Examination of the learning effect with the Dantale II speech material

ELLEN RABEN PEDERSEN* AND PETER MØLLER JUHL

The Maersk Mc-Kinney Moller Institute, University of Southern Denmark, Odense M, Denmark

This study examines the learning effect when using the Danish speech material Dantale II to determine the speech reception threshold (SRT) in noise under three different test conditions. The learning effect is shown by an improvement of the test result, i.e., by a decrease in the value of SRT at repeated measurements until a certain number of measurements has been made. A listening test was performed with 24 normal-hearing subjects. The purpose of the test was to investigate the influence of the target level on the learning effect in an open-set test format, where the subject's task is to orally repeat as much as possible of the sentence just presented. The target level was set to 50% and 80% correctly understood words, respectively. Furthermore, the purpose was to investigate whether using a closed-set test format affects the learning effect. In the closed-set test format the subject had, for each word presented, to select a response from ten alternative words. Statistical analyses of the test results did not show any significant differences in neither the within-visit learning effect nor the inter-visit learning effect for the two target levels or for the different test formats. However, the learning effect was found to be finished faster for the open-set test format with a target level of 80% than for the two other conditions.

INTRODUCTION

Over the years different speech-in-noise tests have been developed for determining the speech reception threshold (SRT). The tests have been applied both in the clinical practice and in hearing research. A commonly known speech material is the Danish Dantale II speech material (Wagener *et al*., 2003), which consists of syntactically fixed but semantically unpredictable test sentences and an almost stationary noise signal. The speech material Dantale II is developed in analogy to the materials for the Swedish Hagerman test (Hagerman, 1982) and the German Oldenburg sentence test (Wagener *et al*., 1999). Within the European HearCom project the material has also been developed for other languages, e.g., Polish (Ozimek *et al*., 2010), Spanish (Hochmuth *et al*., 2012), and French (Jansen *et al*., 2012).

It is known that, when using a speech material as the Dantale II speech material, a learning (or training) effect is present. The learning effect is shown by an improvement of the test result, i.e., by a decrease in the value of SRT at repeated

*Corresponding author: erpe@mmmi.sdu.dk

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

measurements until a certain number of measurements has been made – then the SRT values only vary with the uncertainty of the measurement. At the development of the Dantale II speech material Wagener *et al*. (2003) found a learning effect of 2.2 dB in an open-set test format. Wagener *et al*. (2003) performed eight subsequent measurements of SRT (containing 20 test sentences each) on normal-hearing subjects. The learning effect was determined as the difference between SRTs obtained at the first and eighth measurement. If two lists of 20 sentences were performed as training prior to an actual measurement, the learning effect was found to affect SRT by less than 1 dB (Wagener *et al*., 2003).

Hernvig and Olsen (2005) also investigated the learning effect using the Dantale II speech material in an open-set test format. The study distinguishes between two types of learning effects: the within-visit learning effect and the inter-visit learning effect. The within-visit learning effect corresponds to the learning effect investigated by Wagener *et al*. (2003), whereas the inter-visit learning effect is a learning effect found between SRT measurements that are substantially separated in time. Hernvig and Olsen (2005) performed six subsequent measurements of SRT (containing 30 test sentences each) on hearing-impaired subjects and found a within-visit learning effect (the difference between SRT determined at the first and the sixth measurement) of 3.2 dB. The inter-visit learning effect was found to be 1.6 dB with a median inter-visit period of 27 days (range: 14-43 days).

A within-visit learning effect in an open-set test format has also been found for the Swedish Hagerman test (Hagerman, 1984; Hagerman and Kinnefors, 1995), the German Oldenburg sentence test (Wagener *et al*., 1999) and the corresponding French test (Jansen *et al*., 2012). In a study by Brand *et al*. (2004) the within-visit learning effect has been investigated for the German material in a quasi closed-set test format and compared to that for an open-set test format. In the quasi closed-set test format the subject had, for each word presented, to select a response from ten alternative words (corresponding to the different words in the speech material) or they could answer 'I do not know' (each 'I do not know' answer was interpreted as an incorrect answer). The alternative answers for each word were listed in a matrix on a computer screen. The within-visit learning effect was found to be comparable in the quasi closed-set test format and in the open-set test format. A corresponding finding was made with the Spanish material (Hochmuth *et al*., 2012).

