Towards a spatial speech-in-speech test that takes SNR confounds and ecological validity into account

Authors

Søren Laugesen
Niels Søgaard Jensen
Julie Hefting Pedersen
Renskje K. Hietkamp

Eriksholm Research Centre
Rørtangvej 20
DK-3070 Snekkersten
www.eriksholm.com

Scan to access poster pdf

Contact

Søren Laugesen
slu@eriksholm.com

Three ways of shifting the speech reception threshold (SRT) for five-word sentences were investigated: changing the scoring criterion from 50% words correct to 50% whole sentences correct, changing the masker talker from male to female, and changing among three different same-sex masker talkers.

Background

Adaptive SRT procedures are popular for good reasons, but they have drawbacks related to the unbounded nature of the Target-to-Masker-Ratio (TMR) at which criterion performance (the SRT) is achieved [1].

(Lack of) ecological validity

Often the SRT is much lower than the TMR found in realistic listening conditions [2]. If the test involves aided listening, the hearing aid may therefore be subjected to conditions for which it was never intended. This has the potential to cause misleading results.

SNR confounds

Bernstein and co-workers [3,4,5] suggest that the difference in mean SRT between groups of normally hearing vs. hearing-impaired listeners confound the conclusions from studies of fluctuating masker benefit. Similar problems may occur when testing hearing aids with hearing-impaired listeners who often show a wide spread in SRT. Thus, the hearing aids under test will be subjected to very different TMRs among listeners. This may affect hearing-aid performance and can potentially confound the test results [6].

Aim of the study

The long-term goal of this work is to devise a spatial speech-in-speech test, which includes 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 TMR.

This particular study examines three candidate 'SRT manipulators':
1. Change from scoring individual words to scoring whole sentences.
2. Change the masker talker from male to female (which is the same sex as the target talker).
3. Select among different masker talkers (of the same sex).

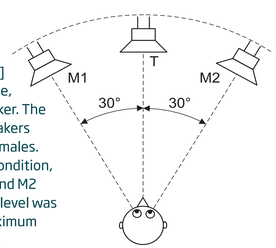
Research questions

What are the effects on measured SRT of each of the proposed SRT manipulators above, in terms of magnitude and consistency?

As a side issue, the effect of spectrally shaping the masker talkers to the target — as opposed to leaving them un-shaped — was investigated for the male maskers.

Method ...

Target (T) speech was the Dantale 2 sentences [7] ("John had three yellow boxes") spoken by a female, played at 70 dB SPL (C) from the centre loudspeaker. The masker speech signals were six recordings of speakers reading from a fairytale: three females and three males. Speech pauses were cut down to 65 ms. In each condition, the same masker signal was played from the M1 and M2 loudspeakers (with a time offset). Masker speech level was varied adaptively. SRTs were determined by a maximum likelihood approach, based on 30 sentences each.



... and material

N = 17 hearing-impaired listeners with sensorineural hearing loss took part. PTA (Pure Tone Average HTL values across 0.5, 1, 2, and 4 kHz) ranged from 26 dB to 63 dB, with a mean of 48 dB and a standard deviation of 10 dB. Subjects were listening binaurally aided using their own hearing aids, which had directionality, noise management etc. disabled during testing.

Table 1. Overview of experimental conditions.

Masker talker	Spectral shaping	Scoring method
Female 1, 2, 3	Shaped as target	Word, *N* = 17, 17, 17; Sentence *N* = 6, 6, 5
Male 1, 2, 3	Shaped as target	Word, *N* = 17, 17, 17; Sentence *N* = 6, 6, 5
Male 1, 2, 3	Un-shaped	Word, *N* = 17, 17, 17; Sentence *N* = 6, 6, 5

Protocol

The listening conditions are outlined in Table 1. Each listener completed 12 conditions, balanced across two visits separated by about one week. Each visit started with a training run.

Results

Data were analysed with a mixed-model ANOVA, with Listener as a random variable. The results in Table 2 show significant effects of the main variables (Scoring method, Masker talker sex, and Spectral shaping). The significant effect of Listener corroborates the considerable spread in individual SRTs, see Figure 1. The significant effects of Visit and Sentences accumulated (counting the number of sentences presented before each SRT determination, at each visit) indicate training effects both between and within visits. The significant Listener*Visit interaction suggests that the between-visits training effect varied among the listeners. No other 2nd-order interactions were significant. The within-in session training effect was corrected for before any further analysis.

