Informational masking in speech intelligibility tests

ELLEN RABEN PEDERSEN[1,*], HODA EL-SAMAIL[2], CARSTEN DAUGAARD[3], PETER MØLLER JUHL[1], AND TURE ANDERSEN[4, 5]

[1] *The Maersk Mc-Kinney Moller Institute, University of Southern Denmark, Odense M, Denmark*

[2] *Audiology and Logopedics Studies, University of Southern Denmark, Odense M, Denmark*

[3] *DELTA Technical-Audiological Laboratory, Odense C, Denmark*

[4] *Institute of Clinical Research, University of Southern Denmark, Odense M, Denmark*

[5] *Department of Audiology, Odense University Hospital, Odense C, Denmark*

It is often challenging to separate speech from a noise – especially for hearing-impaired persons. A particular difficult listening situation is when speech is obscured by speech from one or more simultaneous talkers. The purpose of this study is to investigate the effect of informational masking on the speech reception threshold (SRT) and to compare the SRT values obtained with subjective data from the SSQ questionnaire. A listening test was performed with 20 normal-hearing and 20 hearing-impaired subjects. The subjects were presented to the sentences from the Danish speech material Dantale II in four different speech-shaped interfering maskers. The maskers differ regarding fluctuation and to what extent they represent intelligible speech. The listening test shows that the three fluctuating maskers distinguish better between normal-hearing and hearing-impaired subjects than the almost stationary masker. The test-retest variation was found to be the same for the four maskers. The SRT values for the four maskers were generally found not to correlate with the hearing-impaired subjects' answers to specific questions in the SSQ questionnaire.

INTRODUCTION

Understanding speech in noise is a challenging task for people in general and especially for hearing-impaired persons – a particular difficult listening situation is when the masker is speech from one or more simultaneous talkers. Therefore speech-in-noise tests are routinely carried out in clinics in order to assess the degree of the hearing loss and the effect of treatment. However, the results of the tests are often in disagreement with the problems that the subjects report. One reason for this difference might be that the masker used in the clinical test does not represent real-life maskers well, by which the tests are dominated by energy masking and only to a

*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.

limited amount involve informational masking (for a review on informational masking, see Schneider *et al.*, 2007).

The purpose of this study is to investigate how four different maskers influence the result of a speech-in-noise test – i.e., the speech reception threshold (SRT) – for both normal-hearing and hearing-impaired subjects. The maskers differ regarding fluctuation, to what extent they represent intelligible speech, and were expected to cause different amount of informational masking. Three study questions were addressed: 1) Do the different maskers influence the test sensitivity differently, i.e., are the different maskers equally good/bad at distinguishing between normal-hearing and hearing-impaired subjects?, 2) Is the test-retest variation affected by the different maskers?, and 3) Are the test results obtained with the four maskers in agreement with the subject's own experience of his/her ability to understand speech in noise?

METHODS

Target signal

As a target signal the test sentences from the Danish speech material Dantale II (Wagener *et al*., 2003) were used. The material consists of 16 lists with ten test sentences each. The test sentences are spoken by a female speaker and have a fixed structure of five words from different word classes in the order: name, verb, numeral, adjective, and noun. 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').

Masker signals

During the listening test the subjects were presented to the target signal in four different speech-shaped interfering maskers, which are expected to cause different amount of informational masking. Below is a short description of each of the four signals. The Dantale II noise is almost stationary, whereas the three other signals fluctuate comparably to natural speech. For the listening test the overall RMS level of the three fluctuating signals was adjusted to that of the Dantale II noise. The signal named 2FS was specially generated for this study and is the only of the four signals representing intelligible speech.

<u>Dantale II noise</u>: This signal is included in the Dantale II speech material. It is generated by superimposing the test sentences many times by which the signal becomes almost stationary (Wagener *et al*., 2003).

<u>ICRA-4</u>: This signal is made by the International Collegium for Rehabilitative Audiology (the number four refer to track no. 4 on the ICRA CD). The signal is artificial and represents one female speaker (Dreschler *et al*., 2001).

<u>IFFM</u>: This signal is based on the International Speech Test Signal (ISTS) but with limited pause durations (www.ehima.com). The ISTS contains fragments of recordings from female speakers talking different languages (Holube *et al*., 2010).

2FS: This signal was generated by making a sequence of nine Dagmar sentences and a sequence of nine Asta sentences from the DAT corpus (Nielsen *et al*., 2011) and storing them in different channels. The name 2FS refers to the signal containing '2 Female Speakers'.

