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

Filip M. Rønne*, Søren Laugesen, Niels S. Jensen, Rensje K. Hietkamp and Julie H. Pedersen

*Corresponding author's address: Eriksholm Research Centre, Rørtangvej 20, Snekkersten, 3070, Denmark, Denmark, fmr@eriksholm.com

Measuring Speech-Reception Threshold (SRT) using adaptive procedures is popular, as testing yield results with desirable statistical properties. However, SRT measures have drawbacks related to the unbounded nature of the Signal-to-Noise Ratio (SNR) at which the SRT is achieved. Often the SRT will be a double-digit negative number, which compromises the ecological validity of the result. If testing involves hearing aids, it means that these devices and the signal-processing algorithms in them may be operating in conditions for which they were not intended. Further, the commonly observed large spread in SRT (both between- and within-group) has the possibility to cause SNR confounds that may lead to faulty conclusions. One way to address these issues is to provide the experimenter with SRT manipulators, to control the SNR at which testing takes place for the individual listener. The present work aims at developing a spatial speech-in-speech test with a selection of SRT manipulators for the experimenter to choose from. The manipulators investigated in this study are: the spatial separation between target and maskers, the number of spatially separated maskers, changing the masker gender, and scoring in words versus sentences. The magnitudes of the SRT manipulators were investigated using 20 hearing-aid users as listeners.
INTRODUCTION

Measuring Speech-Reception Threshold (SRT) using adaptive procedures is popular, as testing can yield results obtained at the steepest and thus 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 in this test paradigm not kept constant. Thus, if testing involves the use of hearing impaired (HI) test subjects, the SRT will often be a double-digit negative number, which compromises the ecological validity of the result (Pearson et al., 1977, Smeds et al., 2012). If testing involves normal hearing (NH) test subjects, the SRT will often be a double-digit negative number, which compromises the ecological validity of the result (Pearson 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. Additionally, the observed large spread in SRT (both between- and within-group) has the possibility to cause SNR confounds that may lead to faulty conclusions. This could for instance be the case when testing subjects wearing HAs, as the HA processing is likely to be working differently in different SNRs, and thus testing the benefit of a new HA algorithm, might be confounded with the SNR at which the individual SRTs are obtained.

One way to address these issues 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. As an example, this could be obtained by using word-scoring (adaptation target = 50% correct) for the poorest performers, and sentence scoring (again, adaptation target = 50% correct) for the best performers. Given the hypothesis that sentence scoring will result in a higher SRT than word scoring, the SNR, at which the best performers would be tested, would thus be made higher and the within-group SRT deviations would be reduced.

The long-term goal of this study is to devise a spatial speech-in-speech test with means of addressing ecological validity and SNR confounds. This will be achieved by selecting appropriate test conditions so as to shift the individual listener’s SRT towards a common desired SNR. This particular study examined four candidate SRT manipulators:

1. Change the spatial separation between target and maskers.
2. Change the number of maskers among 2, 4, and 6.
3. Change the adaptation target from 50% words to 50% sentences correct.
4. Change the gender of the masking speech from female to male (different or same gender as male target talker).

METHOD

N = 20 hearing-impaired listeners with symmetrical sensori-neural hearing loss took part. PTA (Pure Tone Average hearing threshold levels at 0.5, 1, 2, and 4 kHz) ranged from 29 dB to 69 dB, with a mean of 51 dB and a standard deviation of 11 dB. Subjects were listening bilaterally aided using their own state-of-the HAs. 7 in-the-canal (ITC) HAs and 13 behind-the-ear (BTE) HAs were used, all of which had directionality, noise management, program change, and volume control disabled during testing. This was done to ensure that all HAs were in the same well-defined state during testing.