Even though previous studies showed that a learning effect exists and that it is needed to present a subject with training lists prior to an actual measurement, it is unknown what causes the learning effect. Some studies indicated that the observed learning effect is due to the sentences having a syntactically fixed structure and the number of different words in the material being limited (Hernvig and Olsen, 2005; Wagener and Brand, 2005). However, the studies by Brand *et al*. (2004) and Hochmuth *et al*. (2012) found no difference in the within-visit learning effect for an open-set and quasi closed-set test format. A difference would have been expected if the learning effect is caused by the composition of the speech material, since the subjects in the quasi closed-set test format were visually presented to the different words in the material.

It is interesting to study the learning effect because the number of lists required for training influences the total test time. To the authors' knowledge no previous study has investigated the influence of the target level (i.e., the level at which the sentences are presented) on the learning effect. If the learning effect is caused by the composition of the speech material, the learning effect could be expected to be influenced when the test sentences are presented at a target level higher than the normal 50% correctly understood words, i.e., when the subject hears more of the words presented. Furthermore, no study on the learning effect has to the authors' knowledge previously been performed with the Dantale II speech material in a quasi closed-set or a closed-set test format. Therefore, this study investigated whether the target level affects the learning effect (in an open-set test format) and whether the learning effect is influenced by a closed-set test format using the Dantale II speech material. In the closed-set test the subject had, for each word presented, to select a response from ten alternative words without the possibility to answer 'I do not know'. For each test condition both the within-visit learning effect and the inter-visit learning effect were determined.

METHODS

Speech material

The Danish speech material Dantale II (Wagener *et al*., 2003), which was used in this study, consists of 16 lists with ten test sentences each. The test sentences have a syntactically fixed structure of five words from different word classes in the order: name, verb, numeral, adjective, and noun. Since the test sentences are semantically unpredictable, the words cannot be predicted from the context. 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'). The noise signal included in the speech material was generated by superimposing the test sentences many times by which the signal became speech-shaped without strong fluctuations.

Test versions

Three test versions were implemented: two with an open-set test format and one with a closed-set test format. In the two versions with the open-set test format the subject had to orally repeat as much heard as possible after each sentence presented. The operator then registered whether the subject's answer for each word was correct or incorrect. In the version with the closed-set test format the subject had to select a response from ten alternative words listed in a matrix for each word presented. The subject did not have the possibility to answer 'I do not know', i.e., the subject was forced to guess when a word had not been heard.

All three test versions were implemented using the adaptive procedure described in Brand and Kollmeier (2002). The presentation level, i.e., the signal-to-noise ratio (SNR) at which the sentences was presented, was adjusted from sentence to sentence depending on the number of correctly answered words given to the previous sentence, and on the advance of the test to stabilize the SNRs near the target level.

The adjusting was done by changing the level of the test sentences, whereas the level of the noise signal was kept constant at 65 dBC. The first sentence was presented at 0 dB SNR. For the two versions with the open-set test format the target level was set to 50% and 80% correctly understood words, respectively. For the version with the closed-set test format the target level was set to 50%.

Equipment

A specially-designed measurement program was developed in MATLAB 6.5 according to the three test versions and the adaptive procedure. Under the listening test a laptop with a touch screen (Acer model TravelMate C300XCi) was used. The subjects who were presented to the closed-set test format had to use the touch screen to give their answers after each sentence presented. The test sentences and the noise signal were presented to the subjects by a loudspeaker (Vifa P13WH00-08 in a 6.6-litres vented cabinet), which was connected to the laptop through a power amplifier (Bruel & Kjaer, type 2706). The subjects were seated 1.2 m in front of the loudspeaker.

Subjects

The listening test was performed with 24 normal-hearing subjects (12 males and 12 females, aged 21-39 years with a mean age of 26 years). The subjects were native speakers of Danish and had not been presented to the Dantale II speech material before the actual listening test. They had no otological problems and their hearing thresholds did not exceed 15 dB HL at the frequencies 0.5, 1, 2, and 4 kHz. The subjects participated in the study voluntarily without getting paid.

Test course

The 24 normal-hearing subjects were divided into three test groups of eight persons each, who were presented the two versions with the open-set test format and the version with the closed-set test format, respectively. For each subject eight subsequent measurements of SRT using two lists each were made to determine the within-visit learning effect. The subjects were presented to each of the 16 test lists in the speech material once. To avoid any effect of the list sequences, the presentation order of the lists was counterbalanced among the subjects. After a period of 12-16 days (mean: 14 days) one more measurement of SRT was made to determine the inter-visit learning effect. As at the first visit, the measurement of SRT included two lists. Within each of the three groups none of the subjects were presented to the same lists at the second visit.

