Stefan BRACHMAŃSKI[1]

4. TEST MATERIAL USED TO ASSESS SPEECH QUALITY IN POLAND

4.1. Introduction

As far as subjective, acoustic speech intelligibility is concerned, one can say that the main goal of it is to value and differentiate the quality of acoustic signals reaching the listener. The process of the acoustic perception can be investigated as a reaction of the listener to a stimulus reaching him. The reaction of the listener depends on the type of the stimulus reaching him but also on the condition in which the listener is located. It can be concluded that the listener's reaction depends not only on external factors affecting him at a specific moment in time but also on internal ones. This statement does not take into account the test material used to assess speech quality in Poland's characteristics of the listener, such as cognitive ability, association skills, quick-wittedness, information processing speed, memory, etc., which also influence the listener's reaction. As a result, the listener's reaction to the stimulus shall be considered in the category of a multidimensional function. From the physiological and psychological point of view, the listener's reaction can be investigated in the category of the impression and emotion.

The impression reaction is caused by the stimulus overcoming the predefined sensitivity thresholds or the thresholds of the hearing impression categories. The emotional reaction, which is not the effect of the features of the received signal but coming from the habits and individuality of the listener, is more complex and harder to analyze. It can be stated that the impression reaction is a reflection of the auditory image created in a human mind, while the emotional reaction is a reflection of the relationship of the human to this image. When respectively stable conditions of assessment are assured, it can be expected that the differences in the reaction between multiple listeners are smaller than the differences in the emotional reactions. As a result, one of the basic auditory objectives of the acoustic signal assessment in particular speech is to reduce as much as possible the impact

[1] Department of Acoustics, Multimedia and Signal Processing, Wroclaw University of Science and Technology, Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland, stefan.brachmanski@pwr.edu.pl

of the emotional reaction to the outcome. This can be achieved by introducing appropriately numerous statistics of the assessment, appropriate choice, and training of the listener group, and appropriate formulation of the test exercises. The test exercises include choosing the test material, defining the methodology of the measurement, and the methodology of analyzing the results obtained in the conducted measurements.

The test material used in the subjective assessment of the speech signal transmission can be created from units having the semantic significance or from elements without it. In the case of the former, these elements can be sentences or words, while in the case of the latter – logatoms[2]. A classic article from Fletcher and Steinberg dated 1929 [9] describes the methodology of speech signal quality using various test material including nonsense syllables with consonant–vowel–consonant structure (CVC), sesquipedalian statements, and sentences in the English language. Single syllable phonetic word tests balanced for the English language (American) were normed in 1960 with ANSI 3.2 recommendation, later in 1989 updated with two additional rhyme tests: Diagnostic Rhyme Test (DRT) and Modified Rhyme Test (MRT) [1, 8, 40]. MRT test is a modification of the test previously created by Fairbanks [8]. In the International Telecommunication Union (ITU) recommendations, requirements related to the test signal are the basis for the telecommuni cation application. In this recommendation, the test signal set, containing many typical speech parameters, with different complexity level was presented [15 – 17]. The signals are meant to be used in subjective and objective speech signal transmission assessment.

On the other hand, according to the Polish Standards referring to the intelligibility measurements, it is recommended to use logatom lists [30, 31]. Using the logatom lists is connected with the fact that subjective results of the speech transmission quality shall depend on the maximum level on the physical parameters of the examined telecommunication channel and not on the structure of the language test. The information elimination on the semantic level is guaranteed by logatom lists based on which the logatom and phonetic intelligibility is evaluated. In similar, as all the other test lists used in speech quality measurements also the logatom lists should comprise a representative sample of the Polish language, i.e. should be balanced structurally[3] and phonetically[4].

The phonetic balancing condition means that the percentage share of the individual phonemes from the test list should be compliant with the frequency of the occurrence of them in the Polish language.

On the other hand, the structural balancing condition means that the percentage share of the individual letter or phoneme combinations being part of the test list should match

[2] Logatom – is a string of letters following one after another according to the certain language rules, deprived of any semantic content (nonsense word).

[3] Balanced structurally means that the percentage contribution of individual phoneme connections in test material should be compliant with the frequency of the occurrence of those in the Polish language.

[4] Balanced phonetically means that the frequency of individual phoneme occurrence in the test material should be compliant with the frequency of the occurrence of those phonemes in the Polish language.

the frequency of the occurrence of those combinations in the Polish language and the percentage share of the test units (e.g. logatoms, words) with a given structure should correspond to the frequency of the occurrence of the words having the same structure in the language.

The test lists should assure the possibility of conducting high precision measurements and satisfactory reproducibility of the results with minimal time, effort, and means a loss to perform them. Hence the test lists should conform to the following criteria:

- should be uniform on an adequate level,
- should be easy to read,
- should be easy to verify,
- should be ordered and written in a form that avoids the generation of errors related to the ambiguity of the record,
- should not be too long (related to the tiredness of the listeners).

Looking at the above criteria, it can be stated that creating the tests compliant with all of the requirements is practically impossible. Therefore, when creating the test list, one should rely on the hierarchy of conditions and reasonable compromises [7, 13, 25, 27-29, 32-39].

