

INTERNATIONAL STANDARD

IEC
60268-16

Third edition
2003-05

Sound system equipment –

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

Équipements pour systèmes électroacoustiques –

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



Reference number
IEC 60268-16:2003(E)

Publication numbering

As from 1 January 1997 all IEC publications are issued with a designation in the 60000 series. For example, IEC 34-1 is now referred to as IEC 60034-1.

Consolidated editions

The IEC is now publishing consolidated versions of its publications. For example, edition numbers 1.0, 1.1 and 1.2 refer, respectively, to the base publication, the base publication incorporating amendment 1 and the base publication incorporating amendments 1 and 2.

Further information on IEC publications

The technical content of IEC publications is kept under constant review by the IEC, thus ensuring that the content reflects current technology. Information relating to this publication, including its validity, is available in the IEC Catalogue of publications (see below) in addition to new editions, amendments and corrigenda. Information on the subjects under consideration and work in progress undertaken by the technical committee which has prepared this publication, as well as the list of publications issued, is also available from the following:

- **IEC Web Site** (www.iec.ch)

- **Catalogue of IEC publications**

The on-line catalogue on the IEC web site (http://www.iec.ch/searchpub/cur_fut.htm) enables you to search by a variety of criteria including text searches, technical committees and date of publication. On-line information is also available on recently issued publications, withdrawn and replaced publications, as well as corrigenda.

- **IEC Just Published**

This summary of recently issued publications (http://www.iec.ch/online_news/justpub/jp_entry.htm) is also available by email. Please contact the Customer Service Centre (see below) for further information.

- **Customer Service Centre**

If you have any questions regarding this publication or need further assistance, please contact the Customer Service Centre:

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

INTERNATIONAL STANDARD

IEC
60268-16

Third edition
2003-05

Sound system equipment –

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

Équipements pour systèmes électroacoustiques –

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

© IEC 2003 — Copyright - all rights reserved

No part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from the publisher.

International Electrotechnical Commission, 3, rue de Varembé, PO Box 131, CH-1211 Geneva 20, Switzerland
Telephone: +41 22 919 02 11 Telefax: +41 22 919 03 00 E-mail: inmail@iec.ch Web: www.iec.ch



Commission Electrotechnique Internationale
International Electrotechnical Commission
Международная Электротехническая Комиссия

PRICE CODE

U

For price, see current catalogue

CONTENTS

FOREWORD	4
1 Scope	6
2 Normative references.....	6
3 Definitions and abbreviations.....	6
4 Description of the methods	7
4.1 General	7
4.2 The STI method	8
4.3 The STITEL method.....	9
4.4 The STIPA method	10
4.5 The RASTI method	10
4.6 Methods of measurement.....	13
5 Methods of determining intelligibility	15
5.1 Word tests	15
5.2 Modified rhyme tests.....	15
5.3 Speech Intelligibility Index	15
5.4 Articulation loss of consonants	15
Annex A (normative) Speech transmission index (STI) and revised (STI _r) methods	16
A.1 Background	16
A.2 The STI method	19
A.3 The test signals	23
Annex B (informative) The STITEL method.....	24
Annex C (informative) The STIPA method	25
Annex D (informative) The RASTI method.....	26
Annex E (informative) Qualification of the STI and relation with some subjective intelligibility measures.....	27
Bibliography.....	28
Figure 1 – Modulation transfer function: input/output comparison	7
Figure 2 – Relationship between the theoretical STI by the RASTI method and the STI measured by a proprietary equipment with a measurement time of 12 s approximately	11
Figure 3 – Conditions under which RASTI results do not differ by more than 0,05	12
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)	16
Figure A.2 – Theoretical expression of the MTF	18
Figure A.3 – The measurement system and frequencies for the STI method.....	19
Figure A.4 – Auditory masking strength of octave band (k – 1) on that above (k).....	20
Figure A.5 – The relationship between effective signal-to-noise ratio and transmission index for a shift of 15 dB and a range of 30 dB.....	22
Figure D.1 – Illustration of a practical RASTI test signal.....	26
Figure E.1 – Qualification of the STI and relation with some subjective intelligibility measures	27

Table A.1 – Octave level specific slope of masking and corresponding auditory masking factor (amf)	21
Table A.2 – STI_f octave band specific male and female weighting factors	23
Table A.3 – Octave band levels (dB) relative to the A-weighted long-term speech level.....	23
Table B.1 – STITEL: modulation frequencies for the seven octave bands	24
Table C.1 – STIPA: modulation frequencies for the seven octave bands	25
Table C.2 – STI_f octave band specific male and female weighting factors adopted to STIPA	25

IECNORM.COM: Click to view the full PDF of IEC 60268-16:2003

Withd 2013

INTERNATIONAL ELECTROTECHNICAL COMMISSION

SOUND SYSTEM EQUIPMENT –

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

FOREWORD

- 1) The IEC (International Electrotechnical Commission) is a worldwide organization for standardization comprising all national electrotechnical committees (IEC National Committees). The object of the 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, the IEC publishes International Standards. 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. The 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 the 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 National Committees.
- 3) The documents produced have the form of recommendations for international use and are published in the form of standards, technical specifications, technical reports or guides and they are accepted by the National Committees in that sense.
- 4) In order to promote international unification, IEC National Committees undertake to apply IEC International Standards transparently to the maximum extent possible in their national and regional standards. Any divergence between the IEC Standard and the corresponding national or regional standard shall be clearly indicated in the latter.
- 5) The IEC provides no marking procedure to indicate its approval and cannot be rendered responsible for any equipment declared to be in conformity with one of its standards.
- 6) Attention is drawn to the possibility that some of the elements of this International Standard may be the subject of patent rights. The IEC shall not be held responsible for identifying any or all such patent rights.

International Standard IEC 60268-16 has been prepared by IEC technical committee 100: Audio, video and multimedia systems and equipment.

This third edition cancels and replaces the second edition, published in 1998. This third edition constitutes a technical revision.

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

FDIS	Report on voting
100/650/FDIS	100/677/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.

