www.czasopisma.pan.pl www.journals.pan.pl

ARCHIVES OF ACOUSTICS
Vol. **45**, No. 3, pp. 393–400 (2020)
DOI: 10.24425/aoa.2020.134056

Research Paper

Evaluation of New Polish Articulation Lists (NAL-93) in the Presence of Various Speech-Like Maskers

Anna SCHELENZ*, Ewa SKRODZKA

Chair of Acoustics, Department of Physics, Adam Mickiewicz University
Poznań, Poland
*Corresponding Author e-mail: anna.schelenz@amu.edu.pl

(*received February 5, 2020; accepted May 15, 2020*)

The main goal of the research was to obtain a set of data for ability of speech in noise recognition using Polish word test (New Articulation Lists – NAL-93) with two different masking signals. The attempt was also made to standardise the background noise for Polish speech tests by creating babble noise for NAL-93. Two types of background noise were used for Polish word test – the babble noise and the speech noise. The short method was chosen in the study as it provided similar results to constant stimuli method using less word material. The experiment using both maskers was presented to 10 listeners with normal hearing.

The mean SRT values for NAL-93 were −3.4 dB SNR for speech noise and 3.0 dB SNR for babble noise. In this regard, babble noise provided more efficient results. However, the SRT parameter for speech noise was more similar to values obtained for other Polish speech tests. The measurement of speech recognition using Polish word test is possible for both types of masking signals presented in the study. The decision as to which type of noise would be better in practice of hearing aid prosthetics remains an open-end question.

Keywords: NAL-93; babble noise; speech noise; speech-in-noise test; Speech Reception Threshold.

1. Introduction

The last decades of developments in the field of psychoacoustics have revealed that the reception of acoustic signals is hardly dependent on purely physical aspects. It is illustrated by the achievements of environmental acoustics (Miedema, Oudshoorn, 2001), in which the term 'signal annoyance' is used much more frequently than the 'signal level', as the first collocation is more respondent to the actual signal perception. There are many reports which suggest that the reception of acoustic information is related to the so called soundscape (Preis *et al.*, 2015). A similar approach has been observed in audiology in which it is claimed that a reliable evaluation of the auditory system is possible only when the proper evaluation tool (in this case – appropriate acoustic stimulus) is used.

Today, the developing societies deal with the cumulating "acoustic pollution". The transition of the soundscape from 'hi-fi', in which only one sound constitutes the main source of information, to 'lo-fi', in which the spate of sounds results in exhausting noise, is an inevitable consequence of civilisational development (Shaffer, 2004). For this reason, speech recognition in noise measurements appears to be no longer complementary but necessary element of hearing diagnostics. The masking signals used in these tests include the environmental, random, or quasi-random noise (i.e., white noise), as well as the noise representing averaged speech spectrum (Voss *et al.*, 2015; Chen *et al.*, 2014; Ozimek *et al.*, 2009a).

This last group of masking signals is most commonly used in auditory tests and includes many types of speech-like noise, such as babble noise, multitalker babble, or speech noise. Each one of them is created in a slightly different manner (Gundmi *et al.*, 2018; Fontan *et al.*, 2015; Hall, Flanagan, 2010; Krishnamurthy, Hansen, 2009; Wilson, 2003). Moreover, the same type of noise might differ depending on the application or the language in which the measurement is carried out. The lack of homogeneity of masking speech-like noise is a substantial drawback, because it disables or substantially inhibits comparisons of results obtained with similar tests. An example of this problem has been described in the article written by Wilson *et al.* (2007), in which the authors presented the results of speech recognition for the WIN (Words-In-Noise) test presented in two different masking signals. The difference in SRT values (Speech Recognition Threshold – the threshold for which the speech recognition equals 50%) for multitalker babble (involving several speakers talking simultaneously) and for speech-spectrum noise (reflecting long term spectrum of speech with 12 dB/oct decrease above 1 kHz) was above 2 dB.

There are many speech-like noise types that are claimed to be of 'universal' use. There are, for example, LTASS (Long Term Averaged Speech Spectrum) – the international averaged speech spectrum (Byrne, 1977); CCITT Rec G.227 – noise representing the averaged spectrum of six languages: English, German, Hungarian, Swedish, Russian, and Italian (Fastl, 1993); Fastl noise – CCITT noise with the amplitude modulation (the modulator – noise band with the central frequency of 4 Hz (Zwicker, Fastl, 1990)); ISTS (International Speech Test Signal) comprised of American English, Arabic, Chinese, French, German, and Spanish language (Holube *et al.*, 2010). The implementation of those kinds of noise in speech signal masking seems to be problematic taking into consideration the character of individual language (the differences in averaged speech spectrum, different phonetic content, different prosodic features). In some cases, it might lead to substantial diversity in test difficulty. For this reason, it seems much better to create masking signals of analogous structure (using homogeneous rules of masker generation), but preserving the spectral envelope representative of a specific language (as in the international matrix tests (Kollmeier *et al.*, 2015)).