Test setup

The SRT measurements were performed with the software HearVal 1.0.0.8, which is developed in LabVIEW 2010 by DELTA. Three active Genelec 1029A loudspeakers were positioned in the horizontal plane at a distance of 1.4 m from the subject at different angles. The target signal was presented frontal to the subject at an angle of incidence of 0^{o}, whereas the masker was presented by two loudspeakers symmetrically located at the angles of incidence of $\pm 45^{\circ}$ (in accordance with the recommendation in DS/EN ISO 8253-3:2012). A laptop (IBM ThinkPad R51 Type 1829-R6G) was used to play and control the level of the target signal, while another laptop (IBM ThinkCentre MT-M type 9210-D1G) was used to play the masker. The masker was played incoherently from the two spatially-separated loudspeakers.

Measurement procedure

In each SRT measurement the presentation levels, i.e., the signal-to-noise ratios (SNRs) at which the sentences were presented, were adjusted (to a speech understanding of 50%) according to the adaptive procedure described in Brand and Kollmeier (2002). The adjustment was done by changing the target level, whereas the level of the masker was kept constant at 65 dBC for the normal-hearing subjects and at 80 dBC for the hearing-impaired subjects. The first sentence was presented at 0 dB SNR. After the presentation of each sentence the subjects orally repeated the words that were perceived, whereupon the test operator registered whether the subject's answer was correct or incorrect to control the SNR of the next sentence. The measurement stopped when one of the two following criteria were met: 1) the subject had been presented to three entire lists, i.e., 30 sentences or 2) ten reversals of the presentation level were attained (a reversal is attained when the change in SNR alters sign). When the measurement stopped the SRT was determined as a mean of the SNRs at the four last reversals.

Subjects

The listening test was performed with 20 normal-hearing subjects (eight males and 12 females, aged 18-26 years with a mean age of 21 years) and 20 hearing-impaired subjects (seven males and 13 females, aged 18-65 years with a mean age of 45 years). The normal-hearing subjects had no otological problems and their hearing thresholds did not exceed 20 dB HL at the octave frequencies from 0.25 to 8 kHz. The hearing-impaired subjects had varying degrees of a bilateral sensorineural hearing loss. Their pure-tone averages (PTAs) for the frequencies 0.5, 1, 2, and 4 kHz were 17.5-66.9 dB HL (mean: 38.4 dB HL). The hearing-impaired subjects were hearing-aid users, but they did not use their hearing aids during the listening test.

Questionnaire

In order to investigate whether the SRT values obtained agreed with subjective data the hearing-impaired subjects were presented to part 1 of a Danish version of the SSQ questionnaire (Gatehouse and Noble, 2004). Part 1 contains 14 questions regarding hearing speech in competing contexts. The subjects were asked to respond on a scale from 0 to 10 ('not at all' to 'perfectly') and to focus on listening situations where they did not use their hearing aids when filling in the questionnaire. The subjects' answers to four of the questions (Q1, Q5, Q9, and Q11) were chosen for comparison with the SRT values obtained – one of the questions (Q5) describes a listening situation with an almost stationary background noise, whereas the three others describe listening situations with fluctuating background noises:

Q1: You are talking with one other person and there is a TV on in the same room. Without turning the TV down, can you follow what the person you're talking to says?

Q5: You are talking with one other person. There is continuous background noise, such as a fan or running water. Can you follow what the person says?

Q9: Can you have a conversation with someone when another person is speaking whose voice is different in pitch from the person you're talking to?

Q11: You are in conversation with one person in a room where there are many other people talking. Can you follow what the person you are talking to is saying?

Test course

The normal-hearing subjects participated in two test sessions, which were separated by 20-62 days (mean: 31 days). The hearing-impaired subjects participated only in one test session. During each test session the subjects were presented to four SRT measurements containing the four different speech-shaped interfering maskers. For each measurement three different lists were randomly chosen. To avoid any effect of the presentation sequences on the results, the presentation order of the maskers was counterbalanced among the subjects. Before each measurement the subjects were presented to sentences for training masked by the same masker as in the subsequent measurement. The training contained three lists before the first measurement and one list before each of the following measurements (for both test sessions). After the SRT measurements the hearing-impaired subjects filled in part 1 of the Danish SSQ questionnaire.