The committee has decided that the contents of this publication will remain unchanged until 2005. At this date, the publication will be

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

A bilingual edition of this standard may be issued at a later date

Withdrawn
IECNORM.COM: Click to view the full PDF of IEC 60268-16:2003

SOUND SYSTEM EQUIPMENT –

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

1 Scope

This part of IEC 60268 defines objective methods for rating the transmission quality of speech with respect to intelligibility. The four methods, which are closely related, are referred to as the “STI,” the “STITEL,” the “STIPA” and the “RASTI” methods (see Clause 3). The methods are intended for rating speech transmission with or without sound systems.

A survey of other methods of determining or predicting speech intelligibility is also included, together with a method of correlating the results of different methods of determination.

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.

ISO 4870:1991, *Acoustics – The construction and calibration of speech intelligibility tests*

ITU-T Recommendation P.51:1996, *Artificial mouth*

3 Definitions and abbreviations

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

3.1

speech transmission index (STI)

physical quantity representing the transmission quality of speech with respect to intelligibility

3.2

speech transmission index for telecommunication systems (STITEL)

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

3.3

speech transmission index for public address systems (STIPA)

index obtained by a condensed version of the STI method but still responsive to distortions found in room acoustics including public address systems

3.4

room acoustics speech transmission index (RASTI)

index obtained by a condensed version of the STI method, to be used for screening purposes and focused on direct communication between persons without making use of a communication system. RASTI accounts for noise interference and distortions in the time domain (echoes, reverberation)

4 Description of the methods

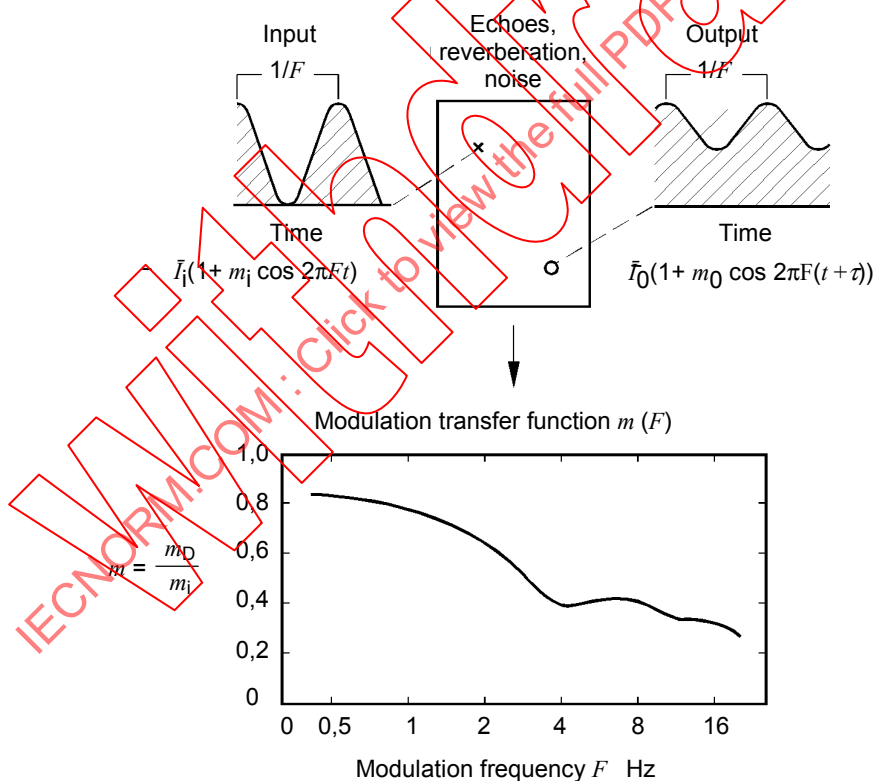
4.1 General

The methods can be used to compare speech transmission quality at various positions and for various conditions within the same listening space, in particular for assessing the effect of changes in the acoustic properties. This includes effects from the presence of an audience or of changes in any sound system [1]¹⁾. The methods are also able to predict the absolute rating of the speech transmission quality with respect to intelligibility when comparing different listening spaces under similar conditions or assessing a speech communication channel. Annex A provides a more detailed description of the basis of the speech transmission index.

The determination of the transmission quality of speech with respect to intelligibility is based on the reduction of the modulation index m_i of a test signal, simulating the speech characteristics of a real talker, when sounded in a room or through a communication channel. The test signal is transmitted by a sound source situated at the talker's position to a microphone at any listener's position, where the modulation index is m_o .

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

The typical test signal consists of a noise carrier with a speech-shaped frequency spectrum and a sinusoidal intensity modulation with modulation frequency F (see Figure 1).



IEC 1572/03

NOTE m_i and m_o are the modulation indices of the input and the output signals, respectively. I_i and I_o are the input and output intensities, the intensities being equal to the square of the sound pressure levels (p^2).

Figure 1 – Modulation transfer function: input/output comparison

¹⁾ Figures in square brackets refer to the bibliography.

The reduction in the modulation index is quantified by the modulation transfer function $m(F)$ which is determined by

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

and is interpreted in terms of an apparent signal-to-noise ratio (SNR), irrespective of the cause of the reduction which can be reverberation, echoes, non-linear distortion components or interfering noise, determined by

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

The values of the apparent signal-to-noise ratio are limited to the range ± 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.

4.2 The STI method

4.2.1 General

The STI method, described in Annex A, is based on the determination of the modulation transfer function $m(F)$ for 98 data points, obtained for 14 modulation frequencies at one-third octave intervals ranging from 0,63 Hz up to and including 12,5 Hz and for seven octave bands with centre frequencies ranging from 125 Hz up to and including 8 kHz (see Figure A.3).

4.2.2 Precision of the STI method

Because the test signal is band-limited random or pseudo-random noise, repetition of measurement does not normally produce identical results, even under conditions of steady interference. The results centre on a mean with a certain standard deviation. This depends, amongst other factors, on the number of discrete measurements of the modulation transfer function (usually 98 for the STI method) and the measuring time involved. Typically, the value of the standard deviation is about 0,02 for a measuring time of 10 s for each $m(F)$ and with stationary noise interference. With fluctuating noise (for example, a babble of voices), higher standard deviations may be found possibly with a systematic error. This can be checked by carrying out a measurement in the absence of the test signal. This should result in a residual STI value less than 0,20. An estimate of the standard deviation should be made by repeating measurements for at least a restricted set of conditions.

