3rd International Symposium on Auditory and Audiological Research

August 24 – 26, 2011

Nyborg Strand, Denmark

Towards a Danish speech material for speech-on-speech masking investigations

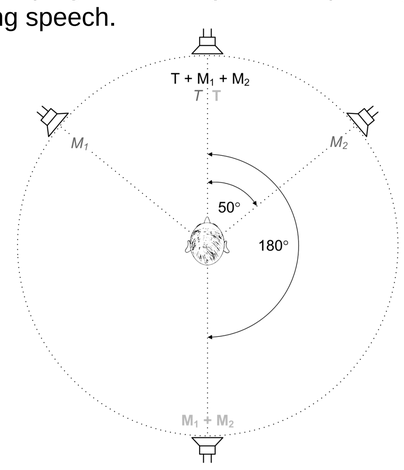
Jens Bo Nielsen and Torsten Dau • CAHR, DTU Elektro • Technical University of Denmark

Tobias Neher • Eriksholm Research Centre • Oticon A/S

Introduction

The objective of the present project is to create a Danish speech material designed specifically for investigations of speech-on-speech masking. Such masking is often divided into an “energetic” and an “informational” effect. The energetic effect corresponds to that of a noise signal with the same spectro-temporal characteristics as the masking speech. Informational masking is the additional effect caused by the confusion and the uncertainty regarding the target that arises when the masker is speech [1]. Investigations of speech-on-speech masking are expected to be useful for developing new strategies for improving speech intelligibility in environments with competing speech.

FIG 1. *Examples of different speech-on-speech masking set-ups with one target and two simultaneous maskers (T = target, M1 = masker 1, M2 = masker 2). The target is presented from the front while the maskers are presented from various directions. The difficulty of the listening task is very dependent on the location of the maskers. From [3].*



Speech-on-speech sentence material

The new material fulfils three requirements that are meant to facilitate effective speech-on-speech masking investigations:

i. Each sentence begins with a call sign that allows cuing of the target sentence. When target and masker sentences are similar, the call sign is essential for separating them.

ii. The material is open-set, i.e. the response options are effectively unlimited. Everyday speech recognition involves processes that are better simulated in open-set than closed-set tests, e.g. finding the target word in the lexical memory [2]. An open-set test also leads to a higher – and arguably more representative – speech recognition threshold (SRT).

iii. The speech rate is similar to natural speech to ensure ecological validity in this respect.

Sentence format

The format of the sentences in the corpus is similar to that of the TVM-corpus developed by Helfer and Freyman [2]. The sentences are based on the carrier:

***Navn* tænkte på ..... og ..... i går,**

where *Navn* is a name, and each blank represents a unique noun. This format has several advantages compared to natural sentences:

i. In listening tests, the scoring of the sentences is less prone to uncertainties on behalf of the test leader.

ii. Context effects are low and relatively consistent as long as frequent word pairs (such as knife and fork) are avoided.

iii. A large number of sentences is relatively easy to produce.

iv. The process of assessing the naturalness of the sentences is eliminated. However, target words must be assessed to avoid conflicts with the carrier sentence.

v. The two target words can be selected to produce sentences of similar duration. This ensures that masker and target sentences overlap.

Call signs

In order to reduce the risk that one of the call signs corresponds to the listener’s name in an actual test, three Danish female names with a low occurrence in the Danish population (yet not strange) were selected:

i. Dagmar (1429 occurrences, Jan. 2011)

ii. Asta (4352)

iii. Tine (7852)

The durations of the spoken names are similar (300-350 ms). The names were judged to be equally identifiable when presented simultaneously.

Target words

The target words were extracted from databases of Danish nouns according to the following requirements:

i. They contain one or two syllables. Three syllables were allowed if the word was shorter than eight letters.

ii. They contain at least one ‘strong’ consonant, e.g. /k/, /t/, /s/, /f/, /p/.

iii. They are not negative, emotional, technical, abstract or slang-like.

iv. They should be neutral and concrete.

Final sentences

The target words were combined in pairs with respect to duration and their number of lexical neighbors. (Two words are lexical neighbors if they only deviate by one phoneme). The number of neighbors was minimized by including at least one noun without any neighbors in each pair of target words.

The noun pairs were inserted into the carrier sentence and divided into three sets of 268 sentences each. Each set was recorded with one of three professional female talkers (AG, KL and MH) in a professional recording studio.

Equalization of sentence intelligibility

The recorded sentences were presented in a custom-made, speech-shaped background noise to seven young normal-hearing (NH) listeners. The initial signal-to-noise ratio (SNR) was -8 dB, increasing to 0 dB. The sentences were presented in rounds, raising the SNR by 2 dB for each round and only repeating the sentences that had not been correctly identified in the previous round. The omitted sentences were assigned a test result value corresponding to the SNR at which they were correctly identified. Sentences that were not identified at 0 dB were not tested further, but assigned a value of 2 dB.

The mean SNR across the seven listeners was calculated for each sentence. The results are shown in the figure below, where sentences have been sorted separately for each talker in terms of the mean SNR.