The next problem is the diversity of material in speech tests. The stimuli that are used in studies are numerals, logatoms, words, and even the whole sentences. For Polish language, the commonly used tests presented in noise are: logatom lists (Brachmański, Staroniewicz, 1999), the Polish Triplet Test (Ozimek *et al.*, 2009b), Polish Sentence Test (Ozimek *et al.*, 2006), and Polish Sentence Matrix Test (Ozimek *et al.*, 2010), the latter two being sentence tests. Their resemblance, except for the speech material, stems from the type of masking signal used. In both tests the dedicated babble noise presented was made by multiple superposition of sentences used in each test and the modification of this material (shift or reversal of sentences in the time domain (Schelenz, Skrodzka, 2018)).

Another commonly used Polish test is NAL-93 (the New Articulation Lists) created in 1993 by Pruszewicz, and since 2011 it has been available in its revised version (Pruszewicz *et al.*, 2011). The new version comprises 10 lists containing 20 monosyllable words each and it is a balanced test, meaning that the results of the test are not dependent on the number of the lists chosen (Pruszewicz, Demenko, 2000). It does not have a dedicated masking signal as it is mainly used to test speech recognition in the quiet. However, it is possible to present NAL-93 using a universal masker, for example, speech noise (broadband averaged speech spectrum of 150–6000 Hz and a decrease of 12 dB/oct above 1 kHz) that is implemented in the audiometer used to carry out the examination.

The implementation of masking signal in NAL-93 might be of great value in hearing loss diagnostics, which is related to different qualities that word and sentence tests provide. The sentence tests are more redundant (they have more excessive information) than word tests, therefore they resemble real communication conditions. Besides, they cut out the time needed for the examination and use less material to obtain the results. On the other hand, word tests require less cognitive skills (i.e. the necessity to remember a chain of words and to recreate them correctly) and they provide great separation between SRT values for persons with normal hearing and persons with hearing impairment, so they are of high diagnostic value. Their main flaw is, however, the low prognostic value, because they poorly reflect the problems with real communication (Wilson *et al.*, 2007). Therefore, it seems that taking into consideration the complementarity of both tests (word and sentence ones), they should be presented both in the quiet and in the noise.

One can find data concerning the SRT parameter determination for NAL-93 presented in the noise in Lorens *et al.* (2006). However, these data are significantly edited so they lack essential information like the type of noise used or the standard deviation (SD) value for the averaged SRT parameter. Hence, it is difficult to draw any conclusions based on these results. For this reason, the authors decided to measure the SRT parameter using the NAL-93 test presented in standard speech noise. The aim of the research was also an attempt to standardise the masking noise for Polish speech tests by generation of babble noise for NAL-93 and comparison of the results with the SRT values obtained for Polish sentence tests using the same type of noise. In this manner, the authors wanted to test whether there are any benefits of using Polish word test in the presence of a masking signal and whether the babble noise dedicated for the test is a more effective masker than the 'universal' one (namely the speech noise). The presented work is the continuation of research described in the article that was previously published in *Archives of Acoustics* (Schelenz, Skrodzka, 2018).

2. Method

2.1. Subjects and measuring equipment

Ten listeners with normal hearing, aged 22–32, took part in the examination. Seven persons from this group were musically educated. Prior to the proper measurements, the participants were tested with the otoscopy, the pure tone audiometry, and the speech audiometry in the quiet.

The PTA (Pure Tone Average) for 500, 1000, 2000, and 4000 Hz was < 25 dBHL for each participant. The results of the speech audiometry fell within the scope of 90% speech recognition in the quiet for the material presented. The speech material used in this test was one list from the NAL-93 test presented at the level of 65 dB SPL.

