



100/3202/CDV

COMMITTEE DRAFT FOR VOTE (CDV)

PROJECT NUMBER:

IEC 60268-16 ED5

DATE OF CIRCULATION:

2019-03-08

CLOSING DATE FOR VOTING:

2019-05-31

SUPERSEDES DOCUMENTS:

100/3012/CD, 100/3171/CC

IEC TA 20 : ANALOGUE AND DIGITAL AUDIO

SECRETARIAT:

Japan

SECRETARY:

Mr Gen Ichimura

OF INTEREST TO THE FOLLOWING COMMITTEES:

PROPOSED HORIZONTAL STANDARD:

☐

Other TC/SCs are requested to indicate their interest, if any, in this CDV to the secretary.

FUNCTIONS CONCERNED:

☐ EMC

☐ ENVIRONMENT

☐ QUALITY ASSURANCE

☐ SAFETY

☒ SUBMITTED FOR CENELEC PARALLEL VOTING

☐ NOT SUBMITTED FOR CENELEC PARALLEL VOTING

Attention IEC-CENELEC parallel voting

The attention of IEC National Committees, members of CENELEC, is drawn to the fact that this Committee Draft for Vote (CDV) is submitted for parallel voting.

The CENELEC members are invited to vote through the CENELEC online voting system.

This document is still under study and subject to change. It should not be used for reference purposes.

Recipients of this document are invited to submit, with their comments, notification of any relevant patent rights of which they are aware and to provide supporting documentation.

TITLE:

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

PROPOSED STABILITY DATE: 2022

NOTE FROM TC/SC OFFICERS:

Copyright © 2019 International Electrotechnical Commission, IEC. All rights reserved. It is permitted to download this electronic file, to make a copy and to print out the content for the sole purpose of preparing National Committee positions. You may not copy or "mirror" the file or printed version of the document, or any part of it, for any other purpose without permission in writing from IEC.

CONTENTS

1		
2		
3	FOREWORD.....	7
4	INTRODUCTION.....	9
5	1 Scope.....	12
6	2 Normative references	12
7	3 Terms and definitions	12
8	4 Description of the STI model	17
9	4.1 Overview	17
10	4.1.1 Applicability of the STI model	18
11	4.2 Theoretical details	18
12	4.2.1 Envelope function and envelope spectrum.....	18
13	4.2.2 Reduction of modulation	19
14	4.2.3 Role of the octave-band noise carriers.....	19
15	4.2.4 Theoretical overview.....	20
16	4.3 Measurement of STI.....	22
17	4.3.1 Direct and Indirect Methods	22
18	4.3.2 FULL STI.....	22
19	4.3.3 STIPA.....	23
20	4.3.4 Choice of method	23
21	5 Direct method of measuring STI – User guidance	25
22	5.1 Overview	25
23	5.2 STIPA	25
24	5.3 Application.....	26
25	5.4 Limitations	26
26	6 Indirect method of measuring STI (impulse response) – User guidance	26
27	6.1 Overview	26
28	6.2 Application.....	27
29	6.3 Limitations (non-linear distortion).....	28
30	7 Limitations of the STI model	28
31	7.1 General.....	28
32	7.2 Frequency shifts	29
33	7.3 Centre clipping.....	29
34	7.4 Dropouts	30
35	7.5 Jitter	30
36	7.6 Digital voice compression systems.....	30
37	7.7 Overestimation of STI under low background noise conditions	30
38	7.8 Frequency response	30
39	7.9 Echoes	31
40	7.10 Fast amplitude compression and expansion.....	33
41	7.11 Non-linear distortion	34
42	7.12 Hearing impaired listeners	34
43	7.13 Impulsive and fluctuating noise	34
44	7.14 Conclusion.....	35
45	8 Measurement procedures, post-processing of data and applications.....	35
46	8.1 General.....	35

47	8.2	Acoustical input	35
48	8.3	Acoustical output	37
49	8.4	Electrical input	37
50	8.5	Electrical output	37
51	8.6	Examples of input/output combinations	37
52	8.6.1	Acoustical input – Acoustical output	37
53	8.6.2	Electrical input – Electrical output (e.g. assessment of wired and 54 wireless) communication systems)	38
55	8.6.3	Acoustical input – Electrical output (e.g. assessment of microphones)	38
56	8.6.4	Electrical input – Acoustical output (e.g. assessment of PA systems)	38
57	8.7	Spatial averaging of STI measurements	38
58	8.8	Post-processing of measured MTF data	39
59	8.9	Issues concerning noise	40
60	8.9.1	General	40
61	8.9.2	Measurement of background noise	40
62	8.9.3	Fluctuating noise	40
63	8.10	Analysis and interpretation of the results	41
64	8.11	Binaural STI measurements	41
65	9	Use of STI as a design prediction tool	42
66	9.1	Overview	42
67	9.2	Statistical predictions	42
68	9.3	Prediction from simulated impulse response	42
69	Annex A	(informative) The basis of the STI concept	44
70	A.1	Introduction to this Annex	44
71	A.1.1	Purpose	44
72	A.1.2	Modulation transfer function (MTF)	44
73	A.1.3	STI model	45
74	A.1.4	STI modulation frequencies	46
75	A.2	Calculation of the STI	46
76	A.2.1	General Equation for STI	46
77	A.2.2	Gender-specific octave band weighting and redundancy factors	47
78	A.2.3	Adjustment of the MTF for ambient noise	48
79	A.2.4	Adjustment of the MTF for auditory masking and threshold effects	48
80	A.3	Calculation of the modulation transfer ratio values	49
81	A.3.1	Direct method: Analysis of the STI test signal	49
82	A.3.2	Indirect method: Determination of the modulation transfer function (MTF) 83 50	49
84	A.4	Auditory effects on the STI	51
85	A.4.1	Overview	51
86	A.4.2	Level-dependent auditory masking	51
87	A.4.3	Absolute speech reception threshold	53
88	A.5	Generation of the STI test signal (direct method)	54
89	A.5.1	Pink noise source signal	54
90	A.5.2	Generating octave band carrier signals	54
91	A.5.3	Intensity modulation of the carrier signals	54
92	A.5.4	Applying the speech spectrum to the STI test signal	55
93	A.6	Spectrum of STI test signal	55
94	A.6.1	Speech shaped noise	55
95	Annex B	(normative) STIPA method	57

96	B.1	Overview	57
97	B.2	Test signal	57
98	Annex C	(normative)	59
99	C.1	Specification of the measuring device	59
100	C.2	Signals for testing STI implementations	59
101	C.3	Testing the dynamic range in the modulation domain	59
102	C.3.1	Modulation depth testing for STIPA direct method	59
103	C.3.2	Modulation depth testing for STI indirect method	60
104	C.4	Testing of cross-talk between octave-band filters	61
105	C.4.1	Flank attenuation slopes	61
106	C.4.2	Octave band filter testing - STIPA direct method	61
107	C.4.3	Performance verification files	62
108	Annex D	(informative) Use of STI Measuring Devices	63
109	D.1	Overview	63
110	D.2	STIPA characterises only the speech transmission channel	63
111	D.3	Examples of test scenarios for STIPA tests	64
112	D.4	Equipment and resources needed for a STIPA test	66
113	D.4.1	Availability of the test signal	66
114	D.4.2	A source of the STIPA test signal	67
115	D.5	Steps in the overall procedure	67
116	Annex E	(informative) Qualification of the STI and relationships with other speech intelligibility measures	68
118	E.1	Relationship between STI and word/sentence scores	68
119	E.2	Relationship between STI and listening difficulty	68
120	Annex F	(informative) Nominal qualification bands for STI	70
121	Annex G	(informative) Examples of STI qualification bands and typical applications	71
122	Annex H	(informative) Non-native listeners	72
123	Annex I	(informative) Effect of age-related hearing loss and hearing impairment on speech intelligibility	73
125	Annex J	(normative) Setting & adjustment of STI test signal level	74
126	J.1	Overview	74
127	J.2	The concept of 'speech level' and the method of measurement	74
128	J.3	Real speech level	74
129	J.4	Corrected speech level derived from real speech level	75
130	J.5	Comparison of dynamic structures of speech and test signals	75
131	Annex K	(informative) Example test report sheet for STI measurements	76
132	Annex L	(normative) Prediction of STI using statistical methods	78
133	Annex M	(informative) Adjustments to STI data to simulate alternative ambient noise spectra and different speech levels	80
135	Annex N	(informative) Other methods of determining speech intelligibility	88
136	N.1	Overview	88
137	N.2	Word tests	88
138	N.3	Modified rhyme tests	88
139	N.4	Speech Intelligibility Index (SII)	88
140	N.5	PESQ	89
141	Annex O	(informative) Alternative direct methods for measuring Full STI	90
142	Annex P	(normative) Information to be provided by manufacturers	91
143	P.1	Purpose of this Annex	91

144	P.2	Form in which the information is to be provided	91
145	P.3	Required information.....	91
146	P.4	Declaration	91
147	Annex Q (informative)	Effect of uncertainties of selected parameters on STI	
148		uncertainty	92
149	Q.1	STI calculation framework.....	92
150	Q.1.1	Overview	92
151	Q.1.2	Statistical MTF	92
152	Q.1.3	Corrections.....	92
153	Q.1.4	Effective SNR.....	93
154	Q.1.5	Modulation Transfer Index MTI	93
155	Q.1.6	Speech Transmission Index STI	93
156	Q.2	The effect of RT uncertainty on STI uncertainty	93
157	Q.2.1	Modulation Transfer Function	93
158	Q.2.2	Uncertainty in STI.....	93
159	Q.2.3	Conclusions:.....	95
160	Q.3	The effect of S/N uncertainty on STI uncertainty	95
161	Q.3.1	Ideal transfer function	95
162	Q.3.2	Reverberation.....	96
163	Q.3.3	Conclusions:.....	97
164	Q.4	The effect of signal level uncertainty on STI uncertainty.....	97
165	Q.4.1	Overview	97
166	Q.4.2	Auditory masking	97
167	Q.4.3	Conclusions:.....	98
168	Bibliography.....		100
169			
170	Figure 1–	Envelope function (panel A) of a 10 s speech signal for the 250 Hz octave	
171		band and corresponding envelope spectrum (panel B)	19
172	Figure 2 –	Modulation transfer function – Input/output comparison.....	21
173	Figure 3 –	Effect of a single delayed arrival on the MTF (idealised conditions).....	32
174	Figure 4 –	Idealised STI (Male speech Spectrum) versus delay and level of secondary	
175		arrival	33
176	Figure A.1 –	Theoretical expression of the MTF	44
177	Figure A.2 –	Measurement system and frequencies for the STI method	46
178	Figure A.4 –	Auditory masking of octave band ($k-1$) on octave band (k).....	52
179	Figure D.1.	Schematic representation of the definition of a speech transmission	
180		channel.....	64
181	Figure E.1	Relationships between some speech intelligibility measures	68
182	Figure E.2 –	Relationship between STI, speech intelligibility scores and listening	
183		difficulty ratings [34], [35].....	69
184	Figure F.1 –	STI qualification bands.....	70
185	Figure Q.1 -	Uncertainty in absolute value of STI vs reverberation time RT with various	
186		degrees of uncertainty in RT	95
187	Figure Q.2 -	Uncertainty in absolute value of STI vs reverberation time RT with 1 dB	
188		uncertainty in SNR at various SNRs.....	97
189	Figure Q.3 -	Uncertainty in absolute value of STI versus reverberation time RT with	
190		various degrees of masking.	98

192	Table 1 How to use this standard	10
193	Table 2 – Comparison of direct and indirect methods	22
194	Table 3 – Suitability of STI test methods for different types of distortion	23
195	Table 4 – Test-method suitability	24
196	Table 5 – Measurement applications	24
197	Table A.1 – MTI octave band weighting factors	48
198	Table A.1 – Auditory masking as a function of the octave band level.....	53
199	Table A.2 – Absolute speech reception threshold level in octave bands	54
200	Table A.4 – Octave band levels (dB) relative to the A-weighted speech level	55
201	Table A.5 – Filter parameters and s-plane polynomials that produce speech-shaped	
202	pink noise.	56
203	Table B.1 – Modulation frequencies for the STIPA method.....	57
204	Table C 1 Specification of an STI measuring device	59
205	Table D.1 Scenario 1. PA with “live” announcer	64
206	Table D.2 Scenario 2. PA with pre-recorded announcements	65
207	Table D.3 Scenario 3. “Live” meetings and conversations	65
208	Table D.4 Scenario 4. Lecture	66
209	Table E.1 – Categories for listening difficulty	69
210	Table G.1. – Examples between STI qualification bands and typical applications	71
211	Table H.1. – Adjusted intelligibility qualification tables for non-native listeners.....	72
212	Table I.1. – Adjusted intelligibility qualification tables for normal listeners and people	
213	over 60 years old with hearing loss	73
214	Table J.2 Comparison of speech and the test signal	75
215	Table K.1 Example test report sheet	76
216	Table K.2 Measurement data record sheet.....	77
217	Table M.1. - Flow chart of post-processing adjustment steps	80
218	Table M.2 Example calculation	84
219		

INTERNATIONAL ELECTROTECHNICAL COMMISSION

SOUND SYSTEM EQUIPMENT –

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

FOREWORD

- 1) The International Electrotechnical Commission (IEC) is a worldwide organization for standardization comprising all national electrotechnical committees (IEC National Committees). The object of IEC is to promote international co-operation on all questions concerning standardization in the electrical and electronic fields. To this end and in addition to other activities, IEC publishes International Standards, Technical Specifications, Technical Reports, Publicly Available Specifications (PAS) and Guides (hereafter referred to as "IEC Publication(s)"). Their preparation is entrusted to technical committees; any IEC National Committee interested in the subject dealt with may participate in this preparatory work. International, governmental and non-governmental organizations liaising with the IEC also participate in this preparation. IEC collaborates closely with the International Organization for Standardization (ISO) in accordance with conditions determined by agreement between the two organizations.
- 2) The formal decisions or agreements of IEC on technical matters express, as nearly as possible, an international consensus of opinion on the relevant subjects since each technical committee has representation from all interested IEC National Committees.
- 3) IEC Publications have the form of recommendations for international use and are accepted by IEC National Committees in that sense. While all reasonable efforts are made to ensure that the technical content of IEC Publications is accurate, IEC cannot be held responsible for the way in which they are used or for any misinterpretation by any end user.
- 4) In order to promote international uniformity, IEC National Committees undertake to apply IEC Publications transparently to the maximum extent possible in their national and regional publications. Any divergence between any IEC Publication and the corresponding national or regional publication shall be clearly indicated in the latter.
- 5) IEC itself does not provide any attestation of conformity. Independent certification bodies provide conformity assessment services and, in some areas, access to IEC marks of conformity. IEC is not responsible for any services carried out by independent certification bodies.
- 6) All users should ensure that they have the latest edition of this publication.
- 7) No liability shall attach to IEC or its directors, employees, servants or agents including individual experts and members of its technical committees and IEC National Committees for any personal injury, property damage or other damage of any nature whatsoever, whether direct or indirect, or for costs (including legal fees) and expenses arising out of the publication, use of, or reliance upon, this IEC Publication or any other IEC Publications.
- 8) Attention is drawn to the Normative references cited in this publication. Use of the referenced publications is indispensable for the correct application of this publication.
- 9) Attention is drawn to the possibility that some of the elements of this IEC Publication may be the subject of patent rights. IEC shall not be held responsible for identifying any or all such patent rights.

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

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

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

The spectrum of the male speech test signal has been changed, with significant reductions in the 125 Hz and 250 Hz bands being implemented. Some corrections to formulae have been made. Additional information has been included on prediction and measurement procedures.

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

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

FDIS	Report on voting
100/XX/FDIS	100/XX/RVD

276
277 Full information on the voting for the approval of this standard can be found in the report on
278 voting indicated in the above table.

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

280 The committee has decided that the contents of this publication will remain unchanged until
281 the stability date¹ indicated on the IEC web site under "<http://webstore.iec.ch>" in the data
282 related to the specific publication. At this date, the publication will be

- 283 • reconfirmed,
- 284 • withdrawn,
- 285 • replaced by a revised edition, or
- 286 • amended.

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

288

INTRODUCTION

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

To quantify the deterioration of the speech intelligibility induced by the transmission channel, a fast and objective measuring method was developed; the Speech Transmission Index (STI).

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

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

The STI method has been the subject of ongoing development and refinement since its introduction in the 1970s. Major improvements of the STI have been consolidated by incorporating them in successive revisions of IEC 60268-16.

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

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

NOTE The STI method was not designed for the measurement and evaluation of speech privacy or speech masking systems and therefore has not been validated for these situations. It is not recommended to use STI below 0.3, but if this is to be undertaken, specialist expertise and techniques beyond the scope of this standard are required.

Potential users of STI

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

- certifiers of voice alarm and other types of emergency systems;
- certifiers of sound reinforcement and audio systems;
- audio and telecommunication equipment manufacturers;
- audio and communication engineers;
- acoustic and electroacoustic consultants;
- audio and telecommunication equipment manufacturers;

- sound system installers;
- researchers into STI methods and developers of instruments to measure STI.

Table 1 How to use this standard

Purpose	Topic	Clauses
All users	Introduction to the STI method	
Routine check of voice-alarm or sound system with STIPA	Direct method of measuring STI	4.0
In-depth check of or to certify sound system with STIPA and/or impulse response methods	Description of the STI method	5.0
	Direct method of measuring STI	4 and 5
	Indirect method of measuring STI using the impulse response	4 and 6
	Measurement procedures, and applications	8.0
	Post-processing of measured MTF data	8.8
	Limitations of the measurement methods	5.4, 6.3
	Optional: Theory and equations governing STI methods	Annexes A and B
	Optional: Relationship between subjective and objective measures of intelligibility	Annex F
	Optional: Measurement uncertainties	Annex E
Measure telecommunication equipment	Direct method only	8.6.2
Manufacturer of STIPA device	Theory and equations governing STI methods	Annexes A and B
	Verification of STI measurement device performance	Annex C
	Information to be provided	Annex D
Manufacturer of acoustical analyser and simulation software	Theory and equations governing STI methods	Annex A
	Calibration of STI instruments	Annex C
	Information to be provided	Annex D
Research into intelligibility	Theory and equations governing STI methods	Annex A and B
Using simulation software	Prediction methods	Annex M
Post processing of STI and STIPA measurement	Post processing measurement results	Annex M
	Optional - As per in-depth measurements of STI listed above	
	Optional -Worked calculation example	Annex N
Evaluation of the potential intelligibility of Assistive Listening Systems	As per in-depth measurements of STI listed above	
	Special process for Assistive Listening Systems	8.6.3

Items that have changed in Revision 5.

Specific changes that have been incorporated in this revision are:

- changes to the male speech spectrum shown in Table A.4
- corrections to an equation in Annex L
- Greater information is given in Annex M about adjustments to the measured STI results to simulate the effects of alternative ambient noise and speech levels.
- Greater displayed precision is given for the results of the example calculation in Annex M.

- Correction of Schroeder equation for MTF in Clause 6.1.
- Spectrum and weighting factors for female speech have been removed.
- Verification information for STI measurement devices added

Revision history

The history of revisions is as follows:

- Revision 1: 1988. In the first version of the STI standard, a gender-independent test signal spectrum was used.
- Revision 2: 1998. Gender specific test signals were introduced, for male and female talkers, each gender relating to a specific set of weighting factors. In addition, weightings were introduced for redundancy factors. The term STI_r was introduced to signify the use of these redundancy factors.
- Revision 3: 2003. Important differences between Revision 2 and Revision 3 are the introduction of:
 - level dependent masking functions,
 - the STI derivative STIPA.
 - STIPA was specially developed as a fast measurement method that could deal with electro-acoustic and acoustic effects while determining the speech transmission quality of PA systems.
- Revision 4: 2011.
 - The terms STI_r and Room Acoustic Speech Transmission Index (RASTI) were discontinued.
 - A new function for the prediction of auditory masking effects was introduced.
 - STI corrections for non-native language listeners and some forms of hearing loss were introduced.

SOUND SYSTEM EQUIPMENT –

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

1 Scope

This standard defines the STI model, test signals, measurement and prediction methods.

The objective of this standard is to provide a comprehensive manual for all types of users of the STI model 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, but some typical application values are provided in Annex G.

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

This standard does not cover the case of fluctuating noise on the STI, although some general comment on dealing with this complex issue is provided in 7.13 and 8.9.3

2 Normative references

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

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

ISO 9921:2003, *Ergonomics — Assessment of speech communication*

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

IEC 60318-1:2009 Electroacoustics - Simulators of human head and ear – Ear simulator for the measurement of supra-aural and circumaural earphones

IEC 60318-7:2011 Electroacoustics - Simulators of human head and ear – Head and torso simulator for acoustic measurement of hearing aids

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

3 Terms and definitions

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

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

3.1**speech intelligibility**

rating of the proportion of speech that is understood

3.2**speech quality**

rating of sound quality of a speech signal

3.3**speech transmission index****STI**

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

3.4**speech intelligibility index SII**

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

3.5**STI model**

framework for quantifying the potential effect that a transmission path between a talker and listener has on speech intelligibility. The model predicts the speech intelligibility based on the degree to which the intensity modulations of speech are preserved during transmission.

3.6**STI model FULL STI**

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

3.7**distortion**

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

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

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

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

3.9**direct method**

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

3.10**indirect method**

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

3.11**speech transmission channel**

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

3.12

public address system PA

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

3.13

emergency sound and communication system

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

3.14

real speech level

sound pressure level of the broadcast speech signal that is measured and is used to derive the corrected speech level (typically by adding 3 dB). The difference between the real speech level and the corrected speech level represents the reduction in the long-term level produced by the pauses and silences between words

3.15 corrected speech level

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

3.16

vocal effort

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

3.17

artificial mouth

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

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

3.18

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

3.19

absolute speech reception threshold

threshold of hearing increased by the minimum required dynamic range to enable recognition of speech

3.20

auditory masking

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

Note 1 to entry Within the STI model, auditory masking is also referred to as the upward spread of masking.

3.21

artificial ear

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

Note 1 to entry See IEC 60318 [2].