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 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. All correlation analyses were consistently made with the non-parametric Spearman's rank-order correlation, even though some could be made with the parametric Pearson's product-moment correlation.

RESULTS

Normal-hearing vs. hearing-impaired

Figure 1 shows the results from the SRT measurements for both the normal-hearing and hearing-impaired subjects. From the figure it is seen that the hearing-impaired subjects obtained higher (poorer) SRT values than the normal-hearing subjects for all four maskers. The standard deviations are also higher for the hearing-impaired subjects than for the normal-hearing subjects indicating that the hearing-impaired subjects were a more inhomogeneous group. The figure shows that the SRT values obtained for the normal-hearing and hearing-impaired subjects differ more for the three fluctuating maskers than for the Dantale II noise, i.e., the three fluctuation maskers are better than the Dantale II noise to distinguish between normal-hearing and hearing-impaired subjects.

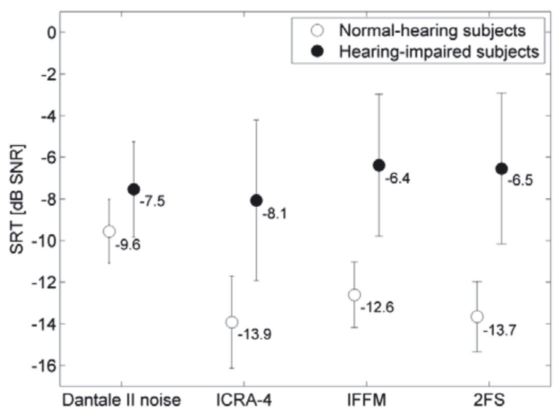

Fig. 1: Mean and one standard deviation of the SRT values for each of the four different maskers obtained with 20 normal-hearing subjects and 20 hearing-impaired subjects, respectively. The SRT values for the normal-hearing subjects are obtained at the first test session.

For the normal-hearing subjects a one-way ANOVA test showed a difference between the mean values of SRT obtained with the four maskers ($F(3,76) = 25.402$, $p = 0.000$). The post hoc Scheffe test revealed that the difference is between the mean SRT values obtained with the Dantale II noise and with the three other maskers. For the hearing-impaired subjects the non-parametric Kruskal-Wallis test showed no statistical difference between the mean SRT values obtained with the four maskers ($X^2(3) = 4.462$, $p = 0.216$).

Test-retest variation

Figure 2 shows the difference between the SRT values obtained at the two test sessions for the normal-hearing subjects. One of the normal-hearing subjects did not complete the second test session, by which the SRT values shown in the figure only are for 19 subjects. The negative differences indicate that the subjects obtained lower (better) SRT values in the second test session than in the first test session. The one-way ANOVA test showed no statistical difference between the SRT difference for the four maskers ($F(3,72) = 0.460$, $p = 0.711$), i.e., the test-retest variation was found to be independent of the type of masker used.

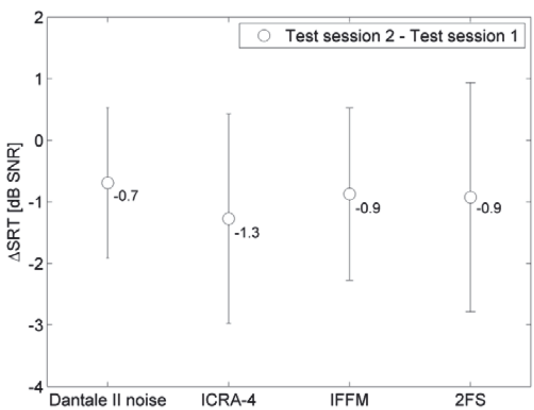

Fig. 2: Mean and one standard deviation of the SRT difference for each of the four different maskers calculated based on measurements from 19 normal-hearing subjects.

Comparison with subjective data

The mean of the answers to the four selected questions given by the hearing-impaired subjects were (one standard deviation is given in the brackets): Q1 = 5.0 (2.1), Q5 = 5.1 (2.0), Q9 = 4.0 (2.2), and Q11 = 3.4 (1.8). In order to compare the SRT values obtained with the subjects' answers to the questions correlation analyses were performed. Two of the subjects did not fill in the questionnaire. Thus the analyses only include data from 18 subjects. For the SRT values obtained with the Dantale II noise and the subjects' answers to Q5 the non-parametric Spearman's rank-order correlation showed no statistical correlation (ρ = −0.368, p = 0.132). Table 1 shows the correlations between the SRT values obtained with the three fluctuating maskers and the subjects' answers to questions Q1, Q9, and Q11. Only the correlation between the SRT values obtained with the 2FS noise and the subjects' answers to Q9 was statistically significant.

