ARCHIVES OF ACOUSTICS
Vol. **43**, No. 4, pp. 603–612 (2018)

DOI: 10.24425/aoa.2018.125154

Speech Reception Thresholds for Polish Language Word and Sentence Tests Presented in Noise

Anna SCHELENZ, Ewa SKRODZKA

Faculty of Physics
Institute of Acoustics
Adam Mickiewicz University
Umultowska 85, 61–612 Poznań, Poland; e-mail: anna.schelenz@amu.edu.pl

(*received March 23, 2018; accepted June 26, 2018*)

The aim of the study was to determine the signal-to-noise ratio (SNR) for the Speech Reception Threshold (SRT) for young persons with normal hearing. The following three tests available for Polish language were used: the New Articulation Lists (NAL-93) version of 2011, the Polish Sentence Test (PST) and the Polish Sentence Matrix Test (PSMT). When using PST and PSMT the masking signal was *babble noise* made of the language material contained in the test. For NAL-93 the masking signal was *speech noise*. The speech reception threshold (SRT) was found to be (-6.8 ± 1.1), (-4.8 ± 1.6), (-3.5 ± 1.8) and (-3.4 ± 2.0) dB SNR for PST, PSMT, NAL-93 (constant stimuli method) and NAL-93 (short method), respectively. The values of SRT depend on semantic redundancy of the language material. Differences in SRT were statistically non-significant only for NAL-93 (constant stimuli method) and NAL-93 (short method). Moreover, it was shown that the time needed for presentation of a single word list (NAL-93, short method) or single sentence list (PST, PSMT) was comparable and equal to 2–3 minutes. The most uniform SRT values were obtained for PST. The PSMT was the least demanding for the listener, experimenter and equipment.

Keywords: speech reception threshold in noise; Polish Sentence Test; Polish Sentence Matrix Test; new articulation lists.

1. Introduction

The speech sounds, permitting effective share of information and capable of expressing feelings, needs and emotions, are the main tool of communication. Speech signals display diversity in the domain of time and frequency and, at the same time, are resistant to changing acoustic conditions (Darwin, 2010). These features are vital for its everyday use and that is why speech signals have been used for diagnostics of hearing loss.

The following things are important:

1) the use of speech signals in the tests permits generation of conditions resembling everyday communication and thus it allows evaluation of the actual problems with hearing,
2) speech signals can be used in diagnostics of auditory problems related not only to hearing sensitivity (absolute hearing thresholds), but also to speech intelligibility (disturbances and distortions of signals presented on supra-threshold levels),
3) tests based on speech signals are the basis for diagnostics of persons who have problems with comprehension of speech against background noise (Wilson, McArdle, 2005; Moore, Skrodzka, 2002; Habasińska *et al.*, 2018).

For the above reasons, the masking signal was introduced in the tests based on recognition of speech signal. Its use allows determination of the signal-to-noise ratio (SNR) at which a given person is able to recognize correctly 50% of the presented speech material, i.e. determination of the Speech Reception Threshold (SRT) (Kociński *et al.*, 2014). Such an approach permitted a more effective diagnostics of hearing loss that is not always possible with the pure-tone audiometry, which has been confirmed by many authors. For instance, McArdle *et al.* (2005a; 2005b) have shown that the persons with hearing loss were able to understand speech signals presented in silence in at least 80%, which is a result for normal hearing persons, while when speech signals were presented in the presence of background noise, for the same persons SRT was by 6.5 dB higher than for the persons with normal hearing. This experiment confirmed that testing of speech recognition presented in masking conditions permits detection of the hearing loss not detectable by standard diagnostic methods (the pure-tone audiometry, the speech audiometry in silence; McArdle *et al.*, 2005a; 2005b). Additional benefits of the tests presented in masking noise are the possibility of carrying out correct prosthetic procedure, definition of actual aims that can be reached with a given hearing aid and hearing aid benefit (Wilson *et al.*, 2007b), and detection of limitations in perception of speech signals in certain acoustic conditions (Brungart *et al.*, 2017).

Despite the well-documented benefits of tests based on the use of speech signals against background noise, they have been rarely used in diagnostic work. The reasons for this include the lack of adequate training of the staff being accustomed to certain procedures used for a long time and reluctance to changes in this procedure (different diagnostic material and different methods of interpretation of the results; Wilson, McArdle, 2005). Moreover, the longer time of diagnostics is discouraging. Till recently another problem was the lack of standardization of speech signals permitting analysis of speech recognition in the presence of background noise (Wilson, 2003). At present there are directives on uniformization of testing material for many languages (Akeroyd *et al.*, 2015). Although much progress has been made in this field, different tests are still used in speech audiometry. The differences are related to the language material (words, sentences), semantic redundancy of the test, choice of a speaker and technique of recording, method of test performance, level of signal presentation and type of background noise (Ozimek *et al.*, 2007; Jansen *et al.*, 2012).