3.22**intensity function**

the squared amplitude signal as a function of time

3.23**envelope function**

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

3.24**envelope spectrum**

relative contribution of spectral components of the envelope function

3.25**modulation frequency F**

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

3.26**modulation frequency f_m**

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

3.27**modulation depth**

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 depths of the intensity functions of the received and original (transmitted) signals – also referred to as the m -value

3.29**modulation transfer function MTF**

modulation transfer ratio as a function of the modulation frequency

3.30**modulation transfer index MTI**

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

3.31**transmission index TI**

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

3.32**octave band weighting factor α**

relative contribution of each octave band to the speech transmission index

3.33**octave band redundancy factor β**

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

537 **3.34**
 538 **background noise**
 539 all sounds comprising stationary, fluctuating and impulsive noise remaining in the absence of
 540 the speech or test signal

541 **3.35**
 542 **stationary noise**
 543 continuous noise with approximately-constant level

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

545 **3.36**
 546 **impulsive noise**
 547 sound or noise characterized by short individual bursts of sound pressure

548 **3.37**
 549 **fluctuating noise**
 550 continuous sound or noise whose sound pressure level varies over time, but not in an
 551 impulsive manner, during the observation period

552 **3.38**
 553 **signal-to-noise ratio SNR**
 554 difference in dB between the sound pressure level of the speech or test signal and the sound
 555 pressure level of the background noise where the sound pressure levels are determined with
 556 a standardized frequency weighting

557 **3.39**
 558 **effective signal-to-noise ratio SNR_{eff}**
 559 The modulation transfer function transformed into the signal-to-noise ratio domain - expressed
 560 in dB

561 **3.40**
 562 **crest factor**
 563 difference in dB between the peak and the RMS sound pressure levels during a given time-
 564 interval

565 **3.41**
 566 **Lombard effect**
 567 spontaneous increase of the vocal effort induced by the increase of the ambient noise level at
 568 the speaker's ear

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

570 **3.42**
 571 **fractional-octave-band filter**
 572 bandpass filter for which the ratio of upper cut-off frequency f_2 to lower cut-off frequency f_1 is
 573 two raised to an exponent equal to the fraction of an octave band

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

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

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

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

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

3.43**reference sound pressure P_0**

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

3.44**sound pressure level L_p**

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

3.45**equivalent continuous sound pressure level $L_{eq,T}$**

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

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

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

with

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

t the integration variable for time;

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

p_0 the reference sound pressure (20 µPa).

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

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

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

4 Description of the STI model**4.1 Overview**

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

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

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

A speech signal level varies rapidly with time producing variations (or fluctuations) in the intensity envelope of the sound. Slower fluctuations of this intensity envelope correspond with

word and sentence boundaries while faster fluctuations coincide with individual phonemes within words, Phonemes are the fundamental elements of speech and connected discourse can be considered as a sequence of phonemes.

The STI concept is based on the empirical finding that these fluctuations carry the most relevant information relating to speech intelligibility, and preservation of the intensity envelope is considered to be of the utmost importance see [5], [6] and [7]. Time-domain distortions within a transmission channel (such as reverberation, echoes and automatic gain control) along with noise can degrade the fluctuating speech-signal and reduce the intelligibility. The extent of degradation in the fluctuations determines the potential speech intelligibility and the STI model measures the degree to which the fluctuations are preserved.

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

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

Research [29] 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 led to the use of redundancy factors. Equations used in the STI model and more technical details are presented in Annex A

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

In contrast to the approach of the articulation index [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 or fluctuations of the speech signal.

4.1.1 Applicability of the STI model

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

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

4.2 Theoretical details

4.2.1 Envelope function and envelope spectrum

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

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

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

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

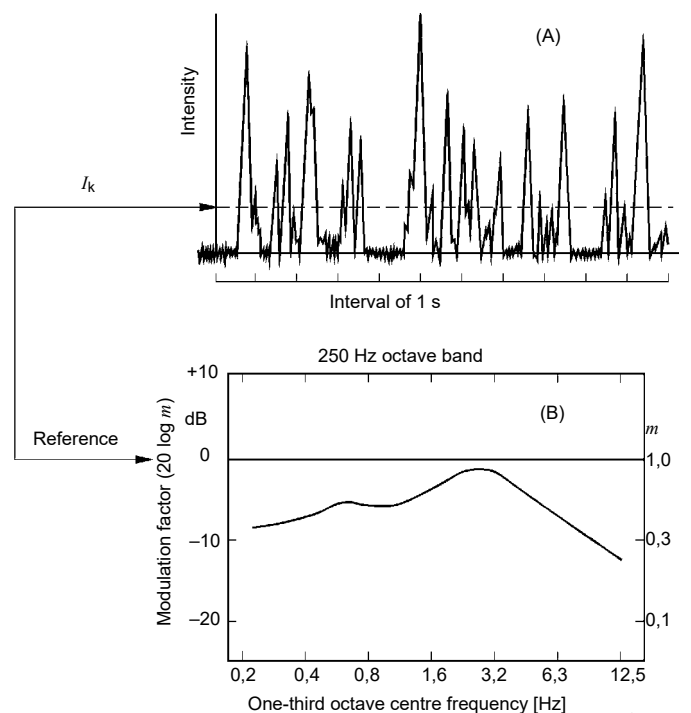


Figure 1– Envelope function (panel A) of a 10 s speech signal for the 250 Hz octave band and corresponding envelope spectrum (panel B)

4.2.2 Reduction of modulation

Any deterioration of the modulation spectrum by the transmission channel is generally considered to result in a reduction of the speech intelligibility. This deterioration of the modulation spectrum corresponds to a reduction of the modulation depth at one or more modulation frequencies and is calculated as a modulation transmission value for each octave band over the speech spectral range. Figure 2 shows the concept of the reduction in modulation that can occur between a talker (input) and listener (output).

4.2.3 Role of the octave-band noise carriers

The STI test signal was developed from parameters derived from speech material. In general, the STI test signal comprises noise signals in the seven octave-bands ranging from 125 Hz to

8 kHz. As the noise signals in these octave-bands carries modulation signals, they are termed noise carriers. Each noise carrier is modulated with one or more modulation frequencies at one-third octave intervals ranging from 0,63 Hz up to and including 12,5 Hz.

The STI model determines the modulation transfer function $m(F)$ of the transmission channel. In the Full STI method, a total of 98 results are obtained, corresponding to the 14 modulation frequencies and the seven octave bands (see Figure A.3). The RMS level of each octave-band carrier matches the relative level of the average, long term spectrum of speech material (see also 4.5 for further information). Each octave band has a contribution to speech intelligibility which is weighted according that band. Using the weighted sum of these transmission index values, the overall STI value for the transmission channel is determined. Its description and the octave- band weighting factors and redundancy factors are given in [29].

4.2.4 Theoretical overview

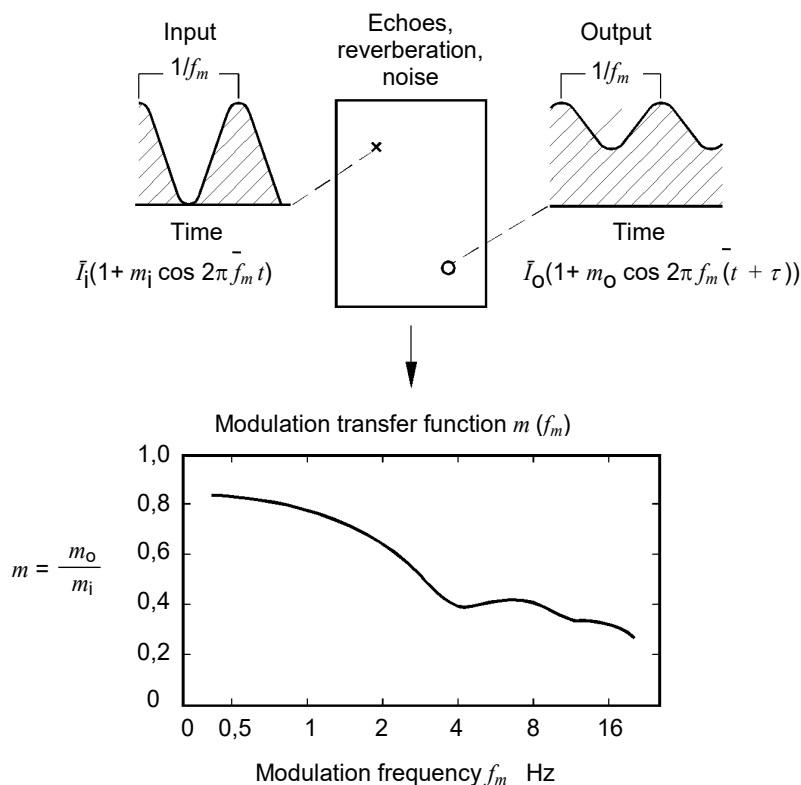
By proper choice of the form of test signal, the effective signal-to-noise ratio can include and allow for distortions in the time domain and non-linearities as well as background noise, etc.

This is modelled in the STI procedure by determining the modulation transfer function for the range of relevant frequencies present in the envelope of natural speech signals. The relevant range for these modulation frequencies extends from 0,63 Hz to 12,5 Hz in 14 one-third octave bands. Figure A.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 modulation depth m_i of a test signal is played into a room or through a communication channel and received at a listener position with degraded modulation depth m_o . The test signal would be transmitted by a sound source simulating a human talker situated at the talker's position with a receiving test microphone located at any listener position.

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

The typical test signal consists of a carrier with a speech-shaped frequency spectrum and a sinusoidal intensity modulation with modulation frequency f_m as illustrated in Figure 2.



IEC 1572/03

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

Figure 2 – Modulation transfer function – Input/output comparison

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

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

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 then limited to the range of –15 dB to 15 dB. Values less than –15 dB are given the value of –15 dB and values greater than 15 dB are given the value of 15 dB.

The effective signal-to-noise ratios are used to calculate the modulation transfer index (MTI) in each octave band. The speech transmission index STI combines the MTI values from measurements in seven octave bands into one overall weighted value.

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

4.3 Measurement of STI

4.3.1 Direct and Indirect Methods

There are two methods to measure STI:

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

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

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

Table 2 – Comparison of direct and indirect methods

Subject	Direct method	Indirect method ^b
Post processing	possible	mandatory
Handheld device	possible	possible
Amplitude nonlinearities	reduce the reliability of the result	reduce the reliability of the result
Frequency response nonlinearities (Uneven spectrum) ^a	possible	possible
Frequency shift	not possible	not possible
Noise suppression	no	yes
Sample rate accuracy between the clock frequencies of the signal source and receiver	errors less than 20×10^{-6}	errors less than $0,5 \times 10^{-6}$
^a See 7.8 for further details		
^b Different methods of deriving the impulse response may produce small differences		

4.3.2 FULL STI

The research described in [5], [6], [7], [18], [19], [20] and [29] developed the basis and method for the Full STI.

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

An alternative version of the Full STI signal contains random modulations in the other octave bands in addition to the modulation frequency and octave band under test.

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

The indirect method is also widely employed for FULL STI.

4.3.3 STIPA

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

The STIPA test signal consists of only one test signal with a predefined set of two modulations in each of the seven octave bands. The 14 modulations are generated simultaneously. One measurement takes between 15 s and 25 s.

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

4.3.4 Choice of method

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

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

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

In addition, the direct STI methods account for non-linear distortion, e.g. clipping, whereas the indirect methods should only be used for linear systems. Additional information about the effects of non-linear distortion is given in chapter 6. Table and Table give an overview of the STI test methods versus the types of linear and non-linear distortion for which they are appropriate. The + and - symbols are a general indication of the suitability of the method.

Table 3 – Suitability of STI test methods for different types of distortion

Method	Type of distortion				
	Noise	Reverberation	Echoes	Non-linear distortion	Spectral distortion ^a
Direct FULL STI	yes	yes	yes	condition dependent	yes
Direct STIPA	yes	yes	limited	condition dependent	yes
Indirect FULL STI ^c	No ^b	yes	yes	no	yes
Indirect STIPA ^c	No ^b	yes	limited	no	yes
<p>NOTE The term 'condition dependent' is used to indicate that the corresponding test signal type might or might not produce sufficiently accurate results, depending on the exact distortion type. For example:</p> <ul style="list-style-type: none"> • Centre clipping is unlikely to have any effect on the modulation depth, whereas peak clipping reduces the modulation depth but generally has little effect on the intelligibility of speech, so the measured STI value might be pessimistic. • STIPA can be used for PA systems that produce non-linear distortion components, unless the signal is severely clipped in various frequency bands. 					
<p>^a The frequency response of the transmission channel might produce a perceived loss of intelligibility that is not adequately accounted for in the result, see 7.8.</p>					

- b Yes, if a MLS test signal is used, however signal averaging of time domain data shall not be employed and the excitation spectrum shall be speech-shaped.
- c This includes time delay spectrometry.
Theoretically, other mathematically deterministic pseudo-noise (random phase) signal could be employed.
The effects of noise should be computed mathematically.

793

794

Table 4 – Test-method suitability

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

795

796 Table 5 provides an overview as to which forms of STI are recommended for various types of
797 application. The + and - symbols are a general indication of the suitability of the method.

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

800

Table 5 – Measurement applications

Application	FULL STI Direct	STIPA	FULL STI Indirect	Comment	Work-arounds/section in text
Assessing suitability of room acoustics for speech communication (no electronic amplification)	++	++	++		
Evaluating PA and VA systems	++	++	++		
Evaluating telecommunication channels (phone, radio)	+	+	--		
Channel features amplitude compression	+	+	--		

Application	FULL STI Direct	STIPA	FULL STI Indirect	Comment	Work-arounds/section in text
Measurements of industrial noise situations with fluctuating noise	+/-	+/-	+/-	Caution required	See 8.9 Measure levels and post process
Speech and noise clearly spatially separated, or a strong direct-field component exists in a highly reverberant environment	+	+/-		To be used with caution. Currently standardised methods are inaccurate.	See 8.11
Channels that do not permit artificial test signals, such as vocoders	+/-	+/-		Currently standardised methods are inaccurate.	Use a speech-based STI test signal or listener tests

5 Direct method of measuring STI – User guidance

5.1 Overview

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

STIPA – consists of only one test signal with a predefined set of two modulations in each of the seven octave bands. The 14 modulations are generated simultaneously. One measurement typically takes between 15 and 25 s.

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

5.2 STIPA

The STI test signal can be simplified if the related modulations in other octave bands that are required for the accurate interpretation of non-linear distortions are omitted [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 5. For each octave frequency band, the modulation transfer function is determined for two modulation frequencies.

The STIPA method, described in Annex B, employs this simplification and has a measurement time of between 15 s and 25 s. The STIPA method is suitable for the measurement of natural speech (room acoustic transmission) as well as sound systems.

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

Without specific corrections, the STIPA method is not a reliable predictor of the intelligibility of speech for hearing-impaired listeners [15]. The measurement of hearing assistive systems or channels is possible, though specific corrections can 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 model described in Clause 4, there are 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 $m(f_m)$ and with stationary noise interference. With STIPA and a measurement time of 15 s, the maximum deviation is approximately 0,03 STI for repeated measurements; see [9],[21],[22],[55].

With fluctuating noise (for example, a babble of voices), higher deviations can be found, possibly with a systematic error (bias). This can be checked by carrying out a measurement in the absence of the test signal, which should result in a residual STI value less than 0,30. An estimate of the deviation should be made by repeating measurements for at least a restricted set of conditions.

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

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

6.1 Overview

The modulation transfer function MTF, as the basis of the STI, can also be computed from the impulse response of a transmission channel, using the process known as the Schroeder method [17]. The impulse response is acquired (usually by computer-based equipment) and the MTF derived from which the STI is subsequently calculated.

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

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

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 can affect the result.

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

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

6.2 Application

When deriving STI values from impulse response measurements, it is usual to make a noise free measurement and then correct this for the effects of background noise and speech level. However, techniques are available that enable the effects of background noise to be directly accounted for within the measurement, for example, using a speech shaped 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) 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 considered by separately measuring the maximum sound pressure level and applying an appropriate correction.
- g) Time variances due to movements of the air (wind) or climatic changes during the measurement process shall be avoided (they also invalidate averaging over longer periods of time). The average wind speed during MLS measurements, for example, should not

exceed 4 m/s. Measurements using maximum length sequences (MLS) are more vulnerable in this respect than measurements performed with sine-sweeps.

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

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

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

6.3 Limitations (non-linear distortion)

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

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

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

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

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

7 Limitations of the STI model

7.1 General

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

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

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

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

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

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

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

Consequently, for certain situations and possible (narrow-band) transmission channels, care shall be taken when using the STI. In some cases, intelligibility can suffer little from a distortion, whilst the STI shows a significant reduction. In other cases, in which the STI shows only minimal changes, the intelligibility can be considerably reduced.

The following clauses discuss specific situations in which the applicability of the STI is limited in more details

7.2 Frequency shifts

This type of distortion can occur with:

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

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

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

7.3 Centre clipping

This type of distortion can occur when low-level parts of a signal are not transmitted faithfully or are silenced. This could happen in amplifiers and corroded connectors. The STI

1002 overestimates the intelligibility for systems that show effects of severe centre clipping, since
1003 the effect on intelligibility is a result of the degraded fine structure of the signal is undetected
1004 by the STI model.

1005 NOTE Centre clipping is also known as crossover distortion and origin distortion.

1006 **7.4 Dropouts**

1007 Signal dropout at regular intervals can result from selective fading patterns in wireless
1008 transmissions and corruption of digital signals. The STI might not be reduced much, but
1009 intelligibility might be very poor. Analysis of the fine structure of the received modulated
1010 signal is recommended to flag dropouts and where possible allow computation of the STI with
1011 the dropouts removed.

1012 **7.5 Jitter**

1013 Time shifts of speech, as applied in digital signal transmission to compensate for variation in
1014 transmission rate, have no effect on intelligibility but can severely reduce the STI, so the STI
1015 can underestimate intelligibility for systems with jitter.

1016 **7.6 Digital voice compression systems**

1017 Digital voice compression systems are often based on models of human speech. STI test
1018 signals, on the other hand, are based on modulated noise, which these systems tend to
1019 suppress rather than reproduce. STI test signals are therefore fundamentally not suited for
1020 digital voice compression systems.

1021 This issue is a limitation inherent to STI measuring model. In addition, voice coders also tend
1022 to affect the fine structure of speech to such a degree that affects intelligibility.

1023 Generally, STI measurements should not be made with channels that include digital voice
1024 compression systems. Exceptions can be made for those cases where it is demonstrated that
1025 the test signal (the fine structure as well as intensity envelope) remains unaffected by the
1026 voice coder. This can be the case with higher bit-rate compression systems

1027 **7.7 Overestimation of STI under low background noise conditions**

1028 It should be noted that the STI model inherently assumes a non-infinite signal to noise ratio in
1029 each octave band, as the hearing reception threshold in the model operates as a source of
1030 background noise. If the background noise levels or the reception threshold values are set to
1031 zero during measurements or simulations, STI values might be too high.

1032 As an example, this issue can arise when investigating the behaviour of STI with changes to
1033 the form of the test signal spectrum. If an MTF matrix having every m value at 1 (i.e. no
1034 degradation from reverberation or background noise) is used with an input signal that deviates
1035 from the specified speech spectrum, the STI result often shows little change, even with large
1036 changes in the input spectrum, see [8].

1037 It is therefore essential that STI predictions and measurements should always incorporate a
1038 level of background noise that is realistic for the application. For example, measurements with
1039 an acoustic output should use a realistic background noise as well as the speech reception
1040 thresholds.

1041 **7.8 Frequency response**

1042 Research so far [8], [9], [10], [11], [12], [13] indicates that the frequency response of the
1043 transmission channel (which is manifest as the perceived tonal balance of speech) is much
1044 more important for perceived intelligibility than is indicated by STI measurements, especially
1045 in the presence of reverberation. If the frequency response is not reasonably flat, it is possible
1046 that the STI can indicate values that are too high compared to the perceived intelligibility.

Systems with measured STIs exceeding 0.5 under conditions of low noise have been reported where the perceived speech intelligibility has been found to be inadequate due to the poor frequency response or tonal balance of the system. The application of equalisation to improve the frequency response substantially improved the perceived intelligibility.

Acknowledging this limitation of the STI-method, a suitable solution for ensuring an even amplitude response is to perform a separate measurement of the amplitude versus frequency response of the system, preferably at a higher resolution than one octave bandwidths. (For example, 1/3 octave bandwidth or 1/3 octave smoothing might be employed). Nonetheless, there are significant factors that might not be included in such measurements.

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

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

7.9 Echoes

Situations have been encountered in which audible echoes (late reflections) cause significant loss of perceived speech intelligibility whilst the corresponding measured STI values are significantly higher than the perceived intelligibility would indicate. The effect is also dependent on the rate of the transmitted speech. This issue is the subject of ongoing research, see e.g. [8], [48]

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

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

NOTE In the example of Figure 2, the frequency response of the MTF extends from 0,5 Hz to below 20 Hz. For STI measurements, frequencies above 12.5 Hz are not taken into account.

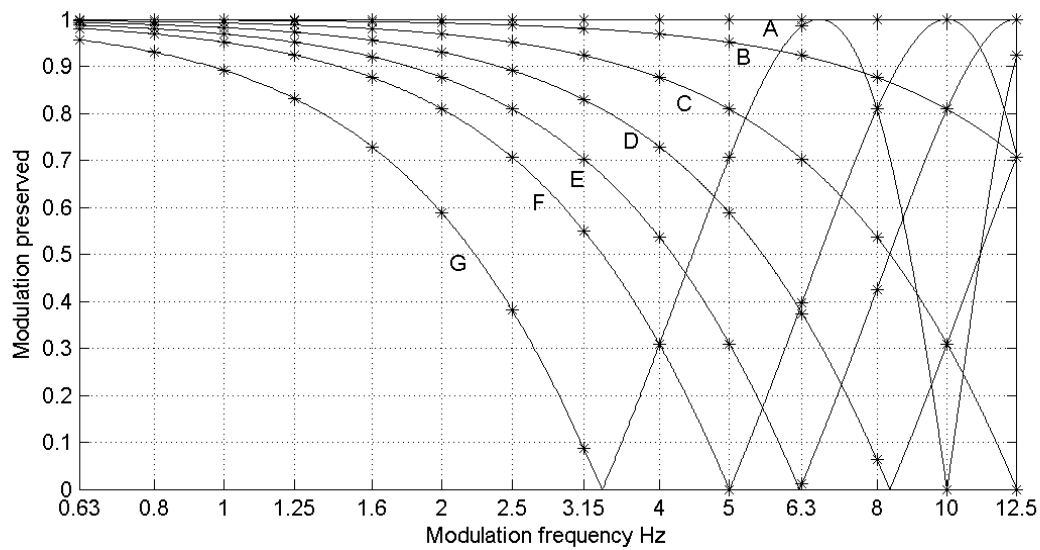
For a simple transmission system that consists only of the direct sound and a single reflection (or secondary arrival), the reflection always produces a reduction in modulation. With very short delay times, up to two or three milliseconds, this reduction appears at high modulation frequencies and is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

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

1095 The frequency at which a notch occurs does not necessarily coincide with any of the third-
1096 octave band frequencies at which the MTF is sampled, so for many delay conditions, the
1097 modulation in a given band is not reduced to zero at the MTF sampling frequency.