The test material used in subjective speech signal quality measurements realized under the intelligibility or expressiveness criterion can be created from the units having semantic meaning or from elements without it. In the former, those elements can be sentences or words, in the latter – logatoms [6].

4.2. Subjective methods of speech quality assessment

4.2.1. Absolute Category Rating

The Absolute Category Rating (ACR) method is suggested by ITU Recommendation P.800 [18] for evaluation of the subjective quality of the speech. The speech material used in this method should consist of simple, short, semantically unrelated sentences. Listeners listen to those sentences and give their opinions on five levels scale (5 – excellent, 4 – good, 3 – fair, 2 – poor, and 1 – bad). The average rating MOS (Mean Opinion Score) is calculated over the listeners and the speakers.

4.2.2. Degradation Category Rating

The Degradation Category Rating (DCR) [18] method is an alternative to the ACR method which is not accurate enough for high-quality systems. The sentence lists are the same as in the ACR method. In the DCR method, the listeners try to answer the question „Please rate the degradation of the second sample relative to the first.”. Listeners hear two sentences (original and transmitted) and give their opinions on a five-point scale (5 – not perceived,

4 – perceived but not annoying, 3 – slightly annoying, 2 – annoying, and 1 – very annoying). The average rating DMOS (Degradation Mean Opinion Score) is calculated over the listeners and the speakers.

4.2.3. Comparison Category Rating

The Comparison Category Rating (CCR) method [18] is similar to the DCR. The process of recording and replaying of the list is the same whereas the model and tested samples are played in random order. The A (reference) – B (assessed) pairs are created randomly. The listener aims to compare two samples A and B and to assess if the quality of the first signal in comparison to the second one is the same or different. There is the seven grades scale from 3 to -3 (3 – the quality of the first signal in comparison to the second one is much better 2 – better, 1 – slightly better, 0 – about the same, -1 – slightly worse, -2 – worse, -3 – much worse). The average rating CMOS (Comparison Mean Opinion Score) is calculated over the listeners and the speakers.

4.2.4. Logatom intelligibility

The measurement of logatom intelligibility is based on the transmission of logatom lists (100 – 300 logatoms) phonetically balanced, read out by a speaker, transmitted through the tested channel, which is then written down by listeners. Listeners write received logatoms on special sheets. Received logatoms may be written in phonetic transcription (the listeners must know it) specific for a given language. 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. The average logatom intelligibility is calculated as a proportion between the number of correct received logatoms and all generated logatoms.

Subjective measurements of logatom intelligibility are recommended by the Polish Standards: PN–T–05100 “Analog Communication Systems. Requirements and Methods for Measurement of Logatom Intelligibility” [30] and PN–V–90001 “Digital Communication Systems. Requirements and Methods for Measurement of Logatom Articulation” [31].

4.3. Objective methods of speech quality assessment

4.3.1. Perceptual Evaluation of Speech Quality

The idea of the Perceptual Evaluation of Speech Quality (PESQ) method is based on the so-called internal representation which reflects a theoretical form of the speech signal in a human brain [19]. As a reference signal, the previously recorded male and female voices (one sentence by each voice) are used. Such a prepared original signal is transmitted via

a telecommunication channel being under investigation, and at the output of this channel, this signal is distorted (degraded). Next, these two signals are compared in a psychoacoustic domain which reflects the human impression of speech. The transformation from the physical form into the psychoacoustic representation appears in three stages time-frequency reflection, frequency-critical bank scaling, and scaling of the signal levels. The output value is the PESQ score. The range of the PESQ score is -0.5 to 4.5 [19]. This PESQ score can be transformed into a subjective listening quality MOS – like scale between 1.0 and 5.0, the normal range of MOS values found in an ACR experiment.

4.3.2. Perceptual Objective Listening Quality Assessment

The POLQA (Perceptual Objective Listening Quality Assessment) method described in the ITU–T P.863 recommendation [20] is intended for objective evaluation of the quality of speech transmitted via narrowband (300 Hz to 3400 Hz), wideband (70 Hz to 7 000 Hz), and super wideband (50 Hz to 14 000 Hz) channels. It should be noted that, although the POLQA operates at a sampling rate of 48 000 samples/s in the super wideband mode, it would be a mistake to apply this method to music signals. For musical signals, the standard method is PEAQ (Perceptual Evaluation of Audio Quality) described in recommendation ITU–R BS.1387 [14].

The general idea of the POLQA algorithm is the same as in the PESQ method. Both methods are based on comparing the reference speech signal with the signal degraded from the output of the tested telecommunication channel and creating an internal representation in the human auditory perception model. The difference between the internal representations comes down to the cognitive model used to predict the quality of the degraded speech signal. In the POLQA method (similar to the PESQ method), the quality of the speech signal is given in the MOS – LQO (Mean Opinion Score Listening Quality Objective) scale, which is well correlated with the subjective MOS score (according to the ACR method) obtained by following per under the recommendation ITU–T P.800.