For Polish language a few tests for evaluation of speech recognition presented in noise or in quiet have been prepared: the Polish Sentence Test (PST; Ozimek *et al.*, 2009), the Polish Sentence Matrix Test (PSMT; Ozimek *et al.*, 2010) and the New Articulation Lists (NAL-93; Pruszewicz *et al.*, 2011). PST and PSMT contain sentences that can be presented against background noise (so-called *babble noise*) generated on the basis of all sentences occurring in the test (see Fig. 1). Test NAL-93 is composed of monosyllabic words and has been devised for the use without masking noise. In our experiment the NAL-93 test was masked with *speech noise* that is a broadband speech-shaped noise implemented as a standard in audiometers (according to the norm IEC 60645-1:2017). Each test contained a different number of lists of words or sentences, the tests also differed in the method of obtaining results. For the sentences (PST, PSMT) the adaptive method was used, whereas for words (NAL-93) the results were obtained by the constant stimuli method. For all tests (PST, PSMT, NAL-93), the result was the value of SRT expressed in terms of SNR.

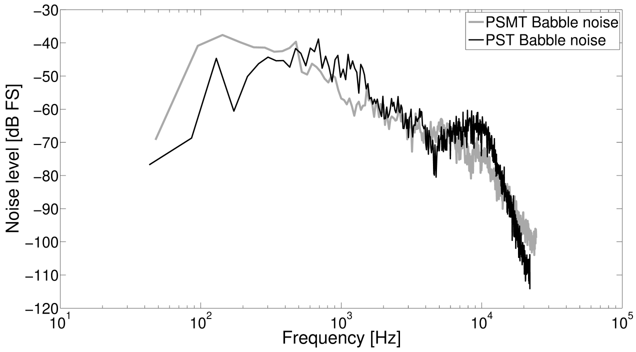

Fig. 1. The *babble noise* spectrum for PST and PSMT.

The main aim of this study was to determine to what extent these three different types of Polish speech-in-noise tests (sentence, matrix, word) yield similar SRTs in the presence of masking noise. Moreover, the comfort of experiment for the subject and for the experimenter was evaluated and duration as well as reproducibility of results obtained for particular lists from each test were determined.

2. The method

2.1. Speech and noise material

The study was performed for three Polish tests: the Polish Sentence Test (PST), the Polish Sentence Matrix Test (PSMT) and the New Articulation Lists (NAL-93) version of 2011. In all tests the speaker was a man. Masking noise was introduced at 300 ms prior to the test signal, the rise time and decay time were 20 ms.

PST contains *everyday sentences* taken from literature, TV, theater, e.g. "firma zatrudnia dwie osoby". The test comprises of 37 sentence lists and each list contains 13 highly redundant sentences. PST was presented in *babble noise* as it is the most effective masker for Polish speech tests. It was generated by a multiple overlapping of all sentences included in the lists, after their earlier modification (some sentences were shifted or reversed in the time domain to obtain a 10-sec masking signal). The study was performed by the adaptive 1-up/1-down method; depending on the listener response the level of the next presentation was increased or decreased, at the constant level of the masker. The response was accepted when the whole sentence was recognized correctly (Ozimek *et al.*, 2009). SRT was calculated from 3 last turning points, including the so called 'virtual point' (the point calculated according to the last answer).

PSMT was made in the form of a matrix composed of 5 columns of 10 words in the order of name-verb-numeral-adjective-object (see Table 1). The sentences were generated at random by choosing one word from each column. The 10-sentence lists obtained in this way were semantically unpredictable and therefore they were less redundant than the everyday sentences from PST. The speech material in PSMT was presented in masking *babble noise* that authors of PST obtained by multiple overlapping of the sentences generated on the basis of the matrix. PSMT was presented by the adaptive 1-up/1-down method, the level of the desired signal was changed in response to the listener's answer. The answer was assumed to be correct if all words in the sentence were correctly recognized (Ozimek *et al.*, 2010). Similarly as in PST, the SRT parameter was determined on the basis of the 3 last turning points, including the 'virtual point'.

Table 1. The 50-word matrix used in PSMT.

Name	Verb	Numeral	Adjective	Object
Tomasz	nosi	pięć	dobrych	piłek
Paweł	woli	sześć	tanich	gazet
Adam	widzi	siedem	drogich	soków
Maciej	bierze	osiem	pięknych	dzwonów
Michał	daje	dziewięć	nowych	opon
Anna	ma	dużo	starych	stołów
Ewa	robi	sto	białych	klocków
Maria	kupi	tysiąc	żółtych	toreb
Zofia	wygra	wiele	czarnych	okien
Julia	sprzeda	kilka	dziwnych	koszy