1098 Figure 3 shows the effect of secondary arrivals on the modulation transfer function and
1099 resultant MTF values for a range of differential arrival times. The intensities of the first and
1100 second arrival are assumed to be identical, resulting in m -values ranging from 0 to 1. The
1101 graphs have been computed with a continuous frequency input, with the heavy dots in each
1102 graph showing the values that would be recorded in the associated MTF matrix.

1103 Notches in the MTF are clearly evident. For the six conditions from Fig. 1, the STI values
1104 resulting from the respective MTFs are 1, 0,92, 0,77, 0,72, 0,71 and 0,64, respectively.



1105 **Figure 3 – Effect of a single delayed arrival on the MTF (idealised conditions)**

1106 Key

- | | | |
|------|---|--------|
| 1107 | A | 0 ms |
| 1108 | B | 20 ms |
| 1109 | C | 40 ms |
| 1110 | D | 60 ms |
| 1111 | E | 80 ms |
| 1112 | F | 100 ms |
| 1113 | G | 150 ms |

1114

Figure 4 shows the resulting STI values when MTFs become notch-filtered by secondary arrivals with delay times between 0 and 1000 ms and various relative levels between the two arrivals.

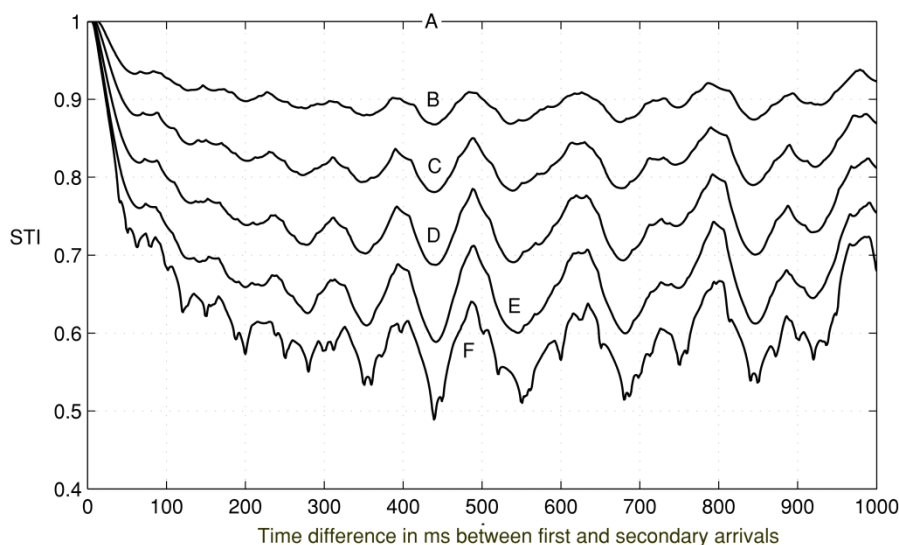


Figure 4 – Idealised STI (Male speech Spectrum) versus delay and level of secondary arrival

Key

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

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

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

7.10 Fast amplitude compression and expansion

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

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

Fast compression reacts on the near-instantaneous amplitude envelopes of a range of frequency bands. With this compression, signal level variations above the compression threshold level (knee point) are reduced according to the compression ratio. As compression reduces the dynamic range of the signal, the modulation depth is also likely to be reduced [51].

On the other hand, automatic gain control (AGC) has slow reaction recovery times and generally does not reduce the short-term dynamic range.

1147 Compression and AGC techniques are often applied to improve speech intelligibility (e.g. for
1148 the hearing impaired who suffer from a limited dynamic range) and can also be applied in
1149 public address systems.

1150 Sentence intelligibility as measured by the speech reception threshold (SRT) has been found
1151 to increase by up to an equivalent of a 4 dB change in effective SNR, but this is dependent on
1152 the amount and type of compression.

1153 The effect of compression on intelligibility at high signal and noise levels, such as in public
1154 address systems, awaits the outcome of further research.

1155 **7.11 Non-linear distortion**

1156 Although the STI is sensitive to distortion, the result is highly dependent on the measurement
1157 method adopted. (This is discussed further in 6.3.)

1158 **7.12 Hearing impaired listeners**

1159 Without specific corrections, the STI model is not a reliable predictor of the intelligibility of
1160 speech for hearing-impaired listeners [15]. The measurement of hearing assistive systems or
1161 channels is possible, though specific corrections might be required [16]. In particular, the
1162 reception thresholds and masking function need to be disabled and the bandwidth considered.

1163 **7.13 Impulsive and fluctuating noise**

1164 Two types of background noise should be distinguished in STI measurements:

- 1165 • impulsive;
- 1166 • fluctuating.

1167 Impulsive noise and undesired short events, such as a hammer dropping, result in inaccurate
1168 STI results, especially with narrow band transmission, as well as in the incorrect diagnosis of
1169 the contribution of frequency bands.

1170 Interpretation of the speech intelligibility in the presence of fluctuating noise is extremely
1171 difficult and cannot be addressed in this edition of the standard. However, it has been found
1172 that listeners listen to speech in the gaps between the fluctuating noise and perceive a higher
1173 intelligibility than the STI would predict, based simply on the L_{eq} of the fluctuating noise.

1174 Fluctuating noise, such as babbling voices or machinery that is repeatedly turned on or off or
1175 is cyclical, can lead to variations in the STI value obtained for repeated measurements and
1176 can also produce considerable underestimation or overestimation of intelligibility
1177 measurements.

1178 Subjectively, the intelligibility of sentences in fluctuating noise is known to be higher than in
1179 stationary noise, with the same time-averaged RMS output [14].

1180 If the impulsive or fluctuating background noise cannot be removed, then an STI
1181 measurement without the test signal should have a value less than 0,3 to ensure that the
1182 temporal variations in the noise will not seriously degrade the STI measurement with the test
1183 signal.

1184 It is preferable that STIPA meters indicate errors resulting from fluctuating and impulsive
1185 noise.

7.14 Conclusion

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

8 Measurement procedures, post-processing of data and applications

8.1 General

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

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

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

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

8.2 Acoustical input

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

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

The following procedure shall be used.

- a) Verify the integrity of the test signal (e.g. via means of a loop back measurement by directly connecting the output of the test signal source to the analyser input). This is particularly important if the test signal is generated from a CD player, although PCM (e.g. .wav file) generators should also be checked. (Digitally compressed signal formats e.g. MP3 should not normally be used, though compression schemes employing at least 128 kbit/s have been shown to work without apparent error). Further information is available in [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 frequency range 88 Hz to 11.6 kHz (the effective limits of the 125 Hz and 8 kHz octave bands) when measured in a free field (free of reflections).
- c) Verify that the individual octave band L_{eq} levels over the range 125 Hz to 8 kHz are within ± 1 dB and preferably ± 0.5 dB of the values for the male spectrum signal given in Table A.4 when using a STIPA or other speech-shaped test signal conforming to the STI spectrum.

1228 NOTE 1 For indirect measurements, the frequency response derived from an MLS or other impulse response
1229 measurement can be processed to calculate an octave-band spectrum.

1230 If necessary, adjust the equalisation (if any) of the artificial mouth or test loudspeaker to
1231 satisfy this requirement.

1232 d) In the absence of an artificial mouth, a suitable transducer such as a small, single-source,
1233 high-quality loudspeaker with cone diameter or aperture not exceeding 65 mm, may be
1234 used and shall be described with the results. The following parameters shall be provided
1235 by the source:

- 1236 • Directionality to match human talker.
- 1237 • The shape of the test spectrum at 50 mm from the source shall not deviate from the
1238 defined STI spectrum shape (see Table A.4) by more than $\pm 2,5$ dB when measured at
1239 the specified reference point of 250 mm or 500 mm (as nominated by the
1240 manufacturer).
- 1241 • The distortion characteristics associated with the system (e.g. driver excursion,
1242 amplifier power capacity, enclosure vibrational modes) shall be sufficiently low that
1243 the m values (in the MTF matrix) are unity when measured under anechoic conditions
1244 at the reference position with the maximum corrected speech level.

1245 Generally, in a listening space, speech intelligibility depends upon the directivity of the
1246 source; therefore, a mouth simulator having similar directivity characteristics to those of
1247 the human head/mouth should be used when assessing the intelligibility of unamplified
1248 talkers or the acoustic pickup from microphones. The directional characteristics of the
1249 acoustic test source can have a significant effect on the results when making STI/STIPA
1250 measurements particularly in reverberant spaces, or when the pick-up microphone is
1251 located at some distance from the talker.

1252 Apertures (cone diameters) not exceeding 65 mm are generally more representative of
1253 the directivity of a human talker. If larger diameter drivers are used to simulate live
1254 talkers, the high-frequency directivity might be too high for accurate STI measurements
1255 especially when using distant microphones. For further information, see
1256 [16],[23],[50],[51],[56].

1257 When the system (source) microphone is situated in either a reverberant or noisy location
1258 or if a close-talking or noise-cancelling microphone is involved, a mouth/talker simulator
1259 should be employed as the test-signal source. Under low noise/low reverberation
1260 conditions, direct injection of the test signal may be suitable.

1261 e) Set the acoustic test source on the axis of the system microphone at the normal talker
1262 position and distance.

1263 f) Set the test signal level at the microphone position to the corrected speech level that is
1264 used in the system. The speech and test signal levels shall be matched according to the
1265 method described in Annex J.

1266 If the corrected speech level is unknown, a default equivalent level of 60 dBA at 1 m in
1267 front of the artificial mouth or test loudspeaker should be used.

1268 Smaller talker distances typically result in speech levels of approximately 86 dBA to
1269 94 dBA for handheld microphones (distances of 5 cm to 2 cm), while speech levels of
1270 approximately 80 dBA to 86 dBA result for gooseneck microphones (distances of 10 cm to
1271 5 cm).

1272 NOTE 3 This test can stress the amplifier driving the source. See 14.9 of IEC 60268-3[24]. It can be convenient to
1273 apply the test signal for 1 min, for example, followed by several minutes of zero signal to allow cooling to take
1274 place.

1275 NOTE 4 The above levels are subject to wide variations in practice.

- 1276 g) Run the STI or STIPA test sequence. Normally, and where available, the “with noise”
1277 option should be selected.
- 1278 h) The sound field of the test signal should be allowed to develop and stabilise in the space
1279 for a minimum of 2 secs before commencing a measurement. In highly reverberant spaces,
1280 e.g. road tunnels, this minimum period may need to be extended to between 5 and 10
1281 secs. An insufficient stabilisation period can lead to over-estimation of the STI.
- 1282 i) If an MLS signal is used to measure the impulse response and if it is required to take
1283 account of the background noise, the excitation spectrum should be adjusted to the
1284 standardised speech spectrum by appropriate filtering. Signal averaging should be
1285 disabled or a single sequence should be employed [25].
- 1286 j) If sine-sweeps, MLS or TDS are used to determine the noise-free impulse response,
1287 appropriate adjustments to speech and noise levels at both the microphone and receiver
1288 locations shall be applied to the noise-free MTF by post processing.
- 1289 k) The test signal shall be fed into the system in such a way as to ensure that all signal
1290 processing components relevant for speech reproduction (equalizers, signal delays, etc.)
1291 are correctly considered during the measurement process.

1292 **8.3 Acoustical output**

1293 The STI model is based on the use of a single omni-directional measurement microphone
1294 which shall be acoustically calibrated with respect to sensitivity and frequency response.
1295 Measurements shall be performed at the listener’s normal location and listening height (or at
1296 a specified listening height). Alternatively, a measurement can be made with an artificial
1297 binaural ear/head simulator with appropriate adjustments as described in 8.11.

1298 **8.4 Electrical input**

1299 Follow the above procedure in 8.2, replacing step d) by the step below, and selecting the
1300 injection point for the signal to be as close as possible to the normal signal input, to include
1301 as much of the system as possible in the test.

1302 The STI test signal at the point of injection shall be adjusted to be equivalent to the level of
1303 speech at that point. The speech level is determined using the speech level measurement
1304 method as described in Annex J.

1305 **8.5 Electrical output**

1306 Since no acoustic conditions are involved at the electrical output, hearing-related effects,
1307 such as masking and the reception threshold, shall be disabled on the measurement device. If
1308 this is not possible, the electrical input to the measurement device shall be adjusted to
1309 simulate a sound pressure level well above the reception threshold but below a point where
1310 level-dependent masking becomes noticeable in the STI results (between approximately 55
1311 and 85 dBA). Broad band output levels should be A-weighted and then reported as A
1312 weighted voltage levels in dB relative to a stated reference, e.g. 1 volt.

1313 **8.6 Examples of input/output combinations**

1314 **8.6.1 Acoustical input – Acoustical output**

1315 In the normal STI measurement set-up for PA systems and in auditoria, a sound source is
1316 used to acoustically generate the STI test signal. The test signal level is calibrated and
1317 corresponds to the nominal speech level. A situation-dependent and representative talking
1318 distance should be employed as described in 7.2. A calibrated STI measuring device is used
1319 at the receiver location to determine the STI of the transmission channel.

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

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

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

To compare microphones with respect to their effect on the intelligibility, the STI test signal level at the microphone should be calibrated as given in 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.

STI may be used to measure the potential intelligibility performance of Assistive Listening Systems (ALS) and in particular, their associated acoustic paths. In most ALS, (e.g. Hearing Loop Systems (HLS), also known as Audio Frequency Induction Loop Systems (AFILS), or Infra Red systems) it is the path between the pick-up microphone and transmission system that is of critical importance. In so called 'soundfield' or voice reinforcement systems, the path between the loudspeaker and listener is more relevant. Further information can be found in IEC 60118-4. [53].

Special methods can be required when measuring the STI of assistive hearing systems and Hearing Loop Systems, e.g.[16]. In particular, the reception thresholds and masking function need to be disabled. However, much of the general guidance given in Clause 7 is applicable.

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

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

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

When assessing headsets, an in-ear microphone or an artificial ear (see IEC 60318) should be used.

8.7 Spatial averaging of STI measurements

In some applications, e.g. evacuation using speech messages, applicable standards give information about how STI measurements are to be executed in detail and how a combined result is to be calculated from multiple measurements representing an average over an area. Often, such standards require a space to be subdivided into areas that are characterised by a common scenario e.g. reverberation time and/or background noise. Such areas are typically termed 'Acoustically Distinguishable Area' or ADA. Most applicable standards also specify the

1368 number of measurements to be taken or alternatively, they specify a typical grid on which
1369 measurements shall be performed (see for example ISO 7240-19).

1370 For the calculation of a single-number result, averaging of individual measurements and some
1371 statistical post-processing is required. The applicable standards specify this process but two
1372 commonly used methods are described below.

- 1373 • Perform STI measurements at each specified location within an ADA according to the
1374 applicable standard. To reduce statistical fluctuations, apply averaging at each individual
1375 location.
- 1376 • One method then takes all positional results and from that dataset, calculates the
1377 arithmetic mean value as well as the standard deviation and then subtracts the standard
1378 deviation from the mean value to produce the result.
- 1379 • Other methods require more complex post-processing of the individual result, e.g.
1380 generating a cumulative distribution function, discarding a certain percentage of samples
1381 and post-processing the remaining values for mean and minimum values.

1382 Some basic rules should be observed when selecting analysis positions across a space or
1383 building:

- 1384 • It is generally required to perform separate measurements for each room.
- 1385 • Performing measurements of example rooms is permitted for time-saving purposes,
1386 provided that they are identical in terms of their room acoustics (including furniture and
1387 equipment), dimensions, sound system implementation and ambient noise.
- 1388 • The selection of the measurement positions shall be representative. Positions immediately
1389 next to loudspeakers should not be included to avoid exceedingly good results. Grid size
1390 should be selected based on the application and the applicable standards, but should
1391 include as a minimum, two positions per room. For further information, refer to ISO 7240-
1392 19.
- 1393 • No more than one-third of the measurements should be made on the axis of the
1394 loudspeakers.
- 1395 • Symmetries in terms of room geometry, surface material and sound system design can be
1396 utilized for reducing the number of necessary measurement positions.
- 1397 • The selection shall consider changing spatial conditions within a room (e.g. variation of
1398 the ceiling height, the acoustical absorption of the surface or shadowing effects).
- 1399 • The microphone position during the measurements should be at assumed ear height. The
1400 height to be assumed is 1 m to 1,2 m for seated persons and 1,5 m to 1,7 m for standing
1401 persons.

1402 **8.8 Post-processing of measured MTF data**

1403 There are a number of corrections that can be made to measured MTF data:

- 1404 • elimination of noise (de-noising) from a measured MTF;
- 1405 • addition of an ambient noise level and spectrum;
- 1406 • consideration of the hearing reception threshold;
- 1407 • adjustment of the speech level and spectrum;
- 1408 • correction for different reverberation times.

1409 The effect of occupancy noise can be determined either:

- 1410 a) by manually entering noise data into the noise data table used by the measuring
1411 equipment or
- 1412 b) by mixing an artificial or recorded noise signal of the correct spectral content and level
1413 with either the direct signal input to the analyser or a recorded signal.

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

8.9 Issues concerning noise

8.9.1 General

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

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

An essentially noiseless situation in which only the auditory hearing threshold acts as a residual noise source, is usually not a realistic assumption for most practical cases. Even in quiet environments, such as libraries or court rooms, a residual noise level of 25 dB to 35 dB SPL is not uncommon and should be considered. This can be achieved by applying a suitable criterion, such as NCB, RC or NR curves (see [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 and impulsive noise is detected by measuring the direct STI in the absence of the test signal. If the STI is too high (e.g. $STI > 0,3$), the measurement results are likely to be erroneous. Preferably, the STI measurement should be carried out without the noise being present. The noise should then be separately measured (see 8.9.2), and the STI computed mathematically.

8.9.2 Measurement of background noise

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

It should be noted that for the corrective calculation described here, it is not sufficient to determine a single broadband value for the background noise (e.g. $L_{A,eq}$) and to use a single A-weighted broadband sound pressure level value for the speech signal in dB. Also see 8.9.3.

8.9.3 Fluctuating noise

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

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

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

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

8.10 Analysis and interpretation of the results

It is important to examine the MTF data in each octave band to determine the reliability of the results. Examples of items that should be examined are as follows:

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

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

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

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

8.11 Binaural STI measurements

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

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

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

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

9 Use of STI as a design prediction tool

9.1 Overview

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

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

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

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

9.2 Statistical predictions

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

9.3 Prediction from simulated impulse response

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

- a) the MTF matrix shall be calculated using the Schroeder equation (see 6.1). The evaluated part of the impulse response shall not be less than half the reverberation time and at least 1,6 s to ensure a reliable calculation of the modulation indices for the lowest modulation frequency of 0,63 Hz;
- b) both the hearing reception thresholds and the measured or estimated background noise sound pressure levels for each octave band shall then be introduced into the MTF matrix;

- 1544 c) the speech spectrum and real speech level shall be selected and the auditory masking
1545 corrections listed in Table A.1 applied to the MTF matrix;
- 1546 d) the octave band specific male weighting factors given in Table A.3 shall be applied to the
1547 MTI values;
- 1548 For each prediction location, access shall be available to the MTI values in each octave band
1549 and the octave band levels of the output speech signal along with the frequency response.
- 1550 For predictions with multiple listener positions, the statistical properties and distribution of the
1551 results over the listening area shall be stated.
- 1552 The results shall also state:
- 1553 • that the STI has been calculated from an MTF derived from a predicted impulse response
1554 with the appropriate male or female weighting applied;
 - 1555 • the background noise levels which have been applied to the prediction.

