BNAM2012
Joint Baltic-Nordic Acoustics Meeting
June 18th - 20th 2012 Odense, Denmark

Speech in Noise Test based on a Ten-Alternative Forced Choice Procedure

Ellen Raben Pedersen
University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark, ellenped@hotmail.com

Peter Møller Juhl
University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark, pmjuhl@iti.sdu.dk

The purpose of this study is to design and evaluate a user operated speech in noise test. The test is based on the Danish speech material Dantale II, which consists of test sentences with a syntactically fixed structure of five words. For each presented word the subject had to select a response from ten alternative words listed in a matrix on a touch screen. To evaluate the test including a possibility for each word to answer "I do not know" (?-button) a listening test was performed with 24 normal-hearing subjects. The listening test shows that when the subjects have the possibility to press the ?-button, they made fewer *motivated* guesses (a guess made when only a part of the word is heard) than when they don't have the ?-button available. Results from the user operated test with the ?-button equals results found in a traditional test, where the subject's task is to orally repeat as much heard as possible. Additionally, the listening test shows that the homogeneity of the speech material is uninfluenced when the material is used in a user operated test.

1 Introduction

Many people, and in particular hearing-impaired persons, find it difficult to hear and understand speech in a noisy environment. Over the years different speech in noise tests have been developed to achieve a measure of hearing-impaired individual's ability to hear speech in background noise. In the Danish clinical practice the commonly used speech test is the Dantale I test, which consists of lists of single monosyllables and numbers of one syllable [1]. Danish speech tests based on sentences also exists. The most common is the Dantale II test [2] and the Danish version of the Hearing In Noise Test (HINT) [3], which consists of nonsense sentences and everyday sentences, respectively. The Dantale II test is developed in analogy to the Swedish Hagerman test [4] and the German Oldenburg sentence test [5, 6, 7], and is used as material in the present study.

The speech material Dantale II consists of syntactically fixed but semantically unpredictable (nonsense) sentences. The five words in each sentence can be regarded as single words in sequence, which makes the material well suited for an adaptive test procedure. In an adaptive test procedure the presentation level of each sentence is depending on the number of correct answered words in the previous sentence. The presentation level that corresponds to a particular speech intelligibility percentage is the Speech Reception Threshold (SRT). Typically SRT corresponds to a probability of 50% correct responses [8]. When the subjects answers two or fewer words correct (below 50%), the presentation level has be to be raised, and when the subjects answers three or more words correct (above 50%), the presentation level has be to be reduced. Another way to test is to present all the sentences at the same presentation level and then determine the percentage of correct answered words at that particular level. This leads to the Discrimination Score (DS) [9].

A traditional test setup involves the subject and an operator. The subject has to repeat (orally) as much heard as possible after each presented word or sentence. The operator then registers whether the subjects answer is correct or incorrect. In the Dantale II test this corresponds to register the number of correct repeated words. The possible benefits of a user operated test are a reduction of the test time for the operator and the elimination of the possible influence of the operator on the evaluation of the subject response. Another purpose could be to design a test for internet use [11]. The German Oldenburg sentence test had been implemented and evaluated in a user operated test version, where the subject had to select a response from ten alternative words listed in a matrix for each word presented [12]. Besides the ten alternative words the subject had the option of answering "I do not know". Each "I do not know" answer was interpreted as an *incorrect* answer. Without the possibility to answer "I do not know" the user operated test was designed as five subsequent ten-Alternative Forced Choice (10AFC) tests. Comparable results were found for a traditional test and the user operated test version [12].

The purpose of this study is to design and evaluate a Danish version of the German user operated test [12] based on the Danish speech material Dantale II. Two versions of a user operated test are investigated: one with the possibility for each word to answer "I do not know" and one without that possibility. The two versions of the user operated test are evaluated and compared with a traditional test. In connection with this, the influence of the possibility to answer "I do not know" on the test result will be evaluated. Further, it will be evaluated whether it is reasonable to use the speech material in a 10AFC test, i.e. whether the homogeneity (the difficulty of the different words) of the speech material is influenced when the material is used in a user operated test instead of being used in a traditional test for which the material original is designed.

