Speech audiometry in Estonia: Estonian words in noise (EWIN) test

Anneli Veispak^a, Sofie Jansen^b, Pol Ghesquière^a & Jan Wouters^b

^a KU Leuven, Parenting and Special Education Research Unit, Leuven, Belgium

^b KU Leuven, Dept. Neurosciences, ExpORL, Leuven, Belgium

Corresponding Author

Anneli Veispak,
Parenting and Special Education Research Unit
KU Leuven
Leopold Vanderkelenstraat 32
Postbus 03765
B-3000 Leuven
Belgium
Anneli.Veispak@ppw.kuleuven.be

Key words

Speech intelligibility, words in noise test

Abbreviations

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

ABSTRACT

Objective: Currently, there is no up to date speech perception test available in the Estonian language that may be used to diagnose hearing loss and quantify speech intelligibility. Therefore, based on the example of the Nederlandse Vereniging voor Audiologie (NVA)-lists (Bosman, 1989; Wouters et al., 1994) an Estonian words in noise (EWIN) test has been developed.

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

Study sample: Thirty-six NH native speakers of Estonian (age range from 17 to 46 years).

Results: The reference psychometric curve for NH subjects was determined, with the slope and speech reception threshold being well in accordance with the respective values of the NVA lists. The 14 lists in noise yielded equivalent scores with a high precision.

Conclusion: The EWIN test is a reliable and valid speech intelligibility test, and is the first of its kind in the Estonian language.

INTRODUCTION

Currently, there is no up to date speech perception test available in the Estonian language suitable for diagnosing hearing loss and quantifying speech intelligibility. Two speech tests were developed during Soviet times (multisyllabic digits and monosyllables for presentation in quiet), but these are rarely if ever used due to the lack of high quality recordings and testing procedures as well as the lack of professionals. The fact that audiology does not (yet) exist as a separate discipline in Estonian universities with the responsibilities of audiologists being divided between otolaryngologists and nurses, explains why new speech tests have not been developed, even though the need has been recognized. Thus, mostly simple pure-tone audiometry is performed in Estonia to diagnose hearing abilities as well as for rehabilitation evaluation purposes.

The most common complaint reported by individuals with hearing loss is difficulty understanding speech in the presence of background noise (Killion, 2002; McArdle et al., 2005). Traditional assessments of hearing loss based on pure-tone thresholds do not adequately predict speech intelligibility in noisy environments (Plomp & Mimpen, 1979a). Speech in noise, consequently, is an important audiometric test procedure as it reflects the functional hearing and communication ability in individuals with hearing impairment (Hällgren et al., 2006), making the diagnostic assessments more relevant and more reliable. Information gained from measuring speech perception in noise is beneficial not only in discriminating between normal and impaired hearing but also for rehabilitation purposes such as selecting amplification strategies, and counselling concerning expectations for hearing aid performance (McArdle et al., 2005). Therefore, based on the example of the Nederlandse Vereniging voor Audiologie (NVA)-lists (Bosman, 1989; Wouters et al., 1994) an Estonian words in noise (EWIN) test has been developed and evaluated.

METHODS

Selection of the stimulus words

The words were selected from widely used first and second grade primary school Estonian language textbooks (Tungal & Hiiepuu, 2007; Jundas et al., 2005). In total 350 well know and widely used

monosyllables were chosen, which should be a part of the vocabulary of children above 6 years of age.

Speakers and recordings

The words were recorded by a native Estonian speaker (female), who was instructed to articulate as clearly as possible without placing undue emphasis on any of the phonemes. 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. All the words were recorded word-by-word in isolation 10 (2 x 5) times in random order. One word was repeated five times with approximately one second in between and then another round of five repetitions followed with other words recorded in between.

Editing, speech rate, and noise

The best of the 10 tokens 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 to avoid clicks. The words were 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 long-term average speech spectrum (LTASS) of the 350 words was determined by removing the silence parts of each word (frames of 20 ms with $\text{RMS} < 0.001$) and the spectrum was calculated with a 4096-points fast Fourier transform using a rectangular window and without overlap. These spectra were then averaged. For the resulting LTASS, a 2048-taps finite impulse response filter was generated and applied on an 11 seconds long white noise. Transients at the start and the end of the speech-shaped noise were removed to allow looping of the noise without any clicks. The average RMS level of the noise was rescaled to the average RMS of the words.