NAL-93 is the word test of low-redundancy and is composed of monosyllabic words (Pruszewicz *et al.*, 2011). The test is divided into 10 lists of 20 words each. The study was performed by the method of constant stimuli; the desired signal was presented at the fixed SNR values permitting speech recognition higher and lower than 50%. The SRT parameter was obtained by interpolation of SNR values for which the speech recognition in noise was a little above and little below 50%. NAL-93 is the most popular test of monosyllabic words in Poland. In our experiment, NAL-93 was masked with broad-band speech-shaped noise, so-called *speech noise*, composed of frequencies in the range of 125–6000 Hz and the signal decay 12 dB/oct for the frequencies higher than 1 kHz (IEC 60645-1:2017). The use of *speech noise* as a masking signal for NAL-93 instead of *babble noise* needs some clarification. At the stage of experiment preparation, *babble noise* similar to standard masking signals used in PST and PSMT (Ozimek *et al.*, 2009; 2010) was generated for NAL-93, and its effectiveness in masking of monosyllabic words was evaluated. For NAL-93 masked by *speech noise* and *babble noise*, the SRT values determined by the short method (Subsec. 2.2) were -3.4 ± 2.0 dB SNR and $+3.0 \pm 2.4$ dB SNR, respectively. One-way ANOVA ($F(1, 18) = 42.67$; $p < 0.001$) confirmed a statistically significant difference in SRT values obtained for two different masking noises (dependent variable – SRT value, fixed factor – noise). A high positive value of SRT for NAL-93 masked with *babble noise* should be noted. Although this value indicates a high effectiveness of *babble noise* generated from the speech material contained in NAL-93, its use in audiological measurements rises doubts that follow from the necessity of using NAL-93 signals at presentation level even by 10 dB higher than the values specified in the norm (PN-EN ISO 8253-2, 2010). Thus, if SRT for persons with normal hearing was close to 7 dB SNR, even higher values should be expected for the persons with problems with speech recognition in the presence of noise. It would imply the need to present the test at a level of 80 dB SPL or higher, which would increase the effect of other factors not directly related to the speech comprehension in the presence of noise (e.g. "roll-over" of the psychometric function for some types of hearing loss; Wilson, McArdle, 2005). Besides, in contrast to PST and PSMT, there is no *babble noise* attached to the carrier for NAL-93. Taking into account the above, we decided that for the experiment described as well as for audiological measurements it is not necessary to use a special type of noise, different from hitherto used *speech noise*.

2.2. Subjects

The listeners were a group of 10 persons, from 22 to 32 years of age. Young persons were asked to take part in the experiment in order to eliminate the effect of age-related factors (Füllgrabe *et al.*, 2014). The participants had no prior knowledge of the tests and were not paid for taking part in the study. Prior to the experiment, the medical interview, otoscopy, the pure-tone audiometry and the speech audiometry in quiet were carried out. Results of the interview and otoscopy revealed no contraindications as to participation in the experiment. The PTA (Pure-Tone Average) did not exceed 25 dB HL and the recognition of monosyllabic words was not lower than 90% at the presentation level of 65 dB SPL (WHO/PDH/97.3; Martin, Clark, 2009).

3. Measuring procedure

Measurements were performed in an acoustically insulated audiometric laboratory at the Institute of Acoustics, Adam Mickiewicz University, Poznań. The pure-tone audiometry and NAL-93 were generated via Interacoustics AC40 audiometer, while the sentence tests were carried out on a Lenovo IdeaPad 700 laptop with earphones Sennheiser HD 600. The measurements were carried out using the special software written for the Institute (Sęk, 2016) that permitted the choice of list and initial parameters for each listener. The adaptive method was used and the change in the presentation level depended on the recognition of the whole sentence. The measurements were made for the better ear of each participant.

At the first stage of the experiment, NAL-93 was presented by two methods:

1) the constant stimuli method in which a few word lists were presented and for each of them a different SNR was chosen, and
2) the short method in which SNR decreased at a step of 5 dB for each subsequent 5 words from a list.

The second method permitted limitation of the material presented to at most 2 lists. The measurement was finished when the participant was not able to repeat any of the 5 subsequent words at the same presentation level.

In the method of constant stimuli, SRT was determined in the way described in Subsec. 2.1 (PN-EN ISO 8253-2:2010). In the short method, SRT was calculated based on the Spearman-Kärber Eq. (1) (Finney, 1952):

$$S = l + \frac{d}{2} - \frac{d \cdot n}{w}, \tag{1}$$

where S – the signal level at which the speech recognition was 50% [dB SPL], l – the lowest level of signal presentation at which the speech recognition was 100% [dB SPL], d – the step at which speech signal level was changed [dB], w – number of words presented at a given signal level, n – number of correctly recognized words from the list.

Substituting the values $d = 5$ dB, $w = 5$ words, and taking into account the level of masking signal N [dB SPL], Eq. (1) was transformed to:

$$\text{SRT} = S - N = \left(l + \frac{5}{2} - n\right) - N. \tag{2}$$

The use of the short method permitted limitation of the presented material to 1–2 words lists, which enabled to determine SRT parameter in a shorter time than by the constant stimuli method. The aim of the experiment was to check if the two methods provide the same results, so if it is possible to shorten the measurements with no compromise to the results. The noise level for all participants was 65 dB SPL, the initial level of the desired signal was 70 dB SPL and it was changed at a step of 5 dB.

SRT was determined for PST and PSMT in the way described in Subsec. 2.1, on the basis of the average from the last three turning points. For PST lists no. 5 and 8 were selected at random. The participants were asked to indicate the words they had heard in the sentence from those presented on the screen. The initial level of the speech signal was 68 dB SPL and for the first 4 sentences it was changed at a step of 2 dB, while for the subsequent ones at a step of 1 dB. The noise level was 70 dB SPL. The level values of noise and speech signals were chosen on the basis of the study by Ozimek *et al.* (2009).

PSMT test was similar to PST, the difference was that the words presented on the screen were replaced with a matrix made of 5 columns of 10 words on the basis of which subsequent sentences were generated. The initial level of speech and noise presentation was the same, of 65 dB SPL, as recommended by Ozimek *et al.* (2010). The lists were chosen at random by the software and each participant was asked to listen to two of them. No preliminary training in which the participants could listen to the material was made, which is a deviation from the procedure recommended for the matrix type tests (Jansen *et al.*, 2012). However, such a decision was motivated by a routine of clinical testing, where testing time restrictions prohibit sufficient training.

