Available online at www.sciencedirect.com

ScienceDirect

Speech Communication 77 (2016) 1–7

www.elsevier.com/locate/specom

Estonian words in noise test for children (EWINc)

Anneli Veispak[a,*], Sofie Jansen[b], Pol Ghesquière[a], Jan Wouters[b]

[a] KU Leuven, Parenting and Special Education Research Unit, Vanderkelenstraat 32, bus- 03765, 3000 Leuven, Belgium

[b] KU Leuven, ExpORL, Dept. Neurosciences, O&N II Herestraat 49 - bus 721, 3000 Leuven, Belgium

Received 20 April 2015; received in revised form 30 November 2015; accepted 30 November 2015

Available online 6 December 2015

Abstract

Objective: Based on the example of the Nederlandse Vereniging voor Audiologie (NVA)-lists (Bosman, 1989; Wouters et al., 1994) and in addition to the Estonian words-in-noise (EWIN) test for adults (Veispak et al., 2015), a words-in-noise test has been developed for Estonian children (EWINc).

Design: Two experimental steps were carried out: (1) selection and perceptual optimization of the monosyllables; construction of 14 lists, (2) an evaluation of the lists in normal hearing (NH) children both in noise and in quiet.

Study sample: Forty-three NH native speakers of Estonian (average age 7.9 years).

Results: The reference psychometric curve for all lists combined both in noise and in quiet for NH children was determined, with the slope and speech reception threshold differing from the respective values of the Estonian adults in accordance with previous research. The 14 lists in noise as well as in quiet were proven to be of equal difficulty with very little variation in the SRTs averaged over the lists.

Conclusion: The EWINc is a precise and reliable test for quantifying speech intelligibility in Estonian children.

Keywords: Speech intelligibility; Words in noise test; Phoneme scoring.

Abbreviations

ANOVA	repeated-measures analysis of variance
CVC	consonant-vowel-consonant
EWIN	Estonian words in noise
ExpORL	experimental oto-rhino-laryngology
IPA	international phonetic alphabet
LTASS	long-term average speech spectrum
NH	normal hearing
NVA	Nederlandse Vereniging voor Audiologie
RMS	root mean square
SNR	signal-to-noise ratio
SRT	speech reception threshold

1. Introduction

Due to historic circumstances and the fact that audiology does not yet exist as a separate discipline in Estonian higher education, there are no current speech perception tests available in Estonian language. Typically, simple pure-tone audiometry is being performed in Estonia to diagnose hearing loss as well as for rehabilitation evaluation purposes. It is generally agreed that traditional assessments of hearing loss based on pure-tone thresholds do not adequately predict speech intelligibility in noisy environments (Plomp and Mimpen, 1979a). Consequently, speech in noise is an important audiometric test procedure reflecting the functional hearing and communication ability in individuals with hearing impairment (Hällgren et al., 2006). The adequate assessment of hearing abilities is relevant, especially for children at school, who spend their days in a noisy classroom environment where listening is a major part of the learning process.

The perception of speech is an essential ability providing relevant information regarding overall auditory perception skills and is useful in outlining the prognosis of speech, language, reading, and cognitive abilities of children (Mendel, 2008). In addition to the child's vocabulary, language competency, chronological age, and cognitive abilities, speech perception performance in a child can also be influenced by the type of the required response, whether reinforcement is used and the inherent memory load in the task (Kirk et al., 1997). To date, a vast number of open- and closed-set tests tapping sensory and/or cognitive capabilities,

* Corresponding author. Tel.: +32 16 32 61 73; fax: +32 16 32 59 33.
E-mail address: Anneli.Veispak@ppw.kuleuven.be (A. Veispak).

http://dx.doi.org/10.1016/j.specom.2015.11.005
0167-6393/

using either monosyllables, sentences or another type of auditory stimuli, have been developed to evaluate speech perception skills in children (for review see Mendel, 2008).

Intelligibility tests based on sentences are thought to comprise utterances that are closer to the natural process of speech communication, reveal better phoneme representativeness, and can yield more accurate intelligibility data (Nilsson et al., 1994; Ozimek et al., 2012; Ozimek et al., 2010; Plomp and Mimpen, 1979b; van Wieringen and Wouters, 2008). On the other hand, in order for a speech recognition task to provide useful information about the auditory capacity of children, the test should seek to minimize the influence of cognitive and linguistic demands (McCreery et al., 2010; Nittrouer and Boothroyd, 1990). However, each test using different type of stimuli, contributes one piece of the puzzle toward an assessment of a child's overall speech perception and hence, at a minimum, the test battery should include an assessment of both word and sentence recognition (Mendel, 2008).

