The Influence of Semantically Related and Unrelated Text Cues on the Intelligibility of Sentences in Noise

Adriana Zekveld, Mary Rudner, Ingrid Johnsrude, Joost M Festen,
Johannes H M van Beek and Jerker Rönnberg

Linköping University Post Print

N.B.: When citing this work, cite the original article.

Abstract

Objectives: In two experiments with different subject groups, we explored the relationship between semantic context and intelligibility by examining the influence of visually presented, semantically related and unrelated 3-word text cues on perception of spoken sentences in stationary noise across a range of signal-to-noise ratios (SNRs). Additionally, in Experiment (Exp) 2, we explored the relationship between individual differences in cognitive factors and the effect of the cues on speech intelligibility.

Design: In Exp 1, cues had been generated by participants themselves in a previous test session (*own)* or by someone else (*alien).* These cues were either appropriate for that sentence (*match*), or were for a different sentence (*mismatch*). A condition with *non-word* cues, generated by the experimenter, served as a control. Experimental sentences were presented at 3 signal to noise ratios (dB SNR) corresponding to the entirely correct repetition of 29%, 50%, or 71% of sentences (speech reception thresholds; SRTs). In Exp 2, semantically *matching* or *mismatching* cues and *non-word* cues were presented prior to sentences at SNRs corresponding to SRTs of 16% and 29%. The participants in Exp 2 additionally performed tests of verbal working memory capacity and the ability to read partially masked text.

Results: In Exp 1, *matching* cues improved perception relative to the *non-word* and *mismatching* cues, with largest benefits at the SNR corresponding to 29% performance in the SRT task. *Mismatching* cues did not impair speech perception relative to the nonword-cue condition, and no difference in the effect of *own* and *alien* matching cues was observed. In Exp 2, *matching* cues improved speech perception as measured using both the percentage of correctly reported words and the percentage of entirely correctly reported sentences. *Mismatching* cues reduced the percentage of repeated words (but not the sentence-based scores) compared to the nonword-cue condition. Working memory capacity and ability to

read partly masked sentences were positively associated with the number of sentences

repeated entirely correctly in the *mismatch* condition at the 29% SNR.

Conclusions: In difficult listening conditions, both relevant and irrelevant semantic context

can influence speech perception in noise. High working memory capacity and good linguistic

skills are associated with a greater ability to inhibit irrelevant context when uncued sentence

intelligibility is around 29% correct.

read partly masked sentences were positively associated with the number of sentences

Introduction

The comprehension of speech involves many cognitive and perceptual processes that map a complex, time-varying acoustic signal onto a meaningful representation. The relative contributions of stimulus-driven and knowledge-driven processes to intelligibility probably depend on audibility factors (c.f., Plomp, 2002; Rönnberg, et al., 2010). The relevance of knowledge-driven processes may increase when listening is challenging because of interfering noise and/or hearing impairment (Rönnberg, 2003; Pichora-Fuller, et al., 1995; Rönnberg, et al., 2008; Wingfield & Tun, 2007). Furthermore, the recruitment of such compensatory processes appears to be associated with increased listening effort, as measured indirectly using memory tests and more directly using pupillometry (Wingfield, et al., 2005; Zekveld, et al., 2010, 2011).

The working memory framework of 'Ease of Language Understanding' describes the interplay between explicit and implicit processes in speech perception (Rönnberg, 2003; Rönnberg et al., 2008, Stenfelt & Rönnberg, 2009). According to the framework, 'implicit' processes are rapid and automatic, and 'explicit' processes are generally more effortful. In the former case, stimulus-driven linguistic input is matched automatically and implicitly to stored phonological representations in long-term memory. In the latter case, sub-optimal conditions interrupt this automatic binding process. The resulting 'mismatches' require explicit, knowledge-driven processing in order to be resolved. In this study, we manipulate semantic context to produce matching and mismatching conditions.

Contextual information can help listeners compensate for ambiguous or degraded acoustic information (Miller, et al., 1951; Miller & Isard, 1963, Kalikow, et al., 1977; Sheldon, et al., 2008; Pichora-Fuller et al., 1995; Wingfield, 1996; Plomp, 2002; MacDonald et al., 2008). Such contextual constraints exploit, for example, the knowledge of words and

grammar of a language, world knowledge, knowledge of the topic of conversation and speaker characteristics (Bronkhorst et al., 1993; Samuelsson & Rönnberg, 1991, 1993; Tesink et al., 2009; Nakano et al., 2010; Hagoort & van Berkum, 2007; Boothroyd & Nittrouer, 1988). Such contextual knowledge may help the listener to build a conceptual representation of the unfolding sentence, or to predict potential words or phrases (e.g., van Rooij & Plomp, 1991; Drager & Reichle, 2001). Several studies have examined the effects of semantic context on the intelligibility of sentence-final words (e.g., Freyman et al., 2004; Sheldon et al., 2008; Obleser & Kotz, 2010; Dubno et al., 2000; Pichora-Fuller et al., 1995). The 'Speech Perception in Noise' (SPIN) sentences (Kalikow et al., 1977; Bilger et al., 1984; Elliott, 1995) were specifically developed to include sentences with final words that are predictable from the sentence context (high-context sentences: e.g., "The sport shirt has short sleeves") or not (low-context sentences: e.g., "John talked about the sleeves"). The effect of semantic context is necessarily larger in suboptimal as compared to optimal listening conditions (c.f., Dubno et al., 2000; Sheldon et al., 2008; Pichora-Fuller et al., 1995). An interaction between stimulus degradation and the effect of semantic information has been observed for other sentence sets (Grant & Seitz, 2000; MacDonald et al., 2008), for the perception of text (Meyer et al., 1975; Becker & Killion, 1977; Holcomb, 1993) and for lipreading (Lansing & Helgeson, 1995). Drager and Reichle (2001) observed that context, consisting of five to eight sentences prior to target sentences, increased target sentence intelligibility. Additionally, Helfer and Freyman (2008) observed that providing information about the sentence topic improved sentence perception in noise.

Mäntylä and Nilsson (1983) and Bäckman et al. (1984) explored the effects of context in a rather different way. They demonstrated that semantically related cues generated by the participants themselves (*'subject-generated cues'*) can dramatically improve recall of memorized words. In their test paradigm, participants were first asked to generate three-word

associations to a set of single words. The rationale for presenting subject-generated cues was

that they likely represent a blend of associations, imagery, ideas, and words that will be

associated with each sentence in a personal and particularly meaningful way. The associations

were later presented as cues and the participants were asked to recall the target word. Even

when subjects were not forewarned of this memory task, or when the session was delayed by

several days or weeks, recall performance was remarkably high (Mäntylä, 1986a; Mäntylä &

Nilsson, 1988). For immediate recall, subject-generated cues resulted in 95% correct

performance, whereas the cues generated most frequently by a pool of subjects resulted in

83% correct recall (Mäntylä & Nilsson, 1983). A lower performance of 60% was observed for

"alien" cues (cues generated by a randomly selected subject).

The main aim of the current study is to examine the influence of semantically related

and unrelated three-word text cues on the intelligibility of complete, natural sentences in

stationary noise, at three (Exp 1) or two (Exp 2) different noise levels. In Exp 1, five types of

text cues were presented: subject-generated (*own*) or someone else's (*alien)* cues that had

been generated for the subsequently presented sentence (*matching*) or for another sentence

(*mismatching*). A baseline condition of experimenter-generated *non-word* cues was also

employed. The cue paradigm allowed us to degrade the auditory target sentence while

keeping the cue quality constant, which is not possible in materials where the context and

target are within the same stimulus (in the SPIN materials, for example). We used the

Mäntylä and Nilsson procedure to generate cues in Exp 1, in order that meaningful contextual

support would be maximal. However, "own" cues turned out to yield similar results to "alien"

cues, so in Exp 2, participants did not generate cues themselves. In addition to the *non-word*

baseline, *matching* and *mismatching alien* cues were presented: these were the cues most

frequently reported by the participants of Exp 1. The rationale for presenting *non-word* cues

as a neutral baseline was that they enabled the assessment of the separate influence of

matching and mismatching cues on speech comprehension (e.g., Cardillo et al., 2004; Aydelott & Bates, 2004; Neely, 1991). Without a neutral baseline, it is not possible to assess whether differences between the matching and mismatching cues are due to a facilitative effect of matching cues, or to inhibitory effects of mismatching cues, or a combination. We used non-word cues (instead of no cues) in order to keep conditions as similar as possible (i.e., every trial involved a visual cue followed by an auditory sentence).