4. Results

The SRT values obtained in the NAL-93 test determined by the method of constant stimuli and the short method are given in Table 2. The mean SRT value obtained for all participants for the constant stimuli method was −3.5 dB SNR (SD = 1.8 dB SNR), while for the shortened method −3.4 dB SNR (SD = 2.0 dB SNR). These values are not in agreement with the value of −5.4 dB SNR given in literature (Lorens *et al.*, 2006), which was confirmed by Student's t-test ($t(9) = 3.38$, $p = 0.01$ for the constant stimuli method

Table 2. SRT values obtained for NAL-93 using the constant stimuli method and short method.

Subject	SRT [dB SNR]	
	Constant stimuli method	Short method
1	−4.0	−3.0
2	−6.0	−6.0
3	−2.0	−3.0
4	−3.0	−1.0
5	−1.0	−1.0
6	−5.0	−2.0
7	−2.0	−2.0
8	−2.0	−4.0
9	−6.0	−6.0
10	−4.0	−6.0
Mean	−3.5	−3.4
SD	1.8	2.0

and $t(9) = 3.15$, $p = 0.01$ for the short method). The differences may come from different acoustic conditions in which the tests were carried out and from no information on the type and level of noise used by LORENS *et al.* (2006).

In order to verify if the method of constant stimuli and the short method provide equivalent results, the one-way analysis of variance (ANOVA) was performed. In the Tukey's test, $F(1, 18) = 0.01$ was obtained at the level of significance of $p = 0.91$, so no statistically significant differences between the results provided by the two methods were found (dependent variable – SRT value, fixed factor – method).

Table 3 presents the SRT values for PST and PSMT, obtained by the participants for the two lists from each test, the mean value and standard deviation. One-way ANOVA revealed statistically significant differences between the reception threshold obtained for PST and PSMT ($F(1, 38) = 19.69$; $p < 0.001$; dependent variable – SRT value, fixed factor – test). This difference can be explained by taking into account that

1) the range of SRT for a single participant was lower than 2.5 dB for PST and over 4 dB for PSMT, and
2) SRT values obtained for PSMT were by about 2 dB higher than those obtained for PST.

Table 3. SRT values obtained for PST and PSMT for two selected lists from each test.

Subject	SRT [dB SNR]			
	PST (list 5)	PST (list 8)	PSMT (list 1)	PSMT (list 2)
1	−5.3	−6.3	−6.7	−5
2	−7.7	−5.3	−6.3	−4.7
3	−7.3	−6.7	−2.3	−5.7
4	−4.3	−6.3	−4.3	−2.7
5	−7.0	−7.3	−4.3	−2.3
6	−8.0	−8.7	−3.3	−7.7
7	−6.7	−7.3	−3.7	−4.0
8	−7.0	−7.3	−5.0	−5.0
9	−7.7	−7.7	−6.7	−8.0
10	−5.3	−6.0	−4.3	−4.0
Mean	−6.6	−6.9	−4.7	−4.9
SD	1.2	1.0	1.5	1.9

5. Discussion

5.1. NAL-93

The possibility of using Eq. (2) to obtain SRT parameter for NAL-93 permits a significant shortening of the time of testing as only one or two word lists need to be presented. The language material in NAL-93 is much limited. A smaller number of lists that need to be presented permits a repeated use of NAL-93 as the participants have no chance to remember the speech material, i.e. there is no effect of training (NIEWIAROWSKI, 2013). The one-way ANOVA results for all the NAL-93 material proved that SRT value is independent of the choice of list ($F(7, 72) = 2.12$, $p = 0.05$; dependent variable – SRT value, fixed factor – word list). Analysis of the mean value and SD obtained for a given list (list no. 7) from NAL-93 revealed greater variety of results between the participants than for the sentence tests (see Table 4). During the experiment we noted a few issues that seem to question the accuracy of the NAL-93 test:

1) Upon presentation of the material (list no. 1), both in quiet and in *speech noise*, some of the words were incorrectly recognized even at a high sound level (65 dB SPL). It was noted in particular for the words /plus/ and /pit/ which were understood as /blus/ and /bit/. The reasons can be the lack of earlier acquaintance with the material (which is the condition of using NAL-93), similarity of the misrecognized sounds (voiceless pronunciation and low energy), poor quality of recording and specific character of the speaker's voice. In case of using a classical constant stimuli method it was not a problem, but it significantly adversely affected the use of the short method.
2) The use of the short method for NAL-93 leads to disturbances to acoustic, phonetic and structural equilibration of the speech material. The lists of words in NAL-93 were generated so that each of them contained the same acoustic-phonetic structures (JASSEM, 1973). Presentation of a complete list at a fixed sound level maintains the features of equilibration. When the intensity of the stimulus within the same list was changed (needed for the use of Eq. (2)), at each sound level only a part of the structures was presented. Although it could affect the speech comprehension, analysis of the results did not confirm this possibility.
3) The influence of experimenter they had on the results was significant, in particular when the participant whispered the words (his/her response) or when the participant gave a few versions of the word they had just heard. The experimenter had to interpret the response and thus influenced the result (1 to 2 words per list). The problem

Table 4. SRT values averaged over all participants obtained for Polish language tests (NAL-93, PST and PSMT).