4.4. The test material used in subjective speech signal quality measurements based on the intelligibility criterion

The results of the subjective speech transmission quality measurements shall depend on the maximum level on the physical parameters of the examined telecommunication channel and not on the structure of the language test. The exclusion of information on the semantic level is guaranteed by the logatom lists, based on which the logatom and phoneme intelligibility is evaluated. Single logatom lists can be grouped into sets, e.g. each containing three lists. In the measurements performed at the Department of Acoustics, Multimedia and Signal Processing, Faculty of Electronics, Wroclaw University of Science and Technology

100-logatom lists are used, grouped into 20 sets, each containing three lists. The examplary logatom list was presented in Fig. 1. The logatom lists were ordered according to the calculated frequencies of occurrence of phonemes in the Polish language. In the properly constructed logatom lists the relative frequency of occurrence of an individual phoneme should be determined with the 0.1% precision [3].

The frequency of occurrence of phonemes on individual logatom lists was determined by phonetic transcription based on 37 phonemes [21, 22], calculating it as a ratio between the number of occurrences of individual phonemes from the analyzed list and all phonemes in this logatom list. First, the frequency of occurrence within a single logatom list was evaluated and compared to the frequency of phoneme occurrence in the Polish language. The next stage was to research the phonetic balance in the context of the whole sets – namely doing appropriate calculations summarized for three 100-logatom lists (in total 300 logatoms). For this reason, all the three lists making up one set were treated as a whole. The calculations were done for all the logatom lists, i.e. for 20 sets, each having 3 logatom lists. The number of individual phoneme occurrences in the set, and next to the frequency of phoneme occurrences, the number of occurrences was related to the number of occurrences of individual phoneme among all the phonemes in the set. In Fig. 2 the distribution of the frequency of phoneme occurrences included in the set 1 (list 1, 2 and 3) is presented. In contrast, in Fig. 3 the distribution of the frequency of phoneme occurrences for sets 1 and 3 is compared with the distribution for the Polish language.

SET 3
LIST 3

1 - 10	11 - 20	21 - 30	31 - 40	41 - 50
pry	zoto	bieje	ra	niczo
das	kor	zy	czwaśni	ptemen
cholap	wtąchmi	żej	isza	nia
chu	dzija	tuzja	rewnu	wot
notrwy	jekło	kopsan	żu	sza
miel	lu	rowe	jalerz	ciajmuł
giemar	czo	cza	sido	cia
mo	zjot	dem	przo	dy
żebu	fi	plozu	mnieja	cyga
jażdoś	sywia	czopi	wes	temr
51 - 60	**61 - 70**	**71 - 80**	**81 - 90**	**91 - 100**
nokas	wias	wtydziam	wdeplwe	chrepki
szaku	glyne	rone	ciuso	zane
naść	ry	gą	kięch	oł
oktu	gan	bi	wotre	wnime
sia	streko	grapnyj	mek	jeł
dłacol	comy	szokła	tusta	en
wiąszy	ner	ła	fy	netnie
esa	ciemań	tni	zdyłże	wep
lo	jo	nawdy	su	dzi
wobań	ne	ko	ce	tnic

Fig. 1. The exemplary logatom list used in subjective logatom intelligibility measurements
Rys. 1. Przykładowa lista logatomowa stosowana w subiektywnych pomiarach wyrazistości logatomowej

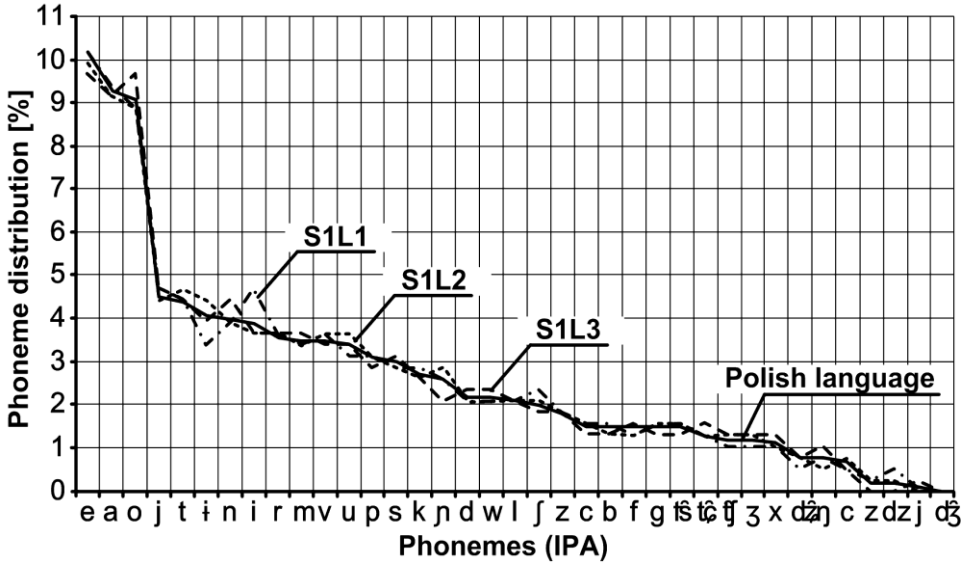

Fig. 2. The frequency of occurrence of Polish phonemes in the natural language and in logatom lists 1, 2, 3 of set 1
Rys. 2. Częstość występowania fonemów języka polskiego w mowie naturalnej oraz w listach logatomowych 1, 2, 3 zestawu 1