Annex A (informative) The basis of the STI concept

A.1 Introduction to this Annex

A.1.1 Purpose

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

A.1.2 Modulation transfer function (MTF)

The rationale underlying the application of the MTF concept to studies of room acoustics has been described elsewhere [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.1 illustrates this for the octave-band centred on 250 Hz for two simple transmission systems, one with exponential reverberation only (case A: $T = 2,5$ s) and the other with only interfering noise (case B; signal-to-noise ratio SNR = 0 dB). (The vertical lines in the envelope spectrum indicate the reduction in modulation index at each modulation frequency)

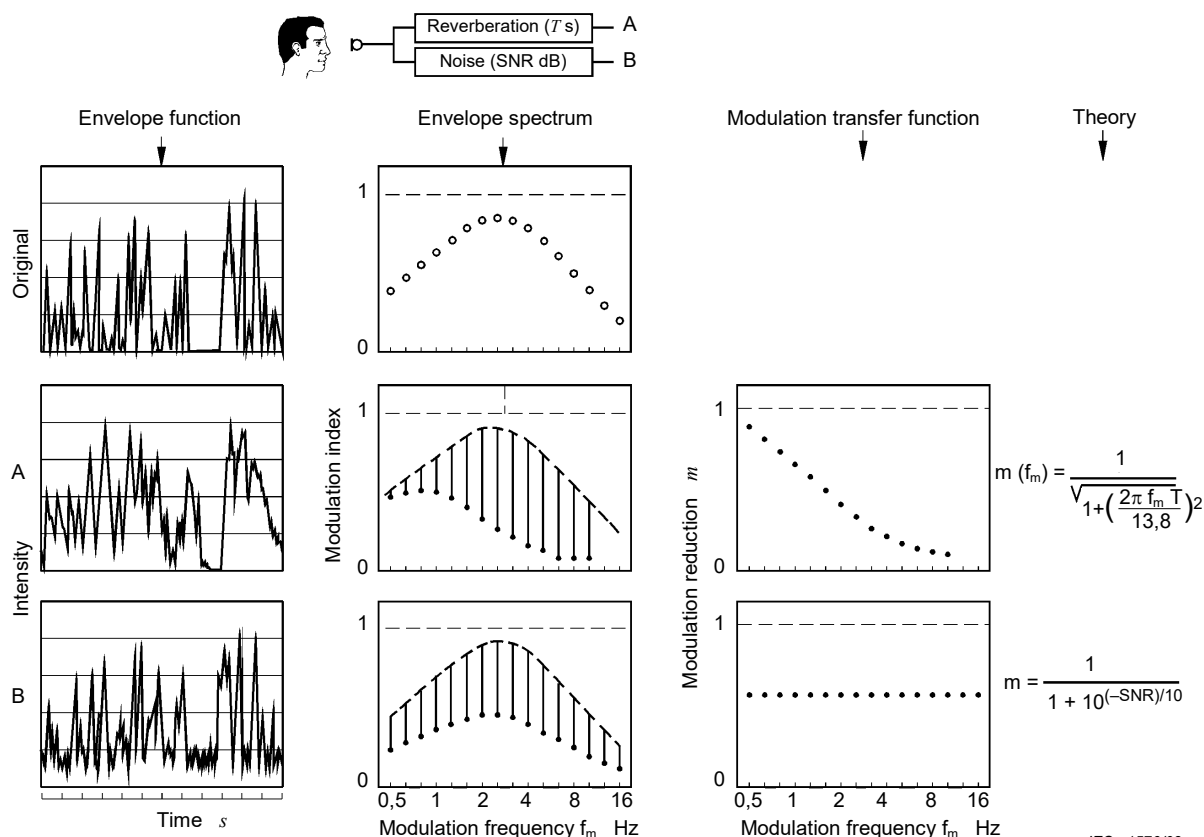


Figure A.1 – Theoretical expression of the MTF

With reverberation, the MTF frequency response has the shape of a low-pass filter: the faster fluctuations being relatively more affected than slower fluctuations. In the theoretical case of a

diffuse sound field with a purely exponential reverberant decay, the MTF can be derived mathematically (see Figure A.1, case A) and the product of f_m and T determines the roll-off as given by:

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

where

f_m is the modulation frequency;

T is the reverberation time in seconds.

For noise, the MTF is defined by the signal-to-noise ratio and is independent of the modulation frequency (see Figure A.1, case B). The noise, by increasing the mean intensity, reduces the modulation depth for all modulation frequencies as given by:

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

where

SNR is the signal to noise ratio in dB.

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

A.1.3 STI model

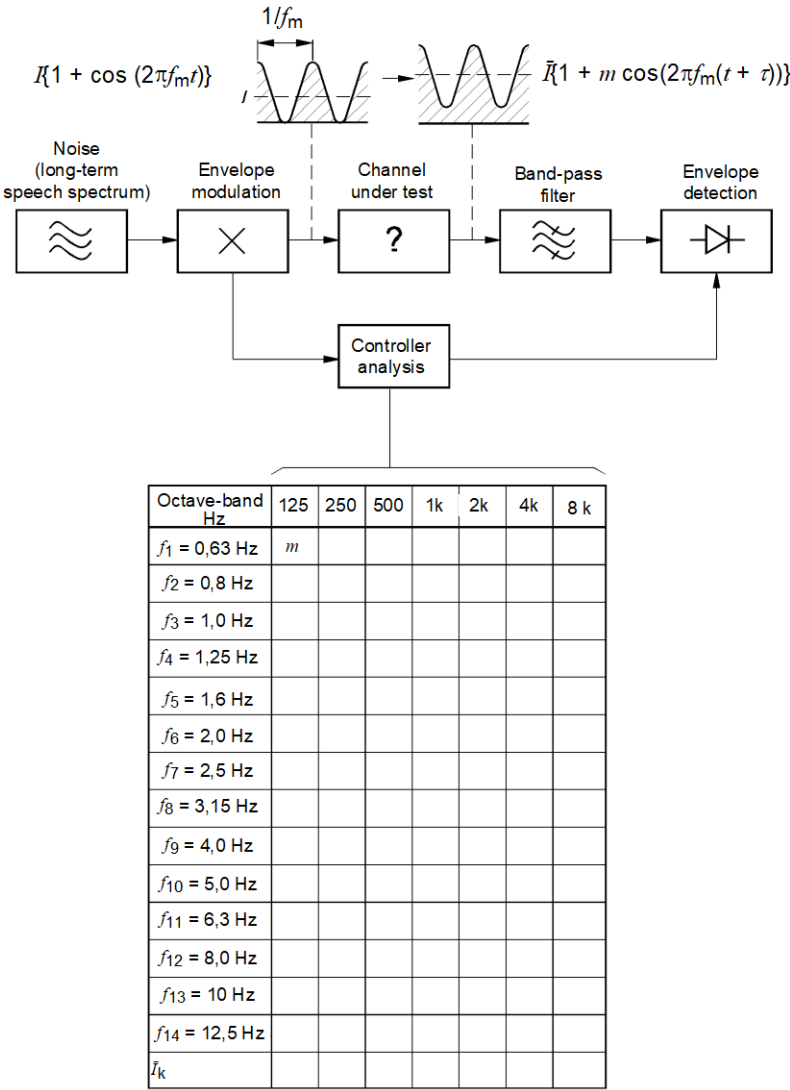
The STI model uses a discretized version of the MTF, with the modulation transfer ratios (also known as m values) being computed at specific modulation frequencies.

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

The STI model was originally developed using a test signal that applied each modulation frequency sequentially to each octave band. Random modulations were used for all other the octave bands not under test. The random modulations were based on energy distributions as found in natural speech and had an instantaneous level which was approximately 3 dB higher than the overall speech level for the specific octave band under evaluation. This model is referred to as the FULL STI method.

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

The concept is shown diagrammatically in Figure A.2.



NOTE The value of the modulation transfer function (m) is determined for all cells of the matrix of 7 octave carrier bands and 14 modulation frequencies. Also, the octave intensity levels (\bar{I}_k , equal to the square of the sound pressure levels) are obtained for use in calculating auditory masking effects.

Figure A.2 – Measurement system and frequencies for the STI method

A.1.4 STI modulation frequencies

The 14 STI modulation frequencies are at nominal one-third octave intervals and consist of 0,63; 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.2 Calculation of the STI

A.2.1 General Equation for STI

The STI is calculated using:

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

where

1626 MTI_k is the modulation transfer index for octave band k ;

1627 α_k is the gender-specific weight factor for octave band k ;

1628 β_k is the gender-specific redundancy factor between octave band k and octave band $k+1$.

1629 In the event that STI values higher than 1,0 are obtained, they should be set at 1,0.

1630 The modulation transfer index (MTI_k) per octave band k is obtained by averaging the
1631 transmission indices (TI) over the modulation frequencies:

1632
$$MTI_k = \frac{1}{n} \sum_{m=1}^n TI_{k,f_m}$$

1633 where

1634 TI_{k,f_m} is the transmission index for each octave band k and modulation frequency f_m ;

1635 m is the index of the modulation frequency.

1636 n is the number of modulation frequencies per octave band.

1637 The transmission index (TI) for each octave band and modulation frequency is calculated
1638 using:

1639
$$TI_{k,f_m} = \frac{SNR_{\text{eff } k, f_m} + 15}{30}$$

1640 where

1641 $SNR_{\text{eff } k, f_m}$ is the effective signal to noise ratio for each octave band k and modulation
1642 frequency f_m expressed in dB.

1643 It is calculated using the corrected modulation transfer ratio values:

1644
$$SNR_{\text{eff } k, f_m} = 10 \times \log \frac{m'_{k, f_m}}{1 - m'_{k, f_m}}$$

1645 where

1646 m'_{k, f_m} is the corrected modulation transfer ratio value for octave band k and modulation
1647 frequency f_m (commonly called m values).

1648 Since the outcome of the signal-to-noise ratio calculation can become infinite, values shall be
1649 limited to the range of -15 dB to +15 dB.

1650 **A.2.2 Gender-specific octave band weighting and redundancy factors**

1651 The STI method can discriminate between male and female speech signals. However, in
1652 practice and to simplify the prediction and measurement process, only male speech shall be
1653 used.

1654 The STI weighting factors (α) and redundancy factors (β) for male speech are shown in Table
1655 A.1 as a function of the octave bands.

Table A.1 – MTI octave band weighting factors

Octave band Hz		125	250	500	1 000	2 000	4 000	8 000
Males	α	0,085	0,127	0,230	0,233	0,309	0,224	0,173
	β	0,085	0,078	0,065	0,011	0,047	0,095	–

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

A.2.3 Adjustment of the MTF for ambient noise

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

where

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

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

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

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

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

NOTE $I_k = I_{s,k} + I_{n,k}$, equals the total acoustic intensity level, which is used in A2.4 below.

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

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

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

where

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

I_k is the total acoustic intensity level for octave band k and includes the received test signal level and noise, ($I_k = I_{s,k} + I_{n,k}$)

1679 $I_{am,k}$ is the total acoustic intensity level for the level dependent auditory masking effect on
 1680 octave band k , as described in Section A.4.2. NB this is the total level of signal and noise.

1681 $I_{rt,k}$ is the acoustic intensity level of the reception threshold for octave band k , as
 1682 described in Section A.4.3

1683 The injected test signal shall have the specified speech spectrum shape.

1684 Modulation transfer ratio values higher than 1,0 shall be truncated to 1,0. An m -value higher
 1685 than 1,3 is very unlikely and is likely to be a result non-sinusoidal fluctuations or impulsive
 1686 noises.

1687 NOTE The auditory masking intensity ($I_{am,k}$) is determined by logarithmically summing the level of the ambient
 1688 noise and speech signal.

1689 **A.3 Calculation of the modulation transfer ratio values**

1690 **A.3.1 Direct method: Analysis of the STI test signal**

1691 **A.3.1.1 Basic processing steps**

1692 The following texts set out the procedure for analysing the received signal and calculating the
 1693 resultant STI. The procedure is broken down into the following basic steps:

- 1694 • filter the input signal with the seven octave band filters;
- 1695 • determine the intensities I_k in each octave band k ;
- 1696 • determine the intensity modulation I at each modulation frequency f_m ;
- 1697 • determine the m_k values in each octave band k ;
- 1698 • calculate the STI according to clause 4.3 and A.2.

1699 **A.3.1.2 Filtering and determination of intensities**

1700 The received modulated-noise test signal shall be band-pass filtered into seven octave-wide
 1701 signals. The filters shall be one octave band wide with centre frequencies ranging from
 1702 125 Hz up to 8000 Hz.

1703 The shape and tolerance limits of the band-pass filters shall comply with IEC 61260-1 class 1
 1704 as stated in IEC 61260-1. The input signal shall be split by the filter bank into output signals
 1705 without power loss. Additionally, to minimize cross-talk between adjacent octave bands, filter
 1706 slopes shall comply with the requirements of Annex C and provide 42 dB minimum attenuation
 1707 at the centre frequency of each adjacent band. Filters can be implemented as IIR or FIR types.

1708 The phase characteristics of the band-pass filters should be as linear as possible to avoid
 1709 distortion of the phase relationship of the amplitude modulations by the settling behaviour of
 1710 the filters. The phase characteristics of the filters shall not give rise to a systematic error
 1711 higher than 0,01 STI for the range between 0,1 and 0,9 STI (between -12 dB and 12 dB SNR).

1712 The intensity envelope shall be calculated by squaring the outputs of the bandpass filters and
 1713 applying a low pass filter at a cut-off frequency of approximately 100 Hz to the intensity signal.

1714 During one measurement, all intensities shall be calculated using the same time segment of
 1715 the input signal; no implicit time weighting is allowed.

1716 **A.3.1.3 Derivation of the modulation transfer function (MTF) using the direct method**

1717 The calculation of the STI is based on the modulation transfer function. The MTF of a
 1718 transmission path can be determined in various ways, the principal being the derivation of the

1719 modulation reduction factor from the comparison of the intensity modulations at the output
1720 and input of the path.

1721 Depending on the test signal, the MTF for each octave band is derived by correlating the
1722 intensity envelope with sine and cosine signals with specific duration and modulation
1723 frequencies. The modulation transfer at each modulation frequency f_m is calculated by first
1724 deriving the modulation depth of the received signal (m_o) for the output of each octave band k
1725 (see Fig. 2).

$$1726 \quad m_o(k, f_m) = 2 \times \frac{\sqrt{\left[\sum I_k(t) \cdot \sin(2\pi f_m t)\right]^2 + \left[\sum I_k(t) \cdot \cos(2\pi f_m t)\right]^2}}{\sum I_k(t)}$$

1727 where

1728 f_m is the modulation frequency in Hz;

1729 t is the time in seconds;

1730 $I_k(t)$ is the intensity envelope as a function of time for octave band k ;

1731 The summation shall be made over the measurement duration using a whole number of
1732 periods for each specific modulation frequency.

1733 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
1734 sinusoidal modulation f_m .

1735 NOTE 2 The use of window functions with non-integer periods leads to inaccuracies.

1736 Using the modulation indices of the received signal and the transmitted signal, the modulation
1737 transfer ratio can be calculated. This value is often referred to as the m -value. All derived
1738 modulation transfer ratio values $m(k, f_m)$ form the so-called MTF matrix. The modulation
1739 transfer ratio is given by:

$$1740 \quad m_{k, f_m} = m_o(k, f_m) / m_i(k, f_m)$$

1741 where

1742 $m_o(k, f_m)$ is the modulation depth of the received test signal for octave band k and
1743 modulation frequency f_m ;

1744 $m_i(k, f_m)$ is the modulation depth of the transmitted test signal for octave band k and
1745 modulation frequency f_m .

1746 NOTE The derivation of the modulation transfer function (MTF) using the indirect method is described in Clause 6.

1747 **A.3.2 Indirect method: Determination of the modulation transfer function (MTF)**

1748 **A.3.2.1 Basic processing steps**

1749 The following sections set out the procedure for analysing the received signal and calculating
1750 the resultant STI using the indirect method. The procedure is broken down into the following
1751 primary steps:

- 1752 • filter the impulse response with the seven octave band filters;
- 1753 • determine the m_k values in each octave band k using equation in Clause A.3.1.3;
- 1754 • calculate the STI according to clause 4.3 and A.2.

A.3.2.2 Filtering and determination of intensities

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

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

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

Filters can be implemented as IIR or FIR types.

A.4 Auditory effects on the STI

A.4.1 Overview

The STI model models two specific hearing-related aspects by applying appropriate noise terms. These two aspects are upward auditory masking (the reduction in aural sensitivity by a stronger, lower frequency sound) [30] and the absolute reception threshold. These auditory effects reduce the effective signal-to-noise ratio in the various octave bands and can be expressed as a reduction of the modulation transfer function, which result in lower STI values.

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

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

A.4.2 Level-dependent auditory masking

Auditory masking is an inherent effect of the human hearing process. When a loud, low frequency sound is presented at the ear, it always masks higher frequencies, possibly rendering them inaudible if the difference between their relative levels exceeds a given threshold. This phenomenon is referred to as upward spread of masking [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 total sound pressure level present in octave band ($k-1$). Note that downward masking is not included in the model.

The auditory masking as modelled in the STI influences the 250 Hz, 500 Hz, 1000 Hz, 2000 Hz, 4000 Hz and 8000 Hz octave bands and extends one octave band upwards.

1800 Accordingly, the 125 Hz octave band masks the 250 Hz octave band, the 250 Hz octave band
1801 masks the 500 Hz octave band, etc. The 125 Hz octave band is not masked at all.

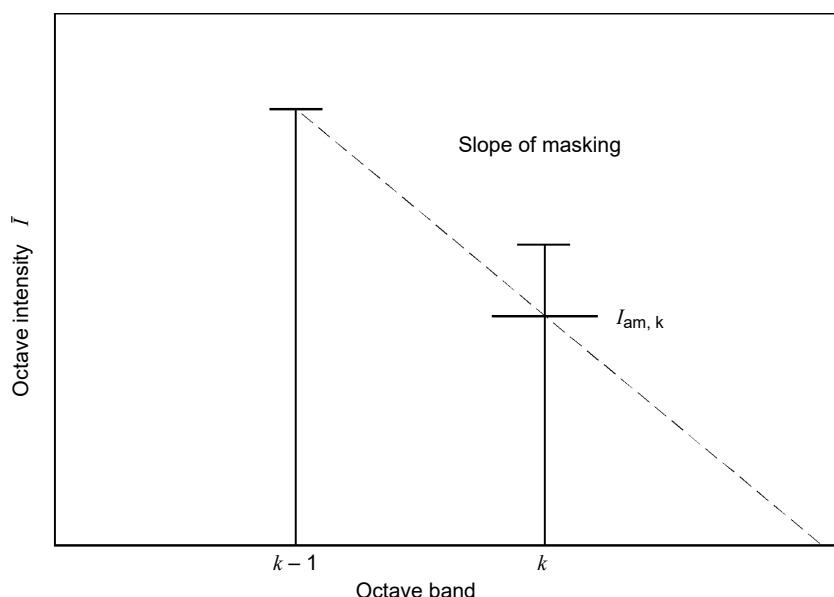


Figure A.4 – Auditory masking of octave band ($k-1$) on octave band (k)

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

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

where

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

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

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

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

where

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

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

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

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

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

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

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

where

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

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

NOTE 2 The masking scheme introduced with Edition 3 was discrete and resulted in non-continuous STI results as a function of the overall sound pressure level. Since Edition 4, the auditory masking scheme presented in Table A.1 yields continuous STI results as a function of the sound pressure level.

The effect of level dependent masking is shown in Figure A.5. This figure shows the STI values for a range of A-weighted speech levels from 0 dB to 120 dB with an MTF matrix produced by seven different reverberation times with equal reverberation times in each octave band. The effect of the reception threshold is also included in this graph.

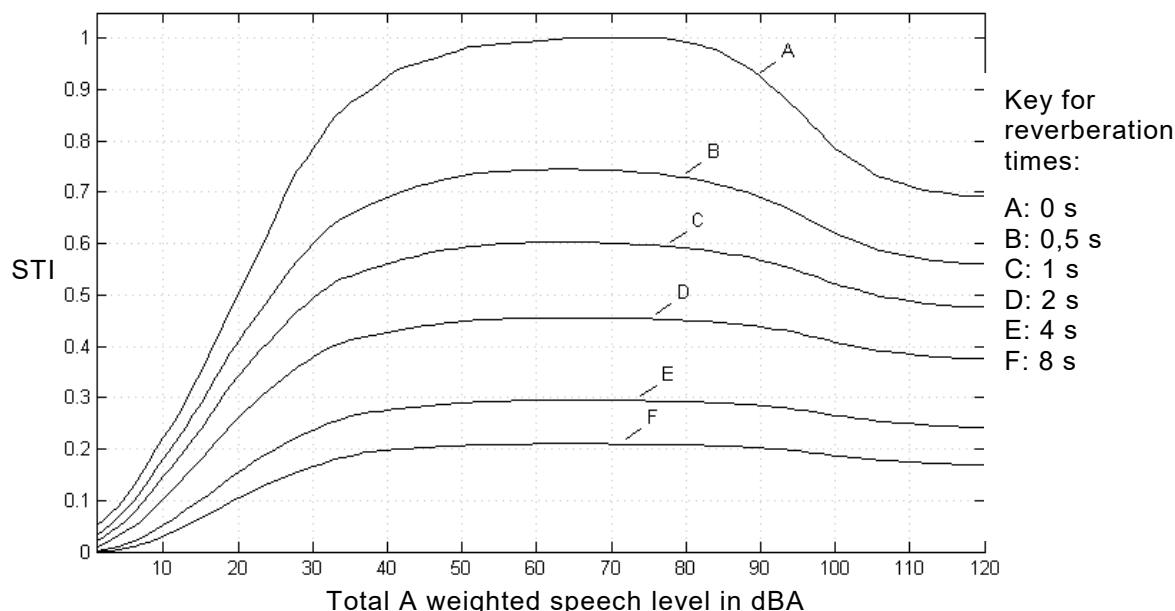


Figure A.5 – Relationship between STI and speech level for different reverberation times.

A.4.3 Absolute speech reception threshold

The absolute speech reception threshold is defined by the absolute threshold of hearing and the minimal required dynamic range for the correct recognition of speech. The absolute speech reception threshold intensity (I_{rt}) is modelled in the STI as an intrinsic noise floor which reduces the effective signal to noise ratio when speech levels are low. The 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 centre frequency Hz	125	250	500	1 000	2 000	4 000	8 000
Absolute speech reception threshold ART dB SPL	46	27	12	6,5	7,5	8	12

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

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

where

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

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

A.5.1 Pink noise source signal

The direct STI method uses test signals that have similar spectral and temporal properties to those found in natural speech. 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 percentile exceedance (1 %) value typically lying between 8 dB and 11 dB.

A.5.2 Generating octave band carrier signals

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

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

A.5.3 Intensity modulation of the carrier signals

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

$$A_{f_m}(t) = \sqrt{0.5 \cdot (1 + m \cdot \cos(2\pi f_m t))}$$

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).

1881 If more than one modulation frequency per carrier band is impressed on a band at the same
 1882 time, the modulation factors shall be equal and the phase relationships and the modulation
 1883 depth shall be selected to prevent over-modulation of the carrier.

1884 For the generation of the STIPA test signal, see Annex B.

1885 **A.5.4 Applying the speech spectrum to the STI test signal**

1886 Each carrier signal shall be assigned a relative octave band level according to the speech
 1887 spectrum given in Table A.4. Since the carrier signals originate from a pink noise signal, the
 1888 levels need only be derived according to Table A.4. All modulated carrier signals shall be
 1889 summed to obtain the STI test signal.

1890 **A.6 Spectrum of STI test signal**

1891 The spectrum of the STI test signal is specified by the mean levels in each octave band given
 1892 in Table A.4. The octave band levels are normalized to give an A-weighted level of 0 dB for
 1893 easy scaling to an overall A-weighted sound pressure level. This spectrum could be
 1894 instantaneously exceeded by 3 dB with a modulated test signal.

