ARCHIVES OF ACOUSTICS
**31**, 4 (Supplement), 171–176 (2006)

# EXPERIMENTAL COMPARISON BETWEEN SPEECH TRANSMISSION INDEX (STI) AND MEAN OPINION SCORES (MOS) IN ROOMS

S. BRACHMAŃSKI

Wrocław University of Technology
Institute of Telecommunications, Teleinformatics and Acoustics
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
e-mail: Stefan.Brachmanski@pwr.wroc.pl

(*received June 15, 2006; accepted September 30, 2006*)

Two methods for assessment of speech quality in rooms were compared: the speech transmission index (STI) and mean opinion scores (MOS). Subjective and objective measurements were carried out in two rooms for different signal to noise ratio within a range from $-15$ dBA to $+15$ dBA. The obtained results are shown as a curve describing a relation between a MOS, STI and signal to noise ratio (S/N).

**Key words:** speech quality, STI, room acoustic.

## 1. Introduction

The methods for assessment speech quality fall into two classes: subjective (listening) and objective methods. From among the different listening tests, nowadays the techniques, which give directly (ACR – absolute category rating, DCR – degradation category rating, CCR – comparison category rating) or indirectly (intelligibility measurements) a MOS (mean opinion scores) rating on a five-grade quality scale are used [1, 2, 4–8, 10–15, 19, 22]. The subjective assesment is natural method to evaluate the quality of speech. In the subjective methods listeners listen to the specific speech material (e.g. logatoms, short words, short sentences), which has been mixed with disturbances and they write what they hear. Unfortunately subjective tests are often very expensive, time-consuming and labour intensive. The designers and manufacturers of the speech transmission systems tend to use the objective techniques of speech quality, not taking into account their limitations and preciseness. Still, the final verification of quality of devices used for speech transmission is done by their human user and the verification is made through subjective measurements of transmission quality. In this study, two methods to measure speech quality are compared, namely absolute category rating and the speech transmission index.

## 2. Speech Transmission Index (STI, RASTI)

Houtgast and Steeneken described an objective method for estimating the speech intelligibility in rooms by calculating a physical index, called the speech transmission index (STI), from the modulation transfer function (MTF) [7, 8, 10, 16–18]. This method was modified by the author of this paper for Polish speech transmitted via analog telecommunication channels [3, 5]. The MTF method uses a test signal (random noise), whose spectrum correspond to the human speech. The intensity of this noise in octave bands is sine-modulated with the modulation frequencies relevant to the envelop of speech. The additive interferences (noise, reverberation) reduce the modulation depth of test signals. The STI value is calculated by weighting the average MTF value for seven octave frequency bands (125 Hz to 8 kHz) and for 14 modulation frequencies (midle frequencies in one-third octave bands from 0.63 Hz to 12.5 Hz) [5, 7, 16].

In 1985, the foreshortened adaptation of Speech Transmission Index (STI) was developed by Houtgast and Steeneken and termed RASTI (Rapid Speech Transmission Index) [20]. The RASTI method is a simpliefied version of STI. This method is restricted to the 500 Hz and 2000 Hz octave bands and to 4 (1, 2, 4 and 8 Hz) and 5 (1, 4, 2.8, 5.6 and 11.2 Hz) different modulation frequencies respectively. Another simplified version of the STI is the STITEL (Speech Transmission Index for Telecommunication Systems). The STITEL method applies the same octave bands as the STI, but in each band only one modulation frequency is used. The test signal includes all seven octave bands, which all are analyzed simultaneously. An efficient form of Speech Transmission Index method for public address systems is STI-PA (Speech Transmission Index-Public Address) [21, 23].

