Validation of a spatial speech-in-speech test that takes signal-to-noise ratio (SNR) confounds into account

SØREN LAUGESEN*, FILIP MARCHMAN RØNNE, NIELS SØGAARD JENSEN,
AND MARIA GRUBE SORGENFREI
Eriksholm Research Centre, Oticon A/S, Rørtangvej 20, 3070 Snekkersten, Denmark

A Spatial Fixed-SNR (SFS) speech-in-speech intelligibility test is presented and the reliability and validity of the test is investigated. As part of the validation the SFS test was used to compare a linear hearing-aid setting to a setting with aggressive compression limiting. Two sub-groups of listeners were tested in a fixed-SNR paradigm at –5 and +5 dB SNR, respectively.

INTRODUCTION

Measuring speech-reception threshold (SRT) using adaptive procedures is popular, as testing yields results at the steepest, most sensitive part of the psychometric functions of individual test subjects. However, the signal-to-noise ratio (SNR) at which the SRT is achieved is not kept constant in this test paradigm. Thus, if testing involves the use of hearing-impaired (HI) test subjects, the variation in SRT measures for a single condition can easily span 10 dB. Further, if testing with normal-hearing (NH) test subjects, the SRT will often be a double-digit negative number, which compromises the ecological validity of the result (Pearsons *et al.*, 1977; Smeds *et al.*, 2012). If testing involves hearing aids (HA), extremely low SRTs mean that these devices and the signal-processing algorithms in them may be operating in conditions for which they were not intended.

Another way of testing speech intelligibility is to score %-correct words or sentences at a fixed SNR. However, as test subjects do not perform equally well at equal SNRs, it may be necessary to vary test SNR across subjects in order to obtain results in the informative 20-90% range. As above, this introduces a potential SNR confound. It would be preferable to test all subjects at the same fixed SNR and at the same time have everybody performing around the steepest part of their psychometric functions.

One way to accomplish this is to provide the experimenter with ‘SRT manipulators’, to control the SNR at which testing takes place for the individual listener. Using such manipulators on an individual basis could potentially reduce the spread of SRTs across a group. In an earlier study (Rønne *et al.*, 2013), three suitable manipulators were identified: changing between male and female masker speakers, changing the scoring method from word-correct to sentence-correct, and changing the spatial separation between target and maskers.

This paper presents a spatial speech-in-speech test with means of addressing ecological validity and SNR confounds. This was achieved by selecting four

---

*Corresponding author: slu@eriksholm.com

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.

appropriate test conditions using the three SRT manipulators mentioned above, making it possible to shift the individual listener's SRT towards a common desired target SNR. Further, this paper presents the results of a perceptual validation study.

METHODS

The SFS test basics and setup

The SFS test is a speech-in-speech intelligibility test, using the Danish HINT corpus (Nielsen and Dau, 2011) as target speech. The masker speech signals are recordings of two different either male or female speakers reading from H.C. Andersen's fairytale *The Nightingale*. The masker signals are approximately 2 minutes long and were looped. Masker-speech pauses were cut down to 65 ms. Both the male target and the male and female maskers were spectrally matched to a female reference spectrum (the Dantale 2 spectrum, Wagener *et al.*, 2003).

For the trials targeting 50% words correct the Dantale 2 (Wagener *et al*., 2003) adaptive procedure is used. For trials targeting 50% sentences correct the HINT adaptive procedure is used (Nielsen and Dau, 2011).

The test is set up in an anechoic chamber. The test subjects are seated in an adjustable chair in the centre of the room and it is ensured that the point between the subject's ears is at the same height as and distance from the surrounding loudspeakers, see Fig. 1.

The target speech is played at 70 dB SPL (C) from 0º, and the two masker signals are used in pairs and arranged symmetrically around the listener, at angles ±15º, ±30º, or ±45º. The ±60º or ±90º loudspeakers are used in conditions with target location uncertainty (see below).

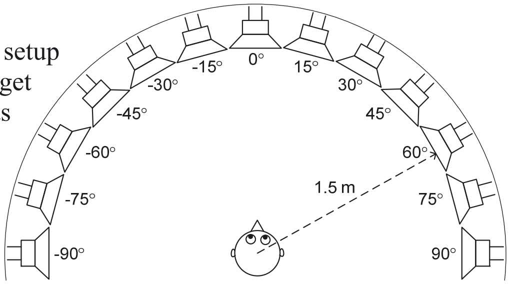


Fig. 1: The loudspeaker setup used for the experiment. Target was presented at 0º, whereas two maskers were presented from different symmetrical configurations.

