Collecting a Corpus of Dutch Noise-induced ‘Slips of the Ear’

Odette Scharenborg[1,2]*, Eric Sanders*[3]*, and Bert Cranen*[1,3]

[1]Centre for Language Studies, Radboud University Nijmegen, The Netherlands
[2]Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, The Netherlands
[3]Centre for Language and Speech Technology, Radboud University Nijmegen, The Netherlands

O.Scharenborg@let.ru.nl, E.Sanders@let.ru.nl, B.Cranen@let.ru.nl

Abstract

When trying to understand how listeners recognise words, listeners’ misperceptions, so-called ‘slips of the ear’, can reveal important aspects of the underlying mechanisms of normal word recognition. Such misperceptions shed light onto how inferences are made by listeners about acoustic details in the speech signal and how these interact with other sound sources in the background. On the other hand, if speech from a particular speaker is more prone to being misperceived than that from another speaker, these misperceptions may also shed light onto speaker characteristics. To study these phenomena, misperceptions that occur consistently are invaluable. Although such confusions are quite rare, within the Marie Curie INSPIRE project, software has been developed to efficiently collect such consistent confusions for different languages. Using this software, we have started to collect Dutch consistent confusions. Single words, embedded in five different types of noise at different SNRs, produced by four speakers were presented to Dutch listeners. In a preliminary analysis, consistent confusions were analysed in terms of phoneme substitutions, insertions, and deletions, reconstructions of words using background noise, and eccentric cases. Moreover, the number and types of consistent confusions obtained in the different noise types and from different speakers are compared.

Index Terms: speech perception, misperception, corpus, noise, Dutch

1. Introduction

‘Slips of the ear’ are generally understood to be speech misperceptions that occur not because of a speaker’s misproduction but because a listener ‘mishears’ the intended word (sequence) (e.g., [1,2]). When listening to speech, listeners map a highly variable signal onto discrete lexical representations (such as words) [3]. Occasionally this process fails, resulting in misperceptions. Misperceptions show that listeners are actively interpreting speech [4] and are trying to reconstruct the perceived sounds to make them fit candidate phonemes, words, or syntactic structures [5]. Investigating the circumstances in which these misperceptions, and in particular consistent misperceptions, occur and analysing these misperceptions in terms of their change from the target [2] can reveal important aspects of the underlying mechanisms of normal word recognition. For instance, consistent misperceptions can shed light onto how inferences are made by listeners about acoustic details in the speech signal [2] and how these interact with other sound sources in the background.

Misperceptions could in principle also inform about speaker characteristics that may deteriorate intelligibility, especially in the presence of background noise. If speech from a particular speaker is more prone to being misperceived in background noise than speech from another speaker, these misperceptions may highlight speaker characteristics or strategies applied by speakers that are more or less successful in making themselves understood.

To study these phenomena, corpora containing misperceptions are invaluable. Misperceptions however are rare, even in noise [6,7]. Most research on misperceptions has focused on naturalistic settings (e.g., [1,2,8-10]). Although these type of misperceptions are obviously most authentic [2,5], the speech signals on the basis of which the misperception occurs is almost never recorded, making it difficult to further analyse these misperceptions or to even replicate these misperceptions with other listeners [5]. To overcome these issues, misperceptions have been induced under controlled conditions in a laboratory setting, for instance using faint speech [11] or in noise [5,6].

In this study, we present a corpus of Dutch noise-induced slips of the ear that is currently being collected, which focuses on consistent misperceptions in the presence of background noise. In the analysis of the consistent misperceptions collected so far we focus on the following questions: What are the types of consistent misperceptions that occur? Are the number and types of consistent misperceptions dependent on the type of background noise and/or on the speaker? Do more difficult listening conditions as shown by a lower recognition rate result in more consistent misperceptions?

We present the word stimuli, the five noise maskers, and the elicitation procedure in Section 2. Section 3 presents an analysis of 113 Dutch word-level consistent misperceptions following [5]. The misperceptions were obtained using software that has been developed to efficiently collect such misperceptions for different languages [5-7].

