ARCHIVES OF ACOUSTICS
33, 3, 341–350 (2008)

AUTOMATION OF SUBJECTIVE MEASUREMENTS OF SPEECH INTELLIGIBILITY IN ANALOGUE TELECOMMUNICATION CHANNELS

Stefan BRACHMAŃSKI

Wrocław University of Technology
Institute of Telecommunications, Teleinformatics and Acoustics
Wybrzeże Wyspiańskiego 24, 50-370 Wrocław, Poland
e-mail: stefan.brachmanski@pwr.wroc.pl

(*received April 03, 2008; accepted July 25, 2008*)

A method of evaluation of speech transmission quality, called "modified intelligibility test with forced choice" (MIT-FC), is presented. This method provides a fully automated measurement of speech intelligibility (the listener's task is to select on the computer monitor which of the alternative utterances presented visually to him was spoken). The MIT-FC method was implemented in a stationary version for measurements of telecommunication systems and chains and in a portable version for the measurements of auditoria and mobile communication systems. The results of application of the MIT-FC method in telecommunications have been presented and the results obtained have been compared with those of traditional intelligibility tests.

Keywords: speech quality, speech intelligibility

1. Introduction

The commonly used voice communication systems with technical devices such as digital and analogue telephone, radio and data transmission (IP networks) create certain technical demands according to the method of speech transmission quality evaluation. The development of modern speech transmission technologies in telecommunication and computer networks has created a demand for the evaluation of the influence of transmission parameters (like: coding techniques, compression methods and transfer characteristics) on the speech transmission quality [2–4, 14, 35]. At the same time, the test result of speech transmission quality evaluation is the assessment of a given telecommunication chain. Most of the speech transmission quality evaluation studies have been focused on telephony transmission [1, 7, 8, 16, 20, 26, 27, 36].

The main goal of telephony transmission is to make the transfer of a message possible by means of voice. The quality of the message transfer depends both on physical parameters of telecommunication devices and on the subjective ones, which are specific to the users of the above devices [12, 15, 37]. All the used methods of speech transmission evaluation should, at a certain stage, consider the subjective factors by applying subjective measurement methods or by the estimation of objective measurement with respect to the subjective parameters [10, 13, 19, 24, 25, 28–30, 34, 38].

The condition of comfortable communication is that the interlocutors can understand each other with insignificant number of reflexive questions. If we remove the condition of naturalness of transmitted speech and restrict the condition to the transmission of linguistic information only, then the speech quality is sufficient if we have good speech intelligibility for adequate loudness of the received signal. The speech loudness and intelligibility are independent from each other. The quality of communication is not satisfactory in cases when both the sound is too low or too high even if transmitted in a full band without distortions and when the loudness is adequate but the transmission band is too narrow or the distortion too large. These two measures, taken separately or together, are the base of subjective methods [22, 23].

Speech intelligibility is one of the basic quality parameters of speech transmission in analogue and digital telephone transmission [3, 31–33, 39, 40], IP networks [4, 7, 8], auditory rooms [9, 11, 24] and in the selection of hearing aids [18]. Despite the significant progress in instrumental techniques of measurement of this parameter, the only reliable method is still time-consuming, expensive and requires high skilled users of the subjective measure of intelligibility along with the participation of a trained team of listeners. It is quite understandable that designers of speech transmission systems or devices tend to use objective measurement techniques not always taking into consideration the limitations of the application and precision which depend on the type of a researched object and measurement conditions. Nevertheless, the final verification of a speech transmission devices has always been done by their user: the man.

The subjective measure results should mostly depend on the physical parameters of the tested transmission chain and do not depend on the structure of the tested language material. The elimination of semantic information is done by means of logatom(1) lists (i.e. pseudo-words) on the basis of which the logatom and phone intelligibilities are obtained [1, 9, 33, 39, 40]. This is a reason for using average logatom intelligibility of a group of listeners as a reference point of of the telecommunication chain quality for the verification of objective methods.