Statistical analyses

For the statistical analyses the computer program SPSS 11.5.1 for Windows was used (www.spss.co.in). All analyses were performed at a 0.05 significance level. The Kolmogorov-Smirnov test was used to ascertain whether data for the different test conditions could be assumed to come from a normal distribution, and the Levene test was used to test for homogeneity of variance. To test for differences between the SRT values obtained, parametric tests were used, provided that the conditions for performing those tests were satisfied. Otherwise corresponding non-parametric tests were used.

RESULTS

Figure 1 shows the results from the SRT measurements for each of the three different test conditions, where the SRT values are given at the representative target level. From the figure it is seen that the mean value of SRT decreased at repeated measurements (indicating that the subjects scored 'better') until a certain number of measurements had been made. The curves for the three conditions have a similar shape but are vertically displaced. The highest SRT values were obtained for the subjects who were presented to the open-set test format with a target level of 80%. The higher target level causes the sentences to be presented at higher SNRs than for a target level of 50%, which results in higher SRT values. The SRT values are higher for the open-set test format than for the closed-set test format both with a target level of 50%. This can be explained by the fact that the subjects who were presented to the closed-set test format had the different words (response alternatives) listed in a matrix as a visual cue, which makes it easier to guess the correct word from the alternatives.

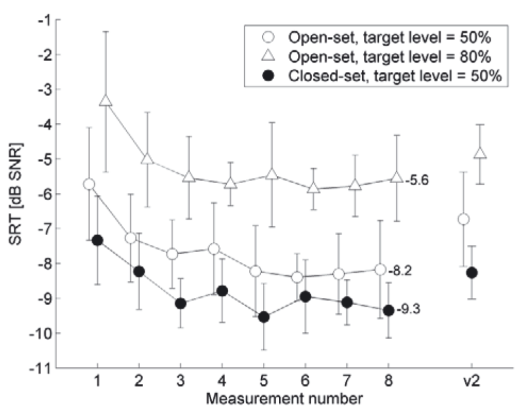

Fig. 1: Results of the SRT measurements as function of measurement number for the three test groups, which were presented to the two versions with the open-set test format and the version with the closed-set test format, respectively. For each measurement the mean SRT and one standard deviation are determined across eight normal-hearing subjects. The results marked 1 to 8 were obtained at the first visit, whereas the results marked v2 were obtained at the second visit, which took place 12-16 days after the first visit.

Figure 2 shows the within-visit learning and inter-visit learning effect. A Kruskal-Wallis test showed for the within-visit learning effect no statistical difference between the three test groups ($X^2(2) = 0.060$, $p = 0.970$). This finding for the open-set test format with a target level of 50% and the closed-set test format is in agreement with previous studies (Brand *et al.*, 2004; Hochmuth *et al.*, 2012). For all three test conditions the within-visit learning effect was comparable to the learning effect of 2.2 dB found by Wagener *et al.* (2003).

For the inter-visit learning effect shown in Fig. 2 a Kruskal-Wallis test showed no statistical difference between the three test groups ($X^2(2) = 1.005$, $p = 0.605$). For all three test conditions the inter-visit learning effect was lower than the inter-visit learning effect found by Hernvig and Olsen (2005) with hearing-impaired subjects. The within-visit learning effect in this study was also lower than in the study by Hernvig and Olsen (2005).

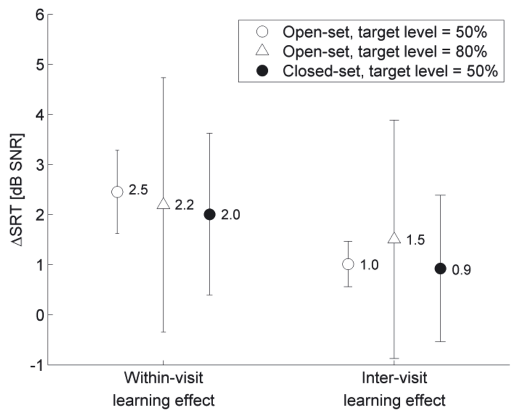

Fig. 2: Mean and one standard deviation of the within-visit and inter-visit learning effect. The within-visit is calculated as the difference between SRT obtained at the first and eighth measurement, whereas the inter-visit learning effect is calculated as the difference between SRTs obtained at the first measurement in the two visits.