The experiments were carried out in a sound treated room at the Institute of Acoustics at the Adam Mickiewicz University in Poznań. The pure tone audiometry and the speech audiometry was presented using the Interacoustics AC40 audiometer. The speech noise signal was generated straight from the audiometer and the babble noise was played using Windows Media Player on a PC computer. Both masking signals were calibrated using the artificial ear Bruel & Kjaer 1613 to provide comparable results. The material was presented monaurally, to the listener's preferred ear (according to the results of the pure tone audiometry), by Interacoustics DD45 headphones.

2.2. NAL-93 test in speech noise

A few lists from the NAL-93 test were chosen for the experiment. The NAL-93 test, as already mentioned, is dedicated to the Polish language. Polish consonant phonemes include the series of affricates: *c* /t͡s/, *dz* /d͡z/ (alveolar), *cz* /t͡ʂ/, *dż* /d͡ʐ/ (retroflex), *ć* /t͡ɕ/, *dź* /d͡ʑ/ (palatal), palatal consonants: *ń* /ɲ/, *ś* /ɕ/, *ź* /ʑ/, *j* /j/, two palatalised plosives /ki/, /gi/, and one palatalised fricative /xi/ (Habasińska *et al.*, 2018). The mentioned phonemes exhibit significant amount of energy in high frequencies. The power spectrum density of the babble noise obtained from Polish male speech has a distinctive increase in level in the region of frequencies higher than 5 kHz. It results from the fact that Polish has the greatest number of affricates from among European languages (Ozimek *et al.*, 2007).

The new version of the NAL-93 test from 2011 was used (Skrodzka, 2014). The test is composed of 10 lists of 20 monosyllable words each (Table 1). Each list was created in a way to include all 24 acoustic-phonetic structures, characteristic to the Polish language. The test is fully balanced, so any word list might be used to obtain correct results. The words are spoken by a male speaker (Demenko, Pruszewicz, 2000).

Although the NAL-93 test was created to determine the speech recognition in the quiet, it is possible to add masking noise to the material. Due to the common availability of the speech noise in many audiometers, the authors decided to use it as a background noise. There is, however, a huge deviation between the spectral structure of this signal and the averaged spectrum of all lists from the NAL-93 test, mainly because

Table 1. The New Articulation Lists (NAL-93 (Pruszewicz *et al.*, 2011)).

No.	List 1	List 2	List 3	List 4	List 5	List 6	List 7	List 8	List 9	List 10
1	twarz	płaz	kwas	dres	plac	głaz	płacz	kraj	krzyż	klej
2	rzecz	sieć	czas	Jaś	zez	jeż	ciecz	dziś	cześć	rzeź
3	pit	kit	typ	tak	kat	byk	tik	bak	gad	bat
4	plus	tchórz	kłus	głos	grosz	klucz	kłos	plusz	klosz	gruz
5	nić	maj	wesz	wyż	widz	nic	raj	las	mysz	mecz
6	bal	dym	dar	pył	pech	dach	dal	tył	pan	bar
7	dreszcz	wieprz	pleśń	płaszcz	pierś	chrzest	wjazd	wrzask	zjazd	wieść
8	mur	wół	wór	woń	muł	rów	chór	loch	mól	łom
9	grad	krzyk	żbik	krzak	grzyb	zbyt	styk	step	brzeg	szpik
10	syn	dzień	dżem	żal	cień	żart	cel	czyn	sen	zew
11	kark	pęk	karp	kant	bank	tynk	pęd	targ	park	garb
12	gmach	kran	plan	krem	drań	gram	krew	gniew	gwar	płyn
13	nos	moc	los	wódz	łódź	wuj	nóż	noc	wóz	mocz
14	biust	grunt	gwóźdź	klops	błąd	sport	brąz	front	prąd	tłuszcz
15	król	dłoń	tłum	broń	groch	proch	grom	plon	tron	dwór
16	łuk	lok	lot	lud	rok	łup	huk	róg	ród	lód
17	ton	ból	gol	dom	tom	bon	puch	dół	duch	koń
18	wstyd	stryj	zwiad	zgryz	kwiat	sklep	spływ	zbieg	spryt	strach
19	walc	filc	hełm	nerw	film	marsz	myśl	węch	chęć	hymn
20	rejs	cyrk	zięć	liść	sens	zysk	zamsz	maść	żart	szewc

of the lack of energetic increase in noise around 5 kHz that is characteristic to the Polish language, Fig. 1.