1895 **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	- 2,4	0,6	0,1	- 5,9	- 11,9	- 17,9	- 23,9	0,0
NOTE 1 For guidance in determining the speech level, see Annex J.4.								

1896
 1897 NOTE The male spectrum has changed in this edition of the standard. Prior to making this change, a theoretical
 1898 investigation was conducted into the extent of changes to STI values with the above spectrum compared to the
 1899 previous spectrum. Approximately 1.5 million scenarios were investigated; refer to [49].

1900 **A.6.1 Speech shaped noise**

1901 To shape a noise signal to the speech spectrum specified in Table A.4, a combination of IIR
 1902 (infinite impulse response) filters with gain can be applied to a pink noise signal. The required
 1903 spectrum can be produced using a combination of high and low pass filters, a biquadratic
 1904 peaking filter and gain settings. An example of filter parameters that produce that spectrum is
 1905 given in Table A.5, and includes two types of common biquadratic peaking filters which have
 1906 different s-plane polynomials.

1907 The spectrum produced by this combination is based on the test signal being integrated into
 1908 octave bands and is not directly equivalent to the frequency response of the filter combination.

1909 The levels in the 31,5 Hz, 63 Hz and 16 kHz octave bands shall be at least 20 dB lower than
 1910 the 125 Hz and 8 kHz bands respectively.

1911

1912 **Table A.5 – Filter parameters and s-plane polynomials that produce speech-shaped pink**
1913 **noise.**

Filter	Type	Parameter	s-plane polynomial
Filter 1:	Second order high-pass filter	resonant frequency $f_h = 142$ Hz $Q = 0.58$	$F(s) = \frac{s_n^2}{s_n^2 + \frac{s_n}{Q} + 1}$ where $s_n = \frac{s}{2\pi f_h}$
Filter 2 Alternative A	Biquadratic peaking filter	Centre frequency $f_c = 500$ Hz $Q = 1.78$ Gain $g = 2.7$ dB	$G(s) = \frac{s^2 + Bs + \omega_0^2}{s^2 + As + \omega_0^2}$ where $\omega_0 = 2\pi f_c$ $G = 10^{\left(\frac{g}{20}\right)}$ $dW = \omega_0 \left(2^{\frac{W}{2}} - 2^{-\frac{W}{2}}\right)$ where $W = 2 \sinh^{-1} \left(\frac{1}{2Q}\right)$ $\frac{1}{\log_e(2)}$ $A = dW \sqrt{\frac{1}{G}}$ $B = GA$
Filter 2 Alternative B	Second order parametric filter	Centre frequency $f_c = 500$ Hz $Q = 2.04$ Gain $g = 2.7$ dB	$G(s) = \frac{s_n^2 + Bs_n + 1}{s_n^2 + As_n + 1}$ where $s_n = \frac{s}{2\pi f_c}$ $G = 10^{\left(\frac{g}{20}\right)}$ $B = \frac{G}{f_c Q}$ where $g \geq 0$ $B = \frac{1}{f_c Q}$ where $g < 0$ $A = \frac{1}{f_c Q G}$ if $g < 0$ $B = \frac{1}{f_c Q}$ if $g \geq 0$
Filter 3:	First order low-pass filter	Turnover frequency $f_l = 315$ Hz	$H(s) = \frac{1}{s_n + 1}$ where $s_n = \frac{s}{2\pi f_l}$
Gain		4,0 dB	

Annex B (normative) STIPA method

B.1 Overview

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

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

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

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

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

Table B.1 – Modulation frequencies for the STIPA method

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

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

B.2 Test signal

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

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

B1

where

- G_k is the octave band weighting factor
- $N_k(t)$ is the bandwidth-limited noise-carrier signal
- $A_k(t)$ is the amplitude modulator
- k is the octave band number

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

1957 The octave band weighting factor G_k is given by:

1958

$$1959 \quad G_k = 10^{\frac{dB_k}{20}} \quad (B2)$$

1960 where

1961 dB_k is the level in dB in the octave band k

1962 The octave band levels are based on the male spectrum from Table A.4.

1963 The modulator $A_k(t)$ for each octave band is described by:

1964

$$1965 \quad A_k(t) = \sqrt{0.5 \left(1 + 0.55 (\sin(2\pi f_{1k} t) - \sin(2\pi f_{2k} t)) \right)} \quad (B3)$$

1966 where

1967 f_{1k} is the first modulation frequency in Hz in the k th octave band

1968 f_{2k} is the second modulation frequency in Hz in the k th octave band

1969

1970 NOTE With STIPA, both sinusoidal oscillations are added in opposite phase so that the modulation signal's crest
1971 factor is minimized. This allows for the maximum modulation factor of 0.55.

Annex C (normative)

Verification of STI measuring devices

C.1 Specification of the measuring device

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

Table C 1 Specification of an STI measuring device

Characteristic	Sub-clause reference	Requirements
Modulation depth for STIPA direct method	C.3.1	The absolute value of the error between the derived and the theoretical m -value shall not exceed 0,05. Overall m -value errors shall not yield a systematic absolute error (offset) in the STI results greater than 0,01. m -values shall be reported for the highest and the lowest possible measuring levels of the measuring device meeting the criteria.
Modulation depth for STIPA indirect method	C.3.2	
Crosstalk between octave-band filters	C.4	Crosstalk shall not increase the effective noise level representing the auditory masking by more than 3 dB. If, in testing, an m -value lower than $0,5 \pm 0,05$ is obtained, the level of the observed modulated carrier shall be increased in 1 dB steps to a level where an m -value of $0,5 \pm 0,05$ results. With the measured relative level, the corresponding sound pressure level as per Table A.1 is determined and reported as the sound pressure level below which the STI is underestimated.

C.2 Signals for testing STI implementations

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

C.3 Testing the dynamic range in the modulation domain

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

C.3.1 Modulation depth testing for STIPA direct method

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

Equation C1 yields a modulated carrier $A(t)$ that allows the testing of the dynamic range of STIPA implementations in the modulation domain for different octave bands.

$$A(t) = \sin(2\pi f_{c_k} t) \sqrt{0.5 (1 + 0.55 m (\sin(2\pi f_{1_k} t) - \sin(2\pi f_{2_k} t)))} \quad (C1)$$

where

- k is the octave band index $k=1\dots7$
- f_{c_k} is the carrier band centre frequency in Hz
- t is the time variable in s
- m is the specified m -value
- f_{1_k} is the lower modulation frequency in Hz as per table B.1.
- f_{2_k} is the higher modulation frequency in Hz as per table B.1.

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

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

NOTE The level of each octave band is identical.

C.3.2 Modulation depth testing for STI indirect method

The indirect method derives the m -values from the impulse response. Equation C2 yields an exponentially decayed sine wave carrier that is used as a substitute impulse response for testing the dynamic range of the modulation domain. The decay of the function C2 is defined by the reverberation time.

$$A(t) = \sin(2\pi f_{c_k} t) 1000^{-\frac{t}{RT_{60}}} \quad (C2)$$

where

- k is the octave band index $k=1\dots7$
- f_{c_k} is the octave band centre frequency in Hz
- t is the time variable in s
- RT_{60} is the reverberation time in s.

For a given reverberation time, the m -values derived from the impulse response generated using C2 should match the theoretical m -values given by C3.

$$m(f_m, RT_{60}) = \frac{1}{\sqrt{1 + \left[\frac{2\pi f_m RT_{60}}{\log(10^6)} \right]^2}} \quad (C3)$$

where

- m is the m -value
- f_m is the modulation frequency in Hz
- RT_{60} is the reverberation time in s
- \log is the natural logarithm, $\log(10^6) \approx 13.8$

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

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

C.4.1 Flank attenuation slopes

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

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

Examination of the auditory masking functions shows that the steepest level-dependent masking slope of relevance is approximately 41 dB/octave. This maximum slope is derived by finding the lowest level in the 500 Hz band (48 dB), which produces a level approximately equal to the lowest reception threshold (6,5 dB at 1000 Hz).

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

C.4.2 Octave band filter testing - STIPA direct method

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

Since filters are likely to be asymmetric, both sides of the band-pass filter shall be investigated. The appropriate test signals for each observed octave band k are generated using equations C4 and C5. Note that the level of the non-modulated sine signal is adjusted by 3 dB to compensate for the fact that it is a non-modulated signal.

$$A(t) = 10^{\left(\frac{-41}{20}\right) \sin(2\pi f_{c_k} t)} \sqrt{0.5 \left(1 + 0.55 \left(\sin(2\pi f_{1_k} t) - \sin(2\pi f_{2_k} t)\right)\right)} + 10^{\left(\frac{-41}{20}\right) \sin(2\pi f_{c_k} t)} \sqrt{0.5 \left(1 + 0.55 \left(\sin(2\pi f_{1_k} t) - \sin(2\pi f_{2_k} t)\right)\right)} + 10^{\left(\frac{-3}{20}\right) \sin(2\pi f_{c_{k-1}} t)} \sqrt{0.5 \left(1 + 0.55 \left(\sin(2\pi f_{1_k} t) - \sin(2\pi f_{2_k} t)\right)\right)} + 10^{\left(\frac{-3}{20}\right) \sin(2\pi f_{c_{k-1}} t)} \sqrt{0.5 \left(1 + 0.55 \left(\sin(2\pi f_{1_k} t) - \sin(2\pi f_{2_k} t)\right)\right)} + 10^{\left(\frac{-3}{20}\right) \sin(2\pi f_{c_{k+1}} t)} \sqrt{0.5 \left(1 + 0.55 \left(\sin(2\pi f_{1_k} t) - \sin(2\pi f_{2_k} t)\right)\right)} + 10^{\left(\frac{-3}{20}\right) \sin(2\pi f_{c_{k+1}} t)} \sqrt{0.5 \left(1 + 0.55 \left(\sin(2\pi f_{1_k} t) - \sin(2\pi f_{2_k} t)\right)\right)}$$

(C4)

$$A(t) = 10^{\left(\frac{-41}{20}\right) \sin(2\pi f_{c_k} t)} \sqrt{0.5 \left(1 + 0.55 \left(\sin(2\pi f_{1_k} t) - \sin(2\pi f_{2_k} t)\right)\right)} + 10^{\left(\frac{-41}{20}\right) \sin(2\pi f_{c_k} t)} \sqrt{0.5 \left(1 + 0.55 \left(\sin(2\pi f_{1_k} t) - \sin(2\pi f_{2_k} t)\right)\right)} + 10^{\left(\frac{-3}{20}\right) \sin(2\pi f_{c_{k-1}} t)} \sqrt{0.5 \left(1 + 0.55 \left(\sin(2\pi f_{1_k} t) - \sin(2\pi f_{2_k} t)\right)\right)} + 10^{\left(\frac{-3}{20}\right) \sin(2\pi f_{c_{k-1}} t)} \sqrt{0.5 \left(1 + 0.55 \left(\sin(2\pi f_{1_k} t) - \sin(2\pi f_{2_k} t)\right)\right)} + 10^{\left(\frac{-3}{20}\right) \sin(2\pi f_{c_{k+1}} t)} \sqrt{0.5 \left(1 + 0.55 \left(\sin(2\pi f_{1_k} t) - \sin(2\pi f_{2_k} t)\right)\right)} + 10^{\left(\frac{-3}{20}\right) \sin(2\pi f_{c_{k+1}} t)} \sqrt{0.5 \left(1 + 0.55 \left(\sin(2\pi f_{1_k} t) - \sin(2\pi f_{2_k} t)\right)\right)}$$

(C5)

where

- k is the observed octave band index $k=1...7$
- f_{c_k} is the observed carrier band centre frequency in Hz
- t is the time variable in s
- f_{1_k} is the lower modulation frequency in Hz as per Table B.1.
- f_{2_k} is the higher modulation frequency in Hz as per Table B.1
- $f_{c_{k-1}}$ is the lower octave non-modulated carrier frequency in Hz
- $f_{c_{k+1}}$ is the higher octave non-modulated carrier frequency in Hz

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

If an m -value lower than $0,5 \pm 0,05$ is obtained, the level of the observed modulated carrier shall be increased in 1 dB steps to a level where an m -value of $0,5 \pm 0,05$ results. With the

2090 measured relative level, the corresponding sound pressure level as per Table A.1 is
2091 determined and reported as the sound pressure level below which the STI is underestimated.

2092 NOTE Testing both sides of each octave-band filter requires non-modulated sine carriers in the 63 Hz and 16 kHz
2093 octave bands.

2094 **C.4.3 Performance verification files**

2095 A number of reference files to verify the performance of STI measuring equipment can be
2096 downloaded from a number of websites including the following URLs: www.aes.org/standards
2097 and www.stipa.info

2098 The reference files enable manufacturers and users to confirm STI performance under a
2099 range of simulated conditions.

2100

Annex D

(informative)

Use of STI Measuring Devices

D.1 Overview

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

D.2 STIPA characterises only the speech transmission channel

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

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

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

In other words, STIPA measurements tell you only what the *speech transmission channel* does to the speech in terms of intelligibility. Before commencing, you should consider what the speech transmission channel you intend to test comprises.

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

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

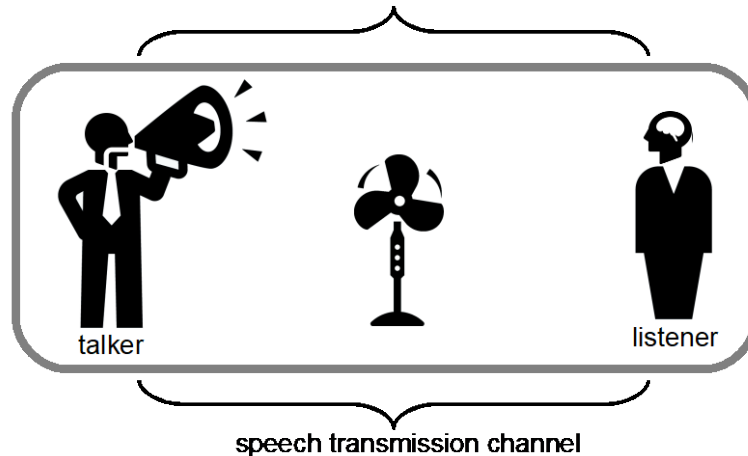


Figure D.1. Schematic representation of the definition of a speech transmission channel.

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

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

When performing STIPA tests:

- We replace the talker with a source of the test signal
- We replace the listener with a STIPA analyzer

D.3 Examples of test scenarios for STIPA tests

Perhaps the most common application of the STIPA test method is to evaluate PA systems – hence the “PA” in STIPA. A few common STIPA testing scenarios are presented below for, which the transmission channel and talkers and listeners are identified, and the factors that would be expected to influence the STI.

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

Table D.1 Scenario 1. PA with “live” announcer

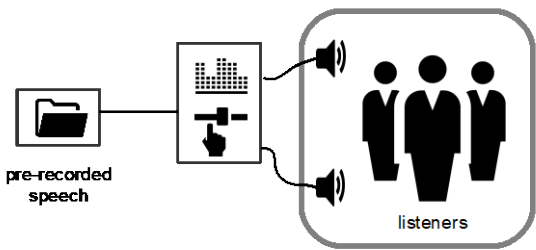
<p><i>Talker</i></p>	<p>The talker is usually a single person making announcements, who might (or might not) have been trained for this purpose. The talker is usually out of range of the PA system and unable to directly hear the announcements.</p>

<i>Listeners</i>	Listeners are all the people present in the venue for whom the announcements are intended. Listener locations to be considered are all spaces where the public is allowed.
<i>Speech transmission channel to be tested</i>	Everything from the paging microphone (in its acoustic environment) to all relevant listener locations and includes sound-system electronics.
<i>Factors influencing the STI</i>	<ul style="list-style-type: none"> • Noise and reverberation at the talker location, • Paging microphone characteristics and speaking distance • Frequency response and distortion of the sound system • Noise and reverberation at the listener location • Overall sound pressure level produced by the sound system

2159

2160

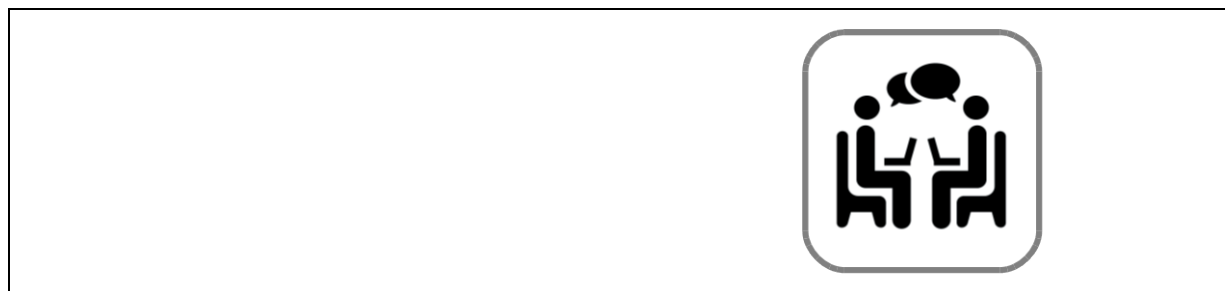
Table D.2 Scenario 2. PA with pre-recorded announcements

	
<i>Talker</i>	Instead of a live talker, recorded speech is used.
<i>Listeners</i>	Listeners are all the people present in the venue for whom the announcements are intended. Listener locations to be considered are all spaces where the public is allowed.
<i>Speech transmission channel to be tested</i>	Everything from (digital) audio storage and playback electronics up to all relevant listener locations and includes sound-system electronics.
<i>Factors influencing the STI</i>	<ul style="list-style-type: none"> • Frequency response and distortion of the sound system • Noise and reverberation at the listener location • Overall sound pressure level due to PA

2161

2162 Scenario 1 and Scenario 2 represent the most common public address and voice evacuation
 2163 scenarios. Other common scenarios are shown below. Scenario 3 is the “classic” application
 2164 of the Speech Transmission Index to pure room acoustics, without the involvement of
 2165 electronics for sound reproduction. STIPA evaluations can be very useful in identifying the
 2166 impact of factors relating to room acoustics (e.g. lack of acoustic absorption materials) and
 2167 ambient noise (e.g. due to air-conditioning systems) on speech intelligibility. Scenario 4 is a
 2168 typical lecture-type situation, where one lecturer speaks to a larger number of listeners in the
 2169 same room.

2170

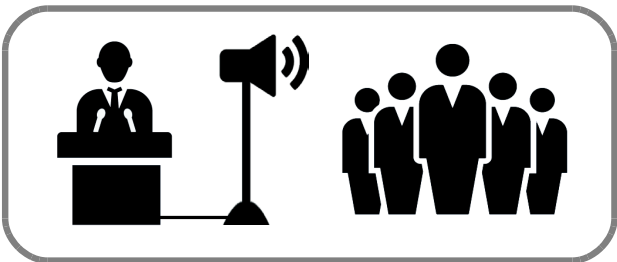
Table D.3 Scenario 3. “Live” meetings and conversations

<i>Talker/listeners</i>	In meetings and conversations, the same people take turns acting as talkers and listeners. All positions around a meeting table are therefore to be considered as talker positions as well as listener positions.
<i>Speech transmission channel to be tested</i>	Each individual talker and listener position combines into a transmission channel.
<i>Factors influencing the STI</i>	<ul style="list-style-type: none"> • Distance between talker and listener • Reverberation in the meeting room • Ambient noise in the meeting room; interfering speech from adjacent rooms • Vocal effort; speaking levels (relaxed vs. raised voice)

2171

2172

Table D.4 Scenario 4. Lecture

	
<i>Talker</i>	A single lecturer usually addresses a room full of people. The talker position is at the lectern, using a fixed microphone, or a somewhat larger presentation area if a wireless microphone is used.
<i>Listeners</i>	All seats in the audience are regarded as listening positions. Generally, there are more seats than can realistically be covered by STIPA measurements. A selection of representative seats (which shall always cover the expected worst-case seats) shall be selected.
<i>Speech transmission channel to be tested</i>	Everything from the microphone up to all listener positions in the room.
<i>Factors influencing the STI</i>	<ul style="list-style-type: none"> • Noise and reverberation in the lecture hall • Microphone characteristics and speaking distance • Frequency response and distortion of the sound system and possible the influence of acoustic feedback • Overall sound pressure level generated by the sound system, which will differ from seat to seat.

2173

2174 **D.4 Equipment and resources needed for a STIPA test**

2175 The following is required for a STIPA test:

2176 **D.4.1 Availability of the test signal**

2177 Access to the test signal is indispensable, but not always easily obtained. For instance, the
 2178 operator of a PA system at a shopping mall might not permit the playback of test signals
 2179 during opening hours. Or, even worse, the complete blocking of ground-to-air radio
 2180 communications used for air traffic control by playing 18 s of test signal is completely
 2181 unacceptable. Therefore, the channel shall be available in a configuration that matches
 2182 "normal operation."

D.4.2 A source of the STIPA test signal

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

3. A STIPA analyser

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

D.5 Steps in the overall procedure

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

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

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 in the so-called “Harvard list” with binaural listening according to [33]. The relation with sentence intelligibility is based on SRT (Speech Reception Threshold) results.

NOTE This graph has been updated from Ed. 4.