In Exp 2, we also measured individual differences on two cognitive tests thought to be relevant to speech-in-noise performance. The Reading Span test (Daneman & Carpenter, 1980; Rönnberg et al., 1989) measures working memory capacity, correlates with performance on a wide range of cognitive tasks, and also correlates with speech in noise performance (Foo et al., 2007; Lunner, 2003; Rudner et al., 2008; 2009; 2011; see Akeroyd, 2008, for a review). The text reception threshold test (Zekveld et al., 2007) evaluates the ability to read partially masked, visually presented text, and also appears to correlate with speech-in-noise performance (Zekveld et al., 2007; George et al., 2007; Kramer et al., 2009). We examined the relation between these variables and cued sentence perception in noise.

Experiment 1

In Exp 1, we examined whether *matching* cues facilitate perception of sentences in noise relative to *mismatching* and *nonword* cues. Furthermore, we examined whether cues generated by the participants themselves for a particular sentence would be more effective than cues generated by others, as shown for recall tests (Mäntylä & Nilsson, 1983). We predicted that, relative to the nonword control condition, speech perception would be impaired in the *mismatching* cue condition and facilitated in the *match* condition, because the processes operating on contextual information would be helpful in the *match* case but unhelpful in the *mismatch* case. We additionally expected that *own* cues would have larger

6

facilitative and inhibitory effects than *alien* cues, as own cues may activate particularly

meaningful associations and thus offer maximally strong supportive context.

Methods Experiment 1

General procedure

Exp 1 consisted of two sessions performed on two consecutive days. Session 1

included hearing and vision screening, determination of speech reception thresholds (SRTs),

and cue generation. Session 2 included determination of SRTs and testing the effect of cues

on intelligibility. All testing took place in a sound-attenuating room. The tests are first

described in general, and then in detail in subsequent sections.

In session 1, hearing and visual acuity were measured by means of a pure-tone

audiogram and a near vision screening test. Subsequently, participants performed a 26-item

practice SRT (Plomp & Mimpen, 1979) in noise test (1-up-1-down procedure). Then, the SRT

test was administered three times in order to determine the SNRs that resulted in the repetition

of 29%, 50%, and 71% of sentences completely correctly. After that, participants generated

cues for 378 sentences selected from the set of 507 sentences spoken by a female. This set of

sentences was developed by Versfeld et al. (2000, see below). These sentences were presented

visually in session 1, but auditorily in session 2.

On day 2, we began session 2 with SRT tests. We expected that the intelligibility of

the cued sentences would be influenced by the fact that participants had seen the sentences the

day before, when they generated cues. Accordingly, we could not use the 29%, 50%, and 71%

SRTs estimated on the previous day, but estimated them again using 78 of the 378 sentences

presented in the cue generation test on the previous day ('familiar sentences'). The SNRs

corresponding to these thresholds were applied in the cued intelligibility test. In that test, 300

familiar sentences were preceded by the text cues. There were five types of cue: *matching*

own, *mismatching own*, *matching alien*, *mismatching alien* and *non-word,* each applied to 20 sentences at each of the three SNRs.

We could examine the effect of hearing familiar sentences on SRTs by comparing the SRT thresholds estimated for novel items in session 1 and for familiar items in session 2. This effect indicates to what degree having seen the sentences previously and having generated cues for each sentence influenced perception in noise.

In both sessions, the order of SRT tests, and the assignment of sentence lists to SRT tests and the cued intelligibility test were counterbalanced across participants. Session 1 was 2.5 hours long and session 2 was 1.5 hours long. These long and demanding test sessions prevented us from also assessing cognitive skills in Exp 1 (see Exp 2).

Participants

Twenty-nine normally hearing students (21 women and 8 men) participated. Their ages ranged from 18 to 30 yrs, with a mean age of 22 yrs (*SD* = 3.3 yrs). They were recruited among students of the Vrije Universiteit (VU University) and Hogeschool InHolland (InHolland University of Applied Sciences), Amsterdam. All participants were native Dutch speakers who reported no hearing problems, normal or corrected to normal vision, no dyslexia, and no history of neurological disease. All participants provided written informed consent in accordance with the Ethics Committee of the VU University Medical Center.

Day 1 Test Session

Near vision screening

Near vision was screened with a visual acuity Snellen equivalent (Bailey & Lovie, 1980). A word-chart is digitally presented at 40 cm and participants are asked to read the words aloud. The words are presented with decreasing font-size on each line, ranging from 80

to 2 points. The smallest font size the participant was able to read had to be equal to or below

eight for inclusion in the current study (i.e., at least 18 points below the smallest point size

used in the tests).

Pure-tone audiometry

Pure-tone hearing thresholds of the participants were measured at the start of the first

test session to ensure that the thresholds of both ears were ≤ 20 dB HL at the octave

frequencies between 125 and 8000 Hz.

Speech Reception Threshold test

The SRT technique was developed by Plomp & Mimpen (1979), but we used

sentences from the large set of 507 sentences recorded by a female speaker developed by

Versfeld et al. (2000). In each SRT test, 26 short, everyday Dutch sentences (Versfeld et al.,

2000) were presented in background noise. The sentences contained eight to nine syllables

each and no word in any sentence contained more than three syllables. The grammatical

structure of the sentences was relatively simple and similar across sentences. Sentences from

this set are pronounced clearly, at a natural rate. An example sentence (translated into

English) is "The shop is within walking distance". The sentences are equalized regarding

intelligibility, and are syntactically and psycholinguistically standardized (see Versfeld et al.,

2000, for more details). Stationary noise with the long-term average spectrum of the speech

served as the masker. Noise-onset was 500 msec before speech onset and noise-offset was

500 msec after speech offset (Versfeld et al., 2000). Participants were asked to report each

sentence aloud. The experimenter scored whether the sentence was reproduced completely

correctly (sentence-scoring). The SNR was varied by adapting both the speech and the noise

levels while keeping the long-term equivalent intensity fixed at 70 dB SPL. The first sentence

was presented at a SNR below threshold (-12 dB SNR) and was presented repeatedly while increasing the SNR with 4 dB steps until the participant reproduced the sentence completely correctly. Each of the other sentences was presented once at an SNR determined by the adaptive procedure. Three adaptive procedures (i.e., for the three different thresholds) were applied: a 1-up-2-down procedure estimating the SNR required for perceiving 29% of the sentences entirely correctly ($SRT_{29\%}$), and 1-up-1-down and 2-up-1-down procedures for estimating the SNR required for repeating 50% or 71% of the sentences entirely correctly, respectively ($SRT_{50\%}$ and $SRT_{71\%}$; Levitt, 1971). These SRTs covered a substantial range in speech perception performance. In the 1-up-1-down procedure, the SNR of each sentence was reduced by 2 dB if the preceding sentence was reproduced correctly and increased by 2 dB if it was reproduced incorrectly. In the 1-up-2-down procedure, participants had to reproduce two consecutive sentences incorrectly before the SNR of the following sentence was increased by 2 dB. In the 2-up-1-down procedure, participants had to correctly report two consecutive sentences for a 2 dB reduction in SNR. In all conditions, the SRT was the mean SNR of sentences 5 to 27; the 27th sentence was not actually presented, but the SNR at which this sentence would have been presented can be calculated from the response to the 26th sentence. The data of the first four sentences were not included in the analyses, as these sentences are used to start the staircase adaptive procedure by approaching the individuals' threshold (Plomp & Mimpen, 1979).

Cue generation test

In the cue-generation phase, 378 sentences were visually presented for 10 seconds each and participants were asked to generate three words that they thought meaningfully summarized the topic of each (Mäntylä & Nilsson, 1983). These words were used as cues in the cued speech perception test on the following day. Words that occurred in the sentence,

parts of compound words occurring in the sentence, or morphological variants of words in the sentence (e.g., the plural form of a noun, or the same verb in another tense) were not valid. Participants responded verbally; the experimenter entered the words in a database. Participants always generated 3 words. The cue words did not form a phrase but were individual words related to the sentence. For example, for the sentence "It smells nice in the bakery", cue words could be "bread-oven-cakes", and for the sentence "I could see everything very clearly", the cue words could be "glasses-close-light".

