INTERNATIONAL STANDARD

ISO 9921

First edition
2003-10-15

Ergonomics — Assessment of speech communication

Ergonomie — Évaluation de la communication parlée

Reference number
ISO 9921:2003(E)

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

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

ISO 9921 was prepared by Technical Committee ISO/TC 159, *Ergonomics*, Subcommittee SC 5, *Ergonomics of the physical environment*.

This first edition of ISO 9921 cancels and replaces ISO 9921-1:1996.

Introduction

The aim of standardization in the field of the ergonomic assessment of speech-communication is to recommend the levels of speech-communication quality required for conveying comprehensive messages in different applications. The quality of speech communication is assessed for the following cases:

— warning of hazard;

— warning of danger;

— information messages for work places, public areas, meeting rooms, and auditoria.

For some applications, direct communication between humans is considered while, in others, the use of electro-acoustic systems (e.g. PA systems) or personal communication equipment (e.g. telephone, intercom) will be the most convenient means of informing and instructing or exchanging information.

The use of auditory warning symbols other than speech is not included in this International Standard but is covered by ISO 7731.

Acoustical danger and warning signals are in general omni-directional and therefore may be universal in many situations. Auditory warnings are of great benefit in situations where smoke, darkness or other obstructions interfere with visual warnings.

It is essential that, in the case of verbal messages, a sufficient level of intelligibility is achieved, in the coverage area. If this cannot be achieved, non-voice warning signals (see ISO 7731, IEC 60849 and [4] in the Bibliography) or visual warning signals (see ISO 11429) may be preferable.

If acoustical signals are too loud, hearing damage or environmental problems may occur (e.g. noise nuisance to dwellings near railway platforms, road traffic, airports, etc.). Good design can minimize these negative aspects. In addition, prediction methods with sufficient accuracy are useful for consultants, suppliers and end-users and may thus reduce costs of necessary adjustments after installation of a system.

The communications might be directly between humans, through public address or intercom systems or by pre-recorded messages. In general, text-to-speech systems are not recommended because of the low intelligibility of these systems.

It is recognized that, in a general-purpose document, simple to apply and easily available tools for prediction and assessment should be described, as well as more sophisticated advanced technological methodologies.

Ergonomics — Assessment of speech communication

1 Scope

This International Standard specifies the requirements for the performance of speech communication for verbal alert and danger signals, information messages, and speech communication in general. Methods to predict and to assess the subjective and objective performance in practical applications are described and examples are given.

In order to obtain optimal performance in a specific application, three stages can be considered:

a) specification of the application and definition of the corresponding performance criteria;

b) design of a communication system and prediction of the performance;

c) assessment of the performance for *in situ* conditions.

The use of auditory warning signals other than speech is not included in this International Standard but is covered by ISO 7731.

2 Normative references

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

ISO/TR 4870:1991, *Acoustics — The construction and calibration of speech intelligibility tests*

IEC 60268-16:1998, *Sound system equipment — Part 16: Objective rating of speech intelligibility by speech transmission index*

3 Terms and definitions

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

3.1
alarm
warning of existing or approaching danger

3.2
danger
risk of harm or damage

3.3
effective signal-to-noise ratio
measure to express the (combined) effect of various types of distortions on the intelligibility of a speech signal in terms of the effect of a masking noise resulting in a speech signal having the same intelligibility

3.4
emergency
imminent risk or serious threat to persons or property

3.5
Lombard effect
spontaneous increase of the vocal effort induced by the increase of the ambient noise level at the speaker's ear

3.6
non-native speaker
person speaking a language which is different from the language that was learned as the primary language during the childhood of the speaker

3.7
speech communication
conveying or exchanging information using speech, speaking, hearing modalities, and understanding

NOTE Speech communication may involve brief texts, sentences, groups of words and/or isolated words.

3.8
speech communicability
rating of the ease with which speech communication is performed

NOTE Speech communicability includes speech intelligibility, speech quality, vocal effort, and delays.

3.9
speech intelligibility
rating of the proportion of speech that is understood

NOTE Speech intelligibility is usually quantified as the percentage of a message understood correctly.

3.10
speech intelligibility index
SII
objective method for prediction of intelligibility based on the Articulation Index

NOTE See [1] in the Bibliography.

3.11
speech interference level
SIL
difference between A-weighted speech level and the arithmetic average of sound-pressure levels of ambient noise in four octave bands with central frequencies of 500 Hz, 1 000 Hz, 2 000 Hz and 4 000 Hz

3.12
speech quality
rating of sound quality of a speech signal

NOTE Speech quality characterizes the amount of audible distortion of a speech signal and is usually rated by a description.

3.13
speech transmission index
STI
objective method for prediction and measurement of speech intelligibility

3.14
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

3.15
warning
important notice concerning any change of status that demands attention or activity

4 Descriptions of speech communications

4.1 General

Speech communication requires three sequential components: speaker, transmission channel and listener(s). Based on this concept, three means of communication are identified.

a) **Direct communication**. This is typical for person-to-person communications, where both persons are in the same environment without making use of electro-acoustic means.

b) **Public address**. In general, an electro-acoustic system that is used to address a group of people in one or more environments.

c) **Personal communication systems**. These include the use of mobile telephones and handheld transceivers and the use of normal telephones, intercoms and hands-free telephones.

4.2 Speaker

Several speaker-related parameters define the contribution of the speaker to the performance of a communication. These parameters include vocal effort, speaking quality, gender, accents, non-native speech, speaking disorders, and distance from the listener or microphone.