## 3. Absolute category rating

The ACR method is recommended by ITU [22] for subjective assessment of speech quality. The speech material (test lists) used in this method should consist of simple, short, semantically unrelated sentences. The test material should be properly prepared and recorded. The speaker should pronounce the sentences fluently and should not have any speech defects. Since the female voice and the male voice have different characteristics, the two types of voice should be included in the measurements. The results obtained for male and female voices should be evaluated separately. They can be averaged only when they do not differ significantly. To reduce the influence of the individual characteristics of the speaker's voice on the obtained result, several speakers should take part in the experiment. The experiment's listening part should take place in a room with a noise level below 30 dBA.

Listeners are chosen at random from the normal telephone using population, with the provisions that:

- they have not been involved in work connected with assessment of performance of telephone systems or speech coding,
- they have not participated in any subjective measurements at least the previous six months,
- they have never heard the same sentences lists before.

Listeners listen to the sentences from test lists and give their opinions in five levels scale. Various scales may be used for different purposes. Operator give the following opinion scales recommended by ITU [14, 22]:

a) listening-quality scale (Excellent speech is rated 5, Good – 4, Fair – 3, Poor – 2, Bad – 1),
b) listening-effort scale (Complete relaxation possible; no effort required is rated 5, Attention necessary; no appreciable effort required – 4, Moderate effort required – 3, Considerable effort required – 2, No meaning understood with any feasible effort – 1),
c) loudness-preference scale (Much louder than preferred is rated 5, Louder than preferred – 4, Preferred – 3, Quieter than preferred – 2, Much quieter than preferred – 1).

The average rating (Mean Opinion Score – MOS) is calculated over the listeners and the speakers for each tested speech transmission condition.

## 4. Experiment

The measurements were performed in two, unoccupied rooms (lecture rooms). In the presented experiment, in each room, four listener locations were selected (Fig. 1). Sound sources (voice and white noise) were positioned in the part of the room normally used for speaking. One loudspeaker was the voice source and second – the noise source.

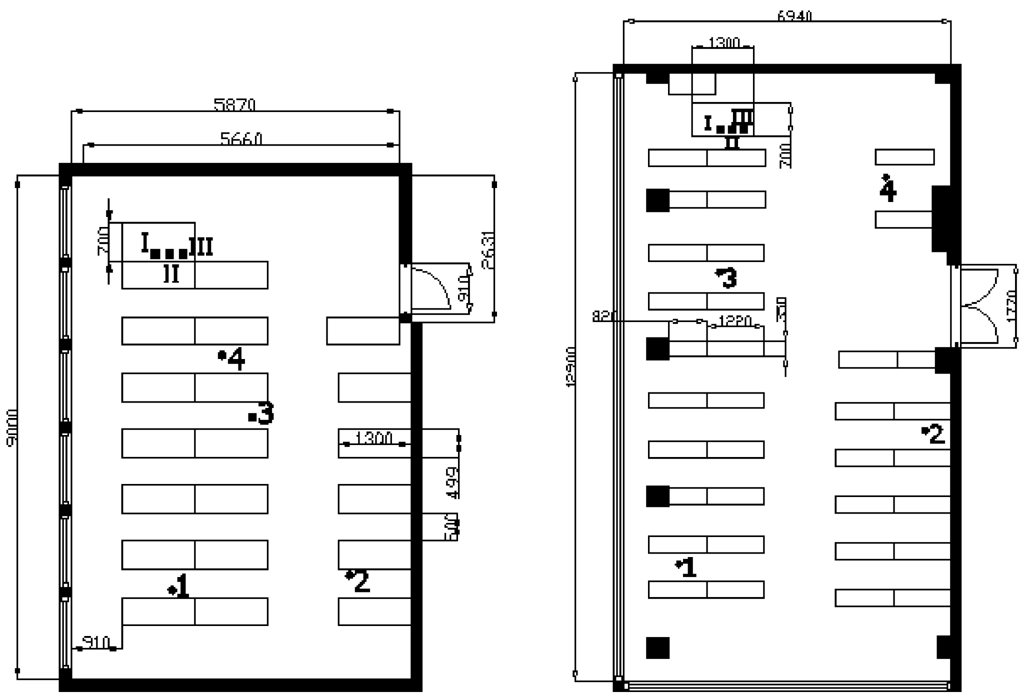


Fig. 1. Plan view of the rooms showing source position and the receiver positions (**I** – loudspeaker – source of the logatoms, **II** – RASTI, **III** – loudspeaker – source of the noise, **1, 2, 3, 4** – measurement points).

