LANGUAGE AND COGNITIVE PROCESSES
0000, 00 (00), 1–33

Speech reductions change the dynamics of competition during spoken word recognition

Susanne Brouwer[1], Holger Mitterer[1], and Falk Huettig[1,2]

[1]Comprehension Group, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[2]Donders Institute for Brain, Cognition, and Behaviour, Radboud University, Nijmegen, The Netherlands

Three eye-tracking experiments investigated how phonological reductions (e.g., "puter" for "computer") modulate phonological competition. Participants listened to sentences extracted from a spontaneous speech corpus and saw four printed words: a target (e.g., "computer"), a competitor similar to the canonical form (e.g., "companion"), one similar to the reduced form (e.g., "pupil"), and an unrelated distractor. In Experiment 1, we presented canonical and reduced forms in a syllabic and in a sentence context. Listeners directed their attention to a similar degree to both competitors independent of the target's spoken form. In Experiment 2, we excluded reduced forms and presented canonical forms only. In such a listening situation, participants showed a clear preference for the "canonical form" competitor. In Experiment 3, we presented canonical forms intermixed with reduced forms in a sentence context and replicated the competition pattern of Experiment 1. These data suggest that listeners penalize acoustic mismatches less strongly when listening to reduced speech than when listening to fully articulated speech. We conclude that flexibility to adjust to speech-intrinsic factors is a key feature of the spoken word recognition system.

Keywords: Spoken word recognition; Reduced forms; Eye tracking; Dutch.

Correspondence should be addressed to Susanne Brouwer, Max Planck Institute for Psycholinguistics, Postbus 310, 6500 AH Nijmegen, The Netherlands. E-mail: s-brouwer@northwestern.edu

The authors thank Mirjam Ernestus, Natasha Warner, and James McQueen for helpful suggestions on an earlier version of this article.

INTRODUCTION

Most research on spoken word recognition has focused on careful speech read aloud by selected speakers (see Cutler, 1998). The advantage of using careful speech materials is that they are highly controllable and intelligible. Such materials have provided valuable insights into key constructs of spoken word recognition such as lexical competition (e.g., Goldinger, Luce, & Pisoni, 1989; Marslen-Wilson, 1987). However, in listeners' everyday communicative exchanges, they most often encounter casual speech, in which words are often pronounced with fewer segments than when they are produced in the laboratory. For example, the word *hilarious* [hilɛriʌs] is realized as [hlɛrɛs] in a corpus of casually spoken English (Johnson, 2004). Nevertheless, people typically do understand each other with ease. Only a few attempts have been made to study speech "in the wild" (e.g., Ernestus, Baayen, & Schreuder, 2002; Mehta & Cutler, 1988). In this article, we investigate whether spoken word recognition during casual speech differs from spoken word recognition during carefully pronounced speech recorded in the laboratory.

Research using laboratory speech has been very successful. It has demonstrated that listeners rapidly analyze the speech signal and that the processing of speech is closely time-locked to the input (e.g., Goldinger et al., 1989; Zwitserlood, 1989). In a cross-modal priming experiment, for instance, Zwitserlood presented gated fragments of Dutch words, such as *kapitein* "captain", which were followed by visually presented target words for lexical decision. The gated fragments were successively longer onsets (/k/, /kɑ/, /kɑp/, etc.) of words. The Zwitserlood study showed that partial information of onset fragments activated different matching candidate words from gate to gate. For example, when hearing *kapi*... listeners responded faster to words with overlapping onsets, such as *kapitaal* "capital" than when they heard the beginning of a phonologically unrelated word. Lexical access thus involves the continuous activation of multiple lexical candidates. As more acoustic evidence becomes available, candidates inconsistent with the speech signal compete less for recognition than candidates that are consistent with the input. Thus the ultimate winner of the word recognition process emerges from a competition process among these candidates (see McQueen & Cutler, 2001, for further discussion).

An important finding of laboratory research is that lexical candidates with initial overlap with the word to be recognised compete more strongly than words with medial or final overlap (Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978). Upon hearing the spoken sequence /kæp../, all words that start with these sounds, such as *captain*, are activated in parallel but words that overlap later in time such as *apple* /æp../ are less activated. Such effects have been particularly clearly demonstrated in eye-tracking studies that used the visual-world paradigm (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). In this paradigm, listeners' eye movements to pictures of objects on a computer screen are measured in response to concurrent speech. Fixation proportions are typically taken to be related to underlying activation levels of word candidates. Eye movements are continuously recorded so that it is possible to evaluate relative competitor activation over time. The paradigm thus provides closely time-locked measures of the ongoing spoken word recognition process.

Using this method, Allopenna, Magnuson, and Tanenhaus (1998) showed that listeners fixate more often on pictures with names similar to the target name than to phonologically unrelated names. In that study, participants' eye movements were tracked as they looked at four pictures on a computer screen (e.g., a "beaker", a "beetle", a "speaker", and a "carriage"). They listened to spoken instructions such as "Pick up the beaker". Participants looked at the pictures of both types of competitors, but more often to competitors matching at word onset (e.g., the "beetle") than competitors matching at word offset (e.g., the "speaker"; see Connine, Blasko, & Titone, 1993). McQueen and Viebahn (2007) replicated these results using printed word displays. In their study, participants' eye movements were recorded as they looked at four printed words on a computer screen. As in the study by Allopenna et al., participants looked more often at phonological competitors than at phonologically unrelated distractors and the effect was stronger for onset-matching competitors (e.g., *buffer* for *buffel* "buffalo") than for offset-matching competitors (e.g., *lotje* "lottery ticket" for *rotje* "firecracker"). In the present study we use this printed word version of the paradigm.

Huettig and McQueen (2007) have recently further validated this method through eye-tracking experiments with both picture and printed word displays. Previous work showed that eye movements in the paradigm can be based on semantic (e.g., Huettig & Altmann, 2005), visual (Dahan & Tanenhaus, 2005; Huettig & Altmann, 2004, 2007), and phonological matches (Allopenna et al., 1998). Huettig and McQueen examined more closely the influence of these three types of matches. When they presented participants with the picture version of the paradigm, they observed a strong influence of all three types of representations on participant's eye movements. Importantly for the present purposes, their study also showed that (when phonological competitors are present) phonological (but not semantic or visual-shape) representations influence eye gaze when printed word displays are used. Huettig and McQueen concluded that the printed word version is more sensitive to phonological manipulations than do the version using pictures. Weber, Melinger, and Lara Tapia (2007) provided further support for this view. They found that written displays produced stronger phonological competition effects than did pictorial displays. The printed word variant of the paradigm thus has been very successful in the investigation of phonological competition during carefully pronounced speech recorded in the laboratory.

To accommodate the finding of strong onset and weak offset competition, it is usually assumed that mismatches lead to strong deactivation of a target word. This is most explicit in the original Shortlist model (Norris, 1994), where the activation of a word candidate increases by one unit for every matching segment but decreases by three units if there is a mismatch. In the TRACE model (McClelland & Elman, 1986), there is no such explicit penalty for a mismatch, but the winner-takes-all competition on a lexical level leads to a strong deactivation of a word if another one matches better.