2 Methods

2.1 Speech material

In this study the Danish speech material Dantale II, which is described in reference [2], was used. The speech material consists of 16 lists based on the same 50 words (ten words for each of five word classes). Each list contains ten test sentences, which have a syntactically fixed structure of five words from different word classes: name, verb, numeral, adjective, and object. The different words presented in a matrix are seen in Table 1. The sentences are not fully meaningful and therefore the words cannot be predicted from the context, hence the words can be regarded as single words in sequence. As an example the first sentence in list 1 is: "Ingrid finds seven red houses" (translation of the Danish sentence: "Ingrid finder syv røde huse").

2.2 Noise signal

Besides the sentence lists, the Dantale II speech material contains a noise signal. The noise signal was generated by superimposing the test sentences many times. This process results in a noise signal, which does not sound like real speech. For this reason an alternative noise signal was made and evaluated [13]. It was found that the two noise signals perform equally well distinguishing between subjects with different abilities to understand speech in noise. Both the original and the alternative noise signal have a long-term spectrum similar to the one of the test sentences, but the alternative noise signal is slightly more modulated and sounds more like real speech babble. This causes the test to reflect a more real-life listening situation and therefore the alternative noise signal was chosen in this study.

2.3 Discrimination functions

The discrimination function describes the speech intelligibility or discrimination, p as a function of the presentation level, here *SNR* (signal-to-noise ratio). The discrimination function for a traditional test, where the subjects orally repeat as much heard as possible, is [8]

$$p_{TT}(SNR, SRT_N, s_{50}) = \frac{1}{1+exp(4 \cdot s_{50} \cdot (SRT_N - SNR))}, \quad (1)$$

where the subscript TT denotes traditional test. The parameter SRT_N refers to the *SNR* level corresponding to a probability of 50% correct responses and s_{50} is the slope of the discrimination function for the traditional test at SRT_N. The reference data (for normal-hearing) is: SRT_N = -8.4 dB SNR and s_{50} = 13.2 %/dB [2].

The procedure for a traditional test is open, which means that the number of alternative responses for each word is in theory infinite. In the following the function for the open test procedure is adapted to a closed test procedure, i.e. an n-Alternative Forced Choice (nAFC) procedure. In a test based on a nAFC procedure the subject has to select a response from n alternative responses. When a word is not heard, the subject is forced to guess. This study distinguishes between two types of guesses: *pure* guesses and *motivated* guesses. A pure guess is a guess made, when a subject has not heard the word at all. If a subject has heard a part of the word, he/she can in some cases guess the right word, since the possible words are presented. The subject then made a motivated guess.

In the limit of low *SNR*'s (no signal) the discrimination function for a nAFC procedure will have the value 1/*n* due to pure guesses, whereas the value will approach 1 for high *SNR* as it is also the case for the traditional test. Hence, the discrimination function for a nAFC procedure corresponds to a vertical compression of the discrimination function for the traditional test. For a nAFC procedure the probability of hearing each word is equal to $p_{TT}(SNR, SRT_N, s_{50})$. The probability of not hearing a word (and having to guess) is $1 - p_{TT}(SNR, SRT_N, s_{50})$. It is assumed, that when a word is heard it is responded correctly. The normalized probability function (discrimination function) of correct responded words, both heard and by pure guessing, is therefore

$$\begin{aligned} p_{nAFC}(SNR, SRT_N, s_{50}) &= p_{TT}(SNR, SRT_N, s_{50}) + \frac{1}{n} \cdot \left(1 - p_{TT}(SNR, SRT_N, s_{50})\right) \\ &= p_{TT}(SNR, SRT_N, s_{50}) \cdot \left(1 - \frac{1}{n}\right) + \frac{1}{n}\,. \end{aligned} \quad (2)$$

Motivated guesses has not been taking into account.