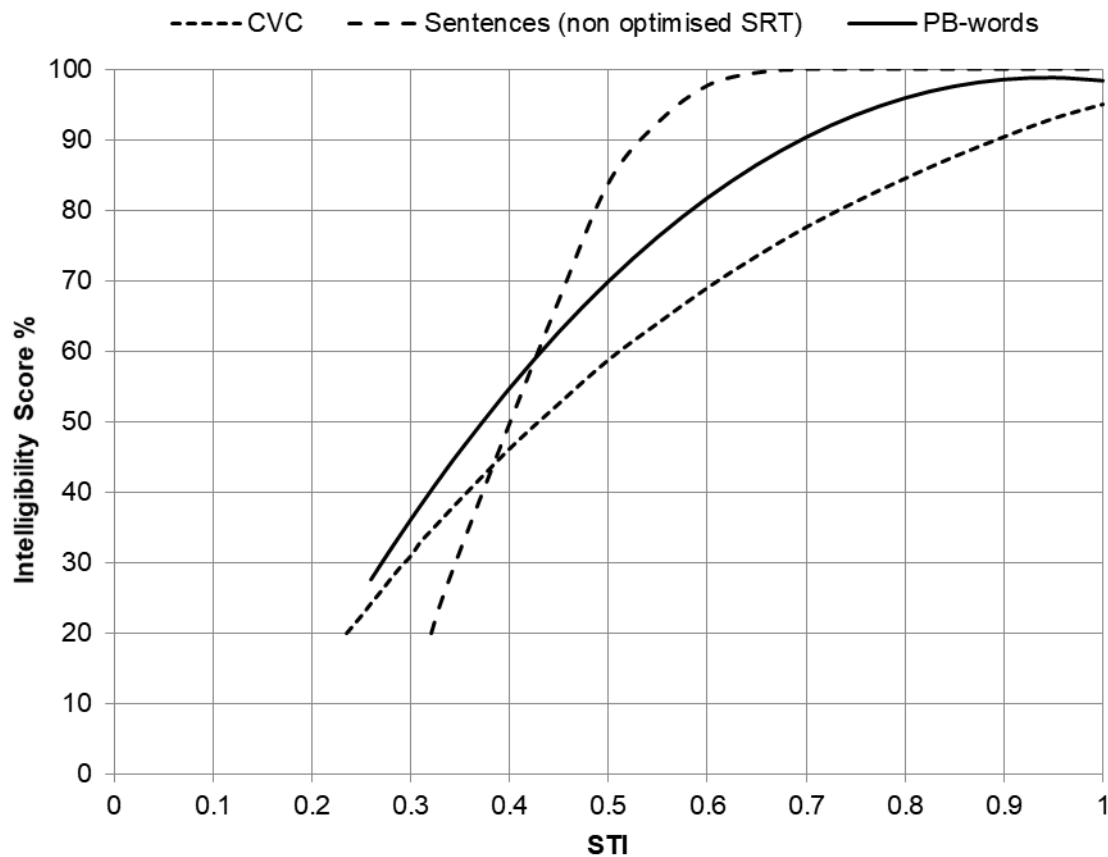


Figure E.1 Relationships between some speech intelligibility measures

E.2 Relationship between STI and listening difficulty

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

2237 The listening difficulty rating is the percentage of responses indicating a certain degree of
 2238 difficulty when listening to speech through a transmission channel. Note that listening
 2239 difficulty ratings decrease for conditions with improved speech transmission, contrary to
 2240 conventional intelligibility scores [31].

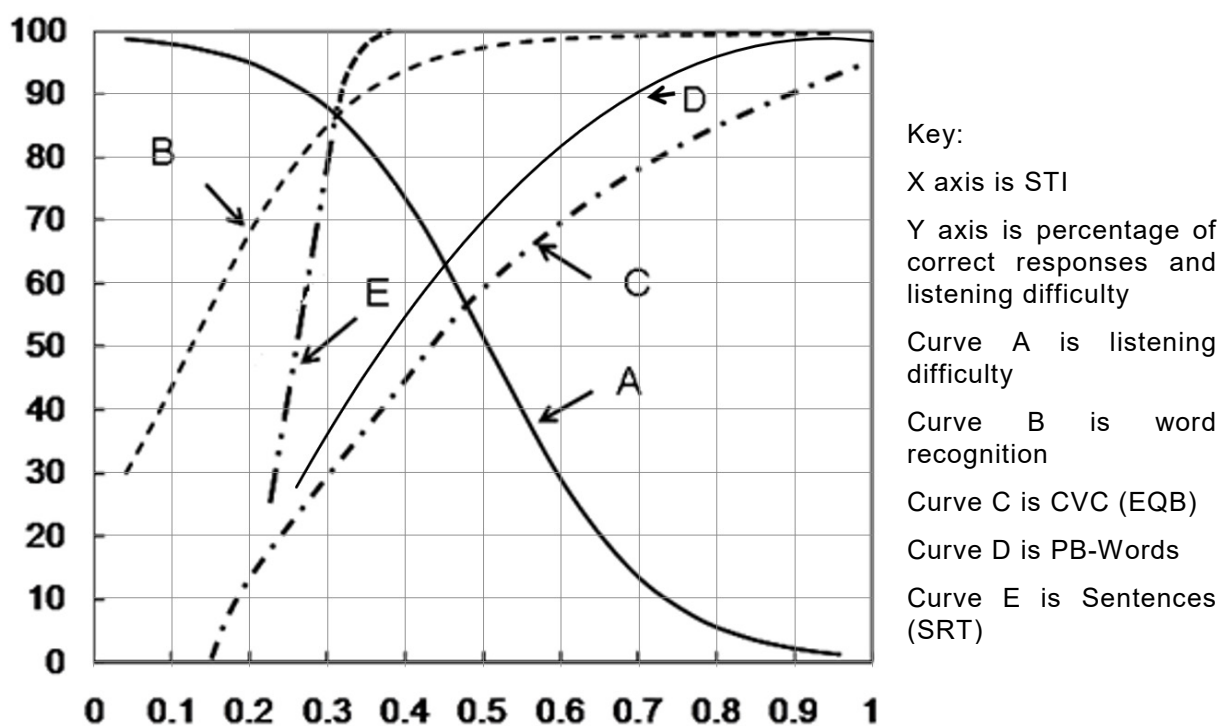
2241 Table E.1 lists the categories which are typically used to describe the listening difficulty.

2242 **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

2243

2244 Figure E.2 shows the relationship between listening difficulty ratings, intelligibility scores in
 2245 Figure E.1 and the STI.

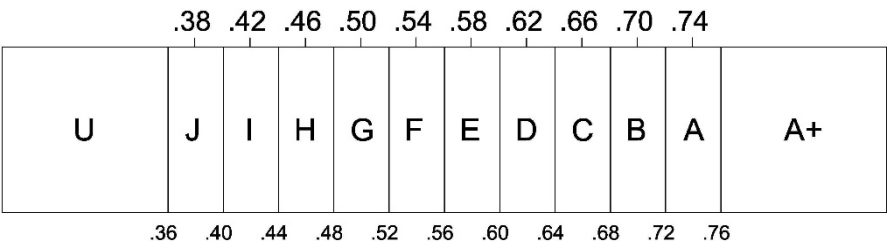


2246 **Figure E.2 – Relationship between STI, speech intelligibility scores and listening**
 2247 **difficulty ratings [34], [35]**

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 can then be obtained from an associated performance table (for an example, see Annex G). Figure F.1 shows the qualification bands:



Key

- Upper row of numbers: STI values at the centre of the bands;
- Row of letters: band designations;
- Lower row of numbers: STI values at the edges of the bands.

Figure F.1 – STI qualification bands

The spacing of the intervals in Fig 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	
These values should be regarded as minimum target values.				
NOTE 1 Perceived intelligibility relating to each category also depend on the frequency response at each listening position.				
NOTE 2 The STI values refer to measured values in sample listening positions or as required by specific application standards.				

Annex H (informative)

Non-native listeners

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

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). 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
<p>NOTE 1 For details on STI label categories, refer to ISO 9921.</p> <p>For a non-native listener of category II and to achieve an intelligibility equivalent to an STI of 0,45, the transmission system needs to achieve a performance of 0,60. For intermediate values between the stated standard STI, interpolation should be used to estimate the adjusted STI.</p>				

Annex I (informative)

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

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

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

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

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 are 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 normal listeners.

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

Table I.1 provides an indication of the adjusted intelligibility qualifications. It should be noted that the maximum intelligibility that can be reached depends on the degree of hearing loss [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.				
NOTE 2 Typical normal listeners have a PTA of between 0 dB and 5 dB				
To achieve an intelligibility equivalent to an STI of 0,45 for an older listener with PTA=15 dB, , the transmission system needs to achieve an STI of 0,57.				

Annex J (normative)

Setting & adjustment of STI test signal level

J.1 Overview

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

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

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

The speech level measurement method was developed by Houtgast and Steeneken ([47]). In order to overcome spectral differences between signals and to have the signal levels closely match a perceived level, the measurement method is based on the A-weighted filtering of the signal. In general, speech signals are not continuous and contain numerous pauses. When specifying the signal-to-noise ratio of speech and noise signals based on the measurement of L_{Aeq} values, it is evident that the number and the duration of the silences between utterances affect the result. For example, speech at the identical nominal level but with a different pattern of pauses lead to a different L_{Aeq} value being measured and therefore to differences in the resultant signal-to-noise ratio.

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

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

Extensive research [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 Real speech level

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

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

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

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

J.4 Corrected speech level derived from real speech level

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

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

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

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

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

Table J.1 Typical speech and test signal dynamics

Signal	L_{Aeq}	L_{Apk}	L_{A1}	L_{A10}
Typical Speech (dB)	60,0	79,9	67,1	63,3
STIPA (dB)	60,0	72,5	61,8	61,0
Difference	0,0	7,4	5,3	2,3

NOTE L_{A1} , and L_{A10} are 1%, 10% percentile exceedance levels over the speech sample and are measured with a FAST time constant.

Table J.2 shows these data in a different format that again illustrates the differences between typical speech and test signals. Further information can be found in [54].

Table J.2 Comparison of speech and the test signal

Signal	Typical crest Factor dB		$L_{A1} - L_{Aeq}$ dB	$L_{A10} - L_{Aeq}$ dB
	A wtd	C wtd		
Typical Speech	20,0	16,7	7,1	3,3
Pink Noise	12,0	11,2	1,8	0,1
STIPA	12,4	11,6	1,8	1,0

Annex K
(informative)

Example test report sheet for STI measurements

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.

Table K.1 Example test report sheet

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

2402

2403

Table K.2 Measurement data record sheet

MEASUREMENT DATA PER POSITION (use a different sheet for each position)								
Position / location reference								
Time of measurement								
Ambient noise levels (without test signal)								
	L_{Aeq} (dB)	Octave band (Hz)						
		125	250	500	1 000	2 000	4 000	8 000
L_{eq} measured over 15 s (dB)								
Description of the ambient noise: e.g. steady, transient, impulsive								
Climatic conditions: wind, rain, temperature, humidity								
STIPA method								
STIPA signal level	L_{Aeq} (dB)	Octave band (Hz)						
		125	250	500	1 000	2 000	4 000	8 000
L_{eq} measured over 15 s (dB)								
STIPA measurement number	Average	1	2	3	4	5	6	
STIPA results for each measurement: take 3 measurements ; if variation > 0,03, take 3 more								
Maximum variation of results								
Impulse response (IR) method								
Length of acquired impulse response Minimum for STI calculation: 1,6 s								
Test signal level	L_{Aeq} (dB)	Octave band (Hz)						
		125	250	500	1000	2000	4000	8000
L_{eq} with test signal (dB)								
Signal to noise ratio SNR (dB)								
STI results	STI or STIPA(IR)	Modulation transmission index MTI						
Transmission indices and STI								
Deviation of frequency response from 250 Hz to 12 kHz relative to 1 kHz.								
Frequency response measurement: Time window applied, smoothing.								
Does IR show arrivals likely to cause audible echoes?								
Results of processing measured data for different signal and noise levels								
Signal levels	L_{Aeq} (dB)	Octave band (Hz)						
		125	250	500	1000	2000	4000	8000
Proposed speech level (dB)								
Proposed ambient noise level (dB)								
STI results	STI	Modulation transmission index MTI						
Calculated STI and MTIs								

2404

Annex L (normative)

Prediction of STI using statistical methods

According to [28], the complete modulation transfer function $m_{f,k}$, at modulation frequency f_m in octave band k including temporal distortion and noise can be written as:

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

where

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

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

f_m is the modulation frequency;

t is the integration variable for time;

SNR_k is the signal-to-noise ratio in dB

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

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

where

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

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

r is the talker to listener distance;

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

T is the reverberation time of the room or space.

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

$$h(t) = I_D \delta(t) + I_R e^{-\frac{13,8t}{T}}$$

where

I_D is the intensity of the direct sound

I_R is the intensity of the reverberant sound

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

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

with

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

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

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

where

$$SNR = 10 \log \left(\frac{I_{s1m}}{I_N} \right)$$

and

I_N is the intensity of the noise

I_{s1m} is the intensity of the source at 1 m

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

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

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

$$C = I_D + I_R + \left[1 + \left(\frac{I_N}{I_D + I_R} \right) \right]^{-1}$$

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

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

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

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

Annex M (informative)

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

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

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

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

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

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

Step	Sub	Situation and action	Comments and further actions
1	a)	Collect all available data: <ul style="list-style-type: none"> - TI-values or the modulation transfer function (MTF) per octave band - The (acoustic) levels per octave band <p>The amount of TI data or MTF data depends on the selected STI method or implementation. Full STI provides 14 values per octave band (in total 98 values), while STIPA yields 2 values per octave band (in total 14 values). There are generally seven octave band levels, one level for each octave band.</p>	<p>The MTF can contain contributions from many different sources such as linear and non-linear distortion, digital (codec) effects, reverberation, echo, ambient noise and auditory masking contributions etc.</p> <p>NOTE An STI value or an MTI (modulation transmission index per octave band) value should not be used to construct MTF data.</p>
	b)	Only the TI values are known <p>If only the TI values are known (and not the MTF), then convert TI values into m-values using the processing of Step 2.</p>	Confirm the appropriate number of TI-values per octave band.
	c)	The MTF data is fully processed MTF and contains auditory masking contributions. <p>If the MTF data is fully processed and MTF contains auditory masking contributions, then remove the auditory contributions from the MTF data using processing Step 3.</p>	<p>When performing this step, the acoustic levels per octave band should also be measured or predicted.</p> <p>If simulations or measurements were done with no acoustic reference at all (e.g. electrical measurements), then skip this step.</p>
	d)	The MTF data is free from auditory contributions but not free from ambient noise contributions. <p>If the MTF data is free from auditory contributions but not free from ambient noise contributions, then remove the ambient noise contributions from the MTF using processing Step 4.</p>	<p>This is likely to be the starting point when m-values are acquired from a measuring device and measured in the <i>presence</i> of ambient noise.</p> <p>To perform Step 4, the signal-to-ambient noise ratios shall be known or predicted.</p>
	e)	The MTF data is free from auditory contributions and free from ambient noise <p>If the MTF data is free from auditory contributions and free from ambient noise contributions, then use</p>	This is likely to be the starting point when m -values are acquired from a measuring device in the <i>absence</i> of ambient noise.

Step	Sub	Situation and action	Comments and further actions
		processing Step 5 to add contributions from an alternative ambient noise spectrum and new speech levels into to MTF.	To perform Step 5, the signal-to-ambient-noise ratios shall be calculated.
	f)	Compute the STI Compute the STI using processing Step 6	Predicted or measured levels per octave band are needed to calculate auditory masking contributions.
2	Convert TI values into an MTF		Go to Step 3 if the MTF is already available
	a)	Calculate the effective signal-to-noise ratio from the TI_{k,f_m} values for octave band k and modulation frequency f_m using the formula: $SNReff_{k,f_m} = 30 TI_{k,f_m} - 15$	This step is used in situations where only transmission indices TI_{k,f_m} are available.
	b)	Calculate the MTF_k from the effective signal-to-noise ratio $SNReff_{k,f_m}$ (from Step 2a) using the formula: $m_{k,f_m} = \frac{1}{10^{\left(\frac{-SNReff_{k,f_m}}{10}\right)} - 1}$	The MTF_k is the series of m_{k,f_m} values corresponding with octave band k . All MTF_k series from all octave bands form the so-called MTF matrix.
3	Removal of auditory contributions from the MTF		If the m -values were obtained <i>before</i> auditory processing or were obtained electrically (and therefore do not contain any auditory contributions), then skip this step and go to Step 4.
	a)	Determine the acoustic octave band levels L_k in dB SPL for octave band k and convert them into intensities I_k . L_k is a combination of the received (deteriorated) test signal (including reflections) <i>and</i> the ambient noise level.	The octave band levels L_k is the sum of all acoustic sources for octave band k at the listener's ear or the measurement microphone. Keep the octave band intensities I_k for further processing.
	b)	By using table A.1, compute the auditory masking slope amf_{dB_k} for octave band k using the acoustic level L_{k-1} from the lower octave band ($k - 1$).	Within the STI concept, octave band 125 Hz has no lower neighbouring masking octave band. Results for the 125 Hz octave band are therefore calculated <i>without</i> auditory masking contributions.
	c)	Compute the auditory masking factor amf_k for octave band k based on the computed amf_{dB_k} using the formula: $amf_k = 10^{\left(\frac{amf_{dB_k}}{10}\right)}$	Note that amf_{dB_k} always is a negative value.
	d)	Compute the auditory masking intensity Iam_k for octave band k using the auditory masking factor amf_k from octave band k and the intensity I_{k-1} from the lower octave band $k - 1$ by using the formula: $Iam_k = amf_k \times I_{k-1}$	Keep the auditory masking intensity Iam_k for octave band k for further processing.
	e)	Compute the intensities of the absolute reception thresholds Irt_k for octave band k using table A.2	<i>Keep</i> the intensities of the absolute reception thresholds Irt_k in octave band k for further processing.
	f)	Compute the intensity ratio correction factor C_k , to be able to remove all auditory contributions from the MTF, using the formula: $C_k = \frac{I_k}{I_k + Iam_k + Irt_k}$	Use the stored intensities from Steps 3a, 3d and 3e.
	g)	Compute the MTF free of auditory contributions by multiplying the m -values with $\frac{1}{C_k}$ (from Step 3f) using the formula: $m'_{k,f_m} = m_{k,f_m} \times \frac{1}{C_k}$	This processing step is explained as: $m_{k,f_m} \times \frac{1}{C_k} = \frac{I_{0k}}{I_k + Iam_k + Irt_k} \times \frac{I_k + Iam_k + Irt_k}{I_k}$

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

Step	Sub	Situation and action	Comments and further actions
		$m_{k,fm}^* = m_{k,fm}'' \times C_k$	measurements, then skip Steps 5e to 5k and go to processing Step 6.
5 cont	e)	Compute the total acoustic level L_k and the acoustic intensity I_k for octave band k using the formulas: $I_k = I_{S_k} + I_{n_k}$ $L_k = 10 \times \log_{10}(I_k)$	The total acoustic octave-band level is needed for calculating auditory masking features. Keep the acoustic intensity for later use.
	f)	By using table A.1, compute the auditory masking slope amf_{dB_k} for octave band k using the acoustic level L_{k-1} from the lower octave band ($k-1$).	Within the STI concept, octave band 125 Hz has no lower neighbouring masking octave band. Results for the 125 Hz octave band are therefore calculated <i>without</i> masking contributions.
	g)	Compute the auditory masking factor amf_k for octave band k based on the computed amf_{dB_k} using the formula: $amf_k = 10^{\left(\frac{amf_{dB_k}}{10}\right)}$	Note that amf_{dB_k} is always a negative value.
	h)	Compute the auditory masking intensity I_{am_k} for octave band k using the auditory masking factor amf_k from octave band k and the intensity I_{k-1} from the lower octave band $k-1$ by using the formula: $I_{am_k} = amf_k \times I_{k-1}$	<i>Keep</i> the auditory masking intensity I_{am_k} for octave band k for further processing.
	i)	Compute the intensities of the absolute reception thresholds I_{rt_k} for octave band k using table A.2	<i>Keep</i> the intensities of the absolute reception thresholds I_{rt_k} for octave band k
	j)	Compute the intensity ratio correction factor C_k to add all auditory contributions to the MTF using the formula: $C_k = \frac{I_k}{I_k + I_{am_k} + I_{rt_k}}$	Combine the results from Step 5e, 5h and 5i.
	k)	Compute the new MTF including auditory contributions by multiplying the MTF (obtained in Step 5d) with C_k (from Step 5j) using the formula: $m_{k,fm}^* = m_{k,fm}^* \times C_k$	Continue with Step 6
6	Compute STI using the adjusted MTF		Get the adjusted MTF from Step 5
	a)	Convert adjusted MTF values into effective signal to noise ratio's $SNRef_{k,fm}$ using the formula: $SNRef_{k,fm} = 10 \times \log_{10}\left(\frac{m_{k,fm}}{1 - m_{k,fm}}\right)$	The adjusted MTF can be obtained from Step 5d (if electrical measurements of STI are to be used) or Step 5k (if acoustical measurements are to be used).
	b)	Truncate the $SNRef_{k,fm}$ values to the range of -15 dB and +15 dB	
	c)	Convert the truncated $SNRef_{k,fm}$ values into transmission indices $TI_{k,fm}$ using the formula: $TI_{k,fm} = \frac{SNRef_{k,fm} + 15}{30}$	
	d)	Compute mean transmission indices MTI_k for octave band k using: $MTI_k = \frac{1}{n} \sum_{m=1}^n TI_{k,fm}$	The value of n depends on the chosen STI method or implementation. For Full STI, n equals 14, for STIPA, n equals 2.
	e)	Compute the STI value using the weight factors of table A.1: $STI = \sum_{k=1}^7 \alpha_k \times MTI_k - \sum_{k=1}^6 \beta_k \times \sqrt{MTI_k \times MTI_{k+1}}$	Eventually, clip STI values higher than 1,00 to 1,00.

