,223 69	0,037 12	-0,037 37	-0,896 07	-3,592 70	-5,374 89
10,00 Hz	1,751 91	-0,403 10	-0,910 00	-0,835 66	-1,192 58	-3,724 21	-5,443 44
12,50 Hz	1,090 03	-1,051 78	-1,375 67	-0,848 82	-1,358 46	-3,823 90	-5,499 97
6b Truncate SNR to +/-15 dB							
0,63 Hz	15,000 00	12,574 16	5,717 56	4,250 96	1,942 14	-1,909 25	-4,817 45
0,80 Hz	15,000 00	10,822 25	5,320 57	4,007 63	1,828 72	-1,960 11	-4,829 06
1,00 Hz	13,923 34	9,192 91	4,788 26	3,676 49	1,670 96	-2,040 31	-4,846 49
1,25 Hz	11,527 53	7,792 79	4,178 97	3,294 94	1,491 97	-2,135 54	-4,875 63
1,60 Hz	9,506 54	6,451 07	3,404 59	2,780 14	1,225 77	-2,283 04	-4,910 73
2,00 Hz	8,038 30	5,505 05	2,710 06	2,283 57	0,950 89	-2,439 31	-4,951 87
2,50 Hz	6,830 47	4,685 27	2,047 63	1,746 81	0,645 23	-2,612 25	-5,010 98
3,15 Hz	5,622 14	3,914 19	1,555 02	1,222 92	0,319 49	-2,802 55	-5,076 49
4,00 Hz	4,816 58	3,177 17	1,188 48	0,826 05	-0,027 01	-3,011 04	-5,148 56
5,00 Hz	4,463 97	2,127 11	0,811 67	0,707 73	-0,330 23	-3,215 01	-5,221 27
6,30 Hz	4,213 79	0,870 39	0,495 25	0,537 27	-0,623 38	-3,422 43	-5,300 78
8,00 Hz	3,270 56	0,223 69	0,037 12	-0,037 37	-0,896 07	-3,592 70	-5,374 89

10,00 Hz	1,751 91	-0,403 10	-0,910 00	-0,835 66	-1,192 58	-3,724 21	-5,443 44
12,50 Hz	1,090 03	-1,051 78	-1,375 67	-0,848 82	-1,358 46	-3,823 90	-5,499 97
6c Convert SNRs to Transmission Indices $TI_{k(f)}$ and compute MTI values							
	125 Hz	**250 Hz**	**500 Hz**	**1000 Hz**	**2000 Hz**	**4000 Hz**	**8000 Hz**
0,80 Hz	1,000 00	0,860 74	0,677 35	0,633 59	0,560 96	0,434 66	0,339 03
1,00 Hz	0,964 11	0,806 43	0,659 61	0,622 55	0,555 70	0,431 99	0,338 45
1,25 Hz	0,884 25	0,759 76	0,639 30	0,609 83	0,549 73	0,428 82	0,337 48
1,60 Hz	0,816 88	0,715 04	0,613 49	0,592 67	0,540 86	0,423 90	0,336 31
2,00 Hz	0,767 94	0,683 50	0,590 34	0,576 12	0,531 70	0,418 69	0,334 94
2,50 Hz	0,727 68	0,656 18	0,568 25	0,558 23	0,521 51	0,412 92	0,332 97
3,15 Hz	0,687 40	0,630 47	0,551 83	0,540 76	0,510 65	0,406 58	0,330 78
4,00 Hz	0,660 55	0,605 91	0,539 62	0,527 53	0,499 10	0,399 63	0,328 38
5,00 Hz	0,648 80	0,570 90	0,527 06	0,523 59	0,488 99	0,392 83	0,325 96
6,30 Hz	0,640 46	0,529 01	0,516 51	0,517 91	0,479 22	0,385 92	0,323 31
8,00 Hz	0,609 02	0,507 46	0,501 24	0,498 75	0,470 13	0,380 24	0,320 84
10,00 Hz	0,558 40	0,486 56	0,469 67	0,472 14	0,460 25	0,375 86	0,318 55
12,50 Hz	0,536 33	0,464 94	0,454 14	0,471 71	0,454 72	0,372 54	0,316 67
MTI_k	**0,750 13**	**0,656 86**	**0,571 36**	**0,556 22**	**0,513 45**	**0,407 21**	**0,330 22**

6d Apply Weightings and calculate STI							
	125 Hz	**250 Hz**	**500 Hz**	**1000 Hz**	**2000 Hz**	**4000 Hz**	**8000 Hz**
alpha weighting factors, males	0,085 00	0,127 00	0,230 00	0,233 00	0,309 00	0,224 00	0,173 00
MTI_k × alpha weighting	0,063 76	0,083 42	0,131 41	0,129 60	0,158 65	0,091 22	0,057 13
beta weighting factors, males	0,085 00	0,078 00	0,065 00	0,011 00	0,047 00	0,095 00	
MTI_k × beta weighting	0,059 67	0,047 78	0,036 64	0,005 88	0,021 49	0,034 84	0,000 00
∑alpha*MTI	0,715 19						
∑beta*MTI	0,206 30						
STI	**0,509**						

Annex N
(informative)

Other methods of determining speech intelligibility

N.1 Overview

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

Another example of the first category (complex perceptual models) is the Speech Recognition Sensitivity model [53], which elegantly works around shortcomings of other models, but has not seen much "field experience" or independent evaluation. Complex models have also been developed to predict speech quality and intelligibility specifically for telecommunication channels (for example, the PESQ model [54], [55]). The added value of the STI, in relation to these models, is the wider applicability (room acoustics and telecommunications), combined with its widespread use and third-party evaluations. The fact that various vendors have implemented the STI method in their measuring devices helps in this respect.