4.2.3 Limitations of the STI method

Due to the form of the test signals and the analysis, the types of distortion not accounted for are frequency shifts (such as those found with devices for preventing acoustic feedback and with single sideband radio transmissions), frequency multiplication (for example, analogue tape recordings played at incorrect speed) and systems such as vocoders that encode speech fragments (for example, linear predictive coding which might use code-book related synthesis or the introduction of errors related to voiced/unvoiced speech fragments and pitch errors).

The method should not be used for transmission channels

- a) which introduce frequency shifts or frequency multiplication, or
- b) which include vocoders (i.e. linear predictive speech coder (LPC), code-excited linear predictive coder (CELP), residually excited linear predictive coder (RELTP), etc.).

Without specific corrections, the STI method is not a reliable prediction measure of the intelligibility of speech for hearing-impaired listeners [17] or to the wearers of ear defenders.

4.3 The STITEL method

4.3.1 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 opens up the possibility of modulating and parallel processing all seven frequency bands simultaneously, thus reducing measuring time. The STITEL method, described in Annex B, employs this simplification and takes 10 s to 15 s for a measurement.

4.3.2 Precision of the STITEL method

As with the STI method (see 4.2.2), results are mean values with a certain standard deviation, due to the randomness of noise. The standard deviation depends on the number of discrete measurements of the modulation transfer function (typically seven for the STITEL method) and the measuring time involved. The standard deviation should be estimated by performing repeated measurements, at least for a restricted number of conditions.

4.3.3 Limitations of the STITEL method

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. Over the range of centre frequencies 125 Hz to 8 kHz, the uniformities of the octave-band early decay times and signal-to-noise ratios should fall within the permitted area shown in Figure 3;
- 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 [2].

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.

4.4 The STIPA method

4.4.1 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 opens up the possibility of modulating and parallel processing of all frequency bands simultaneously, thus reducing measuring time. For each frequency band the modulation transfer is determined for two modulation frequencies. The STIPA method, described in Annex C, employs this simplification and takes 10 s to 15 s for a measurement

4.4.2 Precision of the STIPA method

As with the STI method (see 4.2), results are mean values with a certain standard deviation, due to the randomness of noise. The standard deviation depends on the number of discrete measurements of the modulation transfer function (typically 12 for the STIPA method) and the measuring time involved. The standard deviation should be estimated by performing repeated measurements, at least for a restricted number of conditions.

4.4.3 Limitations of the STIPA method

The STIPA method should not be used for public address systems

- a) which introduce frequency shifts or frequency multiplication;
- b) which include vocoders (i.e. LPC, CELP, RLP, etc.);
- c) if the background noise is impulsive;
- d) which introduce strong non-linear distortion components.

If d) applies, or possibly applies, the STI method should be used instead or used to verify the results obtained by the STIPA method.

4.5 The RASTI method

4.5.1 General

Another simplification that can be applied is a reduction in the number of octave bands. This is the case with the RASTI method, described in Annex D, in which the analysis is restricted to only two octave bands with centre frequencies 500 Hz and 2 kHz, and to only four and five modulation frequencies, respectively, in these bands. This implies that bandpass limiting and background noise with an irregular spectrum are not accounted for correctly, nor is the effect of non-linear distortion included. The RASTI method can, however, be used as a screening approach for most person-to-person communications in room acoustic applications. As with the STI method, certain distortions, particularly those from reverberation, if smooth and monotonic, are accounted for correctly [8].

4.5.2 Precision of the RASTI method

As with the STI method (see 4.2), results are mean values with a certain standard deviation, due to the randomness of noise. The standard deviation depends upon the measuring time involved, amongst other factors. The standard deviation should be estimated by performing repeated measurements, at least for a restricted number of conditions. In practice, a measuring time of 10 s is a useful compromise between speed and accuracy. Figure 2 illustrates the accuracy obtainable with a measuring time of that order.

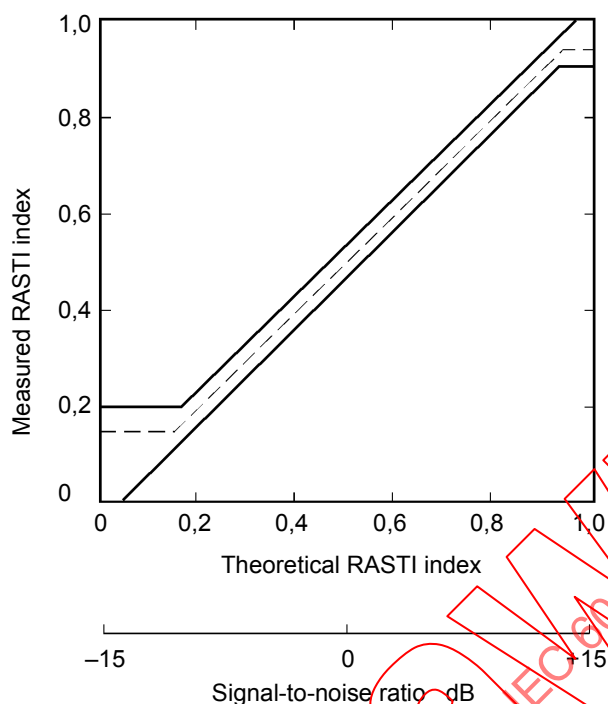
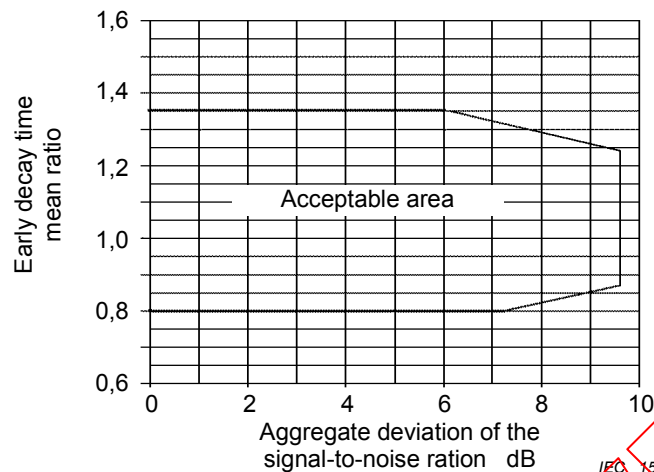


Figure 2 – Relationship between the theoretical STI by the RASTI method and the STI measured by a proprietary equipment with a measurement time of 12 s approximately

4.5.3 Limitations of the RASTI method

The application of the RASTI method is limited by factors concerned with speech transmission, background noise and reverberation. Therefore, its use should be restricted to cases where the following conditions are met.