It is however as yet unknown to what extent the pattern of strong onset and weak offset competition also applies to casual speech. Given the huge amount of variation in casual speech, it might not be beneficial for the listener to weigh mismatches as strongly as some models of spoken word recognition suggest. Additionally, it is conceivable that the competitor words (the "competitor set") may be rather different during casual speech in which speech reduction processes occur very frequently. Johnson (2004), for example, found that over 60% of the words in a spoken English corpus deviated from their citation form by at least one segment and 28% of the words even deviated on two or more segments (see Ernestus, 2000, for convergent evidence for Dutch). To illustrate this phenomenon, Figure 1 shows a waveform and a spectrogram of the same Dutch sentence, once spoken casually and once read out loud. We extracted the sentence *dat staat hier op deze computer, hè?* "that is on this computer, isn't it?" from the Spoken Dutch Corpus (Oostdijk, 2000) and we re-recorded the same sentence in a laboratory setting. Figure 1 shows the waveform and spectrogram of both versions. Figure 1(a) shows the sentence from the spontaneous speech corpus, which is best transcribed as [dɑ sta fiir ɔp dez pjutər ɛ]. The same sentence read out loud was transcribed as [dɑ stad ir ɔ dez kɔmpjutər ɛ]. Clearly, fewer segments are pronounced in the casually uttered sentence than the one recorded in the laboratory, resulting in a durational difference between the two speech fragments. These differences can best be illustrated if we focus on the word *computer* in these sentences (see Figure 2). The segments of the word *computer* in the read utterance are all fully pronounced [see Figure 2(a)]. Figure 2(b) shows this word from the casually produced sentence. As can be seen, the first syllable [kɔm] of *computer* [kɔmpjutər] is missing. This is a clear example of a reduced realization of the target word *computer*.

In this article, we address the question of how such reductions in casual speech impact spoken word recognition. Given our analysis of casual speech, it is likely that word recognition in casual speech differs from word recognition in carefully articulated and fully pronounced speech. Consider,

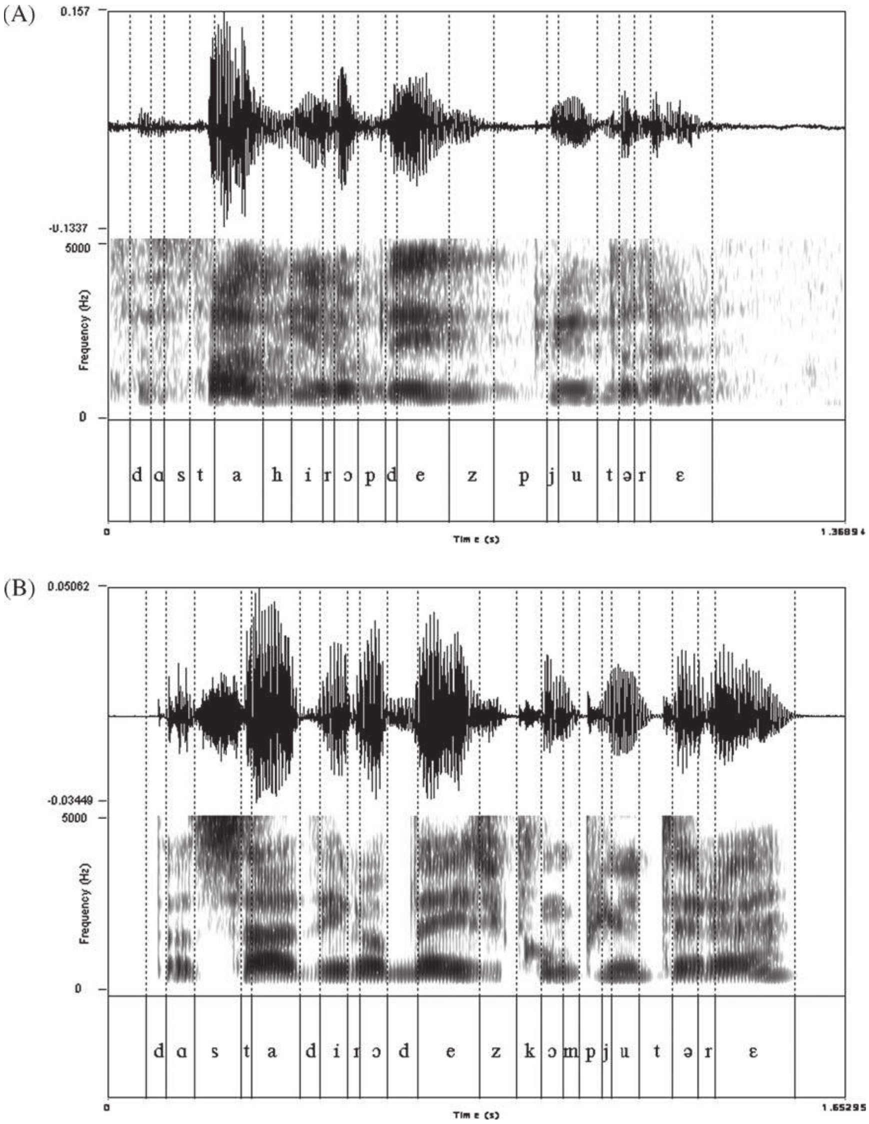

Figure 1. Realizations of the Dutch sentence *dat staat hier op deze computer, hè?* "that is on this computer, isn't it?" as produced in a spontaneous speech corpus (a) and as produced in the laboratory (b).

for instance, which words compete for recognition when the intended word is *computer*. According to the literature reviewed above, /k/-initial words such as *companion* should compete for recognition because they share initial overlap. However, it is unclear whether this is still the case when the word

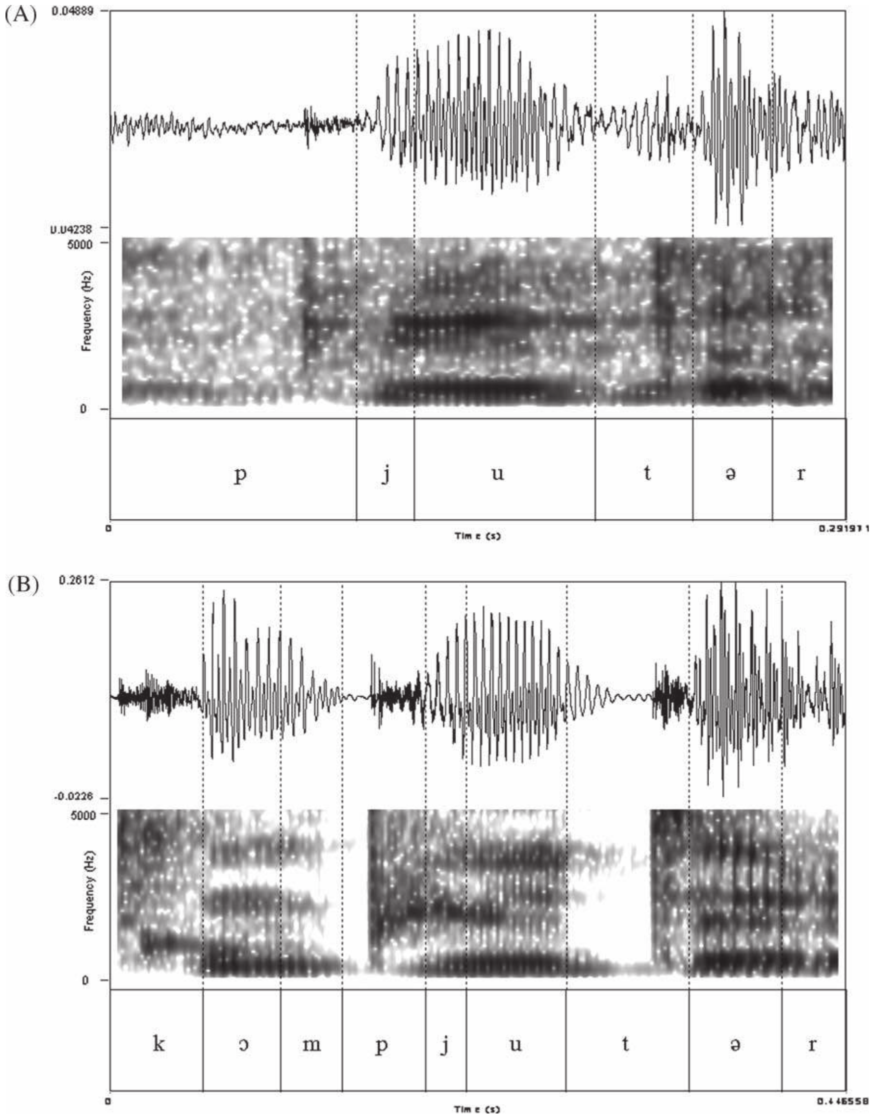

Figure 2. Realizations of the Dutch word *computer* as produced in a spontaneous speech corpus (a) and as produced in the laboratory (b). See text for details.

computer is intended but produced as /pjutər/. In such cases one may predict different competitor sets for canonical and reduced forms.