Given that there are no tests available in the Estonian language for assessing children's speech perception abilities, as a first piece of the puzzle we have developed an Estonian words-in-noise test for 7–9 years old children (EWINc). The test consists of 14 lists of monosyllables, where scoring is based on correctly identified phonemes. In addition to being less influenced by the listener's vocabulary knowledge than whole-word scores (Boothroyd, 1968), phoneme scores increase the number of data points, decrease variability and improve the precision of interpreting small differences in performance across presentations (Gelfand, 2003). Scoring based on phonemes has also been demonstrated to decrease the inter-subject variability and the differences in performance between age groups (Nittrouer and Boothroyd, 1990). McCreery et al. (2010) for example, compared the performance of children in four different age groups (5–6, 7–8, 9–10 and 11–12 year olds) using The Computer Aided Speech Perception Assessment (CASPA; Boothroyd, 2006) and demonstrated that the differences in word scores between 7–8, 9–10 and 11–12 year olds were not significant, but all three groups differed significantly from the 5–6 year olds. However, when pairwise comparisons were made for phoneme scores across age, it emerged that only 5–6 and 11–12 year olds differed significantly from each other.

Several studies have reported that the major age effect on speech recognition is observed in children less than 7 years of age (Hnath-Chisolm et al., 1998; Siegenthaler, 1969, 1975; Wilson et al., 2010). On the other hand, children's speech recognition performance is considered to approach that of adults by around 10 to 12 years of age (Elliot, 1979; Elliot et al., 1979; Stuart, 2005). The age brackets (7–9 years old children) chosen for the EWINc test, therefore, seem well justified by previously conducted research. This age is very important not only for speech and language acquisition, but also for general development as it coincides with the first two years in primary school in Estonia, a period of intense evolution in scholarly learning.

The EWINc test, a relevant addition to the Estonian words in noise (EWIN) test for adults (Veispak et al., 2015) was developed based on the example of the Nederlandse Vereniging voor Audiologie (NVA)-lists (Bosman, 1989; Wouters et al., 1994). In the first phase of the study the monosyllables were selected and perceptually optimized, the 14 list were created. In the second phase of the study the lists were evaluated in normal hearing (NH) children both in noise and in quiet. In general, for a good speech recognition test for children, the basic material should consist of words that are part of the known and used vocabulary of the children at the respective age. The words are combined into lists, which should have equal difficulty quantified by speech reception threshold (SRT), the slope of the performance metric in this SRT, and the precision quantified via test-retest comparisons. All of the abovementioned aspects of what constitutes a good quality speech test for children have been taken into account, analyzed and quantified in the current study.

2. Methods

2.1. Materials

In total, 350 simple and well-known monosyllables, which should be a part of the vocabulary of children above 6 years of age, were selected from extensively used first and second grade primary school Estonian language textbooks (Tungal and Hiiepuu, 2007; Jundas et al., 2005). The words were recorded in isolation one by one (each word 10 times) by a native Estonian speaker (female). No carrier phrase was used. Recordings of the words were made in a double-walled sound-proof booth in the Experimental Oto-rhino-laryngology (ExpORL) Research Group, Department of Neurosciences, University of Leuven (KU Leuven). Recordings were made with an Edirol R-4 PRO recorder and sampled at 44 100 Hz (24 bits resolution), using a Sennheiser HS2 headset microphone.

The best token of each word was selected based on the audibility and clarity of each individual phoneme in words and edited in Cool-Edit (Cool Edit Pro 2002, version 1.2a, Syntrillium Software Corporation, Phoenix, AZ). Words were cut at zero crossings and scaled to their average root-mean-square (RMS) before the first perceptual evaluation. All the words were stored as '.wav' files on the hard disk of a computer. Based on the spectrum of the recorded words stationary speech-shaped noise was generated. The average RMS level of the noise was rescaled to the average RMS of the words. More detailed description on the selection of the stimulus words, recording conditions, editing, speech rate and noise creation can be found by Veispak et al. (2015).

2.2. Subjects, test set-up, and calibration

Three groups of children participated in the evaluations. The participants of Group 1 (n = 11; average age 7.0, range 6.2–8.7 years), Group 2 (n = 20; average age 8.1, range 7.0–8.5 years), and Group 3 (n = 12; average age 8.6, range 7.2–9.9 years) were primary school students. All NH children had pure-tone thresholds below 20 dB HL for all octave frequencies between 250 and 4000 Hz (IS0 389-8, 2004) in both ears and were native speakers of Estonian. The children in Group 1 were tested in Belgium, whereas the children in Group 2 and 3 were tested in Estonia. The majority of the participating children were multilingual. The official training in the second language starts rather early and acquiring another language is difficult to avoid in the face of multilingual pop-culture and media landscape in Estonia. The participant was seated in a quiet room (for testing words-in-noise) or in a double wall sound-treated audiometric suite (for testing words-in-quiet) and heard the words monaurally (right ear) through Sennheiser HDA200 earphones. The words were played directly from a Dell Latitude E6430 portable computer using the software interface APEX 3 (Francart et al., 2008) and passed through an external RME Hammerfall DSP sound card to control the level of presentation. The levels of the words and noise were calibrated with a Type 4153 artificial ear to a sound level meter Brüel & Kjær type 2250. The participants were instructed to listen and repeat aloud whatever was heard. No feedback was provided. The number of correctly identified phonemes was scored manually by one experimenter on the spot. This study was approved by the Medical Ethics Committee of the University of Leuven (KU Leuven), University Hospitals Leuven as well as Medical Ethics Committee of Tallinn (National Institute for Health development in Estonia); written informed consent was obtained from all the parents of our participants.

