

OBJECTIVE MEASURE FOR ASSESSMENT OF SPEECH QUALITY IN ROOMS

Stefan 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, 2008; accepted October 28, 2008)

The methods for assessment of speech quality fall into two classes: subjective and objective methods. This paper includes an overview of the two methods of objective measurements speech intelligibility (Speech Transmission Index and the automatic speech recognition system (QE-ARM)). Experiments in which the methods were used are described. The results are compared with subjective measurements. Subjective and objective measurements were carried out in a room for different signal to noise (white and pink noise) ratio. The obtained results are shown as a curve describing a relation between a STI and signal to noise ratio (S/N) and STI versus logatom speech intelligibility.

Keywords: speech quality, speech intelligibility, room acoustic.

1. Introduction

This paper is a continuation of group research, which was done to find relations between results of objective speech quality measurements and subjective logatom speech intelligibility measurements for different types of noise in a room.

The methods for assessment speech quality fall into two classes: subjective (listening) and objective methods. 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. The results depend on the speech material used. Unfortunately subjective tests are often very expensive, time-consuming and labour intensive.

In this paper two physical, objective methods of assessment of speech intelligibility are compared, namely the speech transmission index (STI) and the automatic speech recognition system (QE-ARM).

2. Objective methods

2.1. Speech transmission index (STI, RASTI)

STI Method, which is recommended by IEC60268-16 [6], was made by Houtgast and Steeneken. This method is designed for assessment of the speech transmission quality. It is realized by calculating parameter of quality from the modulation transfer function (MTF) [5]. This method was modified by the author of this paper for polish speech transmitted via analog telecommunication channels [4]. 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 (middle frequencies in one-third octave bands from 0.63 Hz to 12.5 Hz) [3, 5].

In 1985, the foreshortened adaptation of Speech Transmission Index (STI) was developed by Houtgast and Steeneken and termed RASTI (Rapid Speech Transmission Index). The RASTI method is a simplified 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) [6, 7].

2.2. The QE-ARM method

The automatic speech recognition system (QE-ARM) for objective speech quality evaluation is based on finite state and memoryless automatic speech recognition procedures [1]. The original signals (logatoms), according to their length (i.e. duration), are divided into 2–5 closed sets. The signal is preemphasised using a first order filter with a transfer function $H(z) = 1 - 0.95 \cdot z^{-1}$ and parameterised using FFT, BF-FFT, LPC or Cepstrum method (Fig. 1).

The next step is initial classification of signals according to their duration. The parameters of original signals are stored in n pattern tables corresponding with sets resulting from the classification. In this manner we obtain n finite sets of patterns. The nearest neighbour (NN) algorithm is used as a decision rule. Parameters of every test signal are compared to patterns from a suitable set and nearest of them becomes the answer of the system i.e. the recognised class. Dynamic time warping (DTW) technique is used to appropriately temporally align the features of the test signal with those of the reference

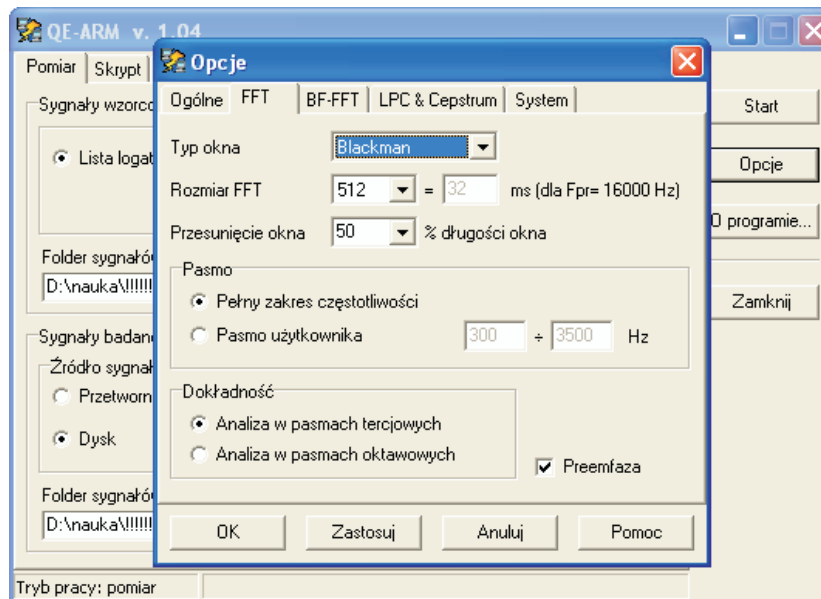


Fig. 1. The example of the window of the QE-ARM.

one before computing a match score. As a similarity function the Hamming distance is used.