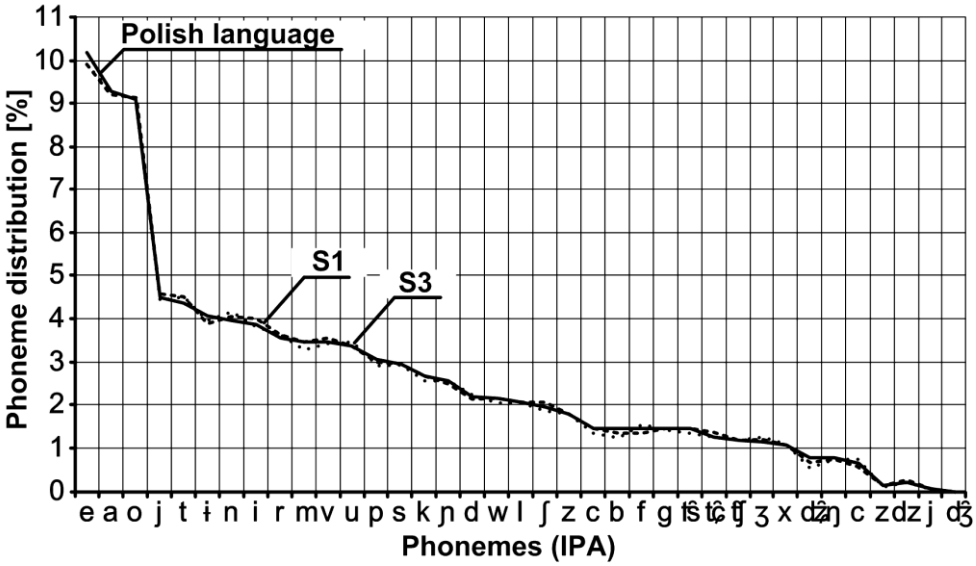

Fig. 3. The frequency of phoneme occurrence for the Polish language in sets 1 and 3
Rys. 3. Częstość występowania fonemów języka polskiego oraz w zestawach 1 i 3

The differences between the frequency of phoneme occurrence given for the Polish language and obtained for the logatom lists exceeded in case of some phonemes the assumptive condition by 0.1% of precision. The hypothesis was raised about the compatibility of the distribution of the frequency of the phoneme occurrence in the Polish language and the analyzed logatom lists. To verify the hypothesis the χ^2 compatibility test was used [41]. The χ^2 compatibility test has shown that at the level of tolerance $\alpha = 0.01$ there is no reason to reject the hypothesis of the compatibility between the distribution of the frequency of the phoneme occurrence.

In many countries, the sentence lists or phonetically balanced word lists are used for speech transmission quality assessment [10, 11, 25, 34, 42]. The sentence lists used in the sentence intelligibility measurements may be the same as in the measurements based on the listener's assessment criteria. In Poland, both types of lists – sentence and word – are mainly used in the audiological research, while less often in the subjective measurements of the speech transmission quality. The sentence lists are created from one or two-syllable words based on the word database from the frequency dictionaries, e.g. the Polish language [12, 23, 24, 35]. The ”inappropriate”, specialist language words, and also the ones related to dialects including student dialects, vernacular, etc are removed from the database.

When creating the sentence lists, the conditions like the numerical amount of words in the lists, phonetic balance, and the level of usage of the most frequent words in the natural language should be considered. The numerical amount of words in the test list is connected with the phonetic balance. According to publications, the ordering of the phonetically balanced list is impossible when the numerical amount is lower than a dozen or so words. The second factor determining the numerical amount is the usage of the word lists, e.g. the test lists used in audiometry should consist of a minimum of 20 words, while the ones used in telephonometry should have at minimum 50 words.

The phonetic balance is related to the usage of the most frequently occurring words in the natural language – namely choosing frequent words causes a significant drop in the phonetic balance. The lack of full balance is especially noticeable when the amount of words in the test list is low. It needs to be compromised between the level of the phonetic balance and the usage of the words most frequently occurring in the natural language. Professor Jassem wrote in [21] that *“for the implementation purposes the 100-word list, which allows for a satisfactory balance of the phoneme numerical amount, seems to be the most beneficial”*

The verification of the phonetic balance is done in analogy to the logatom lists – namely by the ratio between the frequency of the particular phoneme occurrence to all phonemes present in the analyzed word list. The level of the phonetic balance can be checked by using statistical distribution compliance tests in the natural language and a particular word list, e.g. the χ^2 test. In 1993 for the audiometry purposes, the fully balanced articulation lists, consisting of 10 lists each containing 24 words (NLA-93), were worked out [32, 33, 36]. The examplary fragment of the Polish phonetically and structurally balanced one-syllable word list is presented in Fig. 4. In telephonometry, where the compliance between the frequency of the phoneme occurrence in the particular word list and the natural language is an important factor, lists having 50 or 100 words are most frequently used. The ANSI 3.21989 recommendation contains 1000 one–syllable words with the CVC structure (consonant–vowel–consonant) grouped in 20 lists each having 50 words [1]. The words are presented to the listeners with the carrier phrase, e.g. “Now please write down the word...”.