Test	SRT [dB SNR]	
	Mean	SD
NAL-93	−3.4	2.0
PST	−6.8	1.1
PSMT	−4.8	1.6

could be eliminated if the participants were asked to write or choose their responses from a given list of words.

5.2. PST

The usefulness of PST comes from the simplicity of examination (test procedure and SRT determination), short time of measurement and reproducibility of results for different lists. In contrast to NAL-93, the PST results were significantly dependent on cognitive factors. In the experiment we used a panel containing a set of words from which the participants were asked to choose the words they heard. Although it facilitated the collection of results, it demanded greater engagement of the listener. The participants complained that the necessity of finding the right words in the panel made it more difficult for them to memorize the whole sentence. Frequent mistakes in responses were caused by attention distraction and the appearance of words of similar spelling on the screen, e.g. /groszy/ and /gorszy/. Another problem was related to grammatical mistakes, e.g. /tą książkę/ instead of /tę książkę/). In order to avoid the effect of the mistakes on determined SRT, the role of the experimenter was essential as they controlled the procedure of the choice of words that could influence the SRT value.

5.3. PSMT

The results obtained for PSMT were also significantly affected by the cognitive factors:

1) the participants had problems with remembering the whole sentence because of the necessity of checking up each word in the matrix,
2) the use of one set of words for the whole sentence list caused problems following from mixing up the contents of subsequent presentations,
3) the cognitive neutrality of the sentences was questioned as certain elements of the test aroused some associations, distracting the attention from the sentence.

An important problem was the lack of dynamic equilibrium between the subsequent words in the sentence (the first word was evaluated as less loud than the other words). A possible reason for lower reproducibility of the SRT for PSMT for individual participant (in comparison with PST) could be the number of sentences presented. Determination of SRT by the adaptive method is related to the necessity of obtaining of at least three turning points. As the material in the PSMT test was rather modest (10 sentences in a list), each mistake had a significant effect on the final result. In order to be able to make effective use of the adaptive method we tried to resolve this problem. One of the proposed solutions was a greater change in SNR for subsequent sentences. However, it was related to a greater scatter of SRT values. Another proposition was to decrease SNR for the first sentence in order to get SRT quicker. This solution was also ineffective because of incorrect responses of some participants already for the first sentences in the list. We finally decided to double the number of sentences in the list, similarly as it has been proposed by Ozimek *et al.* (2010), which permitted obtaining a greater number of turning points. Although this solution extended the time of examination, it ensured determination of a more accurate SRT value.

5.4. Comparative analysis

In order to verify the usefulness of Polish language tests, their comparative analysis was made. The analysis concerned the SRT values and homogeneity, the effect of redundancy on the results, simplicity and time of experiments. The results from each pair of tests were subjected to one-way ANOVA analysis. In all comparisons the analysis revealed statistically significant differences between the results ($F(1, 28) = 35.67$; $p < 0.001$ for NAL-93 and PST; $F(1, 28) = 4.16$; $p = 0.05$ for NAL-93 and PSMT; $F(1, 38) = 19.69$; $p < 0.001$ for PST and PSMT; dependent variable – SRT value, fixed factor – test)[1]. The results of ANOVA mean that for correct evaluation of the speech reception threshold for Polish language it is necessary to take into account the SRT parameter together with the test used for its determination. The results obtained for NAL-93, PST and PSMT should be interpreted independently.

Table 4 presents the SRT values averaged over all participants for particular tests. The results are strongly correlated with redundancy of the materials analyzed. The greatest scatter of SRT values (SD = 2 dB SNR) was obtained for the low-redundancy monosyllable word test NAL-93, in which the recognition of words was hindered by the lack of context and lack of training prior to the examination. The results of PSMT, in which the understanding of sentences was established on the basis of the choice of words from a matrix, were characterized by higher uniformity (SD = 1.6 dB SNR). The lowest standard deviation (SD = 1.1 dB SNR) was obtained for highly redundant PST, which was explained as related to the presentation of words (to be selected) on the screen and the possibility of making use of the context for correct understanding of sentences. The language material characterized by diversity in the aspect of the contents and redundancy, makes an important tool in audiological examination. The diversity permits the use of tests for examination of different traits of the ear (Surmanowicz-Demenko, 2011).

[1] The difference between the averaged values obtained for the word and matrix tests was bordering on the limit of significance. However, taking into account the use of the tests, this result was evaluated as insufficient to assume that the results are equivalent.

Another element of analysis was the evaluation of the difficulties in taking particular tests. The least demanding was PSMT as:

1) the respondent needed only a computer software and properly calibrated earphones, so no advanced hardware was needed, and
2) experimenter interference was unnecessary.

The experimenter's task was to start the computer program. A similar test was PST, however, it demanded more attention from the experimenter. The NAL-93 test was the most difficult, which was related to the need of using the audiometer and active engagement of the experimenter.

The time of the NAL-93, PST and PSMT tests realization was similar. Presentation of a single word or sentence list took on average 2–3 minutes, depending on the pace of the participants response. For NAL-93 the time of realization was determined for the short method of measurements. The use of the constant stimuli method in NAL-93 or increase in the number of sentences in the matrix test would considerably extend the time of examination.