The category of simpler metrics includes the Speech Interference Level (SIL) as described in ISO 9921, a metric that predicts intelligibility of speech in noise by averaging the speech-to-noise ratio in three octave bands. This second category also includes various measures based on early-to-late energy ratios derived from impulse responses, such as clarity and definition [56]. These are specifically of interest when investigating reverberant environments. Under the conditions and for the type of applications that these measures are intended for, their level of accuracy can approach that of the STI. In more complex situations, the accuracy of the STI outperforms all simpler metrics.

N.2 Word tests

The limitations of word tests are given in ISO/TR 4870 [57]. It should be noted that, because the method is based on the perception of words by listeners, there are no limitations with respect to the characteristics of the sound system or those of the environment. It is essential that the words are embedded in a carrier phrase in case of use in combination with temporal distortions (reverberation, echoes, automatic gain control).

N.3 Modified rhyme tests

The limitations are similar to those given in ISO/TR 4870 [57]. It should be noted that, because the method is based on the perception of words by listeners, there are no limitations in respect of the characteristics of the sound system or those of the environment.

N.4 Speech intelligibility index (SII)

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

NOTE SII does not appear to be significantly more sensitive than STI to the effects of frequency response [18].

N.5 PESQ

The perceptual evaluation of speech quality (PESQ) is especially useful for situations of high intelligibility and is based on mean opinion scores and especially suitable for measuring high quality transmission where speech intelligibility is less of an issue [54], [55].

Annex O
(informative)

Alternative direct methods for measuring Full STI

Historically, direct Full STI measurements have been made by testing all modulation frequencies one at a time, separately in each octave band. A Full STI measurement done this way involves 98 individual measurements (14 modulation frequencies, each in 7 octave bands). The signal generator cycles through the full matrix of all combinations of octave band and modulation frequency, following a pre-set pattern that is balanced to minimize potential errors due to time-variant effects.

Of all the methods to measure the speech transmission index, Full STI is the most generally applicable and most comprehensive method. The main disadvantage of direct Full STI is that it takes considerably longer to measure than other methods.

On measuring platforms with more advanced digital processing capabilities, it is feasible to shorten the measurement time of Full STI by simultaneously measuring multiple modulations, in a fashion similar to STIPA. STIPA measures 14 modulation frequencies at a time (instead of just one). To obtain the full matrix, this pattern can be repeated 7 times (each time with different modulation frequencies per band). In this case, the analyser synchronises with the test signal in order to switch over from one modulation pattern to the next at the right moment.

It is expected that a manufacturer implementing full STI following this approach indicates precisely the applied modulation patterns, modulation frequencies and time constants associated with the measurement. This information enables a user to determine whether full STI test signals and measuring devices are compatible between manufacturers.

Annex P
(normative)

Information to be provided by manufacturers

P.1 Purpose of this annex

Manufacturers of STI measuring solutions generally strive to implement the STI method in accordance with the requirements outlined in this document. However, since the document allows alternative design choices, the manufacturer shall make it unambiguously clear how the equipment was designed, and the restrictions and limitations that apply to its use.

It is expected that manufacturers will carefully validate their measuring solutions across a range of measuring conditions. While the exact form and extent of the validation testing that manufacturers undertake is their responsibility, manufacturers shall declare that their products have been validated to yield accurate results.

P.2 Form in which the information is to be provided

The manufacturer shall provide the information as described in this annex in written form, as part of the operating manual of every STI-measuring device offered by the manufacturer for which compliance with this document is claimed.

A measuring solution shall not be considered compliant with this document if the information specified in this annex is not provided. The information shall be provided as part of the operation manual for the STI measurement device or be easily accessible to the user, either in printed or digital form, and provided with the measuring instrument.

P.3 Required information

The following information is required:

- description of the device for which compliance is claimed (brand, serial number, description of alternative configurations);
- methods supported by the device (direct, indirect or both);
- test signals supported (STIPA, Full STI);
- table of modulation frequencies and octave bands supported, with precise modulation frequencies (rounded to 0,01 Hz);
- maximum tolerable deviation in test signal playback sample frequency (as a percentage, typically 0,1 % or less);
- supported revisions of IEC-60268-16 for backwards compatibility;
- indication as to whether or not a measurement mode enables level-dependent calculations to be disabled, for verification purposes and for making measurements on systems without an absolute sound pressure level reference.

P.4 Declaration

The manufacturer shall declare that the measuring system has been manufactured in compliance with all specifications as described in this document. In addition, the manufacturer shall declare that due diligence has been observed in validating the measuring system, validating not only the accuracy of the measured STI across a range of relevant reference conditions, but also the accuracy of the modulation transfer function matrix.

Annex Q
(informative)

Effect of uncertainties of selected parameters on STI uncertainty

Q.1 STI calculation framework

Q.1.1 Overview

This annex briefly summarizes how the STI is computed in order to establish basic equations and variable names.

Q.1.2 Statistical MTF

In the statistical framework modulation transfer functions m_k are calculated based on the reverberation time T_k of the room. They depend on index m of the modulation frequency f_m and index k of the octave band.

$$m_{k,f_m} = \frac{1}{\sqrt{1+\left(\frac{2\pi f_m T_k}{13{,}8}\right)^2}} \tag{Q.1}$$

We can define the constant

$$\alpha_m = \frac{2\pi f_m}{13{,}8} \tag{Q.2}$$

Therefore

$$m_{k,f_m} = \left(1+\alpha_m^2 T_k^2\right)^{-\frac{1}{2}} \tag{Q.3}$$

Q.1.3 Corrections

The modulation transfer function (Q.1) is also subject to influences other than reverberation. The degradation of intelligibility due to auditory masking, reception thresholds, and ambient noise are accounted for as follows:

$$m'_{k,f_m} = m_{k,f_m} \frac{I_k}{I_k + I_{am,k} + I_{rt,k} + I_{n,k}} \tag{Q.4}$$

Here m'_{k,f_m} represents the degraded (corrected) MTF. The MTF value m_k is reduced by considering the effect of auditory masking $I_{am,k}$, the hearing threshold $I_{rt,k}$, as well as the background noise $I_{n,k}$ – all relative to the signal level I_k.