Testing was carried out in an anechoic chamber. The test subjects were seated in a movable chair in the center of the room and it was ensured that the ears of the subjects were at the same height as and distance from the surrounding loudspeakers. At a distance of 1.5 m from the subject was a ring of 13 loudspeakers equally spaced and positioned at angles, ranging from -90º to +90º, relative to the subject (see figure 1).
The test subjects were instructed to vocally repeat the 5-word sentence or any single words they heard. An experimenter then scored the number of correct words (0 to 5). In each test condition the masker level was varied adaptively across a list of 20 sentences. For trials targeting 50% words correct the Dantale 2 (Wagener et al., 2003) adaptive procedure was used, where each sentence was presented only once and the SNR of the first sentence was based on a qualified guess. Differently, for trials targeting 50% sentences correct the HINT adaptive procedure was used (Nielsen and Dau, 2011), where the first sentence was presented at a very low SNR and then repeated at increasing SNRs until full intelligibility was reached. 20 SRTs were determined for each listener, divided across two visits. Both visits started with two training lists.

Target speech was the Danish HINT corpus (Nielsen and Dau, 2011), played at 70 dB SPL (C) from 0°. The Danish HINT corpus consists solely of 5 words sentences as opposed to the American HINT, allowing easy switches between word scoring and sentence scoring. The masker speech signals were recordings of speakers reading from a fairytale: two females and two males. The two female (or two male) masker signals were used in pairs arranged symmetrically around the listener, at angles ±15°, ±30°, ±45° or ±75°. Additionally, two configurations using 4 loudspeakers (±15°±45° and ±30°±90°) were tested and one configuration using 6 loudspeakers (±15°±45°±75°). Masker-speech pauses were cut down to 65 ms and for the majority of conditions. 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). However, 3 conditions (±15°, ±45°, and ±30°±90°) were included where the target and the maskers (male only) were spectrally shaped to the unfiltered male HINT targets, in order to test whether the overall spectral shape affected the SRTs. A separate analysis of the individual differences in SRT between the HINT-shaped and Dantale 2-shaped conditions revealed no effect of spectral shaping (mean effect 0.02 dB, T(60) = 0.105, p = 0.92), and these results are therefore not considered further.

**Post Processing**

SRTs were found by a maximum likelihood approach in the following way. For all trials where a 50% words-correct criterion was used, psychometric functions were first estimated with both mid-point and slope as free parameters (Brand and Kollmeier, 2002), and then re-estimated using the median slope as a fixed parameter for all psychometric functions. The SRT was taken as the mid-point, corresponding to 50% correct words. For the trials where a 50% sentence-correct criterion was used, psychometric functions were also estimated based on the correct number of words for each sentence using the median slope estimated from the 50% words-correct trials above. In this approach, a key issue is to determine the %-correct value on these word-scoring psychometric functions that correspond to 50% correct on the sentence-scoring psychometric function (Boothroyd & Nittrouer, 1988), see figure 2. This issue is briefly mentioned by Nielsen & Dau (2007) who suggest a value of 70% for normal-hearing listeners and the CLUE material (the predecessor of the Danish HINT). For the present study, this %-correct value, denoted ψ, was estimated directly, based on the assumption that the sentence-SRT estimator using raw sentence scores is a valid (un-biased) estimator, albeit less reliable (higher test-retest variance), than the sentence-SRT estimator that uses the full word scores. Thus, ψ is determined by solving,
where \( M \) is the total number of sentence-scoring trials in the data set, \( \text{SRT}_{\text{sentence} m}^{\text{sentence-based}, 50\%} \) is the sentence-SRT for the \( m \)’th trial determined from the raw sentence scores (at the 50% mid-point), and \( \text{SRT}_{\text{sentence} m}^{\text{word-based}, \psi} \) is the sentence-SRT for the \( m \)’th trial determined from the full word scores at the \( \psi \)% point. The formulation in eq. (1) yields the word-based estimator of sentence-SRTs which has the same mean value as the sentence-based estimator. For the present data, a value of \( \psi = 77\% \) was found based on \( M = 100 \) trials.

\[
\sum_{m=1}^{M} \left( \text{SRT}_{\text{sentence} m}^{\text{sentence-based}, 50\%} - \text{SRT}_{\text{word-based}, \psi}^{\text{sentence} m} \right) = 0
\]

FIGURE 2. Black triangles represent sentence-scored measurements. The solid black line represents a fitted psychometric function and the dashed line illustrates the SNR (0 dB) which is expected to result in 50% correct scores (thus the SRT). Red triangles and the solid line represent hence word-scored measurements and the fitted psychometric function. The red dashed line illustrates that the SNR which corresponds to a 77%-correct word score is comparable (but not the same) to the SNR corresponding to a 50% sentence score.