Test conditions

The post-analysis outcome of the Rønne *et al.* (2013) study was a selection of SRT-manipulator settings resulting in four SFS conditions. In an adaptive-SNR test paradigm these conditions on average lead to successive 2.5-dB shifts of the SRT, as shown in Table 1. When used in a fixed-SNR paradigm, the SFS conditions will allow a group of subjects to be measured at the same target SNR and still be evaluated within the sensitive part of their individual psychometric function. This is true as long as the between-subjects spread in baseline SRT is up to about 10 dB.

SFS condition	Masker gender	Scoring	Masker positions	Expected shift of SRT
15mS	Male	Sentence	±15°	+5 dB
30mS	Male	Sentence	±30°	+2.5 dB
30mW	Male	Word	±30°	0 dB
45fW	Female	Word	±45°	-2.5 dB

Table 1: The four SFS conditions. Condition *30mW* (male maskers at ±30°, Word scoring) is chosen as the baseline.

Target location uncertainty (TGLU)

An option available in the SFS test is to include TGLU. In the SFS test this means presenting the target sentence randomly from three different loudspeaker positions. TGLU was included in the validation study, while data will be reported elsewhere.

Calibration

Calibration of the signals used in the SFS test was done with the test subject absent and a microphone positioned at the centre of the semi-circle in Fig. 1. All SNRs in this paper are referred to this reference condition. Note that the shadow and baffle effect of the head and the pinna changes the SNR at the position of the hearing aid, when the spatial configuration is changed (Rønne *et al.*, 2013), see Table 2.

SFS conditions	$\mathrm{SNR}_{\mathrm{HA}}$-$\mathrm{SNR}_{\mathrm{ref}}$ [dB]
15mS	-0.3
30mS, *30mW*	-0.9
45fW	-1.2

Table 2: Differences between the calibrated $\mathrm{SNR}_{(\mathrm{ref})}$ and the $\mathrm{SNR}_{\mathrm{HA}}$ measured at a BTE hearing aid, averaged across a pool of subjects.

VALIDATION TEST DESIGN

The purpose of the validation test was to validate that the four SFS conditions yielded the expected SNR shifts, and to examine the validity and test-retest reliability of the SFS test.

Test subjects

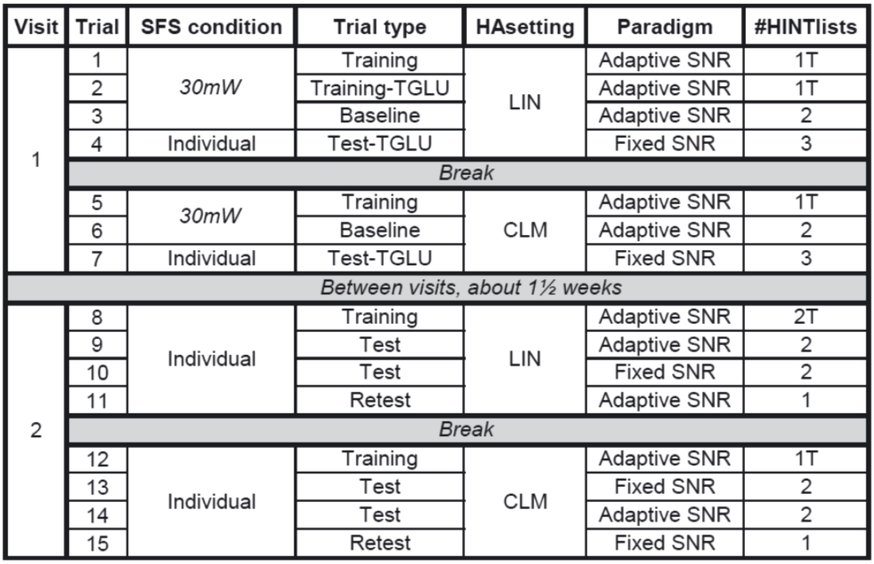
N = 26 hearing-impaired listeners with sensorineural and mixed hearing loss took part. Pure-tone-average (PTA) HTL values across 0.5, 1, 2, and 4 kHz, averaged across ears, ranged from 29 dB to 66 dB, with a mean value of 46 dB. Subjects were listening bilaterally aided through Agil Pro miniRITE hearing aids with closed ‘power domes’. Directionality and noise management were disabled.