Vocal effort is expressed by the equivalent A-weighted sound-pressure level at a distance of 1 m in front of the mouth. The ambient noise level at the speaker's position (causing the Lombard effect) and the wearing of a hearing protector influence the vocal effort. The relation between these parameters and the effect on the speech quality is described in Annex A.

The frequency spectrum of the speech is related to the gender of the speaker and the vocal effort. This may result, in combination with a specific type of noise, in a gender-related performance [see Annex B (B.3) and Annex C].

The effects of strong accents and non-native speakers and listeners reduce the performance of a communication; quantitative data are given in A.6.

4.3 Transmission channel

The transmission path between the speaker’s mouth and the listener’s ear is described by the distribution of the speech signal in a room or by an electro-acoustic system. It affects the deterioration of the speech signal. Important influences are ambient noise, reverberation, echoes, sound radiation, limitation in the frequency response, and non-linearities. In Annex D, an overview is given of the means of communication and related parameters.

4.4 Listener

For the listener, hearing aspects (directional hearing, masking, hearing disorders, reception threshold) and the use of hearing protection define the deterioration. In Annexes A, C, D and E, these listener-related parameters are considered, except for that of directional hearing, which is not considered in this International Standard.

5 Performance of speech communications

5.1 General

A correct recognition of each utterance is required for the understanding of spoken messages. In technical terms, this means that an intelligibility score of 100 % is required for sentences. A sentence intelligibility score

of 100% does not imply that each individual word is clearly understood and that the listening situation is comfortable and relaxed and there are many situations in which a better performance is required. In alert situations under adverse conditions, it is sufficient to fully understand a short message, even if correct understanding requires some effort from the listener. In a meeting room, an auditorium, or at work places where speech communication is a part of the task and where people are normally present for a longer period of time, a more relaxed speaking condition and a good listening condition are required. For the speaker, this is reflected by the low vocal effort required to be understood (see Table A.1). For the listener, the listening effort may be primarily related to the speech intelligibility and speech quality at the listening position (see Table F.1). The range of the classification scales and the number of the intervals is large enough to discriminate between conditions required for different applications (see Table F.1 and Figure F.1).

The quality of speech communication is expressed in terms of intelligibility and vocal effort. In this International Standard, various application and environmental conditions are identified. For each of them, minimal performance criteria are recommended, covering the range from short alert and warning messages under adverse conditions to relaxed communications in a meeting room or auditorium. People with a slight hearing disorder (in general the elderly) or non-native listeners require a higher signal-to-noise ratio (approximately 3 dB).

The different fields of application are described in 5.2 to 5.5 and summarized in 5.6.

5.2 Alert and warning situations

In general, clearly pronounced short messages are required for alert and warning situations, in order to provide guidance for safe evacuation or clearance with minimal risk of panic. Hence, simple sentences should be understood correctly even under adverse conditions, high environmental-noise levels, the speaker shouting, etc.

As seen in Annex F (Figure F.1), the qualification “poor” is just adequate for alert and warning situations. This criterion represents a mean value for listeners with a normal hearing (50 % coverage). For 96 % coverage of the population, an improvement is required that can be expressed by an increase of the signal-to-noise ratio by 3 dB. Therefore, the recommended criterion should be at least “poor”.

With the use of a public-address system, poor-to-fair intelligibility may be recommended in adverse conditions. However, distortions introduced by the electro-acoustic systems and/or the environment (band-pass limiting, non-linear distortion, noise, reverberation and echoes) may also affect the speech intelligibility. This generally results in the need for a better signal-to-noise ratio.

In order to include effects of all the distortions and environmental conditions on the overall intelligibility rating, it is necessary to assess the system performance under representative (*in situ*) conditions.

5.3 Person-to-person communications

For communication in work situations, offices, meeting rooms, auditoria, and in critical situations (ambulance personnel, firemen, etc), a different level of intelligibility is required depending on the purpose of the communication. In critical situations, generally short messages are exchanged which also include a certain number of known critical words. For such communication conditions, at least a “fair” intelligibility is recommended at an increased vocal effort (loud).

In situations of a relaxed type of communication, for example, occurring in offices, during meetings, lectures and performances, which take place over a longer period of time, a good level of intelligibility is recommended allowing for a normal vocal effort.

5.4 Public address in public areas

In public areas, general announcements are made with a short to medium duration at a normal vocal effort. The content of the announcements may consist of numbers, names of destinations, names of persons, etc. For these purposes, a fair-to-good intelligibility is recommended. Typical areas are shopping centres, railway stations, within transportation means, and stadiums.

5.5 Personal communication systems

Communication systems are generally limited in bandwidth and may be used in noisy environments. Examples are the outdoor use of mobile telephones and handheld transceivers, and the indoor use of normal telephones and hands-free telephones. Depending on the type of the communication (complexity of the messages) and intensity of the use, a fair-to-good intelligibility is recommended at a normal vocal effort.

5.6 Summary of recommended minimum performance

The recommended minimal performance rating is summarized in Table 1. However, in certain circumstances, it is advisable to have a higher rating.

Table 1 — Recommended minimal performance ratings for intelligibility and vocal effort in four applications (for examples of rating see Table A.1)

Application	Minimum intelligibility rating	Maximum vocal effort	Description
Alert and warning situations (correct understanding of simple sentences)	Poor	Loud	5.2
Alert and warning situations (correct understanding of critical words)	Fair	Loud	5.2
Person-to-person communications (critical)	Fair	Loud	5.3
Person-to-person communications (prolonged normal communication)	Good	Normal	5.3
Public address in public areas	Fair	Normal	5.4
Personal communication systems	Fair	Normal	5.5

6 Assessment and prediction

6.1 General

Assessment of speech communication includes speech quality, speech intelligibility, speech communicability and vocal effort. For the purpose of this International Standard, only speech intelligibility and vocal effort are considered. The intelligibility can be determined by subjective methods (making use of speakers and listeners) and by objective methods (making use of physical properties and the physical description of the speaking and listening process).

