ISCA Archive
http://www.isca-speech.org/archive

6th European Conference on
Speech Communication and Technology
(EUROSPEECH'99)
Budapest, Hungary, September 5-9, 1999

STANDARDISATION OF ERGONOMIC ASSESSMENT OF SPEECH COMMUNICATION

Herman J.M. Steeneken
Convenor of ISO/TC159/SC5/WG3, CEN/TC122/WG 8
E-mail: steeneken@tm.tno.nl
TNO Human Factors Research Institute
P.O. Box 23
3769 ZG, Soesterberg
The Netherlands

10.21437/Eurospeech.1999-67

ABSTRACT

The purpose of standardisation of the ergonomic assessment of speech communications is to assure a certain level of speech communication quality for various applications. The quality of speech communications is assessed in case of warning, danger, or information messages for work places, public areas, meeting rooms, and auditoria. In many applications direct communication between humans is considered while in other applications the use of electro-acoustic systems (e.g., PA systems) will be the most convenient means of informing and instructing people present.
A standard on this subject, under the responsibility of ISO and CEN, is in preparation. The draft version of this standard covers criteria for speech communication quality in various applications, methods to predict the speech transmission quality, and methods to assess the quality (subjective and objective). Examples of several applications are given in annexes of that standard.

1. INTRODUCTION

A certain level of speech communication quality for different applications will be standardised. This is the goal of the combined working group of ISO and CEN (ISO/TC 159/SC 5/WG 3; CEN/TC 122/WG 8.
The Draft International Standard specifies the *criteria* for speech communication quality in case of verbal alert and danger signals, information messages, and speech communications in general. Methods to *predict* and to *measure* the performance in practical applications will be addressed and examples will be given. For this purpose both subjective and objective methods are presented.

In comparison with visual alert and warning signals, auditory signals are omni-directional and may therefore be universal in many situations (smoke, out of line of sight). It is required however that, in case of verbal messages, a sufficient intelligibility is offered. If this cannot be achieved synthetic warning signals may be considered (see ISO 7731, IEC60849).

The communications may be directly between humans, through public address or intercom systems or by using prerecorded messages. In order to obtain optimal performance for a specific application three items are essential:
1: performance criteria,
2: development and predictive tools,
3: assessment methods.

These items will be covered in the draft standard by separate sections, where it will be recognized that for a general purpose document not only “high technology” solutions should be offered but also methods and tools which are simple to apply or generally available.

It is foreseen that a draft version of the new standard (ISO 9921) will be disseminated by the workgroup early in the year 2000. This standard replaces the present standard (ISO 9921-part1 ed. 1996).

2. SELECTION OF CRITERIA FOR SPEECH COMMUNICATION QUALITY

A requirement for the understanding of spoken messages is a correct recognition of each utterance. In technical terms it means that a sentence intelligibility score of 100% is required for simple sentences. However, there are many situations for which a better performance is required. If we consider alert and warning situations it is sufficient to fully understand a short message under adverse conditions even if it requires some effort from the listener to understand the message correctly. In a meeting room, auditorium, or at work places where speech communication is a part of the task or where people are normally present for a longer period of time, a more relaxed speaking and listening condition is required. For the speaker this may be re-

Table I. Recommended minimum criteria for the intelligibility for the categories of applications.

Application	Intelligibility	L_{SA} -L_{LN} dB (SIL)	STI	effective SNR dB	Maximum vocal effort
Alert and warning	poor/fair	9	0.45	-1.5	Loud
Person-to-person (critical)	poor/fair	9	0.45	-1.5	Loud
Person-to-person (relaxed	good	15	>0.60	3	Relaxed
Public address in public areas	fair/good	11	0.50	0	Normal
Personal communication systems	fair/good	11	0.50	0	Normal

flected in the vocal effort that is required to be understood (quantified as relaxed, normal, raised, loud, very loud, shouting, and maximum shout). For the listener the listening effort may be primarily related to the speech quality offered at the listening position. In Table I minimum criteria for various types of applications are proposed, but these are still under discussion by the workgroup (status April 1999). Five qualification intervals (excellent, good, fair, poor, bad) are proposed and related to various subjective and objective measures. Here we will use these qualifications as yardstick for speech quality at the various applications.

3. METHODS FOR PREDICTION OF THE PERFORMANCE OF SPEECH COMMUNICATION SYSTEMS

The prediction of the performance with respect to the intelligibility of speech communication channels is generally based on the effective signal-to-noise ratio at the listener position. Various methods are developed to calculate this effective signal-to-noise ratio derived from the vocal effort and acoustic aspects of the speaker, the transfer of the speech signal by electro acoustic systems, and the acoustical aspects at the speaker and listener position.