Day 2 Test Session

The second session took place at least 19 and no more than 31 hours following the first session.

Speech Reception Threshold Test

As on the previous day, 3 SRT tests were performed. However, this time, the SRT sentences were all familiar, having been seen on the previous day, and a three-word cue generated to each. SNRs sufficient for repeating 29%, 50% and 71% of the sentences entirely correctly ($SNR_{29\%}$, $SNR_{50\%}$, $SNR_{71\%}$) were once again estimated.

Cued speech perception test

In this test, participants heard 300 sentences that had been presented on the previous day. Sixty of these were presented in each of five cue conditions, 20 at each SNR. *Own mismatching* cues were selected from the participants' own set of cues that did not overlap with the *own*, *matching* cues for that particular sentence. *Alien matching* and *mismatching* cues for a given participant and sentence were selected at random from 3-word cues generated by one of the other previously tested participants, with the restriction that the cue did not

contain any words that were also generated by the target participant him/herself for any sentence. In order to enable the presentation of *alien* cues for the first test subjects, we collected cues from 11 persons before we started Exp 1. The *non-word* cues were created by shuffling the letters of a large set of Dutch words that were randomly selected from the CELEX database (Baayen et al., 1993). We only used non-words that were not pronounceable, did not contain meaningful Dutch words of more than two characters, and had the same distribution of word-lengths as the cues generated by the first 11 participants who completed the cue-generation test. For each participant, and for each sentence in the *non-word* conditions, we randomly selected, without replacement, a 3-non-word cue from the resulting list of 291 3-nonword cues.

On each trial, the words were first presented visually for 2 seconds on a monitor. After the cue disappeared from the screen, the auditory sentence in noise was presented. Participants were asked to repeat each sentence aloud, and were encouraged to make their best guess for sentences they were not able to understand entirely. They were told that the visual cues might or might not help them to understand the sentence. The experimenter recorded the participant's verbal response and whether the sentence was completely correct or not. No feedback was given during the tests. The SNR and cue-type condition were varied unpredictably from trial to trial, with no more than two successive sentences from the same condition.

Statistical analyses

We used repeated measures analysis of variance (2 factors; SNR (3 levels) and cue-type (5 levels)) to examine main effects and interactions. We performed the analysis on both the sentence scores (i.e., the percentage of sentences reported completely correctly) and on the percentage of correctly reported words. We used both word and sentence report scores

because, although the standard procedure in the SRT test is sentence scoring (Plomp & Mimpen, 1979; Versfeld et al., 2000) the cues may have a larger influence on the percentage of correctly repeated words. Also, meaningful sentences consist of several independent parts (Boothroyd & Nittrouer, 1988). The sentences used in the current study consisted of 1.5 to 2.5 independent parts (Versfeld et al., 2000). Word scoring is based upon a larger number of independent test items, resulting in smaller standard error of the scores, if the between sentence-variance is similar for both word and sentence scoring methods. This may also lead to differences between the word and sentence outcome measures.

Results Experiment 1

Pure-tone audiogram

The mean pure-tone hearing threshold at 500, 1000, 2000, and 4000 kHz, averaged over both ears, was 4.2 dB HL ($SD = 3.3$ dB).

SRT tests

Figure 1 shows the mean and standard deviation of SRT thresholds measured in session 1 (novel) and session 2 (familiar).

A repeated-measures analysis of variance testing the main and interaction effects of *SRT condition* ($SRT_{29\%}$, $SRT_{50\%}$, $SRT_{71\%}$) and *familiarity* (*unfamiliar* versus *familiar*) indicated main effects of both factors on the SRT thresholds: *SRT condition*: $F_{(2,56)} = 138$; $p < .001$ and *familiarity*: $F_{(1, 28)} = 53.1$, $p < .001$. SRTs were lower (better) by around 1 dB when participants heard familiar sentences for which they had generated cues on the preceding day (see Figure 1). Figure 1 shows that the change in accuracy in the percentage correctly reported sentences as a function of SNR was around 14 % / dB, for both familiar and unfamiliar

sentences, which is close to the steepness of the psychometric function observed by Versfeld
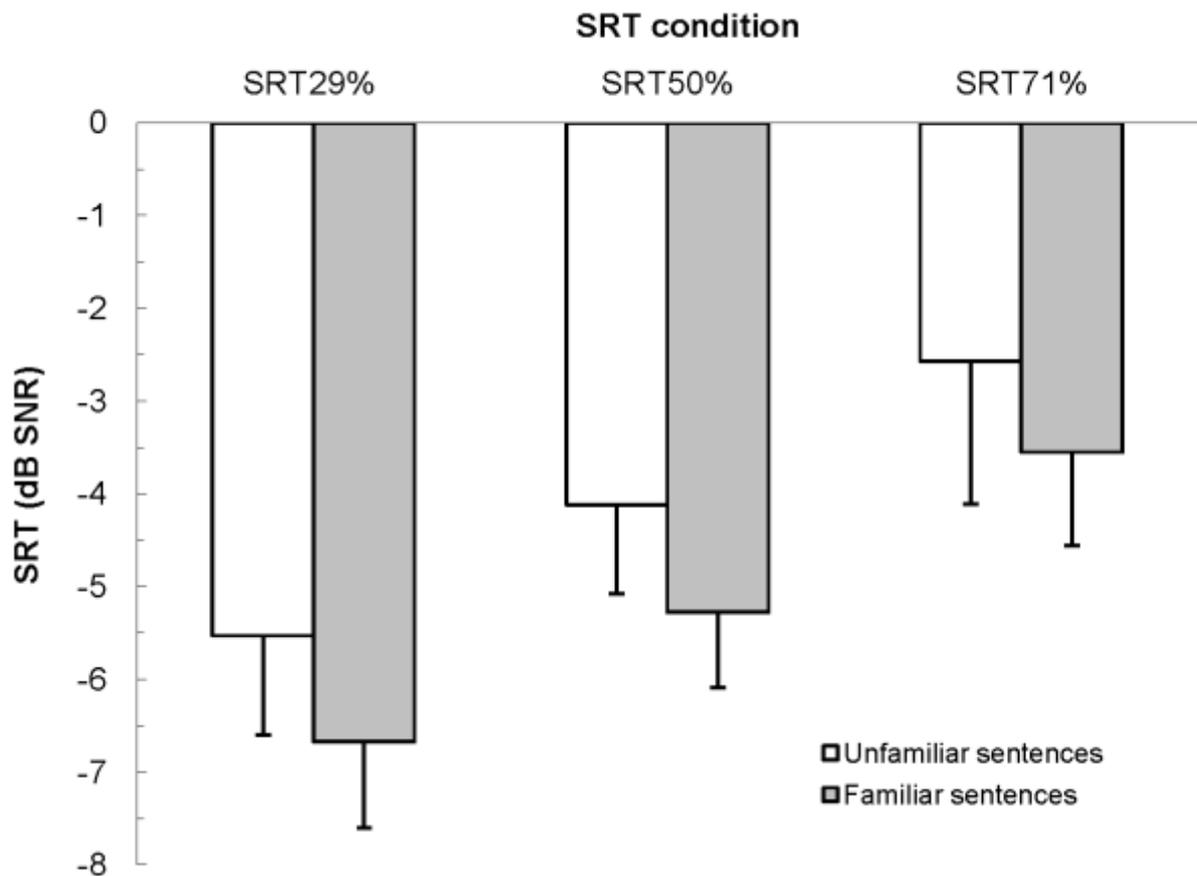
et al. (2000), which was 16.6 % / dB.

Figure 1. Means and standard deviations (error bars) of the Speech Reception Thresholds (SRTs) in the $SRT_{29\%}$, $SRT_{50\%}$, and $SRT_{71\%}$ conditions. Familiar sentences were visually presented on the preceding day in the cue-generation test. SNR = Speech-to-Noise Ratio.

No significant ($p > .05$) interaction between the factors was observed. Pairwise

comparisons with Bonferroni correction were use to evaluate the main effect of SRT

condition on SNR: only adjacent SRTs were tested (i.e., $SRT_{29\%}$ versus $SRT_{50\%}$ and $SRT_{50\%}$

versus $SRT_{71\%}$). Both contrasts were statistically significant ($t_{(28)} = 9.1$, $p < .001$ and $t_{(28)} =$

8.5, $p < .001$, respectively).