5.5. Comparison with previous results

On the basis of literature results, the SRT values obtained for the language material presented in background noise were validated. The speech reception threshold for NAL-93 was −3.4 dB SNR and was by 2 dB higher than −5.4 dB SNR reported by Lorens *et al.* (2006). The reasons for the differences can be:

1) different method of SRT determination – in our experiment (short method) only one list of words was used and SRT was calculated from Eq. (2). Lorens *et al.* (2006) interpolated the results obtained for three selected SNRs, presenting three lists of words at each level,
2) the masking signal used by Lorens *et al.* has not been comprehensively described – the SRT value they obtained was −5.4 dB SNR for NAL-93 but no information has been given on background noise (see Table 5).

SRT obtained for the Polish language word test (NAL-93) was compared with the results of the English language WIN (Words in Noise) test in which monosyllabic words are presented in noise (Wilson, Burks, 2005). For the Polish NAL-93 test masked by *speech noise*, SRT was −3.4 dB SNR. In the English WIN test presented in the presence of *multitalker babble*, SRT was +4.5 dB SNR, in the WIN test masked by speech-spectrum noise, SRT was +6.6 dB SNR (see Table 5). The differences in SRT values for Polish and English tests could be related to differences in language and masking noise applied. In both tests the same method of calculation, based on Eq. (2) was used.

For PST presented in *babble noise*, SRT was −6.8 dB SNR, which was by 0.6 dB lower than −6.2 dB SNR obtained by Ozimek *et al.* (2009). The methods of measurements and masking noise were the same in both tests. Student's *t*-test ($t(19) = -0.86$, $p = 0.40$) confirmed statistically non-significant differences between the results. Small difference between the values can follow from the number of participants and the number of lists. In the experiment described by Ozimek *et al.* (2009) the number of participants was 35 and the mean SRT was determined for 35 sentence lists, while in our experiment 10 participants took part and SRT was determined for 2 sentence lists. The value SRT = −6.8 dB SNR obtained for PST is comparable with −7.4 dB SNR obtained for the French language test FIST (Luts *et al.*, 2008) and with SRT = −6.2 dB SNR obtained for the German language test GÖSA (Kollmeier, Wesselkamp, 1997), in which the speakers were men. The reasons for the differences between SRT values obtained for PST, FIST and GÖSA can be the different number of participants, different masking noise, and scoring method (see Table 6). For the sake of comparison, Table 6 presents SRT values for the other two sentence tests, for Dutch and American language (Ozimek *et al.*, 2009). Although the way of presentation was the same, the values are higher than those obtained for PST. The reason for the differences may result from significant language differences, e.g. Polish language has the greatest number of affricates among all European languages (it results in an exhibition of a significant amount of energy for high frequencies – above 5 kHz; Habasińska *et al.* (2018)).

For PSMT test presented in the presence of *babble noise*, SRT was −4.8 dB SNR, which is by 3 dB higher

Table 5. Comparison of results of a Polish and English word tests and SRT values obtained (reported by different authors and obtained in the current study).

Test	Speaker gender	Masker	Mean SRT [dB SNR]	Subjects
NAL-93 (current study)	male	speech noise	−3.4 (± 2.0)	10
NAL-93 (Lorens *et al.*, 2006)	male	(no information)	−5.4	20
WIN (Wilson *et al.*, 2007a)	female	multitalker babble	4.5 (± 1.3)	24
		speech-spectrum noise	6.6 (± 1.0)	24

Table 6. Comparison of results of sentence tests reported by different authors and obtained in the current study, made for semantically predictable sentences presented by a male speaker.

Language	Mean SRT [dB SNR]	Listeners	Scoring method	Masker
Dutch (Versfeld *et al.*, 2000)	−4.0 (±0.5)	12	sentence scoring	individually shaped white noise
French (Luts *et al.*, 2008)	−7.4 (±0.7)	20	sentence scoring	stationary speech-shaped noise
German (Kollmeier, Wesselkamp, 1997)	−6.2 (±0.3)	20	word scoring	superposition of monosyllabic words
American (Nilsson *et al.*, 1994)	−2.9 (±0.8)	18	sentence scoring	stationary speech-shaped noise
Polish (Ozimek *et al.*, 2009)	−6.2 (±0.2)	35	sentence scoring	babble noise
Polish (current study)	−6.8 (±1.1)	10	sentence scoring	babble noise

Table 7. Comparison of the results of matrix tests reported by different authors and obtained in the current study. SRT values obtained by adaptive method.

Language	Speaker	Mean SRT [dB SNR]	Subjects
German (Brand *et al.*, 2004)	male	−6.3 (±0.6)	20
Italian (Puglisi *et al.*, 2015)	female	−7.4 (±0.8)	55
Polish (Ozimek *et al.*, 2010)	male	−8.0 (±1.3)	30
Polish (current study)	male	−4.8 (±1.6)	10
Russian (Warzybok *et al.*, 2015)	female	−9.4 (±0.8)	77
Spanish (Hochmuth *et al.*, 2012)	female	−7.2 (±0.7)	68

than SRT = −8 dB SNR reported by Ozimek *et al.* (2010). Student's t-test ($t(19) = 9.70$, $p < 0.001$) indicated statistically significant differences between the results. The difference could be related to:

1) different number of participants,
2) different number of lists (in our experiment the SRT value was calculated for 2 sentence lists, while in the Ozimek *et al.* (2010) 10 sentence lists were used),
3) different number of sentences on a single list (in our experiment lists comprised of 10 sentences, while in the Ozimek *et al.* experiment there were 20 sentences on each list).