Different conditions in the room are the result of white noise in different level of volume. In each room the measurements were done in four points where the audience was situated in the back (2 points), in the middle and in the front, to the level of 1.1 m (approximate height of listener's head). The test signal for ACR method (sentence lists), were played on the DAT recorder with loudspeaker set in the front of the hall that is in the place of the rostrum. Next to the emitting set, which issued the test signal, the source of disturbing noise was located. In each measurement point, test signals were recorded on the digital recorders. In the same measurement points, in which test signals were recorded, the RASTI measurement were done. The RASTI values were measured with a Brüel & Kjaer Speech Transmission Meter (Type 3361). The RASTI system consists of a transmitter (Type 4225), which was placed at the speaker's position and a receiver (Type 4419) placed at the listener's position. The averaging time was 32 s. Dependence of STI values in the function of signal to noise ratio for testing room is presented in the Fig. 2. On the basis of the measured STI value the quality of speech assessment (MOS-STI) was calculated in five level scale according to Table 1.

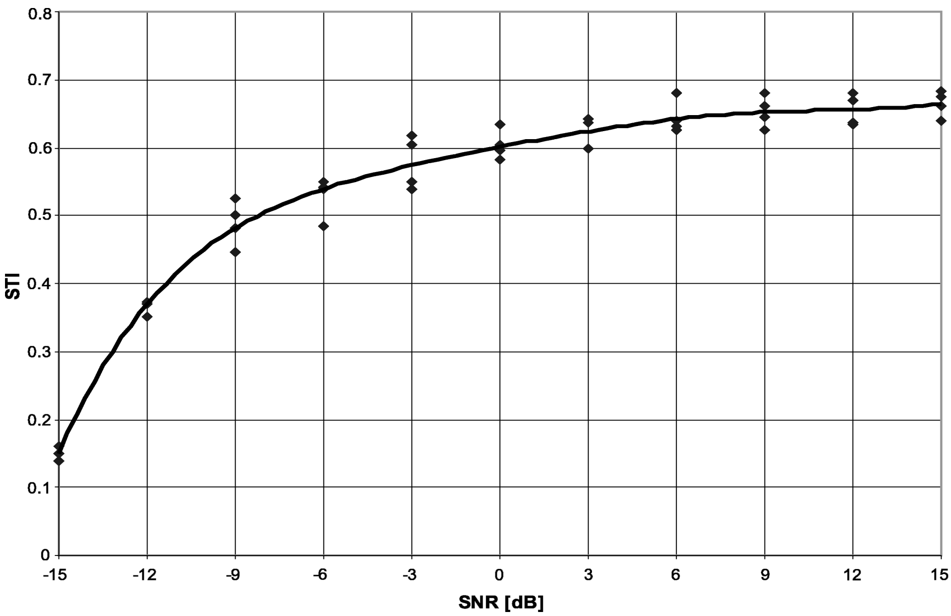


Fig. 2. The example relation between STI and signal to noise ratio in the test lecture room (nrc3).

**Table 1.** The speech quality assessment scale on the basis of RASTI factor [9].

| Quality | Bad (1) | Poor (2) | Fair (3) | Good (4) | Excellent (5) |
|---|---|---|---|---|---|
| RASTI | 0–0.3 | 0.3–0.45 | 0.45–0.6 | 0.6–0.75 | 0.75–1 |

Subjective measurements of MOS factor were done using the ACR method according to the ITU-T P.800 recommendation with a listening group made up of 12 persons in the age from 18 to 25 years. The test material had the form of phonematically balanced sentence lists. It had been recorded by one speaker (male voice) on a R-DAT recorder, divided into sentences and loaded into a computer. The sampling rate of 16 kHz and the resolution of 16 bits were used. There were 50-sentence lists one per each measuring point (different transmission conditions).