The relation of the correctly recognized test signals to all transmitted ones is the measure of speech quality (e.g. logatom intelligibility) and the efficiency of recognition.

3. Experiment

The measurements were performed in one, unoccupied room (lecture room). The lecture room had seats for 70 listeners. The volume of the room was approximately 303 m^3 . The A-weighted background noise levels were then 30 dBA.

In the presented experiment four listener locations were selected [2]. These positions were chosen in the expectation of yielding a wide range of speech intelligibility. Sound sources (voice, white or pink noise) were positioned in the part of the room normally used for speaking. One loudspeaker was the voice source and second – the noise source. The different conditions in the room are the result of white or pink noise in different level of volume. The test signal for speech intelligibility (phonematically balanced logatom 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.

The subjective tests were done according to Polish Standard PN-90/T-05100 [8] with team of listeners made up of 8 listeners in age from 18 to 25 years. The listening team was selected from persons with normal hearing. The qualification was based on audiometric tests. The test material had the form of phonematically balanced and structurally balanced logatoms lists. It had been recorded by one speaker (male voice) on a R-DAT recorder and then loaded into computer. The sampling rate of 16 kHz and the resolution of 16 bits were used. Each measure point (the place where the measure position was situated) consisted of one list of 100 logatoms.

In QE-ARM method measurements the same test lists were used, which were used in subjective measurements. The estimator logatom intelligibility was calculated with QE-ARM method for BF-FFT parametrisation, Blackman Harris window and Hamming distance measures. For those properties the best compatibility is with subjective measurements. The influence of system OE-ARM parameters changes was also measured for the effectiveness of the method. The method effectiveness is the compatibility of the subjective and objective results.

The dependence of logatom intelligibility in function of STI values for testing the room for a disturbing noise (white and pink) is presented in the Fig. 2. The relationships between QE-ARM values, logatom intelligibility and STI values for all speech transmission conditions (white and pink noise, different signal-to-noise ratio) are presented in Fig. 3.

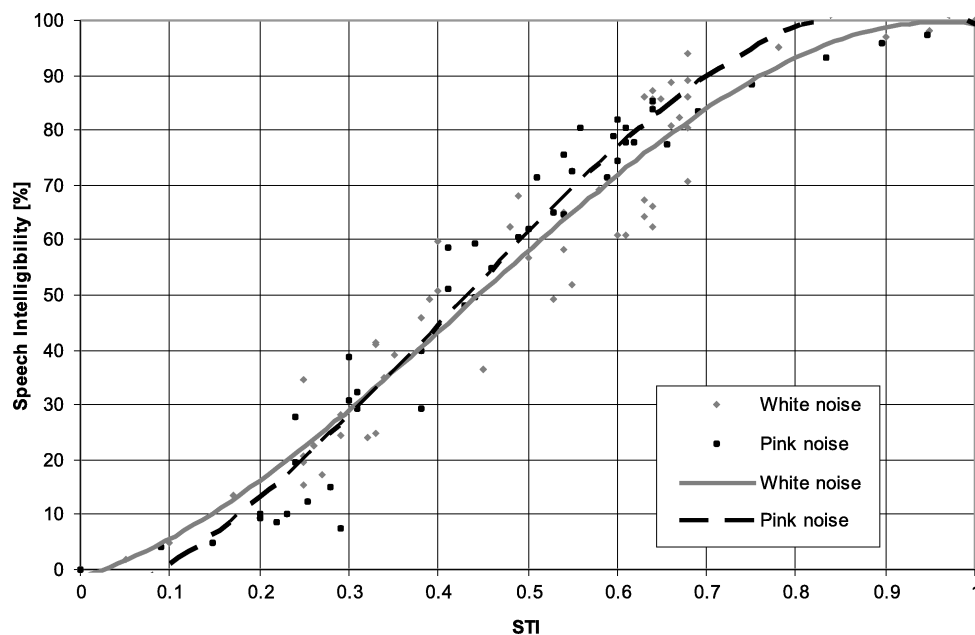


Fig. 2. Relationship between STI and logatom intelligibility for white and pink noise.

