TECHNICAL REPORT

ISO TR 4870

First edition
1991-12-15

Acoustics — The construction and calibration of speech intelligibility tests

Acoustique — Élaboration et étalonnage des tests d'intelligibilité de parole

Reference number
ISO/TR 4870:1991(E)

AVA Documents
Document Sharing Hub

Full document details: ISO_TR 4870_1991

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.

The main task of technical committees is to prepare International Standards, but in exceptional circumstances a technical committee may propose the publication of a Technical Report of one of the following types:

— type 1, when the required support cannot be obtained for the publication of an International Standard, despite repeated efforts;

— type 2, when the subject is still under technical development or where for any other reason there is the future but not immediate possibility of an agreement on an International Standard;

— type 3, when a technical committee has collected data of a different kind from that which is normally published as an International Standard ("state of the art", for example).

Technical Reports of types 1 and 2 are subject to review within three years of publication, to decide whether they can be transformed into International Standards. Technical Reports of type 3 do not necessarily have to be reviewed until the data they provide are considered to be no longer valid or useful.

ISO/TR 4870, which is a Technical Report of type 3, was prepared by Technical Committee ISO/TC 43, *Acoustics*.

It contains data which are valuable in speech intelligibility testing, but it is not expected to become an International Standard.

Annexes A and B of this Technical Report are for information only.

© ISO 1991
All rights reserved. No part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from the publisher.

International Organization for Standardization
Case Postale 56 • CH-1211 Genève 20 • Switzerland
Printed in Switzerland

Documents
Document Sharing Hub
Full document details: ISO_TR 4870_1991

Introduction

A variety of perceptual tests have been developed in the past for the assessment of the intelligibility of speech communications as affected by spectral, amplitude, and temporal distortions of the speech signal and by noise that arises from or in the acoustical, electrical (if any), and ear receptor path used for transmitting and transducing speech from a talker to a listener. The principal tests developed for this purpose have been called speech intelligibility tests and will be later defined in detail.

Beyond factors related to the talkers, listeners, and a given communication system, there are two factors common to all speech intelligibility tests that have a significant influence on the scores obtained from a test evaluation. These two common factors are: (1) the speech material employed in the tests, and (2) for a given type of speech material, the total number of alternative members of that material the listeners expect to be presented during a test. Without some knowledge about the contribution of these two factors to the scores obtained on a given speech intelligibility test meaningful comparisons and interrelations cannot be made with respect to the test scores obtained in different investigations of speech intelligibility.

It is the purpose of this document to standardize fundamental methods for the construction and calibration of speech intelligibility tests in ways that reveal the contributions to the test scores of the two common factors mentioned above. Also, illustrative examples of recommended types of speech test materials possibly suited for such purposes as speech audiometry, the evaluation of room acoustics, or an electro-acoustic transmission system are given.

The communication of thoughts and concepts through spoken languages is a broad and complex operation that is influenced by many other factors than the intelligibility based on the perception of acoustical features of the speech signal. However, the basic feature of speech communication is an acoustical signal, and the greater the understanding of the speech derivable from perceptions of certain acoustical relations within the signals, the more effective and general can be the communication process. Intelligibility tests are aimed at the evaluation of the ability of a communication system or component, including the auditory mechanism of a listener, to effectively transmit basic acoustical information that is instrumental to the correct perception of speech.

AVA Documents
Document Sharing Hub

Full document details: ISO_TR 4870_1991

This page intentionally left blank

Documents
Document Sharing Hub

Full document details: ISO_TR 4870_1991

Acoustics — The construction and calibration of speech intelligibility tests

1. SCOPE

1.1 The present document is concerned with the description of:

(a) methods for the construction of speech tests for the measurement of the intelligibility of speech transmitted by an analog or combination analog and digital communication system;

(b) a reference communication system and test conditions necessary to the development, calibration, and interpretation of the results of standardized intelligibility tests.

1.2 The description of specific speech tests and related test procedures and measurements that are most appropriate for a given test application are beyond the scope of this document.

2. DEFINITIONS FOR PRESENT PURPOSES

2.1 Speech Sound

A speech sound is the smallest identified unit of speech. These units of speech can be categorized into two general classes known as vowels (V) and consonants (C). Consonants that are, on occasion, used as vowels in forming syllables of speech will, for purposes of this document on those occasions, be classified and included under the label V.

2.2 Syllable

A syllable is a pronounceable unit of speech consisting of a vowel or a combination of a vowel with one or more consonants.

2.3 Polysyllable

A polysyllable is a series of more than one syllable pronounced with liaison between syllables.

AVA Documents
Document Sharing Hub

Full document details: ISO_TR 4870_1991

2.4 Word

A word is a monosyllabic or polysyllabic unit of speech that has an accepted meaning to the listeners.

2.5 Logatom

A logatom is a monosyllabic or polysyllabic unit that has no meaning to the listeners.

2.6 Test Sound

A particular speech sound to be used in accordance to defined rules to form test items.

2.7 Test Item

A particular monosyllabic or polysyllabic logatom, or word, to be used in accordance with defined rules for intelligibility measurements.

2.8 Set of Test Sounds

The particular test sounds that have been taken from the total set of possible (in a certain language) or given (according to major frequency in a certain language or according to certain rules) sounds, to form test items. The set of test sounds often is subdivided, dependent on the position of the proper test sound in the test item, into sets of initial, central, and final test sounds.

2.9 Set of Test Items

The particular logatoms or words that have been selected according to defined rules from the total amount of possible or given logatoms or words to be used for intelligibility measurements.

2.10 Subset Item

A speech sound, logatom, or word, to be used in accordance with defined rules for intelligibility measurements.