The aim of Experiment 1 hence is to examine whether phonological competition during casual speech is modulated by the exact phonetic form of the spoken word. In other words, we examine the effect of hearing forms such as the reduced realization [pjutər] or the canonical realization [kɔmpjutər] of *computer* on competition processes during spoken word recognition.

EXPERIMENT 1

We used casual speech in which the same target words appeared in either a canonical or a reduced form. In order to investigate spoken word recognition in casual speech, we have to work with extracts from speech corpora containing ecologically valid examples of casual speech. A disadvantage of using casual speech is that it is difficult to have a similar degree of control over stimulus selection as when creating new stimuli in the laboratory. For example, it is important to establish which acoustic features in the casual speech fragments are precisely produced, a very time-consuming process involving the transcription of a great number of words. For the present study, two independent raters transcribed more than 1400 tokens of 90 words. On the basis of this corpus, we chose words which were produced (at least) once canonically and once in a reduced way.

A requirement for our selected stimuli was that a word exists in the Dutch language that has more phonological onset overlap with the canonical form than with the reduced form, henceforth called a "canonical form" competitor, and another word that has more phonological onset overlap with the reduced form than with the canonical form, henceforth called a "reduced form" competitor. For example, for the canonical form of the English word *computer* [kɔmpjutər] the word *companion* [kɔmpɛnjən] is a canonical form competitor, whereas for the reduced form [pjutər] of *computer* the word *pupil* [pjupɪl] functions as a reduced form competitor.

Note that for 75% of the items (24 out of the 32; see the Appendix), both competitors overlap phonologically at onset. As a result, some reduced form competitors are also to some extent competitors of the target word's canonical form. That is, they function as onset overlap competitors. The canonical form competitor, however, always had more phonological onset overlap with the canonical form than with the reduced form, and the reduced form competitor always had more phonological onset overlap with the reduced form than with the canonical form. For instance, the word *directeur* /dɪrɛktør/ "director" was pronounced canonically and in a reduced way as [dɪktø] in the spontaneous speech corpus. The canonical form competitor *dirigeren* [dɪriχeːrən] "to conduct" shares the first three segments with the canonical form but shares only the first two segments with the reduced form. The reduced form competitor *dictator* [dɪktaːtɔr], however, shares three initial segments with the reduced form but shares only two initial segments with the canonical form. It is therefore crucial to compare the relative strength of the two competitors under different conditions.

The prediction from previous studies using laboratory speech (e.g., Allopenna et al., 1998; McQueen & Viebahn, 2007) therefore is that during listening to canonical forms our reduced form competitors will attract *less* overt attention than our canonical form competitors, because they share no onset overlap (25% of the items) with the target or smaller onset overlap (75% of the items) than the canonical form competitors with the target words. It is, however, unclear what happens during listening to reduced forms. What matters more in such a case? If the acoustic input is crucial, the reduced form competitors should attract more overt attention than our canonical form competitors because in this condition the reduced form competitors overlap to a greater extent with the acoustic signal than the canonical form competitors. If, however, the canonical form of a word is still crucial, even if the input is reduced, then the canonical form competitors should attract more overt attention than the reduced form competitors. This may seem unlikely at first sight, but previous research indicated that listeners may fill in missing phonemes in the input (Kemps, Ernestus, Schreuder, & Baayen, 2004; Samuel, 1996; Warren, 1970) so that the input is restored to its canonical form.

Method

Participants

Twenty-five participants from the Max Planck Institute's subject pool, mostly undergraduates at the Radboud University in Nijmegen, took part in this experiment. All were native speakers of Dutch without any hearing problems and with normal or corrected-to-normal vision. They were paid for their participation.

Materials

We selected 32 polysyllabic, mid-to-high frequency content words for which we could find reduced and canonical pronunciations in the spontaneous speech subcorpora of the Spoken Dutch Corpus (Oostdijk, 2000). For each reduced realization, one or more segments were absent or changed (e.g., [mənеːə] for [bənеːdə] *beneden* "downwards"). Note that there is considerable variation in the reductions (see the Appendix). For example, a reduced form could either deviate from the canonical form in its initial part (first or second segment), such as [mənеːə] for [bənеːdə], or in a later part (third, fourth, or fifth segment), such as [ʋɛs] for [ʋɛtstrɛit] *wedstrijd* "match". The critical criterion for a reduced form was that it shared more initial segments with another existing word than with its own canonical form. In our experiment we compared the recognition of both types of forms in a canonical and a reduced form condition.

All target words were spoken by Dutch (not Flemish) speakers and were not masked by overlapping (speech) sounds. The words of interest were transcribed separately by two transcribers using the software package PRAAT (Boersma, 2001) to observe the signal in auditory and visual, spectrographic form. The independent transcriptions were compared to verify agreement. In case of disagreement, the transcribers were required to reach consensus. The transcribers again examined the spectrogram of the word carefully. Moreover, they listened to the full sentence, parts of the sentence, the target word, and each segment in isolation. Note also that the discrepancies which were encountered were rather minimal. For example, differences were found in where the onset of a segment started.

Note that listeners can hardly recognise reduced word forms on the basis of the acoustic signal for that word alone (e.g., Ernestus et al., 2002). Listeners also find it difficult to recognise highly reduced forms in a limited context in which only the adjacent vowels and intervening consonants around the target word are present. The target forms were therefore presented either in full contexts with several words around the target (e.g., *ook naar beneden die sluit dan aan* "also going downwards this connects then to") or (to reduce the predictability of the target word) in syllable contexts with only the syllables directly neighbouring the target (e.g., *naar beneden die*). Often these single syllables consisted of existing words (e.g., *naar* "to"). We thus compared the recognition of targets in a full versus a syllable context condition.

Note that the context for a canonical item always differed from that of a reduced item, because they occurred in different natural utterances. We conducted an online cloze test to investigate whether the different contexts induce preferences for certain word types (i.e., target, canonical form competitor, reduced form competitor, and distractor), which might have caused confounds in our material. This test measured the predictability of the target word, given the preceding context in canonical and reduced sentences. For both types of sentences, the words preceding the target were presented on the screen. In the first part, participants ($n = 35$) had to finish the sentence freely with three to seven words suitable for the context. In the second part, the sentence was again shown on the screen, but now the potential target, the two competitors, and the distractor were provided. The participants had to rank these words in order of how likely they were to complete the sentence.

In the first open-ended part of the cloze test, participants named the target word on 5.8% of the trials (5.4% in the reduced form sentences and 6.2% in the canonical form sentences; a logistic regression confirms that the small difference is insignificant, $\beta_{\text{Sentence}} = 0.155$, $p > .2$). These results suggest that some target words were indeed somewhat predictable given their linguistic context. The target words were, however, not more predictable in the sentence in which they happened to be reduced. The participants never named a competitor, with the exception of one occurrence of a reduced form competitor ($<1\%$).

