The Integration of Semantic and Phonological Information in Normal, Degraded, and Time-Compressed Speech

Charlotte M. de Blecourt, s2188813

Behavioural and Cognitive Neuroscience Research Master, C-Track

19-11-2015

University Medical Center Groningen, Department of Otorhinolaryngology

Supervisor: dr. Anita Wagner

Second evaluator: prof. dr. Deniz Baskent

Abstract

Understanding speech is a continuous process, which involves integrating the acoustic input signal, the semantic context, and other previous knowledge. A distorted speech signal complicates this process. Contextual information can help building expectations about the upcoming signal; it helps to compensate for loss of acoustic information from distorted speech. However, how contextual information is integrated in a distorted signal is unknown.

The present study involves presenting normal, degraded (CI-simulated), and time-compressed sentences to Dutch participants with normal hearing. Their gaze fixations are captured using an eye-tracker. Listeners were presented with sentences with or without a semantically constraining verb before presenting the actual target. Participants selected one of four images on a computer screen, which contained the target, a phonological competitor, a semantic competitor, and a non-related distractor.

Interaction effects between speech type and semantic condition were observed in both degraded and time-compressed speech. In normal speech, the listeners' gaze reflected the usage of the semantic hints together with the speech signal. In degraded speech, processing and integration of preceding information was delayed, indicating slower facilitation of target words. Compared to normal speech, fewer target fixations were observed in time-compressed speech, especially when no context was present. Participants presented with time-compressed speech fail to inhibit semantically similar distractor words.

Table of Contents

Introduction

Spoken word processing

Understanding speech is based on integrating multiple sources of information. Multiple theories and models have been developed to explain how lexical (knowledge-derived) and sub-lexical (signal-derived) information is integrated (Dahan & Magnuson, 2006; Gaskell & Marslen-Wilson, 1997). Currently, most theories agree that phonological signal and semantic context both drive word recognition (Chen & Mirman, 2012; Dahan & Magnuson, 2006; Dahan, Tanenhaus, & Salverda, 2007; Marslen-Wilson, 1987; Mattys, White, & Melhorn, 2005), but it is unclear if this process is continuous or modular (Apfelbaum, Blumstein, & McMurray, 2011; Chen & Mirman, 2012).

Spoken word-recognition is often broken down into three steps: *access*, *selection* and *integration* (Dahan & Tanenhaus, 2004; Gaskell & Marslen-Wilson, 1997; Marslen-Wilson, 1987). The first step shows up in literature under different names, such as *access* (Marslen-Wilson, 1987) or *activation* (Aydelott & Bates, 2004). This concerns "getting access to" or "activation of" one's lexicon through mapping of speech input onto representations in the lexicon. Because access aids later word-recognition processes, it is regarded as a mainly a facilitatory process (Andruski, Blumstein, & Burton, 1994; Neely, 1991). It acts relatively fast and early in the speech recognition process, and it is not very demanding in terms of attentional resources (Norris, Cutler, McQueen, & Butterfield, 2006). *Selection* involves selecting the best fitting word, after which the word is *integrated* and carried over onto higher levels of processing to make a coherent representation of the general conversation. Selecting a word in the lexicon that fits the auditory input and semantic context are relatively demanding in terms of attentional resources, as well as integrating the new semantic information into the ongoing sentence and context (Frauenfelder & Tyler, 1987; Marslen-Wilson, 1987; Neely, 1991). Also, selection and integration are regarded as inhibitory processes, as word candidates are inhibited from entering the integration process (Aydelott & Bates, 2004; Neely, 1991).

Older theories, such as the cohort-model by Marslen-Wilson and Welsh (1978) see these processing steps as a bottom-up process and argue that semantic information only comes into play in the integration phase. Newer theories reject this (e.g.Gaskell, 2007; Marslen-Wilson, 1987) and argue that semantic information that has been picked up earlier influence the access and selection phases as well. For example, studies suggest that the phonological input does not necessarily need to match with the candidate that is eventually selected (Allopenna, Magnuson, & Tanenhaus, 1998; Connine, Blasko, & Titone, 1993) and word that are more common are more likely to be selected (Marslen-Wilson, 1987). Also, studies focussing on the processing of words in the same domain, such as phonologically or semantically associated words, show that words that have similarities in that domain inhibit or facilitate each other in different ways (Chen & Mirman, 2015; Chen & Mirman, 2012). Words that are strongly related in the same domain inhibit each other, while weaker relations facilitate selection. This means that the three steps in spoken-word processing may not be as clear-cut as they seem to be.

Studies that investigate spoken word recognition as both a bottom-up and a top-down process often use the process of *lexical competition* (e.g. Magnuson, Dixon, Tanenhaus, & Aslin, 2007) to find out where semantic information in integrated with the rest of the input signal. The process is part of the access stage and involves different words competing for activation based on, for example, word frequency (Marslen-Wilson, 1987) or relevance to the rest of the input (Aydelott & Bates, 2004; Frauenfelder & Tyler, 1987; Marslen-Wilson, 1987). A word such as *cat* could be a stand-alone word, but also part of *catastrophe* or *catacomb*, etcetera. By placing semantic constraints on these words, the effect of top-down information can be studied. For example, by placing the semantic constraint of *walking* in a sentence such as *the walking ca…*, the selection of *cat* becomes more likely than selection of *catacomb*. In the access stage, semantic context facilitates responses to targets with a relevant context (Andruski et al., 1994; Huettig & Janse, 2012; Neely, 1991). On the other hand, context inhibits responses to irrelevant targets in the selection and integration stages (Stanovich & West, 1983). Dahan and Tanenhaus (2004) hypothesised that mapping phonological input to meaning is a continuous process with representations that are continuously updated using multiple information sources (see also Gaskell & Marslen-Wilson, 1997). They used a closed-set visual world eye-tracking paradigm (for a review, see Dahan et al., 2007), in which they presented participants with sentences with contextual constraints to track the word-selection mechanism. The contextual constraints consisted of verbs, which would semantically constrain the selection of a subsequent noun. This noun was the target of their trials. For example, in the sentence *Nog nooit klom een bok zo hoog* [*Never before climbed a goat so high*] *klom* is semantically constraining to the target *bok*, but the sentence *Nog nooit is een bok zo hoog geklommen* [*Never before has a goat climbed*

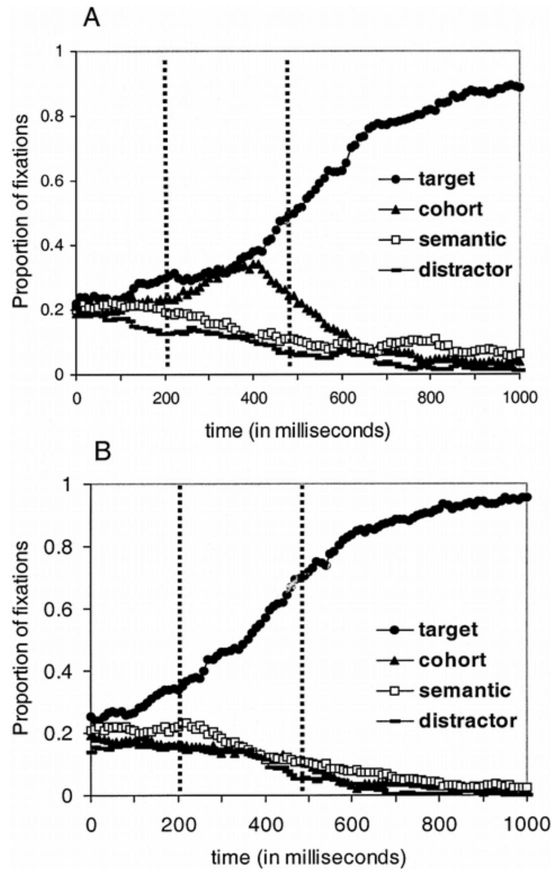

Figure 1: Proportion of fixations to the target, the cohort competitor, the semantic competitor, and the distractor over time from the onset of the target word (in milliseconds) in the neutral-verb condition (A) and in the constraining-verb condition (B). The dotted lines indicate the mean time interval over which fixations reflect the processing of the target word (with a 200-ms delay for saccadic latency). Dahan & Tanenhaus (2004).

so high] is not, as the target is not preceded by a constraining verb. To investigate the selection of the target words and follow lexical competition, participants could choose their target from four pre-selected images. One image was a distractor word being phonologically similar at target onset. For example, an image of a *bot* [*bone]* was presented, as it is similar to *bok* [*goat*]. Dahan and Tanenhaus refer to these images as cohort competitors, as opposite to semantic competitors which are related to the semantically constraining verb (in this case, *climb*). Their results (Figure 1) support their hypothesis that constraining verb guides the word-selection process in a very early stage (Dahan & Tanenhaus, 2004) and in a continuous manner.

Distorted speech

The integration question becomes more complicated, when the input signal is distorted. Distorted speech affects the intelligibility of the speech signal (Bent, Buchwald, & Pisoni, 2009). Information is lost and recovering this information requires more mental resources (Aydelott & Bates, 2004; Farris-Trimble, McMurray, Cigrand, & Tomblin, 2014).

Farris-Trimble, McMurray, Cigrand, and Tomblin (2014) performed studies with CI users (cochlear implant, see Niparko (2009) for an overview) and normal hearing participants using a paradigm similar to Dahan and Tanenhaus (2004). All participants showed patterns similar to Dahan and Tanenhaus'study, but CI users were slower in selecting targets and supressing cohort (phonological) competitors and other distractors. Normal hearing participants, who were tested with CI-simulated stimuli, showed delayed fixations to only the cohort (phonological) competitor and the target. The authors explained that listening to degraded speech increases overall uncertainty about the target word, but that this reason alone does not explain their findings completely.

In a semantic priming study, Aydelott and Bates (2004) used low-pass filtered (degraded) speech and time-compressed speech to investigate the effects of distorted speech on the facilitatory and inhibitory effects of semantic context. As degraded speech affects the identification of words in the (faciliatory) access stage (Andruski et al., 1994; Neely, 1991), low-pass filtered speech reduced the facilitation of responses by semantic context. Degraded speech also had a small effect on inhibition of irrelevant targets. Time–compressed speech increases processing demands in a different way than degraded speech: there is less time to process the information, although the signal remains intact. Intelligibility is less affected (Altmann & Young, 1993; Janse, 2002; Janse, 2003). Aydelott and Bates hypothesised that selection and integration are affected by time-compressed speech, but not lexical access. Their results supported their hypothesis: participants had trouble inhibiting responses to non-target words, which means that selection and integration of information were affected.