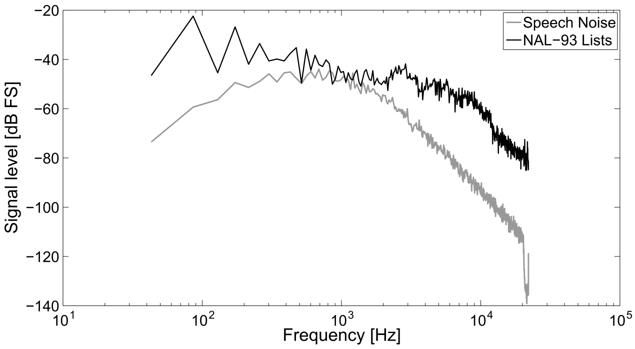

Fig. 1. Comparison of speech noise spectrum and the averaged spectrum of 10 lists taken from the NAL-93 test.

The NAL-93 test might be presented in two ways. The first one, the constant stimuli presentation, is a classic test method in which the whole word list is presented at a chosen signal level. The sound pressure level of the following lists is chosen to obtain two level values – one slightly above and one slightly below the 50% speech recognition. After obtaining both results, the actual SRT value is determined by interpolation of these values.

The second method is a short one, in which the sound pressure level changes after every few words within one list. Based on the transformation of Finney Eq. (1),

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

where S is the speech signal level, l is the lowest signal level for which the 100% speech recognition was obtained [dB SPL], d is the step of the change in signal level [dB], *correct* is the number of all correctly recognised words, w is the number of words presented at each level. It is possible to determine the SRT parameter using only one or two lists of the NAL-93 test at maximum (Wilson *et al.*, 2007; Finney, 1952).

This method has two main advantages; it enables: (1) shortening of the time needed for testing, which improves the measurement procedure, (2) the use of word lists of the NAL-93 test in a greater number of measurements, preventing the emergence of the so called learning effect. Assuming that both methods provide comparable SRT values, the authors of this article decided to use the short method to obtain the experimental results (Schelenz, Skrodzka, 2018).

2.3. Generation of babble noise

Babble noise for the NAL-93 was created based on the recordings of the whole word lists, recorded on a CD ‘Polish numerical and verbal test for hearing diagnostics and tests for auditory training ver. 2’ and consisting of 10 tracks with 20 words each separated with pauses of a few seconds. After putting each word into a separate file (using PreSonus Studio One), a preliminary digital processing was made (the band-pass filter of 50–12 000 kHz and the fade-in and fade-out were inserted). The recordings were then exported to MATLAB software for further processing.

Babble noise was created in a way that is analogous to the generation of babble noise for PST (Ozimek *et al.*, 2006). To do so, on the basis of literature and the information obtained from the authors of babble noise for PST, an empty vector of 15 s was created and each word from the NAL-93 test was placed on this vector in a random way. During the generation process, the positions of the following words in a vector were changed and some of the files were inverted in the time domain. All these operations were made using equal likelihood distribution.

The resultant noise had the following parameters: sampling frequency – 44 100 Hz, bit rate – 24 kbps, time duration – 15 s, iteration number – 10 000, number of words inverted – 50%. The noise was then filtered with band-pass filter of 50–12 000 Hz and normalised to 0 dBFS. To verify the obtained noise, a comparison with the averaged spectrum of all words from the NAL-93 test was carried out, Fig. 2.

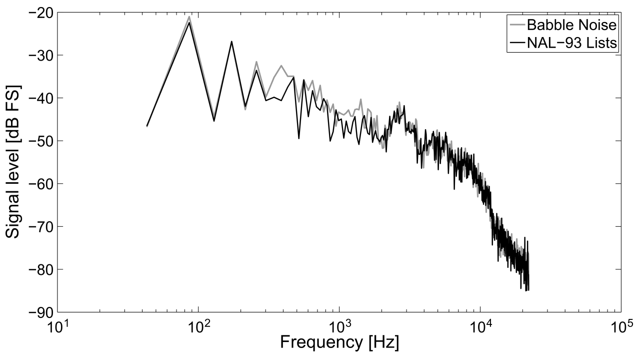

Fig. 2. Comparison of babble noise spectrum and the averaged spectrum of 10 lists taken from the NAL-93 test.

The crest factor values for babble noise and averaged NAL-93 spectrum were 1.40 and 1.46 respectively, no significant difference was therefore observed. Taking into consideration the graphs presented in Fig. 2, the greatest difference of 10–20 dB in signal level might be observed for 400–500 Hz and 10–12 kHz. The abovementioned frequency ranges are not critical to understand Polish as the most important frequency range for speech recognition lies within 500–2000 Hz (Jassem, 1973; Majewski *et al.*, 1977). It is worth mentioning that the reason for these level differences could be the random order of signal generation operations, as some of the words could be inserted to the vector more frequently than the others. It was therefore decided to use the created babble noise as a contrast to the speech noise. Comparison of both types of noise can be seen in Fig. 3.