2. Experimental Set-up

2.1. Dutch spoken word corpus

A set of 3000 words, spoken by 4 speakers (2 male, 2 female), was recorded. These words were selected from an initial list of 3030 Dutch words which adhered to the following criteria:

- All words were mono- or bisyllabic
- No homographs, homophones, homonyms
- No words contained diacritics
- No words that are sensitive to spelling errors
- All words had a frequency of occurrence of at least 13 per 9 million in the e-Lex database [12]
- No potentially offensive or disturbing words

The words were ordered alphabetically, and to avoid list intonation the words were prompted word by word on a computer screen. Additionally, speakers were explicitly instructed to avoid list intonation, and were asked to leave a short pause in between the words. Recordings were made in a well-isolated sound booth using high quality equipment. The speakers, all PhD students of around 25 years of age, were native Dutch speakers without a (clear) regional accent. The recordings of the words were made in two or three sessions of one to two hours per speaker, with a break as often as the speaker desired to avoid fatigue. When a speaker made a mistake, the word was repeated.

The recordings were first crudely split into sound files containing a single word using an mp3/wav splitter based on silence detection [13]. The begin and end points of the words were then automatically detected using an automatic speech recogniser (SPRAAK, [14]), after which Praat [15] was used to detect the nearest zero crossing at 0.2 seconds of silence before the start of the acoustic signal and 0.05 seconds after the end at which points the sound files were cut. Finally, all automatically segmented recordings were judged by a human listener. Items that, due to incidental segmentation errors of the splitter, consisted of two words, were split manually. For each speaker, a couple of words contained flaws (e.g., the recordings were too loud or soft or contained buzzes) or words were mispronounced. These words were discarded for all speakers, yielding four identical sets of 3,000 words, one for each of the four speakers. These 12,000 words constituted the words used for the listening experiments.

2.2. Masking noises

Five different noise types were prepared analogous to [5]: 1. Stationary noise with a spectrum similar to the long term spectrum of speech, i.e., speech-shaped noise (SSN), 2. Speech-shaped noise of which the amplitude was modulated similar to the speech of a single speaker (BMN1), 3. Speech-shaped noise of which the amplitude was modulated similar to the speech of three speakers talking simultaneously (BMN3), 4. Four speaker babble noise, created by mixing four randomly selected speech fragments, one each from four speakers, taken from the Dutch word stimuli described in the previous section (BAB4). 5. Eight speaker babble noise, the same as BAB4, but using eight different speech fragments from different speakers (BAB8). SSN is a stationary masker, while the other four are non-stationary. BAB4 and BAB8 are composed from natural speech and can therefore be expected to induce both energetic and informational masking [17,18].

The noises were artificially mixed with the word stimuli at different signal-to-noise (SNR) levels. The SNR ranges were chosen during a pretest with 16 listeners (mean age: 23.8, SD: 1.9) who carried out a word recognition experiment (see Section 2.4), and such that percentages correct for each noise type were between 50% and 60%. Table 1 shows an overview of the noises, the SNR range used in our experiment, and the average SNR over all tokens presented to our listeners.

Table 1. *The five maskers, their SNR ranges, and the average SNR over all tested tokens in our experiment.*

Name	**Masker**	**SNR range (dB)**	**Avg. SNR**
SSN	Speech-shaped noise	-1 to +3	0.8
BMN1	Speech modulated noise	-2 to +2	-0.2
BMN3	3-talker babble modulated noise	-1 to +3	0.9
BAB4	4-talker babble	+3 to +7	4.9
BAB8	8-talker babble	+3 to +7	4.9

2.3. Participants

Fifty-five native Dutch speakers were drawn from the Radboud University Nijmegen participant pool, and were paid € 5 for their participation in the experiment. The mean age was 23.4 years (SD: 2.8; age range: 19 – 30).

2.4. Experimental software and procedure

The elicitation of the consistent word confusions was carried out using the elicitation software developed by [5-7] which was embedded as a custom Java applet in a webpage [19]. Listeners had to recognise words embedded in noise in blocks of 50 tokens. They were asked to type in their response into a textfield in the Java applet using a keyboard. Listeners were randomly assigned a speaker and noise type combination which was held constant within a block but changed between blocks. In the course of half an hour they finished as many blocks as they could. The experiment took place in a quiet classroom containing up to eight participants. Stimuli were presented binaurally through headphones.

The elicitation software used adaptive token-pruning techniques to determine whether a specific word-in-noise token was worth to keep pursuing (*active* token) or whether it should be dropped (*discarded* token). Responses to discarded tokens were either all different or a pre-set number of listeners all gave the correct answer. A third type of tokens were the *exhausted* tokens, which represent the 'interesting cases'. These tokens constitute the set that were presented a pre-set maximum number of times and did not fulfil the criteria to drop out and therefore warrant further investigation because they might be consistent confusions. We followed the settings of the heuristics as in [5].