Table 2. ANOVA results, whole data set.

Effect	F-test	*p*-value
Scoring method	F(1,166)=562	< 0.00001
Masker talker sex	F(1,166)=78.2	< 0.00001
Spectral shaping	F(1,166)=44.9	< 0.00001
Listener (rand)	F(16,15.6)=55	< 0.00001
Visit	F(1,15.7)=17.4	0.0007
Sentences accumulated	F(1,166)=13.0	0.0004
Listener*Visit	F(16,166)=1.9	0.025

Figure 1. Mean SRT for each listener.

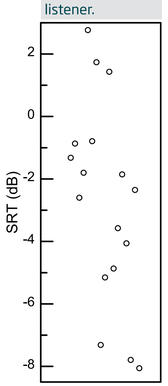


Figure 2. Individual sentence/word scoring differences.

Figure 3. Individual masker female/male differences.

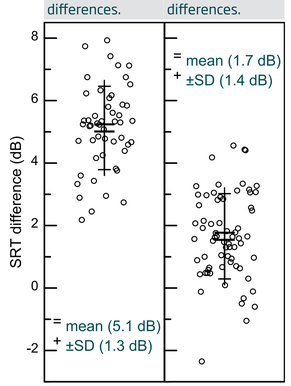
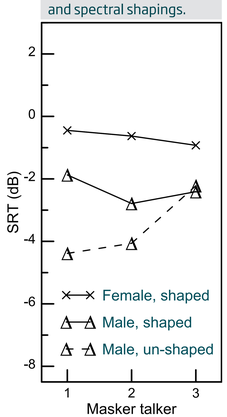


Figure 4. Mean SRTs for individual masker talkers and spectral shapings.



Main findings

THE EFFECT OF SCORING METHOD is illustrated in Figure 2, which shows the paired differences in SRT obtained with sentence and word scoring for each listener, masker talker, and spectral shaping.

THE EFFECT OF MASKER TALKER SEX is seen in Figure 3, which shows the differences in SRT for each listener and scoring method, calculated for pairs of female and male maskers in the shaped-as-target conditions. Pairing was F1-M1, F2-M2, F3-M3, to ensure that paired SRTs were measured at the same visit.

THE VARIATION AMONG INDIVIDUAL MASKER TALKERS is illustrated in Figure 4. The results are least-squares mean SRTs across listeners and scoring methods from three separate ANOVAs. The effect of masker talker was not statistically significant for the female maskers ($p = 0.19$), while there were significant effects for the male maskers, both for the shaped ($p = 0.03$) and un-shaped conditions ($p < 0.00001$).

Discussion

Magnitude and consistency of SRT manipulators

Changing from scoring words to whole sentences shifted the SRT by 5.1 dB on average, which agrees well with literature data [8]. Although Figure 2 reveals some individual variation, the standard deviation of 1.3 dB is small compared to the mean: Cohen's effect size [10] is $d = 5.1/1.3 = 3.9$, which is much larger than the 0.8 value required for a 'large effect'.

Masker talker sex had an average effect of 1.7 dB, which is surprisingly small compared to the about 5 dB found by Helfer & Freyman [9]. However, their results were found with young normal-hearing listeners and older listeners with hearing status ranging from normal to moderate high-frequency hearing loss. Recent research [11] has shown that hearing impairment dramatically reduces the ability to exploit temporal fine structure (TFS) such as voice pitch, which may explain why our hearing-impaired listeners were less able to take advantage of the change in masker talker sex from female (same as target) to male. With a standard deviation of 1.4 dB Cohen's effect size is $d = 1.7/1.4 = 1.2$, which is still a large effect. On the other hand, the 1.7 dB magnitude is coming close to the expected 1-dB test-retest standard deviation of a speech-in-speech test [12].

The magnitude of differences among individual masker talkers were small compared to the expected test-retest standard deviation, when spectral differences between target and background were removed. Therefore this effect is not given further concern.

Training effects

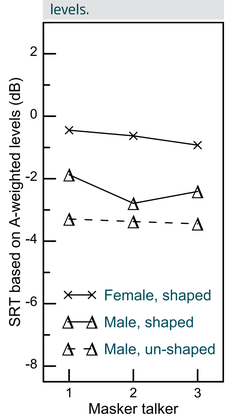
The within-visit training effect was estimated to be 0.0051 dB/sentence. This is in line with data from the literature [7], which suggests about 0.01 dB/sentence. In contrast, the observed between-visit performance improvement, which was estimated to be 0.9 dB, was a surprise. Because this effect might have been the result of an unfortunate balancing of conditions, an estimate was also obtained from the training runs, which were exactly the same at the two visits. Here the improvement was 1.6 dB, confirming the existence of a between-visit training effect. One possible explanation is that this effect is the result of 'procedural learning' [13], which has been found to be particularly likely for taxing listening tasks such as the present one.