In the second forced-choice part, participants rated the target word as the most likely option (in 81.6% of the trials). The mean rank of the target word was hence close to 1, and this did not differ between sentences with reduced forms (1.30) and sentences with canonical forms (1.25). To test whether there was a difference in terms of semantic predictability of the canonical form competitor and the reduced form competitor, we compared the mean rank of both competitors for both types of sentences (i.e., sentences with reduced forms and sentences with canonical forms). The mean rank in all four cases was approximately 3 (canonical form competitors: 3.07 in the canonical form sentences and 2.94 in the reduced form sentences; reduced form competitors: 2.94 in the canonical form sentences and 2.84 in the reduced form sentences). It is hence unsurprising that there were no significant differences as evaluated with a two-by-two repeated measures analysis of variance with competitor and sentence as predictors, F_{Sentence} (1, 30) = 1.68, $p > .1$, all other $F < 1$.

During the experiment the computer screen displayed four different word types: the target word (e.g., *beneden* "downwards"), a phonologically unrelated distractor (e.g., *vakantie* "holiday"), and two types of competitors (see Figure 3). A canonical form competitor shared more initial segments with the canonical form than with the reduced form (e.g., *benadelen* "to disadvantage" [bənaːdeːlə] for [bəneːdə]), whereas a reduced form competitor shared more initial segments with the reduced form than with the canonical form (e.g., *meneer* "mister" [məneːr] for [məneːə]). As a consequence, the

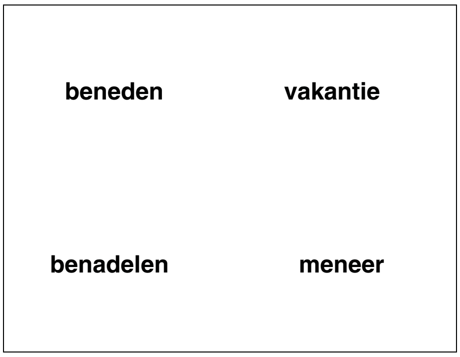

Figure 3. Example of a printed word display presented to participants. The spoken target word in this example was *beneden* "downwards". The four printed words are the target (*beneden* "downwards"), a distractor (*vakantie* "holiday"), a "canonical form" competitor (*benadelen* "to disadvantage"), and a "reduced form" competitor (*meneer* "mister").

display always contained two to three phonologically related words, of which one was the target. To mask this pattern, we used filler trials. On filler trials, displays also contained four printed words, of which two to three were phonologically related. The target appearing in the auditory sentence, however, was not one of the set of phonologically related words on the screen, but rather was the unrelated word. Because of these filler trials, the target word was only phonologically related to other words on the screen on some trials. In this way the fillers discouraged the strategy of limiting one's attention to the phonologically related words on the screen. Moreover, fillers were also included to prevent listeners from predicting the upcoming target word due to repetition of visual displays. The visual displays of the fillers were, as the experimental items, repeated. However, for the fillers, the target word in the visual display did not always end up being the target. For example, the same visual four-word grid (e.g., *familie* "family", *seizoen* "season", *strijden* "to fight", and *strijdlustig* "quarrelsome") was displayed when listeners heard the target word *familie* and when they heard the target word *seizoen*.

We created two different item lists. Both lists had half of the canonical forms and half of the reduced forms with full context and the other half with syllabic context. For example, the first list contained a target word presented in its canonical form in full contexts and the same target word presented in its reduced form in syllable contexts. The second list contained the same target word, but the canonical form was presented in syllable contexts and the reduced form in full contexts. Each subject received one list. The trials in each list were randomized so that each subject received a different order of presentation. Besides 64 fillers (16 fillers in each condition), we also selected 12 practice trials from the spontaneous subcorpora. The positions of the four printed words on the screen were randomized for each participant.

Procedure

Participants were tested individually, seated at a comfortable viewing distance from the computer screen. The eye-tracking system was mounted and calibrated (an SMI EyelinkII system, sampling at 250 Hz). The auditory stimuli were presented over headphones using the NESU software.

Participants received written instructions on the screen. They had to click on the printed word in the visual display representing the word they heard, using the computer's mouse. The location of the printed words was randomized over the four quadrants on the screen to avoid cues to the position of the target. On each trial, the four printed words (24pt Courier) were first presented on the centres of the quadrants on the screen. After 2500 ms, the auditory stimulus was presented. Note that the preview time in the current study was much longer than the one used in the study of McQueen and Viebahn (2007). There are two reasons why we chose to use this longer preview time. First, our target sentences are more complex than their target sentences (e.g., *ook naar beneden die sluit dan aan* "also going downwards that connects then too" versus *Klik op het woord lotje* "Click on the word lottery ticket"). Secondly, the position of the target was unpredictable in our sentences, whereas in McQueen and Viebahn's study it was predictable; that is, the target word always followed after the sentence frame "Click on the word". Hence, we chose the longer preview time to ensure that participants would have enough time to read the four printed words. Preview time is less critical in studies such as that of McQueen and Viebahn (2007) because simple carrier sentences and predictable target word positions allow (presumably) for more concurrent processing of the display. When participants clicked with the mouse on a word, they initiated the next trial. After every five trials, a central fixation cross appeared centred on the screen. Participants were instructed to look at it so that the experimenter could correct drifts in the calibration of the eye tracker. Each participant first completed 12 practice trials. Subsequently, we presented 64 experimental and 64 filler trials (the two lists described above). The experimental session took 20 min.

Design and analysis

For the click responses, we calculated the percentage of correct identifications. The response times (henceforth RTs) on the correct detections were measured from target word offset instead of onset because of the durational differences between the canonical and the reduced form of the same target word. Canonical forms were always longer in duration than reduced forms. The RTs would be confounded if we had measured from target onset. A statistical analysis of the error pattern and the RTs was carried out with linear mixed-effects models. A logistic linking function was used for the error patterns (cf. Dixon, 2008).

For the eye-tracking data, we analyzed only those trials for which the participants clicked on the correct target. We analyzed the data from the right eye of the participants and discarded blinks and saccades. It is estimated that an eye movement is typically programmed about 200 ms before it is launched (Matin, Shao, & Boff, 1993). Thus eye fixations before 200 ms after target onset are unlikely to be driven by acoustic information from the target word. Following Allopenna et al. (1998) and McQueen and Viebahn (2007) we choose to analyze fixation proportions during the 200–800 ms time window after the acoustic onset of the target word. For all four word types (i.e., target, canonical form competitor, reduced form competitor, and distractor) we allowed a deviation of 100 pixels in height and 150 pixels in width around the centre of each printed word in the visual display. The screen resolution was 1024×768 pixels.

For the analysis we first transformed the proportion data with the empirical logit function (Barr, 2008; see formula (6), p. 14), because proportions are problematic in any statistical technique that assumes a linear relation between predictor and outcome variables. From these data, we created three linearly independent measures: (1) looks to the target, to investigate the ease of recognition; (2) mean of looks to both competitors versus looks to the distractor, to assess the existence and strength of overall competition effects; and (3) looks to the canonical form competitor versus looks to the reduced form competitor, to test for the specificity of the competition effects. Note that the latter two are difference measures, so that a difference from zero indicates a preference for one type of stimulus[1].