	Q1	**Q9**	**Q11**
ICRA-4	$\rho = -0.099, p = 0.695$	$\rho = -0.450, p = 0.061$	$\rho = -0.287, p = 0.248$
IFFM	$\rho = -0.005, p = 0.983$	$\rho = -0.307, p = 0.215$	$\rho = -0.103, p = 0.684$
2FS	$\rho = -0.125, p = 0.622$	$\rho = -0.536^*, p = 0.022$	$\rho = -0.410, p = 0.091$

Table 1: Spearman's rank-order correlation between SRT values obtained with the three fluctuating maskers and scores for specific questions in the SSQ questionnaire. Each correlation contains data from 18 hearing-impaired subjects. Significant results at a 0.05 level are marked with an asterisk (*).

DISCUSSION

For the normal-hearing subjects the SRT values were found to be lower for the three fluctuating maskers than for the almost stationary Dantale II noise, whereas no difference was found between the SRT values obtained with the four maskers for the hearing-impaired subjects. This finding may be due to the normal-hearing subjects being able to benefit from the silent intervals in the fluctuating maskers by *listening in the dips*, whereas the hearing-impaired subjects do not seem to benefit from those intervals.

The test-retest variation was found to be independent of the type of masker used for the group of normal-hearing subjects. However, the test-retest variation was expected to differ for the different maskers. This is due to the silent intervals in the fluctuating maskers, where parts of the target signal can be heard through the maskers – sometimes the part heard may contain speech sounds from which the word can be *guessed*, whereas at other times this will not be the case. Hence, a larger variation in the test results was expected, which would also result in a larger test-retest variation.

The correlation analyses show that the lowest p-values were obtained with the 2FS noise. This indicates that, when using a masker representing intelligible speech, the SRT value is in better agreement with the subject's own experience of his/her ability to understand speech in noise, i.e., the test seems to be more valid than when using the other maskers.

CONCLUSIONS

Analysing the results from the listening test gave the following answers to the three study questions: 1) The three fluctuating maskers distinguish better between normal-hearing and hearing-impaired subjects than the almost stationary Dantale II noise, 2) The test-retest variation was not found to be affected by the different maskers used, and 3) The SRT values for the maskers were generally found not to correlate with the hearing-impaired subjects' answers to specific questions in the SSQ questionnaire. Only the correlation between the SRT values obtained with the 2FS noise and the subjects' answers to Q9 was found to be statistically significant.

ACKNOWLEDGEMENTS

The authors would like to thank Chris Vigandt Jørgensen from DELTA Technical-Audiological Laboratory, Odense, Denmark for helping with the test setup. The authors would also like to thank all subjects who participated in the listening test.

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.

Dreschler, W.A., Verschuure, H., Ludvigsen, C., and Westermann, S. (**2001**). "ICRA noises: artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment. International Collegium for Rehabilitative Audiology," Audiology, **40**, 148-157.

DS/EN ISO 8253-3:2012 (**2012**). *Acoustics – Audiometric test methods – Part 3: Speech audiometry*, 2nd Ed. (Copenhagen, Denmark).

Gatehouse, S., and Noble, W. (**2004**). "The speech, spatial and qualities of hearing scale (SSQ)," Int. J. Audiol., **43**, 85-99.

Holube, I., Fredelake, S., Vlaming, M., and Kollmeier, B. (**2010**). "Development and analysis of an International Speech Test Signal (ISTS)," Int. J. Audiol., **49**, 891-903.

Nielsen, J.B., Neher, T., and Dau, T. (**2012**). "Towards a Danish speech material for speech-on-speech masking investigations," in Proceedings of ISAAR 2011: *Speech perception and auditory disorders*. 3rd International Symposium on Auditory and Audiological Research, Nyborg, Denmark. Edited by T. Dau, M.L. Jepsen, T. Poulsen, and J. Christensen-Dalsgaard. ISBN: 87-990013-3-0. (The Danavox Jubilee Foundation, Copenhagen), pp. 175-181.

Schneider, B.A., Li, L., and Daneman, M. (**2007**). "How competing speech interferes with speech comprehension in everyday listening situations," J. Am. Acad. Audiol., **18**, 559-572.

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.