2.3. Experiment 1

The aim of the first evaluation was to select a subset of words that are equally intelligible under the same adverse conditions. The 350 monosyllables were distributed into 25 lists, 14 items in each. The items were distributed with the intention of avoiding category clusters (e.g. animals, body parts etc.) and keeping the representation of different phonemes in each list as equal as possible. The 25 lists were randomly divided into 5 blocks, each presented at 5 different signal to noise ratios (SNR; 1, −2, –5, −8, −11 dB) to Group 1. Every child identified 1 block of words at each SNR, starting out with the block of words, which was at the highest SNR and continued with the consecutive blocks in increasing difficulty. Hence, in total, as there were 11 participants, every block was identified by 2 children at 4 SNRs and by 3 children at 1 SNR. The order of the words was fixed in each list and the same 5 lists were grouped together into one block throughout the whole experiment 1. The order of the lists presented within each block was randomized for every participant.

The noise level was presented continuously at 65 dB SPL. The total testing time together with a break after every block was about an hour and a half per participant. The experiment was described to the children like a computer game with difficulty levels. Children earned stickers with completing each level.

While testing, children's parents were given a list with all the stimulus words and asked to mark down the words they consider unfamiliar for their children. On average 7 words were marked as potentially unfamiliar by parents. After testing, children were asked to describe the meaning of the words noted by their parents as well as the words identified incorrectly in the easiest listening condition (SNR 1 dB). All participating children without exception were fully aware of the meaning of the words indicating that their performance was not influenced by unfamiliarity of the items.

2.4. Construction of the lists

The slopes and SRTs at 50% correct were determined by fitting a logistic function to each of the 350 words separately. As we were aiming to include words with a similar performance curve, a large portion of the words turned out to be unsuitable for the test. Nearly one third of all of the stimulus words were excluded, due to consisting of phoneme clusters which were either perfectly audible even in the most difficult listening condition (e.g. 'kass', 'hunt') or phoneme clusters, which were poorly identifiable also in the easiest listening condition (e.g. 'nälg', 'memm'). The first selection of the potentially includable monosyllables was based on the adults' data. However, as we were aiming to construct one set of words usable in both the adults as well as children's version of EWIN, only the words which were suitable also in the group of children, were selected. In total, 140 words were included based on the individual SRTs at 50% of the items. In the group of adults the SRTs of the 140 monosyllables were ranging ±4 dB around the median SRT (Veispak et al., 2015). In children's data the SRTs of the same words ranged between 5 to −3 dB around the median SRT. The approximate 1 dB shift in SRT values of individual words, when comparing adults and children, is compatible with the resulting SRTs when averaged over every individual subject. As the slope values were highly variable both in the group of adults (4%–16%/dB) as well as in the group of children (3%–20%/dB), no concrete criteria for inclusion based on slope was defined.

As the monosyllables were divided between the lists based on individual SRTs at 50% of the words with the intention of equalizing the average SRTs of all lists and distributing phonemes as equally between the lists as possible, the resulting distribution of words in EWINc (see Appendix) differs slightly from the version for adults. The frequencies of occurrence of 504 phonemes (14 x (3 + 33)) are listed as percentages as well as averages per list in Table 1 in descending order.

2.5. Experiment 2

For words-in-noise testing the 14 lists were randomly divided into 4 blocks (3 + 3+4 + 4) and presented to Group 2 of NH children (n = 20) at 4 SNRs (−2.5, −5, −7.5, and −10 dB). As the average performance of the children from the 1st experiment at −11 SNR was lower than 30% phonemes correct, the SNRs of the 2nd experiment were adjusted to provide detailed data to calculate the SRT at 50% but to minimize zero performance. The speech-weighted noise was presented continuously at 65 dB SPL. Hence, every block was identified by five children at each SNR.

For words-in-quiet, based on a pilot experiment, the stimuli were presented at the following levels: 35, 30, 25, 20 and 15 dB SPL. The 14 lists were divided into 5 blocks (3 + 3+3 + 3+2) and presented to the Group 3 of children. Every child identified one block of words on every presentation level in increasing difficulty. Hence, every list was identified by 3 children at 2 presentation levels and by 2 children at 3 presentation levels. The order of the words was fixed in each list and the same lists were

Table 1
Frequencies over all the lists, average frequency per list and percent frequency of occurrence of the phonemes in descending order.