We tested whether these measures were influenced by Word Form (i.e., canonical vs reduced forms) and Context (i.e., full vs syllable contexts) using linear mixed-effects models (Baayen, Davidson, & Bates, 2008), with participants and items as random effects. This technique is designed to overcome the language-as-fixed-effect problem (Clark, 1973). As Baayen et al. show, the linear mixed effects regression (LMER) technique is more powerful without producing more false positives. Word Form and Context were coded as numeric contrasts (-0.5 and 0.5; cf. Barr, 2008). We estimated p values by using Markov chain Monte Carlo simulations (Baayen et al., 2008). Canonical forms and full context were each separately coded as -0.5, whereas reduced forms and syllable context were each separately coded as 0.5. Thus, we contrasted four conditions: (1) canonical forms in full contexts, (2) canonical forms in syllable contexts, (3) reduced forms in full contexts, and (4) reduced forms in syllable contexts. A positive beta for the variable Word Form (canonical $= -0.5$ vs reduced $= 0.5$) hence indicates that the dependent variable has a higher value for the reduced forms than for the canonical forms, while a positive beta for the Context variable (full $= -0.5$ vs syllable $= 0.5$) indicates that the dependent variable has a higher value for the syllable context condition than the full context condition. Note that the interpretation depends on the dependent measure. In the case of the RT measure, a positive beta would mean longer RTs for reduced forms or for the syllable condition—and hence that these conditions are more difficult, while

[1] Clearly, other contrasts may be of interest, too. For instance, if the competitors are different overall from the distractors, one might wonder if this difference could be driven by one of the competitors. One might then compare each competitor individually with the distractor. There are two reasons not to do this. Firstly, this would generate linearly dependent contrasts and the necessary correction of the statistical tests would reduce the statistical power. Secondly, if only one of the competitors gives rise to competition effects, this should lead to a significant difference between the two competitors. Hence, with the two contrasts—competitors versus distractor and "canonical form" competitor versus "reduced form" competitor—we ascertain whether there are measurable competition effects at all and whether they are mainly carried by one of the competitors.

for target fixation proportions, a positive beta indicates for the Word Form factor that the target is more often fixated in the reduced form condition than in the canonical form condition. For the Context factor, a positive beta indicates more target fixation in the syllable condition than in the full context condition. That is, because greater fixation represents better recognition, but greater RT represents more difficult recognition, effects must be interpreted in opposite directions.

Results

Accuracy and RT measures

Table 1 displays the percentages of mouse-click responses to the different word types and the average RTs per condition. The error analysis showed that participants provided significantly more correct responses for the canonical forms than for the reduced forms ($\beta_{\text{Word Form}} = -5.91$, $p < .01$) as indicated by the negative beta. We found no other main or interaction effects (all $p > .1$).

The analysis of the RT data (measured from target word offset) showed that listeners took significantly more time to recognise reduced versus canonical targets ($\beta_{\text{Word Form}} = 254.5$, $p = .0001$), which is indicated by the positive beta. There were no other main or interaction effects found for this measure (all $p > .1$).

Eye movements

Figure 4 presents the proportion of fixations over time for all four conditions during the first second. In the 200–800 time window we tested the effects of condition on three linearly independent measures: looks to targets (i.e., ease of recognition), looks to competitors versus distractor (i.e., overall competition), and looks to canonical form competitor versus reduced form

TABLE 1
Task performance in Experiment 1

	Canonical forms		*Reduced forms*	
Click responses (%)	*Full context*	*Syllable context*	*Full context*	*Syllable context*
Target	99.8	99.8	92.5	81
Canonical competitor	0	0.3	3	2.5
Reduced competitor	0.3	0	4.5	15.8
Distractor	0	0	0	0.5
RT in ms	977 (467)	974 (368)	1213 (515)	1193 (464)

Note: Standard deviations appear in parentheses.

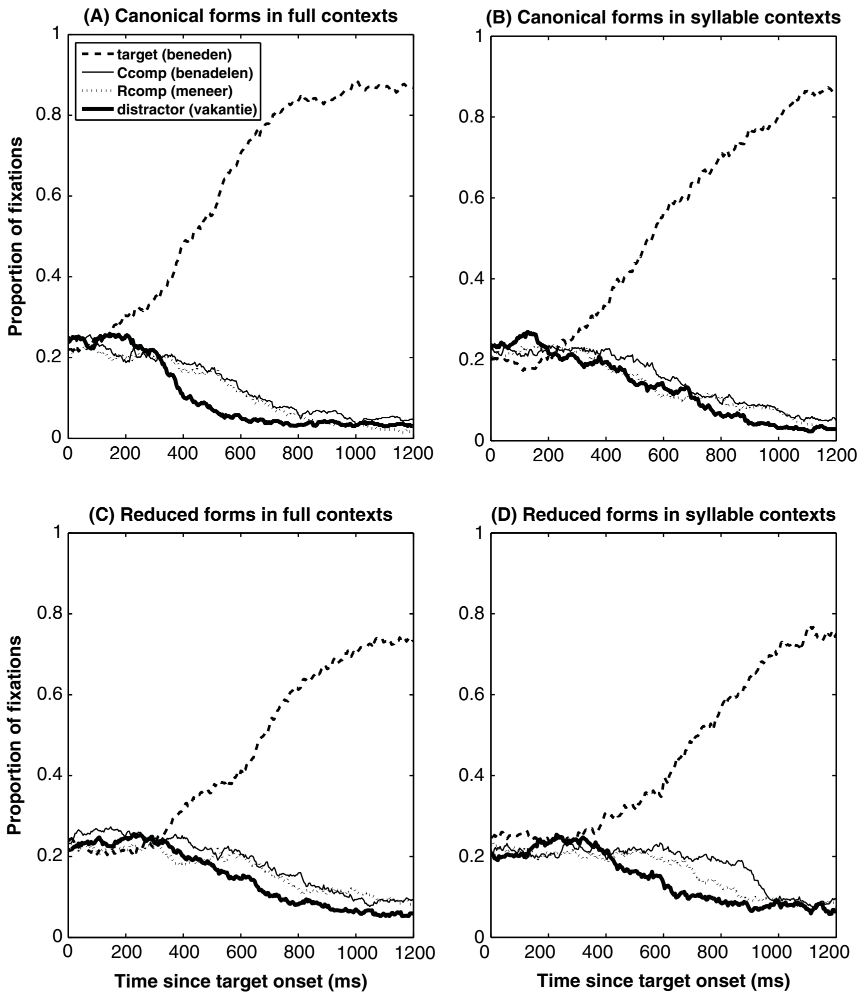

Figure 4. Fixation proportions to the target, the "canonical form" competitor, the "reduced form" competitor, and the distractor in (a) canonical forms presented in full contexts *ook naar beneden die sluit dan aan* "also going downwards this connects then to", (b) canonical forms presented in syllable contexts *naar beneden die* "going downwards this", (c) reduced forms presented in full contexts *buigt het zo af en dan valt het naar beneden dat is echt* "it bends like this and then it falls down, that is really", and (d) reduced forms presented in syllable contexts *naar beneden dat* "going downwards that".

competitor (i.e., specific competition). We first analyzed whether looks to targets differed by condition. We found a main effect of Word Form ($\beta_{\text{Word Form}} = -1.23$, $p_{\text{MCMC}} < .001$). The negative beta reveals that targets attracted more looks in the canonical form condition than in the reduced form condition. Further, we found a main effect of Context ($\beta_{Context} = -0.56$, $p_{MCMC} < .001$). The negative beta reveals that targets attracted more looks in full contexts than in syllable contexts. The analysis also revealed an interaction effect of Word Form by Context ($\beta_{Word\ Form \times Context} = 0.98$, $p_{MCMC} < .05$).

We also analyzed whether the two competitors attracted more looks than the distractor. This analysis (competitors − distractor) showed an effect of overall competition ($\beta_{Intercept} = 0.28$, $p_{MCMC} < .01$), independent of Word Form and Context (all $p_{MCMC} > .1$). A comparison between looks to the canonical form competitor and looks to the reduced form competitor (canonical form competitor − reduced form competitor) showed that the mean difference between looks to the canonical form competitor and looks to the reduced form competitor was not larger than zero; that is, the competitors did not differ from each other ($\beta_{Intercept} = 0.17$, $p_{MCMC} > .1$), and this pattern was not modulated by the phonetic form of the input ($\beta_{Word\ Form} = 0.01$, $p_{MCMC} > .1$).