Experimental contrast

The experimental contrast used in the validation study was the difference between hearing-aid settings with compression limiting (CLM) and linear processing (LIN). This was selected because Naylor and Johannesson (2009) found a significant change in SNR (ΔSNR) from the input to the output of an aggressive compressive hearing aid. Further, this ΔSNR was shown to depend of the input SNR, such that the ΔSNR was positive for negative input SNRs, and negative for positive input SNRs. No change in SNR was expected with a linear hearing aid. Thus, this was a clear example of an experimental contrast with an SNR confound, where different results would be expected if a test subject was tested in positive or negative SNRs. Given this contrast it was decided to include two target SNRs, one at −5dB SNR and one at +5dB SNR. Test subjects were shifted, by choosing an appropriate SFS condition, towards the target SNR that was closest to their baseline performance. The projected SNR confound should be observable in test performance as a significant interaction between target-SNR group and hearing-aid setting.

Protocol

The test protocol is sketched in Table 3. Hearing-aid setting LIN was tested against hearing-aid setting CLM. Measurements were done in either the adaptive-SNR paradigm or the fixed-SNR paradigm. Further, TGLU was included (only in fixed-



Visit	Trial	SFS condition	Trial type	HAsetting	Paradigm	#HINTlists
1	1	*30mW*	Training	LIN	Adaptive SNR	1T
	2		Training-TGLU		Adaptive SNR	1T
	3		Baseline		Adaptive SNR	2
	4	Individual	Test-TGLU		Fixed SNR	3
	Break					
	5	*30mW*	Training	CLM	Adaptive SNR	1T
	6		Baseline		Adaptive SNR	2
	7	Individual	Test-TGLU		Fixed SNR	3
Between visits, about 1½ weeks						
2	8	Individual	Training	LIN	Adaptive SNR	2T
	9		Test		Adaptive SNR	2
	10		Test		Fixed SNR	2
	11		Retest		Adaptive SNR	1
	Break					
	12	Individual	Training	CLM	Adaptive SNR	1T
	13		Test		Fixed SNR	2
	14		Test		Adaptive SNR	2
	15		Retest		Fixed SNR	1

Table 3: The test protocol. The order of hearing-aid settings, the order of test paradigms in trial pairs (9,10) and (13,14), as well as use of HINT test lists were balanced across listeners.

SNR paradigm). The test protocol included two visits. At the beginning of visit 1, two HINT training lists (40 sentences) were included, followed by the baseline *30mW* adaptive-SNR measurement. Based on the performance of the individual test subject, each subject was 'shifted' to one of the two target SNRs in this design, either +5dB SNR or −5dB SNR. One more measurement was done in the baseline setup to determine a baseline performance difference between the two hearing-aid settings.

Within-visit training effects are small (Rønne *et al.*, 2013) and are assumed to be balanced in the present test design. However, Rønne *et al.* (2013) found a 0.3-dB between-visit training effect (better performance at the second visit), which needs to be addressed here because the baseline and the TGLU measurements always were done at the first visit. Thus, between-visit training was corrected for when relevant.

RESULTS

SFS condition performance

All 26 test subjects were measured twice in the baseline condition (*30mW*, adaptive-SNR paradigm), one with hearing-aid setting LIN and one with CLM. Later in the protocol all subjects were also measured in their individually selected SFS condition (trials 3, 6, 9 and 14). In order to assess the effectiveness of the SFS test conditions, Fig. 2 shows the magnitude of the SRT shifts (two from each subject). Some subjects did in their baseline measurement perform close to one of the two target SRTs (−5 or +5 dB SNR), and were thus not shifted. The data points from these subjects are depicted at *30mW*.

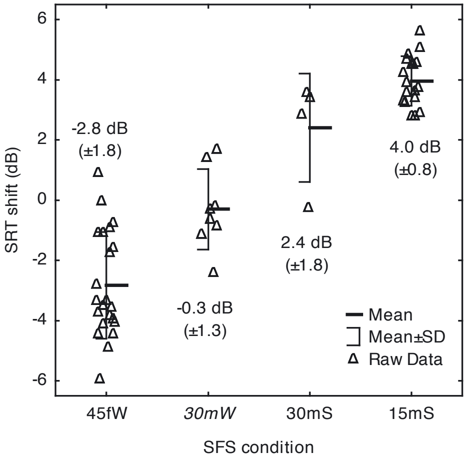


Fig. 2: SRT shifts between the adaptive-SNR baseline conditions (trials 3 and 6 in protocol) and the test subject's individually selected adaptive-SNR SFS condition (trials 9 and 14). Note the very few individual data points in the 30mS condition. Two data points were obtained from each test subject (9-3, 14-6), thus 52 data points are present in the figure.