The matrix tests for the majority of languages are made according to the same rules (Akeroyd *et al.*, 2015), their universal character permits making comparisons. Table 7 presents SRT values obtained for the European matrix tests, in which the SRT value was determined by the adaptive "closed-set" method. For each European language test mentioned in the table, SRT values were smaller than those obtained in our experiment. The closest result was obtained for the German matrix test (−6.3 dB SNR; Brand *et al.*, 2004). It was also the only test of the compared ones in which the speaker was a man. The SRT values for the other matrix tests were in the range from −7.2 dB SNR for the Spanish test (Hochmuth *et al.*, 2012) to −9.4 dB SNR for the Russian test (Kollmeier *et al.*, 2015). By a few dB higher value of SRT obtained in our study (compared to other languages) can be related to a different way of scoring – only in the Polish matrix test not individual words but sentences were scored.

6. Conclusions

On the basis of the results obtained in our experiment for young participants, the following conclusions can be drawn:

- For the Polish Sentence Test (PST) presented in *babble noise* generated from the sentence material used in this test: SRT = −6.8 ± 1.1 dB SNR.
- For the Polish Sentence Matrix Test (PSMT) presented in *babble noise* generated from the

sentence material used in this test: SRT = −4.8 ± 1.6 dB SNR.

- For the New Articulation Lists (NAL-93) presented in *speech noise*, SRT = −3.4 ± 2.0 dB SNR when the examination was performed by the short method. When the examination was performed by the constant stimuli method: SRT = −3.5 ± 1.8 dB SNR. No statistically significant differences in SRT were revealed by these two methods.
- The time needed for presentation of a single word list (NAL-93, short method) or a single sentence list (PST, PSMT) is comparable and equal to 2–3 minutes.

PSMT test is the least demanding as to the hardware and other equipment needed and is the simplest both for the participants and the experimenter. The most difficult and experimentally demanding is the NAL-93 test.

Acknowledgments

We express our deep gratitude to Professor Aleksander Sęk for preparing the software and helpful comments on earlier versions of the text; listeners – for commitment and patience in experiments; and three anonymous reviewers for comments and suggestions that helped in the preparation of the final version of the paper.

References

1. Akeroyd M.A. *et al.* (2015), *International Collegium of Rehabilitative Audiology (ICRA) recommendations for the construction of multilingual speech tests ICRA Working Group on Multilingual Speech Tests*, International Journal of Audiology, **54**, Suppl. 2, 17–22.
2. Brand T., Wittkop T., Wagener K., Kollmeier B. (2004), *Comparison of Oldenburg sentence test and Freiburg word test in closed-set versions*, Proceedings of 7th Congress of the German Society of Audiology, Leipzig, Germany.
3. Brungart D.S. *et al.* (2017), *Development and validation of the Speech Reception in Noise (SPRINT) Test*, Hearing Research, **349**, 90–97.
4. Darwin C. (2010), *Speech Perception*, [in:] *Oxford handbook of auditory science: hearing*, Plack C. [Ed.], Oxford University Press, Oxford.
5. Finney D.J. (1952), *Statistical method in biological assay*, pp. 524–530, C. Griffen, London.
6. Füllgrabe C., Moore B.C.J., Stone M.A. (2014), *Age-group differences in speech identification despite matched audiometrically normal hearing: contributions from auditory temporal processing and cognition*, Frontiers in Aging Neuroscience, **6**, 347, 1–25.
7. Habasińska D., Skrodzka E., Bogusz-Witczak E. (2018), *Development of Polish Speech Test Signal and its comparison with International Speech Test Signal (ISTS)*, Archives of Acoustics, **43**, 2, 253–262.
8. Hochmuth S., Brand T., Zokoll M.A., Castro F.Z., Wardenga N., Kollmeier B. (2012), *A Spanish matrix sentence test for assessing speech reception thresholds in noise*, International Journal of Audiology, **51**, 7, 536–544.
9. IEC 60645–1 (2017), *Electroacoustics – Audiometric equipment – Part 1: Equipment for pure-tone and speech audiometry.*
10. Jansen S. *et al.* (2012), *Comparison of three types of French speech-in-noise tests: A multi-center study*, International Journal of Audiology, **51**, 164–173.
11. Jassem W. (1973), *Fundamentals of phonetic acoustics* [in Polish], PWN, Warszawa.
12. Kociński J., Hafke-Dys H., Preis A. (2014), *Subjective methods for evaluation of speech intelligibility* [in Polish], [in:] *Hearing healthcare profession*, Hojan E. [Ed.], pp. 275–308, Scientific Publishers of AMU, Poznań.
13. Kollmeier B., Wesselkamp M. (1997), *Development and evaluation of a German sentence test for objective and subjective speech intelligibility assessment*, Journal of the Acoustical Society of America, **102**, 2412–2421.
14. Kollmeier B. *et al.* (2015), *The multilingual matrix test: principles, applications, and comparison across languages. A review*, International Journal of Audiology, **54**, 2, 3–16.
15. Lorens A., Obrycka A., Piotrowska A. (2006), *Validation of articulation lists according to Pruszewicz for evaluation of speech intelligibility in the presence of noise* [in Polish], Audiofonologia, **29**, 71–72.
16. Luts H., Boon E., Wable J., Wouters J. (2008), *FIST: A French sentence test for speech intelligibility in noise*, International Journal of Audiology, **47**, 373–374.
17. Martin F.N., Clark J.G. (2009), *Introduction to audiology*, Pearson, Boston.
18. McArdle R., Chisolm T.H., Abrams H.B., Wilson R.H., Doyle P.J. (2005a), *The WHO-DAS II: measuring outcomes of hearing aid intervention for adults*, Trends in Amplification, **9**, 127–143.
19. McArdle R.A., Wilson R.H., Burks C.A. (2005b), *Speech recognition in multitalker babble using digits, words, and sentences*, Journal of the American Academy of Audiology, **16**, 726–739.
20. Moore B.C.J., Skrodzka E. (2002), *Detection of frequency modulation by hearing– impaired listeners: Effects of carrier frequency, modulation rate, and added amplitude modulation*, Journal of the Acoustical Society of America, **111**, 1, 327–335.
21. Niewiarowski J. (2013), *Introduction to variance analysis with repeatable measurements* [in Polish], [in:] *Statistical Signpost 2. Practical Introduction to Variance Analysis*, Bedyńska S., Cypryańska M. [Ed.], pp. 99–112, Academic Publishers Sedno, Warszawa.