2. Subjective methods of assessing the speech transmission quality

2.1. Logatom intelligibility measurement method

Subjective measurements of logatom intelligibility are recommended by ISO "Acoustics – The construction and calibration of speech intelligibility tests" [33] and

(1) Logatom – (*logos* (gr) – spoken phrase, *atom* (gr) – indivisible) vocal sound, generally insignificant, usually made by the sound of a consonant or the first consonant, then by an intermediate vowel, finally by a consonant or a final consonant sound.

the Polish Standards: PN-T-05100 "Analogue Communication Systems. Requirements and Methods for Measurement of Logatom Intelligibility" [39] and PNV90002 "Digital Communication Systems. Requirements and Methods for Measurement of Logatom Articulation" [40].

The measurement of logatom intelligibility is based on transmission of logatom lists (100 logatoms) phonetically balanced, read out by a speaker, transmitted through the tested channel, which are then written down by listeners. Listeners write down (in orthographic form) the received logatoms on a special sheet of paper. The correctness of the record group is checked by a group of experts. The procedure of checking the correctness of the received logatoms begins with a comparison of the orthographical form of the logatom written down by the listener and the source one. In case of no difference, it is assumed that the logatom has been correctly received. If there is a difference, the expert creates a phonetically form of the source logatom. The existence of at least one difference in both forms (at least one error in received logatom) means that the logatom has not been received correctly.

In the next step, the experts, who check the lists, calculate the number of correct responses for each listener and each logatom list, and then an average logatom intelligibility is determined as a proportion between the number of correctly received logatoms and all logatom transmitted. The obtained average logatom articulation value can be used to the determine quality classes according to Table 1 [17, 21, 39].

Table 1. Classes of speech intelligibility quality for analog channels [39].

Quality class	Description of quality class	Logatom intelligibility [%]
I	Understanding transmitted speech without slightest concentration of attention and without subjectively detectable distortions of speech signal	above 75
II	Understanding transmitted speech without difficulty but with subjectively detectable speech distortions	$60 \div 75$
III	Understanding transmitted speech with concentration of attention but without repetitions and return queries	$48 \div 60$
IV	Understanding transmitted speech with great concentration of attention and with repetitions and return queries	$25 \div 48$
V	It is impossible to fully understand transmitted speech (breakdown of communication)	to 25
For each quality class lowest logatom intelligibility values are lowest admissible values		

2.2. Modified intelligibility test with forced choice

The subjective measurement of logatom intelligibility is very time- and cost-consuming, because it is necessary to have a high-specialized group of experts as the listener staff in order to obtain the high reliability of measurement results. To avoid the disadvantages of the subjective evaluation of the logatom intelligibility by means of the "traditional" method, a new measurement method was created and developed at the Institute of Telecommunications and Acoustics of the Wrocław University of Technology. This method is called the *"Modified Intelligibility Test with Forced Choice"* (MIT-FC) [1, 5, 9].

In the MIT-FC method, a computer controls all experiments. The automation of the subjective measurement is connected with the basic change in the generation of logatoms and in making the decision by a listener. The computer generates a list of utterances (for logatom test the list consists of 100 phonetically balanced nonsense words), presents the utterances to the listeners subsequently and for each spoken utterance there are several logatoms visually presented (in orthographic form) that have been previously selected as perceptually similar. It has been found that the optimal number of logatoms presented visually to the listeners is seven (six alternative logatoms and one transmitted logatom to be recognized) [6]. The listener chooses one logatom from the list visually presented on the computer monitor. The computer counts the correct answers and calculates the average logatom intelligibility and the standard deviation. The measurement time for one logatom list consisting of 100 logatoms is 8–9 minutes. All the measurement procedures are fully automated and an operator has flexible possibilities to set the measurement parameters and options. It is also possible to upgrade the application that realizes the MIT-FC method with more sophisticated scores processing.

3. Experiment

The goals of the experiment are as follows:

1) to verify whether the results of the traditional and modified with forced choice methods enable to find the relation which would allow to convert the results from one method to another one and to classify telecommunication chains tested with both methods,
2) to verify whether the relation from point 1. depends or not on the working conditions of the tested telecommunication chain,
3) to measure the approximated relations between the traditional and modified logatom intelligibility methods.

Taking into account the comparative character of the experiment it was planned to be done only in the function of the used logatom intelligibility method. With this end in mind each measurement was done with both traditional and modified methods, not changing:

- listeners,
- surroundings,
- measurement system (only the logatom lists and way of listeners stimulus registration),
- working conditions of the tested chain.