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

Table M.2 Example calculation

1 , Acquire raw MTF data with signal and noise levels							
	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz
Signal level L_{eq}	73,0	70,0	72,0	68,2	62,2	56,2	52,0
Background noise levels L_{eq}	62,0	56,0	50,0	44,0	39,5	35,0	29,0
MTF matrix with noise, temporal effects, masking and threshold factors							
0,63 Hz	0,98168	0,95224	0,95982	0,96882	0,97897	0,98340	0,99414
0,80 Hz	0,96649	0,92808	0,94078	0,95381	0,96897	0,97641	0,99214
1,00 Hz	0,94521	0,89688	0,91373	0,93279	0,95497	0,96541	0,98915
1,25 Hz	0,91887	0,86165	0,88067	0,90777	0,93898	0,95242	0,98416
1,60 Hz	0,88443	0,81937	0,83558	0,87274	0,91498	0,93243	0,97817
2,00 Hz	0,84998	0,78414	0,79250	0,83771	0,88998	0,91145	0,97118
2,50 Hz	0,81452	0,74992	0,74942	0,79868	0,86198	0,88846	0,96120
3,15 Hz	0,77197	0,71468	0,71636	0,75964	0,83198	0,86348	0,95022
4,00 Hz	0,73955	0,67845	0,69131	0,72962	0,79998	0,83649	0,93824
5,00 Hz	0,72436	0,62308	0,66526	0,72061	0,77198	0,81051	0,92626
6,30 Hz	0,71321	0,55262	0,64322	0,70760	0,74498	0,78452	0,91329
8,00 Hz	0,66864	0,51538	0,61116	0,66356	0,71998	0,76354	0,90131
10,00 Hz	0,58962	0,47914	0,54503	0,60251	0,69298	0,74755	0,89033
12,50 Hz	0,55315	0,44190	0,51297	0,60151	0,67798	0,73555	0,88135
2, Remove background noise, auditory masking and reception threshold factors							
	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz
$I_{s(kf)}$ - test signal only	19952623	10000000	15848932	6606934	1659587	416869	158489
$I_{n(kf)}$ - noise only	1584893	398107	100000	25119	8913	3162	794
Correction factor for Ambient Noise: $C_{1k}=I_{s(kf)}/(I_{s(kf)}+I_{n(kf)})$ (see Step 4c)	0,92641	0,96171	0,99373	0,99621	0,99466	0,99247	0,99501
Combined speech and noise level dB	73,332	70,170	72,027	68,216	62,223	56,233	52,022
Intensity of speech and noise I_k	21537516	10398107	15948932	6632053	1668499	420032	159284
Auditory masking factor amf dB		-23,1340	-24,7152	-23,7863	-25,6918	-33,8884	-36,8836
Auditory masking factor amf		0,00486	0,00338	0,00418	0,00270	0,00041	0,00020
$I_{am,k}$	0,000	104663,1	35110,05	66695,63	17884,30	681,536	86,084
Reception Threshold $L_{rt,k}$	46,0	27,0	12,0	6,5	7,5	8,0	12,0
$I_{rt,k}$	39810,72	501,187	15,8489	4,4668	5,6234	6,3096	15,8489
Correction factor for Auditory masking & Reception Threshold: $C_{2k}=I_k/(I_k+I_{am,k}+I_{rt,k})$ (see Step 3f)	0,99815	0,98999	0,99780	0,99004	0,98939	0,99837	0,99936
Inverse product of Correction Factors C_{1k} & C_{2k}	1,08143	1,05033	1,00853	1,01390	1,01615	1,00924	1,00566

3, Adjusted MTF matrix without noise, auditory masking and reception threshold							
0,63 Hz	1,06162	1,00016	0,96800	0,98229	0,99478	0,99249	0,99976
0,80 Hz	1,04519	0,97479	0,94880	0,96706	0,98462	0,98543	0,99775
1,00 Hz	1,02218	0,94202	0,92152	0,94575	0,97040	0,97433	0,99474
1,25 Hz	0,99369	0,90501	0,88818	0,92039	0,95414	0,96122	0,98972
1,60 Hz	0,95644	0,86061	0,84271	0,88487	0,92975	0,94105	0,98370
2,00 Hz	0,91919	0,82360	0,79926	0,84935	0,90435	0,91986	0,97667
2,50 Hz	0,88085	0,78766	0,75581	0,80978	0,87590	0,89667	0,96663
3,15 Hz	0,83483	0,75065	0,72247	0,77020	0,84541	0,87145	0,95559
4,00 Hz	0,79978	0,71259	0,69720	0,73976	0,81290	0,84422	0,94355
5,00 Hz	0,78334	0,65444	0,67093	0,73063	0,78445	0,81799	0,93150
6,30 Hz	0,77129	0,58043	0,64870	0,71743	0,75701	0,79177	0,91845
8,00 Hz	0,72308	0,54132	0,61637	0,67278	0,73161	0,77059	0,90641
10,00 Hz	0,63763	0,50325	0,54968	0,61088	0,70417	0,75445	0,89537
12,50 Hz	0,59819	0,46414	0,51735	0,60987	0,68893	0,74235	0,88633
4, Adjustment factors for operational speech and noise levels, auditory masking and reception threshold							
Operational speech L_{eq}	79,0	82,9	79,2	73,2	67,2	61,2	55,2
Operational background noise L_{eq}	67,0	70,0	72,0	68,0	65,0	63,0	60,0
$I_{S_k(f)}$ - test signal only	79432823	194984460	83176377	20892961	5248075	1318257	331131
$I_{N_k(f)}$ - noise only	5011872	10000000	15848932	6309573	3162278	1995262	1000000
Signal to Noise ratio	12,00	12,90	7,20	5,20	2,20	-1,80	-4,80
Correction factor for Ambient Noise: $C_{3k}=I_{S_k}/(I_{S_k}+I_{N_k})$ (see Step 5c)	0,94065	0,95122	0,83995	0,76805	0,62400	0,39784	0,24876
Combined speech and noise level dB	79,27	83,12	79,96	74,35	69,25	65,20	61,24
Combined intensity I_k	84444696	204984460	99025309	27202535	8410352	3313519	1331131
Auditory masking factor amf dB		-20,167	-18,241	-19,821	-22,627	-25,176	-29,535
Auditory masking factor amf		0,009622	0,014992	0,010420	0,005461	0,003037	0,001113
$I_{am,k}$	0,0	812565,8	3073133,4	1031856,5	148564,2	25540,0	3688,2
Reception Threshold $L_{rt,k}$	46	27	12	6,5	7,5	8	12
$I_{rt,k}$	39810,72	501,187	15,849	4,467	5,623	6,310	15,849
Correction factor for Auditory masking & Reception Threshold: $C_{4k}=I_k/(I_k+I_{am,k}+I_{rt,k})$ (see Step 5j)	0,99953	0,99605	0,96990	0,96345	0,98264	0,99235	0,99723
Product Correction Factors C_{3k} & C_{4k}	0,94021	0,94746	0,81467	0,73998	0,61317	0,39480	0,24807
5, Adjusted MTF matrix for operational speech and noise levels, auditory masking and reception threshold							
0,63 Hz	0,99814	0,94761	0,78860	0,72687	0,60997	0,39183	0,24801
0,80 Hz	0,98269	0,92357	0,77296	0,71561	0,60374	0,38904	0,24751
1,00 Hz	0,96106	0,89252	0,75074	0,69984	0,59502	0,38466	0,24676
1,25 Hz	0,93428	0,85746	0,72357	0,68107	0,58505	0,37949	0,24552
1,60 Hz	0,89925	0,81539	0,68653	0,65479	0,57010	0,37152	0,24402
2,00 Hz	0,86423	0,78033	0,65113	0,62851	0,55452	0,36316	0,24228
2,50 Hz	0,82818	0,74627	0,61573	0,59922	0,53707	0,35400	0,23979
3,15 Hz	0,78492	0,71121	0,58857	0,56994	0,51838	0,34405	0,23705
4,00 Hz	0,75195	0,67515	0,56799	0,54741	0,49844	0,33330	0,23406
5,00 Hz	0,73650	0,62006	0,54659	0,54065	0,48100	0,32294	0,23108
6,30 Hz	0,72517	0,54994	0,52848	0,53089	0,46418	0,31259	0,22784
8,00 Hz	0,67985	0,51287	0,50214	0,49785	0,44860	0,30423	0,22485
10,00 Hz	0,59950	0,47681	0,44781	0,45204	0,43178	0,29786	0,22211
12,50 Hz	0,56242	0,43975	0,42147	0,45129	0,42243	0,29308	0,21987

6, Process MTF matrix to yield STI							
6a, Convert m values into effective SNRs	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz
0,63 Hz	27,29870	12,57416	5,71756	4,25096	1,94214	-1,90925	-4,81745
0,80 Hz	17,54112	10,82225	5,32057	4,00763	1,82872	-1,96011	-4,82906
1,00 Hz	13,92334	9,19291	4,78826	3,67649	1,67096	-2,04031	-4,84649
1,25 Hz	11,52753	7,79279	4,17897	3,29494	1,49197	-2,13554	-4,87563
1,60 Hz	9,50654	6,45107	3,40459	2,78014	1,22577	-2,28304	-4,91073
2,00 Hz	8,03830	5,50505	2,71006	2,28357	0,95089	-2,43931	-4,95187
2,50 Hz	6,83047	4,68527	2,04763	1,74681	0,64523	-2,61225	-5,01098
3,15 Hz	5,62214	3,91419	1,55502	1,22292	0,31949	-2,80255	-5,07649
4,00 Hz	4,81658	3,17717	1,18848	0,82605	-0,02701	-3,01104	-5,14856
5,00 Hz	4,46397	2,12711	0,81167	0,70773	-0,33023	-3,21501	-5,22127
6,30 Hz	4,21379	0,87039	0,49525	0,53727	-0,62338	-3,42243	-5,30078
8,00 Hz	3,27056	0,22369	0,03712	-0,03737	-0,89607	-3,59270	-5,37489
10,00 Hz	1,75191	-0,40310	-0,91000	-0,83566	-1,19258	-3,72421	-5,44344
12,50 Hz	1,09003	-1,05178	-1,37567	-0,84882	-1,35846	-3,82390	-5,49997
6b Truncate SNR to +/-15 dB							
0,63 Hz	15,00000	12,57416	5,71756	4,25096	1,94214	-1,90925	-4,81745
0,80 Hz	15,00000	10,82225	5,32057	4,00763	1,82872	-1,96011	-4,82906
1,00 Hz	13,92334	9,19291	4,78826	3,67649	1,67096	-2,04031	-4,84649
1,25 Hz	11,52753	7,79279	4,17897	3,29494	1,49197	-2,13554	-4,87563
1,60 Hz	9,50654	6,45107	3,40459	2,78014	1,22577	-2,28304	-4,91073
2,00 Hz	8,03830	5,50505	2,71006	2,28357	0,95089	-2,43931	-4,95187
2,50 Hz	6,83047	4,68527	2,04763	1,74681	0,64523	-2,61225	-5,01098
3,15 Hz	5,62214	3,91419	1,55502	1,22292	0,31949	-2,80255	-5,07649
4,00 Hz	4,81658	3,17717	1,18848	0,82605	-0,02701	-3,01104	-5,14856
5,00 Hz	4,46397	2,12711	0,81167	0,70773	-0,33023	-3,21501	-5,22127
6,30 Hz	4,21379	0,87039	0,49525	0,53727	-0,62338	-3,42243	-5,30078
8,00 Hz	3,27056	0,22369	0,03712	-0,03737	-0,89607	-3,59270	-5,37489
10,00 Hz	1,75191	-0,40310	-0,91000	-0,83566	-1,19258	-3,72421	-5,44344
12,50 Hz	1,09003	-1,05178	-1,37567	-0,84882	-1,35846	-3,82390	-5,49997
6c Convert SNRs to Transmission Indices $Tl_{k(f)}$ and compute MTI values							
	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz
0,80 Hz	1,00000	0,86074	0,67735	0,63359	0,56096	0,43466	0,33903
1,00 Hz	0,96411	0,80643	0,65961	0,62255	0,55570	0,43199	0,33845
1,25 Hz	0,88425	0,75976	0,63930	0,60983	0,54973	0,42882	0,33748
1,60 Hz	0,81688	0,71504	0,61349	0,59267	0,54086	0,42390	0,33631
2,00 Hz	0,76794	0,68350	0,59034	0,57612	0,53170	0,41869	0,33494
2,50 Hz	0,72768	0,65618	0,56825	0,55823	0,52151	0,41292	0,33297
3,15 Hz	0,68740	0,63047	0,55183	0,54076	0,51065	0,40658	0,33078
4,00 Hz	0,66055	0,60591	0,53962	0,52753	0,49910	0,39963	0,32838
5,00 Hz	0,64880	0,57090	0,52706	0,52359	0,48899	0,39283	0,32596
6,30 Hz	0,64046	0,52901	0,51651	0,51791	0,47922	0,38592	0,32331
8,00 Hz	0,60902	0,50746	0,50124	0,49875	0,47013	0,38024	0,32084
10,00 Hz	0,55840	0,48656	0,46967	0,47214	0,46025	0,37586	0,31855
12,50 Hz	0,53633	0,46494	0,45414	0,47171	0,45472	0,37254	0,31667
MTI_k	0,75013	0,65686	0,57136	0,55622	0,51345	0,40721	0,33022

6d Apply Weightings and calculate STI							
	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz
alpha weighting factors, males	0,08500	0,12700	0,23000	0,23300	0,30900	0,22400	0,17300
$MTI_k \times \text{alpha weighting}$	0,06376	0,08342	0,13141	0,12960	0,15865	0,09122	0,05713
beta weighting factors, males	0,08500	0,07800	0,06500	0,01100	0,04700	0,09500	
$MTI_k \times \text{beta weighting}$	0,05967	0,04778	0,03664	0,00588	0,02149	0,03484	0,00000
$\sum \text{alpha} \times MTI$	0,71519						
$\sum \text{beta} \times MTI$	0,20630						
STI	0,509						

Annex N (informative)

Other methods of determining speech intelligibility

N.1 Overview

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

Another example of the first category (complex perceptual models) is the Speech Recognition Sensitivity model [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, 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 can approach that of the STI. In more complex situations, the accuracy of the STI outperforms all simpler metrics.

N.2 Word tests

The limitations of word tests are given in ISO/TR 4870 [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 is calculated. It is commonly used by experimental audiologists, because of its higher frequency resolution and its sensitivity to the intelligibility decrease at high vocal efforts.

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

2532 N.5 PESQ

2533 The Perceptual Evaluation of Speech Quality (PESQ) is especially useful for situations of high
2534 intelligibility and is based on mean opinion scores and especially suitable for measuring high
2535 quality transmission where speech intelligibility is less of an issue [42], [43].

Annex O (informative)

Alternative direct methods for measuring Full STI

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

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

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

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

Annex P (normative)

Information to be provided by manufacturers

P.1 Purpose of this Annex

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

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

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

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

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

P.3 Required information

The following information is required:

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

P.4 Declaration

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

Annex Q (informative)

Effect of uncertainties of selected parameters on STI uncertainty

Q.1 STI calculation framework

Q.1.1 Overview

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

Q.1.2 Statistical MTF

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

$$m_{k,f_m} = \frac{1}{\sqrt{1 + \left(\frac{2\pi f_m T_k}{13.8}\right)^2}} \quad (1)$$

We can define the constant

$$\alpha_m = \frac{2\pi f_m}{13.8} \quad (2)$$

Therefore

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

Q.1.3 Corrections

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

$$m'_{k,f_m} = m_{k,f_m} \frac{I_k}{I_k + I_{am,k} + I_{rt,k} + I_{n,k}} \quad (4)$$

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

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

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

$$I_{n,k} = 10^{-\frac{L_{SN}}{10}} I_k \quad (5)$$

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

$$m'_{k,f_m} = \frac{1}{1 + 10^{-\frac{L_{SN}}{10}}} \quad (6)$$

2632 **Q.1.4 Effective SNR**

2633 The effective SNR is calculated as follows:

$$SNR_{eff,k,f_m} = 10 \log \frac{m'_{k,f_m}}{1 - m'_{k,f_m}} \quad (7)$$

2635 Additionally, the effective SNR is limited to ± 15 dB. However, these limits are only exceeded
 2636 for values of m'_{k,f_m} greater than 0,97 or less than 0,03 which equate to reverberation times of
 2637 less than about 0,5 s or more than about 8 s when not considering other factors of influence.

2638 For many practical applications, these limits can therefore be neglected or considered second
 2639 order effects. In particular, they have little influence on the uncertainty estimates that we
 2640 consider in the following.

2641 **Q.1.5 Modulation Transfer Index MTI**

2642 Based on the effective SNR for each octave band k and modulation frequency f_m the
 2643 modulation transfer index (MTI) is calculated by averaging over all modulation frequencies.

$$MTI_k = \frac{1}{n} \sum_{m=1}^n \frac{SNR_{eff,k,f_m} + 15}{30} \quad (8)$$

2645 Due to the limits on the effective SNR the resulting MTI value is limited to the range of 0 to 1.

2646 **Q.1.6 Speech Transmission Index STI**

2647 Now the STI can be calculated as follows:

$$STI = \sum_{k=1}^7 \bar{\alpha}_k MTI_k - \sum_{k=1}^6 \bar{\beta}_k \sqrt{MTI_k MTI_{k+1}} \quad (9)$$

2650 where $\bar{\alpha}_k$ and $\bar{\beta}_k$ are the redundancy weights for male speech.

2651 **Q.2 The effect of RT uncertainty on STI uncertainty**

2652 In this text, we derive how sensitive the STI is when considering small changes to the
 2653 reverberation time (RT). We ignore any corrections for the moment, so $m'_{k,f_m} = m_{k,f_m}$.

2654 **Q.2.1 Modulation Transfer Function**

2655 Given equation (3), small changes of T_k translate to changes in m_k as follows:

$$\Delta m_k = \frac{dm_k}{dT_k} \Delta T_k \quad (10) \text{ with the first order derivative of } m_k \text{ as a function of } T_k:$$

$$\frac{dm_k}{dT_k} = -\alpha_m^2 T_k (1 + \alpha_m^2 T_k^2)^{-\frac{3}{2}} \quad (11)$$

2658 This expression does not depend directly on the octave band k or on the respective frequency.
 2659 This dependency is created only indirectly through the dependency of T on the frequency.

2660 **Q.2.2 Uncertainty in STI**

2661 For the sake of simplicity, let us assume that the reverberation time is approximately
 2662 independent of frequency so that $MTI_k = MTI_j$ for all j, k in equation (8). This might not be

2663 exactly true in practice but for many applications dependence on frequency should be a
2664 secondary effect with respect to the uncertainty.

2665 Accordingly, we can write:

$$2666 \quad STI = \left(\sum_{k=1}^7 \bar{\alpha}_k - \sum_{k=1}^6 \bar{\beta}_k \right) MTI \quad (11)$$

2667 knowing that by definition $\sum_{k=1}^7 \bar{\alpha}_k - \sum_{k=1}^6 \bar{\beta}_k = 1$, we can insert and simplify as follows:

$$2668 \quad STI = \frac{1}{n} \sum_{m=1}^n \frac{15+10 \log \frac{m'_{f_m}}{1-m'_{f_m}}}{30} \quad (12)$$

$$2669 \quad STI = \frac{1}{2} + \frac{1}{3n} \sum_{m=1}^n \log \frac{m'_{f_m}}{1-m'_{f_m}} \quad (13)$$

2670 This means that the only variable is m'_{f_m} which in turn only depends on T . In order to calculate
2671 the uncertainty we have to determine the total derivative

$$2672 \quad \Delta STI = \frac{dSTI}{dT} \Delta T \quad (14)$$

2673 knowing that the STI consists of n separate transfer function contributions

$$2674 \quad \frac{dSTI}{dT} = \sum_{m=1}^n \frac{dSTI}{dm'_{f_m}} \frac{dm'_{f_m}}{dT} \quad (15)$$

2675 one for each modulation frequency. Now we calculate the derivatives of STI

$$2676 \quad \frac{dSTI}{dm'_{f_m}} = \frac{1}{42} \frac{d}{dm'_{f_m}} \left(\log \frac{m'_{f_m}}{1-m'_{f_m}} \right) \quad (16)$$

2677 yielding

$$2678 \quad \frac{dSTI}{dm'_{f_m}} = \frac{1}{(42 \ln 10)} \frac{1}{m'_{f_m}(1-m'_{f_m})} \quad (17)$$

2679 We can now insert the derivatives for MTF eq. (11) and STI eq. (18) into eq. (16) in order to
2680 obtain a single function describing the effect of RT uncertainty.

2681 Following standard GUM methods, we calculate the uncertainty of the STI as the statistical
2682 average over an ensemble of STI values whose spread is defined by the uncertainty of T .

$$2683 \quad u(STI) = \sqrt{\langle \Delta STI^2 \rangle - \langle \Delta STI \rangle^2} = \sqrt{\left\langle \frac{dSTI}{dT} \Delta T^2 \right\rangle - \left\langle \frac{dSTI}{dT} \Delta T \right\rangle^2} \quad (18)$$

$$2684 \quad u(STI) = \left| \frac{dSTI}{dT} \right| \sqrt{\langle \Delta T^2 \rangle - \langle \Delta T \rangle^2} \quad (19)$$

$$2685 \quad u(STI) = \left| \frac{dSTI}{dT} \right| u(T) \quad (20)$$

2686 Inserting yields

$$u(STI) = \left| \sum_{m=1}^n \frac{1}{(42 \ln 10)} \frac{-\alpha_m^2 T (1 + \alpha_m^2 T^2)^{-\frac{3}{2}}}{m'_{f_m} (1 - m'_{f_m})} \right| u(T) \quad (21)$$

which can be simplified to

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

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

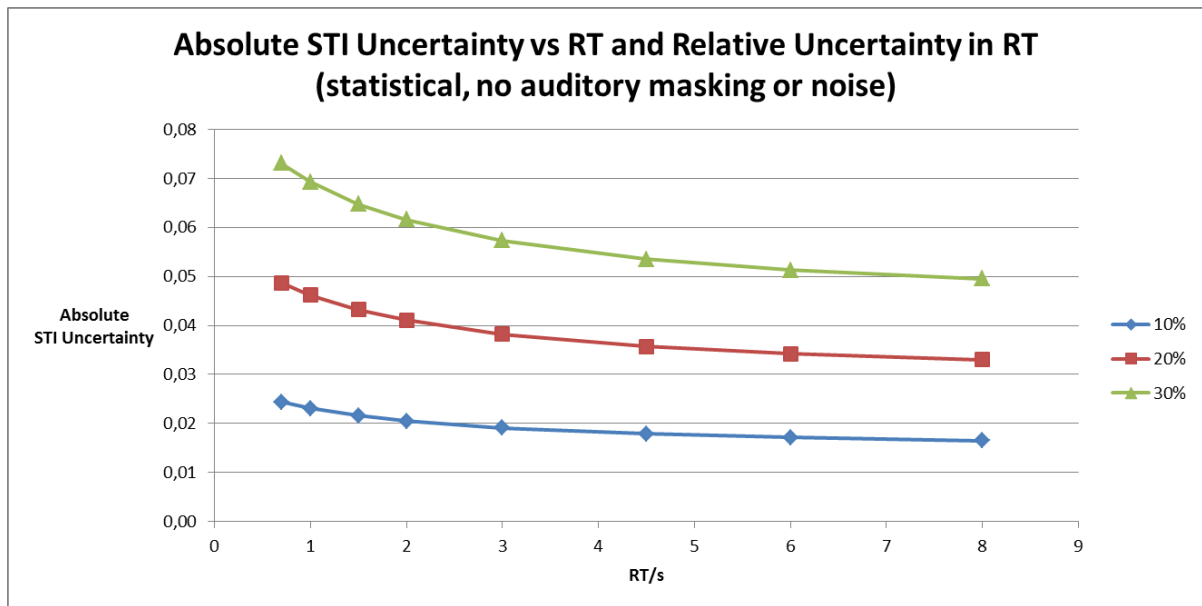


Figure Q.1 - Uncertainty in absolute value of STI vs reverberation time RT with various degrees of uncertainty in RT.