If these effects are negligible or already included in the given m_k value (due to having measured the MTF instead of calculated), then $m'_{k,f_m} = m_{k,f_m}$.

As another example, when considering a perfect transfer function $m_{k,f_m} = 1$ with noise of the intensity

$$I_{n,k} = 10^{-\frac{L_{SN}}{10}} I_k \tag{Q.5}$$

where L_{SN} is the signal-to-noise level in dB, then

$$m'_{k,f_m} = \frac{1}{1 + 10^{-\frac{L_{SN}}{10}}} \tag{Q.6}$$

Q.1.4 Effective SNR

The effective SNR is calculated as follows:

$$\rho_{\mathrm{eff},k,f_m} = 10 \lg \frac{m'_{k,f_m}}{1 - m'_{k,f_m}} \tag{Q.7}$$

Additionally, the effective SNR is limited to ±15 dB. However, these limits are only exceeded for values of m'_{k,f_m} greater than 0,97 or less than 0,03 which equate to reverberation times of less than about 0,5 s or more than about 8 s when not considering other factors of influence.

For many practical applications, these limits can therefore be neglected or considered second order effects. In particular, they have little influence on the uncertainty estimates that are considered in the following.

Q.1.5 Modulation transfer index (MTI)

Based on the effective SNR for each octave band k and modulation frequency f_m the modulation transfer index (MTI) is calculated by averaging over all modulation frequencies.

$$MTI_k = \frac{1}{n} \sum_{m=1}^{n} \frac{\rho_{\mathrm{eff},k,f_m} + 15}{30} \tag{Q.8}$$

Owing to the limits on the effective SNR, the resulting MTI value is limited to the range of 0 to 1.

Q.1.6 Speech transmission index (STI)

Now the STI can be calculated as follows:

$$STI = \sum_{k=1}^{7} \overline{\alpha}_k MTI_k - \sum_{k=1}^{6} \overline{\beta}_k \sqrt{MTI_k MTI_{k+1}} \tag{Q.9}$$

where $\overline{\alpha}_k$ and $\overline{\beta}_k$ are the redundancy weights for male speech.

Q.2 The effect of RT uncertainty on STI uncertainty

Q.2.1 General

In this text, the level of sensitivity of the STI is derived when considering small changes to the reverberation time (RT). Any corrections are ignored for the moment, so $m'_{k,f_m} = m_{k,f_m}$.

Q.2.2 Modulation transfer function

Given equation (Q.3), small changes of T_k translate to changes in m_k as follows:

$$\Delta m_k = \frac{dm_k}{dT_k}\Delta T_k \tag{Q.10}$$

with the first order derivative of m_k as a function of T_k:

$$\frac{dm_k}{dT_k} = -\alpha_m^2 T_k \left(1+\alpha_m^2 T_k^2\right)^{-\frac{3}{2}} \tag{Q.11}$$

This expression does not depend directly on the octave band k or on the respective frequency. This dependency is created only indirectly through the dependency of T on the frequency.

Q.2.3 Uncertainty in the STI

For the sake of simplicity, assume that the reverberation time is approximately independent of frequency, so that $MTI_k = MTI_j$ for all values of j, k in Equation (Q.8). This might not be exactly true in practice, but, for many applications, dependence on frequency should be a secondary effect with respect to the uncertainty.

Accordingly, the following can be written:

$$STI = \left(\sum_{k=1}^{7}\overline{\alpha}_k - \sum_{k=1}^{6}\overline{\beta}_k\right)MTI \tag{Q.12}$$

knowing that by definition $\sum_{k=1}^{7}\overline{\alpha}_k - \sum_{k=1}^{6}\overline{\beta}_k = 1$, we can insert and simplify as follows:

$$STI = \frac{1}{n}\sum_{m=1}^{n}\frac{15+10\lg\dfrac{m'_{f_m}}{1-m'_{f_m}}}{30} \tag{Q.13}$$

$$STI = \frac{1}{2}+\frac{1}{3n}\sum_{m=1}^{n}\lg\frac{m'_{f_m}}{1-m'_{f_m}} \tag{Q.14}$$

This means that the only variable is m'_{f_m} which in turn only depends on T. In order to calculate the uncertainty, the total derivative has to be determined:

$$\Delta STI = \frac{dSTI}{dT}\Delta T \tag{Q.15}$$

knowing that the STI consists of n separate transfer function contributions

$$\frac{dSTI}{dT} = \sum_{m=1}^{n} \frac{dSTI}{dm'_{f_m}} \frac{dm'_{f_m}}{dT} \tag{Q.16}$$

one for each modulation frequency. The derivatives of the STI can now be calculated:

$$\frac{dSTI}{dm'_{f_m}} = \frac{1}{42} \frac{d}{dm'_{f_m}} \left(\lg \frac{m'_{f_m}}{1-m'_{f_m}} \right) \tag{Q.17}$$

yielding

$$\frac{dSTI}{dm'_{f_m}} = \frac{1}{(42 \ln 10)} \frac{1}{m'_{f_m}\left(1-m'_{f_m}\right)} \tag{Q.18}$$

The derivatives for MTF from Equation (Q.11) and the STI from Equation (Q.18) can now be inserted into Equation (Q.16) in order to obtain a single function describing the effect of RT uncertainty.