Equation (1) is inserted into Equation (2):

$$p_{nAFC}(SNR, SRT_N, s_{50}) = \frac{1}{1+exp\left(4 \cdot s_{50} \cdot (SRT_N - SNR)\right)} \cdot \left(1 - \frac{1}{n}\right) + \frac{1}{n}\,, \quad (3)$$

where s_{50} is the slope of the discrimination function for the traditional test. The slope of Equation (3) at $SNR = SRT_N$ is $s_{50} \cdot (1\text{-}1/n)$. This factor corresponds to the factor, by which the discrimination function for the traditional test is vertically compressed. Equation (3) is in agreement with reference [14]. In the following *n* is set to ten, by which all the words within the different word classes in the speech material are represented.

From values of the speech intelligibility at different *SNR*'s the discrimination function with the maximum likelihood was determined by using the Method of Maximum Likelihood. The likelihood of the discrimination function $p(SNR, SRT_N, s_{50})$ is

$$l\left(p(SNR, SRT_N, s_{50})\right) = \prod_{k=1}^{m} p\,(SNR_k, SRT_N, s_{50})^{c_k} \cdot \left(1 - p(SNR_k, SRT_N, s_{50})\right)^{5-c_k}\,, \quad (4)$$

where the number five in the exponent of the last term is the number of words in each test sentence. The parameter *m* is the number of sentences presented and c_k is the number of correct words registered at the *k*'th sentence. To determine the most likely discrimination function the parameters SRT_N and s_{50} are varied until $\log\left(l\left(p(SNR, SRT_N, s_{50})\right)\right)$ is maxi-mized [15].

2.4 User operated test versions

In the user operated test the subjects have to select a response from ten alternative words listed in a matrix for each presented word. The ten alternative words for each of the five words in each sentence are the 50 words, which the 16 lists in the speech material are based on. The user operated test was designed in two versions: one with the possibility for each word to answer "I do not know" (press on a ?-button) and one without that possibility. When the ?-button is available the test procedure is not entirely closed. A press on the ?-button was converted into a 1/10 correct response corresponding to a random guess among the ten possible words. This is contrary to [12], where each "I do not know" answer was interpreted as an incorrect answer. This alternative handling of the ?-button was introduced in order to equalize a potential different use of the ?-button among the subjects. Some subjects would use the ?-button whenever they are not quite sure which word had been presented, whereas others will choose a word instead and almost never use the ?-button. The original intention with the ?-button was to increase the subjects' engagement in the test. It can be frustrating to be forced to choose an answer (without the possibility to answer "I do not know") to a word, which has not been heard at all, and this can potentially draw the attention from the next sentence to be presented.

2.5 Subjects

A listening test was performed with 24 normal-hearing subjects (twelve males and twelve females, aged 21-39 years with an average age of 26 years). Even though all the subjects considered them self to by normal-hearing, their hearing threshold were measured at the frequencies 0.5, 1, 2 and 4 kHz. One of the subjects had a threshold at 15 dB hearing level (HL) at 4 kHz, whereas the remaining subjects had thresholds at maximum 10 dB HL. All the subjects were accepted for the test and voluntary participated without getting paid.

2.6 Test course

The 24 normal-hearing subjects were divided into three test groups of eight persons each, which were presented for the traditional test, and the two versions of the user operated test with and without the ?-button, respectively. The subjects in each test group are assumed to have a similar ability to understand speech in noise. By the traditional test the subjects' task was to orally repeat as much heard as possible. In the user operated tests the subjects had for each presented word to select a response from ten alternative words listed in a matrix on a touch screen. The response alternatives were the same for each sentence presented. They were listed alphabetically except for the numerals which were listed in numerical order, see Table 1. In the test with the ?-button, the subjects were told to use this possibility, when they had not heard the word. No further guidance regarding the ?-button was given.