3. Analysis of the Consistent Confusions

Due to technical problems during one testing session, eight participants did not carry out the experiment for the full half hour, but only for approximately 15 minutes. So far, 29,135 responses to 7,880 different stimulus-noise pairings (tokens) have been collected (3.7 responses per token on average). Of the screened tokens, 6,584 (83.6%) were dropped, 496 (6.3%) resulted in 'interesting cases', and 800 (10.2%) remained active, and will be further tested in the continuation of this experiment in the future.

Interesting cases that had a listener agreement of at least 60% were selected for further analysis and are what we here call consistent confusions. This subset consisted of 113 tokens.

Table 2. *The percentage correctly recognised stimuli, the total number of times a certain background noise was presented, the total number of consistent confusions, and its break-down in types of confusions, per noise type.*

Noise	Recognition		Confusions			
	Acc.	Total	Total	Single	Dual	Other
SSN	44.3	5900	28	22	2	4
BMN1	51.4	4789	21	12	4	5
BMN3	50.8	5402	27	18	4	5
BAB4	49.2	5193	18	15	1	2
BAB8	46.1	4905	19	14	1	4

Table 3. *The percentage correctly recognised stimuli, the total number of times a stimulus of a certain speaker was presented, the total number of consistent confusions, and its break-down in types of confusions, per speaker.*

Speaker	Recognition		Confusions			
	Acc.	Total	Total	Single	Dual	Other
S1	33.0	6077	28	20	1	7
S2	64.3	6756	27	21	3	3
S3	37.6	6751	37	21	8	7
S4	56.7	6605	21	18	0	3

3.1. Consistent confusions per noise type and speaker

Table 2 shows the percentage correctly recognised words (Recognition – Acc(uracy).) and the number of times a word embedded in a certain noise was presented to a listener (Recognition – Total) per noise type. Research on the effect of maskers on speech perception has shown that at equal SNRs, highly-modulated maskers result in higher accuracies than static maskers [20,21]. This is indeed what we observe: The accuracy for the static masker SSN (44.3%) is lower than that of the speech-modulated maskers BMN1 (51.4%) and BMN3 (50.8%), while the average SNRs for these three maskers was (highly) similar. The accuracies for the two maskers with informational masking (BAB4 and BAB8) was similar to that of the two speech-modulated maskers but at a noise level which was approximately 4 dB lower. Babble noise can besides energetic masking also cause informational masking and thus is a more effective masker than the other three purely energetic maskers. Babble noise made of a larger N has been shown to cause less informational masking than babble noise made of fewer talkers [22,23]. In our case, however, the results for BAB8 are worse than those for BAB4. Future research is needed to explain this finding.

Table 2 furthermore shows the total number of consistent confusions and the breakdown into different types of consistent confusions (see Section 3.2) for each noise type separately. The number of consistent confusions per noise type ranged from 18 (BAB4) to 28 for SSN. When comparing the number of consistent confusions with the accuracy for the five noise conditions, it can be seen that there is no clear relationship between number of consistent confusions and the accuracy. A lower number of correctly recognised items does not necessarily lead to a higher number of consistent confusions. Even when taking into account the number of times a noise was presented as background noise, the number of consistent confusions was relatively high for BMN3, while it was relatively low for BAB4. BMN3 and BAB4 were presented a comparable number of times. While BAB4 had a slightly lower accuracy than BMN3, it had far fewer consistent confusions than BMN3. So, in short, the number of consistent confusions seems to be dependent on the type of noise, i.e., the more stationary, purely energetic maskers SSN and BMN3 resulted more often in consistent confusions.

Table 3 shows the accuracy in terms of percentage correctly recognised stimuli per speaker, and the total number of times a stimulus of a certain speaker was presented to a listener. Moreover, the total number of consistent confusions, and its break-down in types of confusions per speaker are shown, these are discussed in Section 3.2. The average SNRs for the four speakers was similar, they ranged from 2.12 (S4) to 2.25 (S1). Table 3 shows that not all speakers were recognised equally well. Speaker S1 was recognised particularly bad at only 33.0% correct, while speaker S2 was recognised best; 64.3% of the words produced by this speaker were recognised correctly. Such large differences in accuracy between individuals is not uncommon: [24] also found large differences in intelligibility of speakers in high noise conditions. We will come back to this finding below.