2.11 Phonemic Structure of Speech

Phonemic structure refers to the relative frequency of occurrence of different speech sounds and their positions relative to other speech sounds, in the syllables and words of a certain language.

2.12 "Phonetically Balanced" Lists

So called "phonetically balanced" (correct definition: phonemically balanced) lists are achieved when each list contains about the same proportion of the various classes or types of speech sounds as are

Documents
Document Sharing Hub
Full document details: ISO_TR 4870_1991

found to be, or presumed to be, present in specified speech communication with a given language.

2.13 Test List

A number of specially selected test items presented and scored as a single test. Typically, for open or pseudo-open test lists, a relatively large set of items is divided into a number of lists each containing an equal number of test items. Typically for the closed-set lists a number of subset ensembles are grouped together on one list.

2.14 Open Test List

Open lists of test items are made of items drawn randomly from the total set each time a list of test items is to be presented to listeners. Typically a listener writes on an answer sheet each test item the listener believes was presented.

Note: In order to insure that the phonemic structure in the total set appears properly in the test lists, it is necessary not to replace the items drawn randomly for one test list back into the total set of items prior to the random selection of items for succeeding lists.

2.15 Pseudo-open Test List

Pseudo-open lists of test items are made of items drawn on the basis of some specified set of phonemic rules, from the total set of items. The groupings of items within each test list, but not their sequential order, thus drawn is maintained for successive uses of the lists. Typically, a listener writes on an answer sheet each test item the listener believes was presented.

2.16 Repeat Test Items

Items within each open and pseudo-open test lists that are presented more than once within a list.

2.17 Pseudo-open List Scrambling

The items assigned to each pseudo-open test list are reordered on a random basis within each scrambling of each test list, to provide a number of sequences of items for each test list which are novel, or seem to be novel, to the listeners.

2.18 Closed-Set List

Closed ensembles of the order of 2 to 10 items per subset, are displayed visually to listeners during the test. One item of each subset is presented acoustically to the listener during a test, at which

AVA Documents
Document Sharing Hub

Full document details: ISO_TR 4870_1991

time the listener indicates, typically by a check mark on an answer sheet, which item of the visually displayed subsets involved was most probably presented acoustically.

Note: The subset ensemble is characterized by one speech sound that is the nucleus of every test item in it. All the test items in a given ensemble are initiated (or terminated) by the same speech sound and are terminated (or initiated) by different speech sounds.

2.19 Apparent Message Set Size

Apparent message set size refers to the number of alternative answers (to the presented items) presumed by a listener to be available as possibly correct answers to each item presented during an intelligibility test on the basis of the listener's knowledge of the total number of test items available to the speaker for presentation.

2.20 Real Message Set Size

Real message set size refers to the number of possible alternative answers by a listener to each item presented during an intelligibility test on the basis of the total number of items having audible phonemic similarities with each test items and which are within the set of test items available to the speaker for presentation.

2.21 Intelligible Speech Sound, Logatom, or Word

A speech sound logatom or word is defined as being intelligible when it is correctly perceived by a listener.

2.22 Percent Speech Intelligibility

Percent speech intelligibility is the percentage of items on a list correctly identified by a listener or group of listeners corrected for chance identifications dictated by the number of alternative answers per item available to the listener. This number is taken for the open or pseudo-open test lists to be the number of items in the total message set from which the test lists are drawn; for the small closed set lists, this number is taken to be the number of subset items or alternatives in a sub-set (note, not list size). In formula this can be expressed as follows:

$$I \text{ in } \% = \frac{100}{T} \left(R - \frac{W}{N-1}\right)$$

where T is the number of items in test and N is number of alternatives to each item. R is number of items right, W the number wrong. The last term is the correction for chance in item identification.

Documents
Document Sharing Hub

Full document details: ISO_TR 4870_1991

Note 1: By chance is meant that the listener is able to correctly guess a certain number of test items inasmuch as the listener knows, because of training or test format, the identity of all the possible alternative answers for each test item presented. For example, if the message-set consists of but 5 words, the listener would, on the average, score one out of five correct, or 20%, merely by guessing the identity of each test item.

Note 2: Under good listening conditions and high intelligibility scores, the size of the real, as opposed to the apparent set size of open or pseudo-open test list format, is of minor concern, because, as reflected in the last term of the formula for calculating percent speech intelligibility, the correction for chance is negligible when most items are correctly perceived. As the listening conditions and, accordingly, the intelligibility scores are degraded, the real message set size approaches the apparent size; i.e., the number of alternative responses is perceived as being much larger in number than is the case under good listening conditions. For tests in which the number of apparent alternative answers to each test item presented to the listener is greater than about 50, the correction for chance becomes negligible and percent speech intelligibility can be taken as the percent items correct on a test.

Example 1. If there are 50 test items on an open or pseudo open test list with each item having 1000 alternatives and 26 of the 50 test items were answered correctly and 24 of the items were answered incorrectly, the percent of speech intelligibility would be 52%. (100/50 (26 - 24/1000)= 51.952% , or, to round off, 52%.)

Example 2. If there are 50 small-closed set test items each consisting of 5 alternative subset items, and 26 of the 50 test items were answered correctly and 24 of the items were answered incorrectly, percent speech intelligibility would be 40%.(100/50 (26 - 24/4) = 40%.)

2.23 Carrier Sentence or Phrase

A sentence or phrase of at least 4 words and that contains a test item but such that the correct understanding of the test item is not dependent upon the context or meaning of the sentence in which it it embedded.

Note 1: The purpose of the carrier sentence is to provide: (1) the talker with means of enunciating the words in a natural manner and a controlled and measurable level of effort; (2) a regular temporal separation of test items of sufficient duration to permit listeners to decide and record

AVA Documents
Document Sharing Hub

Full document details: ISO_TR 4870_1991