Q.2.3 Conclusions:

- Larger RT uncertainties ($\geq 15\%$) lead to significant STI uncertainty (> 0.03).
- Uncertainty is dependent on RT and is more critical for shorter RTs (however, shorter RTs are usually less of a problem in practice).
- This matters for modeling input data (calculated or estimated RT) as well as when measurements are taken at too few locations, such that the locally measured RTs are not consistent with the majority of non-measured locations.

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

In this text we derive how sensitive the STI is against small changes of the signal-to-noise level.

Q.3.1 Ideal transfer function

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

$$m'_{k,f_m} = \frac{1}{1 + \frac{1}{\sigma}} \quad (23)$$

2710

2711 Then the STI equation (14) simplifies to

$$2712 \quad STI = \frac{1}{2} + \frac{1}{3} \log \sigma = \frac{1}{2} + \frac{1}{30} L_{SN} \quad (24)$$

2713 It is immediately clear that a change of 1 dB in the signal-to-noise level corresponds to a
2714 change of 0,033 in STI.

$$2715 \quad \Delta STI = \frac{dSTI}{dL_{SN}} \Delta L_{SN} = \frac{1}{30} \Delta L_{SN} \quad (25)$$

2716 The same is true for the uncertainty

$$2717 \quad u(STI) = \left| \frac{dSTI}{dL_{SN}} \right| u(L_{SN}) = \frac{1}{30} u(L_{SN}) \quad (26)$$

2718 This is true as long as approximately $-15 \text{ dB} \leq L_{SN} \leq 15 \text{ dB}$.

2719

2720 **Q.3.2 Reverberation**

2721 Given eqs. (4) and (6) with a signal-to-noise factor $\sigma = 10^{\frac{L_{SN}}{10}}$ we have

$$2722 \quad m'_{k,f_m} = m_{k,f_m} \frac{1}{1+\frac{1}{\sigma}} \quad (27)$$

2723 The derivative is

$$2724 \quad \frac{dm'_{k,f_m}}{d\sigma} = m_{k,f_m} \frac{1}{(1+\sigma)^2} \quad (28)$$

2725 When only considering uncertainties in S/N, i.e. unknown changes in σ , the differential for the
2726 STI is

$$2727 \quad \Delta STI = \frac{dSTI}{d\sigma} \Delta \sigma \quad (29)$$

$$2728 \quad \frac{dSTI}{d\sigma} = \sum_{m=1}^n \frac{dSTI}{dm'_{f_m}} \frac{dm'_{f_m}}{d\sigma} \quad (30)$$

2729 This assumes that the S/N, the RT and the S/N uncertainty are approximately constant over
2730 frequency. Finally we have for the uncertainty:

$$2731 \quad u(STI) = \left| \frac{dSTI}{d\sigma} \right| u(\sigma) \quad (31)$$

2732 Notice that this error estimate is only valid as long as the +/- 15 dB limit approximately holds.

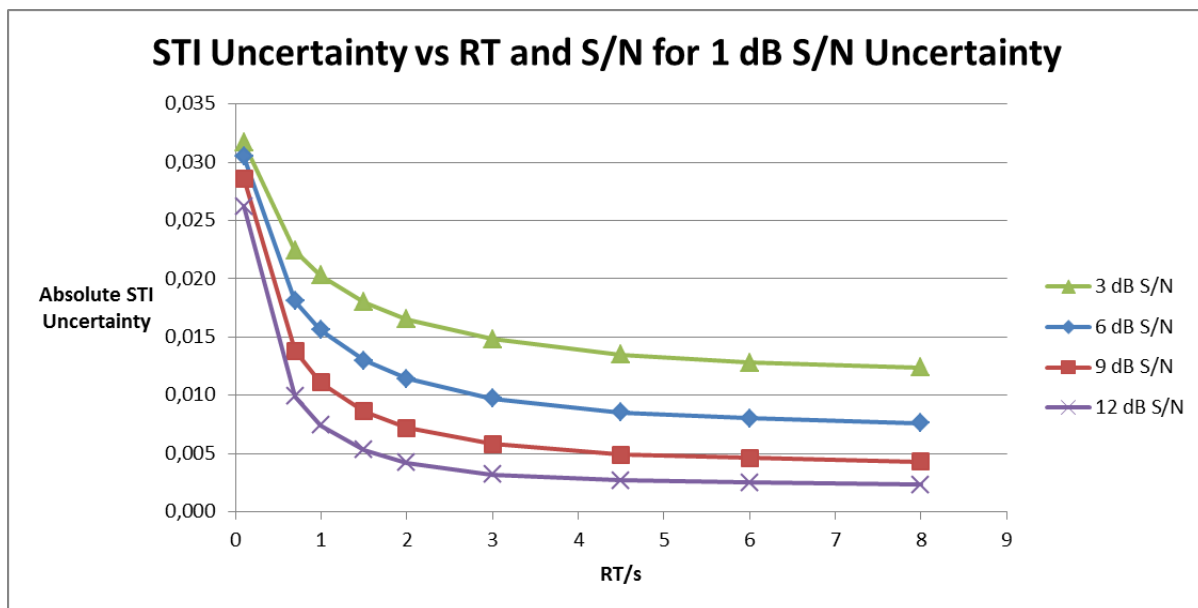


Figure Q.2 - Uncertainty in absolute value of STI vs reverberation time RT with 1 dB uncertainty in SNR at various SNRs.

Q.3.3 Conclusions:

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

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

Q.4.1 Overview

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

Q.4.2 Auditory masking

Given equation (4) and knowing $I_{am,k} = I_{k-1} amf$ for $k > 1$, the auditory masking is accounted for as follows

$$m'_{k,f_m} = m_{k,f_m} \frac{I_k}{I_k + I_{k-1} amf} \quad (32)$$

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

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

$$m'_{k,f_m} = m_{k,f_m} \frac{1}{1 + amf \frac{1}{r_k}} \quad (33)$$

2759 Note that there is no correction for $k = 1$, i.e. $amf = 0$. For $k > 1$ the derivative is

$$2760 \quad \frac{dm'_{k,fm}}{dr_k} = m_{k,fm} \frac{\frac{1}{amf}}{\left(1 + \frac{1}{amf} r_k\right)^2} \quad (34)$$

2761 When only considering an uncertainty in the level difference (i.e. intensity ratio), the
2762 uncertainty for the STI is

$$2763 \quad \Delta STI = \frac{dSTI}{dr} \Delta \quad (35)$$

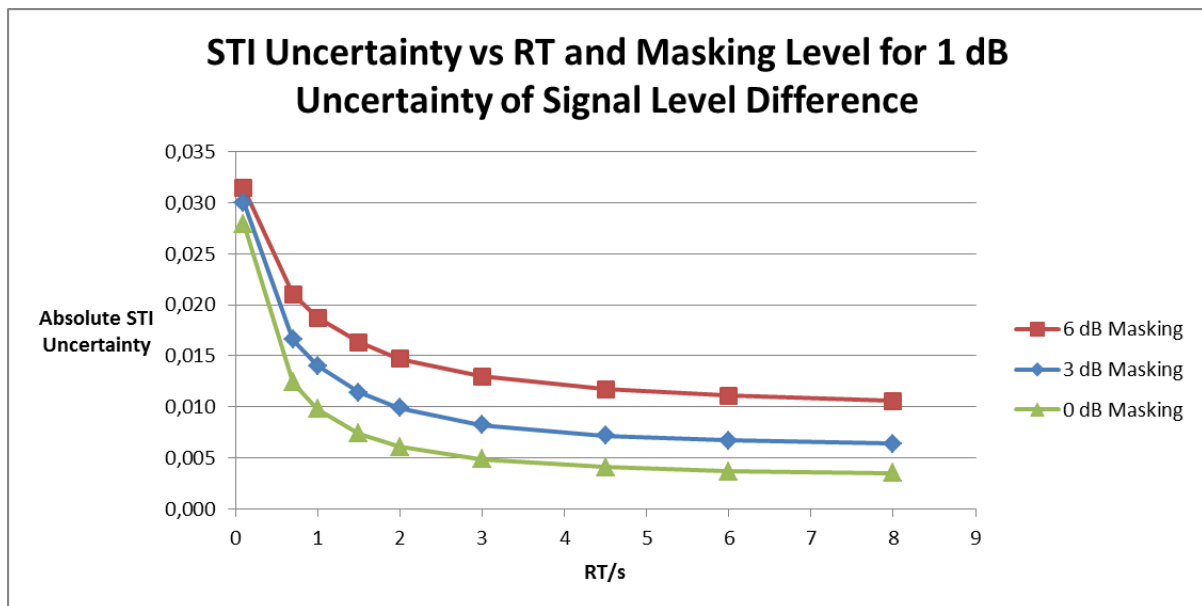
$$2764 \quad \frac{dSTI}{dr} = \sum_{m=1}^n \frac{dSTI}{dm'_{fm}} \frac{dm'_{fm}}{dr} \quad (36)$$

2765 This assumed that the RT and the uncertainty of the intensity ratio as well as signal level
2766 differences are approximately constant over frequency. We have also taken a lower bound for
2767 m by assuming that the lowest octave band is affected by masking in the same way as the
2768 other bands, thus setting $m_1 = m_2 = m_k$. Finally, we have for the uncertainty:

$$2769 \quad u(STI) = \left| \frac{dSTI}{dr} \right| u(r) \quad (37)$$

2770 Expectedly, the resulting curves are similar to the S/N curves. In fact, a curve for a certain
2771 masking level is equivalent to the S/N curve of the same level when negated and increased by
2772 10 dB. Therefore 6 dB masking corresponds to 4 dB SNR, 3 dB masking relate to 7 dB SNR
2773 (all at a maximum auditory masking of $amf = 0,1$). In particular, we also have the same
2774 asymptotic limit for vanishing T :

$$2775 \quad u(STI) = \left| \frac{dSTI}{dL_{Mask}} \right| u(L_{Mask}) = \frac{1}{30} u(L_{Mask}) \quad (38)$$



2776 **Figure Q.3 - Uncertainty in absolute value of STI versus reverberation time RT with**
2777 **various degrees of masking.**
2778

2779 Q.4.3 Conclusions:

- 2780 • The results in the diagram are based on certain assumptions:

- 2781 ○ Signal levels are 100 dB or higher throughout the bandwidth, thus the
2782 maximum *amf* is applied.
- 2783 ○ The masking level represents the differential signal level of a lower octave
2784 band relative to the next higher octave band. This difference is assumed
2785 constant over the entire audio bandwidth (“0 dB” would be pink noise, “6 dB”
2786 has a 6 dB decay per octave in the raw signal, i.e. similar to speech above
2787 1 kHz).
- 2788 • For $RT > 0$, the curves approach the expected upper bound of 0,033 / dB.
- 2789 • A non-zero RT lowers this sensitivity measurably.
- 2790 • Again, the uncertainties are largest where the RT is lowest. But low RT values generally
2791 imply a more controlled environment, so the increase is not as critical.
- 2792 • For less than about 6 dB of masking and typical $RT \geq 2$ the absolute RT value is less
2793 important for the uncertainty of the STI.
- 2794 • For exceptionally high masking levels, consistently greater than 6 dB per octave, the
2795 uncertainty can rise significantly for shorter RTs.
- 2796 • Assuming that most RT measurements are about 10% accurate and level measurements
2797 perhaps about 1 dB, the uncertainty in the RT is more important than the uncertainty of
2798 the masking level (roughly 0,02 versus 0,01).
- 2799
- 2800

Bibliography

- 2801
- 2802 [1] Mapp, P New Techniques for Measuring Speech Privacy & Efficiency of Sound Masking
2803 Systems. AES 122nd Convention, Vienna, May 2007
- 2804 [2] IEC 60318-1:2009, Electroacoustics - Simulators of human head and ear - Part 1: Ear
2805 simulator for the measurement of supra-aural and circumaural earphones
- 2806 [3] IEC 61672 (all parts) *Electroacoustics – Sound level meters – Part 1: Specifications*
- 2807 [4] ANSI (1969). ANSI S3.5-1969, American National Standard, "Methods for Calculation of
2808 the Articulation Index" (American National Standards Institute New York.)
- 2809 [5] Steeneken, H.J.M. and Houtgast, T., "Some applications of the Speech Transmission
2810 Index (STI) in auditoria", *Acustica* 51, 1982, p.229-234.
- 2811 [6] Steeneken, H.J.M. and Houtgast, T., "A physical method for measuring speech
2812 transmission quality", *J. Acoust. Soc. Amer.* 67, 1980, 31, p.318-326.
- 2813 [7] Houtgast, T. and Steeneken, H.J.M., "The modulation transfer function in room
2814 acoustics as a predictor of speech intelligibility", *Acustica* 28, 1973, p.66-73.
- 2815 [8] Steinbrecher, T, "Speech Transmission Index: Too weak in time and frequency?", *Proc*
2816 *IOA Vol 30 Part 6* 2008
- 2817 [9] Mapp, P, "Is STIPA a robust measure of speech intelligibility performance," AES 118th
2818 Convention, Barcelona, 2005.
- 2819 [10] Mapp, P Some Effects of Equalisation on Sound System Intelligibility and Measurement
2820 AES 115th Convention 2003 Preprint
- 2821 [11] Leembruggen, G and A Stacey "Should the Matrix be reloaded *Proc IOA Vol 28 Part 6*
2822 2006
- 2823 [12] Leembruggen, G. "Is SII better than STI at recognising the effects of poor tonal balance
2824 on intelligibility?" *Proc IOA Vol 28 Part 6* 2006
- 2825 [13] Mapp P, Practical application of STI to assessing Public Address and Emergency Sound
2826 Systems. Past, present and future of the Speech Transmission Index, TNO 2002
- 2827 [14] Festen, JM & Plomp, R, Effects of fluctuating noise and interfering speech on the
2828 speech reception threshold for impaired and normal hearing, *J. Acoust. Soc. Amer.*,
2829 88(4) 1990
- 2830 [15] Duquesnoy, A.J.H/M and Plomp, R. (1980) "Effect of reverberation and noise on the
2831 intelligibility of sentences in case of presbycusis". *J. Acoust. Soc. Am.* 68, p.537-544.
- 2832 [16] Mapp, P, "Assessing the Potential Intelligibility of Assistive Audio Systems for the Hard
2833 of Hearing and Other Users", AES 124th convention, Amsterdam, 2008.
- 2834 [17] Schroeder, M, Modulation Transfer Functions: Definition and Measurement, *Acustica*,
2835 49, 1981
- 2836 [18] Houtgast, T. and Steeneken, H.J.M., "The modulation transfer function in room
2837 acoustics as a predictor of speech intelligibility", *Acustica* 28, 1973, p.66-73.
- 2838 [19] Houtgast, T. and Steeneken, H.J.M., "A Multi-lingual evaluation of the Rasti-method for
2839 estimating speech intelligibility in auditoria", *Acustica* 54, 1984, p.185-199

- 2840 [20] Steeneken, H.J.M. and Houtgast, T., "The temporal envelope spectrum of speech and
2841 its significance in room acoustics", Proc. 11th International Congress on Acoustics,
2842 Paris, 1983, Vol. 7, p.85-88.
- 2843 [21] Steeneken, H.J.M., Verhave, J.A., McManus, S., and Jacob, K.D., (2001) "Development
2844 of an Accurate, Handheld, Simple-to-use Meter for the Prediction of Speech
2845 Intelligibility", Proceedings IoA 2001, Reproduced sound (17). Stratford-upon-Avon, UK.
- 2846 [22] Mapp, P. Further thoughts on Speech Transmission Index (STI). IOA Reproduced
2847 Sound 18.Proc. IOA Vol 24 Pt 8. 2002
- 2848 [23] Bozzoli, F, Farina, A, "Influence of artificial mouth's directivity in determining Speech
2849 Transmission Index"., AES 119th Convention, New York, 2005
- 2850 [24] IEC 60268-3:2000, Sound system equipment – Part 3: Amplifiers
- 2851 [25] Rife, D., "Modulation Transfer Function Measurements with Maximum-Length
2852 Sequences", J. Audio Eng. Soc., Oct. 1992, Vol. 40, No. 10.
- 2853 [26] ANSI standard S12.2 1995, Criteria for evaluation of room noise
- 2854 [27] Wijngaarden, S, Drullman, R, "Binaural intelligibility prediction based on the speech
2855 transmission index" J. Acoust. Soc. Amer. 123, 2008, p.4514-4523
- 2856 [28] Houtgast, T, Steeneken, H and Plomp, R. "Predicting Speech Intelligibility in Rooms
2857 from the Modulation Transfer Function. i. General Room Acoustics" Acustica Vol 46
2858 1980
- 2859 [29] Steeneken, H.J.M. and Houtgast, T., "Mutual dependency of the octave-band weights in
2860 predicting speech intelligibility," Speech Communication 28, 1999, p.109-123.
- 2861 [30] Zwicker, E. and Feltkeller, "The ear as an information receiver", (ASA)
- 2862 [31] Wijngaarden, S.J. van and Steeneken, H.J.M. (1999). "Objective prediction of speech
2863 intelligibility at high ambient noise levels using the speech transmission index."
2864 In Eurospeech99 – Proceedings of the 6th European Conference on Speech
2865 Communication and Technology, Budapest, Vol 6, p.2639-2642
- 2866 [32] Steeneken, H.J.M. and Houtgast, T., "Validation of the STIr method with the revised
2867 model," Speech Communication 38, 2002, p.413-425.
- 2868 [33] Morales L, Li F. "A new verification of the speech transmission index for the English
2869 language" Speech Communication 105, 2018 1–11
- 2870 [34] Hiroshi Sato, John S. Bradley, and Masayuki Morimoto, Using listening difficulty ratings
2871 of conditions for speech communication in rooms , J. Acoust. Soc. Am. 117, 1157 (2005)
- 2872 [35] Hayato Sato, Masayuki Morimoto, Hiroshi Sato, and Megumi Wada, Relationship
2873 between listening difficulty and acoustical objective measures in reverberant sound
2874 fields, J. Acoust. Soc. Am. 123, 2087-2093 (2008)
- 2875 [36] Wijngaarden, S.J., Steeneken, H.J.M. and Houtgast, T. (2002) "Quantifying the
2876 intelligibility of speech in noise for non-native listeners," J. Acoust. Soc Am. 112,
2877 p.3004-3013
- 2878 [37] ISO-7029: 2000, Acoustics – Statistical distribution of hearing thresholds as a function
2879 of age.

- 2880 [38] Sato,H, Kurakata,K, Mizunami,T, “Accessible speech messages for the elderly in rooms”
2881 9th Western Pacific Acoustics Conference Seoul, Korea, 2006
- 2882 [39] ISO/TR 22411:2008 Ergonomics data and guidelines for the application of ISO/IEC
2883 Guide 71 to products and services to address the needs of older persons and persons
2884 with disabilities.
- 2885 [40] Steeneken, H.J.M. and Houtgast, T (1986) Comparison of some methods for measuring
2886 speech levels Report IZF1986-20 TNO Institute for Perception Soesterberg Netherlands
- 2887 [41] H. Müsch and S. Buus. ‘Using statistical decision theory to predict speech intelligibility. I.
2888 Model structure’. J.Acoust.Soc.Am. 109, 2896–2909. (2001).
- 2889 [42] ITU-T Rec P.862, Perceptual evaluation of speech quality (PESQ)” An objective method
2890 of end-to-end speech quality assessment of narrow band telephone networks and
2891 speech codecs. (International Telecommunication Union, Geneva Switzerland 2001 Feb)
- 2892 [43] J. G. Beerends, A. P. Hekstra, A. W. Rix and M. P. Hollier. ‘PESQ, the new ITU
2893 standard for objective measurement of perceived speech quality, Part II - Perceptual
2894 model’. J. Audio Eng. Soc., vol. 50, pp. 765-778. (2002 Oct.).
- 2895 [44] ISO 3382-1:2009, *Acoustics – Measurement of room acoustic parameters – Part 1:*
2896 *Performance spaces*
- 2897 [45] ISO/TR 4870:1991, Acoustics — The construction and calibration of speech intelligibility
2898 tests (*Withdrawn*)
- 2899 [46] Steeneken, H.J.M. and Houtgast, T., “Phoneme-group specific octave-band weights in
2900 predicting speech intelligibility,” Speech Communication 38, 2002, p.399-411.
- 2901 [47] Past, present and future of the Speech Transmission Index, TNO 2002
- 2902 [48] Hammond, R, Mapp P, Hill A. “The influence of discrete arriving reflections on perceived
2903 intelligibility and Speech Transmission Index measurements.” AES 141st Convention,
2904 Los Angeles, 2016, Convention Paper 9629
- 2905 [49] Leembruggen, G, Verhave J. et all. “The effect on STI results of changes to the male
2906 test-signal spectrum.” Proc IOA Vol 38 Part 2 2016
- 2907 [50] Wijngaarden, S, Verhave,J. Designing an acoustic source of the STIPA signal. Proc IOA
2908 Vol 38 Part 2 2016
- 2909 [51] Mapp,P “Simulating Talker Directivity for Speech Intelligibility Measurements”, AES
2910 137th Convention, Los Angeles 2014
- 2911 [52] Mapp, P, “Some practical aspects of STI measurement & prediction”, AES 134th
2912 convention, Rome, 2013, convention paper 8864
- 2913 [53] IEC 60118-4 : Electroacoustics – Hearing aids – Part 4 : Induction-loop systems for hearing aid
2914 purposes – system performance requirements
2915
- 2916 [54] Mapp, P, “Some effects of speech signal characteristics on PA system performance and
2917 design”, AES 139th convention, New York, 2015, convention paper 9477.
- 2918 [55] ITU-T Recommendation P.51 (08/96) Objective measuring apparatus : Artificial Mouth