The group of listeners consisted of six people with normal hearing. The classification was done based on the audiometric hearing tests. Before the proper measurements, the listeners were trained for 6 hours (2 measurement sessions, 3 for hours each). The subjective logatom intelligibility measurements were done monaurally with electrodynamics earphones and the optimal speech signal level, i.e. 80 dB(A) [39]. Each measurement was done with one 100-logatom list. The lists were chosen at random.

The logatom intelligibility measurements were done for an analogue chain for various transmission conditions with the telephone channel model MKT-1 designed at the Institute of Telecommunications, Teleinformatics and Acoustics of Wrocław University of Technology [3]. In the model the following interference and distortions can be introduced into the transmitted signal:

1) additive interference whose level is adjusted continuously: white noise, pink noise, hum and impulse interference with four different average values of the random distribution generating a time interval between successive pulses;
2) frequency band limitation – 78 combinations set stepwise in a band of 100–4500 Hz and the so-called full band position; the limit frequency of the high-pass filter can be changed from 100 Hz to 500 Hz in steps of 100 Hz, whereas that of the low-pass filter can be within 2000 Hz to 4500 Hz in steps of 500 Hz; the full-band position represents a bandwidth of 100–6000 Hz;
3) attenuation-diagram linear distortions (9 independently switched on linear equalizers);
4) external interfering signals (intelligible crosstalk, unintelligible crosstalk, etc.); to introduce the interference of this type, a separate input for an adder was included in the channel model.

In the presented experiment, the effect of: white and pink noise in the 300–3400 Hz and 400–2500 Hz band, hum in the 400–2500 Hz band and intelligible crosstalk (for Polish and English) in the 400–2500 Hz was studied. The crosstalk was produced by taping a text in Polish and English, which was read by a speaker and then reproduced during articulation measurements. The reproduced text was fed into the telecommunication channel MKT-1 model through the input assigned for external interfering signals. The speech and interfering signal levels were controlled by means of the 2606 Bruel & Kjaer measuring instrument by measuring them on a logarithmic scale according to the A-weighting curve. The speech signal level and the interfering signal rms values corresponding to the assumed sound intensity values at the listener's ear were measured experimentally [3].

After the setting up the measurement conditions, the logatom intelligibility measurement was done with both the traditional method and with the forced choice one with the modification of the same position (i.e. the operator was changing options in measurement system application). The measurements were done with the system shown in Fig. 1.

The results of logatom intelligibility measurements done with traditional and modified methods in the function of the disturbing noise level for the analogue telecommunication chain are presented in Fig. 2. In this figure the relation between the logatom intelligibility measured by the traditional method and the speech-to-noise given by ISO [33] is presented. A very high compliance of the intelligibility verus signal to noise ratio

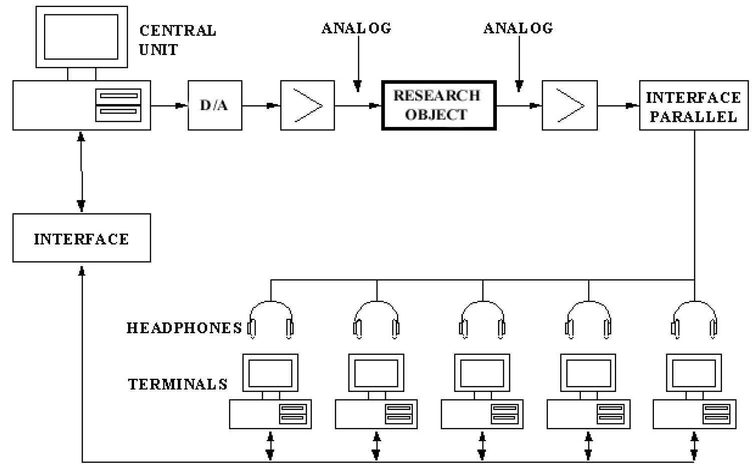

Fig. 1. The measuring system for assessment of logatom intelligibility of telecommunication.