The subject in each test group performed both a training and a test session. The training session contained two test lists (i.e. 100 words) and was carried out according to the procedure described in [8]. The purpose of the training session was to make the test subjects familiar with the test and the test situation. Data from the training session was not included in the results. In the test session eight test lists (i.e. 400 words) were presented at eight different *SNR*'s, from -15 to -1 dB SNR with an increment of 2 dB. The *SNR*'s, which were the same for all three test groups, were set from the reference data for the traditional test [2] to cover the entire discrimination function. The noise level was held constant at 65 dBC, whereas the speech level was adjusted. The presentation order of the sentences at the different *SNR*'s was chosen randomly.

Table 1: The ten alternative Danish words for each presented word. The table is for the version of the user operated test with the possibility to answer "I do not know" (?-button). The five columns represent the different word classes: name, verb, numeral, adjective, and object. For each presented word the subjects had to choose one word from the corresponding column or use the ?-button at the bottom. The words in each column are listed alphabetically except for the numerals which are listed in numerical order. During the test the matrix was visible on a touch screen.

Anders	ejer	tre	fine	biler
Birgit	finder	fem	flotte	blomster
Henning	får	seks	gamle	gaver
Ingrid	havde	syv	hvide	huse
Kirsten	købte	otte	nye	jakker
Linda	låner	ni	pæne	kasser
Michael	ser	ti	røde	masker
Niels	solgte	tolv	sjove	planter
Per	valgte	fjorten	smukke	ringe
Ulla	vandt	tyve	store	skabe
?	?	?	?	?

2.7 Equipment

A special designed measurement program was developed in MATLAB 6.5 according to the test methodic. Under the listening test a laptop with a touch screen (Acer model TravelMate C300XCi) was used. The touch screen was activated by using a special pen. The test sentences and the noise signal were presented monaurally to the subjects with TDH-39P headphones. The headphones were free-field equalized according to [16].

The system was calibrated by means of an artificial ear (Bruel & Kjaer, type 4152), a 1/1 inch microphone (Bruel & Kjaer, type 4132), a preamplifier (Bruel & Kjaer, type 2619), a power supply (Bruel & Kjaer, type 2804) and a measuring amplifier (Bruel & Kjaer, type 2636). The amplification characteristic and the frequency response were verified and found to be linear. The level of the background noise satisfies the tolerances in [16].

2.8 Statistical analyses

For the statistical analyses the computer program SPSS 11.5.1 for Windows was used. Only parametric statistical models were used. The Kolmogorov-Smirnov test was used to ascertain, that different samples could be assumed to come from a normal distribution. Before the ANOVA tests were made, the Levene's test was used to assess the equality of variances in the different samples. All analyses were made at a 0.05 significance level. The different tests used are descrybed in the help-function in the SPSS program.

3 Results

3.1 The influence of the ?-button

For the traditional test and for the two versions of the user operated test with and without the ?-button, the mean values of SRT_N and s_{50} across the eight subjects were determined, see Table 2. The mean values of SRT_N and s_{50} within each test group were found by computing the average values of SRT_N and s_{50}, respectively, based on the subject's individual values. If the mean values of SRT_N and s_{50} were found by adjusting the logistic function direct to the overall average of the correct answers across subjects, the average value of s_{50} would be lower than the ones in the table due to the variations in SRT_N between subjects.

Table 2: Mean values for SRT_N and s_{50} found with the traditional test, and the two versions of the user operated test with and without the ?-button. The values in brackets are one standard deviation. Each value represents data from eight subjects. Each subject was presented to ten sentences of each five word at each of the eight presentation levels, i.e. in total the values for each of the three test groups represents data from a presentation of 3200 words.