The number of consistent confusions also differed quite substantially between the four speakers. The number of consistent confusions ranged from 21 for speaker S4 to 37 for speaker S3. Also for the speakers, there does not seem to be a clear relationship between accuracy and the number of consistent confusions, even when taking into account the number of times a token of a certain speaker was presented. For instance, speaker S2 and speaker S4 showed an opposite pattern: S2's speech resulted in fewer misrecognitions than S4's speech but it resulted in a higher number of consistent confusions, while the number of presentations of speech from these two speakers was similar. So, the number of consistent confusions is dependent on the speaker, analogous to what was found for the different noise types.

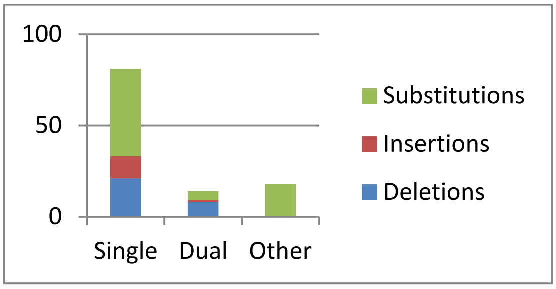

Figure 1. *Breakdown of the types of consistent confusions in terms of the number of tokens per confusion category.*

3.2. Types of consistent confusions

The consistent confusions are first analysed without taking into account noise type or speaker. Following [5], we analysed the confusions in terms of (simple or dual, i.e., two consecutive phonemes) phoneme substitutions, insertions, and deletions, reconstructions of words using background noise (referred to as *compounds* by [5]), and eccentric cases. Compounds are misperceived words where the listener appeared to have constructed the misperception by combining information from the target word and the background noise in ways more complex than the simple case of single or dual phoneme deletions, substitutions, or insertions. Regarding the simple or dual type, we also looked at the types of phonemes and phonetic features involved in the misperceptions, and the location within the word where the insertion, deletion, or substitution occurred.

3.2.1. Single phoneme cases

Our analyses showed that of the 113 consistent confusions, by far the largest part of misperceptions (71.7%) occurred due to the insertion (12 tokens), deletion (21), or substitution (48) of a single phoneme. Figure 1 shows a breakdown of the types of confusions in terms of the total number of tokens per confusion category. Of the 12 single phoneme insertions, eight occurred at word offset. Single phoneme insertions mainly involved /r/ (3 times; e.g., *kerken*: /kɛrkə/, E: churches; *kerker:* /kɛrkər/, E: dungeon) and /k,t,ə/ (all twice). Of the 21 deletions, only two occurred at word onset, the remaining deletions occurred at word offset. The most frequently occurring word offset deletion was that of /ə/ (e.g., *schoenen*: /sxunə/, E: shoes; *schoen:* /sxun/, E: shoe; 10 occurrences), followed by that of /t/-deletion (e.g., *gevoeld*: /xəvult/, E: felt; *gevoel:*, /xəvul/, E: feeling; 4 times).

Phoneme substitutions mainly involved consonants, only in one case a vowel was substituted. Moreover, except for two cases, all substitutions occurred at word onset (33 times) or word offset (12 times). Thirty of the 48 substitutions (62.5%) could be explained by the change or misperception of a single phonetic feature. Most often, place of articulation was misperceived (e.g., *kop:* /kɔp/, E: cup; *pop:* /pɔp/, E: doll; 14 times), followed by manner of articulation (e.g., *neef:* /nef/, E: male cousin; *leef:* /lef/, E: live; 10 times), and voicing (e.g., *pet:*/pɛt/, E: cap; *bed:* /bɛt/, E: bed; 6 times). Interestingly, where the place and manner of articulation misperceptions occurred for speech from all speakers, the voicing misperception occurred for 5 out of 6 observations for speaker S2. Since our sample is still rather small, it is not possible to draw any conclusions yet, but in future analyses we will pay special attention to speaker-specific misperceptions.

3.2.2. Dual phoneme cases

Fourteen of the 113 consistent confusions (12.4%) consisted of two consecutive phonemes that were misperceived. Dual deletions were the majority of this type of misperception, which occurred equally often at word onset and offset (e.g., *langer:* /lɑηər/, E: taller; *lang*: /lɑη/, E: tall; 8 times). Twice two consecutive insertions occurred, and four cases involved combinations of phoneme changes, e.g., a substitution followed by an insertion (e.g., *schema:/* sxema/, E: scheme; *schemer:/*sxemər/, E: dusk).