Discussion

The accuracy data show that it is harder to recognise reduced forms than canonical forms and that listeners benefit from more linguistic context. Similarly, the RT data reveal that listeners need more time to recognise reduced forms than canonical forms. The eye movement data also support the conclusions drawn from the offline data. Listeners looked more often to targets in the canonical than in the reduced conditions. All these findings replicate earlier findings that listeners find it difficult to recognise reduced forms on the basis of the acoustic signal alone (cf. Ernestus et al., 2002; Kemps et al., 2004).

More interestingly, our eye movement data suggest that differences in the exact phonetic form of the acoustic input have no detectable influence on phonological competition. While we anticipated that the phonetic form of the input might not influence the pattern of competition, we had at least expected to replicate the pattern found in other eye-tracking studies (Allopenna et al., 1998; McQueen & Viebahn, 2007), with a preference for onset overlap competitors. In the current case, we therefore had expected that the canonical form competitor would attract more overt attention than the reduced form competitor, at least when the target word was pronounced canonically. The data show, however, that the canonical form competitor attracted as much overt attention as the reduced form competitor when the target word was pronounced canonically (i.e., when hearing *beneden* participants directed as much attention to the canonical form competitor *benadelen* as to the reduced form competitor *meneer*). This finding contrasts with the results from laboratory-speech research that candidates with initial phonological overlap with the target word compete more strongly than candidates with medial or final overlap (e.g., Allopenna et al., 1998; McQueen & Viebahn, 2007). Why do our results using spontaneous speech differ from the results predicted based on laboratory speech?

One possibility is that the style of speech changes listeners' tolerance for mismatch. If listeners are confronted with casual speech (such as the corpus speech in our experiment), they may be more tolerant of acoustic mismatches in the speech signal. As discussed in the Introduction, previous research has interpreted listeners' preference for competitors with an onset overlap over competitors with an offset overlap as evidence for intolerance to acoustic mismatch. It is conceivable however that listeners are more tolerant of such mismatches when the speech style indicates that reductions are possible. In such a listening situation, *overall* match between the input and the candidate words may be the prime influence on phonological competition rather than the amount of *onset* overlap.

Can such an assumption explain why the canonical form and reduced form competitors in Experiment 1 attracted similar levels of attention? In first instance, it seems surprising that the reduced form competitor was as active as the canonical form competitor when the target form was pronounced canonically. An analysis of whether the reduced form and the canonical form competitors differ with respect to their total segmental overlap with the target forms was therefore performed. The overlap of number of phonemes between the reduced form competitors and their target forms was first calculated. This analysis took the segmental order into account, but did not require an exact match of the position. For example, the reduced form competitor *persoon* "person" [pərson]—matching the reduced form [pəsipə]—shares 3 out of 6 phonemes with its target form *principe* "principle" [prɪnsipə]. The shared phonemes between the reduced form competitor and the target form are [p], [r], and [s], which appear in the same order in both words. If the order of the phonemes were not taken into account, the segment [n] as well as the schwa would have also been included in this calculation. The number of matching phonemes was then divided by the total number of phonemes of the reduced form competitor. Similar comparisons were made between the canonical form competitors and their target forms. A t test showed no differences in segmental overlap between the overlap values for the reduced form and the canonical form competitors, $t(62) = -0.18$, $p > .1$. Thus this result is consistent with the notion that overall match between input and candidate words rather than onset overlap is of prime importance when listening to casual speech. This could then explain why there was no difference in looks to the two types of competitors.

A second possibility is that the results of experiment 1 reflect a lack of power. There are two factors that may have reduced experimental power.

First, the cloze test showed that the target words are to some extent predictable. As we needed valid examples of strong reduction, we were forced to use sentences from a speech corpus. It was hence not possible to prevent some predictability of the target word. There is, however, evidence that contextual predictability can constrain lexical activation (e.g., Tabossi, 1988; Tabossi, Colombo, & Job, 1987; Tabossi & Zardon, 1993). A second potential problematic issue is that our manipulation of the canonical form versus reduced form competitor is less strong than the manipulation of onset versus offset overlap in previous experiments (Allopenna et al., 1998; McQueen & Viebahn, 2007). Three quarters of our pairs of canonical form and reduced form competitors both shared the initial segment with the target, and often the difference in the amount of onset overlap was small. This may also make a difference between the two types of competitors less likely. Experiment 2 was designed to test these two possible explanations for the results of Experiment 1.

EXPERIMENT 2

Listeners were again presented with the canonical forms in the full context condition of Experiment 1 (henceforth, canonical forms in casual speech condition). Note that this condition is identical to the canonical form in full context condition of Experiment 1. For a second condition, we re-recorded these same spontaneous sentences under laboratory conditions such that all (target) words were carefully pronounced (henceforth, canonical forms in laboratory speech condition). These conditions hence differ neither in the amount of reduction on the target words—the target word is always fully pronounced—nor in the predictability of the target word based on lexical content in the sentences; the sentences were after all identical. However, there may have been some differences in predictability based on prosody (e.g., speech rate) across conditions (e.g., Dilley & McAuley, 2008). The main difference we focused on in this experiment was the speech style.

Importantly, the experiment was blocked by speech style. The laboratory speech condition was presented before the casual speech condition. These conditions enable us to distinguish the two accounts for the results of Experiment 1. According to the first account, listeners are more tolerant of acoustic mismatch when they hear casual speech (reducing the preference for the canonical form competitor). If this account is correct, the canonical form competitor should attract more overt attention than does the reduced form competitor in the laboratory speech condition but not in the casual speech condition. According to the second account of the data in Experiment 1, the lack of a preference for the canonical form competitors in Experiment 1 was due to lack of power (because of target predictability and/or lack of sufficient difference in onset overlap between canonical form and reduced form competitors). If this account is correct, both conditions should replicate the finding of Experiment 1: Competition effects should be as strong for the canonical form as for the reduced form competitors.

Method

Participants

Twenty-six native Dutch speakers from the Max Planck Institute's subject pool participated in this experiment. They reported normal hearing and vision and were paid for their participation. None of them participated in Experiment 1.

Materials and procedure

We used the same 32 sentences of the canonical forms in full context condition of Experiment 1 for the casual speech condition of Experiment 2. For the laboratory speech condition, we re-recorded these sentences in the laboratory. To do this, the casual speech sentences were orthographically transcribed. We took typical casual speech characteristics such as hesitations (e.g., uh) and repetitions out of the sentences to make them clearer and to make it easier for the speaker to pronounce the target words fully. A female native speaker of Dutch was asked to read the sentences carefully out loud while being recorded in a sound-attenuated booth. Her speech was recorded directly into a computer (sampling rate at 44.1 kHz). The speaker was naive to the purposes of the experiment and did not hear the casual speech sentences beforehand, so she was unable to mimic the speech rate, prosody, or intonation of the original sentences.

We used the same procedure as in Experiment 1. Each participant listened to half of the laboratory and half of the casual speech sentences, counterbalancing this assignment across participants. Note thus that no participant ever heard the same sentence twice. Trials were blocked by speech style (i.e., laboratory vs casual speech). The casual speech block immediately followed after the laboratory speech block. Before each block, participants completed 3 practice trials. Next, 16 experimental and 16 filler trials were presented. The order of presentation within each block was randomized. The total duration of the experimental session was 10 min.