22. Nilsson M., Soli S.D., Sullivan J.A. (1994), *Development of the Hearing In Noise Test or the measurement of speech reception thresholds in quiet and in noise*, Journal of the Acoustical Society of America, **95**, 1085–1099.

23. Ozimek E., Kutzner D., Sęk A., Wicher A. (2007), *Polish Sentence Test for speech intelligibility measurements in masking conditions*, 19th International Congress on Acoustics, 2–7 September 2007, Madrid, Spain.

24. Ozimek E., Kutzner D., Sęk A., Wicher A. (2009), *Polish sentence tests for measuring the intelligibility of speech in interfering noise*, International Journal of Audiology, **48**, 433–443.

25. Ozimek E., Warzybok A., Kutzner D. (2010), *Polish sentence matrix test for speech intelligibility measurement in noise*, International Journal of Audiology, **49**, 444–454.

26. PN-EN ISO 8253-2 (2010), *Acoustics. Audiometric test methods. Part 2: Sound field audiometry with pure-tone and narrow-band test.*

27. Pruszewicz A., Surmanowicz-Demenko G., Jastrzębska M. (2011), *Polish tests for speech audiometry* [in Polish], [in:] *Selected problems of speech audiometry*, Obrębowski A. [Ed.], pp. 95–96, Scientific Publishers of Karol Marcinkowski Medical University in Poznań, Poznań.

28. Puglisi G.E. *et al.* (2015), *An Italian matrix sentence test for the evaluation of speech intelligibility in noise*, International Journal of Audiology, **54**, 2, 44–50.

29. Sęk A. (2016), *Authors software for the Institute of Acoustics, AMU* [Computer Software], version 1.0.6.0, Poznań.

30. Surmanowicz-Demenko G. (2011), *Linguistic and phonetic basis of word tests* [in Polish], [in:] *Selected problems of speech audiometry*, Obrębowski A. [Ed.], pp. 69–72, Scientific Publishers of Karol Marcinkowski Medical University in Poznań, Poznań.

31. Versfeld N.J., Daalder L., Festen J.M., Houtgast T. (2000), *Method for the selection of sentence materials for efficient measurement of the speech reception threshold*, Journal of the Acoustical Society of America, **107**, 3, 1671–1684.

32. Warzybok A., Zokoll M., Wardenga N., Ozimek E., Boboshko M., Kollmeier B. (2015), *Development of the Russian matrix sentence test*, International Journal of Audiology, **54**, 2, 35–43.

33. WHO/PDH/97.3, *Report of the Informal Working Group on Prevention of Deafness and Hearing Impairment Programme Planning WHO (Geneva, 1991). With adaptation from Report of the First Informal Consultation on Future Programme Developments for the Prevention of Deafness and Hearing Impairment*, World Health Organization, 23–24 January 1997, Geneva.

34. Wilson R.H. (2003), *Development of a speech-in-multitalker-babble paradigm to assess word-recognition performance*, Journal of the American Academy of Audiology, **14**, 9, 453–470.

35. Wilson R.H., Burks C.A. (2005), *Use of 35 words for evaluation of hearing loss in signal-to-babble ratio: A clinic protocol*, Journal of Rehabilitation Research & Development, **42**, 6, 839–852.

36. Wilson R.H., Carnell C.S., Cleghorn A.L. (2007), *The Words-in-Noise (WIN) test with multitalker babble and speech-spectrum noise maskers*, Journal of the American Academy of Audiology, **18**, 522–529.

37. Wilson R.H., McArdle R. (2005), *Speech signals used to evaluate functional status of the auditory system*, Journal of Rehabilitation Research & Development, **42**, 4, 79–94.

38. Wilson R.H., McArdle R.A., Smith S.L. (2007), *An Evaluation of the BKB–SIN, HINT, QuickSIN, and WIN Materials on listeners with normal hearing and listeners with hearing loss*, Journal of Speech, Language, and Hearing Research, **50**, 844–856.