- a) No frequency shifts or frequency multiplication.
- b) No use of vocoders (i.e. LPC, CELP, RELP, etc.).
- c) Essentially linear speech transmission (any amplitude compression or expansion limited to 1 dB) and no peak clipping of a sinusoidal signal giving the same sound pressure level at the measuring position as the test signal.
- d) Overall system frequency response between the octave bands centred on 125 Hz and 8 kHz is uniform, i.e. the difference in transmission between any two adjacent octave bands should not exceed 5 dB.
- e) Background noise is free of audible tones and of marked peaks or troughs in the octave-band spectrum.
- f) Background noise is not impulsive and the space is substantially free of discrete echoes, particularly flutter echoes whose repetition frequency is an integral multiple of one or more of the modulation frequencies [2].
- g) Reverberation time is not strongly frequency-dependent. Over the range of centre frequencies 125 Hz to 8 kHz, the uniformities of the octave-band early decay times (first 5 dB) and signal-to-noise ratios should fall within the permitted area shown in Figure 3.
- h) Background noise does not vary substantially with time.



NOTE 1 The absolute aggregate deviation of the signal-to-noise ratio is the algebraic sum of the differences of the octave-band signal-to-noise ratios in the five octave bands centred on 125 Hz, 250 Hz, 1 kHz, 4 kHz and 8 kHz from the arithmetic mean of the signal-to-noise ratios of the 500 Hz and 2 kHz octave bands. Signal-to-noise ratios which exceed ± 15 dB are set to +15 dB or –15 dB, respectively.

NOTE 2 The early decay time mean ratio is the average early decay time over the octave bands centred on 125 Hz, 250 Hz, 1 kHz, 4 kHz and 8 kHz, divided by the average early decay time of the 500 Hz and 2 kHz octave bands.

Figure 3 – Conditions under which RASTI results do not differ by more than 0,05

The STI, STITEL and the STIPA methods may be considered where conditions c), d) or e) are not met. Where f) or g) are not met, the STI and STIPA methods should be used.

In cases where conditions a), b) or h) are not met, it is necessary to use another method such as subjective test methods. These can be based on words in a carried phrase as phonetically balanced or equally balanced words or modified rhyme tests.

4.6 Methods of measurement

Any compression or non-linear amplitude or non-stationary frequency or temporal processing should be bypassed before carrying out RASTI measurements, but it is essential to ensure that any consequent effects on the sound pressure levels produced by the system under test are compensated.

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) should be used for the highest accuracy when assessing the intelligibility of unamplified talkers. When speech is relayed through a sound system, a simulator is not normally required unless a close talking or noise cancelling microphone is involved.

4.6.1 Method of measurement using an acoustic excitation signal via microphone input

- Set the source (artificial mouth or suitable test loudspeaker) on the axis of the appropriate microphone at the normal speaking distance (measured from the lip-circle for the artificial mouth) and direct it in the normal speaking direction.
- Set the test signal level at the microphone position to equal that of the speech level under normal operating conditions. The sound pressure level (SPL) should be set using A-weighting and the level should be 60 dB at 1 m in front of the artificial mouth or test loudspeaker (66 dB at 0,5 m).
- Check that the test signal spectrum at the input microphone position is correct to within ± 1 dB over the range 88 Hz to 11,3 kHz (the limits of the 125 Hz and 8 kHz octave bands). Adjust the equalization (if any) of the artificial mouth or test loudspeaker, as necessary, to satisfy this requirement.
- Run the STI, STITEL or STIPA test sequence. Normally, and where available, the “with noise” option should be selected.

4.6.2 Method of measurement using direct (electrical) injection of the test signal

- Follow the above procedure, replacing step b) by step b) below, and selecting the injection point for the signal to be as close as possible to the normal microphone input, so as to include as much of the system as possible in the test.
- Set the test signal level so that an A-weighted sound pressure level, equal to the $L_{Aeq,14}$ of the system for continuous speech, is produced.

4.6.3 Simulation of occupancy noise

The effect of occupancy noise can be determined either

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

4.6.4 Measuring method using maximum length sequence (MLS) analysis equipment

The limitations given in 4.2.3 (STI), 4.3.3 (STITEL), 4.4.3 (STIPA) or 4.5.3 (RASTI) also apply with this equipment.

NOTE If the MLS test signal is first passed through a linear filter of sufficient impulse-response duration to result in an approximately Gaussian amplitude distribution (for example, the speech filter that is used when compensating for noise (see 4.5.6b)) distortions, such as those due to slewing rate limitations in the amplifier, are correctly treated as noise, rather than as a false echo. See also [4].

- a) Measurements may be made direct or by digital recordings of the test signals and the system response (from the test microphone) for post-processing. If the system input is via a close-talking microphone, the test signal should be applied from an artificial mouth (see ITU-T Recommendation P.51). If the design of the microphone makes this physically impracticable, a suitable transducer, such as a small, single-source, high-quality loud-speaker (cone diameter not exceeding 100 mm), should be used and described with the results.
- b) The test equipment should be set up to provide a sample length of at least 1 s, and the speech-shaping filter should be used. For an STI test, the 1 dB bandwidths of both the test signal (prior to the speech-shaped filter) and of the receiving section should be at least 88 Hz to 11,3 kHz.
- c) Specific types of measuring equipment may require the imposition of other special measuring conditions.

NOTE Averaging and ripple-reduction techniques are not appropriate for this measurement.

4.6.5 Simulation of occupancy noise

This may be accomplished as follows.

- a) If direct (cabled) recovery of the system response is used, noise of the same spectral distribution as the expected occupancy noise should be mixed with the recovered signal at the correct level to give the expected ambient signal-to-noise ratio (see note).
- b) If digital recording of the system response is used, noise of the same spectral distribution as the expected occupancy noise should be mixed with the recorded signal, prior to post-processing and at the correct level to give the expected ambient signal-to-noise ratio (see note).

NOTE Depending on particular artefacts of certain equipment/software, it may be necessary to set the noise level 3 dB higher than that needed simply to give the expected signal-to-noise ratio.