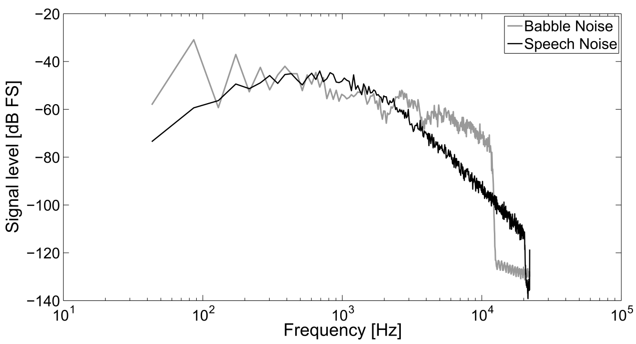

Fig. 3. Comparison of speech noise and babble noise spectra.

3. Experimental procedure

Each of the participants was presented with two randomly chosen lists, one masked with babble noise and the other masked with speech noise. Both speech signal and background noise were presented ipsilaterally (to the preferred ear). The noise signal level had a constant value of 65 dB SPL and the initial value of the speech signal was determined individually, according to the lowest signal level at which the listener was able to understand the whole presented speech material. The level of speech signal decreased 5 dB every 5 words and the material was presented until the participant was not able to understand any presented word at a given level. The SRT value was determined from Eq. (1), subtracting the obtained result from the level of the masking signal (in dB SPL).

4. Results

The results obtained for 10 listeners for two randomly chosen lists of NAL-93 test (one presented against babble noise and one presented against speech noise) are given in Table 2.

Table 2. The SRT values obtained in the NAL-93 test for two types of noise – speech noise and babble noise.

Listener	SRT[dB SNR]		
	Speech Noise	Babble Noise	Difference
1	–3	5	8
2	–6	3	9
3	–3	4	7
4	–1	4	5
5	–1	6	7
6	–2	3	5
7	–2	–2	0
8	–4	4	8
9	–6	3	9
10	–6	0	6
Mean	–3.4	3.0	6.4
SD	2.0	2.4	0.4

For both maskers the mean value of the SRT parameter and the standard deviation was determined as well as the difference in results for each participant concerning both types of noise. The SRT values obtained for babble noise were much higher than for speech noise and the mean difference of the parameter between the two maskers was about 6 dB.

5. Discussion

5.1. Analysis of differences between speech noise and babble noise

The results presented in Table 2 show that generally babble noise was a more effective masker for the NAL-93 test. The results of respective listeners are less homogeneous for the masker generated by the authors of this paper (babble noise) as indicated by a greater standard deviation. The main reason for this outcome might be the difference in structure of both noise types that stems from their generation pattern. Speech noise reflects the averaged speech spectrum, however, it is generated in an artificial way and there is a substantial decrease in the power spectrum density above 1 kHz. Therefore, it is not an effective masker for signals with an energetic maximum above this frequency (which is the case of Polish). On the contrary, babble noise reflects the speech signals of the NAL-93 test in terms of the averaged spectrum as well as the visible lower harmonic frequencies of the laryngeal tone (it is worth mentioning that lower frequency bands are the most responsible for the effective signal masking). For this reason, the SRT values obtained for babble noise surpass the ones obtained for speech noise.

The experiments indicated also the difference in standard deviation values. In the case of greater resemblance of speech signal and the masker, more visible statistical dispersion of values was observed as a result of lower redundancy of the NAL-93 test with babble noise. In the case of speech noise, the listeners could make use of the "spectral gaps" appearing above 1 kHz that came from the incompatibility of the averaged speech spectrum and noise spectrum. In this respect, the NAL-93 test presented in babble noise was much more difficult to understand as both spectra concurred. The effective level of both masking signals was not significant in this case as maskers were calibrated according to the RMS value.

The objective assessment of the results presented in Table 2 was made using the analysis of variance (ANOVA). In Tukey's test ($F = 42.667$, $p < 0.001$, dependent variable: SRT, fixed factor: masker type), statistically significant differences were observed between the values of the SRT parameter for speech noise and babble noise. This implies that the type of noise must be taken into account when the SRT parameter for NAL-93 test is considered.

