BS EN 60268-16:2011

BSI Standards Publication

# Sound system equipment

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



raising standards worldwide™

## National foreword

This British Standard is the UK implementation of EN 60268-16:2011. It is identical to IEC 60268-16:2011. It supersedes BS EN 60268-16:2003 which is withdrawn.

The UK participation in its preparation was entrusted to Technical Committee EPL/100, Audio, video and multimedia systems and equipment.

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

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



ISBN 978 0 580 76163 8

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 September 2011.

**Amendments issued since publication**

| Date | Text affected |
|---|---|

EUROPEAN STANDARD

NORME EUROPÉENNE

EUROPÄISCHE NORM

# EN 60268-16

August 2011

ICS 33.160.01

Supersedes EN 60268-16:2003

English version

## Sound system equipment - Part 16: Objective rating of speech intelligibility by speech transmission index
(IEC 60268-16:2011)

Equipements pour systèmes électroacoustiques - Partie 16: Evaluation objective de l'intelligibilité de la parole au moyen de l'indice de transmission de la parole
(CEI 60268-16:2011)

Elektroakustische Geräte - Teil 16: Objektive Bewertung der Sprachverständlichkeit durch den Sprachübertragungsindex
(IEC 60268-16:2011)

This European Standard was approved by CENELEC on 2011-08-02. 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 Central Secretariat 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 Central Secretariat 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, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland and the United Kingdom.

CENELEC

European Committee for Electrotechnical Standardization

Comité Européen de Normalisation Electrotechnique

Europäisches Komitee für Elektrotechnische Normung

**Management Centre: Avenue Marnix 17, B - 1000 Brussels**



Ref. No. EN 60268-16:2011 E

## Foreword

The text of document 100/1812/FDIS, future edition 4 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 was approved by CENELEC as EN 60268-16 on 2011-08-02.

This European Standard supersedes EN 60268-16:2003.

EN 60268-16:2011 includes the following technical changes with respect to EN 60268-16:2003:

— development of more comprehensive, complete and unambiguous standardization of the STI methodology;

— the term $STI_r$ is discontinued. A new function for the prediction of auditory masking effects is introduced;

— the concept of 'speech level' and the setting of the level of the test signal have been introduced;

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

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

The following dates were fixed:

| | | |
|---|---|---|
| – latest date by which the EN has to be implemented at national level by publication of an identical national standard or by endorsement | (dop) | 2012-05-02 |
| – latest date by which the national standards conflicting with the EN have to be withdrawn | (dow) | 2014-08-02 |

Annex ZA has been added by CENELEC.

## Endorsement notice

The text of the International Standard IEC 60268-16:2011 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:

| | | | |
|---|---|---|---|
| [2] | IEC 60318-1:2009 | NOTE | Harmonized as EN 60318-1:2009 (not modified). |
| [3] | IEC 61672 series | NOTE | Harmonized in EN 61672 series. |
| [37] | ISO 7029:2000 | NOTE | Harmonized as EN ISO 7029:2000. |
| [44] | ISO 3382-1:2009 | NOTE | Harmonized as EN ISO 3382-1:2009. |

# Annex ZA
(normative)

## Normative references to international publications with their corresponding European publications

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

NOTE When an international publication has been modified by common modifications, indicated by (mod), the relevant EN/HD applies.

| Publication | Year | Title | EN/HD | Year |
|---|---|---|---|---|
| IEC 61260<br>+ A1 | 1995<br>2001 | Electroacoustics - Octave-band and fractional-octave-band filters | EN 61260<br>+ A1 | 1995<br>2001 |
| ISO 18233 | 2006 | Acoustics - Application of new measurement methods in building and room acoustics | EN ISO 18233 | 2006 |

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: Multimedia equipment and systems.

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

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

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

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

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

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

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

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

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

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

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

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

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

# INTRODUCTION

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

The history of revisions is as follows.

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

  STIPA was specially developed as a fast measurement method that could deal with electro-acoustic and acoustic effects while determining the speech transmission quality of PA systems.
- Revision 4: 2010. The aim of Revision 4 (this revision) is to provide a more comprehensive, complete and unambiguous standardization of the STI methodology. The term $STI_r$ is now discontinued. A new function for the prediction of auditory masking effects is introduced.

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

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

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

**Items that have changed in this revision**

Specific changes that have been incorporated in this revision are:

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

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

**Potential applications of STI**

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

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

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

**Potential users of STI**

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

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

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

# SOUND SYSTEM EQUIPMENT –

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

## 1 Scope

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

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

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

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

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

The following information is included:

- measurement techniques;
- prediction techniques.

## 2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies:

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

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

## 3 Terms and definitions

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

**3.1**
**speech intelligibility**
rating of the proportion of speech that is understood

---

1 Figures in square brackets refer to the Bibliography.

**3.2**
**speech quality**
rating of sound quality of a speech signal

**3.3**
**speech transmission index**
**STI**
metric ranging between 0 and 1 representing the transmission quality of speech with respect to intelligibility by a speech transmission channel

**3.4**
**speech intelligibility index**
**SII**
objective method for prediction of speech intelligibility based on the Articulation Index

**3.5**
**STI method**
**FULL STI**
objective method for prediction and measurement of the speech transmission index that uses 14 modulation frequencies over a range of 7 octave bands

**3.6**
**distortion**
any unintentional and generally undesired change of the form of a signal occurring in a speech transmission channel

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

**3.7**
**speech transmission index for public address systems**
**STIPA**
method obtained by using a condensed version of the STI method but still responsive to distortions found in room acoustics and/or public address systems

NOTE STIPA is applied as a direct method.

**3.8**
**speech transmission index for telecommunication systems**
**STITEL**
method obtained by using a condensed version of the STI method but still responsive to distortions found in communication systems

NOTE STITEL is applied as a direct method.

**3.9**
**room acoustical speech transmission index**
**RASTI**
method obtained by using a condensed version of the STI method, to be used for screening purposes only and focused on direct communication between persons without making use of an electro-acoustic communication system

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

NOTE 2 RASTI is now obsolete.

**3.10**
**direct STI method**
method using modulated (speech like) test signals to directly measure the modulation transfer function

**3.11**
**indirect STI method**
method using the impulse response and forward energy integral (Schroeder integral) to derive the modulation transfer function

**3.12**
**speech transmission channel**
acoustic or electro-acoustic signal path between a talker and a listener

**3.13**
**public address system**
**PA**
electronic sound distribution system, employing microphones, amplifiers and loudspeakers, used to reinforce or amplify a given sound (such as an announcement or a pre-recorded message) and distributing the sound within a building or a space

**3.14**
**voice alarm system**
**VAS**
sound distribution system that broadcasts speech messages or warning signals, or both, in an emergency

**3.15**
**real speech level**
signal level of a speech signal in dB A where only the segments that contribute to the speech signal are taken into account; pauses and silences between words and sentences are ignored

NOTE See also Annex J.