In the experimental setup, the target and the maskers were calibrated according to a reference microphone positioned in the center of the setup, with the test subject absent. However, in the actual test the head of the test subject created baffle and shadowing effects. The SNR observed at either HA was thus different from the SNR measured in the reference point (SNR\(_{\text{ref}}\)), as the target and maskers were not co-located and thus were subject to different shadowing effects. In a previous study (Neher et al., 2009; Jensen et al, 2013) sound levels were recorded through ITC and BTE HAs worn by 18 test subjects in the same setup. The stimulus was a pink noise signal at 50 dB SPL presented in turn from each of the 13 loudspeakers. Using these data it was possible to derive the SNR at the HA microphone location (SNR\(_{HA}\)), when worn by an average test subject. Table 1 shows the articulation-index (AI) weighted SNR compensation factors that were added to the SNR\(_{\text{ref}}\)’s to turn them into SNR\(_{HA}\)’s. The between-subject variation in these compensation factors was small; the common standard deviation was 0.5 dB.

<table>
<thead>
<tr>
<th>Masker configuration</th>
<th>SNR(<em>{HA})-SNR(</em>{\text{ref}}) [dB], BTE</th>
<th>SNR(<em>{HA})-SNR(</em>{\text{ref}}) [dB], ITC</th>
</tr>
</thead>
<tbody>
<tr>
<td>±15°</td>
<td>-0.3</td>
<td>-0.3</td>
</tr>
<tr>
<td>±30°</td>
<td>-0.9</td>
<td>-0.8</td>
</tr>
<tr>
<td>±45°</td>
<td>-1.2</td>
<td>-1.1</td>
</tr>
<tr>
<td>±75°</td>
<td>-1.9</td>
<td>-1.2</td>
</tr>
<tr>
<td>±30°, ±90°</td>
<td>-1.3</td>
<td>-1.1</td>
</tr>
<tr>
<td>±15°, ±45°</td>
<td>-0.8</td>
<td>-0.8</td>
</tr>
<tr>
<td>±15°, ±45°, ±75°</td>
<td>-1.2</td>
<td>-1.0</td>
</tr>
</tbody>
</table>
RESULTS

Measured SRTs are shown in Figure 3 in terms of the SNR at the HA (SNR_{HA}), thus taking into account the difference in baffle/shadow effects of the head for different angles of masker incidence, as well as the number of maskers. Each line represents data across conditions for one subject. The first seven conditions were recorded to investigate the SRT change observed when changing the spatial separation between target and maskers (manipulator 1), and to investigate the SRT change observed when adding extra maskers to the setup (manipulator 2). Further, these seven conditions acted as a baseline performance for comparing them to conditions 8 to 12 in order to investigate the effect of changing the masker gender from male to female (manipulator 4) and to conditions 13 to 17 to investigate the effect of scoring 50% correct sentences rather than 50% correct words (manipulator 3).

Data were analyzed with a mixed-model main-effects ANOVA, see table 2. The results show significant effects of the main variables (Word/sentence, Masker gender, and Masker configuration). The significant effect of Listener corroborates the considerable spread in individual SRTs, see figure 3. The significant Between-visit and Within-visit training effects are discussed below and were corrected for in the data presented in the following, by adjusting each SRT to the value that would have been obtained if it had been measured as the first condition at the first visit.

FIGURE 3. Measured SRTs given in dB SNR_{HA}, where each line represents data across conditions for one subject. The name tag for each condition describes the number and the gender of the maskers, followed by the spatial configurations of the maskers, followed by the adaptation target (50% words or 50% sentences).
TABLE 2. Summary results of the mixed-model main effects ANOVA of the measured data shown in figure 3.