6.2 Subjective assessment methods

Subjective intelligibility tests require trained speakers to read lists of test words and listeners who write down what they thought they heard. Normally lists are 50 words long and the result is scored out of 100. Test words should be embedded in a carrier phrase in order

a) to let the speaker control his vocal effort,

b) to account for temporal distortion during pronunciation of the test word, and

c) to get the attention of the listener at each utterance.

Test words may be meaningful words or nonsensical words, and phonetically balanced (phoneme distribution representative for the language) or equally balanced (phoneme distribution equal for all phonemes). The type of words used in the test defines the relation with other types of tests such as STI (Speech Transmission Index) or SIL (Speech Interference Level). An informative description of subjective intelligibility tests is given in Annex B and ISO/TR 4870.

6.3 Objective assessment and prediction methods

There are several objective methods to predict speech intelligibility. Depending on the method, either results of objective measurements or specifications of a system and space are used to calculate an index to predict intelligibility. These may include

— spectrum of the speech signal,

— spectrum of environmental noise,

— spatial distribution of these sound fields,

— reverberation,

— associated selection of listener positions, and

— evaluation of the resulting intelligibility score.

Commonly used methods are the Speech Interference Level (SIL), the Speech Transmission Index (STI), and the Speech Intelligibility Index (SII). A normative description of the SIL is given in Annex E, a normative description of STI is given in IEC 60268-16 and an informative description in Annex C. The SII is described in ANSI S3.5 [1].

Annex A
(normative)

Speaker and listener characteristics

A.1 Vocal effort

The level of the speech signal depends on the vocal effort of the speaker. The vocal effort is expressed by the equivalent continuous A-weighted sound-pressure level of speech measured at a distance of 1 m in front of the mouth. The relation between vocal effort and the corresponding level is given in Table A.1 for a typical male speaker.

Table A.1 — Vocal effort of a male speaker and related A-weighted speech level (dB re 20 µPa) at 1 m in front of the mouth

Vocal effort	$L_{S, A, 1\,m}$ dB
Very loud	78
Loud	72
Raised	66
Normal	60
Relaxed	54

A.2 Effect of ambient noise on vocal effort

Ambient noise above a certain level influences the vocal effort (this in known as the Lombard effect). In Figure A.1 the relation between speech level and ambient-noise level is given. The hatched area indicates the variability of the Lombard effect among speakers.

A.3 Decrease of speech quality with loud speech

The quality of loud speech, above the level of $L_{S, A, 1\,m}$ = 75 dB, is substantially reduced, making it more difficult to understand in comparison with speech produced at a lower vocal effort. This is taken into account by reduction of the speech level in calculations: ($L_{S, A\,1\,m}$) shall be reduced by ΔL = 0,4 ($L_{S, A, 1\,m}$ – 75) dB for $L_{S, A, 1\,m}$ > 75 dB.

NOTE Certain symbols used in this annex are defined in Annex G.

A.4 Effect of hearing protection on vocal effort

A speaker wearing hearing protectors will reduce his vocal effort by about 3 dB compared to the unprotected situation, if the ambient noise level $L_{N, A}$ exceeds 75 dB.

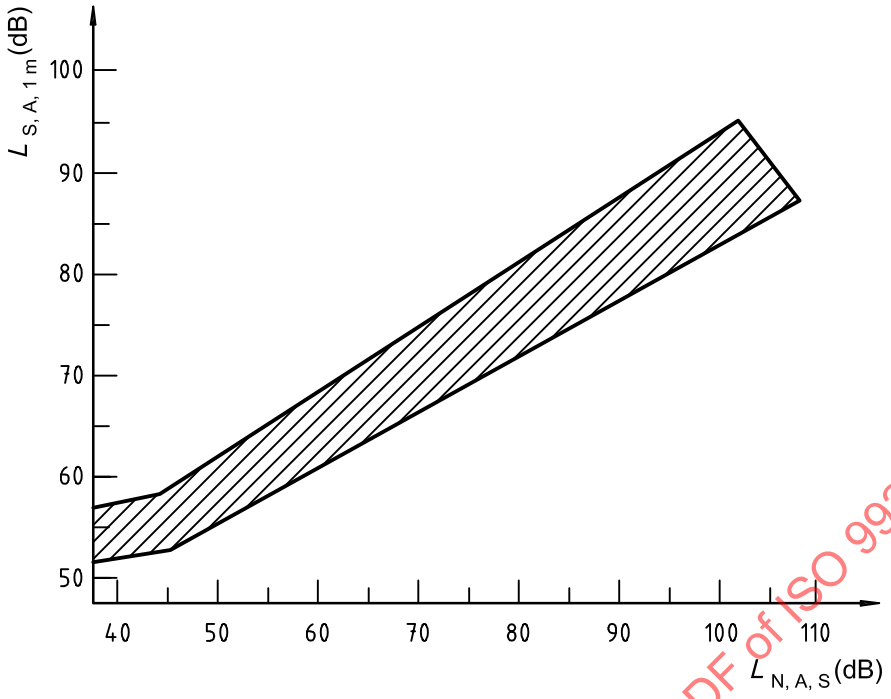


Figure A.1 — Relation between the range of vocal effort (equivalent continuous speech sound level) and the ambient-noise level at the speaker's position

A.5 Effect of distance between speaker and listener

From the speech level at the speaker position ($L_{S, A, 1 m}$), the speech level at the listener position ($L_{S, A, L}$) may be approximated using the equation:

$$L_{S, A, L} = L_{S, A, 1 m} - 20 \lg \frac{r}{r_0}$$

where

r is the distance in metres between the speaker and listener;

r_0 = 1 m

Hence, the decrease in speech level is assumed to be 6 dB for each doubling of the distance. This relation is valid for indoor and outdoor conditions up to about 2 m. For conditions with a reverberation time smaller than 2 s at 500 Hz, a maximum distance of 8 m is valid.