Experiment 1: Degraded Speech

When the results of Dahan and Tanenhaus (2004), Ayedelott and Bates (2004), and Farris-Trimle et al. (2014) are taken together, we can make separate hypotheses about the influence of degraded on the integration of phonetic and semantic information. It is hypothesised that spectrally degraded speech interrupts the facilitation of target-related responses when confronted with semantic context. This is because degraded speech is essentially a loss of input information, which affects the access stage of spoken word processing. Translating this hypothesis to the visual world paradigm used by Dahan and Tanenhaus (2004), this would result in fewer gaze fixations towards the target, and more to the other competitors. When presented with a constraining verb condition, it is hypothesised that the semantic context is already facilitating selection of lexical candidates. Thus, the sound-to-meaning mapping could be, at least partially, omitted (Dahan & Tanenhaus, 2004). This would result in a gaze-pattern similar to Figure 1, where gaze shifts to the cohort competitor or target in a semantically neutral context (Figure 1A) and towards the semantic competitor when contextual information is present (Figure 1B). The situation of using contextual hints when presented with degraded speech is interesting: is facilitation of the contextually constraining verb finished (and thus can be used to omit the mapping of the target word) or not? Then, the integration would fail, and participants would rely more on the acoustic signal, resulting in similar gaze patterns as in a neutral context. From now the cohort competitor will be called phonological competitor, as the target and this competitor are phonologically similar at the onset.

Methods

Participants. Thirty-four native Dutch participants (28 females, mean age 22,9 years old (SD 5,2)) participated in this experiment. Image settings were changed after the first five participants, making the acquired data slightly biased and unfit for analysis. These five participants were removed from analysis for this reason. Another participant was removed due to recording problems. Finally, one participant was removed from analysis, as she participated despite meeting one of the exclusion criteria. Thus, the total number of participants that are included in analysis was twenty-seven.

All participants had self-declared normal hearing. Normal hearing was defined as hearing thresholds of 20 dB HL or lower at audiometric frequencies between 1000 and 8000 Hz. All participants had normal or corrected-to-normal eyesight. All participants signed an informed consent form. Participants received a monetary compensation or course credit for their participation.

Materials. Dahan and Tanenhaus (2004) assembled a specific set of items, which was extended for the purpose of this experiment to 142 nouns. Line-drawing pictures were assigned to the nouns, selected from Cycowicz, Friedman, Rothstein, and Snodgrass (1997) and Snodgrass and Vanderwart (1980). The line-drawing set consisted of 196 pictures, 142 items and 54 distractors that were not used in the experiment.

Forty-four target items were created. The target words could be a target in both context and neutral context; the context-present target words and neutral context target words were identical. Two nouns of the target items overlapped at the onset, creating a target-pair. Thus, 22 item-pairs were created. Target-pairs could overlap in three different ways. First, pairs could be monosyllabic and diverge after the syllable's nucleus (e.g., *pot–pop* [*jar-doll*]). Second, pairs could consist of two- or three-syllable nouns with overlapping first syllables (e.g., *kabouter–kameel* [*gnome-camel*]). Finally, some pairs (*schilder–schildpad* [*painter-turtle]*) diverged at the coda of the first syllable. Each word was assigned with a verb that semantically fits only one of the words in a pair. For example, the sentence *Gisteren bevatte de pot alleen wat meel [Yesterday contained the jar only some flour]* was assigned to the *pop-pot* pair. The semantically coherent word (*pot [jar]*) is the target. The semantically incoherent word (*pop [doll]*) is referred to as the phonologic competitor. A semantically coherent distractor that does not overlap at the onset was included as well. In our example, this is *gieter [sieve]*. Finally, an unrelated distractor was included too (here *rits [zipper]*). Furthermore, twenty-nine context filler items and twenty-five neutral filler items were added to the set. The filler items could be semantically coherent, phonologically similar, or phonologically unrelated subjects of the verb chosen for the set.

The items were digitally recorded in a male voice. Afterwards, the items underwent spectral degradation using a noise-band vocoder (Dudley, 1939) in MATLAB, to simulate CI processing (Pals, Sarampalis, & Başkent, 2013; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). This resulted in two identical sets of items: one in natural speech and one in degraded speech.

Procedure. Participants filled out a questionnaire and an informed consent form. The questionnaire (Appendix I) asked about the language spoken at home, knowledge of other languages, possible communication-, speech-, and language problems, hearing, and sight. A short hearing test was administered to check their hearing thresholds. Participants were familiarised with 196 images and their names, which were the experimental items and additional distractors.

During the familiarisation, the participant would guess the name of each image. The experimenter indicated if the guess was correct or not, and corrected participants if they guessed wrong. This way, the participant was familiar with the words used in the experiment.

Participants were seated at a comfortable distance from the screen, approximately sixty centimetres from the screen. Then, the eye-tracker was mounted and calibrated. Eye-movements were monitored using Eye-Link II at a sampling rate of 250 Hz. The sentences were presented by a loudspeaker in front of the participant at a distance of 140 centimetres at a comfortable volume level.

The experiment consisted of two subsequent sessions in a soundproof booth: one session with the natural stimuli and one session with the degraded stimuli. Participants were assigned to natural speech first or degraded speech first depending on their participant number. Depending on the order of presentation, participants either started with setting up the Eyelink and proceeding with the natural speech session, or with a training to understand degraded speech. This training consisted of ten sentences in the degraded speech that has been used during the experiment. If the participant felt sufficiently familiar with the degraded speech, the Eyelink would be set up. If not, the participants could opt for another set of sentences.

After three practice trials, participants were presented with 47 trials in in either natural and degraded speech. Then, the participant would run another session in the other speech type. Trials followed a fixed structure. Recording started 0.1 seconds before the start of each trial. First, a fixation cross appeared in the middle of the screen for 500 ms, after which a blank screen was presented for 500 ms. Next, four images were presented along with a sentence in natural or degraded speech, depending on the session type. During a target trial the screen would contain the target, the phonologic competitor, the semantic competitor, and an unrelated distractor

(Figure 2). Participants selected their choice by clicking on the image that was mentioned in the sentence. There was no time limit for giving an answer. Participants had a self-paced blink brake after each trial. Drift correction was performed every five trials.

Figure 2: Example stimulus. Stimuli were Dutch sentences with or without a semantically constraining verb (here *shouts*). The verb was followed by the target (*clown*), which was depicted together with a phonologic competitor which was similar in onset (*claw*), a sematic distractor which was semantically coherent with the verb (*child*), and an unrelated distractor (*strawberry*).

Analysis. Two methods have been used to analyse the data acquired in the visual-world paradigm. One is the more traditional method, where the acquired data is aggregated as the proportion of fixations over time and trials (Barr, 2008; Tanenhaus et al., 1996).This method will be used to visualise the acquired eye-tracking data. Grow curve analysis, a variation of multilevel regression, will be used to break down the effects of context, input, and their interactions (Barr, 2008; Mirman, 2014; Mirman, 2015).

Only correct responses to targets were included in the analysis. Analysis was performed in R (R Core Team, 2014), using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015).

The screen was divided into four fixation regions; one for each image. Each fixation region is considered as a binomial variable; participants did (1) or did not (0) fixate to the image.

The fixation proportion towards each fixation region was calculated afterwards by aggregating it over time and trials.

Analysis was further performed by means of a grow curve analysis (Barr, 2008; Mirman, 2014; Mirman, 2015) using fourth-order orthogonal polynomials. The gaze data was analysed from stimulus onset. The models for target fixations, phonologic competitor fixations, and semantic competitor fixations included fixed effects of semantic condition and speech type on the first, second, third and/or fourth term, together with their interactions. Random effects of participants and items on the fixations were added as well. Random effects on item was only applied on the first and second order terms. The random effects were included as such in every model, but the fixed effects differed for each model. Model selection was based on the lowest Bayesian information criterion (BIC) (Bates et al., 2015). Each term of the fourth-order orthogonal polynomial has different effect on the regression model. The first (linear) term represents the intercept. It gives an indication of general shift in focus after target onset (Chen & Mirman, 2012; Mirman, Dixon, & Magnuson, 2008; Mirman, 2015). A steep linear slope means a fast withdrawal from one image fixation to another, presumably to the target. The second (quadratic) term represents the slope, which gives an indication of differences in curvature, or the pace of changes in gaze fixations (Chen & Mirman, 2015; Mirman et al., 2008). Third (cubic) and fourth (quartic) terms influence the curvature of the regression model, which is visible as a 'bump' or 'curve' in the plots. They give information about gaze shifts as additional curves atop of the ones explained by the first and second order terms. Fourth order (quartic) terms are hard to interpret (Barr, 2008; Mirman et al., 2008). Because of this, the first, second, and third order terms will be the main focus when discussing the models.

The overall time-course of the target fixations was modelled using a fourth-order orthogonal polynomial function and fixed effects of speech type (degraded/neutral) and semantic condition (neutral/context) on the first, second, and third order terms, together with their interactions (Figure 5). The model that has been used to model the gaze data can be formulated as *target ~ (intercept + slope + 3*rd *term) * semantic condition * speech condition + (intercept +slope + 3*rd *term + 4*th *term| participant) + (intercept +slope | item). target* is the dependent variable of target fixations, *intercept* is the first linear term of the model, *slope* is the second quadratic term, and *3*rd *term* is the cubic term. These three terms interacted with semantic condition (context/neutral) and speech condition (natural/degraded). The two final components of the formula represent the random effects mentioned earlier.

The models are a method to define statistical differences. In our case this is done using the Bayesian Information Criterion (BIC, see for example Mirman, 2015). The time-course of phonologic competitor fixations with the lowest BIC was modelled using a fourth-order orthogonal polynomial function and included fixed effects of speech type (degraded/neutral) and semantic condition (neutral/context) on the first, second, and fourth order terms, together with their interactions. The model that has been used to model the gaze data can be formulated as *phonologic competitor ~ (intercept + slope + 4*th *term) * semantic condition * speech condition + (intercept +slope + 3*rd *term + 4*th *term| participant) + (intercept +slope | item) + 3*rd *term. phonologic competitor* is the dependent variable of phonologic competitor fixations, *intercept* is the first linear term of the model, *slope* is the second quadratic term, and *4*th *term* is the quartic term. These three items interacted with semantic condition (context/neutral) and speech condition (natural/degraded). Random effects were added as well. The third term is not used to model the data, but it is added at the end of the formula to ensure that the general form of the grow curve remains intact.

The semantic competitor model with the lowest BIC was simpler than the target and phonologic competitor models. This model included fixed effects of semantic condition and speech type on the first and second term, together with their interactions. The model that has been used to model the gaze data can be formulated as *semantic competitor ~ (intercept + slope) * semantic condition * speech condition + (intercept +slope + 3rd term + 4th term| participant) + (intercept +slope | item) + 3rd term + 4th term*. *semantic competitor* is the dependent variable of semantic competitor fixations, *intercept* is the first linear term of the model and *slope* is the second quadratic term. These two items interacted with semantic condition (context/neutral) and speech condition (natural/degraded). Random effects were added as well. The third term and fourth terms are added at the end of the formula to ensure that the general form of the grow curve remains intact.