**3.16**
**reference speech level**
speech level equivalent to 60 dB A at 1 m distance in front of the talker's mouth

**3.17**
**vocal effort**
exertion of the speaker, quantified objectively by the A-weighted speech level at 1 m distance in front of the mouth and qualified subjectively by a description

**3.18**
**artificial mouth**
device consisting of a loudspeaker mounted in an enclosure and having a directivity and radiation pattern similar to those of the average human mouth

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

**3.19**
**non-native speaker**
person speaking a language which is different from the language that was learned as primary language during the childhood of the speaker

**3.20**
**absolute speech reception threshold**
absolute threshold of hearing increased by the minimal required dynamic range for the correct recognition of speech

**3.21**
**auditory masking**
process by which the threshold of hearing (audibility) for one sound is raised by the presence of another (masking) sound

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

**3.22**
**artificial ear**
device with similar characteristics as the human ear for the reception of acoustic signals

NOTE See IEC 60318 [2].

**3.23**
**intensity function**
the squared amplitude signal as a function of time

**3.24**
**envelope function**
envelope of the intensity function

**3.25**
**envelope spectrum**
relative contribution of spectral components of the envelope function

**3.26**
**modulation frequency**
frequency of the sinusoidal variation of the envelope function

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

**3.27**
**modulation index**
value between 0 and 1 that describes the depth of a sinusoidal modulation of the intensity function

**3.28**
**modulation transfer ratio**
ratio between the modulation depth of the received and the original (transmitted) modulation depth of the intensity function

**3.29**
**modulation transfer function**
**MTF**
modulation transfer ratios as a function of the modulation frequency

**3.30**
**octave band weighting factor**
***α***
relative contribution of each octave band to the speech transmission index

**3.31**
**octave band redundancy factor**
***β***
fraction of information overlap between two adjacent octave bands with respect to the speech intelligibility

**3.32**
**background noise**
all sounds including noise remaining in the absence of the speech or test signal

**3.33**
**fluctuating noise**
continuous sound or noise whose sound pressure level varies significantly, but not in an impulsive manner, during the observation period

**3.34**
**impulsive noise**
sound or noise characterized by brief bursts of sound pressure

**3.35**
**signal-to-noise ratio**
***SNR***
difference between the sound pressure level of the speech or test signal and the sound pressure level of the background noise where the sound pressure levels are determined with a standardized frequency weighting

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

**3.36**
**effective signal-to-noise ratio**
**$\mathrm{SNR}_{\mathrm{eff}}$**
difference between the level of the intensity modulation and the level of the intensity of all the distortions of a received STI test signal

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

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

**3.37**
**crest factor**
difference between the peak and the RMS sound pressure levels during a given time-interval

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

**3.38**
**Lombard effect**
spontaneous increase of the vocal effort induced by the increase of the ambient noise level at the speaker’s ear

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

**3.39**
**fractional-octave-band filter**
bandpass filter for which the ratio of upper cut-off frequency $f_2$ to lower cut-off frequency $f_1$ is two raised to an exponent equal to the fraction of an octave band

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

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

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

NOTE 2 For further information, refer to IEC 61260.

**3.40**
**reference sound pressure**
***$p_0$***
sound pressure, conventionally chosen to be equal to 20 µPa for airborne sound

**3.41**
**sound pressure level**
twenty times the logarithm to the base ten of the ratio of RMS sound pressure to the reference sound pressure

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

**3.42**
**equivalent continuous sound pressure level**
ten-fold logarithm to the base ten of the ratio of the squared RMS sound pressure level for a given time-interval to the squared reference sound pressure.

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

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

where

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

$t$ is the integration variable for time;

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

$p_0$ the reference sound pressure (20 µPa).

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

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

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

**3.43**
**percentile level**
ten-fold decimal logarithm of the ratio of the squared RMS sound pressure level being exceeded for a given part of the measurement time to the squared reference sound pressure where the RMS sound pressure is determined with a standardized time and frequency weighting, e.g., $L_{10}$ or $L_{A10}$

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

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

**3.44**
**modulation transfer index**
**MTI**
unweighted mean of the scaled effective signal to noise ratios for a given octave band

**3.45**
**operational speech level**
sound pressure level of speech signal that will be used or is found in the applicable situation

**3.46**
**operational background noise level**
sound pressure level of background that will be present or is found in the applicable situation