<table>
<thead>
<tr>
<th>Effect</th>
<th>F-test</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word/sentence</td>
<td>F(1,370)=341</td>
<td>&lt; 0.00001</td>
</tr>
<tr>
<td>Masker gender</td>
<td>F(1,370)=84.7</td>
<td>&lt; 0.00001</td>
</tr>
<tr>
<td>Masker configuration</td>
<td>F(6,370)=136</td>
<td>&lt; 0.00001</td>
</tr>
<tr>
<td>Listener (rand)</td>
<td>F(19,370)=68</td>
<td>&lt; 0.00001</td>
</tr>
<tr>
<td>Between-visit training</td>
<td>F(1,370)=7.8</td>
<td>0.005</td>
</tr>
<tr>
<td>Within-visit training</td>
<td>F(1,370)=6.2</td>
<td>0.0004</td>
</tr>
</tbody>
</table>

Examination of SRT Manipulators

Changing Target-Masker Spatial Separation and Number of Maskers

The variation among masker configurations in terms of spatial separation and number of maskers is illustrated in figure 4. The results are shown as shifts in SRT away from an (arbitrarily) chosen reference condition: 2male_+/-30º. Thus each circle represents a single SRT obtained for one subject relative to the SRT obtained by the same subject in the reference condition. Only conditions with male maskers and 50%-word adaptation target are considered (conditions 1-7 in figure 3, where condition 2 is the reference). From left to right in figure 4, the mean SRT shifts and standard deviations are 2.8±1.4 dB (15º); 0±0dB (30º); -1.7±1.3 dB (45º); -2.4±1.6 dB (75º); 0.1±1.4 dB (30º,90º); 2.0±1.3 dB (15º,45º); and 1.1±1.1 dB (15º,45º,75º).

FIGURE 4. Individual SRT differences for each masker configuration relative to the 2male_+/-30º reference condition.
Changing Adaptation Target

The effect of changing adaptation target from 50% correct words to 50% correct sentences is illustrated in figure 5. Each circle (100 in total) represents paired sentence/word differences in SRT for each individual listener (\(N = 20\)) and masker configuration (\(N = 5\)), such that conditions 1 and 13 are compared, conditions 2 and 14, etc. The mean SRT shift and standard deviation in figure 5 are 2.6±1.4 dB.

![FIGURE 5. Individual sentence/word SRT differences across listeners and masker configurations.](image)

Changing Masker-Talker Gender

The SRT differences associated with changing masker gender was subjected to a separate ANOVA, which showed significant effects of Listener and Masker configuration. (Similar analyses were conducted for the other SRT manipulators, where no significant effects were found). The significant effect of Listener on the masker-gender induced SRT differences suggests a systematic variation in the magnitude of the gender effect among listeners, which will be investigated later in a correlational analysis. The significant effect of Masker configuration is illustrated in figure 6, which shows the configuration-specific paired differences in SRT, obtained with male and female maskers for each listener. Each circle thus represents an SRT shift (\(SRT_{\text{male}} - SRT_{\text{female}}\)) for a single subject at the given masker configuration, when the masker-talk gender is changed from male (similar to target talker) to female. From left to right in figure 6, the mean SRT shifts and standard deviations are 2.5±1.2 dB (15º); 1.6±1.2 dB (30º); 1.1±1.3 dB (45º); 0.5±1.3 dB (30º,90º); and 1.0±1.3 dB (15º,45º).

![FIGURE 6. Individual Masker gender male/female SRT differences for each masker configuration.](image)
DISCUSSION

SRT Manipulators

Changing Target-Masker Spatial Separation and Number of Maskers

The variation in target-masker spatial separation, with the ±30º condition as reference, showed SRT shifts of +2.8 dB (15º), -1.7 dB (45º), and -2.4 dB (75º). While figure 4 indicates some individual variation, the standard deviations are small compared to the mean: Cohen’s effect size is $d = 2.0, 1.3$ and $1.5$ for the ±15º, ±45º and ±75º respectively, which is well above the 0.8 value required for a ‘large effect’ (Cohen, 2009). Moreover, with the reported 0.9-dB test-retest standard deviation of the Danish HINT (Nielsen & Dau, 2011) the minimal expected standard deviation of any SRT difference measure is $0.9 \cdot \sqrt{2} = 1.3$ dB. Thus, the standard deviations of all SRT shifts in figure 4 are close to the expected minimum. All together, this makes changing target-masker spatial separation an excellent candidate for the conjectured test.