Following standard GUM methods, calculate the uncertainty of the STI as the statistical average over an ensemble of STI values whose spread is defined by the uncertainty of T.

$$u(STI) = \sqrt{\left\langle \Delta STI^2 \right\rangle - \left\langle \Delta STI \right\rangle^2} = \sqrt{\left\langle \frac{dSTI}{dT}^2 \Delta T^2 \right\rangle - \left\langle \frac{dSTI}{dT} \Delta T \right\rangle^2} \tag{Q.19}$$

$$u(STI) = \left| \frac{dSTI}{dT} \right| \sqrt{\left\langle \Delta T^2 \right\rangle - \left\langle \Delta T \right\rangle^2} \tag{Q.20}$$

$$u(STI) = \left| \frac{dSTI}{dT} \right| u(T) \tag{Q.21}$$

Inserting yields

$$u(STI) = \left| \sum_{m=1}^{n} \frac{1}{(42 \ln 10)} \frac{-\alpha_m^2 T \left(1+\alpha_m^2 T^2\right)^{-\frac{3}{2}}}{m'_{f_m}\left(1-m'_{f_m}\right)} \right| u(T) \tag{Q.22}$$

which can be simplified to

$$u(STI) = \left| \frac{T}{(42 \ln 10)} \sum_{m=1}^{n} \frac{\alpha_m^2 m_{f_m}^2}{1 - m_{f_m}} \right| u(T) \tag{Q.23}$$

From this expression, the results in Figure Q.1 below are obtained, which show how the uncertainty of the RT affects the uncertainty of the STI as a function of the absolute reverberation time.

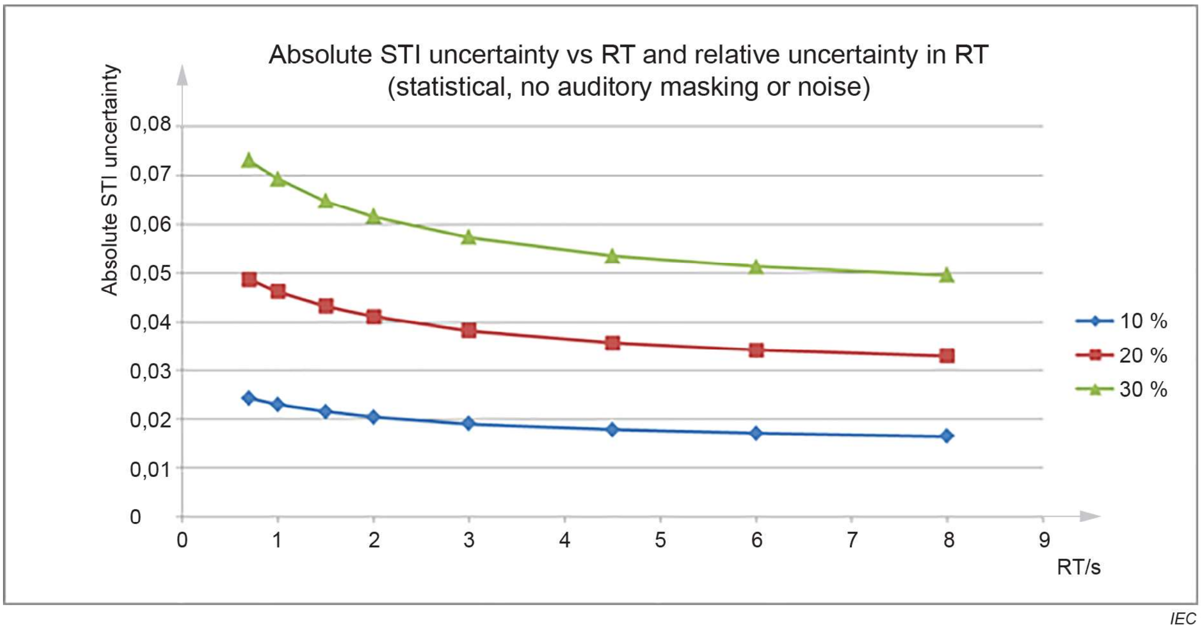


Figure Q.1 – Uncertainty in absolute value of STI vs reverberation time RT with various degrees of uncertainty in RT

Q.2.4 Conclusions:

- Larger RT uncertainties (≥ 15 %) lead to significant STI uncertainty (> 0,03).
- Uncertainty is dependent on RT and is more critical for shorter RTs (however, shorter RTs are usually less of a problem in practice).
- This matters for modelling input data (calculated or estimated RT) as well as when measurements are taken at too few locations, such that the locally measured RTs are not consistent with the majority of non-measured locations.

Q.3 The effect of S/N uncertainty on STI uncertainty

Q.3.1 General

In this text, the sensitivity of the STI is derived against small changes of the signal-to-noise level.

Q.3.2 Ideal transfer function

First consider an ideal transfer function namely $m_{k,f_m} = 1$. Given Equation (Q.6) assuming broadband background noise with the signal-to-noise factor $\sigma = 10^{\frac{L_{SN}}{10}}$ we have

$$m'_{k,f_m} = \frac{1}{1+\frac{1}{\sigma}} \tag{Q.24}$$

Then the STI equation (Q.14) simplifies to

$$STI = \frac{1}{2} + \frac{1}{3}\lg\sigma = \frac{1}{2} + \frac{1}{30}L_{\mathrm{SN}} \tag{Q.25}$$

It is immediately clear that a change of 1 dB in the signal-to-noise level corresponds to a change of 0,033 in STI. This is illustrated in Figure Q.2.

$$\Delta STI = \frac{dSTI}{dL_{\mathrm{SN}}}\Delta L_{\mathrm{SN}} = \frac{1}{30}\Delta L_{\mathrm{SN}} \tag{Q.26}$$

The same is true for the uncertainty

$$u(STI) = \left|\frac{dSTI}{dL_{\mathrm{SN}}}\right| u(L_{\mathrm{SN}}) = \frac{1}{30}u(L_{\mathrm{SN}}) \tag{Q.27}$$

This is true as long as approximately −15 dB ≤ L_{SN} ≤ 15 dB.