A.6 Effect of non-native speakers and listeners

A reduced intelligibility is observed with non-native but fluent speakers and listeners of a second language. For non-native speakers or listeners, or for both in combination, a 4 dB to 5 dB improvement in the signal-to-noise ratio is required for a similar intelligibility as is obtained with native speakers and/or listeners [15]. This 4 dB signal-to-noise ratio improvement corresponds with an improvement of the STI of 0,13 and of the SIL of 4 dB.

Annex B
(informative)

Subjective speech-intelligibility tests

B.1 Basic conditions for testing

The speaking ability of speakers and the hearing capacity of listeners shall be sufficient to provide an efficient direct communication, communication by means of a public-address system or personal communication device (see Figure D.1).

The speakers and the listeners shall be familiar with the language used, as far as to pronounce and understand a verbal message. It is best to use native speakers of the language.

Listeners should be protected from risks to health and safety. This means that a safe speech level should not be exceeded. The recommended maximum speech level is 80 dB A-weighted for an exposure of maximum 8 h per working day.

B.2 Test material

B.2.1 General

The speech-intelligibility test should be such as to obtain valid, reliable results allowing for an analysis of errors in listeners' responses. The test material must use samples of speech sounds, which are typical for the communication system being tested, and representative of the type of message transmitted through the system. Economy of testing should be considered, i.e., possible automation to simplify test administration.

A number of methods have been proposed for the measurement of speech intelligibility (see F.4). In this document, three types of intelligibility tests are included:

— an open-set nonsensical CVC_{EQB} word test;

— an open-set meaningful PB-word test;

— a sentence test.

B.2.2 Open-set lists

Open-set lists of test items are made using items drawn randomly from a total set of test items. In the case of nonsensical CVC word tests, a test item is generated randomly from a set of initial consonants, vowels and final consonants. The CVC_{EQB} nonsensical words are balanced to represent all phonemes of the test language in equal proportion. In the CVC test generation, language-dependent restrictions may apply in the conjunction of specific phonemes.

The meaningful phonetically balanced word test (PB-words) is constructed as a set of monosyllabic words. For phonetically balanced tests, different phonemes occur in the test in the proportion in which they occur in natural language.

The nonsensical CVC-word test and the meaningful phonetically balanced word test (PB-words) typically comprises 50 words per list. The total number of required test items is at least 1 000 words, to avoid listeners adapting to frequently used lists (see ISO/TR 4870). The CVC_{EQB} test requires about a 6 dB higher signal-to-noise ratio to obtain a similar percentage correct score as does the meaningful PB-word test (see Figure F.1).

The lists spoken by speakers are presented to a panel of listeners. Since open format is used, the listeners typically respond by writing down the response on a response sheet (or using a silent keyboard). The intelligibility score is the percentage of words correctly identified in the test. With nonsensical CVC-words, separate scores for the initial consonant, the vowel, and the final consonant can also be determined, this then allows for the construction of a confusion matrix. For details see Annex F, ISO/TR 4870 and [13].

B.2.3 Sentence tests

Usually, sentence tests are not recommended for evaluating transmission systems because the listener's knowledge of grammar, meaning and syntax of the sentence influences the results. Another difficulty is creating a large number of sentences that are phonetically representative of speech and with a well-defined complexity. However, for specific uses, the SRT method [10] can be used which determines the noise level which provides 50 % sentence intelligibility. Depending on the speech material, this corresponds to a signal-to-noise ratio of − 4 dB to − 6 dB (see Figure F.1). Hence, conversion to other conditions is possible.

B.3 Speakers and listeners

Speakers and listeners should be selected to be representative of the user population of a system under test. In selecting the speakers and listeners, age, gender, education, relevant experience and linguistic background should be taken into account. The group of speakers and listeners, the size and training shall be selected in accordance with ISO/TR 4870.

ISO/TR 4870 recommends the following:

— at least one male and one female speaker typical of a given nationality and language;

— five well-motivated listeners for small closed-set test formats, and ten for large open-format tests;

— normal experience in use and spelling of the language to be used, good hearing, that is a pure tone audiogram not exceeding a hearing level of 10 dB at any test frequency up to 4 000 Hz, and 15 dB at any frequency up to 6 000 Hz;

— training time between 5 min and 24 h depending on the test format, see ISO/TR 4870:1991, 3.10.

The speech samples may be spoken directly, or prerecorded. Recordings of test material should be made according to ISO/TR 4870. The electrical parameters of a recording system such as frequency response, non-linear distortions, and the signal-to-noise ratio should be good enough to be considered ideal in comparison with the respective parameters of the system under test. For recording, the speaker should be placed in a quiet and sound absorbing environment. The distance of the speaker's mouth to the microphone should be reported.

The speaker should be familiar with the grammar of the text material. The speaker should be given visual feedback to control the level, and timing, of spoken items. The same kind of feedback should be used in the case of live and recorded speech. Speakers must be trained until they attain a stable sound-pressure level of pronounced speech (65 dB ± 3 dB) on the average, at a distance of 1 m in front of the speaker's lips. For details see ISO/TR 4870.