4.6.6 Repetition of measurements

The measurements should be repeated several times and the results averaged. Also, an estimate of the standard deviation should be derived from them and included with the results.

4.6.7 Analysis and interpretation of the results

It is important to examine the modulation reduction matrix to determine the reliability of the results.

As a rule, the values in each octave-band column should decrease with increasing modulation frequency. Constant or slightly reducing values in a column indicate the presence of noise. Large reductions indicate that reverberation is the main effect. Values that first reduce and then increase with modulation frequency indicate the presence of periodic or strong reflections, which may produce an over-optimistic conclusion. It is recommended that if this effect is detected, it should be reported with the results and an estimated correction applied.

5 Methods of determining intelligibility

NOTE 1 In spoken tests, embedding the words in carrier phrases results in representative reverberation/echoes during the presentation of the test words.

NOTE 2 See also [5].

5.1 Word tests

The limitations are given in ISO 4870. 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. 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).

5.2 Modified rhyme tests

The limitations are similar to those given in ISO 4870. 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.

5.3 Speech Intelligibility Index

NOTE The speech intelligibility index is also referred to as the articulation index (AI).

The limitations are given in [6].

5.4 Articulation loss of consonants

The limitations are similar to those given in ISO 4870. It should be noted that the measurement procedure does not include vowels. This may cause a systematic error with respect to word tests [12]. As the test is based on the reception of words by listeners, there are no limitations in respect of the characteristics of the sound system or those of the environment.

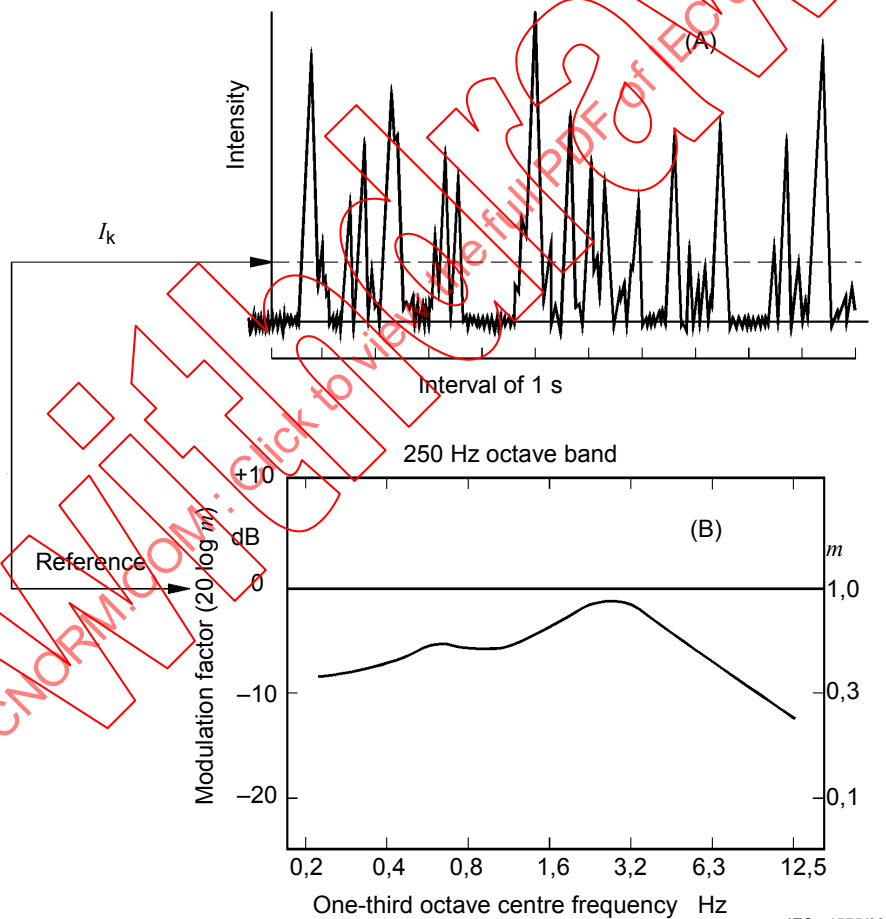
Annex A (normative)

Speech transmission index (STI) and revised (STI_r) 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. Distortion of the speech signal, such as by noise or reverberation, results in a reduction of the spectral differences between phonemes and this is reflected in a reduction in the fluctuations in the envelope function.



IEC 1575/03

NOTE The spectrum is normalized with respect to the mean intensity I_k .

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)

Figure A.1 (panel A) shows the envelope function for the 250 Hz octave frequency band. The shape of the envelope function is unique for a specific sequence of phonemes. The envelope spectrum gives a more general description of the fluctuations of the envelope function and results from a one-third octave-band analysis of the envelope function. Typically a 1 min speech excerpt is analysed to reflect the spectral distribution of the envelope fluctuations about the mean intensity so giving the modulation index as a function of modulation frequency (see Figure A.1 (panel B)).

Comparison of the envelope spectra obtained direct from the speaker with corresponding spectra obtained via a transmission path gives the reduction in fluctuations due to the path and leads to the modulation transfer function or MTF, which represents the reduction factor of the modulation index as a function of modulation frequency.