Surprisingly, subjective opinions of the noise made by the listeners were inconsistent with the results of objective measurements (determined as SNR for SRT). The participants were asked to tell in which type of noise it was easier for them to recognise the presented speech material. The answer was unambiguous indicating babble noise as the easier one. The listeners justified their choice naming lower noise annoyance, better speech material clarity and lower hearing exhaustion when consecutive words from the list were presented. As the results seem to be at odds with the objective measurements (that unambiguously indicate better speech recognition for the NAL-93 test presented in speech noise), the reason for this might be the difference in both maskers' spectra transferring to experienced annoyance. Considered as more 'friendly' babble noise is marked by abovementioned harmony for lower frequencies that is visible in the noise's mellifluousness.

5.2. Practical application of noise in NAL-93 test

While choosing the most effective masker for the NAL-93 test, the listeners' assessment as well as practical utility must be taken into consideration. The speech-like noise used in the test should enable legible analysis of results and ensure sufficient SRT values separation between persons with and without hearing impairment. Moreover, the noise signal should be commonly available for persons using the NAL-93 test in their everyday practice.

The results of experiments using speech noise as the masker (that were carried out by us) were characterised by negative SRT parameters whose mean value was −3.4 dB SNR. This was a value similar to that obtained by the authors of Polish sentence tests (−6.2 dB SNR for PST and −8 dB SNR for PSMT; (Ozimek *et al.*, 2009a; 2010)). Although the analysis using t-Student's test indicated statistically significant differences between each test pair ($t = 4.403$, $p = 0.002$ for NAL-93 and PST; $t = 7.233$, $p < 0.001$ for NAL-93 and PSMT), the mean SRT parameter for persons with normal hearing had negative value. If the babble noise was used for the NAL-93 test, whose results for persons with normal hearing oscillated within 3 dB SNR, it would lead to some mistakes in the interpretation of results by persons that use sentence tests in their daily practice (for whom positive SRT values indicate hearing disorder).

Thanks to its low redundancy, the NAL-93 test is characterised by the greatest scattering of SRT values obtained in the measurements from among the abovementioned Polish speech tests (Table 3). In our judgement, the addition of babble noise with greater standard deviation (compared to speech noise) would lead to insufficient separation of results obtained for persons with normal hearing and persons with hearing impairment using the NAL-93 test. This hypothesis requires, however, further investigation.

Table 3. The SRT values determined for three Polish speech tests presented in noise.

Test	Masker	SRT [dB SNR] Mean	SRT [dB SNR] SD
NAL–93 (current study)	speech noise	−3.4	2.0
PST (Ozimek *et al.*, 2009a)	babble noise	−6.2	0.2
PSMT (Ozimek *et al.*, 2010)	babble noise	−8.0	1.3

For practical reasons, the speech noise is considered as a better masker as it is commonly available in audiology. Being a diagnostic signal, it constitutes an integral element of many audiometers. The signal itself was also placed on a CD player with the NAL-93 test (in the original version from 1993). The implementation of babble noise to the diagnostic battery would require the audiologists to possess a new masker and implement it to audiometers (i.e. the calibration would be required), not to mention the additional costs connected with the whole process of babble noise adaptation.

5.3. Comparison with literature

The results of the experiments presented in speech noise for the NAL-93 test are close to the results presented in literature for Polish sentence tests (Polish Sentence Test and Polish Sentence Matrix Test; (Ozimek *et al.*, 2009a; 2010)) meaning that the mean SRT parameter for the abovementioned tests is a single number with a negative value. However, the SRT values obtained for the same test (NAL-93) presented in babble noise reflect the literature results for English test WIN. In the experiment conducted by Wilson *et al.* (2007), two types of noise were compared as well. The analogue of the aforementioned speech noise (used as a stimulus in the experiment described in this paper) was speech-spectrum noise and the analogue of babble noise was multitalker babble (although in the case of the second pair of signals, a greater difference in the spectra could be observed).

In Table 4, the SRT values for both tests (the NAL-93 and the WIN) were juxtaposed, taking into consideration the masking signal used. The results obtained for babble noise/multitalker babble have similar values and the difference between the mean SRT values is 1.5 dB. On the other hand, the results presented for structurally alike maskers (speech noise/speech-spectrum noise) are −3.4 dB and 6.6 dB respectively, providing a separation of 10 dB.

To check whether it is possible to compare the SRT values obtained for both languages, the analysis of variance was conducted. In t-Student's test for bab-

Table 4. Comparison of the SRT values determined for the Polish NAL-93 test and the English WIN test, both presented in noise.