Listeners should be familiar with the communication system under testing. They must also become familiar with the test procedure. The listeners should be given written instructions.

The listeners should be trained until they become familiar with the test procedure and the test words. The training should include hearing all the words from a list under quiet conditions, using an undistorted communication system. The training should be conducted until listeners achieve 100 %, or nearly 100 % performance in ideal conditions. Listeners should be trained by hearing the voices of all the speakers used. There should be no visual contact between the speaker and the listener in order to prevent the listener from lip reading.

B.4 Administration of the intelligibility test

Usually, intelligibility testing involves a number of test conditions because several communication systems or several states of a communication system (e.g. various speech-to-noise ratios) are to be measured, resulting in different intelligibility ratings. However, if only one test condition is to be assessed, the use of reference conditions is recommended.

If several conditions are measured, they should be presented using a balanced experimental design that will neutralize the influence of various random factors, that are not fully controlled in measurements such as the effect of learning by the listeners. Other information relevant to the listener's performance should be collected. This includes information about the confidence of the listener's responses as well as the listener's opinions about the measured system. All variables important for the conditions of testing should be chosen in advance or measured.

In the case of live speech, the speaking level, rate of speech and vocal effort should be controlled and reported. The speech and noise level both on the speaker's side and at the listener's ears should be measured and reported. In the case of prerecorded speech, the speech and noise level at the listener's ears should be measured and reported.

If the communication device creates constraints of the mouth and lips (e.g. special helmet with a microphone), it should be reported and described.

B.5 Statistical analysis and documenting results

For a simple test, the mean score (percent correct responses) and the corresponding standard deviation should be calculated, thus allowing for prediction of the 96 % confidence interval. Depending on the construction of the test (i.e., number of speakers, number of listeners, number of conditions, number of replicas), statistical analysis such as an analysis of variance (ANOVA) can be applied.

Annex C
(informative)

Speech transmission index, STI

The STI-method [7], [11], [12], [14] assumes that the intelligibility of a transmitted speech signal is related to the preservation of the original spectral differences between speech sounds. These spectral differences may be reduced by band-pass limiting, masking noise, temporal distortion (echoes, reverberation, and automatic gain control), and non-linear distortion (system overload, quantization noise). The reduction of these spectral differences can be quantified by the effective signal-to-noise ratio obtained for a number of frequency bands. Also human-related hearing aspects such as masking, the reception threshold, hearing disorders, and non-native speakers and listeners may reduce the effective signal-to-noise ratio. The method is based on the calculation of the effective signal-to-noise ratio in seven relevant frequency bands (octave bands, centre frequencies ranging from 125 Hz to 8 kHz). Weighted contributions of the quantified information transfer function in seven octave bands results in a single index, the STI_r.

Originally the STI-method was developed for measurements. For this purpose a specific test signal was designed, which, after transmission through the channel under test, was analysed in order to determine the effective signal-to-noise ratios in different frequency bands and to calculate the STI_r. The test signal was so designed that, after analysis, information could be obtained on most types of distortion mentioned above. In particular, temporal distortion and non-linear distortion require a specific test signal and analysis.

It is possible to predict the STI_r value for transmission channels with band-pass limiting and noise, based on the signal-to-noise ratio in the seven octave bands. However, the prediction of the effect of temporal distortion on the STI_r is limited to single echoes and reverberation. For reverberation, a simple algorithm is used and only continuous exponential decay curves can be accounted for. This excludes prediction, for acoustically coupled enclosures and very complex environments[1)]. The effect of non-linear distortion on the STI_r cannot be predicted by a simple algorithm.

The measurement of the STI is described in IEC 60268-16.

Prediction of the STI-value can be performed in the following nine steps.

Step 1: Determine the speech spectrum in seven octave bands at the listener's ear.

This includes the determination of the vocal effort (including the Lombard effect and the effect of wearing a hearing protector, see Annex A), the male/female speech spectrum, the distance between speaker and listener, and the effect of band-pass limiting.

Step 2: Determine the noise spectrum in seven octave bands at the listener's ear.

Step 3: For each band, determine the signal-to-noise ratio, based on the speech and noise spectra and convert these signal-to-noise ratios to the corresponding m-values.

$$m = 10 \exp \frac{10\,S}{S+N}$$

where

S is the speech level, in decibels;

N is the noise level, in decibels.

1) With prediction algorithms such as ray-tracing, more complex environments can be included.

If no temporal distortion has to be accounted for, then proceed with Step 6.

Step 4: Determine the early decay reverberation time for the listening environment, and calculate the (octave-band specific) modulation transfer function using the formula given in IEC 60268-16:1998, A.2.1 and Annex D. This will result in 14 m-values per octave band.

Step 5: For each octave band, correct the seven m-values obtained in Step 3 with the modulation transfer functions obtained in Step 4. This is performed by multiplication of the modulation transfer function with the octave-band-specific m-value from Step 3.

Step 6: Correct the m-values for auditory effects (masking, reception threshold).

Step 7: Determine effective signal-to-noise ratios within range limits (– 15 dB to + 15 dB).

Step 8: Determine the modulation transfer indices (MTI) from these effective signal-to-noise ratios.

Step 9: Calculate STI_r from the MTIs.