Results

Behavioural results. Only correct trials were included in the analysis. The behavioural performance in the natural speech condition was high with an overall accuracy of 0,998 (*sd* = 0,048). Participants performed slightly worse in the degraded speech condition, with an overall accuracy of 0,967 (*sd* = 0,177). McNemar's chi-square test for symmetry (McNemar, 1947; R Core Team, 2014) showed that participants performed significantly worse in the degraded speech condition: $\chi^2 = 33.065$, df = 1, p-value < 0,001.

Eye-tracking results. Only correct trials were included in the analysis. Figure 3 shows the proportion of fixations towards targets, phonologic competitors, semantic distractors, and unrelated distractors from target onset. Taking into account that participants need roughly 200 ms for programming their eye-movements (Matin, Shao, & Boff, 1993), four different patterns can be distinguished. Participants are already moving their gaze towards the target (Figure 3, upper left image) if they are provided with semantic context. In a neutral context condition (Figure 3, lower left image), this distinction happens around the 200 ms window. The clear distinction between context and neutral conditions is not visible in the degraded speech condition. Gaze remains around chance level for about 400 ms in the semantic context condition (Figure 3, upper right image) and for over 500 ms in the neutral condition (Figure 3, lower right image). However, participants clearly shift their gaze earlier towards the target when provided with semantic context.

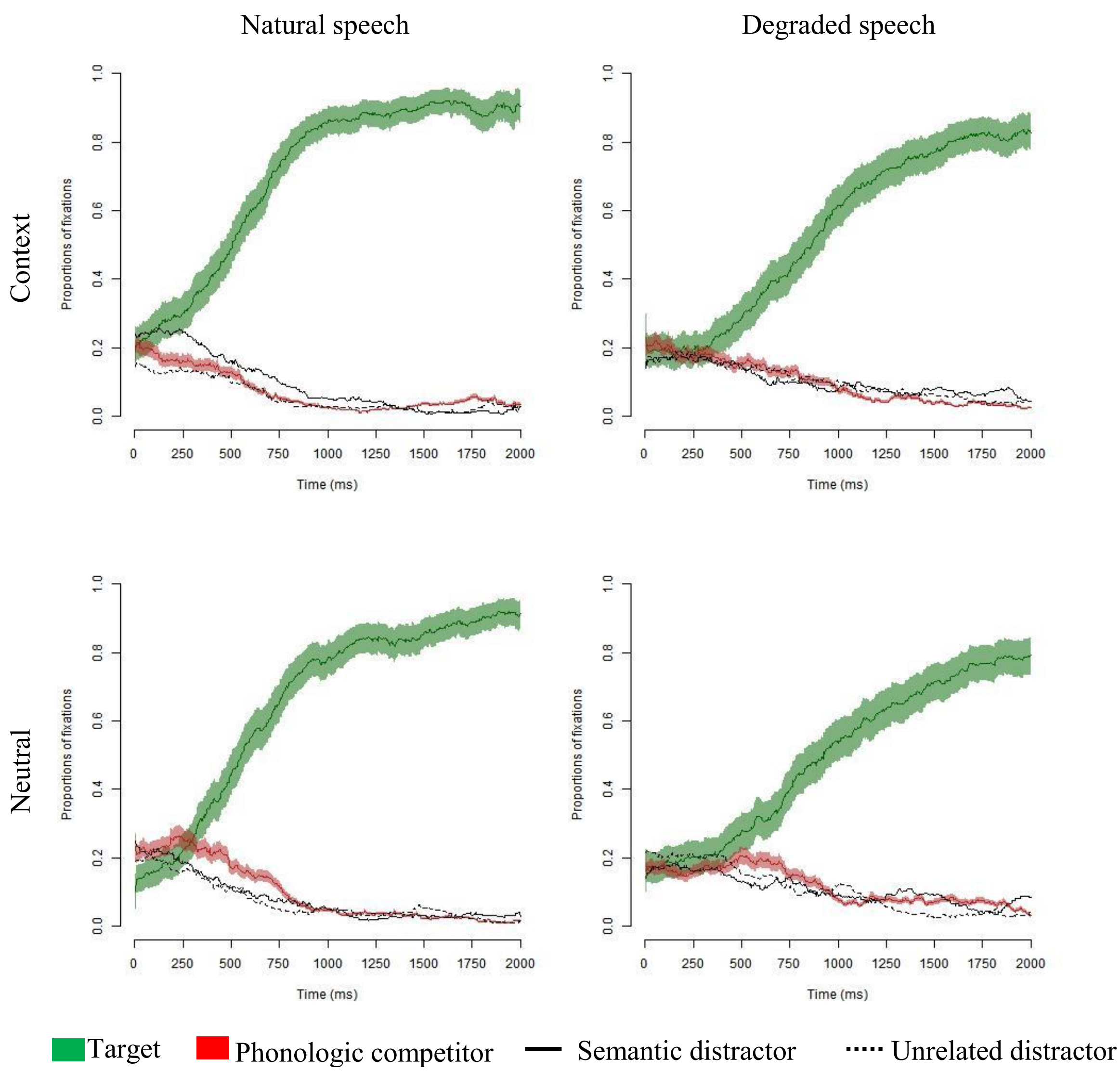

Figure 3: Proportion fixations over time after target onset for target, phonologic competitor, semantic competitor and the unrelated distractor in the natural speech condition (left) and the degraded speech condition (right).

The regression model of the target fixations (Figure 4) shows significant effects of semantic condition on the first (*Estimate* = -2,01, *SE* = 0,243; $p < 0,001$), second (*Estimate* = 2,38, *SE* = 0,233; $p < 0,001$), and third (*Estimate* = 1,44, *SE* = 0,229; $p < 0,001$) order terms. Speech session only on the second term (*Estimate* = -7,44, *SE* = 1,86; $p < 0,001$). Speech type and semantic condition also have interaction effects on the first (*Estimate* = 5,47, *SE* = 0,395; *p* <

0,001) and third order term (*Estimate* = 3,11, *SE* = 0,355; $p < 0{,}001$), indicating a different use of context between the two types of speech. For a complete overview of all estimates, see Table 1.

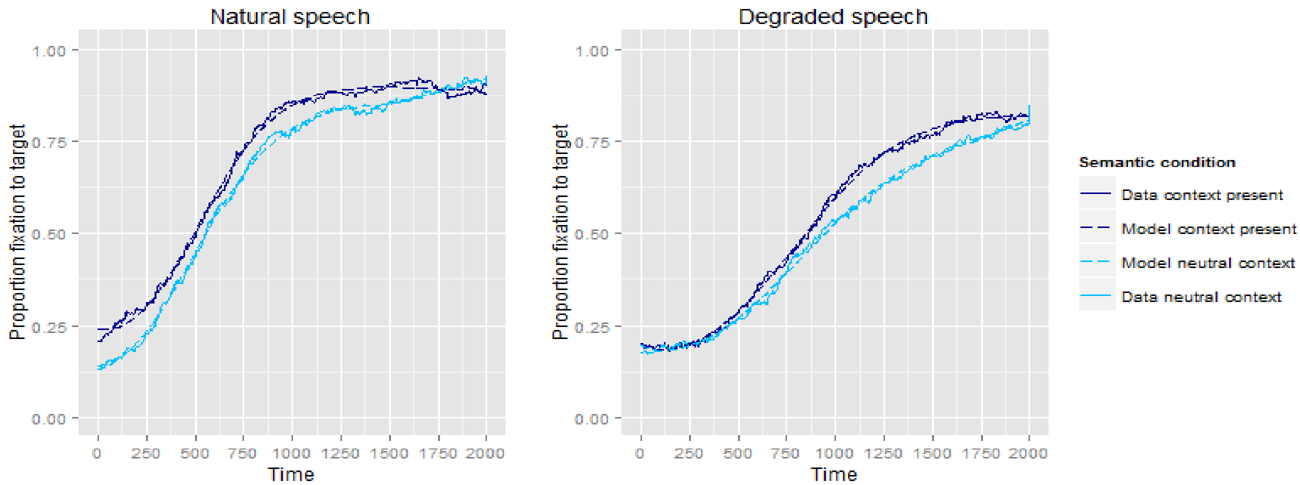

Figure 4: Visualisation of data (straight line) and model (dotted line) of the target fixations in natural and degraded speech.

Table 1: Parameter estimates of target fixations model. The context present condition and the degraded speech condition are the default conditions. Significance codes: < 0,001 '***' < 0,01 '**' < 0,05 '*'.

	Estimate	Std. Error	Pr(>|z|)
Intercept	0,2271	0,1650	0,1690
1st term	30,40	2,478	< 0,001 ***
2nd term	-3,402	1,490	0,02251 *
3rd term	-4,405	1,090	< 0,001 ***
Semantic condition	-0,2558	0,01002	< 0,001 ***
Speech session	1,209	0,2068	< 0,001 ***
1st : Semantic condition	-2,007	0,2429	< 0,001 ***
2nd : Semantic condition	2,380	0,2328	< 0,001 ***
3rd : Semantic condition	1,435	0,2292	< 0,001 ***
1st : Speech session	3,028	3,195	0,3433

2nd : Speech session	-7,439	1,858	< 0,001 ***
3rd : Speech session	-1,160	1,543	0,4520
Semantic condition : Speech session	-0,05945	0,01608	< 0,001 ***
1st : Semantic condition : Speech session	5,469	0,3949	< 0,001 ***
2nd : Semantic condition : Speech session	0,5087	0,3783	0,1787
3rd : Semantic condition : Speech session	3,111	0,3553	< 0,001 ***

The regression model of the phonologic competitor fixations (Figure 5) shows a different pattern. Semantic condition has a significant effect on the first (*Estimate* = 5,00, *SE* = 0,392; *p* < 0,001), second (*Estimate* = -1,31, *SE* = 0,350; *p* < 0,001), and fourth order term (*Estimate* = -2,54, *SE* = 0,313; *p* < 0,001). Speech condition has a significant effect on the second (Estimate = -11,1, SE = 4,22; p < 0,001) and fourth order term (*Estimate* = -5,93, *SE* = 1,41; *p* < 0,001). Speech type and semantic condition have interaction effects on the first (*Estimate* = -15,0, *SE* = 0,675; *p* < 0,01), second (*Estimate* = -9,26, *SE* = 0,577; *p* < 0,001) and third order terms (*Estimate* = -2,84, *SE* = 0,467; *p* < 0,001), indicating a different use of context between the two types of speech. For a complete overview of all parameter estimates, see Table 2.

