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].