Cued speech perception test

Data were scored both as proportion of words correct, and proportion of sentences

correct. Figure 2a shows mean word-scores and Figure 2b show mean sentence-scores as a

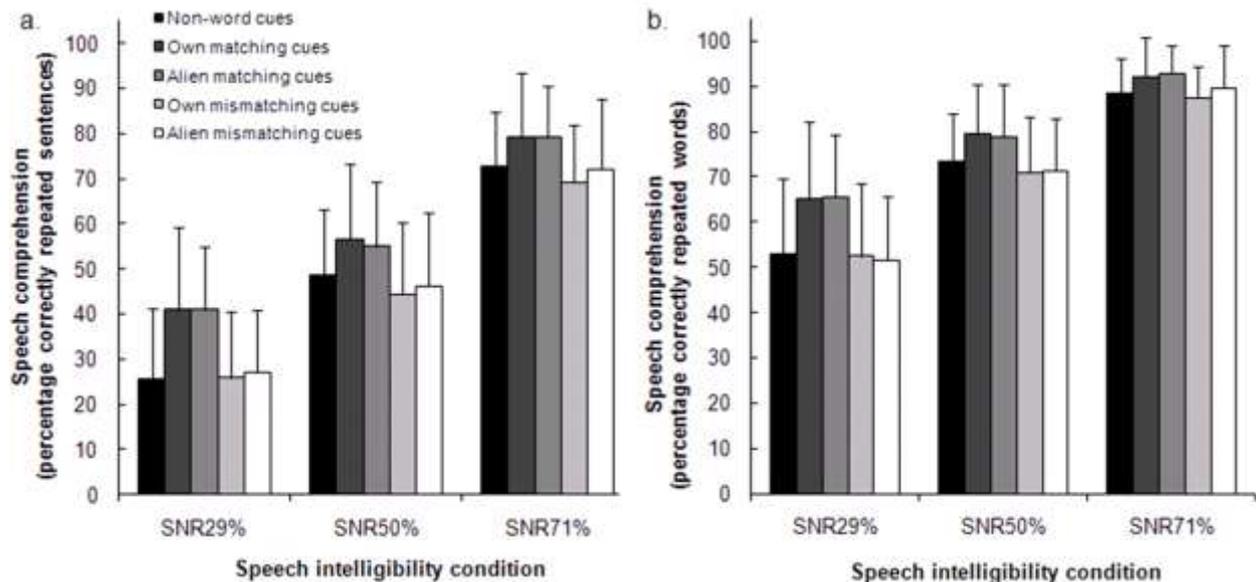function of cue condition and SNR.

Figure 2. Means and standard deviations (error bars) of the percentage of sentences that were entirely correctly reproduced (Figure 2a) and the percentage correctly repeated words (Figure 2b) in each of the 15 conditions of the cued speech-perception test. SNR = Speech-to-Noise Ratio; these levels corresponded to 29, 50, or 71% correct sentence-perception and were individually determined in the Speech-Reception-Threshold test.

The repeated-measures analyses showed an interaction between *SNR* and *cue-type* on

the word-scores ($F_{(8,224)} = 4.1$; $p < .001$), but not on the sentence-scores ($p > .05$). The

interaction was probably due to larger benefit obtained from the *matching* cues at lower SNRs

(see Fig. 2). To examine the origin of the interaction effect of *SNR* and *cue-type* on the word-

scores, we tested the performance-differences between the *non-word* cues (baseline) and each

of the four other cue-types, at each of the three SNRs using Bonferroni-corrected paired *t*-

tests. The results showed that word perception performance was significantly better for both

own and *alien matching* cues at the SNR$_{29\%}$ level (*own*: $t(28) = 4.72$, $p < .01$; *alien*: $t(28) =$

5.15, $p < .01$) and at the SNR$_{71\%}$ level (*own*: $t(28) = 3.67$, $p < .05$; *alien*: $t(28) = 3.61$, $p < .05$).

Additionally, *own matching* cues supported perceiving the words at the $SNR_{50\%}$ level ($t(28) =$ 3.18, *p* < .05) as compared to the *non-word* baseline. Additionally, we observed main effects of *SNR* and *cue-type*, for both the word and sentence scores (words: $F_{(2,56)} = 89.6$; *p* < .001 and $F_{(4,112)} = 26.0$, *p* < .001; sentences: $F_{(2,56)} = 107,4$; *p* < .001 and $F_{(4,112)} = 19.4$; *p* < .001, respectively). Figure 2 suggests that the main effect of cue-type was mainly based on facilitative effects of *matching* cues, both *own* and *alien*. The percentage correctly repeated words was between 4% − 13% higher in the matching-cue compared to the nonword-cue conditions; this corresponded to a benefit of 7% - 15% using sentence scoring.

The main effect of *SNR* indicates that participants were better able to perceive the speech when the SNR of the speech in noise was higher. Finally, the results of separately performed pairwise *t*-tests indicated no statistically significant differences (Bonferroni corrected *p* > .05) between the effects of *own* and *alien* cues on both word and sentence perception.

Note that disadvantages of using proportional scales like percent correct scores are that the variances are correlated to the means, the data are not normally distributed, and the scale values are not linear in relation to test variability (Studebaker, 1985). To examine whether the distribution of the scores influenced the results, the same analyses were performed on the rationalized arcsine-transformed (Studebaker, 1985) scores. The same significant interaction and main effects were observed: an interaction effect *cue-type* x *SNR* on the word scores: $F_{(8,224)} = 2.32$; *p* < .05; and main effects of *SNR* and *cue-type*, for both the word and sentence scores (words: $F_{(2,56)} = 105.5$; *p* < .001 and $F_{(4,112)} = 28.0$, *p* < .001; sentences: $F_{(2,56)} = 110.6$; *p* < .001 and $F_{(4,112)} = 19.5$; *p* < .001, respectively).

Discussion Experiment 1

Not surprisingly, SRT thresholds were lower in session 2, due to participants having seen the sentences and generated cues for them the day before. More interestingly, listeners benefited from semantically related contextual cues when perceiving degraded sentences across intelligibility levels. The results of Exp 1 were surprising in two respects. First, 'alien' matching cues provided just as much benefit to intelligibility as 'own' matching cues. Second, the data did not provide evidence of a negative effect of *mismatching* cues. This was surprising since inappropriate context should impair performance to the extent that a listener attempts to use it.

Experiment 2

Given the absence of differences in efficacy between *alien* and *own* cues in Exp 1, and given that it is more realistic to assess the intelligibility of novel sentences than of familiar ones, Exp 2 involved the assessment of cued speech intelligibility using novel materials in a single session. We used the most commonly generated cues from Exp 1 as *matching* and *mismatching* cues. We hypothesized that the most-frequently reported cues would provide a relatively general and strong semantic context. Secondly, Exp 1 showed that the benefit obtained from matching cues diminished at more advantageous SNRs. Accordingly, we used the $SNR_{29\%}$ level again, as well as an even less intelligible one resulting in 16% entirely correct repetition of uncued sentences in noise. Finally, since the association between language understanding and cognitive abilities including working memory capacity appears to increase in degraded listening conditions (Rönnberg et al., 2008) we also examined performance on tests of reading span (Daneman & Carpenter, 1980) and masked reading thresholds (Zekveld et al, 2007) and related these to the speech perception performance.

Sarampalis et al. (2009) and McCoy et al. (2005) showed that semantically related context provided during speech perception improves later recall, suggesting that contextual support leaves more resources available for encoding of the words in memory (McCoy et al., 2005). Daneman & Carpenter (1983) demonstrated that high working memory capacity is associated with a better ability to cope with misleading context, and better ability to draw inferences from text. This is consistent with the Ease of Language Understanding model, which predicts that the relevance of working memory processes increases under degraded conditions. We therefore expected that working memory capacity (as assessed using the reading span test; Daneman & Carpenter, 1980; Rönnberg et al., 1989) and good ability to read partially masked text (as assessed using the text reception threshold test, Zekveld et al., 2007) would be related to speech perception, specifically when irrelevant context is presented.

<div align="center">Methods Experiment 2</div>

In Exp 2, naive participants were tested in one session. We used the most frequently reported cues for each of the sentences as generated by the participants of Exp 1 as *matching* or *mismatching* cues. As in Exp 1, we also included *non-word* cues. The methods of Exp 2 were the same as those applied in Exp 1, except for the adaptations described in detail below.