Spectral shaping

The results in Figure 4 shows that removing spectral differences among the (male) masker talkers reduces the measured differences in SRT.

Using masker signals that are spectrally matched to the target has the advantage that TMR (and SRT) can be faithfully represented by one number, since TMR is constant across frequency. This is not the case if the masker spectrum is not matched to the target. The potential consequence of this is illustrated in Figure 5, which replicates Figure 4 but uses A-weighted levels to calculate the SRTs instead of the C-weighting used for Figure 4. As seen in Figure 5, this has no effect for the shaped conditions, but entirely changes the relation among the three un-shaped conditions.

Figure 5. As Figure 4 but based on A-weighted levels.



Conclusions

- Change of scoring method provided a mean SRT shift of 5.1 dB, with good consistency. This method recommends itself for use in a future test protocol.
- Change of masker talker sex provided a mean SRT shift of 1.7 dB, with reasonable consistency. However, the magnitude of the effect is too small to be of real value for the present purpose.
- Selecting among different same-sex maskers provided too small SRT shifts to be of value.
- An example was given of how measured SRTs can be biased by the choice of frequency weighting (C or A), when the spectrum of the masker is not matched to that of the target.

REFERENCES [1] Naylor G (2010). Limitations of Speech Reception Threshold (SRT) as an outcome measure in hearing-aid research. IHCON, oral presentation. [2] Pearsons KS, Bennett RL, Fidell S (1977). Speech levels in various noise environments, Bolt, Beranek and Newman. Inc. Canoga Park, California. [3] Bernstein JGW, Grant KW (2009). Auditory and auditory-visual speech intelligibility in fluctuating maskers for normal-hearing and hearing-impaired listeners. J. Acoust. Soc. Am. 125, p. 3358-3372. [4] Bernstein JGW, Brungart DS (2011). Effects of spectral smearing and temporal fine-structure distortion on the fluctuating-masker benefit for speech at a fixed signal-to-noise ratio. J. Acoust. Soc. Am. 130, p. 473-488. [5] Bernstein JGW (2011). Controlling signal-to-noise ratio effects in the measurement of speech intelligibility in fluctuating maskers, In: Dau T, Jepsen ML, Poulsen T, Dalsgaard JC. (Eds.) Speech Perception and Auditory Disorders. Proceedings of the Third International Symposium on Audiological and Auditory Research, ISAAR. Centertryk A/S: Denmark. [6] Naylor G, Burmand Johannesson R (2009). Long-term signal-to-noise ratio at the input and output of amplitude-compression systems. J. Am. Acad. Audiol. 20, p. 161-171. [7] Wagener K, Josvassen JL, Ardenkjær R (2003). Design, optimization and evaluation of a Danish sentence test in noise. Int. journ. Aud. 42, p. 10-17. [8] Boothroyd A, Nittrouer S (1988). Mathematical treatment of context effects in phoneme and word recognition. J. Acoust. Soc. Am. 84, p. 101-114. [9] Helfer KS, Freyman RL (2008). Aging and speech-on-speech masking. Ear Hear 29, p. 87-98. [10] Cohen J (2009). Statistical Power Analysis for the Behavioral Sciences. 2nd ed. New York: Psychology Press. [11] Lunner T, Hietkamp RK, Andersen MR, Hopkins K, Moore BCJ (2011). Effect of speech material on the benefit of temporal fine structure information in speech for young normal-hearing and older hearing-impaired participants. Ear Hear 33, p. 377-388. [12] Nielsen JB, Neher T, Dau T (2011). Towards a Danish speech material for speech-on-speech masking investigations, In: Dau T, Jepsen ML, Poulsen T, Dalsgaard JC. (Eds.) Speech Perception and Auditory Disorders. Proceedings of the Third International Symposium on Audiological and Auditory Research, ISAAR. Centertryk A/S: Denmark. [13] Robinson K, Summerfield QA (1996). Adult auditory learning and training. Ear Hear 17, p. S51-S65.

EriksholmResearchCentre
PART OF OTICON