3.2.3. More complex cases and eccentric cases

The remaining 18 of the 113 consistent confusions (15.9%) consisted of misperceived words where the listener appeared to have constructed the misperception by combining information from the target word and the background noise in ways more complex than the simple case of single or dual phoneme deletions, substitutions, or insertions, e.g., *wonder:* (/wɔndər/, E: wonder) recognised as *grondig* (/xrɔndəx/, E: thorough). This particular misperception can be reconstructed through a substitution followed by an insertion at word onset, and another substitution at word offset.

The number of differences between the target word and its misperception can be rather large, e.g., *somber* (/sɔmbər/, E: sad) was consistently misperceived as *zonde* (/zɔndə/, E: sin), which constitutes a voicing of the first consonant, two subsequent changes in place of articulation in the middle of the word and a deletion at word offset. Inspection of the background noise will reveal whether the misperception is indeed a particularly complex case or a word that occurred in the background noise, in which case the misperception would be classified as an 'eccentric case'.

3.3. Distribution of the types of consistent confusions over speakers and noise types

Table 2 shows the total number of consistent confusions, and its break-down in types of confusions per noise type. For all noise types, the single phoneme cases constitute by far the largest part. However, the distribution of the different confusion types over the five noise types seems to show two different trends. The number of dual cases (and rest category containing the more complex cases of misperceptions) seems to be relatively higher for the two speech-shaped babble-modulated noise types (BMN1 and BMN3) compared to the other three noise types. Future data collection will show whether this trend is statistically significant.

Table 3 shows the total number of consistent confusions, and its break-down in types of confusions per speaker. Although there are clear differences in number of consistent confusions resulting from the different speakers (compare S4 with 21 confusions to S3 with 37 confusions), the distribution of the types of consistent confusions is fairly similar for the four speakers, i.e., the single phoneme cases are by far the most frequent type of consistent confusion for all speakers. For S3, however, dual phoneme and more complex cases are relatively more frequent compared to the other three speakers. Future data collection will show whether this trend is statistically significant.

A more detailed analysis of the single phoneme cases in relation to speaker also showed no difference in distribution of the types of single phoneme misperceptions over the speakers, apart from the finding that the voicing misperceptions were, apart from one case, all related to speaker S2. In Dutch, voiced obstruents tend to be devoiced, but since for this particular speaker voiceless obstruents tended to be interpreted as their voiced counterparts, these misperceptions are most likely not due to a simply sloppier speaking style for this speaker. Overall, these results seem to suggest that a certain speaker's pronunciation can be prone to more misperceptions, but that the type of misperception is not speaker specific.

4. Concluding Remarks

We presented a corpus of Dutch noise-induced slips of the ear that is currently being collected within the framework of the Marie Curie INSPIRE project [25]. Our corpus will be part of a large multi-lingual consistent confusions corpus that is being collected in collaboration with INSPIRE partners from the University of the Basque Country and the University of Sheffield. Upon completion, this consistent confusion corpus will be distributed via the INSPIRE website [25].

Thus far, we analysed 113 Dutch noise-induced consistent confusions collected so far with several questions in mind. Although our data set is still relatively small, several interesting observations could be made. First, regarding the type of consistent confusions found: the bulk of the consistent confusions were caused by single phoneme insertions, deletions, or substitutions. Like was found in other studies, a majority of these consistent confusions was caused by substitutions of consonants [5,8].

Secondly, the number of confusions was speaker dependent; speech of some speakers seems inherently more prone to be misperceived. The types of misperception, however, are not speaker specific, i.e., all types of confusions occurred relatively equally often for the four speakers. A more detailed analysis is needed to find out whether specific speaker characteristics, e.g., long-term average spectra characterising the speaker's voice, or particularly weak consonant sounds, can be identified that explain these differences.

Thirdly, the number of confusions seems to be dependent on the type of background noise. More consistent confusions occurred in the more stationary, purely energetic maskers SSN and BMN3 compared to the more complex maskers. Possibly, more complex maskers give rise to more different misperceptions for the different listeners causing less agreement between the listeners. Moreover, different noise types seem to cause a different distribution of the types of consistent confusions, although the future will tell whether the observed pattern will hold in a larger data set. Finally, more difficult listening conditions, for instance due to a more difficult background noise or an inherently less intelligible speaker, did not result in more consistent confusions.