Q.3.3 Reverberation

Given Equations (Q.4) and (Q.6) with a signal-to-noise factor $\sigma = 10^{\frac{L_{\mathrm{SN}}}{10}}$ we have

$$m'_{k,f_m} = m_{k,f_m}\frac{1}{1+\frac{1}{\sigma}} \tag{Q.28}$$

The derivative is

$$\frac{dm'_{k,f_m}}{d\sigma} = m_{k,f_m}\frac{1}{(1+\sigma)^2} \tag{Q.29}$$

When only considering uncertainties in S/N, i.e. unknown changes in σ, the differential for the STI is

$$\Delta STI = \frac{dSTI}{d\sigma}\Delta\sigma \tag{Q.30}$$

$$\frac{dSTI}{d\sigma} = \sum_{m=1}^{n}\frac{dSTI}{dm'_{f_m}}\frac{dm'_{f_m}}{d\sigma} \tag{Q.31}$$

This assumes that the S/N, the RT and the S/N uncertainty are approximately constant over frequency. Finally, we have for the uncertainty:

$$u\left(STI\right)=\left|\frac{dSTI}{d\sigma}\right|u\left(\sigma\right) \quad (Q.32)$$

Notice that this error estimate is only valid as long as the ±15 dB limit approximately holds.

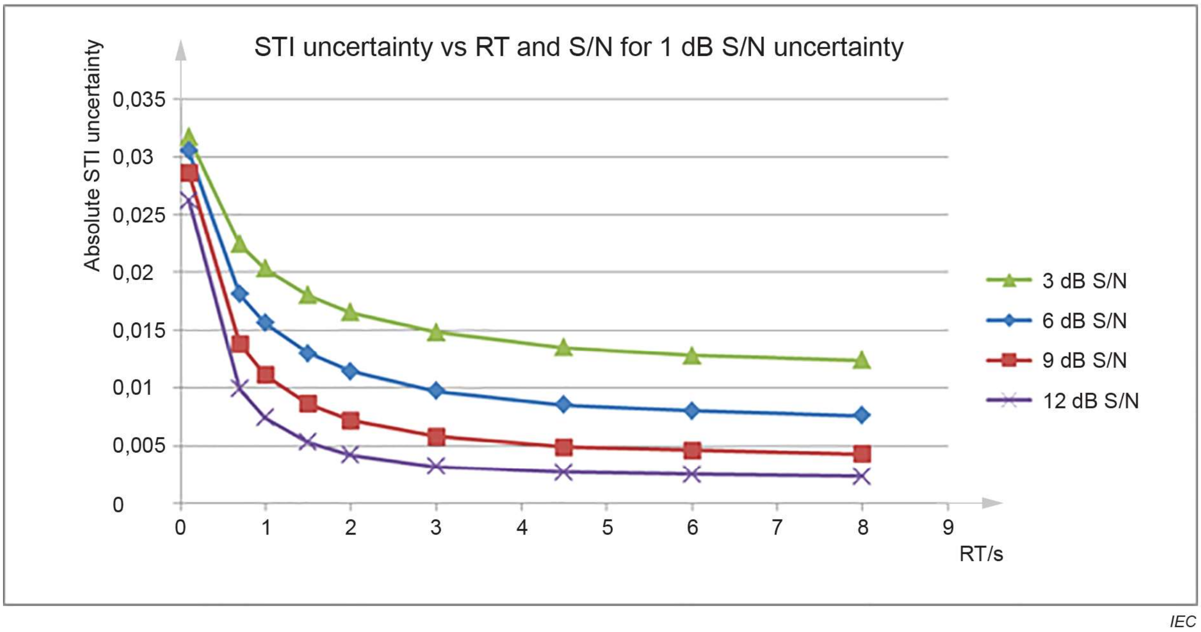


Figure Q.2 – Uncertainty in absolute value of STI vs reverberation time RT with 1 dB uncertainty in SNR at various SNRs

Q.3.4 Conclusions:

- For RT near 0, the curves approach the expected upper bound uncertainty of 0,033/dB.
- A non-zero RT lowers this sensitivity measurably.
- Again, the uncertainties are largest where the RT is lowest. But low RT values generally imply a more controlled environment, so the increase is not as critical.
- For more than about 6 dB S/N and typical RT ≥ 2, the absolute RT value is less important for the uncertainty of the STI.
- For practical 3 dB to 6 dB S/N, the uncertainty can rise significantly for shorter RTs.
- Assuming that most RT measurements are accurate within about ±10 % and level measurements within perhaps about ±1 dB, the uncertainty in the RT is more important than the uncertainty of the S/N level (roughly 0,02 versus 0,01).

Q.4 The effect of signal level uncertainty on STI uncertainty

Q.4.1 Overview

This text derives the sensitivity of the STI to small changes of the signal level measurements.

Q.4.2 Auditory masking

Given Equation (Q.4) and knowing $I_{am,k} = I_{k-1}a$ for k > 1, the auditory masking is accounted for as follows:

$$m'_{k,f_m} = m_{k,f_m} \frac{I_k}{I_k + I_{k-1}a} \tag{Q.33}$$

where the auditory masking factor a depends on the level $L_{k\text{-}1}$ of the masking band $k - 1$, but at maximum $a = 0{,}1$ for $L_{k\text{-}1} \geq 100$ dB. This worst case is considered in the following.

By defining the intensity ratio $r_k = \frac{I_k}{I_{k-1}}$ the expression (Q.33) becomes very similar to the S/N term used earlier:

$$m'_{k,f_m} = m_{k,f_m} \frac{1}{1 + a\frac{1}{r_k}} \tag{Q.34}$$

Note that there is no correction for $k = 1$, i.e. $a = 0$. For $k > 1$ the derivative is

$$\frac{dm'_{k,f_m}}{dr_k} = m_{k,f_m} \frac{\frac{1}{a}}{\left(1 + \frac{1}{a} r_k\right)^2} \tag{Q.35}$$