In Fig. 3 the speech quality measurement results in the lecture room for different signal to noise ratio (SNR) values were presented. The presented MOS values were obtained by means of the subjective speech quality measurements using ACR (MOS – ACR) method and on the basis of the objective measurements RASTI (MOS-STI).

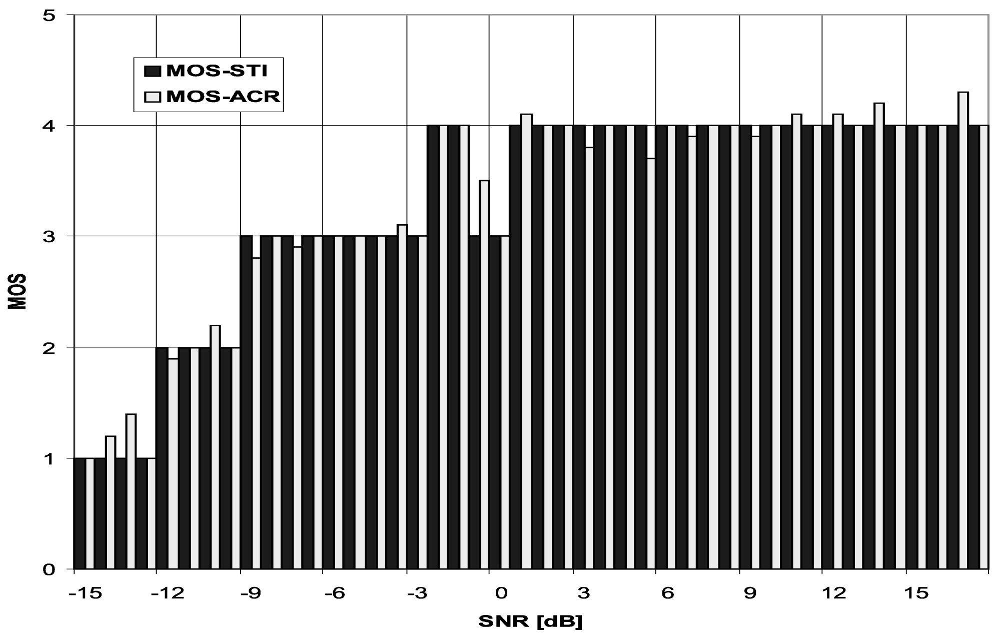


Fig. 3. Relation between MOS and speech to noise ratio in the test lecture room. The MOS value was obtained on the basis of STI (MOS-STI) and ACR (MOS-ACR).

## 5. Conclusion

The carried out experiment has shown that the STI method gives a very close MOS assessments to the subjective measurement results in the indoor conditions for Polish language for a white noise as an interfering signal. The obtained speech assessment results are in the range from one to four in the MOS scale in view of high delays. In this range there was an almost 100% compatibility achieved between subjective MOS and the results, obtained basing on the objective measurements RASTI.

The aim of the further research is the calculation the relationship between logatom intelligibility and STI for speech transmission in indoor environment for Polish language.

## References


[1] ANDERSON W. B., KALB J. T., *English verification of STI method for estimating speech intelligibility of a communications channel*, JASA, **81**, 6, 1982–1985 (1987).

[2] BRACHMAŃSKI S., *Effect of additive interference on speech transmission*, Archives of Acoustics, **27**, 2, 95–108 (2002).

[3] BRACHMAŃSKI S., *Modulation Transfer Function (MTF) as measure of Polish speech transmission quality* [in Polish], Doctoral Dissertation, Wrocław 1982.