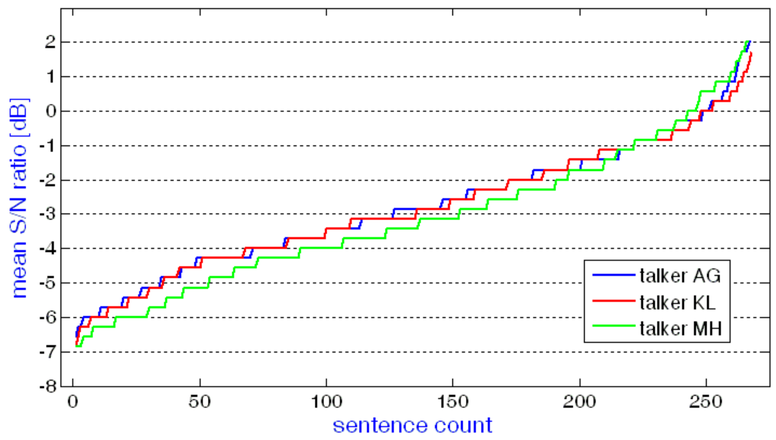


FIG 2. *The sentences of each talker are sorted separately in terms of the average SNR at which they were correctly identified. A high SNR indicates that the nouns in the sentence are difficult to identify.*

Sentences with a mean SNR above −1 dB and below −5.5 dB (−6 dB for MH) were discarded. The remaining 3 x 200 sentences were considered to be of approximately equal intelligibility and were randomly distributed into 10 lists of 20 sentences for each talker. These preliminary sentence lists formed the basis for a pilot validation of the material.

Pilot validation

SRTs were determined using a HINT-like adaptive procedure [4]. Seven new NH listeners participated, each listening to the 10 lists of one talker. The set-up was free-field with the target in front and the two maskers at ±50° as shown in Fig. 1. The two non-target talkers were used as maskers. Each masker was presented at 60 dB SPL and the SNR was calculated relative to this individual level. The following mean SRTs were obtained for the three talkers:

TABLE 1. *Mean SRT and mean within-subject standard deviation based on 20 SRTs for each talker (30 SRTs for KL). The results indicate a significant difference between the intelligibility of the talkers.*

	SRT	within-subject SD
Target AG	0.0 dB	1.2 dB
Target KL	0.7 dB	1.1 dB
Target MH	−1.4 dB	1.2 dB

Psychometric functions for the sentence intelligibility in the speech-on-speech setup were estimated. Each function below is based on the responses from one listener to the 10 test lists of one talker. The intelligibility was calculated as the percentage of correctly identified sentences (both nouns) for the last 16 levels of the adaptive procedure.

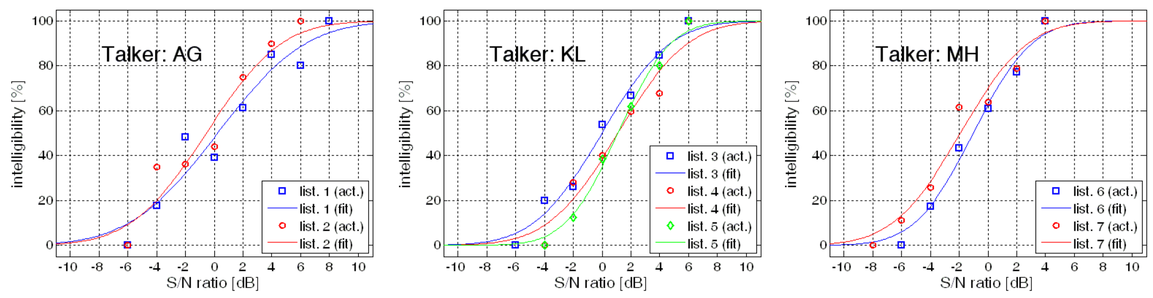


FIG 3. *Individual psychometric functions for the seven NH listeners in the pilot validation experiment. Each listener listened to10 lists from one (target) talker. The curves are best fit cumulative normal distributions to the actual measurements. The steepest slope of the curves vary from 8.3 %/dB (listener 1) to 13.9 %/dB (listener 5).*

Perspectives

The pilot validation shows that the present speech-on-speech material is characterized by

- a much higher SRT compared to other speech tests
- a relatively low within-subject standard deviation
- relatively steep psychometric functions

We thus expect that the material will be an effective tool in investigations of speech-on-speech masking. The material will be made publicly available.

Acknowledgements

We would like to thank Lise Bruun Hansen, Oticon A/S, Morten Løve Jepsen, CAHR, and Thomas Ulrich Christiansen, CAHR, for their valuable contributions to this project.

References

[1] K. S. Helfer and R. L. Freyman (2008). Lexical and indexical cues in masking by competing speech. *J. Acoust. Soc. Am.* 125 (1): 447-456.

[2] C. G. Clopper et al. (2006). Effects of open-set and closed-set task demands on spoken word recognition. *J. Am. Acad. Audiol.* 17 (5): 331-349.

[3] T. Neher et al. (2009). Benefit from spatial separation of bilateral hearing-aid users: Effects of hearing loss, age, and cognition. *Int .J. Audiol.* 48, 758-774.

[4] M. Nilsson et al. (1994). Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise. *J. Acoust. Soc. Am.* 95 (2): 1085-1099.