Test-rest reliability

Test-rest data were derived from the first list (20 sentences) of trials 9 and 13, to allow direct comparison with trials 11 and 15. Note that each subject thus contributes one set of test-retest data points in the adaptive-SNR paradigm and one set in the fixed-SNR paradigm. The protocol was balanced such that half of the data points for each paradigm were measured with each hearing-aid setting.

First, the variance of the difference measure (trial pair 9-11 for the adaptive-SNR paradigm) is found as

$$V_{\Delta\mathrm{adaptive-SRT}} = \frac{1}{N}\sum_{n=1}^{N}(SRT_{\mathrm{test}} - SRT_{\mathrm{retest}})^2 \,. \qquad \text{(Eq. 1)}$$

The test-retest standard deviation (SD) of a single measurement is then

$$\mathrm{SD}_{\mathrm{adaptive-SRT}} = \sqrt{\tfrac{1}{2}V_{\Delta\mathrm{adaptive-SRT}}} = 0.95\ \mathrm{dB}. \qquad \text{(Eq. 2)}$$

Similarly, for the fixed-SNR paradigm:

$$V_{\Delta\mathrm{fixed-SNR}} = \frac{1}{N}\sum_{n=1}^{N}(\%correct_{\mathrm{test}} - \%correct_{\mathrm{retest}})^2 \qquad \text{(Eq. 3)}$$

$$\mathrm{SD}_{\mathrm{fixed-SNR}} = \sqrt{\tfrac{1}{2}V_{\Delta\mathrm{fixed-SNR}}} = 8\ \%. \qquad \text{(Eq. 4)}$$

Validity

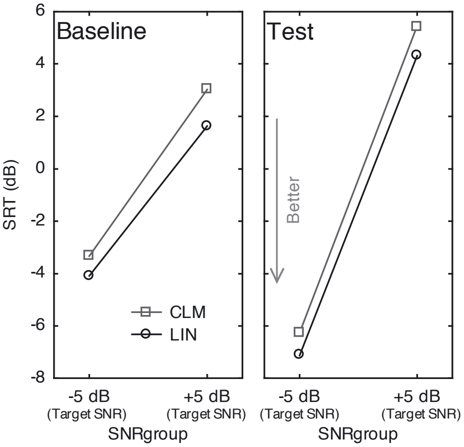
Figure 3 (left panel) shows the baseline performance with the two hearing-aid settings in the two subgroups of good performers (labelled −5 dB, typically small hearing losses) and poor performers (labelled +5 dB, typically larger hearing losses). Figure 3 (right panel) shows the performance of the same subjects in their individually selected SFS condition. The Test SRTs are forced further apart than the Baseline SRTs, and the Test SRTs are closer to the target SNRs in the +5-dB group. This indicates that the SFS conditions are working as expected. Two mixed-model ANOVAs (one for each panel) indicate that hearing-aid setting was significant for both Baseline ($p = 0.0008$) and Test ($p = 0.004$), whereas the expected interactions between SNR group and hearing-aid setting were not significant.

Figure 4 shows the average performance for the fixed-SNR paradigm. Hearing-aid setting was significant ($p = 0.00009$) and the interaction between hearing-aid setting and SNR group was significant ($p = 0.02$).

DISCUSSION

Test-retest reliability

The test-retest within-subject SD of the adaptive-SNR paradigm SFS test was determined to be 0.95 dB (Eq. 2). This is comparable to the test-retest SD of the original HINT material that was found to be 0.92 dB for hearing-impaired listeners (Nielsen and Dau, 2011). Thus, it seems that the inclusion of a spatial setup, speech maskers, and different SFS conditions, does not increase the SD of the test. For the fixed-SNR paradigm the test-retest SD was determined to be 8% (Eq. 4). For this measure no relevant literature comparison exists. However, the gradient of the



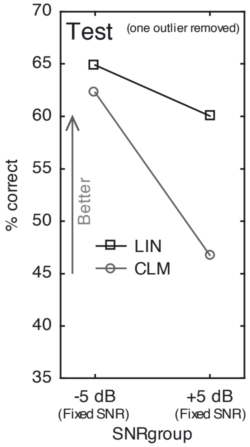


Fig. 3: Left panel shows the average performance of the test subjects in the baseline condition for each the two hearing-aid settings (trials 3 and 6). The subjects were based on their performance divided into the groups, labelled −5 or +5 dB target SNR. Right panel shows the average performance of the same subjects when measured in their individually selected SFS conditions (trials 9 and 14).