When only considering an uncertainty in the level difference (i.e. intensity ratio), the uncertainty for the STI is

$$\Delta STI = \frac{dSTI}{dr} \Delta \tag{Q.36}$$

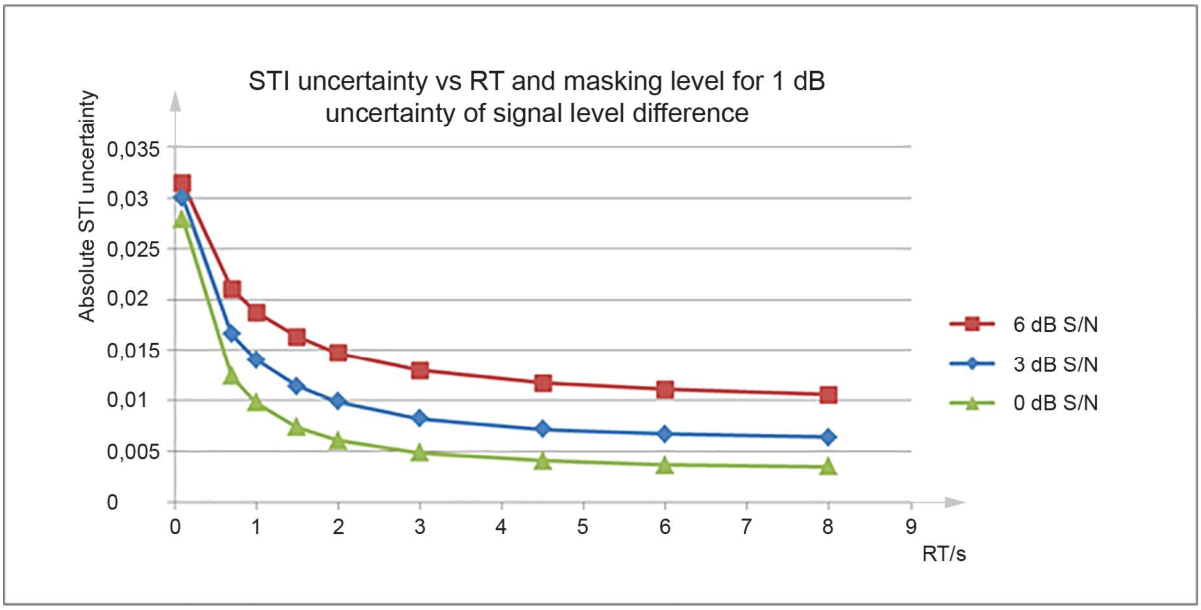
$$\frac{dSTI}{dr} = \sum_{m=1}^{n} \frac{dSTI}{dm'_{f_m}} \frac{dm'_{f_m}}{dr} \tag{Q.37}$$

This assumed that the RT and the uncertainty of the intensity ratio as well as signal level differences are approximately constant over frequency. A lower bound for m is also taken by assuming that the lowest octave band is affected by masking in the same way as the other bands, thus setting $m_1 = m_2 = m_k$. Finally, we have for the uncertainty:

$$u(STI) = \left| \frac{dSTI}{dr} \right| u(r) \tag{Q.38}$$

Expectedly, the resulting curves are similar to the S/N curves. In fact, a curve for a certain masking level is equivalent to the S/N curve of the same level when negated and increased by 10 dB. Therefore, 6 dB masking corresponds to 4 dB SNR, 3 dB masking relates to 7 dB SNR (all at a maximum auditory masking of $a = 0{,}1$). In particular, the same asymptotic limit is also obtained for vanishing T: These relationships are illustrated in Figure Q.3.

$$u(STI) = \left| \frac{dSTI}{dL_{\text{Mask}}} \right| u(L_{\text{Mask}}) = \frac{1}{30} u(L_{\text{Mask}}) \tag{Q.39}$$



IEC

Figure Q.3 – Uncertainty in absolute value of STI versus reverberation time RT with various degrees of masking.

Q.4.3 Conclusions

- The results in the diagram are based on certain assumptions:
  - Signal levels are 100 dB or higher throughout the bandwidth, thus the maximum *amf* is applied.
  - The masking level represents the differential signal level of a lower octave band relative to the next higher octave band. This difference is assumed constant over the entire audio bandwidth ("0 dB" would be pink noise, "6 dB" has a 6 dB decay per octave in the raw signal, i.e. similar to speech above 1 kHz).
- For RT > 0, the curves approach the expected upper bound of 0,033/dB.
- A non-zero RT lowers this sensitivity measurably.
- Again, the uncertainties are largest where the RT is lowest. But low RT values generally imply a more controlled environment, so the increase is not as critical.
- For less than about 6 dB of masking and typical RT ≥ 2 the absolute RT value is less important for the uncertainty of the STI.
- For exceptionally high masking levels, consistently greater than 6 dB per octave, the uncertainty can rise significantly for shorter RTs.
- Assuming that most RT measurements are about 10% accurate and level measurements perhaps about 1 dB, the uncertainty in the RT is more important than the uncertainty of the masking level (roughly 0,02 versus 0,01).

Bibliography

[1] ITU-T Recommendation P.51 (08/96) Objective measuring apparatus: Artificial Mouth

[2] IEC 60318-1:2009, *Electroacoustics – Simulators of human head and ear – Part 1: Ear simulator for the measurement of supra-aural and circumaural earphones*

[3] IEC 61672-1, *Electroacoustics – Sound level meters – Part 1: Specifications*

[4] Steeneken, H.J.M. and Houtgast, T., "Some applications of the Speech Transmission Index (STI) in auditoria", Acustica 51, 1982, p.229-234.

[5] Steeneken, H.J.M. and Houtgast, T., "A physical method for measuring speech transmission quality", J. Acoust. Soc. Amer. 67, 1980, 31, p.318-326.

[6] Houtgast, T. and Steeneken, H.J.M., "The modulation transfer function in room acoustics as a predictor of speech intelligibility", Acustica 28, 1973, p.66-73

[7] ANSI (1969). ANSI S3.5-1969, American National Standard, "Methods for Calculation of the Articulation Index" (American National Standards Institute New York)

[8] Steeneken, H.J.M. and Houtgast, T., "Mutual dependency of the octave-band weights in predicting speech intelligibility," Speech Communication 28, 1999, p.109-123.

[9] Steeneken, H.J.M. and Houtgast, T., "Phoneme-group specific octave-band weights in predicting speech intelligibility," Speech Communication 38, 2002, p.399-411.