No	List 1	List 2	List 3	List 4	List 5
1	**plac** *(place)*	**dres** *(tracksuit)*	**kwas** *(acid)*	**płaz** *(amphibian)*	**twarz** *(face)*
⋮	⋮	⋮	⋮	⋮	⋮
5	**skurcz** *(shrink)*	**biust** *(breast)*	**ksiądz** *(priest)*	**zrost** *(adhesion)*	**złość** *(anger)*
6	**kat** *(hangman)*	**tak** *(yes)*	**typ** *(type)*	**kit** *(putty)*	**byt** *(existence)*
⋮	⋮	⋮	⋮	⋮	⋮

Fig. 4. The fragment of the word list for the Polish language (NLA-93) [32, 33]
Rys. 4. Fragment listy wyrazowej dla języka polskiego (NLA) [32, 33]

4.5. The word lists used in subjective speech quality measurements based on the listener's assessment

The test lists used in the subjective speech transmission quality measurements carried out according to the speech judgment criteria should be built out of simple and short sentences. The sentences should not contain any expressions or phrases hard to understand or infrequently encountered in the colloquial speech of the language for which those were created, in the considered case for the Polish language. The sentences should be ordered in a way to eliminate any semantic connections between succesive sentences. The duration of the spoken sentence should not exceed 3 seconds.

In the ITU-T P.501 recommendation, two sentences spoken by two women and two men were given for different languages including the Polish language (Fig. 5) [17]. Sentences in the P.501 can be used in objective techniques of the speech signal quality assessment. Unfortunately, for subjective ones this base is too small.

Głos żeński 1 *Female voice 1* Pielęgniarki były cierpliwe. *The nurses were patient* Przebiegał szybko przez ulicę. *He ran quickly across the street.*	Głos męski 1 *Male voice 1* On był czarującą osobą. *He was a charming person.* Lato wreszcie nadeszło. *Summer finally came*
Głos żeński 2 *Female voice 2* Ona była jego sekretarką od lat. *She was his secretary for years.* Dzieci często płaczą kiedy są głodne *Children often cry when they are hungry.*	Głos męski 2 *Male voice 2* Większość dróg było niezmiernie zatłoczonych. *Most of the roads were very crowded* Mamy bardzo entuzjastyczny zespół. *We have a very enthusiastic team.*

Fig. 5. Test sentences for the Polish language according to the ITU-T P.501 Recommendation [17]
Rys. 5. Zdania testowe dla języka polskiego wg zalecenia ITU-T P.501 [17]

Based on the assumptions for the creation of the test lists for the speech signal quality assessment was mapped out in the Department of Acoustics, Multimedia and Signal Processing, Faculty of Electronics, Wroclaw University of Science and Technology a test set

containing 10 lists. Each list was divided into 10 groups each with 5 tasks [7]. The example of one of the created sentence lists is shown in Fig. 6.

The statistical analysis is done analogically to the logatom lists. As an example, the graphical interpretation of the obtained results for one test list was presented in Fig. 7. For the hypothesis about the compliance of the frequency of the phoneme occurrence in the test lists and the Polish language, the t-Student test was used. It was observed that on the assumptive relevance level $\alpha = 0.3$ there are no reasons to reject the hypothesis: *The frequencies of the phoneme occurrence given for the Polish language and calculated for individual lists originate from the same general population*.

List 7	
Grupa 1 *Group 1*	Grupa 6 *Group 6*
1. Na święta jedziemy do babci. *For the holidays we go to Grandma's*.	1. Proszę podać numer telefonu. *Please specify your phone number*
2. Zamiast szynki wezmę dżem. *Instead ham, I'll take jam*	2. Jeden zawodnik został w tyle. *One player was behind*
3. Gdzie jest dział rachunkowości? *Where is the accounting department?*	3. Żaden z elementów nie pasował. *None of the parts fitted*
4. Działy się wtedy różne rzeczy. *Different things happened then.*	4. Nie ma jeszcze potwierdzenia wpłaty. *There is currently no confirmation of the payment*
5. Popracuj nad tym jeszcze trochę. *Work on this some more*.	5. Wyglądał na bardzo zakłopotanego. *He looked very uncomfortable*
...	...
Grupa 5 *Group 5*	Grupa 10 *Group 10*
1. Zioła są dobre na wszelkie choroby. *Herbs are good for all diseases.*	1. Zamówiłem golonkę z musztardą. *I ordered a pork knuckle with mustard*
2. Zmarł małżonek pani prezes. *Ms. president's huseband died*	2. Oni są kompletnie szaleni. *They are completely crazy*
3. Spotkanie było bardzo interesujące. *The meeting was very interesting*	3. Wyniki działalności firmy są obiecujące. *The results of the company are promising.*
4. Reklama przynosi dobre rezultaty. *Publicity brings good results*	4. To był tylko zły sen. *It was just a bad dream.*
5. Urwał rozmowę w pół słowa. *He stopped talking in mid-sentence.*	5. Gdzie chcesz dzisiaj iść? *Where do you want to go today?*