Annex D
(informative)

Overview of the means of communication and related parameters

D.1 General

A modular overview of the three major means of communication between persons is given in Figure D.1. Each module is discussed and reference is made to the appropriate section in this International Standard. It is advised to make an inventory for each module in the communication channel and to identify the relevant issues that determine the performance of the complete system. Three systems are described in D.2, D.3 and D.4, by their modules or components, organized by

— input: speaker,

— channel: environment (room), transmission system,

— output: listener.

D.2 Direct communication without the use of electro-acoustic means

For person-to-person communication, the major parameters are: speaker, listener, and the acoustic environment in which the speaker and listener are positioned. The following parameters are identified.

a) Speaker:

— speaker characteristics: gender, age, speaking disorders;

— language (native, non-native, see A.6);

— type of speech (complexity, see Annex F, Figure F.1);

— vocal effort, including Lombard effect and loud speech (see A.1, A.2 and A.3);

— speaking direction (directivity, restricted in this International Standard to face-to-face);

— wearing a hearing protector (see A.4).

b) Environment:

— ambient noise (level, spectrum, see Annexes A, C and E);

— temporal distortion (reverberation, echoes, see Annex C);

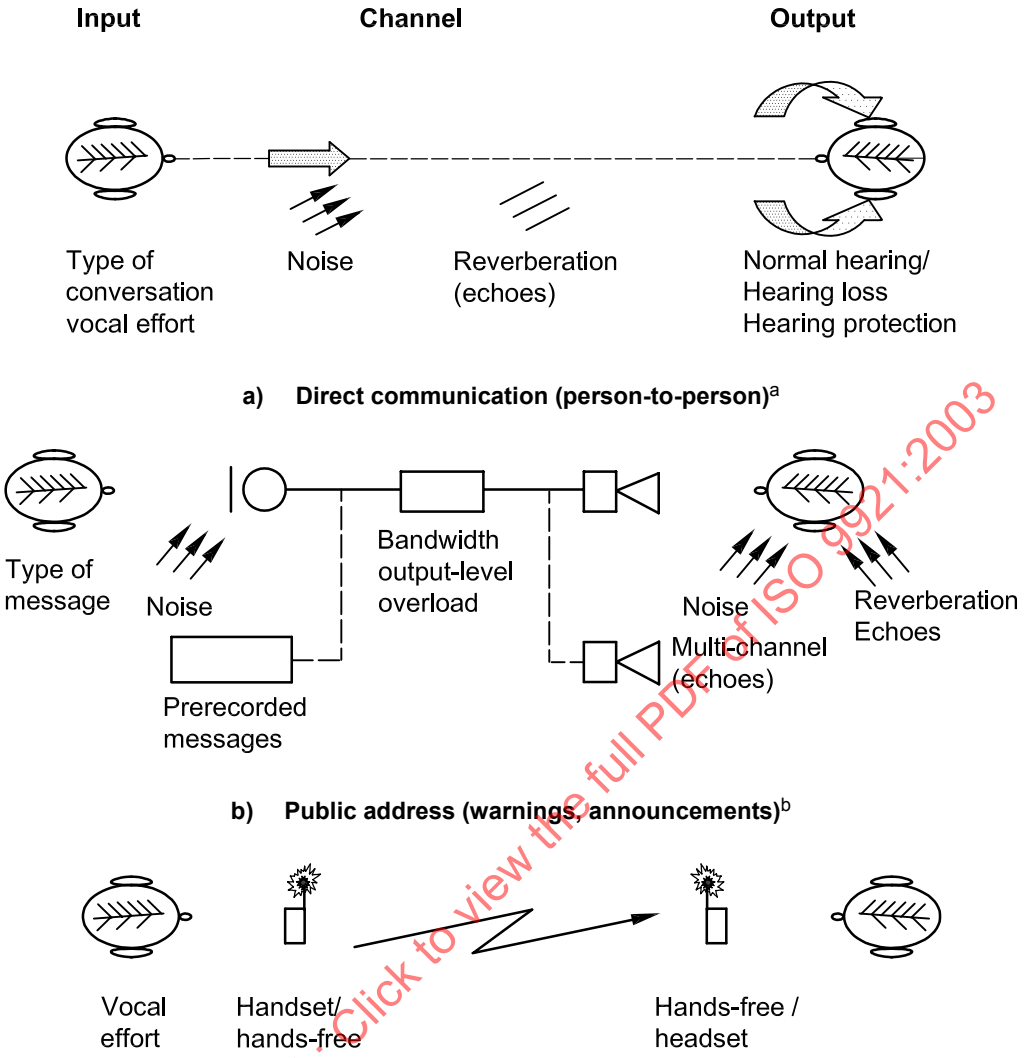
— distance between speaker and listener (see A.5).

c) Listener:

— listener characteristics: gender, age, hearing disorders;

— language (native, non-native, see A.6);

— hearing protection (earmuffs, earplugs, headsets, see A.4).



a Workshop, office, conference room, auditorium.

b Alert/warning in public areas, public address, offices, shops, railway station, inside transportation means.

c Telephone, intercom, mobile telephone, command and control sites.

Figure D.1 — Overview of the three means of communication: direct, public address, and via a personal communication system

D.3 Communication via a public-address system

For public address (PA), electro-acoustic means are used, such as microphones, amplifiers and loudspeakers. The specifications of the following modules have to be taken into account when estimating the performance.

a) Speaker:

— see D.2;

— prerecorded messages: (initial intelligibility of reproduced speech).

b) Microphone:

— speaking distance and position (see D.2 and A.5);

— frequency response (see Annex C);

— noise suppression (determine gain in signal-to-noise ratio improvement);

— noise spectrum in speaking area (see Annexes C and E).

c) (Power) amplifier:

— frequency response (see Annex C);

— non-linear distortion (overload, see Annex C);

— adjustment for microphone and loudspeaker sensitivities (determine output level).

d) Loudspeaker (cluster):

— acoustic output level;

— frequency response (see Annex C);

— directivity (area to be covered);

— multi-channel output (delay between clusters and related echoes, see Annex E).

e) Environment:

— acoustic transfer to listener (noise, reverberation, echoes, see Annex C).

f) Listener:

— see D.2.