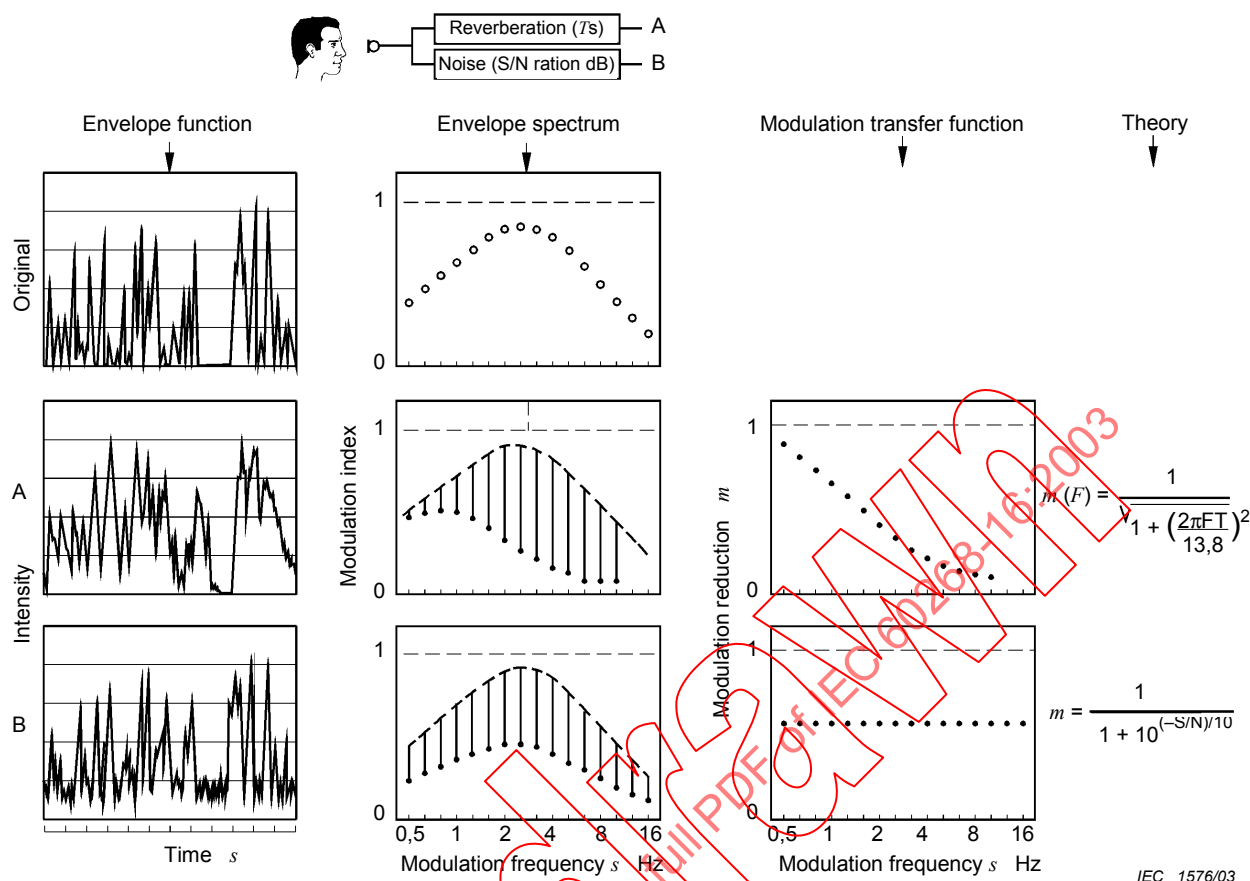
A.1.2 Modulation transfer function (MTF)

The rationale underlying the application of the MTF concept to studies of room acoustics has been described elsewhere [3], [7], [8], [9]. 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 reverberation only (case A; $T = 2,5$ s) and the other with only interfering noise (case B; signal-to-noise ratio $S/N = 0$ dB).

With reverberation, the MTF has the shape of a low-pass filter; the faster fluctuations being relatively more affected. In the theoretical case of purely exponential reverberation, the MTF can be derived mathematically (see Figure A.2, case A) and the FT product ($F =$ modulation frequency, $T =$ reverberation time) determines the roll-off. For noise, the MTF is defined by the signal-to-noise ratio and is independent of the modulation frequency. The noise, by increasing the mean intensity, reduces the modulation index for all modulation frequencies.

A.1.3 Determining the MTF

The MTF of a sound 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 the input to the path. Thus, a) speech signals, or b) the impulse response, or c) special test signals can be used.



NOTE The reductions in the octave-band specific envelope of an output signal (A or B) relative to the original signal can be expressed by the MTF. The two conditions considered (reverberation, A, or noise interference, B) lead to characteristic MTFs according to the theoretical expressions shown.

Figure A.2 – Theoretical expression of the MTF

- The use of speech signals has the advantage that the MTF can be determined under fully realistic conditions, for example, a live performance, but is only applicable for temporal distortion and noise and less accurate [9].
- The impulse response can be used to determine the MTF and the effect of reverberation or echoes on it.

NOTE Impulse signals are not suitable where background noise, bandpass limiting and non-linear distortion (sound distribution systems) are significant since the average speech spectrum and level distribution are not represented in the test signal.

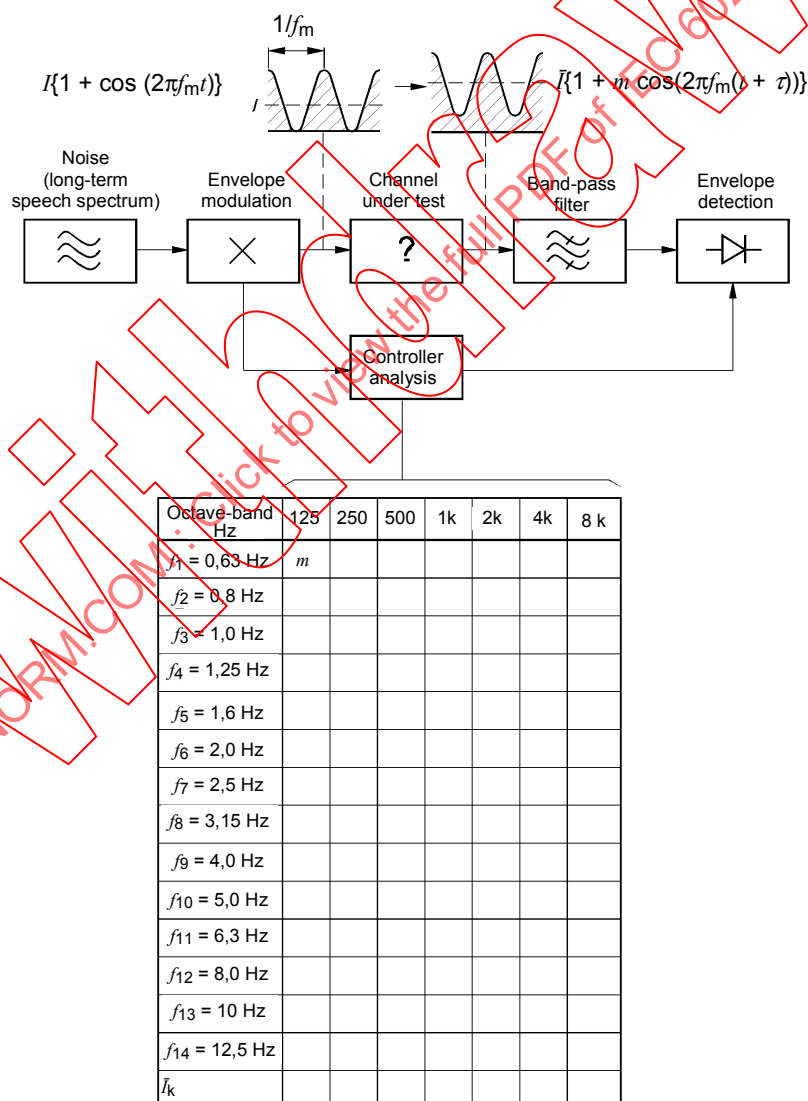
- Appropriate artificial test signals allow the determination of the modulation reduction factor for each modulation frequency and measurements are performed for octave-band centre frequencies typically from 125 Hz to 8 kHz. For a determination representative of the signal-to-noise ratio, the mean intensity of the test signal should be equivalent to the normal speech level at the test position, i.e. the L_{eq} of the test signal is adjusted to the typical L_{eq} of on-going speech in that position.