	Traditional test	User operated test with ?-button	User operated test without ?-button
SRT_N [dB SNR]	-8.0 (1.10)	-8.2 (0.71)	-9.1 (0.80)
s_{50} [%/dB]	14.6 (2.63)	14.6 (1.62)	13.1 (2.33)

An ANOVA test shows a statistical significant difference ($p = 0.044$) between the values for SRT_N for the three tests in Table 2. A *post hoc* analysis, the Sheffe test reveals, that the difference is between SRT_N for the user operated test without the ?-button and the two other tests. This statistical significant difference is caused by motivated guesses, which will be described further below. The mean value of SRT_N found with the traditional test and the user operated test with the ?-button are comparable with the reference value, SRT_N = -8.4 dB SNR [2].

From the mean values in Table 2 an average discrimination function for each of the three test groups are determined and shown in Figure 1. Figure 1 also shows a theoretically predicted curve for a 10AFC test which is based on using the SRT_N and s_{50} estimated in the traditional test in Equation (3). The difference between the curve for the traditional test and the theoretically predicted curve illustrates correct responded words due to pure guesses (a guess made when the subject had not heard the word at all). The speech intelligibility at a given presentation level is consistently higher for the user operated test without the ?-button than for the theoretically predicted curve, in particular for speech intelligibility in the range of 20% to 80%. The difference between these two curves describes the amount of correct responded words due to motivated guesses (a guess made when a part of the word is heard).

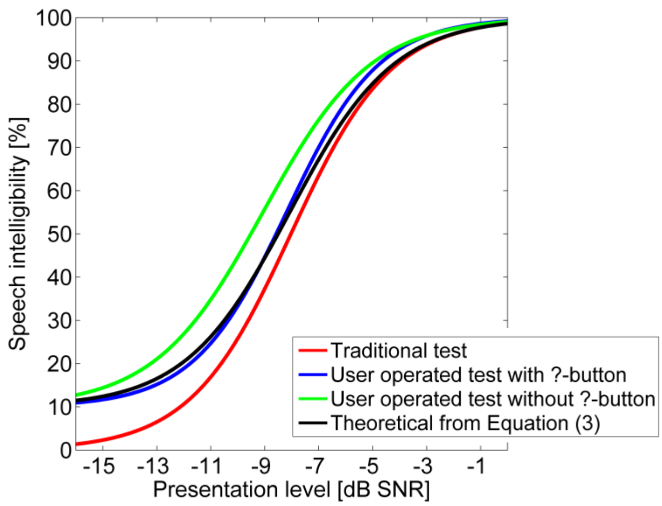

Figure 1: The average discrimination function for each of the three test groups: the traditional test, the user operated test with and without the ?-button. The discrimination functions are plotted from the data in Table 1. The black curve is theoretically determined from the data for the traditional test by setting the values for SRT_N and s_{50} into Equation (3). It shows the expected discrimination function for a 10AFC test.

Figure 1 shows that the curve for the user operated test with the ?-button is almost coincident with the theoretically predicted curve, i.e. the results from the user operated test with the ?-button are comparable with results from the traditional test. Because of the difference between the two curves for the user operated test with and without the ?-button, it can be concluded, that when the subjects have the possibility to press the ?-button, they make fewer motivated guesses, than when they do not have the ?-button available.

3.2 The homogeneity of the speech material

The homogeneity of the speech material can be described by the slope of the discrimination function; the steeper the slope the more homogeneous is the material, i.e. the more even is the difficulty of the different words. By the generation of the speech material done by [2] the sound pressure level of particular words were adjusted in order to optimize the homogeneity of the material. The adjustments were done based on result from a traditional test with 16 normal-hearing subjects. To investigate whether the homogeneity of the speech material is influenced, when the material is used in a user operated test, the values of s_{50} in Table 2 are compared. An ANOVA test shows no statistical significant difference ($p = 0.362$) between the values of s_{50} for the three tests. From that it can be concluded, that the homogeneity of the speech material is not influenced, when the material is used in a user operated test. For all three test groups the mean value of s_{50} is comparable with the reference value, $s_{50} = 13.2$ %/dB found by [2].