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

# 4 Description of the STI method

## 4.1 General

### 4.1.1 Rationale for the STI method

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

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

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

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

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

### 4.1.2 Applicability of the STI method

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

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

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

### 4.2 Background of the STI method

#### 4.2.1 General

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

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

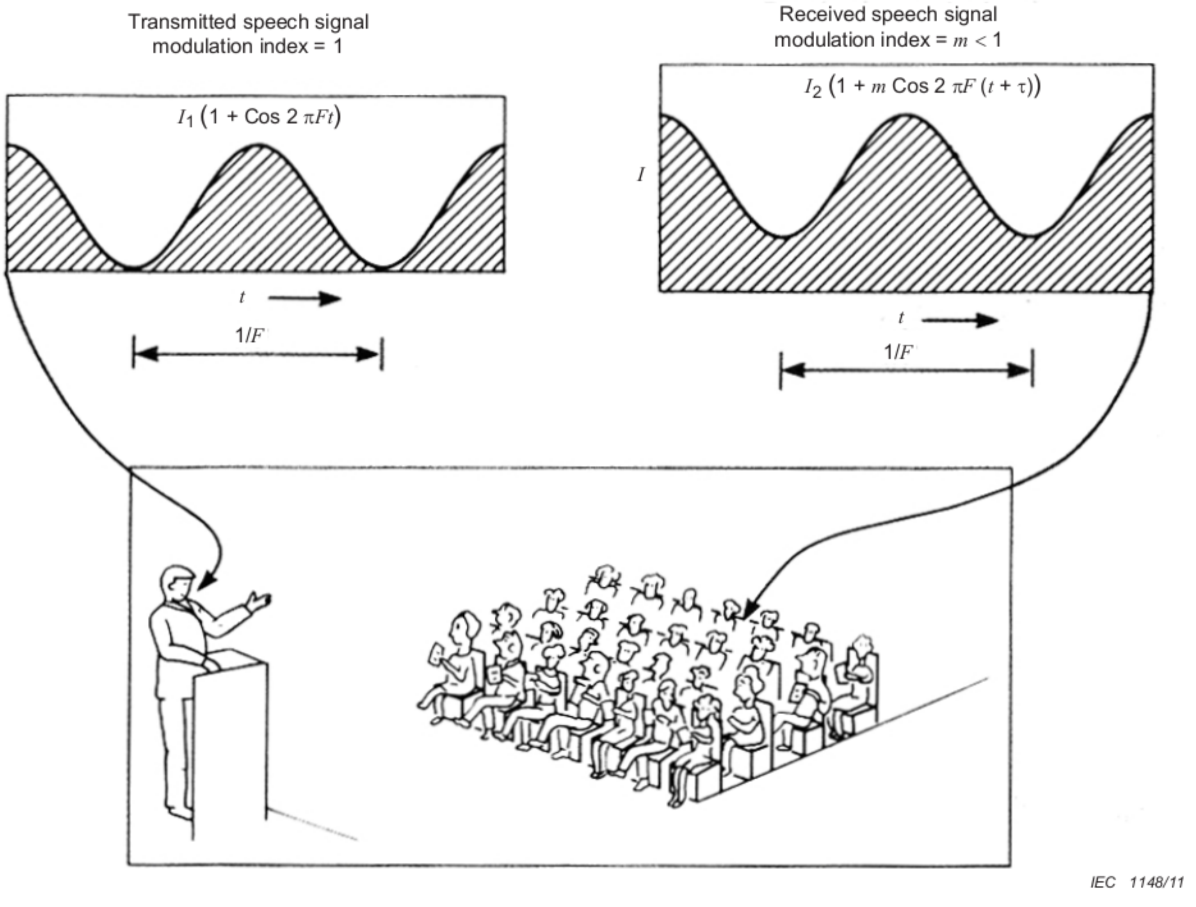



**Figure 1 – Concept of the reduction in modulation due to a transmission channel**

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

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

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

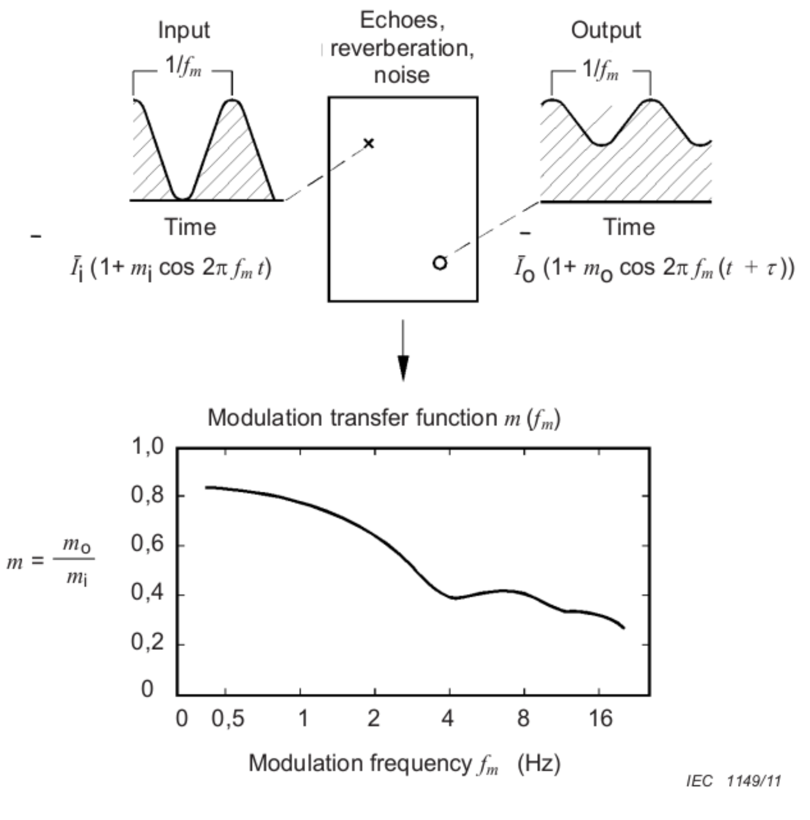
Research [4] has shown that adjacent octave bands contain redundant information with respect to speech intelligibility. If one octave band does not contribute to intelligibility (e.g. by masking from reverberation or background noise) then neighbouring octave bands can partly compensate for this missing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

### 4.2.2 Theoretical overview

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

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

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




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

**Figure 2 – Modulation transfer function – Input/output comparison**

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

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

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

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

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

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

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

### 4.2.3 Measurement of STI

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

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

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

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

There are two methods to measure STI:

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

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

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

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

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

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

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

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

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

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

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

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

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

[c] This includes time delay spectrometry.

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

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

## 4.3 Applicability of STI test methods

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

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

**Table 2 – Applicability of test**

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

## 4.4 Use of direct and indirect methods

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

**Table 3 – Choice of method**

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

## 4.5 Limitations of the STI method

### 4.5.1 General

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

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

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

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

### 4.5.2 Frequency shifts

This type of distortion may occur with

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

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

#### 4.5.3 Centre clipping

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

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

#### 4.5.4 Drop outs

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

#### 4.5.5 Jitter

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

#### 4.5.6 Vocoders

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

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

#### 4.5.7 Overestimation of STI under low background noise conditions

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

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

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

#### 4.5.8 Frequency response

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

Systems with measured STIs exceeding 0,5 have been reported where the perceived speech intelligibility has been found to be inadequate due to the poor frequency response / tonal balance of the system. The application of equalisation to improve the frequency response improved the perceived intelligibility.

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

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

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

#### 4.5.9 Echoes

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

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

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

#### 4.5.10 Fast amplitude compression and expansion

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

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

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

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

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

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

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

#### 4.5.11 Non-linear distortion

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

#### 4.5.12 Impulsive and fluctuating noise

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

- impulsive;
- fluctuating.

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

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

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

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

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

#### 4.5.13 Hearing impaired listeners

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

### 4.6 Conclusion

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

## 5 Direct method of measuring STI

### 5.1 Overview

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

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

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

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

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

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

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

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

## 5.2 STIPA

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

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

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

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

### 5.3 Application

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

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

### 5.4 Limitations

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

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

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

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

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

## 6 Indirect method of measuring STI using the impulse response

### 6.1 Overview

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

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

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

where

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

$f_m$ is the modulation frequency;

$SNR_k$ is the signal-to-noise ratio in dB.