A.2 The STI method

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 are given in [11]. 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.

Distortion in the time domain (reverberation, echoes and automatic gain control) may degrade the fluctuating speech signal and reduce 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, diagrammatically, a measuring arrangement in which the modulation transfer function, $m(F)$, is determined separately for each modulation frequency in each octave band.



IEC 1577/03

NOTE 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 (I_k , equal to the square of the sound pressure levels) are obtained for use in calculating the auditory spread of masking, and the absolute reception threshold.

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

A.2.2 Non-linearity distortions

A speech signal that passes through a system with a non-linear transfer function (for example, with peak clipping) acquires harmonic and intermodulation components in other frequency bands. For this reason, the test signal should not be modulated with the same modulation frequency simultaneously for all the octave bands, because, in that case, the distortion components due to non-linearity cannot be distinguished from the modulated test signal in the frequency band under consideration.

However, by modulating with uncorrelated signals all the other frequency bands except that under test the distortion components appear as noise, reducing the effective signal-to-noise ratio, just as with linear distortion (uneven frequency response).

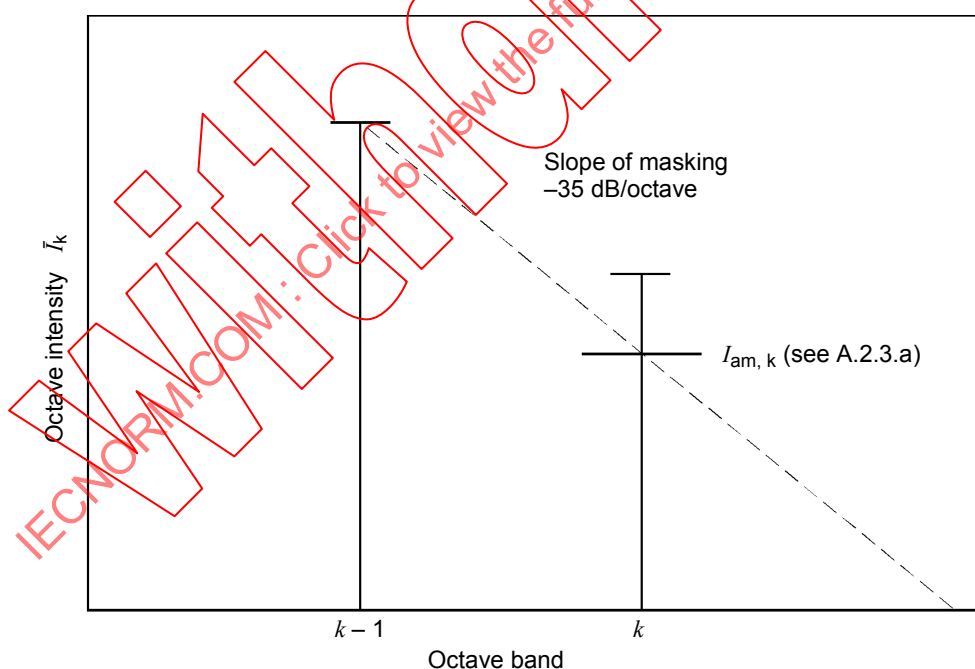
A.2.3 Other auditory effects

Auditory masking (the reduction in aural sensitivity by a stronger, lower frequency sound) [10] and the absolute reception threshold are allowed for by modelling appropriate noise terms and adding them in, thus reducing further the effective signal-to-noise ratio and the modulation transfer function. For this, the signal levels in each of the frequency bands have to be considered.

NOTE The auditory masking depends on the signal level. See Table A.1 and [14].

NOTE The absolute reception threshold is defined by the absolute threshold of hearing and the minimal required dynamic range for correct recognition of speech sounds.

Figure A.4 shows the effect of masking in one octave band by the octave band immediately lower and for an octave level between 56 dB and 65 dB SPL.



IEC 1578/03

NOTE The slope of -35 dB per octave equates to an auditory masking factor (amf) of 0,000 316.

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

- a) Modelled in this way, the effect is independent of the band frequency but depends on the absolute signal level [14]. The masking response is seen to be a straight line with a slope depending on the signal level (see Table A.1). The corresponding auditory masking factor (*amf*) is also given in Table A.1. This factor, applied to the intensity in one octave band, gives the masking effect on the octave band immediately above, namely:

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

where

$I_{am,k}$ is the intensity of the audio masking in octave band k ;

I_{k-1} is the intensity of the masking signal in band $(k - 1)$.

- b) The modelling of the absolute reception threshold is as the lower limit of the masking noise level within each octave band ($I_{rs,k}$) and only applies if the intensity, I_k , is the operational, i.e. normal (listening) level (see Table A.2).

Table A.1 – Octave level specific slope of masking and corresponding auditory masking factor (*amf*)

Octave level dB	46-55	56-65	66-75	76-85	86-95	>95
Slope of masking	–40	–35	–25	–20	–15	–10
Auditory masking factor	0,000100	0,000316	0,003162	0,010000	0,031622	0,100000

Applying the corrections for masking and threshold gives:

$$m'_{k,f} = \frac{m_{k,f} I_k}{I_k + I_{am,k} + I_{rs,k}}$$

where

$m_{k,f}$ is the modulation index for octave band k and modulation frequency f ;

$m'_{k,f}$ is the corrected index.