The various methods differ in complexity. Simple methods just compare the speech spectrum and the noise spectrum at the listener position. Advanced methods also take into account the effect of temporal distortion, non-linear distortion and hearing aspects.

The SIL-method (Speech Interference Level, [2]) is based on the speech and the noise spectrum described in four octave bands. A predictive measure of the intelligibility is obtained by substracting the mean noise level from the estimated speech level (expressed in dB, A-weighted, L_{SA} -L_{LN}).

The STI (Speech Transmission Index, [5, 9, 10, 11] and SII (Speech Intelligibility Index, ANSI) take into account the speech and noise spectrum and additionally the bandwidth, speech production and hearing aspects.

The SII is designed to predict various subjective intelligibility measures such as: nonsense syllables, phonetically balanced words, monosyllables, DRT words (Diagnostic Rhyme Test), short passages of easy reading material and monosyllables of speech in presence of noise. A slightly different calculation scheme is used in order to predict the scores related to the various subjective measures. SII takes also into account hearing aspects such as masking and hearing disorders.

The STI is designed for prediction of nonsense syllables, and also gives a qualification of the predicted speech intelligibility. Additionally to the other methods the STI accounts for temporal distortions by making use of the so-called Modulation Transfer Function (MTF), male and female speech signals, and for non-linear distortions.

None of the predictive intelligibility measures take into account the ability of a person to focus on speech sounds from a specific direction (directional hearing). Directional hearing might improve, under certain conditions, the intelligibility. This may be related to an improvement of the effective signal-to-noise ratio by approximately 3 dB.
The STI and SII are well described in various standards [IEC 60268-16 2nd edition, ANSI 305.2].

Table II. Qualification and relation between various intelligibility measures. The sentence score is related to simple sentences, CVC_{EB}-nonsense words with an equally balanced phoneme distribution, and the PB-word score (related to the phonetically balanced Harvard list).

Qualification	Sentence score %	CVC_{EB} non-sense word score %	Meaningful PB-word score %	STI	$L_{SA} - L_{LN}$ dB (SIL)	SII[1]
Excellent	100	>81	> 98	>0.75	21	>0.75
Good	100	70-81	93-98	0.60-0.75	15 - 21	
Fair	100	53-70	80-93	0.45-0.60	9 - 15	
Poor	70-100	31-53	60-80	0.30-0.45	3 - 9	<0.45
Bad	<70	<31	<60	< 0.30	< 3	

4. ASSESSMENT METHODS

Quantification of the speech quality first requires the specification of qualification intervals which cover the potential use, selection of measuring methods should comply with the qualification intervals. Also the selected measures must be applicable by the potential users of the standard. Therefore, the selected measures should cover the following specifications:

1 described in a standard or at least published and generally accepted,

2 reproducible,

3 producing results which comply with the required qualifications, such that scores from one method can be converted to another without ceiling effects (saturation),

4 including subjective and objective measuring methods,

5 applicable to the (acoustical) conditions covered by the standard.

In Table II three subjective [see 1, 3, 5, 7, 8, and 10] methods are compared. Some of these are described in standard ISO 4870. It is of great importance that the test material for a subjective test used in reverberating environments makes use of test words embedded in a carrier phrase in order to make sure that a representative reverberation is present during the presentation of the test word.

Also three objective methods [see 5, 6, 9, 11] are given in Table II. These methods are similar to those discussed for prediction purposes.

We compared, for a number of noise conditions, the scores of the STI, SII and SIL. In Fig. 1 the relation between STI and SII is given for 40 noise conditions with random spectrum distributed over a wide range of octave levels. The rank-order correlation between both measures is r = 0.93. The figure shows a slight saturation of the SII at higher values.

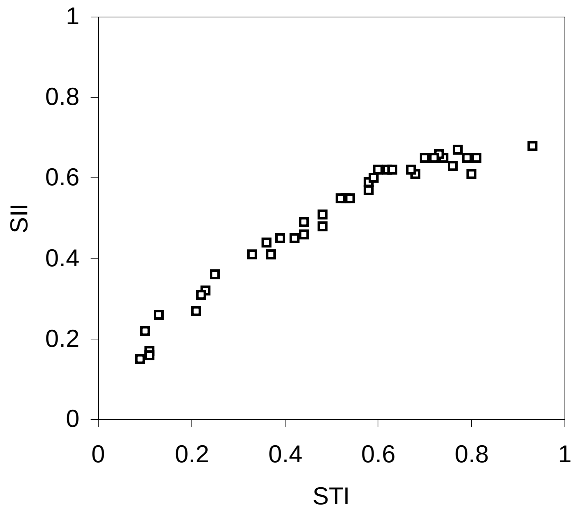

Fig. 1 relation between STI and SII. The correlation coefficient amounts r = 0.93.