Subjects, test set-up, and calibration

Three groups of subjects participated in the evaluations. The subjects of the group 1 (n=10) and group 2 (n=16), were university students with the average of 22 years of age. The subjects of group 3 (n=10) were adults with the average of 36 years of age. All subjects were screened for normal hearing (<20 dB HL for octave frequencies from 500 Hz to 4000 Hz) in both ears and were native speakers of Estonian. The subject was seated in a quiet room 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 subjects were instructed to listen and repeat aloud whatever was heard or understood. No feedback was provided. The number of correctly identified phonemes was scored manually. This study was approved by the Medical Ethics Committee of the University of Leuven (KU Leuven) and University Hospitals Leuven, written informed consent was obtained from all participants.

First evaluation and construction of the lists

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. 25 lists were divided into 5 blocks, each presented on 5 different signal to noise ratios (SNR +1, -2, -5, -8, -11 dB) to the group 1 of NH Estonian adults. Hence, each word was presented at each SNR level to two subjects. The noise level was presented continuously at 65dB SPL. The total testing time was about an hour per subject.

From the total of 350 included words, 140 monosyllables were selected. As the determination of slope and SRT were based on fitting a logistic function to each word separately (2 x 5 data points), the slope values were highly variable ranging from 4 to 16% per dB. Therefore the selection of the words was mainly based on the SRT (+/- 4 dB around the median SRT) of the performance curve. No concrete criteria for inclusion based on slope was defined.

While in the Flemish NVA-lists (Wouters et al., 1994) simple CVC (consonant-vowel-consonant) words are used, the structure of the monosyllables chosen for the EWIN test is more varied (CVC, CVCC,

CVV, CVC, VVC, or VCC). This is due to the highly consistent nature of the language. Estonian belongs to the group of Finno-Ugric languages. It is a transitional form from an agglutinating to a fusional language, thus being heavily inflected. There are 14 productive cases, no grammatical gender for nouns or personal pronouns, no articles, and a differentiation between three quantities both in vowels and consonants in Estonian language. There are 26 phonemes in Estonian, 9 of which are vowels (a, e, i, o, u, õ, ä, ö, ü). The language is orthographically phonemic (one grapheme equals one phoneme), while the spelling does not distinguish between 2nd and 3rd quantity, nor is a palatalization marked (e.g. Erelt, 2003; Sutrop, 2004).

The 140 words were divided with the intention of distributing the different phonemes as equally between the 14 lists as possible. The frequencies of occurrence of 504 phonemes (14 times 3 + 33 phonemes) are listed as percentages as well as averages per list in table 1 in descending order.

Structure of the test and principle of scoring

In the NVA test, there are 15 lists of 12 monosyllabic CVC words, consisting of 1 practice and 11 testing items (Bosman, 1989; Wouters et al., 1994). The performance test score is based on the percentage of correctly identified phonemes. As the 11 testing items consist of 33 phonemes in total, each phoneme equals 3%. Hence, the total percentage score equals the number of correctly heard phonemes multiplied by 3, and 1% is added if the total score is higher than 50%. The construction of the EWIN test is the same in principle. 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. As the 9 testing items comprise of 33 phonemes in total, the scoring of the test is identical to the NVA-lists, 3% per phoneme and 1% is added if the total score is higher than 50%.

Second evaluation of the lists

For words-in-noise the speech-weighted noise was presented at 65 dB SPL and the SNR levels were -5, -7.5, -10 and -12.5 dB. The 14 lists were divided into 4 blocks and presented to group 2 of NH subjects. Hence, every list was identified by four different subjects on each SNR level.

A pilot experiment with words in quiet showed intelligibility scores to range between 0% and 100% at the following intensity levels: 35, 30, 25, 20 and 15 dB SPL. The 14 lists were divided into 5 blocks

and presented to the group 3 of NH subjects. Every list was identified by two different subjects at each intensity level.

RESULTS

SRTs and slopes at 50% scores are based on non-linear regression fits to a logistic function (SAS 9.3) of the performance intensity curve of each individual subject obtained at fixed levels in quiet and in noise with data (except for the data of the practice items), averaged over the lists. The resulting SRTs and slopes are the arithmetic average of the individually fitted SRTs and slopes of the different subjects. 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 9.3%/dB for words-in-noise and considerably shallower for words-in-quiet (4.9%/dB). Performance intensity functions of words-in-noise and in quiet are presented in Figure 1.