[10] Houtgast, T. and Steeneken, H.J.M., "A Multi-lingual evaluation of the Rasti-method for estimating speech intelligibility in auditoria", Acustica 54, 1984, p.185-199

[11] Steeneken, H.J.M. and Houtgast, T., "The temporal envelope spectrum of speech and its significance in room acoustics", Proc. 11th International Congress on Acoustics, Paris, 1983, Vol. 7, p.85-88.

[12] Steeneken, H.J.M., Verhave, J.A., McManus, S., and Jacob, K.D., (2001) "Development of an Accurate, Handheld, Simple-to-use Meter for the Prediction of Speech Intelligibility", Proceedings IoA 2001, Reproduced sound (17). Stratford-upon-Avon, UK.

[13] Duquesnoy, A.J.H/M and Plomp, R. (1980) "Effect of reverberation and noise on the intelligibility of sentences in case of presbycusis". J. Acoust. Soc. Am. 68, p.537-544.

[14] Mapp, P, "Assessing the Potential Intelligibility of Assistive Audio Systems for the Hard of Hearing and Other Users", AES 124th convention, Amsterdam, 2008.

[15] Mapp, P, "Is STIPA a robust measure of speech intelligibility performance," AES 118th Convention, Barcelona, 2005.

[16] Mapp, P. Further thoughts on Speech Transmission Index (STI). IOA Reproduced Sound 18.Proc. IOA Vol 24 Pt 8. 2002 Mapp P, Practical application of STI to assessing Public Address and Emergency Sound Systems. Past, present and future of the Speech Transmission Index, TNO 2002

[17] Schroeder, M, Modulation Transfer Functions: Definition and Measurement, Acustica, 49, 1981

[18] Leembruggen, G. "Is SII better than STI at recognising the effects of poor tonal balance on intelligibility?" Proc IOA Vol 28 Part 6 2006

[19] Mapp, P Some Effects of Equalisation on Sound System Intelligibility and Measurement AES 115th Convention 2003 Preprint

[20] Leembruggen, G and A Stacey "Should the Matrix be reloaded Proc IOA Vol 28 Part 6 2006

[21] Steinbrecher, T, "Speech Transmission Index: Too weak in time and frequency?", Proc IOA Vol 30 Part 6 2008

[22] Mapp, P, "Some practical aspects of STI measurement & prediction", AES 134th convention, Rome, 2013, convention paper 8864

[23] Hammond, R, Mapp P, Hill A. "The influence of discrete arriving reflections on perceived intelligibility and Speech Transmission Index measurements." AES 141st Convention, Los Angeles, 2016, Convention Paper 9629

[24] Festen, JM & Plomp, R, Effects of fluctuating noise and interfering speech on the speech reception threshold for impaired and normal hearing, J. Acoust. Soc. Amer., 88(4) 1990

[25] Bozzoli, F, Farina, A, "Influence of artificial mouth's directivity in determining Speech Transmission Index"., AES 119th Convention, New York, 2005

[26] Wijngaarden, S, Verhave, J. Designing an acoustic source of the STIPA signal. Proc IOA Vol 38 Part 2 2016

[27] Mapp, P "Simulating Talker Directivity for Speech Intelligibility Measurements", AES 137th Convention, Los Angeles 2014

[28] IEC 60268-3:2000, Sound system equipment – Part 3: Amplifiers

[29] Rife, D., "Modulation Transfer Function Measurements with Maximum-Length Sequences", J. Audio Eng. Soc., Vol 40, Oct. 1992

[30] IEC 60118-4, Electroacoustics – Hearing aids – Part 4: Induction-loop systems for hearing aid purposes – system performance requirements

[31] ISO 7240-19:2007, Fire detection and alarm systems: Design, Installation, Commissioning and Service of Sound Systems for Emergency Purposes

[32] ANSI standard S12.2 1995, Criteria for evaluation of room noise

[33] Wijngaarden, S, Drullman, R, "Binaural intelligibility prediction based on the speech transmission index" J. Acoust. Soc. Amer. 123, 2008, p.4514-4523

[34] ITU-T P.58, *Head and torso simulator for telephonometry* (International Telecommunication Union, Geneva Switzerland 2011)

[35] Houtgast, T, Steeneken, H and Plomp, R. "Predicting Speech Intelligibility in Rooms from the Modulation Transfer Function. i. General Room Acoustics" Acustica Vol 46 1980

[36] Zwicker, E. and Feltkeller, "The ear as a communication receiver", Pub ASA 1999.

[37] Wijngaarden, S.J. van and Steeneken, H.J.M. (1999). "Objective prediction of speech intelligibility at high ambient noise levels using the speech transmission index." In Eurospeech99 – Proceedings of the 6th European Conference on Speech Communication and Technology, Budapest, Vol 6, p.2639-2642

[38] Leembruggen, G, Verhave J. et all. "The effect on STI results of changes to the male test-signal spectrum." Proc IOA Vol 38 Part 2 2016

[39] www.aes.org/standards

[40] www.stipa.info.

[41] Steeneken, H.J.M. and Houtgast, T., "Validation of the STIr method with the revised model," Speech Communication 38, 2002, p.413-425.

[42] Morales L, Li F. "A new verification of the speech transmission index for the English language" Speech Communication 105, 2018 1–11

[43] Hiroshi Sato, John S. Bradley, and Masayuki Morimoto, Using listening difficulty ratings of conditions for speech communication in rooms , J. Acoust. Soc. Am. 117, 1157 (2005)

[44] Hayato Sato, Masayuki Morimoto, Hiroshi Sato, and Megumi Wada, Relationship between listening difficulty and acoustical objective measures in reverberant sound fields, J. Acoust. Soc. Am. 123, 2087-2093 (2008)

[45] ISO 9921:2003, *Ergonomics – Assessment of speech communication*

[46] Wijngaarden, S.J., Steeneken, H.J.M. and Houtgast, T. (2002) "Quantifying the intelligibility of speech in noise for non-native listeners," J. Acoust. Soc Am. 112, p.3004-3013

[47] ISO-7029:2000, *Acoustics – Statistical distribution of hearing thresholds as a function of age*

[48] Sato,H, Kurakata,K, Mizunami,T, "Accessible speech messages for the elderly in rooms" 9th Western Pacific Acoustics Conference Seoul, Korea, 2006

[49] ISO/TR 22411:2008 Ergonomics data and guidelines for the application of ISO/IEC Guide 71 to products and services to address the needs of older persons and persons with disabilities.