The relation between SIL and STI / SII is given in Fig. 2. Here the rank order correlation of SIL and STI amounts r = 0.97 and between SIL and SII r = 0.95. It should be noted that SIL is only applicable for noise conditions.

The field of application of the objective methods depends on the ability to cope with the distortions which are relevant for a specific application. In brief possible distortions are: background noise, reverberation, echoes, increased vocal effort of the speaker. In case that electronic means are used also band-pass limiting and non linear distortions have to be included.

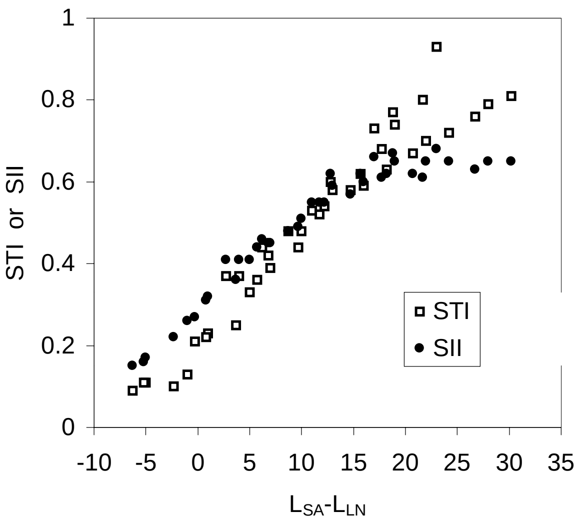

Fig. 2. Relation between SIL and STI and SII. The correlation coefficients between SIL and STI is r=0.97 and between SIL and SII r=0.95.

5. CONCLUSIONS

Dissemination of a new standard on the ergonomic assessment of Speech Communication is in progress. The standard will give criteria for acceptable speech communication quality in various conditions from warning and alert conditions to the more relaxed meeting room. Additional to the criteria methods to assess the performance of existing situations or to predict the performance for applications under development are also given.

The methods proposed in the standard are not restricted to advanced “high technology” methods but include also simple methods which are easy to apply in “in situ” conditions as long as basic requirements are fulfilled.

6. ACKOWLEDGMENTS

The author is indebted to Mr. Sander van Wijngaarden for the calculations of the comparison between STI, SII and SIL.

7. REFERENCES

[1] Anderson, B.W., and Kalb, J. T. (1987). "English verification of the STI method for estimating speech intelligibility of a communications channel," J. Acoust. Soc. Am. 81, 1982-1985.

[2] Beranek, L.L. (1947) “Airplane quieting II specification of acceptable noise levels”. Trans. Amer. Soc. Mech. Engrs. 69:97-100.

[3] Fletcher, H., and Steinberg, J.C. (1929). “Articulation testing methods”, Bell Sys Tech. J. 8, 806.

[4] House, A.S., Williams, C. E., Hecker, M.H.L., and Kryter, K.D. (1965). "Articulation testing methods: Consonantal differentiation with a closed-response set," J. Acoust Soc. Am. 37, 158-166.

[5] Houtgast, T., and Steeneken, H.J.M. (**1984**). "A multi-lingual evaluation of the RASTI-method for estimating speech intelligibility in auditoria," Acustica **54**, 185-199.

[6] Lazarus, H., (1990). “New methods for describing and assessing direct speech communication under disturbing conditions”. Environment International, 16, pp. 373-392.

[7] Miller, G. A., and Nicely, P. E. (1955). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. 27, 338-352.

[8] Plomp, R., and Mimpen, A.M., (1979). “Improving the reliability of testing the speech reception threshold for sentences”. Audiology 8, 43-52.

[9] Steeneken, H.J.M., and Houtgast, T. (**1980**) “A physical method for measuring speech transmission quality”. J. Acoust. Soc. Am. **67**, 318-326.

[10] Steeneken, H.J.M. (1992). "Quality evaluation if speech processing systems," Chapter 5 in Digital Speech Coding: Speech coding, Synthesis and Recognition, edited by Nejat Ince, (Kluwer Norwell USA), 127-160.

[11] Steeneken, H.J.M., and Houtgast, T. (1999). "Mutual dependence of the octave-band weights in predicting speech intelligibility". accepted for publication in Speech Communication 1999.