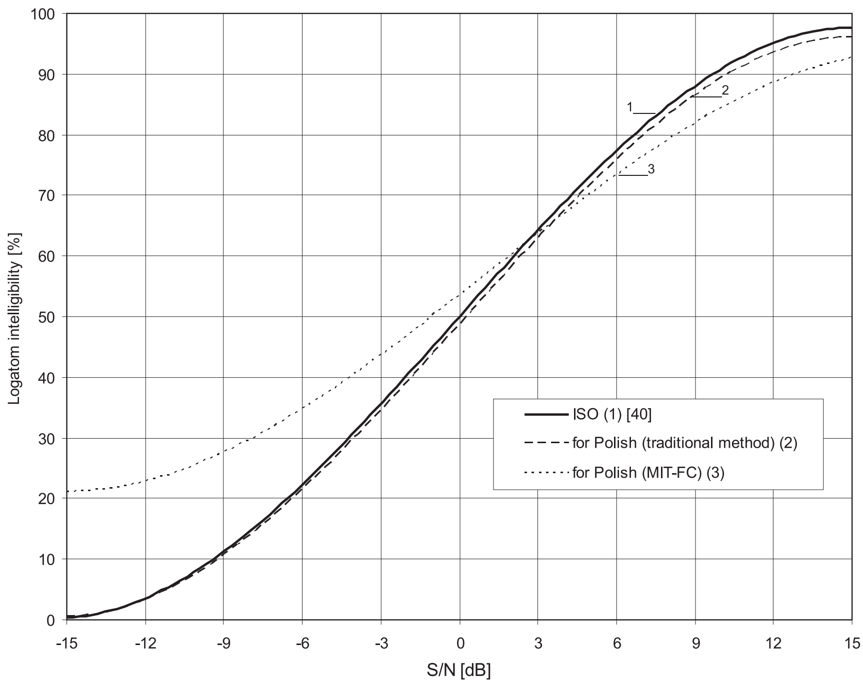

Fig. 2. Relationship between logatom intelligibility and signal/noise ratio

(S/N) is given by the ISO (for the English language) and the curve obtained in the presented experiment (for the Polish language) can be observed. It means that the relationship between the logatom intelligibility obtained by the traditional method and that one with the forced choice is correct either for the Polish language and for English. For signal to noise ratio (S/N) from range $\langle -3$ dB, -15 dB$\rangle$ there is a big divergence between the curve obtained by the traditional method and that one with the forced choice. The divergence is caused by the difference in the idea of the logatom intelligibility measurements in both methods. In the method with the forced choice (MIT-FC), the listener chooses from a limited dictionary, picking 1 from 7; it means that the probability of random choice of the correct logatom equals approximately 17%. The fact that in MIT-FC methods, the listener receives not only the sound but also the visual information which is also important. These factors have an influence on the growth of the logatom intelligibility value in not favourable conditions (high level of distortions). The existence of the disproportion in results obtained with the traditional and forced choice methods is irrelevant to allow the device to be used because it should belong to I or II quality class, which means that the logatom intelligibility should be over 60%.

It was stated that both tested methods gave significant differences in the statistical results and the differences are not constant. In order to find relations between intelligibility obtained with tested methods, all the results were plotted on the logatom intelligibility plane (traditional versus modified method) (Fig. 3). Although the obtained relationship between the logatom intelligibility measured by the tradional method and that one by the forced choice is not linear, the relationship is monotonical in the full range of the logatom intelligibility values. It allows both methods to be used interchangeably. For example, after doing the logatom intelligibility measurements using the method with the forced choice and doing the recalculation of the results – according to the curve presented in Fig. 3 – to the logatom intelligiblity value measured with the traditional method one can define the quality of the device in accordance to Polish [39].