4 Discussion

In this study the number of alternative words for each presented word in the user operated tests was set to ten. The more alternative words the lower is the probability for responding the correct word by a pure guess cf. Equation (3). However, the number of alternative words should not be too high due to a possible effect on the short time memory. It may then not be the subjects hearing which is tested, but the person's ability to remember what had been heard when finding the presented word among the alternatives. Ten alternative words for each presented word corresponding to the ten words within the different word classes in the speech material seem reasonable. Using the same response alternatives and listing them in the same order for each sentence presented, makes it easier for the subject to get familiar with the response matrix and thereby to find the word to answer among the alternatives.

If the number of alternative words in a user operated test is reduced, the probability for responding the correct word by a pure guess increases. It also becomes easier to make a motivated guess and by that choose the correct word. The number of correct responded words will then rise due to both pure and motivated guesses. If the number of alternative words is reduced, it has to be decided which words should be the alternative words and whether the alternative words should be the same at repeated presentations of the certain words. If the alternative words sound or are spelled like the presented word, it is harder to make a motivated guess and thereby choose the correct word than if the alternative words are not phonetically like the presented word. The difficulty of a presented word would therefore depend on the alternative words.

Beyond the number of alternative words and the words used as alternative words, it should also be decided whether the possibility to answer ”I do not know” should be available. The advantage of inducing a ?-button is that the subject is not forced to choose an answer to a word, which had not been heard. However, a disadvantage is that different subjects will use a ?-button differently. Some subjects would use the ?-button whenever they are not quite sure which word had been presented, whereas others will choose a word instead and almost never use the ?-button. To equalize a potential differrent use of the ?-button, each press on the button was converted into a 1/10 correct answer in this study. A statistical significant difference between SRT_N was found with the user operated test without the ?-button and the two other tests (the traditional test and the user operated test with the ?-button). When the subjects have the possibility to press the ?-button, they make fewer motivated guesses, than when they do not have the ?-button available. Hence, including the ?-button seems favorable.

The subjects' answers depend not only on the presentations level but also on variations in the subjects' concentration level, fluctuations in the noise signal and a possibly uneven difficulty for the different words in the speech material. As described above the difficulty of the different words in the user operated tests is depending on the alternative words. To equalize the difficulty of the different words the phonetic likeness of the alternatives must then be comparable across the presented words. In this study there was found no statistically difference between s_{50} and therefore the homogeneity of the speech material of the traditional test seems to be preserved in the two versions of the user operated test.

A user operated test designed as the one in this study requires that the subject can read. Another disadvantage is that the test has a longer test time for the subject than a traditional test. However, the test time for the operator and a possible influence of the operator on the evaluation of the subject response are reduced compared with a traditional test. Another advantage is that a user operated test can be designed for use for hearing screening e.g. on the internet due to the less central role of the operator in the test course.

5 Conclusions

The Danish speech material Dantale II can be used for a user operated speech in noise test. In the user operated test designed in this study the subject had to select a response from ten alternative words listed in a matrix for each presented word. The user operated test was designed in two versions: one with the possibility for each word to answer "I do not know" (?-button) and one without that possibility. A press on the ?-button was converted into a 1/10 correct response corresponding to a random pure guess among the ten possible words. This conversion was done to equalize a potential different use of the ?-button among the subjects.

A listening test among 24 normal-hearing subjects was performed to compare the two versions of the user operated test with a traditional test. The listening test shows that results from the user operated test with the ?-button equals results found in a traditional test, and that not was the case for the user operated test without the ?-button. When the subjects have the possibility to press the ?-button, they then make fewer motivated guesses than when they don't have the ?-button available. Additionally, the listening test shows that the homogeneity of the speech material is uninfluenced when the material is used in a user operated test compared to when the speech material is used in a traditional test.