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

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

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

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

## 6.2 Application

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

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

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

c) In order to produce a "noise-free" impulse response, a $SNR$ of at least 20 dB should be obtained in all seven octave bands. If necessary, signal averaging can be used to achieve this.

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

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

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

g) Time variances due to movements of the air (wind) or climatic changes during the measurement process shall be avoided (they also invalidate averaging over longer periods of time). The average wind speed during MLS measurements, for example, should not exceed 4 m/s. Measurements using maximum length sequences (MLS) are more vulnerable in this respect than measurements performed with sine-sweeps.

h) It should be ensured that the components involved in the transmission of sound (loudspeakers, room surfaces, reflectors, measurement microphone, people) do not move during the measurement cycle.

i) Under critical conditions, the repeatability of the measurement results shall be proven by repeated measurements.

j) The impact of background noise ($L_n$) and operational speech level ($L_s$) in each octave band $k$ shall be incorporated into the result by post-processing (see Annex M).

### 6.3 Limitations (non-linear distortion)

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

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

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

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

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

## 7 Measurement procedures, post-processing of data and applications

### 7.1 General

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

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

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

### 7.2 Acoustical input

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

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

The following procedure shall be used.

a) Verify the integrity of the test signal (e.g. via means of a loop back measurement). This is particularly important if the test signal is generated from a CD player, although PCM (e.g. .wav file) generators should also be checked. (Digitally compressed signal formats e.g. MP3 should not normally be used, though compression schemes employing at least 128 kbit/s have been shown to work without apparent error). Further information is available in [9].

b) Verify that the 1/3 octave frequency response of the test signal source (artificial mouth or suitable test loudspeaker) is within ±1 dB over the required frequency range when measured in a free field (free of reflections) for either of the following measurement techniques:

   - over the range 88 Hz to 11,3 kHz using a FULL STI or MLS or other impulse response measurement signal (the limits of the 125 Hz and 8 kHz octave bands) or
   - individual octave band levels over the range 125 Hz to 8 kHz when using a STIPA or other speech shaped test signal.

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

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

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

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

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

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

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

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

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

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

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

e) Run the STI, STIPA (or STITEL) test sequence. Normally, and where available, the “with noise” option should be selected.

f) If an MLS signal is used to measure the impulse response and if it is required to take account of the background noise, the excitation spectrum should be adjusted to the standardised speech spectrum by appropriate filtering. Signal averaging should be disabled or a single sequence should be employed [25].

g) If sine-sweeps, MLS or TDS are used to determine the noise-free impulse response, appropriate adjustments to speech and noise levels at both the microphone and receiver locations shall be applied to the noise-free MTF by post processing.

h) The test signal shall be fed into the system in such a way as to ensure that all signal processing components relevant for speech reproduction (equalisers, signal delays, etc.) are correctly taken into account during the measurement process.

## 7.3 Acoustical output

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

## 7.4 Electrical input

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

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

## 7.5 Electrical output

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

## 7.6 Examples of input/output combinations

### 7.6.1 Acoustical input – Acoustical output

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

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

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

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

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

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

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

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

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

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

### 7.7 Post-processing of measured MTF data

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

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

The effect of occupancy noise can be determined either

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

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

### 7.8 Issues concerning noise

#### 7.8.1 General

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

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

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

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

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

### 7.8.2 Measurement of background noise

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

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

### 7.8.3 Fluctuating noise

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

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

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

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

## 7.9 Analysis and interpretation of the results

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

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

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

### 7.10 Binaural STI measurements

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

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

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

## 8 Use of STI as a design prediction tool

### 8.1 Overview

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

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

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

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

### 8.2 Statistical predictions

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

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

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

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

note that:

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

### 8.3 Prediction from simulated impulse response

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

a) the MTF matrix shall be calculated using the Schroeder equation (see 6.1). The duration of the impulse response shall not be less than half the reverberation time and at least 1,6 s to ensure a reliable calculation of the modulation indices for the lowest modulation frequency of 0,63 Hz;

b) both the hearing reception thresholds and the measured or estimated background noise sound pressure levels for each octave band shall then be introduced into the MTF matrix;

c) the speech spectrum and operational speech level shall be selected and the auditory masking corrections listed in Table A.1 applied to the MTF matrix;

d) the octave band specific male and female weighting factors given in Table A.3 shall be applied to the MTI values.

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

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

The results shall also state:

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

# Annex A
(normative)

# Speech transmission index (STI) and revised STI methods

## A.1 Background

### A.1.1 Envelope function and envelope spectrum

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

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

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

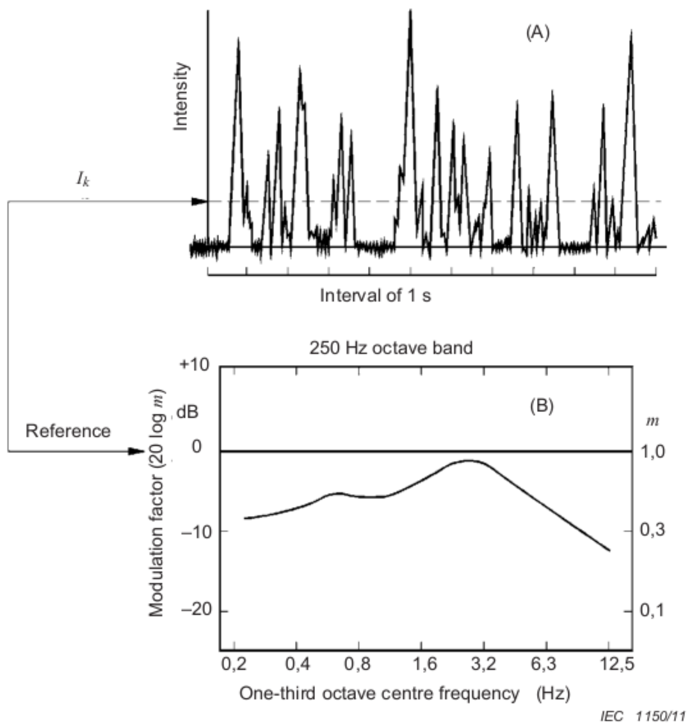



**Figure A.1 – Envelope function (panel A) of a 10 s speech signal for the 250 Hz octave band and corresponding envelope spectrum (panel B)**