Design and analysis

We examined whether the results were influenced by Speech Style (i.e., laboratory vs casual speech), using linear mixed-effects models with participants and items as random effects (Baayen et al., 2008). Speech Style was coded as a numerical contrast (-0.5 and 0.5; cf. Barr, 2008) in which laboratory speech was coded as -0.5 and casual speech as 0.5. We used the same measures as in Experiment 1 (i.e., errors, RTs, target activation, overall competition, and specific competition), and we analyzed fixation proportions during the 200–800 ms time window after the acoustic onset of the target word (cf. Allopenna et al., 1998; McQueen & Viebahn, 2007).

Results

Accuracy and RT measures

Table 2 shows the error rate and the average RTs per Speech Style. Listeners made no errors. The reaction time analysis (measured from target word offset) revealed that listeners clicked faster on canonical targets in the laboratory speech condition than in the casual speech condition ($\beta_{\text{Speech Style}} = 59.52$, $p < .05$). This can be explained by differences between the two speech styles. In laboratory speech, the sentences were clearly and carefully pronounced by one speaker only, whereas in casual speech the sentences contained more noise and were uttered by multiple speakers (e.g., Bradlow, Nygaard, & Pisoni, 1999; Smiljanic & Bradlow, 2008a).

Eye movements

Figure 5 presents the fixation proportions over time for (a) laboratory speech and (b) casual speech for the first second after target word onset. We analyzed whether looks to targets differed between the two conditions. The analysis showed no difference in target looks between the conditions ($\beta_{\text{Speech Style}} = -0.32$, $p_{\text{MCMC}} > .1$).

An analysis of whether listeners looked more often to the competitors than to the distractor showed an effect of overall competition ($\beta_{\text{Intercept}} = 0.31$, $p_{\text{MCMC}} < .01$). The significant intercept indicates that the

TABLE 2
Task performance in Experiments 2 and 3

	Experiment 2		*Experiment 3*	
	Canonical forms		*Casual speech*	
	Laboratory speech	*Casual speech*	*Canonical forms*	*Reduced forms*
Correct (%)	100	100	99	93
RT in ms	906 (536)	975 (475)	1008 (430)	1192 (479)

Note: Standard deviations appear in parentheses.

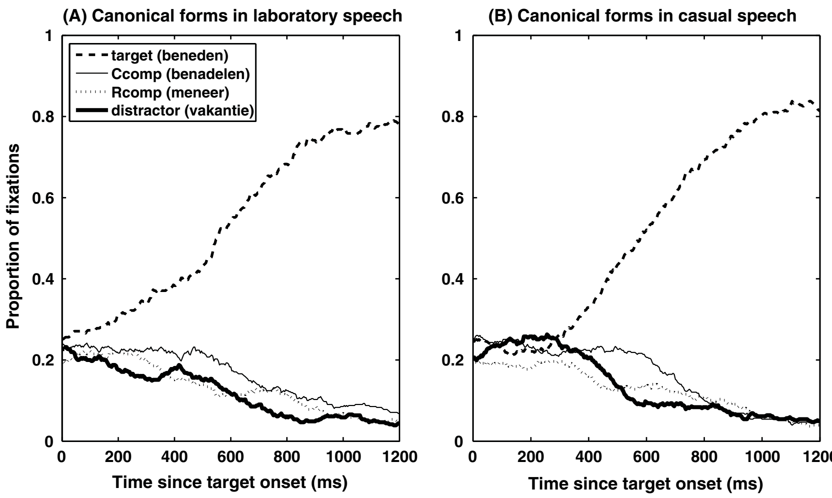

Figure 5. Fixation proportions to the target, the "canonical form" competitor, the "reduced form" competitor, and the distractor in (a) canonical forms in casual speech and (b) canonical forms in laboratory speech for the sentence *ook naar* <u>*beneden*</u> *die sluit dan aan* "also going downwards, this connects then to".

mean difference between looks to competitors and looks to distractor is larger than zero, and hence that the competitors attracted more looks than the distractors. A marginally significant difference was found between the laboratory speech condition and the casual speech condition ($p_{MCMC} = .06$).

Finally, a comparison between looks to the competitors (canonical form competitor − reduced form competitor) revealed that the canonical form competitor attracted more looks than the reduced form competitor ($\beta_{Intercept} = 0.53$, $p_{MCMC} < .001$). The significant intercept shows that the mean difference between looks to the canonical form competitor and looks to the reduced form competitor is larger than zero, indicating that the canonical form competitor is more strongly activated than the reduced form competitor. No main effect was found for Speech Style ($p_{MCMC} > .1$).

In sum, the data of Experiment 2 are very clear. We observed a significant preference for the canonical form competitor over the reduced form competitor in both the laboratory and the casual speech condition. There were no differences on any other measure between the two conditions of Experiment 2.

Discussion

The data of Experiment 2 reveal a preference for the canonical form competitor in both the casual speech condition and the laboratory speech condition. These results are both expected and unexpected. On one hand, this pattern replicates earlier results showing a preference for onset over offset overlap competitors (cf. Allopenna et al., 1998, McQueen & Viebahn, 2007). As such, it is expected. Therefore, it is possible with the present materials—despite predictable targets and relatively small differences in onset overlap between the two types of competitors—to obtain differentiating competition effects. We observed a clear preference for the canonical form competitor in both conditions of Experiment 2.

On the other hand, the results are unexpected because we had predicted either no preference in both conditions or a preference for the canonical form competitor only in the laboratory speech condition. The latter prediction was driven by the expectation that the casual speech condition of Experiment 2 would replicate the results of the full context–canonical form condition of Experiment 1 (i.e., no preference for the canonical form competitor). This expectation was based on the fact that the stimuli in these two conditions were *identical*. The only difference between the two conditions was that in Experiment 1, reduced forms were presented randomly intermixed with the canonical forms, while in Experiment 2, participants heard only canonical pronunciations.

To ascertain that the difference caused by the experimental situation is real, we performed a statistical comparison between the two experiments. We compared the results of the casual speech condition of Experiment 2 with the canonical forms in full context condition of Experiment 1. To reiterate, the stimuli in these two conditions are identical, and only the experimental context varies. This cross-experiment analysis examined whether the results were different for the canonical forms in Experiment 1 and Experiment 2, using linear mixed-effects models with participants and items as random effects (Baayen et al., 2008). Experiment was coded as a numerical contrast (-0.5 and 0.5; cf. Barr, 2008) in which Experiment 1 was coded as -0.5 and Experiment 2 as 0.5. We used the same measures and analyzed the same time window as in the within-experiment analysis.

The RT analysis revealed no difference between the two identical conditions ($\beta_{Experiment} = -2.17$, $p > .1$). The analysis of target fixations showed that listeners looked more often to the target words in the canonical forms in full context condition of Experiment 1 than in the casual speech condition of Experiment 2 ($\beta_{Experiment} = -0.92$, $p_{MCMC} < .01$) as indicated by the negative regression weight. There was also an effect of overall competition ($\beta_{Intercept} = 0.83$, $p_{MCMC} < .05$), independent of the experiment ($p_{MCMC} > .1$). Importantly, a comparison in strength between the two types of competitors showed a significant difference between the two conditions ($\beta_{Experiment} = 0.59$, $p_{MCMC} < .05$). The positive beta indicates that listeners looked more often to the canonical form competitor than to the reduced form competitor in the casual speech condition of Experiment 2 than in the canonical forms in full context condition of Experiment 1. This is a crucial result. It shows that there is a preference for the canonical form competitor in Experiment 2, but not in Experiment 1.