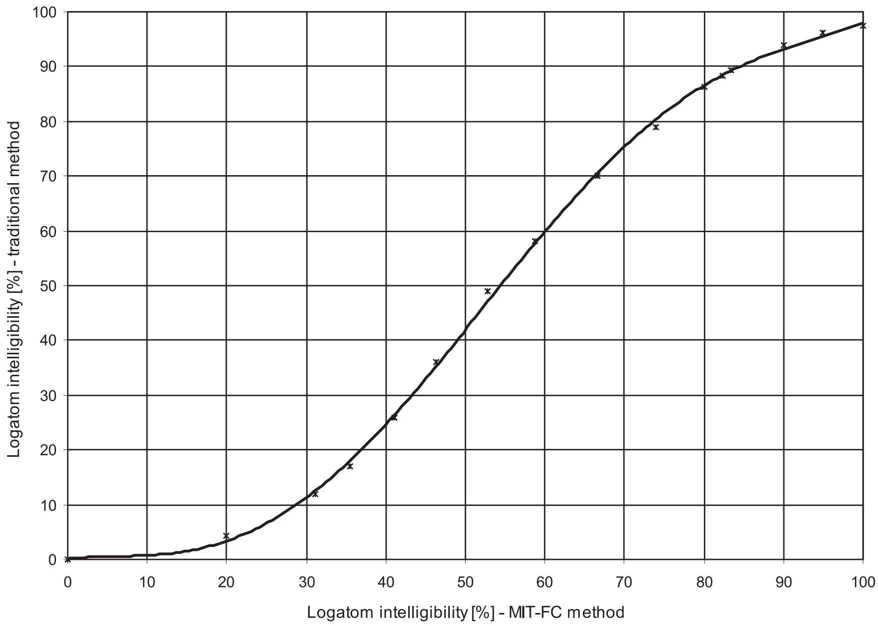

Fig. 3. Relationship between logatom intelligibility measured with traditional and MIT-FC method for analog telephone channels.

4. Conclusions

The experiments carried out to find the relations between the logatom intelligibility measured with the traditional and the semi-automatic with forced choice methods for the analog telecommunication chain have shown that there exists the multi-value and repetitive relation between them. It allows both methods to be used interchangeably and converting the results between them.

The characteristics (line 1 and 2) presented in Fig. 2 show a very close relation between the results obtained for the English and Polish. It can be assumed that the relation between the logatom intelligibility measured with traditional method and the one with forced choice obtained may be used for Polish as well as for English.

In contrast to the aditional measurements the MIT-FC test satisfies the demands of fast and effective intelligibility measurements and exhibits the following features, that are especially useful under field conditions:

1) it permits individual measurements with multiple repetitions for different listeners,
2) the test signal has a stable and repeatable quality under conditions of multiple reproductions,
3) the time needed to carry out the measurement with the MIT-FC method is the same as in the traditional one but we obtain the results right after finishing the measurement process,
4) the scores are obtained just after the measurements are completed,
5) the software applied enables an easy installation of upgradeable applications as a recalculation of the intelligibility scores to the MOS quality scale, the graphical presentation of the results and a comparison of subjective scores with objective predictors,
6) the intelligibility measurements may be carried out under conditions of direct binaural reception of the test material or via headphones (single or double-ear reception).

References

[1] BAŚCIUK K., BRACHMAŃSKI S., *The automation of the subjective measurements of logatom intelligibility*, Proc. 102nd Conv. AES, Munich, Preprint 4407, 1997.

[2] BERGER J., *Future work on objective speech quality measurement in ITU-T*, Proc. Workshop on Wideband Speech Quality in Terminals and Networks: Assessment and Prediction, Mainz 2004.

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

[4] BRACHMAŃSKI S., *Influence of lost packet reconstruction on speech quality in Internet network*, Archives of Acoustics, **30**, 4, 115–118 (2005).

[5] BRACHMAŃSKI S., *Automatic measuring equipment for subjective determination of speech intelligibility* [in Polish], Proc. XLI Open Seminar on Acoustics, pp. 311–314, Wrocław-Szklarska Poręba 1994.

[6] BRACHMAŃSKI S., *Choosing optimum number of test items in subjective logatom measurements* [in Polish], Proc. XLII Open Seminar on Acoustics, pp. 423–428, Warszawa-Białowieża 1995.

[7] BRACHMAŃSKI S., *Assessment of quality of speech transmitted over IP networks*, Internet Technologies, Applications and Societal Impact, WITASI 2002, Wrocław, Kluwer Academic Publishers, 1–14, 2002.

[8] BRACHMAŃSKI S., *Effect of disturbances on speech transmission over internet*, [in:] Proceedings of the 4th European Workshop on Image Analysis for Multimedia Interactive Services, London, World Scientific Publishing, 310–313, 2003.

[9] BRACHMAŃSKI S., *The subjective measurements of speech quality in rooms*, [in:] Proc. of Subjective and Objective Assessment of Sound, Poznań 2004.