D.4 Communication via a personal communication system

Personal communication systems use wired or radio-based technology for communication between the users. In principle, the speaker- and listener-related aspects are similar to direct communications and to public address. Personal communication systems may use techniques for hands-free operation and this may increase the influence of the acoustical environment. The characteristics of the following modules have to be specified:

a) Speaker:

— see D.2.

b) Microphone:

— speaking distance and position (see also D.2);

— frequency response (see Annex C);

— noise suppression (determine gain in signal-to-noise ratio improvement);

— noise spectrum in speaking area (see Annexes C and E).

c) Transmission path:

— frequency response (see Annex C);

— automatic gain control (dynamic range, for fast attack and decay time use MTF, see Annex C);

— non-linear distortion (overload, specific speech coding, see Annex C);

— noise introduced by transmission (see Annexes C and E);

— acoustic output level (noise, reverberation, listener distance).

d) Loudspeaker:

— output level;

— frequency response (see Annex C);

— directivity.

e) Listener:

— signal level (determine signal-to-noise ratio);

— ambient noise (determine signal-to-noise ratio);

— temporal distortion (reverberation, echoes, see Annex C).

Annex E
(normative)

Speech interference level, SIL

E.1 General

The speech interference level offers a simple method to predict or to assess the speech intelligibility in cases of direct communication in a noisy environment [3], [8]. It takes into account a simple average of the noise spectrum (no frequency-dependent contributions), the vocal effort of the speaker and the distance between the speaker and listener. The method should only be used in situations where other assessment and prediction methods of speech intelligibility cannot be applied.

E.2 Ambient noise

For the determination of the speech interference level of noise (L_{SIL}), the sound-pressure levels in the octave bands 500 Hz, 1 000 Hz, 2 000 Hz and 4 000 Hz shall be determined at the listener's position in noise conditions which are typical for the communication period. Under normal conditions, an equivalent sound-pressure level shall be determined and, for safety reasons, this should be the maximum value of a sound-pressure level determined with a time-weighting "slow" of the sound-level meter.

The speech interference level of noise (L_{SIL}) is calculated as the arithmetic mean of the sound-pressure levels of the ambient noise in four octave bands with the central frequencies 500 Hz, 1 000 Hz, 2 000 Hz and 4 000 Hz. The following equation defines this relation.

$$L_{SIL} = \frac{1}{4} \sum L_{N,\,oct,\,i}$$

for i = 1,4

NOTE Certain symbols used in this annex are defined in Annex G.

E.3 Speech level

The level of the speech signal is determined by the vocal effort of the speaker (A.1), taking into account: ambient-noise level (see A.2), the effect of loud speech (see A.3), the use of a hearing protector (see A.4), distance (A.5), and non-native speakers (A.6). The speaker's vocal effort is described by the equivalent continuous A-weighted sound-pressure level of the speech at a distance of 1 m in front of the speaker's mouth as given in Table A.1.

E.4 Parameter defining the intelligibility

The SIL is given by the difference between the speech level $L_{S,\,A,\,L}$ and the speech interference level of noise L_{SIL}, both determined at the listener's position. Fair speech-communication intelligibility is ensured if the difference in levels, SIL = $L_{S,\,A,\,L} - L_{SIL}$, is $\geqslant$ 10 dB at the listener's position. For the rating of SIL, see Table F.1.

Annex F
(informative)

Intelligibility ratings for speech communications

F.1 General

The performance of a speech-communication channel can be determined by using subjective tests (based on speakers and listeners), or by objective methods (based on physical properties of the transmission path and the physical description of the speaking and listening process).

F.2 Subjective test methods

A number of subjective intelligibility tests have been developed for the evaluation of speech-communication systems (see [2] [3] [5] [6] [8] [9] [10]).

Subjective intelligibility tests can be categorized by the speech items and the response procedure used in the test. The smallest items tested are at the segmental level, i.e. phonemes. Other test items are CV (consonant-vowel), VC (vowel-consonant), and CVC (consonant-vowel-consonant) combinations, nonsensical words, meaningful words, sentences, and the validation of short conversations (see Annex B).

The test can be designed in a closed-set or in an open-set response procedure. In the closed-set format, the listener must select the most likely candidate from a number of alternatives that are presented in the test. This is normally applied with rhyme tests (e.g. the MRT, Modified Rhyme Test [6]). An open-set response allows the listener to respond freely with the items that he or she is convinced have been presented.

Besides intelligibility, speech quality and vocal effort may be determined by making use of questionnaires or scaling methods. Subjective qualities to be assessed include overall impression, naturalness, noisiness, clarity, etc. Speech-quality assessment is normally used for high-intelligibility communication channels for which most intelligibility tests cannot be applied because of ceiling effects.

F.3 Objective test methods

Objective ratings of intelligibility are generally based on relevant physical properties of the transmission path between the speaker and listener. From measured degradations, a prediction can be made of the related intelligibility. Objective ratings take into account speaking and hearing aspects such as the Lombard effect, vocal effort, masking, reception threshold, and hearing protection (covered by STI, SII and partly by SIL). In addition, temporal distortions (reverberation, echoes and automatic gain control) and non-linearity are covered by STI.