Fig. 6. The exemplary sentence list number 7 for the Polish language
Rys. 6. Przykładowa 7 lista zdaniowa dla języka polskiego

The statistical, structural balance research was done for the same test material (7 test lists) as for the phonetic balance research [4, 6, 7]. The statistics of the interphoneme connections for the Bible text were used as the reference material. It contains around 4 million interphoneme connections and is written in the colloquial language, which makes it reliable as reference material for Polish language. In theory, for the Polish language, 1444 possible

interphoneme connections (for 38 phonemes taking into account the so-called silence phoneme) should be considered; however, after removing the forbidden connections and the ones not occurring in Polish the number can be reduced to around 1000 [2]. The test lists make a quite limited set, which results in the difficulty to obtain the number of interphoneme connections, which would ensure their natural frequency of occurrence. It was stated that the compliance of the structure of the interphoneme connection for the reference text and the 7 test lists is high and at the assumed relevance level $\alpha = 0.5$, there are no reasons to reject the hypothesis: *The frequencies of the phoneme occurrence given for the Polish language and calculated for individual lists originate from the same general population*.

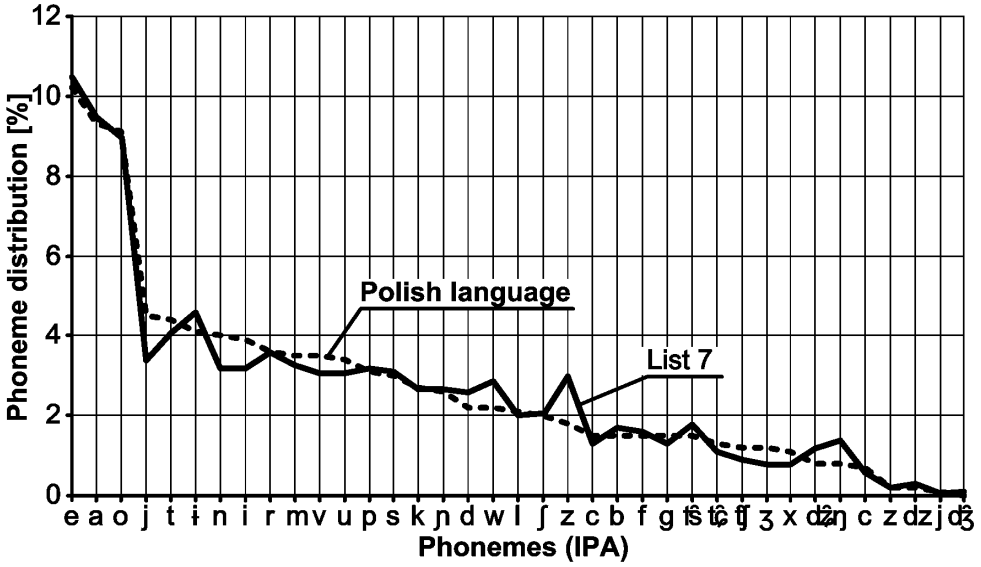

Fig. 7. The frequency of phone occurrence for the Polish language and list number 7
Rys. 7. Częstość występowania fonemów dla języka polskiego i w liście 7

It was observed that between individual discrepancies in the frequency of the given interphoneme connection occurrence sets were minimal, not exceeding 1%. Bigger differences occur between the reference material and logatom sets – namely the differences reach the level of 2.5%. As a result of using the t-Student test, it was observed that at the assumed relevance level $\alpha = 0.5$ there are no reasons to reject the hypothesis – the obtained results calculated for the Polish language and calculated for individual lists originate from the same general population. It can be accepted that the discrepancies between the frequency of the given phoneme connection occurrences in the reference material and sentence lists are negligible.

4.6. Conclusions

The results of statistical analysis of the logatom lists and sentence lists used for Polish speech quality assessment have confirmed that all lists are phonetically balanced.

Nonetheless, it can be observed that in the case of logatom lists the balance of the whole set (three 100-logatom lists) was more complete than in the case of a single logatom list. It should be accepted that in subjective logatom intelligibility measurements for each studied speech signal transmission condition (measurement point), there should be three 100-logatom lists making one measurement set. The relationship between the white noise level and the logatom intelligibility value has been examined using three 100-logatom lists. The results are given in [5, 6].