Adding additional maskers to the ±15º and ±30º conditions changed SRT only marginally. The expectation was that increasing the number of maskers would increase SRT (Brungart and Simpson, 2007). This could be due to an increase in informational masking and a decreased possibility of dip listening (as the envelope of the combined masker would be less fluctuating). However, the data shows that hearing-impaired test subjects in these conditions are not penalized by increasing the number of maskers, when the trivial change in energetic masking has been accounted for. Thus, changing the number of maskers is not a recommended SRT manipulator; at least for the present hearing-impaired listeners.

Changing Adaptation Target

The effect of changing the adaptation target from 50% words to 50% sentences was 2.6 dB on average. This is less than the 5.1 dB found with the Dantale 2 corpus (Danish Matrix test) in a previous study (Laugesen et al, 2012). However, this was expected due to the HINT sentences’ greater redundancy. The standard deviation is 1.4 dB, which again is close to the 1.3 dB expected minimum standard deviation of the Danish HINT and yields a Cohen’s $d$ of 1.9. This makes changing adaptation target an excellent candidate for the conjectured test.

Changing Masker-Talker Gender

The effect of masker gender depended on Masker configuration. In the ±30º condition, the magnitude was 1.6 dB which is very close to the 1.7 dB found with the Dantale 2 corpus (Laugesen et al, 2012). As above, the standard deviations are close to the 1.3 dB lower limit value. The results in figure 6 indicate that the benefit of having an opposite-gender masker is greatest when the spatial cues are least powerful, that is, where the target-masker separation is smallest. At a glance, it is surprising to observe a smaller masker-gender effect in the ±15º±45º condition with four maskers. However, in the four-masker conditions, the combined masker signal is dominated by the ‘wide’ maskers due to the greater head-baffle effect at these angles of incidence. Using Cohen’s $d = 0.8$ as a criterion, the Masker gender SRT manipulator is only relevant for the ±15º and ±30º two-masker configurations.

Training Effects

The within-visit training effect, that is, the average SRT improvement from one sentence to the next (excluding the two training lists in the beginning of each visit), was estimated to be 0.0019 dB/sentence (corresponding to a 0.38 dB improvement from the first sentence to the last of the 200 sentences used at one visit). This agrees well with the within-visit training-effect of 0.0025 dB/sentence reported by Nielsen and Dau (2011). In a previous study (Laugesen et al., 2012), where the Dantale 2 corpus was used in a setup identical to the one used in this study, a larger within-visit training effect of 0.0051 dB/sentence was reported.

The between-visit training effect was estimated to be 0.3 dB, which is comparable to the 0.36 dB reported by Nielsen and Dau (2011). The Dantale 2-based Laugesen et al. (2012) study reported a higher between-visit training effect of 0.9 dB. Thus, this study confirms that the HINT corpus shows smaller training effects than the Dantale 2 corpus. Adding this to the higher ecological validity of the open set HINT corpus, makes the HINT material a good choice for spatial speech-in-speech testing. Note that the long-term (three or more visits) between-visit training effect when HINT sentences are re-used has not been investigated.
CONCLUSION

Three useful SRT manipulators have been identified that will allow an experimenter to shift the SRT of an individual listener over an 8-dB range. This holds considerable promise regarding the development of a speech test that includes means of addressing SNR confounds and some control of the ecological validity of the SNR at which testing takes place.

REFERENCES


Jensen NS, Neher T, Laugesen S, Johannesson RB, Kragelund L. Laboratory and field study of the benefits of pinna cue-preserving hearing aids. Submitted to Trends in Amplification.