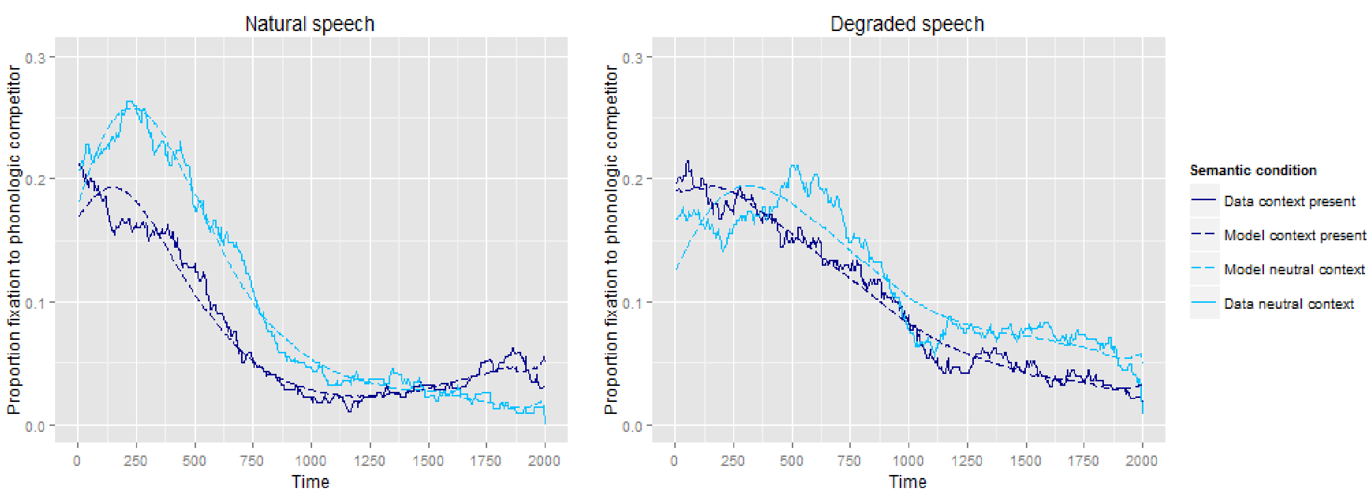

Figure 5: Visualisation of data (straight line) and model (dotted line) of the phonologic competitor fixations in natural and degraded speech.

Table 2: Parameter estimates of phonologic competitor fixations model. The context present condition and the degraded speech condition are the default conditions. Significance codes: < 0.001 '***' < 0.01 '**' < 0.05 '*'.

	Estimate	Std. Error	Pr(>|z|)
Intercept	-3,558	0,3644	< 0,001 ***
1st term	-33,09	6,136	< 0,001 ***
2nd term	-4,405	3,679	0,2312
4th term	0,1012	1,559	0,9483
Semantic condition	0,2965	0,01646	< 0,001 ***
Speech session	-0,7674	0,4122	0,06261
3rd term	2,542	2,458	0,3010
1st : Semantic condition	5,004	0,3928	< 0,001 ***
2nd : Semantic condition	-1,314	0,3498	< 0,001 ***
4th : Semantic condition	-2,539	0,3131	< 0,001 ***
1st : Speech session	-1,639	7,327	0,8229
2nd : Speech session	11,10	4,221	0,008578 **
4th : Speech session	-5,386	1,413	<0,001***
Semantic condition : Speech session	0,03740	0,02854	0,1900
1st : Semantic condition : Speech session	-14,96	0,6752	< 0,001 ***
2nd : Semantic condition : Speech session	-9,256	0,5768	< 0,001 ***
4th : Semantic condition : Speech session	2,843	0,4665	< 0,001 ***

The regression model of the semantic competitor fixations (Figure 6) is simpler than the target and phonologic competitor models. Semantic condition has a significant effect on the second term (*Estimate* = -2,08, *SE* = 0,327; p < 0,001). Speech condition has a significant effect on the first (Estimate = -30,9, SE = 6,23; p < 0,001) and second order term (*Estimate* = -6,63, *SE* = 2,82; p < 0,05). Speech type and semantic condition have interaction effects on both the first (*Estimate* = 7,02, *SE* = 0,670; p < 0,01), second order term (*Estimate* = 8.34, *SE* 0,541; p <

0,001), indicating a different use of context between the two types of speech. For a complete overview of all parameter estimates, see Table 3.

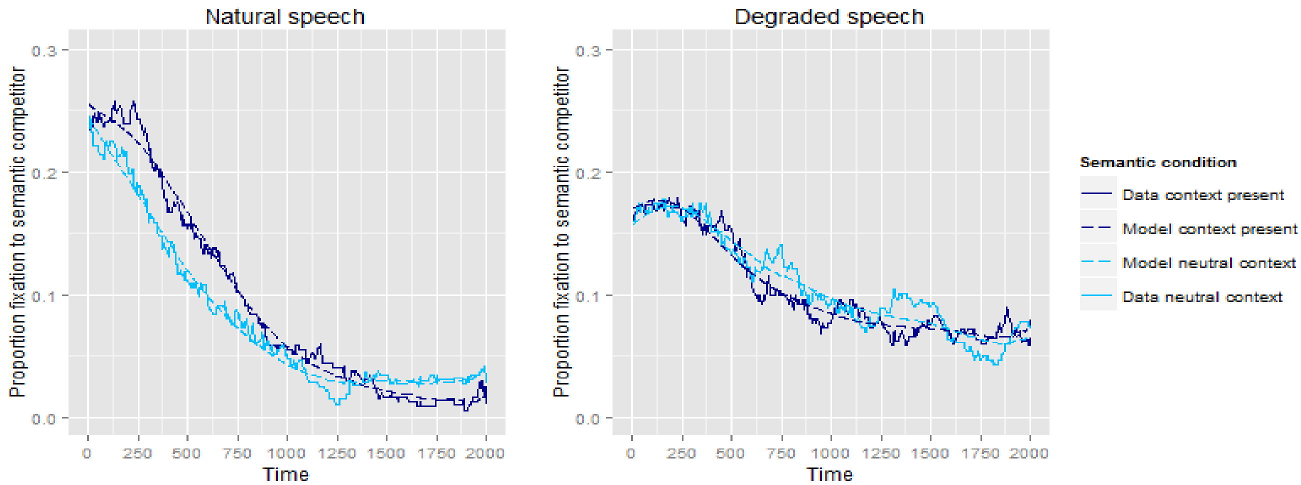

Figure 6: Visualisation of data (straight line) and model (dotted line) of the semantic competitor fixations in natural and degraded speech.

Table 3: Parameter estimates of semantic competitor fixations model. The context present condition and the degraded speech condition are the default conditions. Significance codes: 0 '***' 0.001 '**' 0.05 '*'.

	Estimate	Std. Error	Pr(>|z|)
Intercept	-3,137	0,3077	< 0,001 ***
1st term	-15,49	4,821	0,001313 **
2nd term	-2,781	2,728	0,3077
Semantic condition	0,04908	0,01480	0,000913 ***
Speech session	-0,7530	0,3797	0,047360 *
3rd term	0,7766	1,237	0,5300
4th term	-0,2780	1,086	0,7980
1st : Semantic condition	-0,4614	0,3379	0,1720
2nd : Semantic condition	-2,083	0,3266	< 0,001 ***
1st : Speech session	-30,89	6,230	< 0,001 ***
2nd : Speech session	-6,625	2,821	0,01884 *
Semantic condition : Speech session	-0,1467	0,02814	< 0,001 ***
1st : Semantic condition : Speech session	7,022	0,6701	< 0,001 ***
2nd : Semantic condition : Speech session	8,399	0,5407	< 0,001 ***

Discussion results Experiment 1

As mentioned earlier, the linear, quadratic, and cubic terms will have the main focus. The implications of the quartic term is hard to relate to the research question.

When presented with degraded speech, context and speech type showed interaction effects on different aspects of the models, indicating that they differ between conditions and in several aspects of the speech recognition process (Barr, 2008; Mirman et al., 2008; Mirman, 2015). As hypothesised, it seems that participants fail to use semantic context when presented with degraded speech: their fixations are delayed by about 200 ms (Figure 3) compared to the natural speech condition. This delay is not as late as in the degraded speech condition without context, meaning that context is still used, but only later.

Target fixations. The target fixations model is fairly straightforward. Target fixations are different in both the speech condition and the semantic condition (Figure 4 and Table 1). The gaze patterns showed in Figure 3 and the significant interaction on the first term show that participants fixated less to target images when presented with degraded speech, as was hypothesised earlier. It was also hypothesised that the semantic context aids selection of lexical candidates. This is clearly visible in the upper left image of Figure 3, where target selection starts after activation. However, Figure 3 also shows that in degraded speech, the selection of the target image takes longer; participants fixate much more to the competitors, even when presented with context. However, fixations to the target start even later when no context is present. Thus, semantic context indeed aids selection of the target word, but it is delayed. This is also supported by the significant interaction on the third term, which indicates that participants are later in ending their fixation shift (Figure 4) when presented with degraded speech.

Phonologic competitor fixations. It was hypothesised participants will look more to other competitors when presented with degraded speech. This is confirmed by the phonologic competitor model, which predicts fewer fixations to the phonologic competitor in a neutral context when presented with natural speech (Figure 5 and Table 2). On the other hand, the model predicts more gaze fixations to the phonologic competitor in a context present condition when presented with degraded speech, which indicates that the semantic context is not as helpful as in natural speech. There is also an interaction effect between semantic condition and speech session on the second term, which gives an indication of slope, or the pace of changes in gaze fixations (Chen & Mirman, 2015; Mirman et al., 2008). The interaction effect on the second term shows that the shift in amount of fixations is faster in the context present condition while simultaneously presented with degraded speech. So, although looking more to the phonologic competitor in general, the shift from phonologic competitor to the target is still present in degraded speech, meaning that contextual information is used. Participants also use semantic information in natural speech: when presented with natural speech, the pace at which participants withdraw from the phonologic competitor is also faster in a condition where participants were presented with a semantically constraining verb. The main effects of semantic condition and speech type on intercept and slope of the phonologic competitor gaze fixation model are further confirmed by their significant main effects on respectively the first and second term, and the second term. Figure 5 confirms that the change from fixating to the competitor to the target is slower in the degraded speech condition.

Semantic competitor fixations. The semantic competitor model only needed a linear and quadratic term to explain the general course of the model. Again, interaction effects are observed (Table 3). The parameter estimates of the semantic competitor are reversed in sign compared to the phonologic competitor, meaning that effects of speech session and semantic condition are reversed.