Fig. 4: Average performance across subjects tested in the fixed-SNR paradigm (trial 10 and 13). All individual %-correct scores were in the 10-84% range.

psychometric curve at the 50% correct point is 13.7%/dB for the SFS test, and it could thus be speculated that the test-retest SD of the fixed SNR paradigm should be somewhere slightly below 13.7*0.95 = 13%. That it is in fact 8% indicates that the test-retest reliability potentially is better for the fixed-SNR paradigm compared to adaptive-SNR.

The experimental contrast

According to Naylor and Johannesson (2009) hearing-aid compression affects the long-term SNR, such that the SNR at the output of the hearing aid is improved by compression in negative input SNRs and is made worse in positive SNRs. This change in long-term SNR from input to output was denoted ΔSNR. This study replicated the setup of Naylor and Johannesson (2009) and did actual measurements in a test-box to determine the magnitude of the ΔSNR for each individually-fitted hearing aid programmed to be first in LIN and then in the CLM setting. For all subjects the ΔSNR was measured to be approximately 2 dB in the expected direction, positive or negative.

The adaptive-SNR trials in this study showed a constant influence of CLM of about −1 dB independent of input SNR, whereas the fixed SNR trials showed a small but significant interaction between SNR group and hearing-aid setting. Neither of the two methods showed the expected ±2 dB ΔSNR swing and strong dependence on SNR group. Thus a major question mark has to be raised regarding the perceptual relevance of the Naylor and Johannesson (2009) output-SNR measure. Also, the results from this study contradict the perceptual correlations found between speech intelligibility performance and ΔSNR by Naylor *et al.* (2008).

It is also interesting that the two test paradigms yield different results regarding the interaction between hearing-aid setting and SNR group in Figs. 3 (right) and 4. One explanation could be that all test subjects in the fixed-SNR paradigm are tested at the same SNR, whereas subjects in the adaptive-SNR paradigm are tested at a range of SNRs around the target SNR. It can be speculated that the latter approach has made the results more variable, and thus made it harder for a contrast to be visible.

CONCLUSION

A Spatial Fixed-SNR (SFS) speech intelligibility test was designed and validated. The unique asset of the SFS test is the way individual test subjects can be evaluated in different conditions such that the SNR at which they are evaluated is the same. This study found that the SFS test conditions provide SNR shifts of the expected magnitude, that reliability is on par with the standard HINT, and that the test is able to detect relevant experimental differences with high statistical significance.

REFERENCES

Naylor, G., Rønne, F.M., and Johannesson, R.B. (**2008**). **“**Perceptual correlates of the long-term SNR change caused by fast-acting compression,” International Hearing Aid Research Conference (IHCON), Lake Tahoe, CA, USA (poster).

Naylor, G., and Johannesson, R.B. (**2009**). “Long-term signal-to-noise ratio at the input and output of amplitude-compression systems,” J. Am. Acad. Audiol., **20**, 161-171.

Nielsen, J.B., and Dau, T. (**2011**). “The Danish hearing in noise test,” Int. J. Aud., **50**, 202-208.

Pearsons, K.S., Bernett, R.L., and Fidell, S. (**1977**). “Speech levels in various noise environments,” Bolt, Beranek and Newman Inc., Canoga Park, California.

Rønne, F.M., Laugesen, S., Jensen, N.S., Hietkamp, R.K., and Pedersen, J.H. (**2013**), “Magnitude of speech-reception-threshold manipulators for a spatial speech-in-speech test that takes signal-to-noise ratio confounds and ecological validity into account,” Proc. Meet. Acoust., **19**, International Congress on Acoustics, Montréal, Canada, pp. 050069.

Smeds, K., Wolters, F., and Rung, M. (**2012**). “Realistic signal-to-noise ratios,” International Hearing Aid Research Conference (IHCON), Lake Tahoe, CA, USA (poster).

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