	IPA	504	per list	100%
K	[k]	45	3.21	8.93
L	[l]	37	2.64	7.34
R	[r]	35	2.50	6.94
A	[ɑ]	33	2.36	6.55
P	[p]	28	2.00	5.56
N	[n]	27	1.93	5.36
I	[i]	24	1.71	4.76
T	[t,]	24	1.71	4.76
E	[e]	23	1.64	4.56
S	[s]	23	1.64	4.56
O	[o]	19	1.36	3.77
V	[v]	19	1.36	3.77
U	[u]	18	1.29	3.57
H	[h]	18	1.29	3.57
Õ	[ɤ]	17	1.21	3.37
M	[m]	16	1.14	3.17
D	[d̥]	14	1.00	2.78
G	[g]	14	1.00	2.78
Ä	[æ]	11	0.79	2.18
Ii	[iː]	8	0.57	1.59
Oo	[oː]	7	0.50	1.39
J	[j]	6	0.43	1.19
Uu	[uː]	5	0.36	0.99
Nn	[nn]	5	0.36	0.99
Ü	[y]	4	0.29	0.79
Ll	[ll]	4	0.29	0.79
Ee	[eː]	3	0.21	0.60
B	[b]	3	0.21	0.60
Tt	[tt,]	3	0.21	0.60
Ss	[ss]	3	0.21	0.60
Aa	[ɑː]	2	0.14	0.40
Öö	[øː]	2	0.14	0.40
Mm	[mm]	2	0.14	0.40
Kk	[kk]	1	0.07	0.20
Pp	[pp]	1	0.07	0.20

grouped together into one block throughout either the testing in quiet and in noise. However, the order of the lists presented within each block was randomized for every participant. The total testing time was about half an hour per subject for testing both in noise as well as in quiet.

2.6. Structure of the test and principle of scoring

The EWINc test for children is identical in principle to the EWIN adult version (Veispak et al., 2015), consisting of 14 lists. The performance test score is based on the percentage of correctly identified phonemes. Each of the 14 lists consists of one 3-phoneme practice item and 9 testing items, of which 3 are 3-phoneme and 6 are 4-phoneme words. In every list, there's one three-phoneme practice item, which precedes the 9 test items and is not scored. As the 9 test items in each list consist of 33 phonemes in total, each phoneme equals 3%. Hence, the total percentage score equals the number of correctly identified phonemes multiplied by 3, and 1% is added if the total score is higher than 50%.

Table 2
Average SRTs (dB) and slopes, together with their precision values.

	Average SRT	Precision	Average slope	Precision
Noise, fitted	−8.5 dB SNR	±0.7 dB	8.6%/dB	±1.7%/dB
Quiet, fitted	22.0 dB SPL	±0.7 dB SPL	5.2%/dB	±0.7%/dB

3. Results

3.1. Norm values and reference psychometric function

Slopes and SRTs at 50% are based on non-linear regression fits to a logistic function (SAS 9.3) using the following Eq. (1):

$$P = \frac{100}{\left(1 + e\left(4 \times s \times \frac{SRT-i}{100}\right)\right)} \tag{1}$$

where P is a percentage of correct recognition, e is the base of natural logarithms (~2.72), s is the regression slope, and i is the level of presentation either in dB SPL or dB SNR. The starting values of the parameters for iterations were defined as $s = 8\%$/dB; SRT $= -3$ dB SNR for words-in-noise; and $s = 8\%$/dB; $SRT = 15$ dB SPL for word-in-quiet.

The slopes and SRTs at 50% (Table 2) are the arithmetic average of the individually fitted SRTs and slopes of the different subjects obtained at fixed levels in quiet and in noise with data aggregated over the lists. The precision (error) bars on both parameters were deducted from the quadratically averaged error bars of the fit to the data of each individual subject.

Table 2 shows that the slopes at 50% point are 8.6%/dB for words-in-noise and substantially shallower for words-in-quiet (5.2 %/dB). The fact that the items of the test were selected based on noise data and the lists were optimized for the use in noise also explains the shallower slope in quiet. Psychometric functions of words-in-noise and in quiet based on phoneme scores are presented in Fig. 1.

The differences between mean phoneme- and word scores at each presentation level both in quiet as well as in noise are shown in Table 3. Across presentation levels the phoneme scoring gave on average 22% higher speech intelligibility scores over the whole-word scoring in quiet and 20% higher scores in noise. Comparing the SRTs and slopes based on the phoneme scoring (Table 2) with the respective values of the word scoring (Noise: SRT –5.7 SNR, slope 9.0 %/dB; Quiet: SRT 28.1 dB SPL, slope 5.3%/dB) shows, that while the slopes remain approximately the same, the difference in SRTs is 6.1 dB in quiet and 2.8 dB in noise.