### A.1.2 Modulation transfer function (MTF)

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

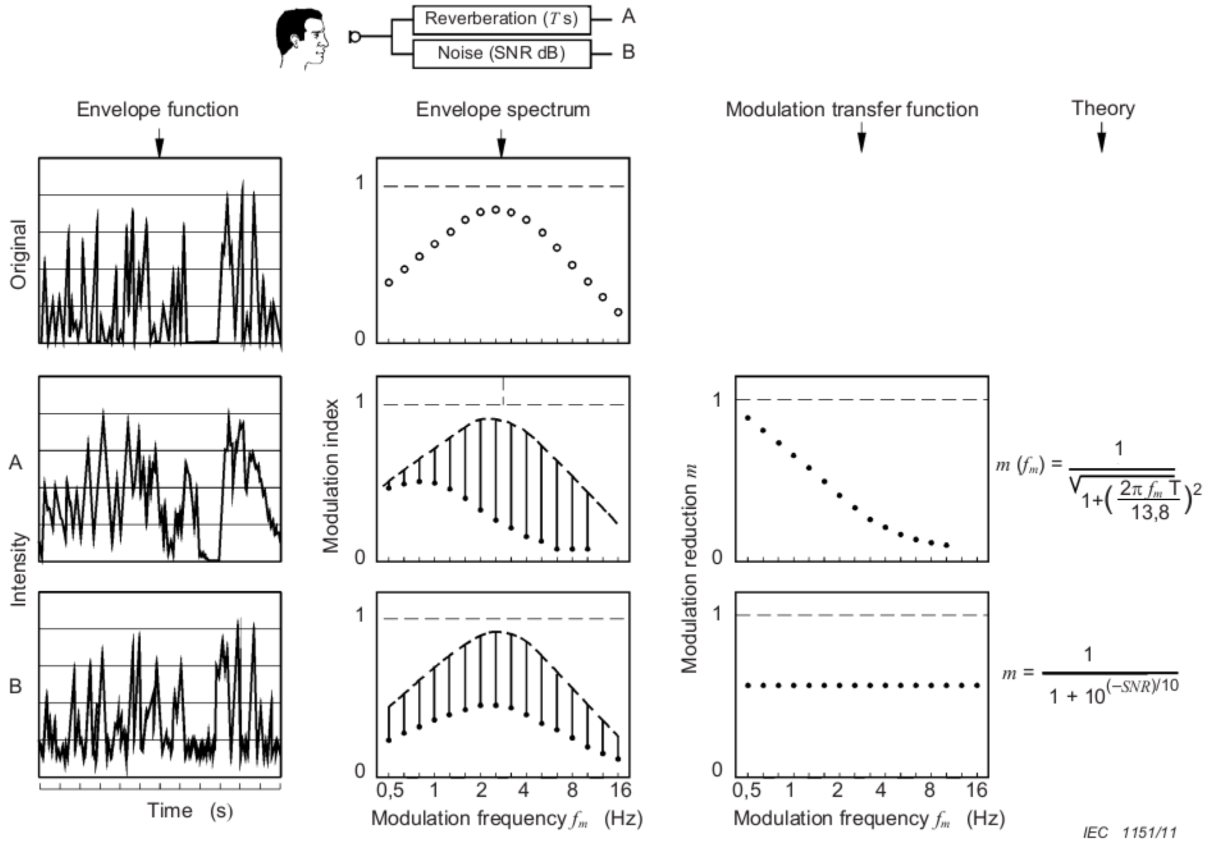



**Figure A.2 – Theoretical expression of the MTF**

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

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

where

$f_m$ is the modulation frequency;

$T$ is the reverberation time in seconds.

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

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

where

$SNR$ is the signal to noise ratio in dB.

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

## A.2 STI technique

### A.2.1 General

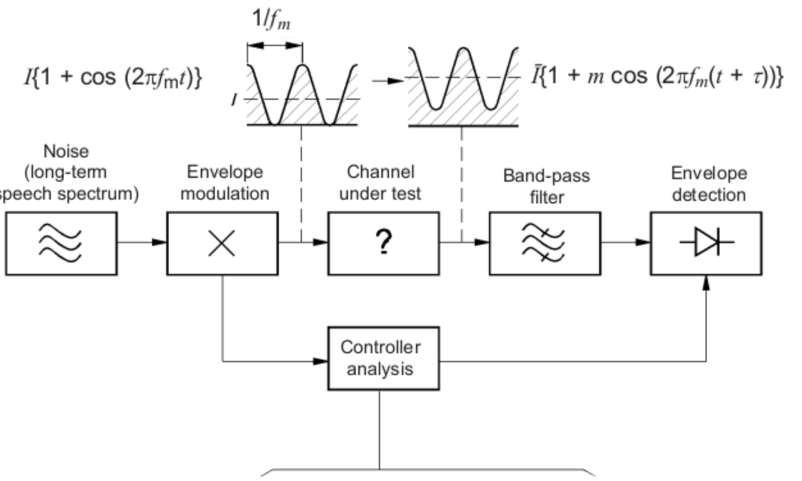
The speech transmission index (STI) is an objective measure, based on the weighted contribution of a number of frequency bands within the frequency range of speech signals, the contributions being set by the effective signal-to-noise ratio. Its description and the octave-band weighting factors and redundancy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