The linear parameter estimate predicts more fixations to the semantic competitor in the neutral context condition when presented with degraded speech and natural speech. This supports the hypothesis that participants will fixate more to competitors when presented with degraded speech. The quadratic term of the semantic competitor model indicates that there are differences in slope (Chen & Mirman, 2015; Mirman et al., 2008; Mirman, 2015). The change in gaze fixations is slower when presented with degraded speech and a semantically constraining verb. This would mean that participants were overall slower in changing their gaze from the semantic competitor to the target when presented with degraded speech while also presented with a semantically constraining verb, compared to the neural context condition. Participants consider the semantic constraining verb, but are later in withdrawing from this incorrect competitor to the target. This effect is reversed when presented with natural speech. The pace at which participants withdraw from the semantic competitor is slower in a condition where participants were not presented with a semantically constraining verb. Although not captured by the model, the participant data shows a remarkable gaze pattern in the degraded speech condition. Gaze fixations fluctuate especially when no context is present. This can be interpreted as participants repeatedly shifting their gaze between the semantic competitor and the other images, perhaps indicating that they were uncertain about the correct target image.

In summary, Experiment 1 shows that degraded speech affects the overall amount of fixations to the target. In addition, the moment when the fixation starts is delayed. The models show that speech type and semantic condition have interaction effects. The phonologic and semantic fixation models both show interaction effects on the second term, meaning that participants differ in the change from competitor to target, depending speech condition and semantic condition. In neutral speech, participants tend to fixate more to the phonologic competitor, but only when no context is present, which is a result that was also found by Dahan and Tanenhaus (2004). In degraded speech, participants fixate more to both phonologic and semantic competitors. Participants focus more on the semantic competitor when presented with degraded speech, but the change from fixating to the competitor to the target is slower, indicating slower suppression of these competitors.

Experiment 2: Time-Compressed Speech

Time-compressed speech (see for example Janse, 2003; Poldrack et al., 1998) increases processing demands compared to natural speech, but these additional demands are different from degraded speech (Adank & Janse, 2009; Aydelott & Bates, 2004; Janse, 2002). Time-compressing increases demands by reduces processing time, as opposed to spectrally degraded speech which reduces intelligibility. Time-compressed speech still holds the full phonologic input, which is one of the reasons people tend to adapt quickly to time-compressed speech (Adank & Janse, 2009; Adank & Devlin, 2010; Boyle, Nunn, O'Connor, & Moore, 2013; Janse, 2002; Janse, 2003).

In accordance to the results of Aydelott and Bates (2004), it is hypothesised that time-compressed speech will reduce inhibition of non-related target words, and that this effect is stronger when semantic context is present. Thus, it is expected that more irrelevant semantic distractors are selected for integration compared to the normal hearing condition. The phonologic competitor fixation patterns would be similar in both natural and time-compressed speech condition, or perhaps slightly faster in the time-compressed condition (Janse, 2002). When context is present, the sound-onto-word mapping might be omitted (Dahan & Tanenhaus, 2004; Huettig & Janse, 2012), resulting fewer fixations to the phonological competitors.

Methods

Participants. Twenty-seven native Dutch participants (16 females, mean age 22,3 years old (*SD* 2,3)) participated in this experiment. All participants had self-declared normal hearing. Normal hearing was defined as hearing thresholds of 20 dB HL or lower at audiometric frequencies between 1000 and 8000 Hz. All participants had normal or corrected-to-normal eyesight. All participants signed an informed consent form. Participants received a monetary compensation for their participation.

Materials. Natural speech items and time-compressed natural speech items were used during the experiment. Note that the degraded speech and natural speech items were identical. Consequently, experimental items used in Experiment 2 were identical to Experiment 1, except for the speech type. The items in natural speech were processed with Praat (Boersma & Weenink, 2015). Effects of time-compressed speech are visible from about 60% compression (Boyle et al., 2013; Janse, 2002); people tend to report increasing effort while still able to attain reasonably. The items were compressed to 60% of their original length, resulting in a 1.67x acceleration of the original stimulus, and then equalised. The standard settings were minimum pitch of 50 Hz and a maximum pitch of 700 Hz. Time-compression may cause pitch changes. Individual settings were applied to stimuli to reduce the difference between the original and the time-compressed stimulus to 7 Hz or less.

Procedure. The procedure was identical to Experiment 1, except that no familiarisation session was administered before starting the time-compressed session. Participants filled out a questionnaire and an informed consent form. A short hearing test was administered to check the hearing thresholds. Participants were then familiarised with 196 images and their names, which were the experimental items and additional distractors. The experiment consisted of two subsequent sessions in a soundproof booth: one session with the natural stimuli and one session with the time-compressed stimuli. Participants were assigned to natural speech first or time-compressed speech first depending on their participant number.

Participants were seated at a comfortable distance from the screen, approximately sixty centimetres from the screen. Then, the eye-tracker was mounted and calibrated. Eye-movements were monitored using Eye-Link II at a sampling rate of 250 Hz. The sentences were presented by a loudspeaker in front of the participant at a distance of 140 centimetres at a comfortable volume level. Participants performed three practice trials before starting with the session. They then completed 47 trials in both natural and time-compressed speech.

The trials structure was identical to Experiment 1. Recording started 0.1 seconds before the start of each trial. First, a fixation cross appeared in the middle of the screen for 500 ms, after which a blank screen was presented for 500 ms. The subsequent target screen would contain the target, the phonologic competitor, the semantic competitor, and an unrelated distractor (Figure 2), while simultaneously a sentence was presented in natural or time-compressed speech, depending on the session. Participants selected their choice by clicking on the image that was mentioned in the sentence. There was no time limit for giving an answer. Participants had a self-paced blink brake after each trial. Drift correction was performed every five trials.

Analysis. Analysis was similar to Experiment 1. Only correct responses to targets were included in the analysis. Analysis was performed in R (R Core Team, 2014), using the lme4 package (Bates et al., 2015). Two methods have been used to analyse the data acquired in the visual-world paradigm. For visualisation, acquired data is aggregated as the proportion of fixations over time and trials (Barr, 2008; Tanenhaus et al., 1996). Grow curve analysis has been used to break down the effects of context, input, and their interactions (Barr, 2008; Mirman, 2014; Mirman, 2015).

The screen was divided into four fixation regions; one region for each image. Each fixation region is considered as a binomial variable. In order to create the figures, the fixation proportion towards each fixation region was calculated afterwards by aggregating it over time and trials.

Analysis was further performed by means of a grow curve analysis (Barr, 2008; Mirman, 2014; Mirman, 2015) using fourth-order orthogonal polynomials. The gaze data was analysed from stimulus onset. The models for target fixations, phonologic competitor fixations, and semantic competitor fixations included fixed effects of semantic condition and speech type on the first, second, third and/or fourth term, together with their interactions. Random effects of participants and items were added as well. Random effects on item was only applied on the first and second order terms. The random effects were included as such in every model, but the fixed effects differed for each model. Model selection was based on the lowest Bayesian information criterion (BIC) (Bates et al., 2015). Each term of the fourth-order orthogonal polynomial has different effect on the regression model. The first (linear) term represents the intercept. The second (quadratic) term represents the slope. Third (cubic) and fourth (quartic) terms influence the curvature of the regression model, which is visible as a ‘bump’ or ‘curve’ in the plots. They give information about gaze shifts as additional curves atop of the ones explained by the first and second order terms. Fourth order (quartic) terms are hard to interpret (Barr, 2008; Mirman et al., 2008). Because of this, the first, second, and third order terms will be the main focus when discussing the models.

The overall time-course of the target fixations was modelled using a fourth-order orthogonal polynomial function and fixed effects of speech type (natural/time-compressed) and semantic condition (neutral/context) on the first, second, third, and fourth order terms, together with their interactions. The model that has been used to model the gaze data can be formulated as *target ~ (intercept + slope + 3rd term + 4th term) * semantic condition * speech condition + (intercept +slope + 3rd term + 4th term| participant) + (intercept +slope | item)*. *target* is the dependent variable of target fixations, *intercept* is the first linear term of the model, *slope* is the second quadratic term, *3rd term* is the cubic term, and *4th term* is the quartic term. These terms interacted with semantic condition (context/neutral) and speech condition (natural/time-compressed). The two final components of the formula represent the random effects mentioned earlier.

The phonologic competitor model with the lowest BIC was simpler than the target and semantic competitor models. This model included fixed effects of semantic condition and speech type on the first and second term. The model that has been used to model the gaze data can be formulated as *phonologic competitor ~ (intercept + slope) * semantic condition * speech condition + (intercept +slope + 3rd term + 4th term| participant) + (intercept +slope | item) + 3rd term + 4th term*. *phonologic competitor* is the dependent variable of phonologic competitor fixations, *intercept* is the first linear term of the model and *slope* is the second quadratic term. Interactions with semantic condition (context/neutral) and speech condition (natural/time-compressed) and random effects were added as well. The third term and fourth terms are added at the end of the formula to ensure that the general form of the grow curve remains intact.

The overall time-course of the semantic competitor fixations was modelled using a fourth-order orthogonal polynomial function and fixed effects of speech type (natural/time-compressed) and semantic condition (neutral/context) on the first, second, third, and fourth order terms. The model that has been used to model the gaze data can be formulated as *semantic competitor ~ (intercept + slope + 3rd term + 4th term) * semantic condition * speech condition + (intercept*

+slope + 3rd term + 4th term| participant) + (intercept +slope | item). *semantic competitor* is the dependent variable of semantic competitor fixations, *intercept* is the first linear term of the model, *slope* is the second quadratic term, *3rd term* is the cubic term, and *4th term* is the quartic term. These terms interacted with semantic condition (context/neutral) and speech condition (natural/time-compressed). The two final components of the formula represent the random effects mentioned earlier.

Results

Behavioural results. The overall accuracy in the natural speech condition was 0,993 (*sd* = 0,082). Participants showed an accuracy of 0,989 (*sd* = 0,102) in the time-compressed condition. McNemar's chi-squared test for symmetry showed no significant difference in performance (McNemar, 1947; R Core Team, 2014): $\chi^2 = 0.7619$, df = 1, p-value = 0.3827.

Eye-tracking results. Figure 7 shows the general gaze patterns towards targets, phonologic competitors, semantic competitors, and unrelated distractors from target onset. Again, it is estimated that participants need roughly 200 ms for programming their eye-movements (Matin et al., 1993). When not presented with context, participants seem to shift their gaze slightly (100 ms) later to the target. Participants also tend to look more to the phonological competitor. Gaze shifting starts the latest in the time-compressed speech condition when no context is provided. A distinctive pattern is observed in the time-compressed speech condition when context is provided: the gaze patterns are similar to the normal speech condition, except for the semantic distractor condition. Here, participants temporarily look more to the semantic distractor image than the target. After 300 ms, the participants look relatively more to the target again.