Direct objective measurement using specific test signals is possible with the STI method which allows determination of the relevant physical properties [MTF, (Modulation Transfer Function), non-linearity] of the transmission channel. The algorithm for objective measurements and prediction are similar, but the effect of non-linear distortions is only provided by *in situ* measurements.

For any noise spectrum, the quantified STI and SII have an accuracy of 1 dB to 2 dB. The SIL has an accuracy in the order of 2 dB to 3 dB. The SIL may give systematic errors, especially for noise signals with a non-contiguous frequency spectrum.

F.4 Relation between various intelligibility ratings

The relation between intelligibility rating and some subjective and objective intelligibility ratings are given in Table F.1. Figure F.1 indicates the effective range of each test method.

The CVC_{EQB}-nonsensical words[2)] discriminate over a wide range, while meaningful test words[3)] have a slightly smaller range [2]. Sentences (and related digits and the alphabet) show saturation at poor intelligibility levels, which implies that testing with these types of speech material cannot be used to assess conditions with a qualification better than “poor”. These ceiling effects may be due to

a) the redundancy of the words in a sentence,

b) the limited number of test words for digits and the alphabet, and

c) conditions in which correct recognition of words is mainly determined by recognition of vowels and subsequently by consonants.

Intelligibility ratings and scores are obtained for listeners with normal hearing. The effect of the vocal effort of the speaker, accents and non-native speakers (see Annex A) can also be accounted for, but not for speaker deficiencies (pathological defects).

Table F.1 — Intelligibility rating and relations between various intelligibility indices

Intelligibility rating [a]	Sentence score [b] %	Meaningful PB-word [c] score %	CVC_{EQB} -non-sensical word score %	STI [d]	SIL [d] dB	SII [e]
Excellent	100	> 98	> 81	> 0,75	21	—
Good	100	93 to 98	70 to 81	0,60 to 0,75	15 to 21	> 0,75
Fair	100	80 to 93	53 to 70	0,45 to 0,60	10 to 15	—
Poor	70 to 100	60 to 80	31 to 53	0,30 to 0,45	3 to 10	< 0,45
Bad	< 70	< 60	< 31	< 0,30	< 3	—

[a] Qualification according a five-point scale, see [7] [8] [14].

[b] The sentence score refers to simple sentences [10], CVC_{EQB}-nonsensical words with an equally balanced phoneme distribution [12, 13], and the PB-word score (related to the phonetically balanced Harvard list) [2].

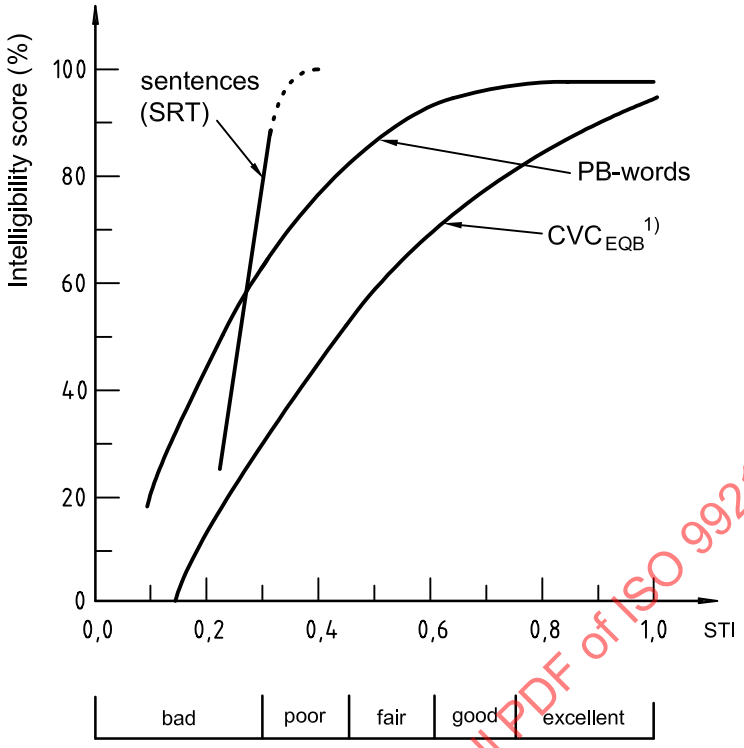
[c] According to Anderson and Kalb (1987) [2].

[d] The SIL (Annex E) and SII (Annex C) only refer to noise conditions.

[e] The SII procedure does not provide qualification intervals. The ANSI standard [1] does provide two benchmarks: good > 0,75, poor < 0,45.

2) EQB (equally balanced phoneme distribution), is frequently used for the compilation of word lists. The advantage is that the phoneme error rate of each phoneme is determined with equal accuracy and that a balanced nonsensical confusion matrix is obtained [13].

3) Meaningful test words are naturally phonetically balanced (PB), hence the frequency distribution of the phonemes is representative for the language used. Word scores are generally higher than those obtained with nonsensical words, as word familiarity increases the score.



1) See F.4.

Figure F.1 — Relation between qualification and some subjective and objective intelligibility ratings
see references [2] [10] [11] [13] [14]

Annex G
(normative)

Definition of symbols

$L_{S, A, 1\,m}$	Equivalent continuous A-weighted sound-pressure level of the speech at a distance of 1 m in front of the mouth.
$L_{S, A, L}$	Equivalent continuous A-weighted sound-pressure level of the speech at the listener's ear.
$L_{N, oct, i}$	Octave pressure level of the ambient noise at the listener's ear in octave band "i".
L_{SIL}	Speech interference level of the noise at the listener's ear.