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

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

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

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




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

IEC 1152/11

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

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

### A.2.2 STI modulation frequencies

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

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

## A.3 Auditory effects on the STI

### A.3.1 Overview

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

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

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

### A.3.2 Level dependent auditory masking

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

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

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

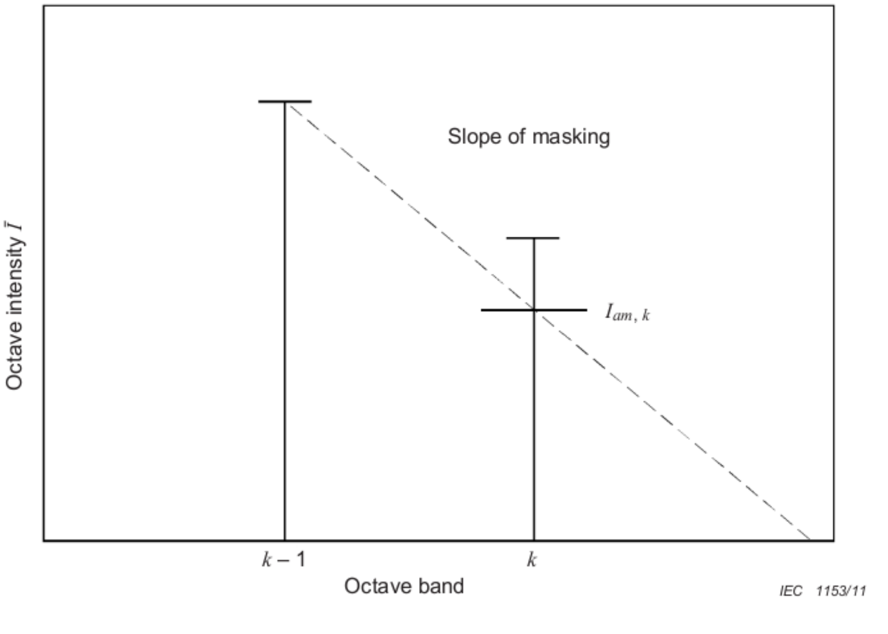



**Figure A.4 – Auditory masking of octave band ($k$ – 1) on octave band ($k$)**

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

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

where

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

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

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

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

where

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

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

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

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

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

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

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

where

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

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

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

### A.3.3 Absolute speech reception threshold

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

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

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

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

$$I_{\mathrm{rt},k} = 10^{\left(ART_k/10\right)}$$

where

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

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

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

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

**Table A.3 – MTI octave band weighting factors**

| **Octave band**<br>Hz | | 125 | 250 | 500 | 1 000 | 2 000 | 4 000 | 8 000 |
|---|---|---|---|---|---|---|---|---|
| **Males** | $\alpha$ | 0,085 | 0,127 | 0,230 | 0,233 | 0,309 | 0,224 | 0,173 |
| | $\beta$ | 0,085 | 0,078 | 0,065 | 0,011 | 0,047 | 0,095 | – |
| **Females** | $\alpha$ | – | 0,117 | 0,223 | 0,216 | 0,328 | 0,250 | 0,194 |
| | $\beta$ | – | 0,099 | 0,066 | 0,062 | 0,025 | 0,076 | – |

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

### A.3.5 Gender-specific spectra of STI test signals

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

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

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

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

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

## A.4 STI test signal generation

### A.4.1 Pink noise source signal

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

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

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

### A.4.2 Generating octave band carrier signals

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

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

### A.4.3 Intensity modulation of the carrier signals

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

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

where

$f_m$ is the modulation frequency in Hz;

$t$ is the time in seconds;

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

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

#### A.4.4 Applying the speech spectrum to the STI signal

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

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

### A.5.1 Signal processing

#### A.5.1.1 Basic processing steps

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

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

#### A.5.1.2 Filtering and determination of intensities

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

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

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

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

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

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

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

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

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

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

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 $k$;

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

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

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

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

where

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

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

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

### A.5.3 Correction of the MTF using auditory masking

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

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

where

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

$I_k$ is the acoustic intensity level for octave band $k$;

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

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

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

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

### A.5.4 Calculation of the effective signal to noise ratio

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

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

where

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

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

### A.5.5 Calculation of the Transmission Index (TI)

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

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

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.

### A.5.6 Calculation of the STI

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

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

where

$TI_{k,f_m}$ is the transmission index for each octave band $k$ and modulation frequency $f_\mathrm{m}$;

$m$ is the index of the modulation frequency.

$n$ is the number of modulation frequencies per octave band.

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

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

where

$MTI_k$ is the modulation transfer index for octave band $k$;

$\alpha_k$ is the weight factor for octave band $k$;

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

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

# Annex B
(normative)

## STIPA method

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

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

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

**Table B.1 – Modulation frequencies for the STIPA method**

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

# Annex C
(normative)

# STITEL method

## C.1 General

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

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

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

**Table C.1 – Modulation frequencies for the STITEL method**

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

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

## C.2 Limitations

The STITEL method should not be used for transmission channels:

a) which introduce frequency shifts or frequency multiplication;

b) which include vocoders (i.e. LPC, CELP, RELP, etc.);

c) which introduce strong non-linear distortion components;

d) for which reverberation time is strongly frequency-dependent;

e) having echoes stronger than –10 dB referred to the primary signal;

f) if the background noise has audible tones and/or marked peaks or troughs in the octave-band spectrum;

g) if the background noise is impulsive and/or the space is not substantially free of discrete echoes, particularly flutter echoes whose repetition frequency is an integral multiple of one or more of the modulation frequencies.

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

# Annex D
(informative)

## RASTI method (obsolete)

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

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

**Table D.1 – Modulation frequencies for the RASTI method**

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

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

Figure D.1 illustrates a practical RASTI test signal.