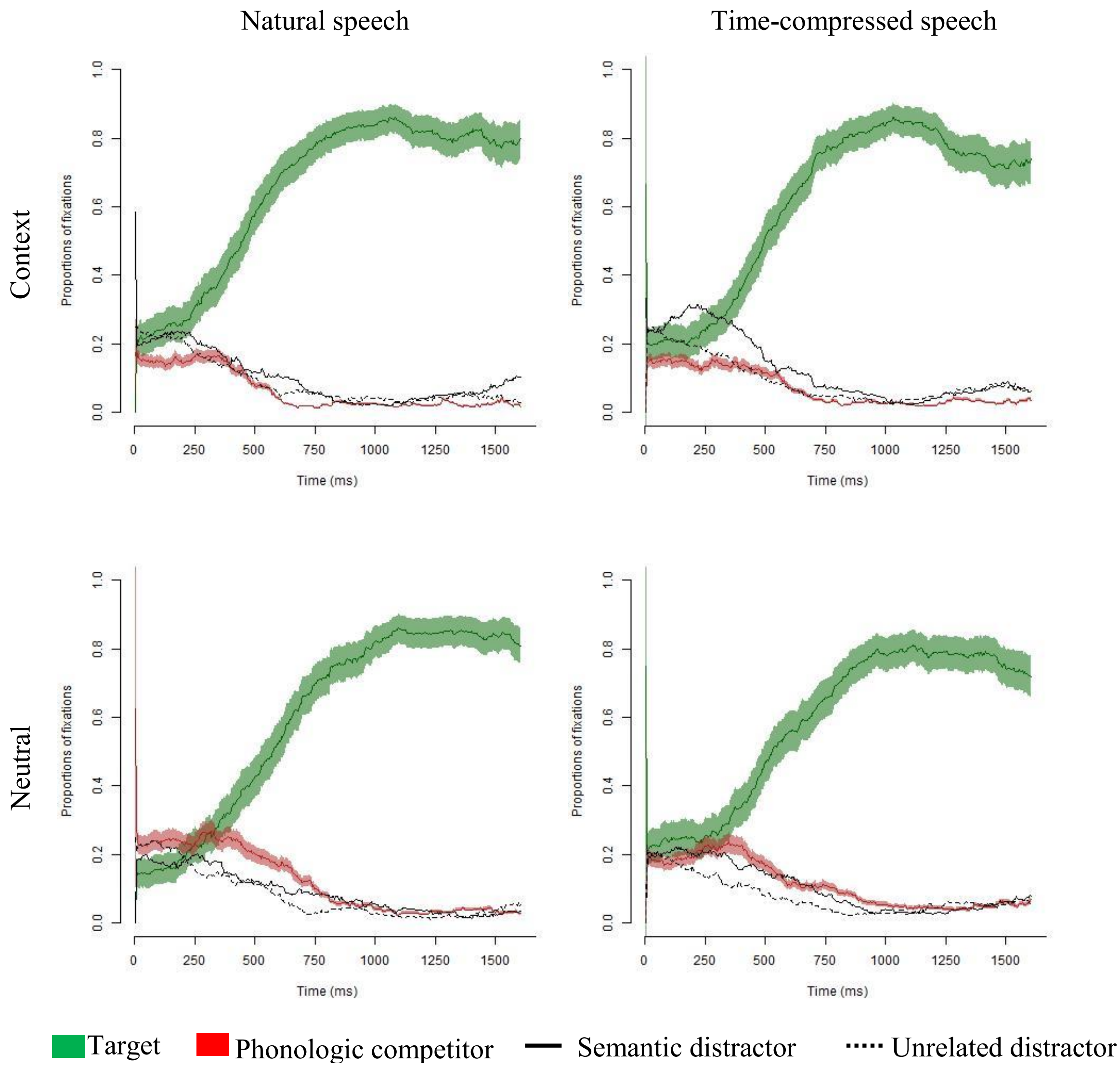

Figure 7: Proportion fixations over time after target onset for target, phonologic competitor, semantic competitor and the unrelated distractor in the natural speech condition (A and C) and the time-compressed speech condition (B and D).

The regression model of the target fixations (Figure 8) shows significant effects of semantic condition on the first (*Estimate* = 0,868, *SE* = 0,388; $p < 0{,}05$), second (*Estimate* = 4,28, *SE* = 0,388; $p < 0{,}001$), third (*Estimate* = -2,81, *SE* = 0,337; $p < 0{,}001$), and fourth (*Estimate* = -2,06, *SE* = 0,445; $p < 0{,}001$) order terms. Speech session showed significant effects on the first

(*Estimate* = 5,49, *SE* = 2,33; $p < 0{,}05$), second (*Estimate* = 6,78, *SE* = 1,96; $p < 0{,}001$), and fourth term (*Estimate* = -2,63, *SE* = 0,885; $p < 0{,}05$). Speech type and semantic condition also have interaction effects on the first (*Estimate* = 5,76, *SE* = 0,654; $p < 0{,}001$), second (*Estimate* = -6,69, *SE* = 0,595; $p < 0{,}001$), and third order term (*Estimate* = -3,73, *SE* = 0,517; $p < 0{,}001$), indicating a different use of context between the two types of speech. For a complete overview of all estimates, see Table 4.

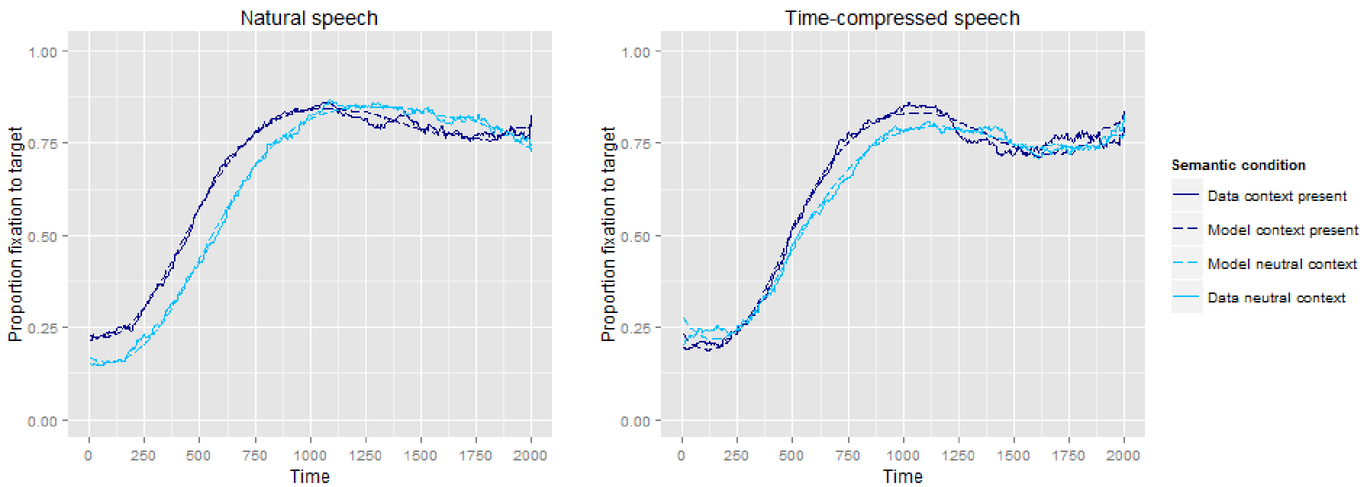

Figure 8: Visualisation of data (straight line) and model (dotted line) of the target fixations in natural and time-compressed speech.

Table 4: Parameter estimates of target fixations model. The context present condition and the time-compressed speech condition are the default conditions. Significance codes: < 0,001 '***' < 0,01 '**' < 0,05 '*'.

	Estimate	Std. Error	Pr(>|z|)
Intercept	0,9470	0,2396	< 0,001 ***
1st term	25,22	1,601	< 0,001 ***
2nd term	-14,55	2,305	< 0,001 ***
3rd term	4,876	0,8065	< 0,001 ***
4th term	10,19	0,6538	< 0,001 ***
Semantic condition	-0,08119	0,01275	< 0,001 ***

Speech session	0,2811	0,3678	0,4448
1st : Semantic condition	0,8676	0,3878	0,02527 *
2nd : Semantic condition	4,280	0,3878	< 0,001 ***
3rd : Semantic condition	-2,812	0,3368	< 0,001 ***
4th : Semantic condition	-2,062	0,4445	< 0,001 ***
1st : Speech session	5,485	2,329	0,01850 *
2nd : Speech session	6,777	1,962	< 0,001 ***
3rd : Speech session	2,528	1,495	0,09083
4th : Speech session	-2,626	0,8852	0,003008 **
Semantic condition : Speech session	-0,1039	0,01873	< 0,001 ***
1st Semantic condition : Speech session	5,757	0,6548	< 0,001 ***
2nd : Semantic condition : Speech session	-6,690	0,5950	< 0,001 ***
3rd : Semantic condition : Speech session	-3,725	0,5165	< 0,001 ***
4th : Semantic condition : Speech session	0,03758	0,5238	0,9428

The regression model of the phonologic competitor fixations (Figure 9) shows that semantic condition only has a significant effect on the second order term (*Estimate* = -5,15, *SE* = 0,562; $p < 0{,}001$). Speech condition has a significant effect on the first (Estimate = -15,1, SE = 1,32; p < 0,001) and second order term (*Estimate* = -5,05, *SE* = 0,878; $p < 0{,}001$). Speech type and semantic condition have interaction effects on the first (*Estimate* = -6,44, *SE* = 0,771; $p < 0{,}001$) and second order terms (*Estimate* = -3,43, *SE* = 0,761; $p < 0{,}001$), indicating a different use of context between the two types of speech. For a complete overview of all parameter estimates, see Table 5.