List equivalency

Figure 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 check the list equivalence statistically, the SRTs were calculated for each word separately (except for the practice item of each list) and averaged across the lists. One sample t-tests demonstrated that the SRTs of the 14 lists measured in noise do not differ from the overall mean. However, the SRTs of lists number 2 [$t(8) = -2.49$, $p = .04$] and 12 [$t(8) = 3.35$, $p = .01$] measured in quiet differ moderately from the overall mean.

DISCUSSION

The importance of assessing speech perception in noise in the adult population has led to the development of precise and standardized speech tests, using both sentence (e.g. Bilger et al, 1984; Cox et al, 1987; Nilsson et al, 1994; Plomp & Mimpen, 1979; Versfeld et al., 2000; van Wieringen & Wouters, 2008; Kollmeier & Wesselkamp, 1997; Vaillancourt et al., 2005; Hällgren et al., 2006; Bevilacqua et al., 2008; Killion & Villchur, 1993; Killion et al, 2004) as well as word materials (e.g.

Hirsch et al., 1952; Tillman & Carhart, 1966; Wilson, 2003; Boothroyd, 2006; Magnusson, 1995; Bosman, 1989; Wouters et al., 1994). While the speech tests using word stimuli are intended for diagnostic applications, speech tests with sentence materials are rather used for evaluation and rehabilitation such as selecting suitable amplification strategies (McArdle et al., 2005). However, repeating sentence materials involves many issues beyond a simple recognition task including recognition versus recall (Salthouse, 1985; Craik, 1994), top-down versus bottom-up processing (Wingfield, 1996), recency and primacy effects (Murdock, 1962), and the multiplicative effects of various degradations on recognition of the speech signal (Helfer, 1995). Monosyllabic stimuli in speech testing paradigms, on the other hand, depend largely on bottom up processing, minimizing the effects on performance of working memory and linguistic context (McArdle et al., 2005).

Although the majority of speech tests using word stimuli are developed for testing speech intelligibility in quiet, there are examples of well-known speech tests where monosyllables are presented in speech shaped noise or multitalker babble. These examples include the CID W-22 Auditory tests (Hirsch et al., 1952), Northwestern University Auditory Test No. 6 (NU No.6; Tillman & Carhart, 1966), the Words-in-Noise test (WIN; Wilson, 2003), The Computer Aided Speech Perception Assessment (CASPA; Boothroyd, 2006), Swedish PB words (Magnusson, 1995), and the NVA lists in Dutch (Bosman, 1989; Wouters et al., 1994).

The EWIN test was created based on the example of the NVA lists (Wouters et al., 1994). As Dutch is orthographically relatively transparent, the structure of the NVA lists was considered most suitable for Estonian language. Also, creating a speech test based on phoneme scores in a language where no other alternative speech test is available, was thought to be reasonable. Scoring based on correctly identified phonemes increases the number of data points, decreases variability and improves the precision of interpreting small differences in performance across presentations (Gelfand, 2003). Additionally, the phoneme score is less influenced by the listener's vocabulary knowledge than the whole-word score and is, therefore, a more valid measure of auditory resolution (Boothroyd, 1968).

The SRTs of the 14 EWIN lists measured in noise were demonstrated not to deviate more than 0.24 dBs from the overall mean, indicating list equivalence with high precision. However, the SRTs measured in quiet showed more variability, with 2 lists differing significantly from the overall mean. As the 14 lists were optimized for testing in noise, not in quiet, the finding is not surprising. Given that speech recognition in noise correlates to high frequency pure-tone thresholds and signal analytic abilities such as frequency sensitivity and gap detection, whereas speech recognition in quiet rather correlates with low-frequency thresholds (Dreschler & Plomp, 1980), the differences in list equivalency in noise and in quiet are explainable.