Test	Masker	SRT [dB SNR]	
		Mean	SD
NAL-93 (current study)	babble noise	3.0	2.4
	speech noise	–3.4	2.0
WIN (Wilson *et al.*, 2007)	multitalker babble	4.5	1.3
	speech-spectrum noise	6.6	1.0

ble noise/multitalker babble ($t = -2.012$, $p = 0.075$) no statistically significant differences were observed, however, for speech noise/speech-spectrum noise ($t = -15.724$, $p < 0.001$) the differences between the results were statistically significant. In this respect, babble noise provided better results. In other words, the use of babble noise enabled comparison of the SRT values determined for two different languages (Polish and English), making it a strong argument for using this particular noise in the NAL-93 test.

The reason for which the NAL-93 test presented in babble noise worked out much better when juxtaposed with the English monosyllable test (compared to the NAL-93 test presented in speech noise) was the spectrum of masking signals. In the case of babble noise/multitalker babble, their frequency characteristics was close to the speech material presented to the listeners. In the case of speech noise/speech-spectrum noise, in which there is a decrease in the spectrum above 1 kHz of 12 dB/oct, the masking efficiency of the WIN test was greater due to the lack of energetic maximum in averaged speech spectrum of English in higher frequencies (which, again, is the case in Polish).

Taking into consideration the abovementioned premises, the use of speech noise enables creation of a practical word test for the Polish language that determines the speech recognition in noise. However, in a universal test, it should be possible to compare the SRT values for different languages. In this case, the presentation of the NAL-93 test in babble noise provides better results.

6. Conclusions

Although, in the listeners' judgement, babble noise is a more effective masking signal, the majority of foregoing arguments unambiguously suggest using speech noise as the masker for the NAL-93 test. The attempt to unify the speech-like noise for Polish speech tests turned out to lead to many difficulties, including problems with results' interpretation, poor separation of results, or poor availability of noise recording for audiologists. For the aforementioned reasons, speech noise seems to be sufficient to evaluate speech recognition in the noise using word test. Despite the low redundancy, it provides sufficient homogeneity of results for persons with normal hearing. Besides, it is a universal noise that most of the hearing-aid prosthetics have in their offices.

The use of speech noise inhibits comparison of results for different languages. For this purpose, babble noise seems to be a better choice, as it represents averaged Polish speech spectrum in a much better way and hence, it is a more effective masker. The decision to what extent the constraints of both maskers are important in the particular SRT measurement are up to the audiologist and depends on the aim of the measurement.

References

1. Brachmański S., Staroniewicz P. (1999), Phonetic structure of a test material used in subjective measurements of speech quality [in Polish], *Speech and Language Technology*, **3**: 71–80.
2. Byrne D. (1977), The speech spectrum-some aspects of its significance for hearing aid selection and evaluation, *British Journal of Audiology*, **11**(2): 40–46, doi: 10.3109/03005367709078831.
3. Chen M.S., Liu J.S., Chen W.R. (2014), Differences in auditory discrimination ability between visually impaired and normally sighted adults, *Journal of Industrial and Production Engineering*, **32**(44): 255–262, doi: 10.1080/21681015.2015.1049226.
4. Fastl H. (1993), A masking noise for speech intelligibility tests, *Proceedings of TC Hearing*, Acoustic Society of Japan, H-93-70.
5. Finney D.J. (1952), *Statistical method in biological assay*, London: C. Griffen, pp. 524–530.
6. Fontan L., Tardieu J., Gaillard P., Woisard V., Ruiz R. (2015), Relationship between speech intelligibility and speech comprehension in babble noise, *Journal of Speech, Language and Hearing Research*, **58**(3): 977–986, doi: 10.1044/2015_JSLHR-H-13-0335.
7. Gundmi A., Himaja P., Dhamani A. (2018), Effectiveness of multitalker babble over speech noise and its implications: A comparative study, *Indian Journal of Otology*, **24**: 88–90, doi: 10.4103/indianjotol.INDIANJOTOL-24-18.