3.2. Within-group variability

The performance of young children in identifying words presented in noise as well as in quiet has been shown to be poorer compared to older children and adults (Elliot et al., 1979). As the age of the children in Group 2 (n = 20; average age 8.1, range 7.0–8.5 years) and Group 3 (n = 12; average age 8.6, range 7.2–9.9 years) varied, a simple linear regression was calculated to predict children's performance in noise (Group 2) and in quiet (Group 3) based on age. The results show that age does not

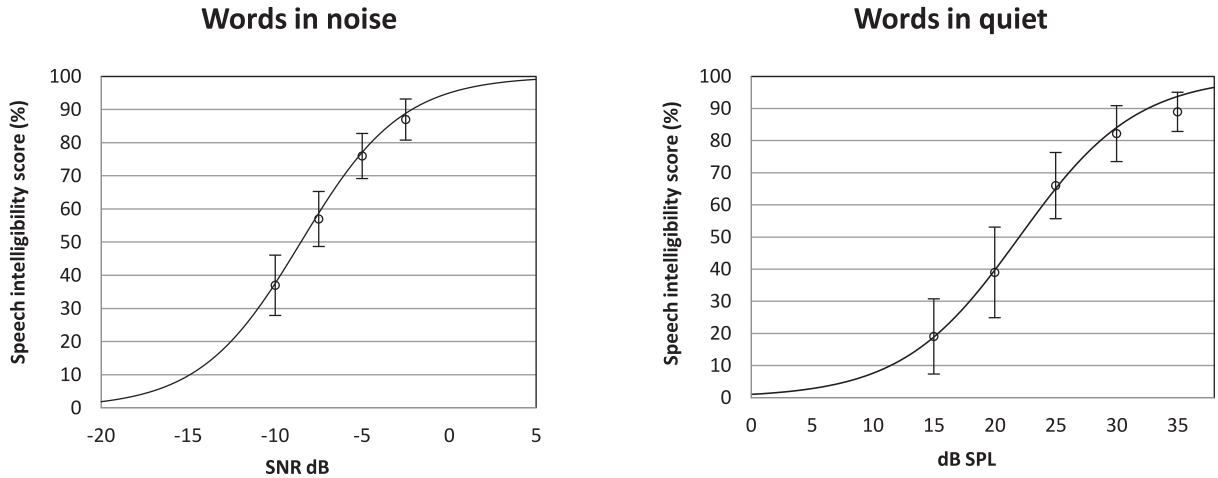

Fig. 1. Measured and fitted psychometric functions of the words in noise and in quiet together with the average performance (SD) at each presentation level.

Table 3
Speech intelligibility scores (%) in quiet and in noise.

In quiet						In noise					
dB SPL	Phoneme scores (%)			Word scores (%)		dB SNR	Phoneme scores (%)			Word scores (%)	
	SD	Mean	(–)	Mean	SD		SD	Mean	(–)	Mean	SD
35	6.1	89.3	15.4	73.9	11.7	−2.5	9.1	87.2	12.9	74.3	9.7
30	8.7	83.5	23.8	59.7	14.3	−5	8.3	76.0	21.5	54.5	9.1
25	10.3	66.1	27.8	38.3	12.7	−7.5	6.8	57.4	23.4	34.0	11.2
20	14.1	40.2	28.5	11.7	9.0	−10	6.2	36.7	21.0	15.7	11.6
15	11.7	19.5	15.3	4.2	4.3						

Note: The difference between the mean phoneme- and mean word score is indicated in the (–) column.

significantly predict the performance either in noise ($F(1,2798) = 1.86$, $p = .18$) nor in quiet ($F(1,1677) = 1.20$, $p = .29$).

3.3. List equivalency

Fig. 2 illustrates the variation in SRTs of the 14 lists in noise and in quiet. The values are plotted in terms of a deviation score from the overall mean, together with their respective standard errors. In order to examine the list equivalency, a repeated-measures analysis of variance (ANOVA) was conducted. The effect of test list on the SRT was found to be insignificant both for measurements in noise [$F(13,11.72) = 0.10$, $p = 1.00$] as well as in quiet [$F(13,36.15) = 0.14$, $p = 0.99$].