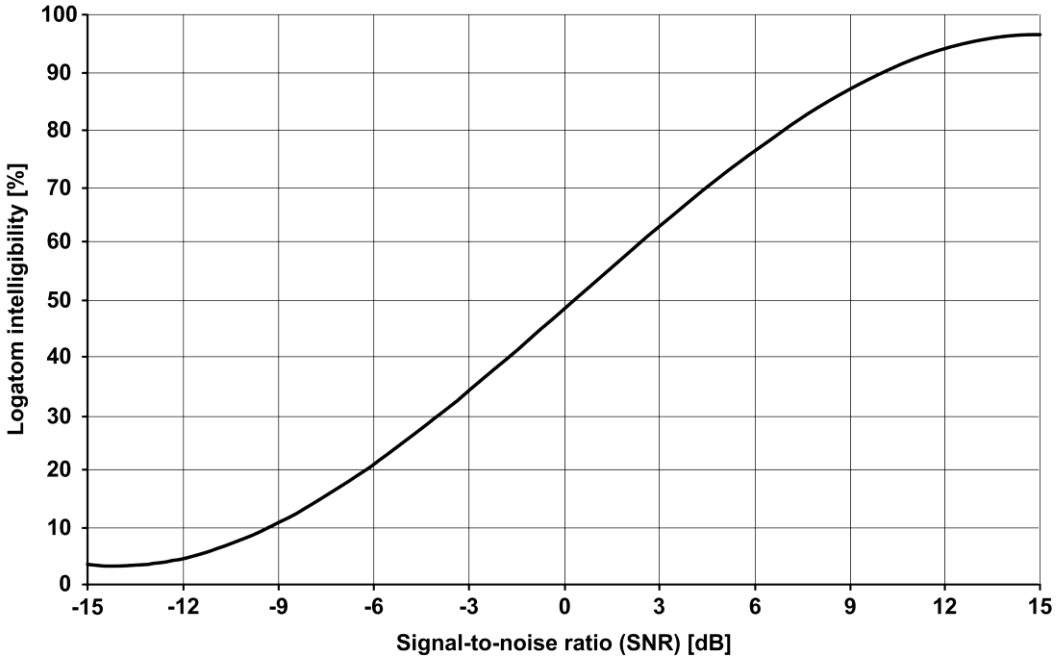

Fig. 8. Relationship between logatom intelligibility and signal-to-white noise ratio (SNR) for the Polish language

Rys. 8. Związek między wyrazistością logatomową, a stosunkiem sygnału do szumu (SNR) dla języka polskiego

The obtained results are compliant with the outcomes of the research done for the English language. Brain Moore wrote: *"...For accurate communication, the average speech level should exceed that of the noise by 6 dB (i.e., the S/N ratio should be +6 dB). When speech and noise levels are equal (0 dB S/N ratio), the word intelligibility score usually reaches about 50%"* [26].

Bibliography

1. ANSI S 3.2: Methods for measuring the intelligibility of speech over communication systems. American National Standards Institute, 1989, Revision ANSI S 3.2–1960 American Standard Method for Measurement of Monosyllabic Word Intelligibility.

2. Basztura Cz., Staroniewicz P.: The diphones statistics of English and Polish speech with relationships to ASR (in Polish). Proceedings XLIV Open Seminar on Acoustics, 1997, 129 – 134.
3. Brachmański S.: A statistic investigation of logatom lists used in subjective measurements of speech intelligibility (in Polish). Proceedings XL Open Seminar on Acoustics, 1993, 329 – 332.
4. Brachmański S.: Phonetic structure of test material used in measurements of speech quality of ACR method (in Polish). Proceedings XLVIII Open Seminar on Acoustics, 2001, 127 – 132.
5. Brachmański S.: Effect of additive interference on speech transmission. Archives of Acoustics, Vol. 27, No. 2, 2002, 95 – 108.
6. Brachmański S.: Chosen problems of speech transmission quality assessment (in Polish). Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław 2015.
7. Brachmański S., Staroniewicz P.: Phonetic structure of test material used in subjective measurements of speech quality (in Polish). Speech and Language Technology, Vol. 3, 1999, 71 – 80.
8. Fairbanks G.: Test of phonemic differentiation: The rhyme test. J. Acoust. Soc. Am., Vol. 30, No. 7, 1958, 596 – 600.
9. Fletcher H., Steinberg J. C.: Articulation Testing Methods. Bell System Technical Journal, Vol. 8, No. 4, 1929, 806 – 854.
10. Haslam V.N.: Psychometrically Equivalent Monosyllabic Words for Word Recognition Testing in Mongolian. Brigham Young University, 2009. http://scholarsarchive.byu.edu/cgi/viewcontent.cgi?article=2735&context=etd (accessed 12.08.2020).
11. Hochmuth S., Brand T., Zokoll M.A., Zenker C.F., Wardenga N., Kollmeier B.: A Spanish Matrix Sentence Test for assessing speech reception thresholds in noise. Int J Audiol. Vol. 51, No. 7, 2012, 536 – 544.
12. Imiołczyk J.: Subjective probability words. Basic polish language attendance dictionary (in Polish), PWN, Warszawa 1983.
13. ISO/TR 4870: Acoustics – The construction and calibration of speech intelligibility tests. International Organization for Standardization, 1991.
14. ITU–T Recommendation BS.1387–1: Method for the objective measurements of perceived audio quality. International Telecommunication Union, Geneva 2001.
15. ITU–T Recommendation P.59: Artificial conversational speech. International Telecommunication Union, Geneva 1993.
16. ITU–T Recommendation P.50: Artificial voice. International Telecommunication Union, Geneva 1999.
17. ITU-T Recommendation P.501: Test signals for use in telephonometry. International Telecommunication Union, Geneva 2009.