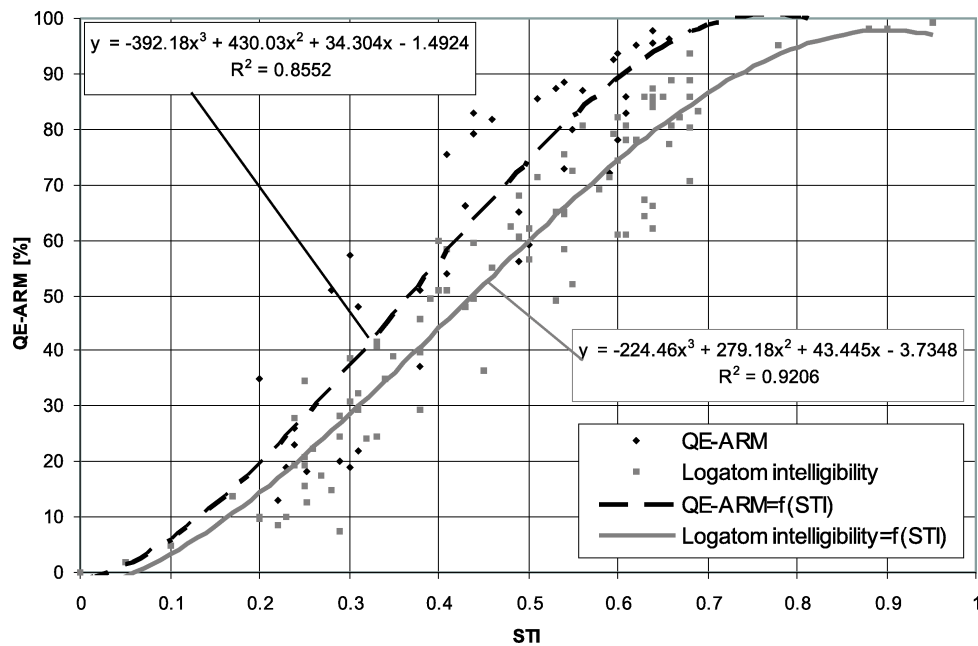


Fig. 3. Relationship between STI and logatom intelligibility and QE-ARM.

4. Conclusions

Within the confines of this paper lots of experiments were made to measure speech transmission quality in rooms in which there were different kinds of noise like white and pink. The tests were measured with objective methods: QE-ARM and RASTI (STI). Results of those experiments were compared to the subjective assessments, which were made with logatom intelligibility. QE-ARM method measurements were made for many system properties but the most effective results were observed for the properties presented in this paper.

The experiments carried out in finding the relations between logatom intelligibility, QE-ARM value and STI for the rooms have shown that there exists the multi-value and repetitive relation between them. It allows using both methods interchangeably and converting results between them. The presented STI and QE-ARM methods offer a simple, easy to use, stable and fully automatized system to assess speech transmission quality.

References

- [1] BAŚCIUK K., BRACHMAŃSKI S., OLACZEK A., *Objectivisation of perceptual evaluation of speech transmission quality*, The 106th Convention AES, Monachium, 1999, Preprint 4954.
- [2] BRACHMAŃSKI S., *Experimental comparison between speech transmission index (STI) and mean opinion scores (MOS) in rooms*, Archives of Acoustics, **31**, 4 (Supplement), 171–176 (2006).

- [3] BRACHMAŃSKI S., *Estimation of logatom intelligibility with the STI method for polish speech transmitted via communication channels*, Archives of Acoustics, **29**, 4, 555–562 (2004).
- [4] BRACHMAŃSKI S., *Modulation Transfer Function (MTF) as measure of Polish speech transmission quality* [in Polish], Doctoral Dissertation, Wrocław 1982.
- [5] HOUTGAST T., STEENEKEN H.J.M., *The Modulation Transfer Function in room acoustics as a predictor of speech intelligibility*, Acustica, **28**, 66–73 (1973).
- [6] IEC 60268-16, *Sound system equipment – Part 16: Objective rating of speech intelligibility by speech transmission index*, Ed. 3, International Electrotechnical Commission, 2003.
- [7] Polish Standard PN-EN-60268-16, *Sound system equipment – Part 16: Objective rating of speech intelligibility by speech transmission index*, 2002.
- [8] Polish Standard PN-90/T-05100, *Analog telephone chains. requirements and methods of measuring logatom articulation* [in Polish], 1990.