4. Discussion

We have developed an Estonian words-in-noise test for 7–9 years old children consisting of 14 lists of monosyllables, where

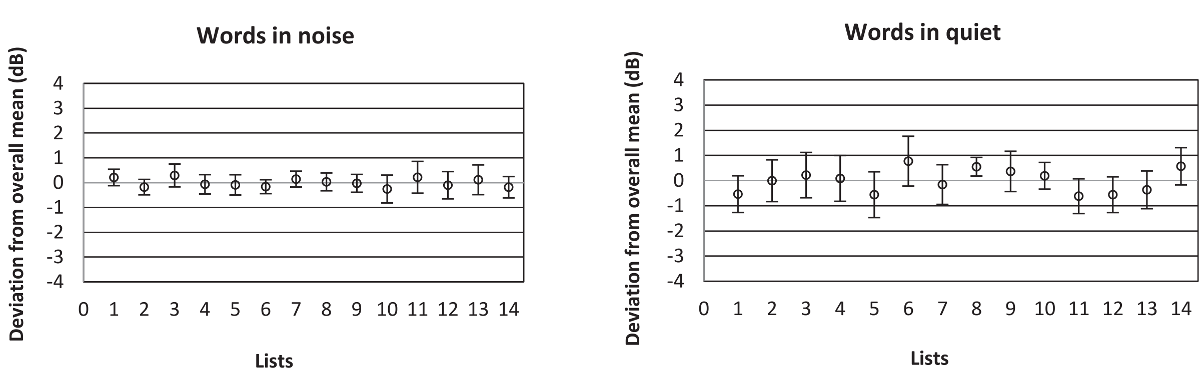

Fig. 2. SRTs of the 14 lists (+ standard error) of the words in noise and quiet. The data are plotted in terms of a deviation score from the overall mean.

scoring is based on correctly identified phonemes. Phoneme scoring has been shown to provide several advantages over whole-word scoring. In addition to increasing the number of data points and decreasing variability, phoneme scoring is also demonstrated to provide a more precise measure of the reception of the acoustical cues of speech (Boothroyd, 1968; Gelfand, 1998; Gelfand, 2003; Markides, 1978). In principle, a person listening to a EWINc list can achieve up to 73% discrimination score based on correctly identified phonemes before registering a single percent of correctly identified words. In reality, however, the difference is never so wide.

When comparing the phoneme- and word scores in the current study both in noise as well as in quiet, phoneme scores were consistently higher across all presentation levels. The difference between phoneme- and word scores peaked at 23.4% at −7.5 dB SNR in noise and at 28.5% at 20 dB SPL in quiet, being smaller in the easiest as well as in most difficult listening conditions. While the difference in slopes calculated either on phoneme- or word scoring data is less than 1%/dB, the difference between SRTs is 6.1 dB in quiet and 2.8 dB in noise in favor of phoneme scoring. The same pattern with closely resembling values is observable in the data of the Estonian adults as well as in Dutch speaking adults who were tested with the NVA lists (Wouters et al., 1994). The average difference in performance across presentation levels was 19.7% in noise and 22.2% in quiet in favor of phoneme scoring. These results are well in accordance with previous research where the average difference between phoneme- and word scoring across presentation levels either in noise or in quiet has been shown to slightly range around 20% in children (McCreery et al., 2010) as well as in adults (Gelfand, 1998; Keidser, 1993; Markides, 1978; Olsen et al., 1997).

The EWINc test was developed for 7–9 years old Estonian children. Despite the fact that there is variability in speech perception performance in this age range, consistent with previous studies (Elliot et al., 1979; McCreery et al., 2010; Stuart, 2005; Wilson et al., 2010), age did not significantly predict the performance either in noise nor in quiet. Even though 6 years old children participated in the optimization phase, the reported reference psychometric functions were calculated based on the data of children aged 7 years and older. The inclusion of the 6 year olds in the stimuli selection helped to ensure that the final set of monosyllables chosen for the test would be highly familiar and comfortably within the receptive vocabularies of the Estonian children starting school at the age of 7.

It is generally accepted that the capacity of NH children to recognize speech, relative to the ability of adults, is poorer in the quiet as well as more adversely affected in conditions of competing noise (Elliot, 1979; Elliot et al., 1979; McCreery et al., 2010; Neuman & Hochberg, 1983; Stuart, 2005; Stuart et al., 2006, Stuart, 2008). Comparing the SRTs at 50% of the Estonian adults (Veispak et al., 2015) and children obtained in noise (50% phonemes correct: adults 0.81 dB SNR better; 50% words correct: adults 0.9 dB SNR better) and in quiet (50% phonemes correct: adults 4.6 dB SPL better; 50% words correct: adults 3.7 dB SPL better), the differences in values appear to be in accordance with previous research. Elliot et al. (1979) for instance, who used the adaptive procedure both in noise and in quiet for identifying monosyllabic nouns by children and adults, depicted the mean SPL levels for 71% words correct performance for different age groups. The difference in the performance between adults and 7–8 years old children, relative to the difference between Estonian adults and children, appears to be of the same order of magnitude both in noise (i.e. ≈ 1–2 dB SPL) and in quiet (i.e. ≈ 4–5 dB SPL).