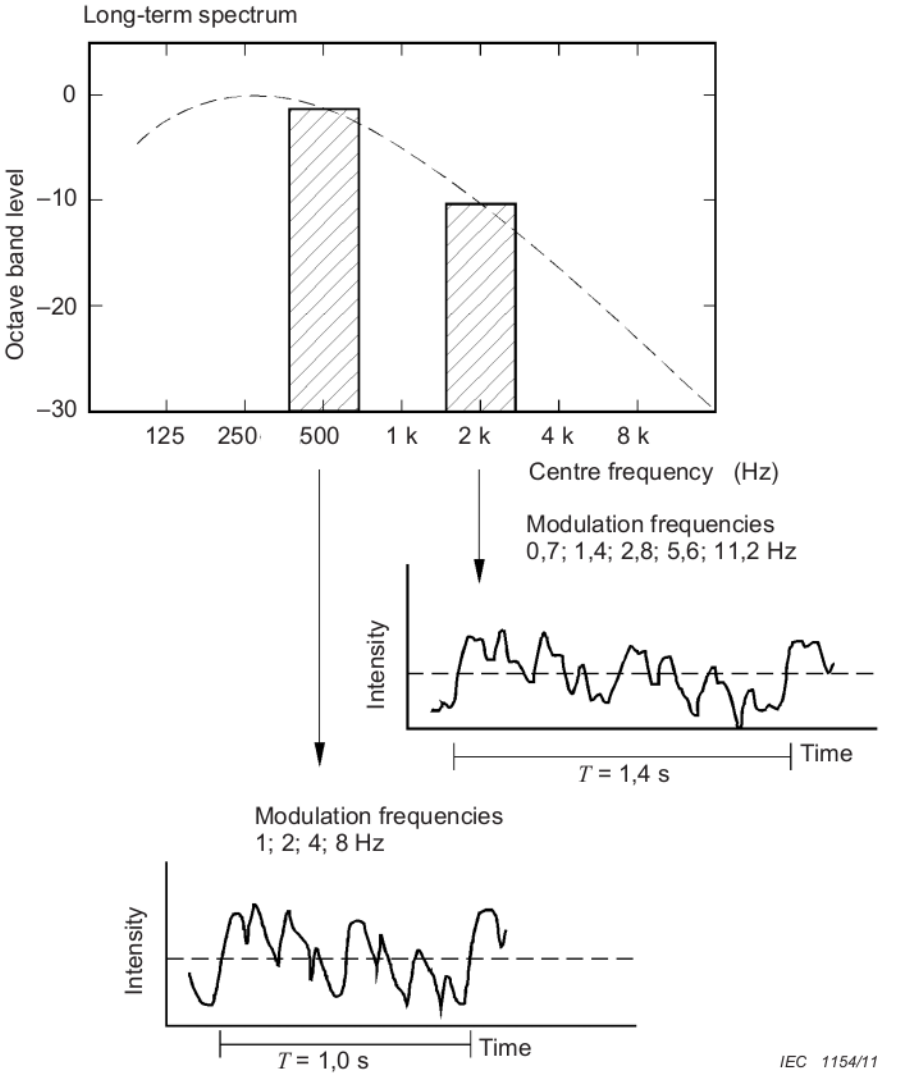



**Figure D.1 – Illustration of a practical RASTI test signal**

# Annex E
(informative)

# Qualification of the STI and relationships with other speech intelligibility measures

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

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

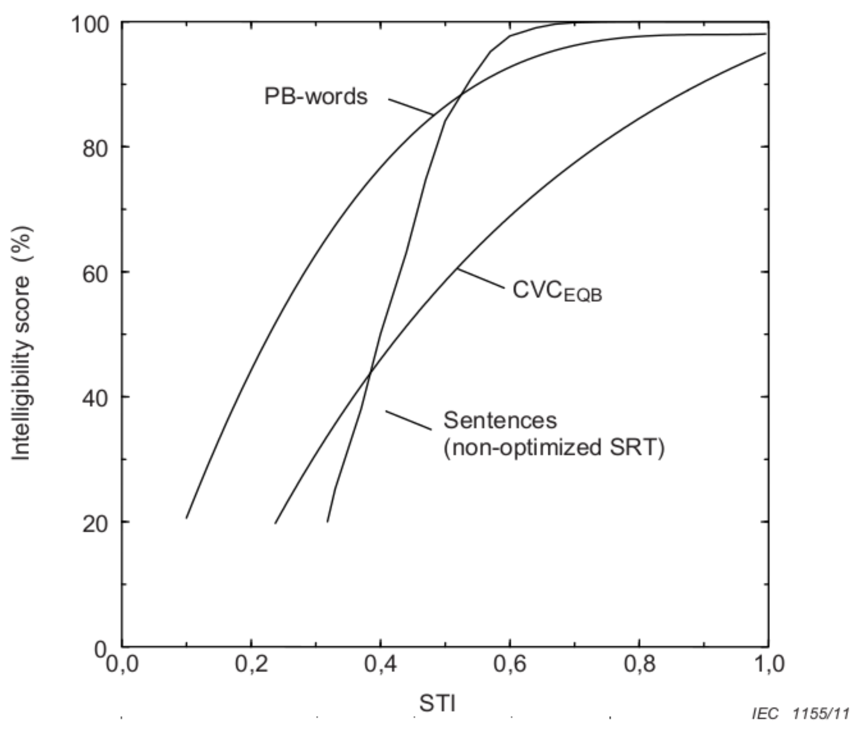



**Figure E.1 – Relationships between some speech intelligibility measures**

## E.2 Relationship between STI and listening difficulty

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

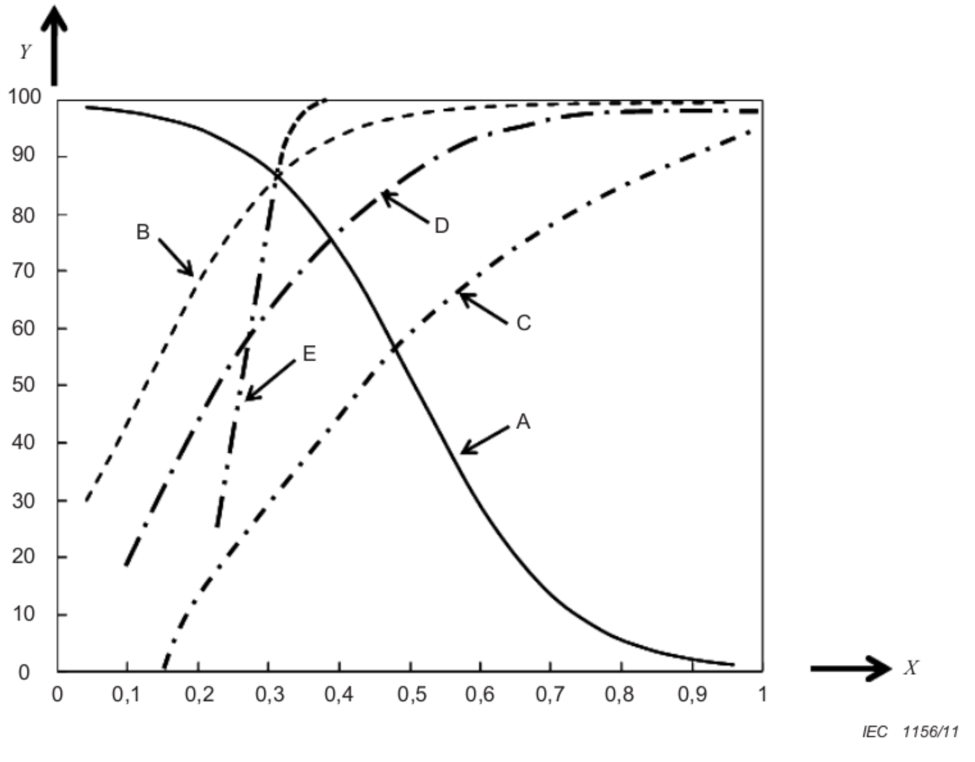
The listening difficulty rating is the percentage of responses indicating a certain degree of difficulty when listening to speech through a transmission channel. Note that listening difficulty ratings decrease for conditions with improved speech transmission, contrary to conventional intelligibility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

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

**Table E.1 – Categories for listening difficulty**

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

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




**Key**

*X* axis is STI

*Y* axis is percentage of correct responses and listening difficulty

Curve A is listening difficulty

Curve B is word recognition

Curve C is $CVC_{EQB}$

Curve D is PB-Words

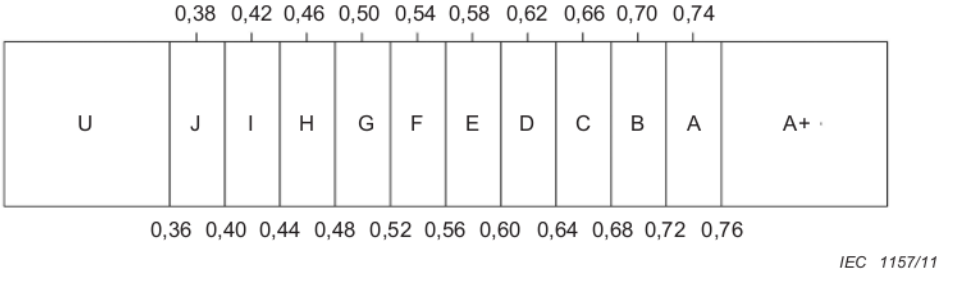
Curve E is Sentences (SRT)

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

# Annex F
(informative)

## Nominal qualification bands for STI

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




**Key**

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

Row of letters: band designations;

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

**Figure F.1 – STI qualification bands**

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

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

# Annex G
(informative)

## Examples of STI qualification bands and typical applications

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

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

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

NOTE 1 These values should be regarded as minimum target values.