To analyse when the within-visit learning effect can be assumed to be finished, paired-sampled *t*-tests (2-tailed) were performed between SRT values obtained at different measurements. For the SRT values obtained at the third and eighth measurement the tests showed no statistical difference for any of the three test conditions ($t(7) = 2.091$, $p = 0.075$; $t(7) = 0.024$, $p = 0.982$; $t(7) = 1.061$, $p = 0.324$), i.e., the learning effect can be assumed to be finished after two measurements. The difference between the second and eighth measurement were also analysed. For the open-set test format with a target level of 50% and for the closed-set test format a statistical difference was found ($t(7) = 2.927$, $p = 0.022$; $t(7) = 2.721$, $p = 0.030$), i.e., the learning effect cannot be assumed to be finished after only one measurement. For the open-set test format with a target level of 80% no statistical difference were found ($t(7) = 0.692$, $p = 0.511$), i.e., for this test condition the learning effect can be assumed to be finished after only one measurement.

DISCUSSION

The learning effect was found to be finished after only one measurement for the open-set test format with a target level of 80%. However, the learning effect for the closed-set test format was not found to be finished until after two measurements. Therefore, it is not possible to conclude whether the cause of the learning effect is dominated by the composition of the speech material or by the subjects having to adapt to the test situation and to listening for the words in the noise signal.

It could be interesting to investigate the learning effect in further details in a future study in order to obtain more insight in its causes, e.g., it could be interesting to investigate whether there is any difference between the learning effect obtained with a speech material as the Danish Dantale II speech material and a speech material containing everyday sentences (sentences without a syntactically fixed structure and with an unlimited number of words).

CONCLUSIONS

No statistical differences were found either in the within-visit or in the inter-visit learning effect for the three conditions tested. However, for the open-set test format with a target level of 80%, the learning effect was found to be finished faster than for the two other conditions.

Like previous studies this study shows the need for presenting the subjects with training lists prior an actual measurement to remove the effect of learning on the test result. Two training lists of 20 sentences seem sensible. The number of training lists might be reduced to one for an open-set test format with a target level of 80%. If the subject has been presented to the material within a short period of time training can be reduced.

ACKNOWLEDGEMENTS

The authors would like to thank Carsten Daugaard and Søren L. Jørgensen from DELTA Technical Audiological Laboratory, Odense, Denmark for providing some of the equipment and for valuable discussions. The authors also thank all the subjects who participated in the tests.

REFERENCES

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

Brand, T., Wittkop, T., Wagener, K., and Kollmeier, B. (**2004**). "Vergleich von Oldenburger Satztest und Freiburger Wörtertest als geschlossene Versionen," (in German), Deutsche Gesellschaft für Audiologie (DGA).

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

Hagerman, B. (**1984**). "Some aspects of methodology in speech audiometry – Studies of reliability, computer simulations and development of a new speech material for measuring speech reception threshold in noise," Scand. Audiol. Suppl., **21**, 1-25.

Hagerman, B., and Kinnefors, C. (**1995**). "Efficient adaptive methods for measuring speech reception threshold in quiet and in noise," Scand. Audiol, **24**, 71-77.

Hernvig, L.H., and Olsen, S.O. (**2005**). "Learning effect when using the Danish Hagerman sentences (Dantale II) to determine speech reception threshold," Int. J. Audiol., **44**, 509-512.

Hochmuth, S., Brand, T., Zokoll, M.A., Castro, F.Z., Wardenga, N., and Kollmeier, B. (**2012**). "A Spanish matrix sentence test for assessing speech reception thresholds in noise," Int. J. Audiol., **51**, 536-544.

Jansen, S., Luts, H., Wagener, K.C., Kollmeier, B., Del Rio, M., Dauman, R., James, C., Fraysse, B., Vormes, E., Frachet, B., Wouters, J., and van Wieringen, A. (**2012**). "Comparison of three types of French speech-in-noise tests: a multi-center study," Int. J. Audiol., **51**, 164-173.

Ozimek, E., Warzybok, A., and Kutzner, D. (**2010**). "Polish sentence matrix test for speech intelligibility measurement in noise," Int. J. Audiol., **49**, 444-454.

Wagener, K.C., and Brand, T. (**2005**). "Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: Influence of measurement procedure and masking parameters," Int. J. Audiol., **44**, 144-156.

Wagener, K., Brand, T., and Kollmeier, B. (**1999**). "Entwicklung und Evaluation eines Satztests für die deutsche Sprache – Teil III: Evaluation des Oldenburger Satztests," (in German), Zeitschrift für Audiologie, **38**, 86-95.

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