The SRT of the Estonian children obtained in noise (i.e. −8.5 dB SNR) resembles closely to the SRT at 50% phonemes correct of the 7–8 years old children depicted by McCreery et al. (2010) (i.e. ≈ −8 dB SNR) and to the SRT at 50% phonemes correct (−8.6 dB SNR) reported by Poelmans et al. (2011), who used NVA lists (Wouters et al., 1994) to measure speech perception performance in a group of 11 years old NH children.

Comparison of all the parameters of the EWINc test with other existing speech tests is not feasible due to different speech and noise materials, language differences, as well as diverse methods for measuring and reporting speech levels. However, the results of the current study appear to fit nicely with the existing evidence from other studies using speech tests with a similar concept, procedure and a way of reporting speech levels (McCreery et al., 2010; Poelmans et al., 2011). As no other speech-in-noise test is available in Estonian language, and given the unanimous recognition of the necessity of such a test in clinical practice, Estonian hearing specialists have already expressed their interest in the test materials. By making the EWINc test available with a data sharing agreement, the authors of the current paper are aiming to evaluate the test in larger groups of NH children in different ages as well as in a group of children with a hearing loss.

5. Conclusion

The EWINc test is a precise and valid speech intelligibility test, which consists of 14 lists of monosyllables, with 1 practice- and 9 testing items (33 phonemes) in each.

The EWIN test material can be obtained through the website of the ExpORL research group: https://exporl.med.kuleuven.be/web/index.php/Public:Software/Speech_Materials. Following a registration procedure, the speech materials as well as free software to run the test can be downloaded.

Acknowledgements

The authors would like to thank all those who participated in the development of the Estonian words in noise test: Tom Francart, Heleen Luts from ExpORL, Department Neuroscience, KU Leuven, Dr. Liina Luht from Ida Tallinna Keskhaigla, Mrs. Evelin Nappus from Haapsalu Linna Algkool, all our little motivated participants and their parents. We would also like to thank Ave Schank Lukas for lending us her voice. The first author of the paper is funded by the Fonds voor Wetenschappelijk Onderzoek FWO-Vlaanderen.

Appendix

The lists of the EWIN children test

List 1	List 2	List 3	List 4	List 5
aed	taat	mees	soe	pott
nahk	kuld	särg	kõrs	saun
liiv	õhk	lõik	kann	ahv
käed	neid	riik	vaht	palk
nõel	poks	kamp	luup	koht
kirp	org	niit	king	nõu
soov	tilk	kurt	pai	naer
kurg	tuul	konn	lend	järv
kolm	lõhn	halb	silm	pall
kuus	värv	talv	rühm	leib
List 6	**List 7**	**List 8**	**List 9**	**List 10**
rott	siis	auk	seen	hiir
kael	hea	nurk	laud	sõlm
õis	laul	jälg	keel	raud
kurb	põis	nööp	hing	paat
märg	tund	vait	kurk	virk
lind	mäed	roos	vars	hell
vann	kook	lehm	nõid	poeg
kell	rong	koll	jook	loom
poiss	kuiv	pliit	külm	küps
kõht	mai	tulp	ruut	trenn
List 11	**List 12**	**List 13**	**List 14**	
riis	pea	tass	nokk	
mänd	kõhn	hein	lomp	
peal	nöör	põlv	pirn	
karp	jalg	roog	aeg	
jonn	triip	koer	komm	
pilv	pood	päev	vaip	
nutt	salm	kamm	õng	
müts	pluss	või	reis	
hall	täpp	kord	lõug	
nõrk	korv	juust	täht	

Note: Underlined words are practise items only.