References

[1] C. Elberling, C. Ludvigsen and P. E. Lyregaard, Dantale, A New Danish Speech Material, *Scand. Audiol.*, 18(3), 1989, 169-175.

[2] K. C. Wagener, J. L. Josvassen and R. Ardenkjaer, Design, optimization and evaluation of a Danish sentence test in noise, *Int. J. Audiol.*, 42(1), 2003, 10-17.

[3] J. B. Nielsen and T. Dau, The Danish hearing in noise test, *Int. J. Audiol.*. 50(3), 2011, 202-208.

[4] B. Hagerman, Sentences for testing speech intelligibility in noise, *Scand. Audiol.*, 11(1), 1982, 79-87.

[5] K. Wagener, T. Brand and B. Kollmeier, Entwicklung und Evaluation eines Satztests für die deutsche Sprache Teil II: Optimierung des Oldenburger Satztests (Development and evaluation of a German sentence test part II: Optimization of the Oldenburg sentence test), *Zeitschrift für Audiologie,* 38(2), 1999, 44-56.

[6] K. Wagener, T. Brand and B. Kollmeier, Entwicklung und Evaluation eines Satztests für die deutsche Sprache Teil III: Evaluation des Oldenburger Satztests (Development and evaluation of a German sentence test part II: Evaluation of the Oldenburg sentence test), *Zeitschrift für Audiologie*, 38(3), 1999, 86-95.

[7] K. Wagener, V. Kühnel and B. Kollmeier, Entwicklung und Evaluation eines Satztests für die deutsche Sprache Teil I: Design des Oldenburger Satztests (Development and evaluation of a German sentence test part I: Design of the Oldenburg sentence test), *Zeitschrift für Audiologie,* 38(1), 1999, 4-15.

[8] T. Brand and B. Kollmeier, Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests, *J. Acoust. Soc. Am.*, 111(6), 2002, 2801-2810.

[9] A. Boothroyd, Statistical Theory of the Speech Discrimination Score, *J. Acoust. Soc. Am.*, Volume 43(2), 1968, 362-367.

[10] M. C. Leensen, J. A. de Laat and W. A. Dreschler, Speech-in-noise screening tests by internet, part 1: test evaluation for noise-induced hearing loss identification, *Int. J. Audiol.*, 50(11), 2011, 823-834.

[11] M. C. Leensen, J. A. de Laat, A. F. Snik and W. A. Dreschler, Speech-in-noise screening tests by internet, part 2: improving test sensitivity for noise-induced hearing loss, *Int. J. Audiol.*, 50(11), 2011, 835-848.

[12] T. Brand, T. Wittkop, K. C. Wagener and B. Kollmeier, Vergleich von Oldenburger Satztest und Freiburger Wörtertest als geschlossene Versionen (Comparison of the Oldenburg sentence test and the Freiburg word test as closed versions), *Proceedings of the 7th annual meeting of the Deutsche Gesellschaft für Audiologie (DGA),* 2004, Leipzig, 3 pages.

[13] E. R. Pedersen and P. M. Juhl, An Alternative Noise Signal for Danish Speech in Noise Tests. *Proceedings of Forum Acusticum 2011*, Aalborg, Denmark, 6 pages.

[14] D. M. Green, V. M. Richards and T. G. Forrest, Stimulus step size and heterogeneous stimulus conditions in adaptive psychophysics, *J. Acoust. Soc. Am.*, Volume 86(2), 1989, 629-636.

[15] W. K. Newey and D. McFadden, Chapter 36: Large sample estimation and hypothesis testing, In Engle, R., and McFadden, D. *Handbook of Econometrics*, Vol. 4. Elsevier Science, 1994, 2111–2245.

[16] IEC 60645-2 Ed. 1.0 B:1993, Audiometers – Part 2: Equipment for speech audiometry, *International Electrotechnical Commission*, Switzerland.