8. 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, doi: 10.24425/122373.
9. Hall J.L., Flanagan J.L. (2010), Intelligibility and listener preference of telephone speech in the presence of babble noise, *The Journal of the Acoustical Society of America*, **127**(1): 280–285, doi: 10.1121/1.3263603.
10. Holube I., Fredelake S., Vlaming M., Kollmeier B. (2010), Development and analysis of an International Speech Test Signal (ISTS), *International Journal of Audiology*, **49**(12): 891–903, doi: 10.3109/14992027.2010.506889.
11. Jassem W. (1973), *Fundamentals of phonetic acoustics* [in Polish], Warszawa: PWN.
12. Kollmeier B. *et al.* (2015), The multilingual matrix test: principles, applications, and comparison across languages: preview, *International Journal of Audiology*, **54**(Supp. 2): 3–16, doi: 10.3109/14992027.2015.1020971.
13. Krishnamurthy N., Hansen J.H.L. (2009), Babble noise: modeling, analysis and applications. *IEEE Transactions on Audio, Speech, and Language Processing*, **17**(7): 1394–1407, doi: 10.1109/TASL.2009.2015084.
14. 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.
15. Majewski W., Rothman H.B., Hollien H. (1977), Acoustic comparison of American English and Polish, *Journal of Phonetics*, **5**(3): 247–251.
16. Miedema H.M., Outshoorn C.G. (2001), Annoyance from transportation noise: relationships with exposure metrics DNL and DENL and their confidence intervals, *Environmental Health Perspectives*, **109**(4): 409–416, doi: 10.1289/ehp.01109409.
17. Ozimek E., Kutzner D. (2006), The Polish sentence test for speech intelligibility evaluations, *Archives of Acoustics*, **31**(4): 431–438.
18. Ozimek E., Kutzner D., Sęk A., Wicher A. (2007), Polish Sentence Test for speech intelligibility measurements in masking conditions, *19th International Congress on Acoustics*, September 2–7, 2007, Madrid, Spain.
19. Ozimek E., Kutzner D., Sęk A., Wicher A. (2009a), Polish sentence tests for measuring the intelligibility of speech in interfering noise, *International Journal of Audiology*, **48**(7): 433–443, doi: 10.1080/14992020902725521.
20. Ozimek E., Kutzner D., Sęk A., Wicher A. (2009b), Development and evaluation of Polish digit triplet test for auditory screening, *Speech Communication*, **51**(4): 307–316, doi: 10.1016/j.specom.2008.09.007.
21. Ozimek E., Warzybok A., Kutzner D. (2010), Polish sentence matrix test for speech intelligibility measurement in noise,*International Journal of Audiology*, **49**(6): 444–454, doi: 10.3109/14992021003681030.
22. Preis A., Kociński J., Hafke-Dys H., Wrzosek M. (2015), Audio-visual interactions in environment assessment, *Science of the Total Environment*, **523**: 191–200, doi: 10.1016/j.scitotenv.2015.03.128.
23. Pruszewicz A., Demenko G. (2000), Speech audiometry [in Polish], [in:] *Outline of Clinical Audiology*, Pruszewicz A. [Ed.], Poznań: Scientific Publishers of Karol Marcinkowski Medical University in Poznań, pp. 248–254.
24. Pruszewicz A., Surmanowicz-Demenko G., Jastrzębska M. (2011), Polish tests for speech audiometry [in Polish], [in:] *Selected problems of speech audiometry*, Obrębowski A. [Ed.], Poznań: Scientific Publishers of Karol Marcinkowski Medical University in Poznań, pp. 95–96.
25. Schafer M.R. (2004), The Music of the Environment, [in:] *Audio Culture: Readings in Modern Music*, Cox C., Warner D. [Ed.], New York, NY: Continuum International Publishing Group, pp. 29–38.
26. Schelenz A., Skrodzka E. (2018), Speech reception thresholds for Polish language word and sentence tests presented in noise, *Archives of Acoustics*, **43**(4): 603–612, doi: 10.24425/aoa.2018.125154.
27. Skrodzka E. (2014), Speech audiometry [in Polish], [in:] *Hearing healthcare profession*, Hojan E. [Ed.], Poznań: Scientific Publishers of AMU, pp. 275–308.
28. Voss P., Tabry V., Zatorre R.J. (2015), Trade-off in the sound localization abilities of early blind individuals between the horizontal and vertical planes, *The Journal of Neuroscience*, **35**(15): 6051–6056, doi: 10.1523/jneurosci.4544-14.2015.
29. 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.
30. 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**(6): 522–529, doi: 10.3766/jaaa.18.6.7.
31. 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, doi: 10.1044/1092-4388(2007/059).
32. Zwicker E., Fastl H. (1990), *Psychoacoustics. Facts and Models*, Berlin: Springer-Verlag.