The effective signal-to-noise ratio for octave band k and modulation frequency f then becomes:

$$SNR_{k,f} = 10 \lg \frac{m'_{k,f}}{1 - m'_{k,f}} \text{ dB}$$

According to the STI concept, signal-to-noise ratios in the range –15 dB to +15 dB are linearly related to contributions to intelligibility in the range 0 to 1. Therefore the effective signal-to-noise ratio is converted to a corresponding transmission index ($TI_{k,f}$), specific for octave band k and frequency f , by the equation:

$$TI_{k,f} = \frac{SNR_{k,f} + \text{shift}}{\text{range}}$$

This makes $0 \leq TI_{k,f} \leq 1,0$ if shift is 15 dB and range 30 dB.

Figure A.5 shows this graphically.

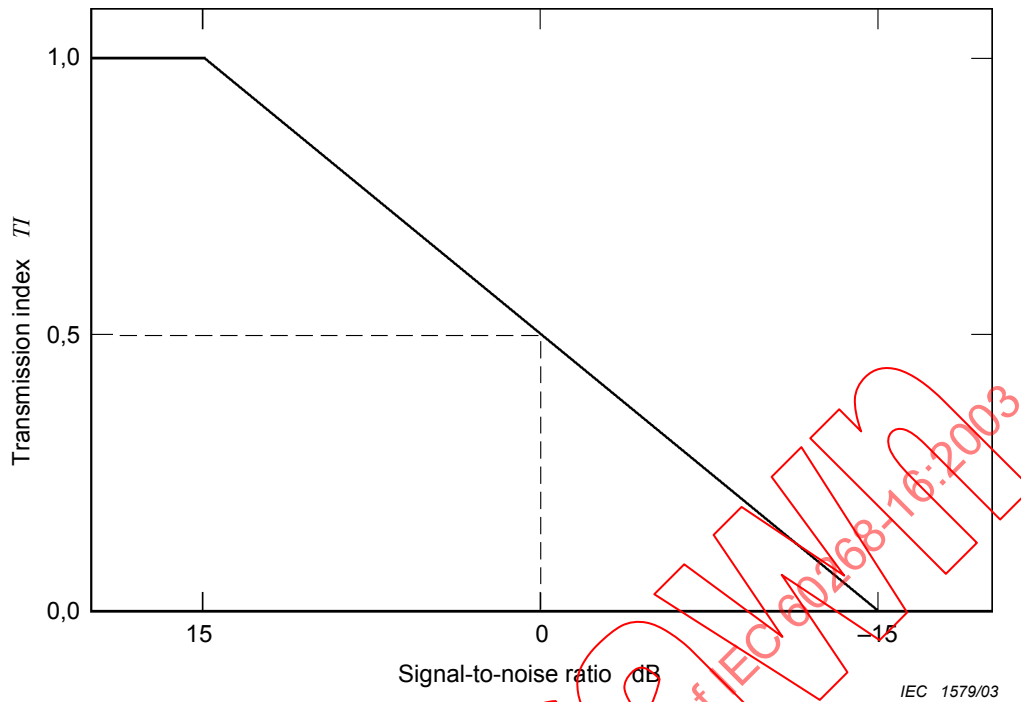


Figure A.5 – The relationship between effective signal-to-noise ratio and transmission index for a shift of 15 dB and a range of 30 dB

All 14 transmission indices for the modulation frequencies from 0,63 Hz to 12,5 Hz are obtained for each octave band and averaged for each band to give the modulation transfer index (MTI_k) specific for the contribution of octave band k, namely:

$$MTI_k = \frac{1}{14} \sum_{f=1}^{14} TI_{k,f}$$

Finally, the revised speech transmission index, STI_r , which takes account of auditory masking and the reception threshold, is obtained by a weighted summation of the modulation transfer indices for all seven octave bands and the corresponding redundancy corrections.

$$STI_r = \sum_{n=1}^7 \alpha_n MTI_n - \sum_{n=1}^6 \beta_n \sqrt{MTI_n \times MTI_{(n+1)}}$$

where

factors α_k represent the octave weighting factors;

β_k are the so-called redundancy correction factors.

This redundancy correction is related to the contribution of adjacent frequency bands.

A.2.4 Factors for male and female speech

The optimal weighting factors and redundancy factors for male and female speech are shown in Table A.2 along with the threshold correction.

Table A.2 – STI_r octave band specific male and female weighting factors

Octave band Hz		125	250	500	1 k	2 k	4 k	8 k
Males	α	0,085	0,127	0,230	0,233	0,309	0,224	0,173
	β	0,085	0,078	0,065	0,011	0,047	0,095	–
Females	α	–	0,117	0,223	0,216	0,328	0,250	0,194
	β	–	0,099	0,066	0,062	0,025	0,076	–
Absolute reception threshold dB	$L_{rs,k}$	46	27	12	6,5	7,5	8	12

A.3 The test signals

A.3.1 Spectra of STI, STITEL, and STIPA signals

The test signal spectra are specified by octave band mean levels (which may instantaneously be exceeded by 3 dB by a modulated test signal). The octave band levels are normalized to an A-weighted level of 0 dB. This coincides with the nominal long-term A-weighted speech level. For male speech, seven, and for female, six, octave bands are included in the test signals.

Table A.3 – Octave band levels (dB) relative to the A-weighted long-term speech level

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

A.3.2 STI modulation frequencies

The STI modulation frequencies are 0,63; 0,80; 1,00; 1,25; 1,60; 2,00; 2,50; 3,15; 4,00; 5,00; 6,30; 8,00; 10,0; 12,5 Hz with all 14 frequencies applied successively to all seven octave bands.

A.3.3 Form of signal for STI (and STITEL, STIPA)

The noise carrier for each octave band has a bandwidth of half an octave to avoid spillover from adjacent bands and to allow for practical filters in the analyser. Therefore, for male speech there are seven, and for female six, half-octave components at octave intervals.

The intensity envelope of the noise carrier for each octave band is sinusoidally modulated with the modulation frequency or frequencies. In the amplitude domain, the modulation function is the square root of the desired intensity envelope. For example:

$$\sqrt{1 + \cos 2\pi \cdot f_m t}$$