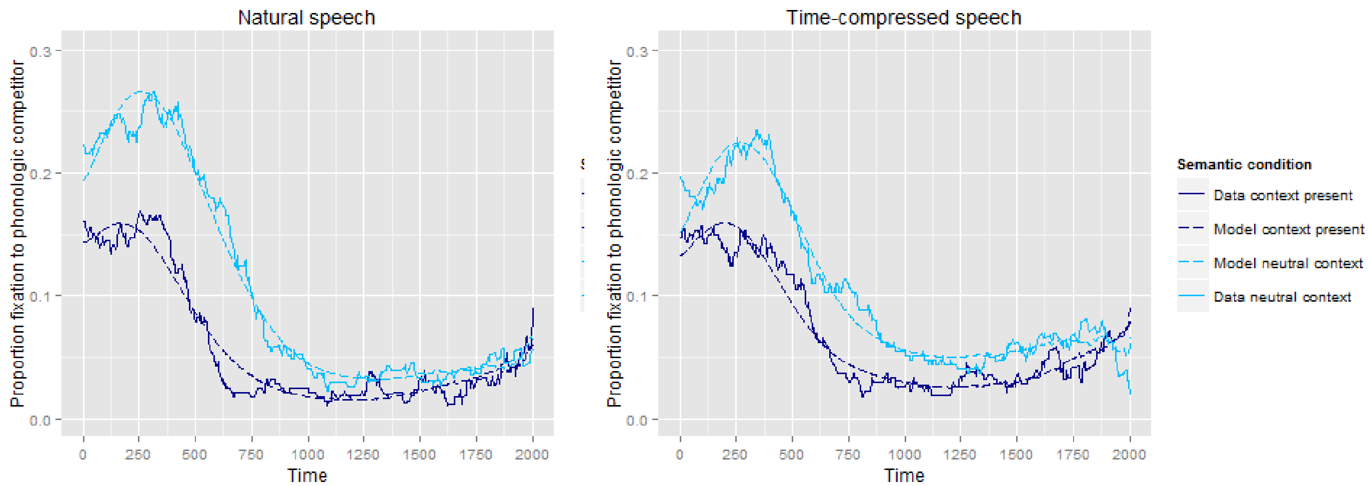

Figure 9: Visualisation of data (straight line) and model (dotted line) of the phonologic competitor fixations in natural and time-compressed speech.

Table 5: Parameter estimates of phonologic competitor fixations model. The context present condition and the time-compressed speech condition are the default conditions. Significance codes: < 0.001 '***' < 0.01 '**' < 0.05 '*'.

	Estimate	Std. Error	Pr(>|z|)
Intercept	-5.011	0,1786	< 0,001 ***
1st term	-49,34	1,440	< 0,001 ***
2nd term	0,5511	0,9218	0,5499
Semantic condition	0,6471	0,02049	< 0,001 ***
Speech session	-0,7047	0,1899	< 0,001 ***
3rd term	6,834	2,247	0,002357 **
4th term	-2,283	1,838	0,2141
1st : Semantic condition	-0,03416	0,4819	0,9435
2nd : Semantic condition	-5,157	0,5618	< 0,001 ***
1st : Speech session	-15,12	1,322	< 0,001 ***
2nd : Speech session	-5,047	0,8779	< 0,001 ***
Semantic condition : Speech session	0,08596	0,0312	0,005877 **
1st : Semantic condition : Speech session	-6,442	0,7714	< 0,001 ***
2nd : Semantic condition : Speech session	-3,433	0,7610	< 0,001 ***

The regression model of the semantic competitor fixations (Figure 10) shows that semantic condition has a significant effect on the second (*Estimate* = -3,01, *SE* = 0,487; $p <$ 0,001) and fourth order term (*Estimate* = 1,70, *SE* = 0,380; $p <$ 0,001). There are no significant effects of speech condition on any term. However, speech type and semantic condition have interaction effects on the first (*Estimate* = -6,38, *SE* = 0,622; $p <$ 0,001), third (*Estimate* = 7,34, *SE* = 0,619; $p <$ 0,001), and fourth order terms (*Estimate* = 7,72, *SE* = 0,539; $p <$ 0,001), indicating a different use of context between the two types of speech. For a complete overview of all parameter estimates, see Table 6.

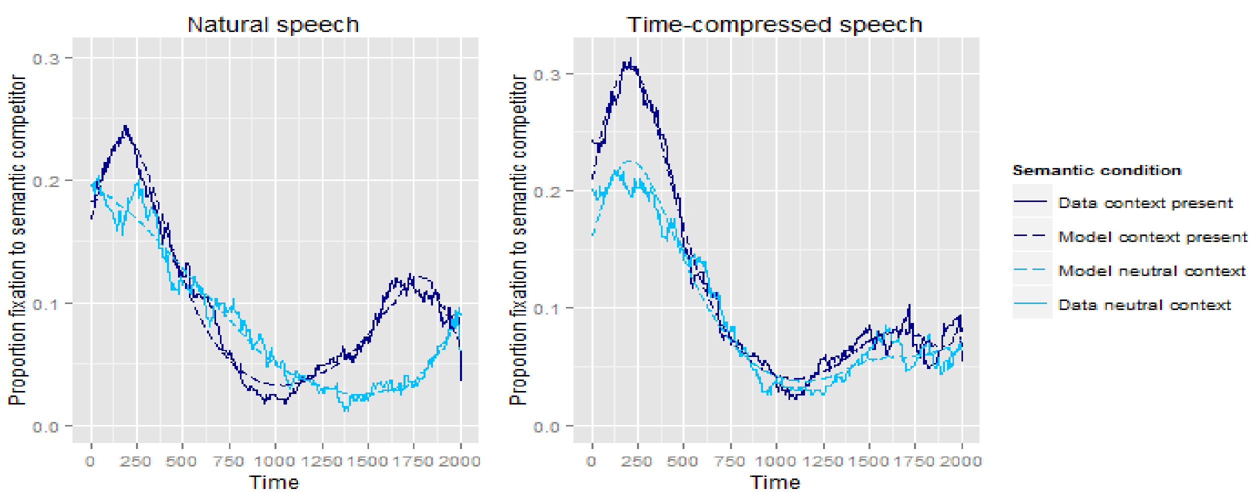

Figure 10: Visualisation of data (straight line) and model (dotted line) of the semantic competitor fixations in natural and degraded speech.

Table 6: Parameter estimates of semantic competitor fixations model. The context present condition and the time-compressed speech condition are the default conditions. Significance codes: 0 '***' 0.001 '**' 0.05 '*'.

	Estimate	Std. Error	Pr(>|z|)
Intercept	-3,846	0,6421	< 0,001 ***
1st term	-49,10	17,34	0,00463 **
2nd term	-15,34	12,19	0,2082
3rd term	-12,70	5,730	0,02667 *

4th term	-13,63	2,394	< 0,001 ***
Semantic condition	-0,2613	0,01887	< 0,001 ***
Speech session	-1,209	0,9142	0,1858
1st : Semantic condition	-0,2958	0,4554	0,5160
2nd : Semantic condition	-3,007	0,4872	< 0,001 ***
3rd : Semantic condition	0,6826	0,4468	0,1265
4th : Semantic condition	1,695	0,3796	< 0,001 ***
1st : Speech session	-23,73	24,86	0,3398
2nd : Speech session	-19,85	17,47	0,2558
3rd : Speech session	-7,737	8,280	0,3501
4th : Speech session	1,123	3,418	0,7426
Semantic condition : Speech session	-0,00632	0,02695	0,8146
1st Semantic condition : Speech session	-6,378	0,6216	< 0,001 ***
2nd : Semantic condition : Speech session	-0,3569	0,6485	0,5821
3rd : Semantic condition : Speech session	7,342	0,6194	< 0,001 ***
4th : Semantic condition : Speech session	7,718	0,5392	< 0,001 ***

Discussion Experiment 2

The linear, quadratic, and cubic terms will have the main focus again, as implications of the quartic terms are hard to relate to the research question. When presented with time-compressed speech, both context and speech type showed interaction effects on different aspects of the models, indicating that their effects differ between conditions and in several aspects of the speech recognition process (Barr, 2008; Mirman et al., 2008; Mirman, 2014; Mirman, 2015).

Target fixations. It was hypothesised that time-compressed speech will reduce inhibition of non-related target words, and that this effect is stronger when semantic context is present. The target model shows that gaze fixations are different between both the speech condition and the semantic condition (Figure 8 and Table 4). There were indeed less target fixations when time-compressed speech was presented along with semantic context. Participants were faster in changing their fixations to the target when presented with natural speech and no context, as indicated by interactions on the second and third term.

Phonologic competitor fixations. The phonologic competitor model needed only the first and second term to predict gaze fixations. It was hypothesised that phonologic competitor fixation patterns would be similar in both natural and time-compressed speech condition, or perhaps slightly faster in the time-compressed condition. There is a main effect of speech session (Table 5), as well as interaction effects on the first and second term, meaning that there are differences in gaze pattern between natural and time-compressed speech. The model predicts fewer gaze fixations to the phonologic competitor when neutral speech and no context are present (Figure 9 and Table 5). There is also an interaction effect between semantic condition and speech session on the second term, which gives an indication of slope, or the pace of changes in gaze fixations (Chen & Mirman, 2015; Mirman et al., 2008). The interaction effect on the second term shows that the general change in amount of fixations is slower in the neural context present condition while the stimuli are presented in natural speech. This means that participants, when presented with time-compressed speech, were overall faster in changing their gaze from the phonologic competitor to the target when a semantically constraining verb was present.

Semantic competitor fixations. It was expected that more irrelevant semantic distractors are selected for integration compared to the natural speech condition. There were no significant main effects of speech type in the semantic competitor model, indicating that speech type does not influence the gaze semantic competitor gaze fixation pattern. Also, semantic condition showed no main effect on the first term, meaning that semantic condition did not influence the amount of fixations to the semantic competitor. Still, interaction effects are observed on the first term between semantic condition and speech session, but not on the second (Figure 10 and Table 6). The linear parameter estimate predicts more fixations to the semantic competitor in the context present condition when presented with time-compressed speech, which is visible as the "bumps" in Figure 7 and Figure 10. This is also the case in natural speech; it is predicted that participants fixated more to the semantic competitor when presented with a semantically constraining verb. This means that sound-onto-word mapping has been omitted. Also, note that the gaze shift finishes much earlier in context present condition when stimuli are presented in natural speech (Figure 7), compared no difference between the semantic conditions when presented with time-compressed speech.

In short, both context and speech type affect the gaze fixation patterns to the targets and competitors. Time-compressed speech results in fewer target fixations, as was hypothesised. Although time-compression does not affect the phonological input, there is a difference between the phonologic competitor fixation patterns in natural or time-compressed speech. The change from fixating to the phonologic competitor to the target is faster in the time-compressed condition when a constraining verb is presented, while the change is faster in the natural speech condition when no context is present. Few main effects of speech type or semantic condition are observed. However, interaction effects on the linear term indicate that more semantic competitors were considered in the time-compressed speech condition when semantic context was present. This supports the hypothesis that time-compression affects inhibition of semantic distractors.

General discussion

In the present study, the effects of context on the processing of natural, degraded, and time-compressed speech have been investigated. See Table 7 for an overview of the most salient findings.