NOTE 2 Perceived intelligibility relating to each category will also depend on the frequency response at each listening position.

NOTE 3 The STI values refer to measured values in sample listening positions or as required by specific application standards.

# Annex H
(informative)

## Non-native listeners

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

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

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

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

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

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

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

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

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

# Annex I
(informative)

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

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

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

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

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

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

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

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

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

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

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

# Annex J
(normative)

# Calibration of STI test signal level

## J.1 Overview

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

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

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

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

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

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

## J.3 Test speech level

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

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

## J.4 Corrected speech level

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

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

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

# Annex K
(informative)

## Example test report sheet for STI measurements

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

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

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

# Annex L
(normative)

## Prediction of STI using statistical methods

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

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

where

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

$f_m$ is the modulation frequency;

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

$t$ is the integration variable for time;

$SNR$ is the signal to noise ratio in dB.

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

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

where

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

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

$r$ is the talker to listener distance;

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

$T$ is the reverberation time of the room or space.

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

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

with

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

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

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

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

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

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

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

# Annex M
(informative)

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

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

**Table M.1 – Example calculation**

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

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

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

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

# Annex N
(informative)

# Other methods of determining speech intelligibility

## N.1 Overview

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

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

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

## N.2 Word tests

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

## N.3 Modified rhyme tests

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

## N.4 Speech Intelligibility Index (SII)

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

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

## N.5 Articulation loss of consonants (%ALcons)

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

## N.6 PESQ

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

# Bibliography

[1] Mapp, P, *New Techniques for Measuring Speech Privacy & Efficiency of Sound Masking Systems. AES 122nd Convention, Vienna, May 2007*

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

[3] IEC 61672 (all parts), *Electroacoustics – Sound level meters*

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

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

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

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

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

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

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

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

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

[13] Mapp P, *Practical application of STI to assessing Public Address and Emergency Sound Systems. Past, present and future of the Speech Transmission Index, TNO 2002*

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

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

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

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

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

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

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

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

[22] Mapp, P. *Further thoughts on Speech Transmission Index (STI). IOA Reproduced Sound 18.Proc. IOA Vol 24 Pt 8. 2002*

[23] Bozzoli, F, Farina, A, *“Influence of artificial mouth’s directivity in determining Speech Transmission Index”., AES 119th Convention, New York, 2005*

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

[25] Rife, D., *“Modulation Transfer Function Measurements with Maximum-Length Sequences”, J. Audio Eng. Soc., Oct. 1992, Vol. 40, No. 10*

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

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

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

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

[30] Zwicker, E. and Feltkeller, *“The ear as an information receiver”, (ASA)*

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

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

[33] Anderson, B.W., and Kalb, J.T. 1987. *“English verification of the STI method for estimating speech intelligibility of a communications channel,” J. Acoust. Soc. Am. 81, 1982-1985*

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

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

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

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

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

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

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

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

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

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

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

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

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

[47] ITU-T P.50, *Telephone transmission quality, telephone installations, local line networks – Objective measuring apparatus – Artificial voices*

[48] ITU-T P.51, *Telephone transmission quality – Objective measuring apparatus – Artificial mouth*

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

___________

*This page deliberately left blank*

*This page deliberately left blank*

# British Standards Institution (BSI)

BSI is the independent national body responsible for preparing British Standards and other standards-related publications, information and services. It presents the UK view on standards in Europe and at the international level.

It is incorporated by Royal Charter.

## Revisions

British Standards are updated by amendment or revision. Users of British Standards should make sure that they possess the latest amendments or editions.

It is the constant aim of BSI to improve the quality of our products and services. We would be grateful if anyone finding an inaccuracy or ambiguity while using this British Standard would inform the Secretary of the technical committee responsible, the identity of which can be found on the inside front cover.

**Tel: +44 (0)20 8996 9001  Fax: +44 (0)20 8996 7001**

BSI offers Members an individual updating service called PLUS which ensures that subscribers automatically receive the latest editions of standards.

**Tel: +44 (0)20 8996 7669 Fax: +44 (0)20 8996 7001**
**Email: plus@bsigroup.com**

## Buying standards

You may buy PDF and hard copy versions of standards directly using a credit card from the BSI Shop on the website **www.bsigroup.com/shop.** In addition all orders for BSI, international and foreign standards publications can be addressed to BSI Customer Services.

**Tel: +44 (0)20 8996 9001 Fax: +44 (0)20 8996 7001**
**Email: orders@bsigroup.com**

In response to orders for international standards, it is BSI policy to supply the BSI implementation of those that have been published as British Standards, unless otherwise requested.

## Information on standards

BSI provides a wide range of information on national, European and international standards through its Knowledge Centre.

**Tel: +44 (0)20 8996 7004  Fax: +44 (0)20 8996 7005**
**Email: knowledgecentre@bsigroup.com**

Various BSI electronic information services are also available which give details on all its products and services.

**Tel: +44 (0)20 8996 7111  Fax: +44 (0)20 8996 7048**
**Email: info@bsigroup.com**

BSI Subscribing Members are kept up to date with standards developments and receive substantial discounts on the purchase price of standards. For details of these and other benefits contact Membership Administration.

**Tel: +44 (0)20 8996 7002  Fax: +44 (0)20 8996 7001**
**Email: membership@bsigroup.com**

Information regarding online access to British Standards via British Standards Online can be found at **www.bsigroup.com/BSOL**

Further information about BSI is available on the BSI website at **www.bsigroup.com/standards**




**BSI Group Headquarters**

389 Chiswick High Road London W4 4AL UK

Tel +44 (0)20 8996 9001
Fax +44 (0)20 8996 7001
www.bsigroup.com/standards


*raising standards worldwide™*