Participants

Twenty normal-hearing subjects (13 women and 7 men) participated. Their ages ranged from 18 to 32 yrs, with a mean age of 22 yrs ($SD = 3.4$ yrs). They had not participated in Exp 1. The same inclusion criteria as in Exp 1 were applied. All participants provided written informed consent in accordance with the Ethics Committee of the VU University Medical Center.

Procedure and Tests

SRT test

After vision and hearing screening (Snellen equivalent test of visual acuity, and pure-tone audiogram, respectively), participants performed a 26-item practice SRT test, and then the $SRT_{16\%}$ and $SRT_{29\%}$ thresholds were determined. In the $SRT_{16\%}$ test, a 1-up-4-down procedure estimating the SNR required for the entirely correct repetition of 16% of the sentences was applied ($SNR_{16\%}$). Participants had to reproduce four subsequent sentences incorrectly before the SNR of the following sentence was increased by +2 dB. Materials and test order were counterbalanced across conditions and participants.

Text reception threshold test

After the SRT tests, participants performed one practice and three consecutive text reception threshold tests. In the visual text reception threshold test, which is conceptually equivalent to the SRT test, partly masked written sentences (Versfeld et al., 2000) are presented (Zekveld et al., 2007), and the ability to report sentences as a function of degree of masking is determined.

The sentences used for the text reception threshold tests were different than those presented in the other tests. Text was masked with a bar pattern. A bitmap image was created that was proportionally filled with black bars, depending on the required percentage of unmasked text. For details, see Zekveld et al. (2007). The field background color was white, text color was red, and the color of the mask was black. At the start of each trial, the mask became visible and the text appeared 'behind' it in a word-by-word fashion. Display-onset of each word in the sentence was equal to the timing of the start of the utterance of each word in the corresponding audio file (Versfeld et al., 2000). All words remained on the screen for 3500 ms after the sentence was complete. A 1-up-1-down adaptive procedure was applied,

targeting the percentage of unmasked text required to read 50% of the sentences entirely

correctly. The step-size applied in the test was a 6% change of the percentage unmasked text.

The text reception threshold was the mean percentage of unmasked text of sentence 5 to 27.

We conducted this test three times, and used the average threshold in the analysis.

Reading Span test

      Next, participants performed the Reading Span Test (Daneman & Carpenter, 1980). In

this test, short (5-word) sentences are presented visually. Half of the sentences are nonsense

(e.g., the table sings a song); the other half are coherent (e.g., the friend told a story). The

Dutch sentence material and test were developed (see Besser et al., 2011) to be equivalent to

the Swedish version described by Andersson et al. (2001; see also Rönnberg et al., 1989).

First, three sets of three sentences are presented, followed by three sets of four sentences,

three sets of five sentences, and three sets of six sentences. Immediately after each sentence,

participants verbally indicate whether the sentence makes sense or not. After each set of

sentences, participants are prompted to recall, in serial order, either the first or the last word

of each sentence. The experimenter records the number of words correctly recalled, regardless

of order.

Cued Speech Perception test

      Finally, subjects performed the cued speech reception threshold test. Sixty sentences

were presented at each of the two SNRs, ($SRT_{16\%}$ and $SRT_{29\%}$), 20 with each type of cue. The

SNR and cue-type condition were varied unpredictably from trial to trial, with no more than

two successive sentences from the same condition.

      Matching cues in Exp 2 were created by taking, for each sentence, the three most

frequently reported cue words generated in Exp 1. (i.e., from the 11 participants who

generated cues prior to Exp 1 and the first 20 participants tested in Exp 1). First, we selected

those 180 sentences that yielded the most consistently reported cues, determined by counting

the number of times participants generated the same individual words for a particular

sentence. As a measure of consistency, we calculated the sum of the frequencies for the three

most frequently generated words for each sentence. The theoretical maximum would be 93 if

each word was reported by all 31 subjects in Exp 1. In fact, this frequency sum ranged from

20 to 46 for the sentences selected in Exp 2. The newly created "most-frequent" three-word

cues were new combinations of those 3 individual words that were most often generated for

each sentence. We did not expect that creating new combinations of cues would affect the

effect of the cues on speech intelligibility. The selected sentences were grouped into nine lists

of 20 sentences. Lists were created such that the frequency-sums for matching cues were, on

average, similar for each list. Only 60 of these sentences were presented with their matching

cues; 60 were presented with nonword cues, and 60 with mismatching cues (these were the

matching cues from the 60 'left over' sentences out of the total set of 180). The allocation of

particular sentences to conditions was counterbalanced across subjects. The duration of the

entire test session was 1.5 hours.

Statistical analyses

We used repeated measures analyses of variance to test the effects of *cue-type* (3

levels) and *SNR* (2 levels) on the sentence and word scores. Spearman correlation coefficients

were calculated to examine the relation between performance on the text reception threshold

and reading span tests and the effect of cue type on speech-perception performance (both

word and sentence-scores). Additionally, we calculated the correlation coefficient between the

text reception threshold and reading span scores and the difference between the scores in the

non-word cue condition and the *match* and *mismatching* cue condition, separately for both

SNRs. This indicates whether cognitive abilities are related to interindividual differences in the facilitating or interfering effects of the cues. We report non-parametric Spearman correlation coefficients as the sample size was relatively small (n = 20).

Results

Pure-tone audiogram

The mean pure-tone hearing threshold at 500, 1000, 2000, and 4000 kHz, averaged over both ears, was 4.3 dB HL (SD = 4.0 dB).

SRT tests

The mean $SRT_{16\%}$ was -6.7 dB SNR (SD = .9 dB SNR) and the mean $SRT_{29\%}$ was -5.6 dB SNR (SD = .8 dB SNR). A paired samples t-test indicated that the two SRT thresholds differed significantly from each other ($t(19)$= -5.5, $p < .001$).

Cognitive tests: Text reception threshold and reading span

The mean text reception threshold (the mean percentage of unmasked text required to read 50% of the sentences entirely correctly) was 54.9%, SD = 2.6%, range 49-59%. The mean reading span was 26.1, SD = 5.7, range 16-40.

Cued speech perception test

Figures 3a and 3b show the means and standard deviations of the sentence and word-scores (percentage correct) in each of the six conditions of the cued-speech perception test.
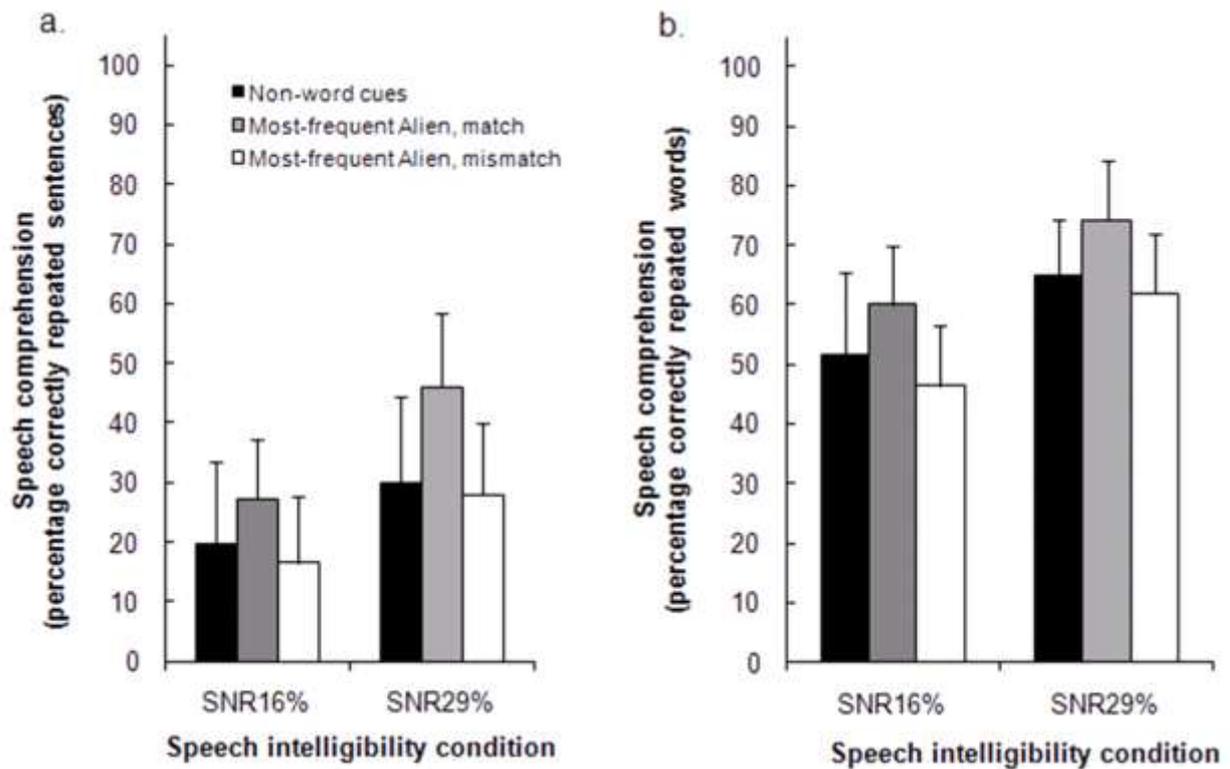
Figure 3. As for Figure 2, but for six conditions and at Speech-to-Noise Ratios (SNRs) corresponding to 16 or 29% correct sentence-perception.