[4] BRACHMAŃSKI S., *Subjective measurement of speech transmission quality in telecommunication chanels* [in Polish], Proceedings of 46-th Open Seminar on Acoustics, pp. 99–104, Kraków – Zakopane 1999.

[5] BRACHMAŃSKI S., *Estimation of logatom intelligibility with STI method for Polish speech transmitted via communication channels*, Archives of Acoustics, **29**, 4, 555–562 (2004).

[6] FARINA A., *Acoustic quality of theatres: correlations between experimental measures and subjective evaluations*, Applied Acoustic, **62**, 889–916 (2001).

[7] HOUTGAST T., STEENEKEN H. J. M., *The Modulation Transfer Function in room acoustics as a predictor of speech intelligibility*, Acustica, **28**, 66–73 (1973).

[8] HOUTGAST T., STEENEKEN H. J. M., *Predicting speech intelligibility in rooms from the Modulation Transfer Function. I. General room acoustics*, Acustica, **46**, 60–72 (1980).

[9] JACOB K. D., BIRKLE T. K., ICKLER C. B., *Accurate prediction of speech intelligibility without the use of in room measurements*, J. Audio Eng. Soc., **39**, 4, 232–242 (1991).

[10] LAM P., HONGISTO V., *Experimental comparison between speech transmission index, rapid speech transmission index, and speech intelligibility index*, J. Acoust. Soc. Am., **119**, 2, 1106–1117 (2006).

[11] MACKIE K. II, *Assesment of evaluation measures for processed speech*, Speech Comm., **6**, 309–316 (1987).

[12] MAJEWSKI W., MYŚLECKI W., BRACHMAŃSKI S., *Methods of assessing the quality of speech transmission* [in Polish], Proceedings of 47-th Open Seminar on Acoustics, Rzeszów – Jawor, 66–75 (2000).

[13] MAPP P., *A comparison between STI and RASTI Speech Intelligibility Measurement Systems*, The 111-th AES Convention, Los Angeles, Preprint 5668 (2002).

[14] MAPP P., *Limitations of current sound system intelligibility verification techniques*, The 113-th AES Convention, Los Angeles, Preprint 5668 (2002).

[15] SOTSCHEK J., *Methoden zur Messung der Sprachgüte I: Verfahren zur Bestimmung der Satz- und der Wortverständlichkeit*, Der Fernmelde Ingenieur, **10**, 1–31 (1976).

[16] STEENEKEN H. J. M., HOUTGAST T., *A physical method for measuring speech-transmission quality*, JASA, **67**, 1, 318–326 (1980).

[17] STEENEKEN H. J. M., HOUTGAST T., *Mutual dependence of the octave-band weights in predicting speech intelligibility*, Speech Communication, **28**, 109–123 (1999).

[18] STEENEKEN H. J. M., HOUTGAST T., *Validation of revised STI method*, Speech Communication, **38**, 413–425 (2002).

[19] TISSEYRE A., MOULINIER A., ROUARD Y., *Intelligibility in various rooms: Comparing its assessment by (RA)STI measurement with a direct measurement procedure*, Applied Acoustic, **53**, 1-3, 179–191 (1998).

[20] WIJNGAARDEN S. J., STEENEKEN H. J. M., *Objective prediction of speech intelligibility at high ambient noise level using the speech transmission index*, Proc. of the 6-th Europeen Conf. on Speech Comm. and Technol., Eurospeech 99, Budapest, **6**, 2639–2642 (1999).

[21] IEC 60268-16, *Sound system equipment – Part16: Objective rating of speech intelligibility by speech transmission index*, Ed. 3, Int. Electrotechnical Commission, 2003.

[22] ITU-T, *Recommendation P800* (1996).

[23] POLISH STANDARD PN-EN-60268-16, *Sound system equipment – Part16: Objective rating of speech intelligibility by speech transmission index*, Warszawa 2002.