5. Acknowledgements

This project has been funded with support from the European Commission under Contract FP7-PEOPLE-2011-290000. The research by O.S. was partly sponsored by a Vidi-grant from the Netherlands Organisation for Scientific Research (NWO). The authors would like to thanks the four speakers for providing all the stimuli, Martin Cooke for providing the experimental software and creating the noise files, Gies Bouwman for building the experimental website, and Esther Kroese for running the listening experiments.

6. References

[1] Bond, Z.S., “Slips of the ear: Errors in the perception of casual conversation”, New York: Academic Press, 1999.

[2] Tang, K.., Nevins, A., “Measuring Segmental and Lexical Trends in a Corpus of Naturalistic Speech”, Proceedings of the 43rd Meeting of the North East Linguistic Society, 2013.

[3] Perkell, J.S., Klatt, D.H. (Eds.), “Invariance and variability of speech processes”, Hillsdale, NJ: Lawrence Erlbaum Associates, Inc, 1986

[4] Cutler, A., “The reliability of speech error data”, In :Slips of the tongue and language production, (Ed: A. Cutler), 7-28, 1982.

[5] Garcia Lecumberri, M.L., Tóth, A.M., Tang, Y., Cooke, M., “Elicitation and analysis of a corpus of robust noise-induced word misperceptions in Spanish”, Proceedings of Interspeech, Lyon, France, 2013.

[6] Cooke, M., “Discovering consistent word confusions in noise,” Proceedings of Interspeech, Brighton, UK, 1887-1890, 2009.

[7] Cooke, M., Barker, J., Garcia Lecumberri, M.L., “Crowdsourcing in speech perception”, In Crowdsourcing in language and speech, (Ed. J. Wiley), 137-172, 2013.

[8] Browman, C.P., “Perceptual processing: Evidence from slips of the ear. In Errors in linguistic performance: Slips of the tongue, ear, pen and hand”, (Ed. V.A. Fromkin), 213-230. New York: Academic Press, 1980.

[9] Meringer, R., “Aus dem Leben der Sprache”, Berlin: B. Behr, 1908.

[10] Labov, W., “Principles of linguistic change, volume III: Cognitive and cultural factors”, Malden, Massachusetts: Wiley-Blackwell. Amsterdam: Walter de Gruyter/Mouton, 2010.

[11] Cutler, A., Butterfield, S., “Rhythmic cues to speech segmentation: Evidence from juncture misperception”, Journal of Memory and Language, 31:218-236, 1992.

[12] Available online at: http://tst-centrale.org/nl/producten/lexica/e-lex/7-25

[13] Available online at: http://www.pistonsoft.com/mp3-splitter.html

[14] Demuynck, K., Roelens, J., Van Compernolle, D., Wambacq, P., “SPRAAK: an open source ‘SPeech Recognition and Automatic Annotation Kit’”, Proceedings of Interspeech, Brisbane, Australia, 22-26, 2008.

[15] Boersma, P., Weenink, D., “Praat. Doing phonetics by computer [Computer program]”, retrieved from http://www.praat.org, 2013.

[16] Bird, H., “Slips of the ear as evidence for the postperceptual priority of grammaticality”, Linguistics, 36(3):469-515, 1998.

[17] Carhart, R., Tillman, T., Greetis, E., “Perceptual masking in multiple sound backgrounds”, J. Acoust. Soc. Am., 45:694-703, 1969.

[18] Brungart, D., Simpson, B., Ericson, M., Scott, K., “Informational and energetic masking effects in the perception of multiple simultaneous talkers”, J. Acoust. Soc. Am., 100:2527– 2538, 2001.

[19] http://biglisten.cls.ru.nl/nl/

[20] Festen, J.M., Plomp, R., “Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing”, J. Acoust. Soc. Am., 88:1725, 1990.

[21] Simpson, S., Cooke, M., “Consonant identification is N-talker babble is a nonmonotonic function of N”, J. Acoust. Soc. Am., 118:2775-2778, 2005.

[22] Carhart, R., Johnson, C., Goodman, J., “Perceptual masking of spondees by combinations of talkers”, J. Acoust. Soc. Am., 58: 535, 1975.

[23] Freyman, R. L., Balakrishan, U., Helfer, K., “Effect of number of masking talkers and auditory priming on informational masking in speech recognition”, J. Acoust. Soc. Am., 115:2246–2256, 2004.

[24] Barker, J., Cooke, M., “Modelling speaker intelligibility in noise”, Speech Communication, 49:402-417, 2007.

[25] http://www.inspire-itn.eu/