18. ITU-T Recommendation P.800: Method for subjective determination of transmission quality. International Telecommunication Union, Geneva 1996.
19. 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. International Telecommunication Union, Geneva 2001.
20. ITU-T Recommendation P.863: Perceptual objective listening quality assessment. International Telecommunication Union, Geneva` 2011.
21. Jassem W.: Fundamentals of phonetic acoustics (in Polish). PWN, Warszawa 1973.
22. Jassem W.: Speech and communication science (in Polish). PWN, Warszawa 1974.
23. Jassem W.: Polish phonetically balanced occurrence-frequency weighted word lists (in Polish). Speech and Language Technology, Vol. 1, 1997, 71 – 99.
24. Jassem W., Gembiak D.: Subjective probability of Polish word (in Polish). PWN, Warszawa 1980.
25. Kryter K.D., Whitman E.C.: Some comparisons between rhyme and PB-Word intelligibility tests. J. Acoust. Soc. Am. Vol. 37, No. 6, 1965, 1146.
26. Moore B.C.J.: An Introduction to the Psychology of Hearing. Koninklijke Brill NV, Leiden 2013, www.ee.iitb.ac.in/~spilab/material/literature/An Introduction to the Psychology of Hearing by Brian Moore 6th Edition.pdf (accessed 10.08.2020).
27. Ozimek E., Kutzner D., Sęk A.P., Wicher A.: Development and evaluation of Polish digit triplet test for auditory screening. Speech Communication, Vol. 51, No. 4, 2009, 307 – 316.
28. Ozimek E., Kutzner D., Sęk A.P., Wicher A., Szczepaniak O.: The Polish sentence test for speech intelligibility measurements. Archives of Acoustics, Vol. 31, No. 4, 2006, 431 – 438.
29. Ozimek E., Warzybok A., Kutzner D.: Polish sentence matrix test for speech intelligibility measurement in noise. International Journal of Audiology, Vol. 49, No. 6, 2010, 444 – 454.
30. Polish Standard PN-90/T-05100: Analog telephone chains. Requirements and methods of measuring logatom articulation (in Polish). Polski Komitet Normalizacyjny, Warszawa 1990.
31. Polish Standard PN-V-90002: Digital Communication Systems. Requirements and Methods for Measurement of Logatom Articulation (in Polish). Polski Komitet Normalizacyjny, Warszawa 1992.
32. Pruszewicz A., Demenko G., Richter L., Wika T.: New articulation lists for speech audiometry (in Polish). Otolaryng. Pol., Vol. 38, No. 1, 1994, 50 – 55.
33. Pruszewicz A., Surmanowicz-Demenko G., Jastrzębska M.: Polish tests for speech audiometry (in Polish). Wydawnictwo Naukowe Uniwersytetu Medycznego im. Karola Marcinkowskiego, Poznań 2011.

34. Puglisi G.E., Astolfi A., Prodi N., Visentin Ch., Warzybok A., Hochmuth S., Kollmeier B.: Construction and first evaluation of the Italian Matrix Sentence Test for the assessment of speech intelligibility in noise. Forum Acusticum, Kraków 2014.
35. Richter L.: Remarks on the lexical bases of vocabularies for an acoustic-phonetic Polish database (in Polish). Speech and Language Technology, Vol. 1, 1997, 35 – 50.
36. Schelenzen A., Skrodzka E.: Evaluation of New Polish Articulation Lists (NAL-93) in the Presence of Various Speech-Like Maskers. Archives of Acoustics, Vol. 45, No. 3, 2010, 393 – 400.
37. Śliwińska-Kowalska M., Kotyło P., Soli S.S.: Development of Polish language Hearing in Noise Test (in Polish). Otorynolaryngologia, Vol. 12, No. 4, 2013, 176 – 182.
38. Voiers W. D.: Diagnostic evaluation of speech intelligibility, [in:] M.E. Hawley, (ed.): Speech intelligibility and speaker recognition. Dowden, Hutchinson, and Ross, 1977, 377 – 387.
39. Voiers W.D.: Diagnostic Acceptability Measure for Speech Communication Systems. ICASSP-77, IEEE International Conference on Acoustics, Speech, Signal Processing, Hartford, 1977, 204 – 207.
40. Voiers W.D.: Evaluating Processed Speech Using the Diagnostic Rhyme Test. Speech Technology, Vol. 1, No. 4, 1983, 30 – 39.
41. Volk W.: Applied Statistics for Engineers (in Polish). WNT, Warszawa 1973.
42. Wagener K., Josvassen J.L., Ardenkjaer R.: Design, optimization, and evaluation of a Danish sentence test in noise. Int J Audiol. Vol. 42, No. 1, 2003, 10 – 17.