For both the word and the sentence scores, the interaction between *cue type* and *SNR* was not statistically significant. However, a trend was observed for the sentence scores ($F_{(2,38)}$ = 2.93, *p* = .067). We did observe statistically significant main effects of *SNR* for both the word and sentence scores (words: $F_{(1,19)}$ = 20.0; *p* < .001 and sentences: $F_{(1,19)}$ = 18.0; *p* < .001). A main effect of *cue-type* was observed as well (words: $F_{(2,38)}$ = 26.2, *p* < .001 and sentences: $F_{(2,38)}$ = 21.0; *p* < .001). The same pattern of results was observed for the analyses based on the rationalized-arcsine-transformed scores: statistically significant main effects of *SNR* and *cue-type* for the word and sentence scores (words: $F_{(1,19)}$ = 19.6; *p* < .001 and $F_{(2,38)}$ = 27.4, *p* < .001; sentences: $F_{(1,19)}$ = 18.9; *p* < .001 and $F_{(2,38)}$ = 19.9; *p* < .001, respectively). We used Bonferroni-corrected simple contrasts to examine the main effect of cue type on the percentage correctly perceived words and sentences. Compared to the *non-word* cues, word and sentence-perception was significantly higher for *matching cues* (*t*(19) = 4.5, *p* < .001 and

$t(19) = 4.7$, $p < .001$, respectively). Only for the word-scores was performance significantly

lower when *mismatching* cues were presented as compared to the *non-word* baseline ($t(19) =$

2.8, $p < .05$). Thus, at these rather low SNRs, word and sentence perception increased when

semantically related cues were presented, whereas semantically unrelated cues reduced the

percentage of correctly repeated words (but not the percentage of entirely correctly repeated

sentences).

Correlation analysis

Reading span and text reception threshold were both significantly associated with the

percentage of entirely correctly reproduced sentences when mismatching cues were presented

at the $SNR_{29\%}$ level (Spearman $r = .60$, $p < .01$ and $r = -.50$, $p < .05$, respectively). This

indicates that both working memory capacity and the ability to read partly masked text are

associated with better sentence perception when unrelated semantic context is presented. We

did not observe statistically significant associations between the percentage of correctly

repeated words and performance on cognitive tests at this SNR. At $SNR_{16\%}$ we did not

observe significant associations between the percentage of correctly repeated words or

sentences and performance on cognitive tests.

Discussion Experiment 2

Semantically appropriate cues facilitated speech perception in noise, even when

listeners were unfamiliar with the sentences. At SNRs corresponding to 16% and 29% SRTs,

matching cues aided both sentence and word perception. *Mismatching* cues impaired word

perception as compared to the *non-word* baseline. At these relatively low SRTs, perception

was apparently influenced by the available (relevant and irrelevant) context. At the $SNR_{29\%}$

level, sentence perception was better for listeners with better text reception thresholds and

better reading span performances (larger working memory capacity) when irrelevant semantic context was provided. These skills may be related to an ability to ignore the irrelevant semantic cues.

General Discussion

The main result of the present study is that visually presented, semantically matching and mismatching 3-word cues influence speech perception in noise. Consistent with other studies (e.g., Obleser et al., 2007; Grant & Seitz, 2000; Pichora-Fuller et al., 1995), congruent semantic context (i.e., the *matching* cues) consistently supported speech perception. Importantly, in the current study, the quality of the visually presented semantic context was not affected by changing the signal-to-noise ratio of the sentences in noise. Hence, in contrast to studies that examined the effect of sentence context on the perception of the sentence-final word, the visual semantic context was always clearly presented, and perception of this context was not confounded by the speech intelligibility manipulations.

At the low SNRs and with unfamiliar sentences (Exp 2), cues with unrelated semantic content impaired intelligibility as measured by word repetition. Why such an effect was noted in Exp 2, but not in Exp 1, is uncertain. It could be the lower SNRs, the novelty of the sentences, a combination of both of these, or some other factor that revealed a deleterious effect of irrelevant context on speech intelligibility in Exp 2.

As expected, and consistent with Daneman & Carpenter (1983) and the predictions derived from the Ease of Language Understanding framework (Rönnberg, 2003), higher working memory capacity and better ability to read partly masked sentences were associated with an ability to cope with mismatching context, although this was only evident at the $SNR_{29\%}$ level and for the percentage correctly repeated sentences (not the percentage of repeated words). The results partly support the idea that working memory and inference

making are involved in coping with misleading context when listeners are trying to understand degraded speech. Other studies have consistently shown that working memory is associated with the ability to inhibit irrelevant information (Lunner, 2003; Conway et al., 2001; Rosen & Engle, 1998; Kane & Engle, 2000). Importantly, the current data suggest that large working memory capacity and good linguistic skills are not related to an overall increase in the use of the contextual cues: if that were the case, these abilities would have been related to *worse* performance in *mismatch* conditions, and better performance in the *match* conditions. It would be interesting to further examine which conditions make speech perception in noise more dependent on inference-making and working memory.

We analysed two types of scores: the percentage of correctly repeated words and the percentage of entirely correctly repeated sentences. The rationale for analyzing both scores was that the semantic cues might have different effects on both performance measures. Indeed, we only observed an interaction effect between cue-type and SNR in the word-scores (Exp 1), and only observed an effect of the misleading semantic cues on word-scores (Exp 2). However, the relation between working memory capacity, the ability to read partly masked text and cued speech perception was only observed for the sentence scores. The different outcomes for the two scoring methods suggest that, although the scores are highly correlated, the number of words and sentences that are correctly repeated may reflect potentially differentiable processes. These may reflect different strategies aimed either at repeating sentences a word at a time, or holistically as a single unit. Possibly, listeners differ in the use of these strategies, as suggested by the present relation between the sentence scores and cognitive skills. However, the differences between the two scoring methods could also be based upon smaller standard errors of the mean word scores as compared to the sentence scores, due to the larger number of independent test items for the word scoring method. On

average, standard errors were at least 1.2 times smaller for the word scores, which is

consistent with the reduction expected based on 1.5 to 2.5 independent parts in each sentence.

Both Exp 1 and Exp 2 included $SNR_{29\%}$ conditions (on average, the SNR applied in

the cued speech perception test was -6.7 dB in Exp 1 and -5.6 dB in Exp 2). Rather

surprisingly, the mean word scores were greater in the nonword cue condition of Exp 2 (74%)

than in Exp 1 (52%), and these were significantly different as indicated by an independent-

samples t-test ($t_{(47)} = -2.92$, $p < .01$). The lower performance in Exp 1 could have been caused

by the longer test session and larger number of sentences in the cued speech perception test as

compared to Exp 2. Presumably, the relatively low performance is not related to sentence

familiarity in Exp 1, as we controlled for familiarity by estimating the SNRs required for the

entirely correct repetition of 29% of familiar sentences.

Despite differences between the two experiments in match cues, and in familiarity of

the sentences, match cues resulted in comparable facilitative effects. This is surprising, since

subjects of Exp 2 could not actively reconstruct the target sentence from memory as the

listeners in Exp 1 might have done. Furthermore, in Exp 1, we did not observe an effect of

whether or not participants had generated the cues themselves. Thus, neither stimulus

familiarity, nor the autogeneration of semantic associations is required for facilitative effects

of 3-word visual cues on the perception of speech in noise.