References

Boothroyd, A., 1968. Statistical theory of the speech discrimination score. J. Acoust Soc Am 43, 362–367.
Boothroyd, A., 2006. CASPA *5.0. Software*. A Boothroyd, San Diego.
Bosman, A.J., 1989. Speech Perception by the Hearing Impaired. University of Utrecht Doctoral dissertation.
Elliot, L.L., 1979. Performance of children aged 9 to 17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability. J Acoust Soc Am 66, 651–653.
Elliot, L.L., Connors, S., Kille, E., Levin, S., Ball, K., Katz, D., 1979. Children's understanding of monosyllabic nouns in quiet and in noise. J Acoust Soc Am 66, 12–21.
Francart, T., van Wieringen, A., Wouters, J., 2008. APEX3: a multi-purpose test platform for auditory psychophysical experiments. J Neurosci Methods 172, 283–293.
Gelfand, S.A., 2003. Tri-word presentations with phonemic scoring for practical high-reliability speech recognition assessment. J Speech Lang Hear Res 46, 405–412.
Gelfand, S.A., 1998. Optimizing the reliability of speech recognition scores. J Speech Lang Hear Res 41, 1088–1102.
Hällgren, M., Larsby, B., Arlinger, S., 2006. A Swedish version of the Hearing in Noise Test (HINT) for measurement of speech recognition. Int J Audiol 45, 227–237.
Hnath-Chisolm, T.E., Laipply, E., Boothroyd, A., 1998. Age related changes on a children's test of sensory-level speech perception capacity. J Speech Lang Hear Res 41, 94–106.
Jundas, E., Kippak, R., Kumberg, K., Põder, S., 2005. Ilus Emakeel. 2. klassi Eesti keele õpik. (1. + 2. osa). Koolibri, Tallinn.
Keidser, G., 1993. Normative data in quiet and in noise for "DANTALE"- a Danish speech material. Scand Audiol 22, 231–236.
Kirk, K.I., Diefendorf, A.O., Pisoni, D.B., Robbins, A.M., 1997. Assessing speech perception in children. In: Mendel, L.L., Danhauer, J.L. (Eds.), Audiologic Evaluation and Management and Speech Perception Assessment. Singular Publishing Company, Inc, San Diego, CA, pp. 101–132.
Markides, A., 1978. Whole-words scoring versus phoneme scoring in speech audiometry. Br J Audiol 12, 40–46.
McCreery, R., Ito, R., Spratford, M., Lewis, d., Hoover, B., Stelmachowicz, P.G., 2010. Performance-intensity functions for normal-hearing adults and children using CASPA. Ear Hear 31 (1), 95–101.
Mendel, L.L., 2008. Current consideration in pediatric speech audiometry. Int J Audiol 47 (9), 546–553.
Neuman, A.C., Hochberg, I., 1983. Children's perception of speech in reverberation. J Acoust Soc Am 73, 2145–2149.
Nilsson, M., Soli, S.D., Sullivan, J.A., 1994. Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. J Acoust Soc Am 95 (2), 1085–1099.
Nittrouer, S., Boothroyd, A., 1990. A context effects in phoneme and word recognition by young children and older adults. J Acoust Soc Am 87, 2705–2715.
Olsen, W.O., Van Tasell, D.J., Speaks, C.E., 1997. Phoneme and word recognition for words in isolation and in sentences. Ear Hear 18, 175–188.
Ozimek, E., Kutzner, D., Libiszewski, P., 2012. Speech intelligibility tested by the pediatric matrix sentence test in 3–6 years old children. Speech Commun 54, 1121–1131.
Ozimek, E., Warzybok, A., Kutzner, D., 2010. Polish sentence matrix test for speech intelligibility measurement in noise. Int J Audiol 49, 444–454.
Plomp, R., Mimpen, A.M., 1979a. Speech-reception threshold for sentences as a function of age and noise level. J Acoust Soc Am 66, 1333–1342.
Plomp, R., Mimpen, A.M., 1979b. Improving the reliability of testing the speech reception threshold for sentences. Audiology 18, 43–52.
Poelmans, H., Luts, H., Vandermosten, M., Boets, B., Ghesquière, P., Wouters, J., 2011. Reduced sensitivity to slow-rate dynamic auditory information in children with dyslexia. Res Dev Disabil,s 32, 2810–2819.
Siegenthaler, B.M., 1969. Maturation of auditory abilities in children. Int Audiol 8, 59–71.
Siegenthaler, B.M., 1975. Reliability of the TIP and DIP speech-hearing tests for children. J Commun Disord 8, 325–333.
Stuart, A., 2005. Development of auditory temporal resolution in school-age children revealed by word recognition in continuous and interrupted noise. Ear Hear 26, 78–88.
Stuart, A., 2008. Reception threshold for sentences in quiet, continuous noise, and interrupted noise in school-age children. J Am Acad Audiol 19, 135–146.
Stuart, A., Givens, G.D., Walker, L.J., Elangovan, S., 2006. Auditory temporal resolution in normal hearing pre-school children revealed by word recognition in continuous and interrupted noise. J Acoust Soc Am 119, 1946–1949.
Tungal, L., Hiiepuu, E., 2007. Eesti keele õpik 1. klassile. Avita, Tallinn.
van Wieringen, A., Wouters, J., 2008. LIST and LINT: sentences and numbers for quantifying speech understanding in severely impaired listeners for Flanders and the Netherlands. Int J Audiol 47, 348–355.
Veispak, A., Jansen, S., Ghesquière, P., Wouters, J., 2015. Speech audiometry in Estonia: Estonian words in noise (EWIN) test. *Int J Audiol*. Early online. http://dx.doi.org/10.3109/14992027.2015.1015688 .
Wilson, R.H., Farmer, N.M., Gandhi, A., Shelburne, E., Weaver, J., 2010. Normative data for the words-in-noise test for 6- to 12-year-old children. J Speech Lang Hear Res 53, 1111–1121.
Wouters, J., Damman, W., Bosman, A.J., 1994. Vlaamse opname van woordenlijsten voor spraakaudiometrie. Logopedie 6, 28–33.