[50] Past, present and future of the Speech Transmission Index, Pub TNO Human Factors Soesterberg Netherlands 2002

[51] Steeneken, H.J.M. and Houtgast, T (1986) Comparison of some methods for measuring speech levels Report IZF1986-20 TNO Institute for Perception Soesterberg Netherlands

[52] Mapp, P, "Some effects of speech signal characteristics on PA system performance and design", AES 139th convention, New York, 2015, convention paper 9477.

[53] H. Müsch and S. Buus. 'Using statistical decision theory to predict speech intelligibility. I. Model structure'. J.Acoust.Soc.Am. 109, 2896–2909. (2001).

[54] ITU-T Rec P.862, Perceptual evaluation of speech quality (PESQ)" An objective method of end-to-end speech quality assessment of narrow band telephone networks and speech codecs. (International Telecommunication Union, Geneva Switzerland 2001 Feb)

[55] J. G. Beerends, A. P. Hekstra, A. W. Rix and M. P. Hollier. 'PESQ, the new ITU standard for objective measurement of perceived speech quality, Part II – Perceptual model'. J. Audio Eng. Soc., vol. 50, pp. 765-778. (2002 Oct.).

[56] ISO 3382-1:2009, *Acoustics – Measurement of room acoustic parameters – Part 1: Performance spaces*

[57] ISO/TR 4870:1991, *Acoustics – The construction and calibration of speech intelligibility tests (Withdrawn)*

[58] IEC TS 60318-7:2011, *Electroacoustics – Simulators of human head and ear – Head and torso simulator for acoustic measurement of hearing aids*

____

This page deliberately left blank

British Standards Institution (BSI)

BSI is the national body responsible for preparing British Standards and other standards-related publications, information and services.

BSI is incorporated by Royal Charter. British Standards and other standardization products are published by BSI Standards Limited.

About us

We bring together business, industry, government, consumers, innovators and others to shape their combined experience and expertise into standards -based solutions.

The knowledge embodied in our standards has been carefully assembled in a dependable format and refined through our open consultation process. Organizations of all sizes and across all sectors choose standards to help them achieve their goals.

Information on standards

We can provide you with the knowledge that your organization needs to succeed. Find out more about British Standards by visiting our website at bsigroup.com/standards or contacting our Customer Services team or Knowledge Centre.

Buying standards

You can buy and download PDF versions of BSI publications, including British and adopted European and international standards, through our website at bsigroup.com/shop, where hard copies can also be purchased.

If you need international and foreign standards from other Standards Development Organizations, hard copies can be ordered from our Customer Services team.

Copyright in BSI publications

All the content in BSI publications, including British Standards, is the property of and copyrighted by BSI or some person or entity that owns copyright in the information used (such as the international standardization bodies) and has formally licensed such information to BSI for commercial publication and use.

Save for the provisions below, you may not transfer, share or disseminate any portion of the standard to any other person. You may not adapt, distribute, commercially exploit or publicly display the standard or any portion thereof in any manner whatsoever without BSI's prior written consent.

Storing and using standards

Standards purchased in soft copy format:

- A British Standard purchased in soft copy format is licensed to a sole named user for personal or internal company use only.
- The standard may be stored on more than one device provided that it is accessible by the sole named user only and that only one copy is accessed at any one time.
- A single paper copy may be printed for personal or internal company use only.

Standards purchased in hard copy format:

- A British Standard purchased in hard copy format is for personal or internal company use only.
- It may not be further reproduced – in any format – to create an additional copy. This includes scanning of the document.

If you need more than one copy of the document, or if you wish to share the document on an internal network, you can save money by choosing a subscription product (see 'Subscriptions').

Reproducing extracts

For permission to reproduce content from BSI publications contact the BSI Copyright and Licensing team.

Subscriptions

Our range of subscription services are designed to make using standards easier for you. For further information on our subscription products go to bsigroup.com/subscriptions.

With **British Standards Online (BSOL)** you'll have instant access to over 55,000 British and adopted European and international standards from your desktop. It's available 24/7 and is refreshed daily so you'll always be up to date.

You can keep in touch with standards developments and receive substantial discounts on the purchase price of standards, both in single copy and subscription format, by becoming a **BSI Subscribing Member.**

PLUS is an updating service exclusive to BSI Subscribing Members. You will automatically receive the latest hard copy of your standards when they're revised or replaced.

To find out more about becoming a BSI Subscribing Member and the benefits of membership, please visit bsigroup.com/shop.

With a **Multi-User Network Licence (MUNL)** you are able to host standards publications on your intranet. Licences can cover as few or as many users as you wish. With updates supplied as soon as they're available, you can be sure your documentation is current. For further information, email cservices@bsigroup.com.

Revisions

Our British Standards and other publications are updated by amendment or revision.

We continually improve the quality of our products and services to benefit your business. If you find an inaccuracy or ambiguity within a British Standard or other BSI publication please inform the Knowledge Centre.

Useful Contacts

Customer Services
Tel: +44 345 086 9001
Email: cservices@bsigroup.com

Subscriptions
Tel: +44 345 086 9001
Email: subscriptions@bsigroup.com

Knowledge Centre
Tel: +44 20 8996 7004
Email: knowledgecentre@bsigroup.com

Copyright & Licensing
Tel: +44 20 8996 7070
Email: copyright@bsigroup.com

BSI Group Headquarters

389 Chiswick High Road London W4 4AL UK