The current results are in contrast with the distinct effects of own and alien cues on

word recall from memory (Mäntylä, 1986a; Mäntylä & Nilsson, 1988). Mäntylä (1986b) has

additionally shown that "own" cues are more effective than "alien" cues in sentence recall,

suggesting that the general absence of an effect of cue "ownership" as observed in the current

study is likely related to task characteristics. Stimulus identification may be less dependent on

the more specific, private, stimulus associations than stimulus recall. Furthermore, no

"mismatching" cues can be provided for recall tasks. The presentation of mismatching cues in

the current study may have influenced performance, even in the matching cue conditions. In the current study, it was easy to detect that for some of the sentences, the cues were not related to the sentences. Most of the subjects commented afterwards that they detected the presence of mismatching cues after several sentences and thereafter mainly focused on the auditory information without relying too much on the cues. It may be interesting to assess individual differences in the degree to which individuals were able to disregard the cues.

To examine the extent to which the current data reflected processing of the cues, the processing of the auditory stimuli, and the interaction between the two, we examined the degree to which the responses were associated with the cues presented in the mismatch conditions. We reasoned that guessing based on mismatching cues would result in responses that were incorrect and semantically associated with the cues. In the present study, we recorded the complete response of the participants, including the incorrect words. The responses of the participants did not indicate large mismatch effects, supporting the idea that participants indeed attempted to report the sentences, without basing their responses solely on the visual cues.

The results of the current study show that providing semantic cues to the content of degraded speech influences intelligibility in noise. This is relevant for people coping with everyday hearing difficulties. If a speaker explicitly introduces a change of topic during a conversation, this is likely to improve the listener's comprehension. The present study shows that even short "introductions" can be effective. Furthermore, the matching cues had a very similar effect on speech intelligibility irrespective of whether cues were own or alien (Exp 1) or whether sentences were familiar or novel. This indicates that providing semantic context affects speech perception independent of the cue characteristics. On the other hand, the relatively large variability in the effect of the cues between listeners (see error bars in Figures 2 and 3), the association of working memory and linguistic skills with cued sentence

perception in one of the conditions, and the differences between the word and sentence-based scoring methods (see above) indicate that people may differ in the strategy they apply when semantic context is provided. It would be interesting to further assess these differences, as knowledge about such potential strategy differences may provide more insight into how listeners most successfully use contextual information. This would be relevant for the development of strategies aimed to reduce difficulties caused by hearing impairment.

In both Experiments, the benefit obtained from matching cues corresponded to an increase in the percentage of correctly repeated sentences of around 15% at most, which corresponds to an increase in SNR of around 1 dB. This is similar to the size of the observed effect of sentence familiarity on speech perception in Exp 1. A 1-dB improvement in SNR may be valuable in some everyday listening conditions. Nachtegaal et al. (2009) recently examined the relation between hearing status and the ability to recover from fatigue and distress at work. They observed that each dB improvement in SNR led to a 9% decrease in the odds of having a high need for recovery after work. Furthermore, the benefit of preceding semantic context may be underestimated here, since the effect of the matching cues may have been reduced by the inclusion of mismatching cues in the two experiments.

In conclusion, we observed facilitative effects of relevant semantic contextual information on the perception of speech in noise across SNR levels. Misleading context was associated with reduced word perception at low SNRs in Exp 2. At the SNR corresponding to 29% SRT, sentence perception preceded by irrelevant cues was associated with working memory processes and linguistic skills. This suggests that these abilities are relevant when perceiving speech in noise and may be related to the ability to ignore irrelevant semantic cues.

References

Akeroyd, M. A. (2008). Are individual differences in speech reception threshold related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *Int J Audiol, 47, Suppl. 2.,* S53-S71.

Andersson, U., Lyxell, B., Rönnberg, J., Spen, K-E. (2001). Cognitive correlates of visual speech understanding in hearing-impaired individuals. *J Deaf Stud Deaf Educ, 6,* 103-116.

Aydelott, J. & Bates, E. (2004). Effects of acoustic distortion and semantic context on lexical access. *Lang Cogn Process, 19,* 29-56.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX Lexical Database (Version Release 2) [CD-ROM]. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

Bäckman, L., Mäntylä, T., & Erngrund, K. (1984). Optimal recall in early and late adulthood. *Scand J Psychol, 25,* 306-314.

Bailey, I. L., & Lovie, J. E. (1980). The design and use of a near-vision chart. *Am J Optom Physiol Opt, 57,* 378-387.

Becker, C. A. & Killion, T. H. (1979). Interaction of visual and cognitive effects in word recognition. *J Exp Psychol Hum Percept Perform, 3,* 389-401.

Besser, J., Zekveld, A.A., Kramer, S.E., et al. (2011). New measures of masked text recognition in relation to speech-in-noise perception and their associations with age and cognitive abilities. J Speech Lang Hear Res, *in revision.*

Bilger, R. C., Neutzel, J. M., Rabinowitz, W. M., et al. (1984). Standardization of a test of speech perception in noise. *J Speech Hear Res,* 27, 32–48.

Boothroyd, A. & Nittrouer, S. (1988). Mathematical treatment of context effects in phoneme and word recognition. *J Acoust Soc Am, 84,* 101-114.

Bronkhorst, A. W., Bosman, A. J., & Smoorenburg, G. F. (1993). A model for context effects in speech recognition. *J Acoust Soc Am, 93,* 499-509.

Cardillo, E. R., Aydelott, J., Matthews, P. M., et al. (2004). Left inferior prefrontal cortex activity reflects inhibitory rather than facilitatory priming. *J Cogn Neurosci, 16,* 1552-1561.

Conway, A. R. A., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psych Bull Rev, 8,* 331-335.

Daneman, M. & Carpenter, P. A. (1980). Individual differences in working memory and reading. *J Verbal Learn Verbal Behav, 19,* 450-466.

Daneman, M., & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *J Exp Psychol Learn Mem Cogn, 9,* 561-583.

Drager, K. D. R. & Reichle, J.E. (2001). Effects of discourse context on the intelligibility of synthesized speech for young adult and older adult listeners: Applications for AAC. *J Speech Lang Hear Res, 44,* 1052-1057.

Dubno,J.R., Ahlstrom, J.B., & Horwitz, A.R. (2000). Use of context by young and aged adults with normal hearing. *J Acoust Soc Am, 107,* 538–546.

Elliot L. L. (1995). Verbal auditory closure and the Speech Perception in Noise (SPIN) test. *J Speech Hear Res, 38,* 1363–1376.

Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *J Acoust Soc Am, 115,* 2246-2256.

Foo C, Rudner, M, Rönnberg J, & Lunner T. (2007). Recognition of speech in noise with new hearing instrument compression release settings requires explicit cognitive storage and processing capacity. *J Am Acad Audiol, 18,* 553-66.

George, E.L.J., Zekveld, A.A., Kramer, S.E., et al., (2007). Auditory and nonauditory factors affecting speech reception in noise by older listeners. *J Acoust Soc Am, 121*, 2362-2375.

Grant, K. W., & Seitz, P. F. (2000). The recognition of isolated words and words in sentences: Individual variability in the use of sentence context. *J Acoust Soc Am, 107,* 1000-1009.

Hagoort, P. & van Berkum, J. (2007). Beyond the sentence given, *Philos Trans R Soc Lond B, 362*, 801-811.

Helfer, K.S. & Freyman, R.L. (2008). Aging and speech-on-speech masking, *Ear Hear, 29,* 87–98.

Holcomb, P.J. (1993). Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology, 30,* 47-61.

Kalikow, D. N., Stevens K. N., & Elliott, L.L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability, *J Acoust Soc Am, 61,* 1337–1349.

Kane, M. J., & Engle, R. W. (2000). Working memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *J Exp Psychol Learn Mem Cogn, 26,* 336-358.

Kramer, S.E., Zekveld, A.A., & Houtgast, T. (2009). Measuring cognitive factors in speech comprehension: The value of using the text reception threshold test as a visual equivalent of the SRT test. *Scand J Psychol, 50,* 507-515.