Comparing the SRTs at 50% speech recognition and slopes of the EWIN test and the NVA lists (Wouters et al., 1994) obtained in noise (-9.3 ± 0.4 dB SNR, 9.3 ± 1.6 %/dB and -9.1 ± 0.6 dB SNR, 5.5 ± 0.6 %/dB respectively) and in quiet (17.4 ± 0.8 dB SPL, 4.9 ± 0.8 %/dB and 19.0 ± 1.1 dB SPL, 4.8 ± 0.7 %/dB respectively), the values are well in accordance. The slope measured in noise, however, is somewhat steeper for the EWIN compared to the NVA lists. The SRT at 50% speech recognition of the EWIN in noise is also of the same order of magnitude as the CASPA (Boothroyd, 2006) values depicted by McCreery et al (2010) (i.e. ≈ -10 dB SNR). As in most of the abovementioned speech tests using word stimuli, the evaluation of performance is based on the percentage of correctly perceived words, not phonemes, and which use multitalker babble instead of speech shaped noise, the comparison of the results is not entirely feasible. Wilson et al. (2007) for example compared the results of WIN stimuli (Wilson, 2003) in noise versus in multitalker babble, and reported 50% words correct scores to be 4.0 dB SNR for multitalker-babble and 6.3 dB SNR for noise, differing from the EWIN and NVA (Wouters et al., 1994) results by more than 10 dB. The great differences reported in obtained SNR at 50% speech recognition with normal-hearing subjects are considered to be partly due to different speech and noise material, language differences, as well as diverse methods for measuring and reporting speech levels (Ludvigsen, 1992). Consequently, to make comparisons possible, the use of standardized speech- and noise-level definitions as well as calibration protocols would be essential (Magnusson, 1995).

In that regard, the EU-funded project HearCom, aiming to harmonize hearing diagnostics for European languages had the purpose to develop speech intelligibility tests based on words, digit-triplets and sentences in different languages allowing across-country comparisons and homogeneity of results/outcomes (Vlaming et al., 2011). Based on procedures that have been standardized and certified within HearCom, many different other materials —tools for auditory diagnosis and evaluation as well as for fitting auditory prostheses— have since been developed in several additional languages (Jansen et al., 2010; Zokoll et al., 2012). The current study fits within this international framework. Several Estonian hearing specialists have already expressed their interest in the study and test materials.

CONCLUSION

The EWIN test is a reliable 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 is the first of its kind in the Estonian language.

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 and all our subjects. We would also like to thank Ave Schank Lukas for being the 'voice' in the test. The first author of the paper is funded by the Fonds voor Wetenschappelijk Onderzoek FWO-Vlaanderen.

Appendix

The lists of the Estonian words in noise test

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

Note: underlined words are practise items only.

REFERENCES

- Bevilacqua, M.C., Banhara, M.R., Da Costa, E.A, Vignoly, A.B., Alvarenga, K.F. 2008. The Brazilian Portuguese Hearing in Noise Test. *Int J Audiol*, 47, 364-365.
- Bilger, R.C., Nuetzel, J.M., Rabinowitz, W.M., Rzeczkowski, C. 1984. Standardization of a test of speech perception in noise. *J Speech Hear Res*, 27, 32-48.
- Bosman, A.J. 1989. Speech perception by the hearing impaired. Doctoral dissertation, University of Utrecht.
- Boothroyd, A. 1968. Statistical theory of the speech discrimination score. *J Acoust Soc Am*, 43, 362-367.
- Boothroyd, A. 2006. *CASPA 5.0. Software*. San Diego: A Boothroyd.
- Cox, R., Alexander, G., Gilmore, C. 1987. Development of the Connected Discourse Test (CST). *Ear Hear*, 8, 119S–126S.
- Craik, F.I.M. 1994. Memory changes in normal aging. *Am Psychol Soc*, 3, 155–158.
- Dreschler, W.A., Plomp, R. 1980. Relation between psychophysical data and speech perception for hearing-impaired subjects. *J Acoust Soc Am*, 68, 1608-1615.
- Erelt, M. 2003. Estonian language. Tallinn: Estonian academy publishers.
- 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.
- 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.
- Helfer K. 1995. Auditory perception by older adults. In: Huntley R, Helfer K, eds. *Communication in Later Life*. Boston: Butterworth-Heinemann, 41-84.
- Hirsh, I.J., Davis, H., Silverman, S.R., Reynolds, E.G., Eldert, E., Benson, R.W. 1952. Development of materials for speech audiometry. *J Speech Hear Disord*, 17, 321-337.
- Jansen S., Luts H., Wagener K.C., Frachet B. & Wouters J. 2010. The French digit triplet test: a hearing screening tool for speech intelligibility in noise. *Int J Audiol*, 49 (5), 378–87.