[10] DAVIES D.D., DAVIES C., *Application of speech intelligibility to sound reinforcement*, J. Audio Eng. Soc., **37**, 12, 1002–1018 (1989).

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

[12] FRENCH N., STEINBERG J., *Factors governing the intelligibility of speech sounds*, JASA, **19**, 1, 90–119 (1947).

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

[14] KOCIŃSKI J., *Speech intelligibility improvement using convolution blind source separation assisted by denoising algorithms*, Speech Communication, **50**, 29–37 (2007).

[15] KRYTER K.D., *Methods for the calculation and use of the articulation index*, JASA, **34**, 11, 1689–1697 (1962).

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

[17] MAJEWSKI W., BASZTURA CZ., MYŚLECKI W., *Relation between speech intelligibility and subjective scale of speech transmission quality*, Proc. 7th FASE Symposium, Proceedings SPEECH'88, pp. 719–726, Edinburgh 1988.

[18] MAJEWSKI W., MYŚLECKI W., BAŚCIUK K., BRACHMAŃSKI S., *Application of modified logatom intelligibility test in telecommunications, audiometry and room acoustics*, Proceedings 9th Mediterranean Electrotechnical Conference Melecon'98, pp. 25–28, Israel, Tel-Aviv 1998.

[19] MAPP P., *A comparison between STI and RASTI Speech Intelligibility Measurement Systems*, Proc. 111th AES Convention, Preprint 5668, Los Angeles 2002.

[20] MÖLLER S., BOURLARD H., *Analytic assessment of telephone transmission impact on ASR performance using a simulation model*, Speech Communication, **38**, 441–459 (2002).

[21] MYŚLECKI W., MAJEWSKI W., BASZTURA CZ., *Quantitative and qualitative standards of transmission evaluation for Polish*, Proc. 9th Conference on Acoustics, pp. 126–131, Budapest 1988.

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

[23] SOTSCHEK J., *Sprachverständlichkeit bei additiven Störungen*, Acustica, **57**, 257–267 (1985).

[24] SRINIVASAN S.H., *Speech quality measure based on auditory scene analysis*, Proc. IEEE 6th Workshop on Multimedia Signal Processing, pp. 371–374, 2004.

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

[26] SUN L., IFEAACHOR E.C., *Perceived Speech Quality Prediction for Voice over IP-based Networks*, Proc. IEEE International Conference on Communications, ICC 2002, **4**, 2573–2577 (2002).

[27] SUN L., IFEAACHOR E.C., *New Methods for Voice Quality Evaluation for IP Networks*, Proc. of the 18th International Teletraffic Congress, ITC 18, pp. 1201–1210, Berlin 2003.

[28] VORAN S., *Objective estimation of perceived speech quality – Part I: Development of measuring normalizing block technique*, IEEE Trans. on Speech and Audio Processing, **4**, 4, 371–382 (1999).

[29] VORAN S., *Objective estimation of perceived speech quality – Part II: Evaluation of measuring normalizing block technique*, IEEE Trans. on Speech and Audio Processing, **7**, 4, 383–390 (1999).

[30] ŻERA J., *Speech intelligibility measured by adaptive maximum likelihod procedure*, Speech Communication, **42**, 313–328 (2004).

[31] ANSI S 3.2, *Methods for measuring the intelligibility of speech over communication systems*, American National Standard, 1989.

[32] ANSI S 3.5, *Methods for the calculation of the speech intelligibility index (SII)*, 1997.

[33] ISO/TR 4870, 1991, *Acoustics – The construction and calibration of speech intelligibility tests*, Technical Report, 1991.

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

[35] ITU-T, Recommendation P.800, *Method for subjective determination of transmission quality*, 1996.

[36] ITU-T Recommendation P.830, *Method for objective and subjective assessment of quality*, 1996.

[37] ITU-T Recommendation P.862, *Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs*, 2003.

[38] Polish Standard PN-EN-60268-16, *Sound system equipment – Part16: Objective rating of speech intelligibility by speech transmission index*, 2002.

[39] Polish Standard PN-90/T-05100, *Analog telephone chains. requirements and methods of measuring logatom articulation* [in Polish], 1990.

[40] Polish Standard PNV90002, *Digital Communication Systems. Requirements and Methods for Measurement of Logatom Articulation* [in Polish], 1992.