Lansing, C. R. & Helgeson, C. L. (1995). Priming the visual recognition of spoken words. *J Speech Hear Res, 38,* 1377-1386.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *J Acoust Soc Am, 49*, 467-477.

Lunner, T. (2003). Cognitive function in relation to hearing aid use. *Int J Audiol, 42,* S49-S58.

MacDonald, H., Davis, M.H., Pichora-Fuller, K. et al. (2008). Contextual influences: Perception of sentences in noise is facilitated similarly in young and older listeners by meaningful semantic context; neural correlates explored via functional magnetic resonance imaging (fMRI). *J Acoust Soc Am, 123,* 3887.

Mäntylä, T. (1986a). Optimizing cue effectiveness: Recall of 500 and 600 incidentially learned word. *J Exp Psychol Learn Mem Cogn, 12,* 66-71.

Mäntylä, T. (1986b). How do you cue: Self-generated retrieval cues and successful recall. Dissertation, University of Umeå, Umeå, Sweden.

Mäntylä, T. & Nilsson, L-G. (1983). Are my cues better than your cues? Uniqueness and reconstruction as prerequisites for optimal recall of verbal materials. *Scand J Psychol, 24,* 303-312.

Mäntylä, T. & Nilsson, L-G. (1988). Cue distinctiveness and forgetting: Effectiveness of self-generated retrieval cues in delayed recall. *J Exp Psychol Learn Mem Cogn, 14,* 502-509.

McCoy, S.L., Tun, P.A., Cox, L.C., et al. (2005). Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech. *Q J Exp Psychol, 58,* 22-33.

Meyer, D.E., Schvaneveldt, R.W. & Ruddy, M.G. (1975). Loci of contextual effects on word recognition. In: Rabbitt, P.M.A. and Dornic, S., Eds, *Attention and Performance* Vol. 5**,** New York: Academic Press.

Miller, G.A., Heise, G.A., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *J Exp Psychol, 41,* 329-335.

Miller, G.A. & Isard, S. (1963). Some perceptual consequences of linguistic rules. *J Verb Learn Verb Behav, 2,* 217-228.Moll., K., Cardillo, E., & Aydelott-Utman, J. (2001). Effects of competing speech on sentence-word priming: Semantic, perceptual, and attentional factors. In J. D.Moore & K. Stenning (Eds.), *Proceedings of the Twenty-*

third Annual Conference of the Cognitive Science Society (pp. 651–656). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Nachtegaal, J., Kuik, D.J., Anema, J.R., et al. (2009). Hearing status, need for recovery after work, and psychosocial work characteristics: Results from an Internet based National Survey on Hearing. *Int J Audiol, 48,* 684-691.

Nakano, H., Saron, C., & Swaab, T.Y. (2010). Speech and span: Working memory capacity impacts the use of animacy but not of world knowledge during spoken sentence comprehension. *J Cogn Neurosci, 22,* 2886-2898.

Neely, J. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264–336). Hillsdale, NJ: Erlbaum.

Obleser, J., & Kotz, S. A. (2010). Expectancy constraints in degraded speech modulate the language comprehension network. *Cereb Cortex, 20*, 633-640.

Obleser, J., Wise, R. J. S., Dresner, et al., (2007). Functional integration across brain regions improves speech perception under adverse listening conditions. *J Neurosci, 27,* 2283-2289.

Pichora-Fuller, M. K., Schneider, B. A., Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *J Acoust Soc Am, 97,* 593-608.

Plomp, R. (2002). *The Intelligent Ear.* Mahwah, NJ: Lawrence Erlbaum Associates.

Plomp, R., & Mimpen, A. M. (1979). Improving the reliability of testing the speech reception threshold for sentences. *Audiology, 18,* 43-52.

Rönnberg, J. (2003). Cognition in the hearing impaired and deaf as a bridge between signal and dialogue: a framework and a model. *Int J Audiol, 42,* S68-S76.

Rönnberg, J., Arlinger, S., Lyxell, B., et al. (1989). Visual evoked potentials: Relation to adult speechreading and cognitive function. *J Speech Hear Res, 32,* 725-735.

Rönnberg, J., Rudner, M., Foo, C., et al. (2008). Cognition counts: A working memory system for ease of language understanding (ELU). *Int J Audiol, 47, S2,* S99-S105.

Rönnberg, J., Rudner, M., Lunner, T., et al. (2010). When cognition kicks in: Working memory and speech understanding in noise. *Noise Health, 12,* 263-269.

van Rooij, J. C. G. M., & Plomp, R. (1991). The effects of linguistic entropy on speech perception in noise in young and elderly listeners. *J Am Soc Am, 90,* 2985–2991.

Rosen, V. M. & Engle, R. W. (1998). Working memory capacity and suppression. *J Mem Lang, 39,* 418-436.

Rudner M., Foo, C., Sundewall-Thorén, E., et al. (2008). Phonological mismatch and explicit cognitive processing in a sample of 102 hearing-aid users. *Int J Audiol, 47*, S91-S98.

Rudner, M., Foo, C., Rönnberg, J. et al. (2009). Cognition and aided speech recognition in noise: Specific role for cognitive factors following nine-week experience with adjusted compression settings in hearing aids. *Scand J Psychol, 50*, 405-418.

Rudner, M., Rönnberg. J. & Lunner, T. (2011). Working memory supports listening in noise for persons with hearing impairment. *J Am Acad Audiol, 22*, 1-12.

Samuelsson, S., & Rönnberg, J. (1991). Script activation in lipreading. *Scand J Psychol*, *32*, 124–143

Samuelsson, S. & Rönnberg, J. (1993). Implicit and explicit use if scripted constraints in lip-reading. *Eur J Cogn Psychol, 5,* 201-233.

Sarampalis, A., Kalluri, S., Edwards, B., et al. (2009). Objective measures of listening effort: Effects of background noise and noise reduction. *J Speech Lang Hear Res, 52, 1230-1240.*

Sheldon, S., Pichora-Fuller, M.K., & Schneider, B.A. (2008). Priming and sentence context support listening to noise-vocoded speech by younger and older adults. *J Acoust Soc Am, 123,* 489-499.

Stenfelt, S. & Rönnberg, J. (2009). The Signal-Cognition interface: Interactions between degraded auditory signals and cognitive processes. *Scan J Psychol*, *50,* 385-393.

Studebaker, G.A. (1985). A "rationalized" arcsine transform. *J Speech Hear Res, 28,* 455-462.

Tesink, C.M.J.Y., Magnus Petersson, K.M., Van Berkum, J.J.A., Van den Brink, D., Buitelaar, J.K., & Hagoort, P. (2009). Unification of speaker and meaning in language comprehension: An fMRI study. *J Cogn Neurosci, 21,* 2085-2099.

Tulving, E. (1983). *Elements of Episodic Memory*. Oxford: Oxford University Press.

van Rooij, J. C. G. M., & Plomp, R. (1991). The effects of linguistic entropy on speech perception in noise in young and elderly listeners, *J Acoust Soc Am, 90*, 2985–2991.

Versfeld, N.J., Daalder, L., Festen, J.M., et al. (2000). Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *J Acoust Soc Am, 107,* 1671-1684.

Wingfield, A. (1996). Cognitive factors in auditory performance: context, speed of processing, and constraints of memory. *J Am Acad Audiol, 7,* 175–182.

Wingfield, A. & Tun, P. A. (2007). Cognitive supports and cognitive constraints on comprehension of spoken language. *J Am Acad Audiol, 18,* 548-558.

Wingfield, A., Tun, P. A., & McCoy, S. L. (2005). Hearing loss in older adulthood. What it is and how it interacts with cognitive performance. *Curr Dir Psychol Sci, 14,* 144-148.

Zekveld, A. A., George, E. L. J., Kramer, S. E., et al. (2007). The development of the Text Reception Threshold test: A visual analogue of the Speech Reception threshold test. *J Speech Lang Hear Res, 50,* 576-584.

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2010). Pupil response as an indication of effortful listening: The influence of sentence intelligibility. *Ear Hear, 31,* 480-490.

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2011). Cognitive Load during Speech

Perception in Noise: The Influence of Age, Hearing Loss, and Cognition on the Pupil

Response. *Ear Hearing* (Epub ahead of print).

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2011). Cognitive Load during Speech

Perception in Noise: The Influence of Age, Hearing Loss, and Cognition on the Pupil

Response. *Ear Hearing* (Epub ahead of print).