Table 7: Findings of Experiment 1 (degraded speech) and Experiment 2 (Time-compressed speech) by dependent variable

Dependent variable	Experiment 1	Experiment 2
Semantic context	• Context aids selection in both conditions, although stronger in natural speech	• Stronger influence of context on selection in natural speech. • Less inhibition of semantic competitor in time-compressed speech
Speech session	• Degraded speech increases fixations to non-target competitors.	• Time-compressed speech increased fixations to non-target competitors

In the normal speech conditions, we replicated Dahan and Tanenhaus' (2004) results (compare Figure 1 to Figure 3 and Figure 7). When participants are presented with a neutral context, they tend to fixate more to the phonologic competitor before shifting to the target. When the semantically constraining verb is present, participants tend to fixate to the semantic competitor before shifting to the target. In the degraded speech condition, processing of degraded speech is different from natural speech (Aydelott & Bates, 2004; Farris-Trimble et al., 2014). Contrary to the results of Dahan and Tanenhaus (2004), the mapping of degraded speech onto meaningful representations is delayed and not continuous. However, it is not that participants do not use the semantic context at all. In the degraded speech condition, participants shift their gaze earlier to the target when semantic context is present. Generally, we can conclude that degraded speech delays facilitatory process of (lexical) activation (Andruski et al., 1994; Aydelott & Bates, 2004; Farris-Trimble et al., 2014;

Neely, 1991). It seems that participants first gather enough information from the signal, before they can integrate it with information that was already available.

In the time-compressed speech condition, participants tend to fixate more to competitor words. This effect is stronger when semantic context is present, indicating that participants fail to inhibit irrelevant words (Aydelott & Bates, 2004). It is thought that time-compression puts additional strain on speech processing, as there is less time to process the input (Aydelott & Bates, 2004; Janse, 2002; Janse, 2003). Still, Janse (Janse, 2002) found that participants tend to be slightly faster in detection when presented with time-compressed speech. Our results show this as well. The results of Experiment 2 show that speech recognition in time-compressed speech depends on the semantic context. When semantic context is present, participants tend to show the same fixation patterns to targets and phonological competitors in natural and time-compressed speech (Figure 8). Participants seem to have more difficulty with inhibiting responses to the semantic competitor in time-compressed speech (Figure 7 and Figure 10). In time-compressed speech, context-present and neutral context condition show similar target fixation patterns. In natural speech, participants are clearly later in fixating to the target when there is no context. The shift from competitor to target, however, is faster in the natural speech condition. Generally, we can conclude that time-compressed speech affects the inhibitory selection process and slightly affects the facilitatory activation process (Aydelott & Bates, 2004). Adaption and compensation seems to depend on the semantic context of the stimuli (Adank & Janse, 2009; Adank & Devlin, 2010; Boyle et al., 2013)

Spectrally degraded speech and time-compressed speech both result in fewer target fixations. In degraded speech, all fixations are delayed compared to natural speech (Figure 3). In time-compressed speech, fixations are delayed if they are presented in a neutral context (Figure

7). Thus, it seems that participants still use context when presented with either distortion, but make better use of semantic context when presented with time-compressed speech. This is also supported by the failure to inhibit semantic distractors in time-compressed speech (Figure 10). Unfortunately, direct model comparison between the distortions was not performed, but may show more precisely how processing differs between these two types of distortions.

It might be interesting to look at the pupil dilation data as well. Pupil dilation is associated with effort; dilation increases as mental effort increases (Kahneman, Tursky, Shapiro, & Crider, 1969; Mulder, 1986). Although not the original scope of this experiment, the pupil dilation data of the distorted signals can show understand to what extend effort is reduced by context.

References

Adank, P., & Devlin, J. T. (2010). On-line plasticity in spoken sentence comprehension: Adapting to time-compressed speech. *NeuroImage, 49*(1), 1124-1132.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*(4), 419-439. doi:10.1006/jmla.1997.2558

Altmann, G. T., & Young, D. (1993). Factors affecting adaptation to time-compressed speech. *Third European Conference on Speech Communication and Technology,*

Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition, 52*(3), 163-187.

Apfelbaum, K. S., Blumstein, S. E., & McMurray, B. (2011). Semantic priming is affected by real-time phonological competition: Evidence for continuous cascading systems. *Psychonomic Bulletin & Review, 18*(1), 141-149.

Aydelott, J., & Bates, E. (2004). Effects of acoustic distortion and semantic context on lexical access. *Language and Cognitive Processes, 19*(1), 29-56.

Barr, D. J. (2008). Analyzing 'visual world'eyetracking data using multilevel logistic regression. *Journal of Memory and Language, 59*(4), 457-474.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). *Lme4: Linear mixed-effects models using eigen and S4*

Bent, T., Buchwald, A., & Pisoni, D. B. (2009). Perceptual adaptation and intelligibility of multiple talkers for two types of degraded speech. *The Journal of the Acoustical Society of America, 126*(5), 2660-2669.

Boersma, P., & Weenink, D. (2015). *Praat: Doing phonetics by computer*

Boyle, P. J., Nunn, T. B., O'Connor, A. F., & Moore, B. C. (2013). STARR: A speech test for evaluation of the effectiveness of auditory prostheses under realistic conditions. *Ear and Hearing, 34*(2), 203-212. doi:10.1097/AUD.0b013e31826a8e82 [doi]

Chen, Q., & Mirman, D. (2015). Interaction between phonological and semantic representations: Time matters. *Cognitive Science, 39*(3), 538-558.

Chen, Q., & Mirman, D. (2012). Competition and cooperation among similar representations: Toward a unified account of facilitative and inhibitory effects of lexical neighbors. *Psychological Review, 119*(2), 417-430. doi:10.1037/a0027175; 10.1037/a0027175.supp (Supplemental)

Connine, C. M., Blasko, D. G., & Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language, 32*(2), 193-210. doi:10.1006/jmla.1993.1011

Cycowicz, Y. M., Friedman, D., Rothstein, M., & Snodgrass, J. G. (1997). Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology, 65*(2), 171-237.

Dahan, D., & Magnuson, J. S. (2006). Spoken word recognition. *Handbook of Psycholinguistics, 2*, 249-284.

Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: Immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(2), 498.

Dahan, D., Tanenhaus, M. K., & Salverda, A. P. (2007). The influence of visual processing on phonetically driven saccades in the 'visual world' paradigm. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, R. L. Hill, R. P. G. (. van Gompel, M. H. (. Fischer, . . . R. L. (. Hill (Eds.), (pp. 471-486). Amsterdam, Netherlands: Elsevier.

Dudley, H. (1939). The automatic synthesis of speech. *Proceedings of the National Academy of Sciences of the United States of America, 25*(7), 377-383.

Farris-Trimble, A., McMurray, B., Cigrand, N., & Tomblin, J. B. (2014). The process of spoken word recognition in the face of signal degradation. *Journal of Experimental Psychology: Human Perception and Performance, 40*(1), 308-327. doi:10.1037/a0034353; 10.1037/a0034353.supp (Supplemental)

Frauenfelder, U. H., & Tyler, L. K. (1987). The process of spoken word recognition: An introduction. *Cognition, 25*(1), 1-20.

Gaskell, M. G. (2007). Statistical and connectionist models of speech perception and word recognition. *The Oxford Handbook of Psycholinguistics,* , 55-69.

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes, 12*(5-6), 613-656.

Huettig, F., & Janse, E. (2012). Anticipatory eye movements are modulated by working memory capacity: Evidence from older adults. *The 18th Annual Conference on Architectures and Mechanisms for Language Processing (AMLaP 2012),*

Janse, E. (2002). Time-compressing natural and synthetic speech. *Interspeech,*

Janse, E. (2003). *Production and perception of fast speech* Netherlands Graduate School of Linguistics.

Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science, 31*(1), 133-156. doi:10.1080/03640210709336987

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25*(1), 71-102.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10*(1), 29-63.

Matin, E., Shao, K., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics, 53*(4), 372-380.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General, 134*(4), 477-500.

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika, 12*(2), 153-157.

Mirman, D. (2014). *Growth curve analysis and visualization using R* CRC Press.

Mirman, D. (2015). Growth curve analysis. Retrieved from http://www.danmirman.org/gca

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language, 59*(4), 475-494.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. *Basic Processes in Reading: Visual Word Recognition, 11*, 264-336.

Niparko, J. K. (2009). *Cochlear implants: Principles & practices* Lippincott Williams & Wilkins.

Norris, D., Cutler, A., McQueen, J. M., & Butterfield, S. (2006). Phonological and conceptual activation in speech comprehension. *Cognitive Psychology, 53*(2), 146-193.

Pals, C., Sarampalis, A., & Başkent, D. (2013). Listening effort with cochlear implant simulations. *Journal of Speech, Language, and Hearing Research, 56*(4), 1075-1084.

Poldrack, R., Protopapas, A., Nagarajan, S., Tallal, P., Merzenich, M., Temple, E., & Gabrieli, J. (1998). Auditory processing of temporally compressed speech: An fMRI study. *Journal of Cognitive Neuroscience, , 10* 126-126.

R Core Team. (2014). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science (New York, N.Y.), 270*(5234), 303-304.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6*(2), 174.

Stanovich, K. E., & West, R. F. (1983). On priming by a sentence context. *Journal of Experimental Psychology: General, 112*(1), 1.

Tanenhaus, M. K., Sedivy, J. C., Spivey-Knowlton, M. J., Allopenna, P. D., Eberhard, K. M., & Magnuson, J. S. (1996). Eye movements and spoken language comprehension. *Acl,* 48-54.

Appendices

Appendix I

Questionnaire about background

Proefpersooninformatie

Mentale representaties van spraak

Proefpersoon ID	Datum

1. Geboortedatum:
2. Geslacht: M / V
3. Opleiding: Jaar:
4. Wat is/zijn de eerste taal/talen die u geleerd heeft?

5. Welke tal(en) werd(en) bij u thuis gesproken als kind?

6. Welke andere talen beheerst u? Kunt u hierbij aangeven in hoeverre u deze taal begrijpt en spreekt (slecht / matig / goed)?

7. Heeft u ooit last van gehoorproblemen gehad? Zo ja, graag toelichten.

8. Heeft u communicatie-, taal-, en/of spraakproblemen als dyslexie?
Ja / Nee
9. Bent u muzikaal? Zo ja, heeft u een muzikale opleiding (graag toelichten)?

10. Is uw zicht gecorrigeerd? Zo ja, graag toelichten (bril, lenzen, laserbehandeling)?

11. Heeft u een – met correctie – normaal zicht?
Ja / Nee

Afterword

I would like to thank dr. Anita Wagner for her guidance and support. Her helpful advice strengthened my enthusiasm for a topic that was previously quite unknown to me. Also many thanks to prof. dr. Deniz Başkent, who volunteered as second evaluator of this report. I would also like to thank rest of the CI research department of the University Medical Center Groningen for their helpful support.