- Jundas, E., Kippak, R., Kumberg, K., Põder, S. 2005. Ilus Emakeel. 2. klassi Eesti keele õpik. (1. + 2. osa). Tallinn: Koolibri
- Killion, M.C., Niquette, P.A., Gundmundsen, G.I., Revit, L.J., Banerjee, S. 2004. Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *J Acoust Soc Am*, 116, 2395–2405.
- Killion, M.C., Villchur, E. 1993. Kessles was right-partly: but SIN test shows some aids improve hearing in noise. *Hear J*, 46(9), 31-35.
- Killion, M.C. 2002. New thinking on hearing in noise: a generalized articulation index. *Semin Hear*, 23, 57-75.
- Kollmeier, B., Wesselkamp, M. 1997. Development and evaluation of a German sentence test for objective and subjective speech intelligibility assessment. *J Acoust Soc Am*, 102, 2412-2421.
- Ludvigsen, C. 1992. Comparison of certain measures of speech and noise level. *Scand Audiol*, 21, 23-29.
- Magnusson, L. 1995. Reliable clinical determination of speech recognition scores using Swedish PB words in speech weighted-noise. *Scand Audiol*, 24, 217-223.
- McArdle, R., Wilson, R.H., Burks, C.A. 2005. Speech recognition in multitalker babble using digits, words, and sentences. *J Am Acad Audiol*, 16, 726-739.
- 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.
- Murdock, B.B. Jr. 1962. The serial position effect of free recall. *J Exp Psychol*, 64, 482–488.
- 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.
- 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.
- Salthouse, T.A. 1985. *A Theory of Cognitive Aging*. Amsterdam: North-Holland.

- Sutrop, U. (2004). *Estonian language*. Tallinn: Estonian Institute.
- Tillman, T.W., Carhart, R. 1966. *An Expanded Test for Speech Discrimination Utilizing CNC Monosyllabic Words*. Northwestern University Auditory Test No. 6. USAF School of Aerospace Medicine Technical Report. Brooks Air Force Base, TX: US Air Force.
- Tungal, L., Hiiepuu, E. 2007. Eesti keele õpik 1. klassile. Tallinn: Avita
- Vaillancourt, V., Laroche, C., Mayer, C., Basque, C., Nali, M., et al. 2005. Adaption of the HINT (hearing in noise test) for adult Canadian Francophone populations. *Int J Audiol*, 44, 358-369.
- 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.
- 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. *J Acoust Soc Am*, 107, 1671-1684.
- Villchur, E. 1982. The evaluation of amplitude-compression processing for hearing aids. In: Studebaker, G. Bess, F. (eds). *The Vanderbilt Hearing Aid Report*. Upper Darby, PA: Monographs in Contemporary Audiology.
- Vlaming, M.S.M.G., Kollmeier, B., Dreschler, W.A., Martin, R., Wouters, J., Grover, B., Mohammad, Y., Houtgast, T. 2011. HearCom: Hearing in the Communication Society. *Acta Acustica United with Acoustica*, 97, 175-192.
- Wilson, R.H. 2003. Development of a speech in multitalker babble paradigm to assess word-recognition performance. *J Am Acad Audiol*, 14, 453-470.
- Wingfield, A. 1996. Cognitive factors in auditory performance: context, speed of processing, and constraints of memory. *J Am Acad Audiol*, 7, 175-182.
- Wouters, J., Damman, W., Bosman, A.J. 1994. Vlaamse opname van woordenlijsten voor spraakaudiometrie. *Logopedie*, 6, 28-33.
- Zokoll, M.A., Wagener, K.C., Brand, T. Buschermöhle, M., Kollmeier, B. 2012. Internationally comparable screening tests for listening in noise in several European languages: The German digit triplet test as an optimization prototype. *Int J Audiol*, 51, 697-707

TABLES

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	[a]	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	[ɣ]	17	1.21	3.37
m	[m]	16	1.14	3.17
d	[d]	14	1.00	2.78
g	[g]	14	1.00	2.78
ä	[æ]	11	0.79	2.18
ii	[i:]	8	0.57	1.59
oo	[o:]	7	0.50	1.39
j	[j]	6	0.43	1.19
uu	[u:]	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	[e:]	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	[a:]	2	0.14	0.40
öö	[ø:]	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

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

	<i>Average SRT</i>	<i>Precision</i>	<i>Average slope</i>	<i>Precision</i>
Noise, fitted	-9.3 dB	±0.4 dB	9.3%/dB	±1.6%/dB
Quiet, fitted	17.4 dB SPL	±0.8 dB SPL	4.9%/dB	±0.8%/dB

FIGURE LEGENDS

Figure 1. Measured and fitted performance intensity functions of the words in noise and in quiet together with the performance on each intensity level averaged across the subjects.

Figure 2. SRTs of the 14 lists (+ standard error) of the words in noise and quiet. Data is plotted in terms of a deviation score from the overall mean.