However, the cross-experiment comparison showed another difference, with more looks to targets in Experiment 1 than in Experiment 2. Interestingly, these looks to targets do not decrease the looks to the competitors in Experiment 1. This indicates that the participants are not faster in recognizing the targets in Experiment 1, a conclusion in line with the nearly identical RT results. A possible explanation may be the presence of the syllable context condition, in which the target occurs shortly after the speech onset. This may have led participants to be more eager in looking for a target during the (longer) sentences in the full context condition.

Note that differences for identical conditions in different experimental contexts are not unprecedented (for a classical example, see Van der Heijden, Hagenaar, & Bloem, 1984). To account for such effects, modelling approaches (see Phaf, Van der Heijden, & Hudson, 1990) typically assume that participants adjust their processing strategy to the most difficult condition. Such an interpretation fits well with our data. In Experiment 1, listeners had to deal with reduced forms and, therefore, put less confidence in mismatches between input and canonical form. This led to similar levels of activation for canonical form and reduced form competitors. In Experiment 2, listeners encountered little reduction and hence took mismatches more seriously, leading to a preference for canonical form over reduced form competitors.

This raises the question of how the presence versus absence of the sentences with reduced forms could have such a strong impact on the spoken word recognition system. To answer that question, we analyzed how faithfully the sentences were produced overall. That is, the complete sentences, including content and function words, were transcribed independently by the first and the second author. From these transcriptions and the sentence-level transcriptions with the intended words provided in the CGN corpus, we calculated how many underlying phonemes in a given sentence were realized (coded as 0 = deleted, 0.5 = changed, e.g., a /b/ realized as /m/, 1 = fully realized). The measurements from the two transcribers agreed reasonably ($r = 0.91$) and showed that there was much less reduction overall in the casual speech sentences with canonical forms (93% of the phonemes realized) than in the sentences with the reduced forms [78% of the phonemes realized, $t(31) = 11.8$, $p < .001$]. It is conceivable that the absence versus presence of sentences with such massive reductions in Experiment 2 and Experiment 1, respectively, may have influenced the strategy that listeners were using to recognize spoken words.

Rather than relying only on our current post hoc explanation and cross-experiment analyses, we conducted an additional experiment to provide further evidence that our interpretation of the data is correct. Experiment 3 was designed to test directly that listeners are more tolerant of acoustic mismatches in a listening situation in which they encounter reduced speech. We again presented the corpus sentences with the canonical forms of the target words (casual speech condition of Experiment 2) but now intermixed with reduced forms. If our interpretation of the data of Experiments 1 and 2 is correct, we should again observe no preference for the canonical form competitor. In other words, the same target words in canonical form, which led to increased eye gaze to canonical form competitors when intermixed with clearly spoken sentences in Experiment 2, should produce no such preference when intermixed with sentences containing reduced forms. Additionally, this experiment also allows us to see whether the presence of a syllable context condition triggers faster looks to the target, as hypothesized above. If so, we should see a slower convergence on the targets in this experiment than in Experiment 1.

EXPERIMENT 3

Method

Participants

Twenty-four native Dutch speakers from the Max Planck Institute's subject pool participated in this experiment. They reported normal hearing and vision and were paid for their participation. None of them took part in the previous experiments.

Materials and procedure

We used the same 32 sentences of the casual speech condition of Experiment 2 (henceforth, canonical forms in casual speech) and intermixed these sentences with the reduced forms in full context condition of Experiment 1 (henceforth, reduced forms in casual speech). The same procedure was used as in the previous experiments.

Participants were exposed to either the canonical or the reduced form of each target word. The four-word display thus appeared only once, as in Experiment 2, in the course of the experiment. Note that this presentation is different from Experiment 1, in which the four-word display was presented twice to participants. An anonymous reviewer was concerned that the increased target predictability in Experiment 1 might have reduced participants' consideration of either competitor, thereby washing out any differences in their consideration of either competitor as a function of the phonetic realization of the target word. Experiment 3 tested this possibility.

The presentation order of the stimuli in Experiment 3 was randomized. Participants started with three practice trials after which 32 experimental and 32 filler trials were presented. The total duration of the experimental session lasted 15 min.

Design and analysis

We examined whether the results were influenced by Word Form (i.e., canonical vs reduced) using linear mixed-effects models with participants and items as random effects (Baayen et al., 2008). Word Form was coded as a numerical contrast (-0.5 and 0.5; cf. Barr, 2008), in which canonical forms were coded as -0.5 and reduced forms as 0.5. We used the same measures as in the previous experiments (i.e., errors, RTs, target activation, overall competition, and specific competition), and we analyzed fixation proportions during the 200–800 ms time window after the acoustic onset of the target word (cf. Allopenna et al., 1998; McQueen & Viebahn, 2007).

Results

Accuracy and RT measures

Table 2 shows the error rate and the average RTs per Word Forms. Listeners made more errors in the reduced form condition than in the canonical form condition ($\beta_{\text{Word Form}} = -4.13$, $p < .05$) as indicated by the negative beta. The reaction time analysis (measured from target word offset) showed that listeners took significantly more time to recognize reduced versus canonical targets ($\beta_{\text{Word Form}} = 211.4$, $p < .001$).

Eye movements

Figure 6 presents the fixation proportions over time for (a) canonical forms in casual speech and (b) reduced forms in casual speech for the first second after target word onset. We first analyzed whether looks to targets differed by conditions. We found a main effect of Word Form ($\beta_{\text{Word Form}} = -0.52$, $p_{\text{MCMC}} < .05$), indicating that listeners looked more often to targets in the canonical form condition than in the reduced form condition.

Secondly, we analyzed whether there is an effect of overall competition (competitors – distractor). We found an effect of overall competition ($\beta_{\text{Intercept}} = 0.40$, $p_{\text{MCMC}} < .01$), independent of Word Form ($p_{\text{MCMC}} > .1$). The significant intercept indicates that the mean difference between looks to competitors and looks to distractor is larger than zero, and hence that the competitors attracted more looks than did the distractors.

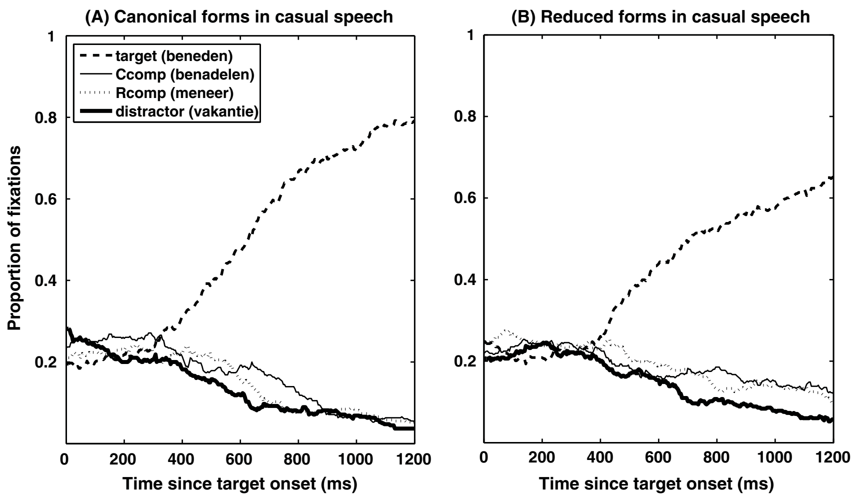

Figure 6. Fixation proportions to the target, the "canonical form" competitor, the "reduced form" competitor, and the distractor in (a) canonical forms in casual speech *ook naar beneden die sluit dan aan* "also going downwards, this connects then to" and (b) reduced forms in casual speech *buigt het zo af en dan valt het naar beneden dat is echt* "it bends like this and then it falls down, that is really".

Most importantly, we compared listeners' fixations to the competitors (canonical form competitor − reduced form competitor). The analysis revealed no difference between looks to the canonical form competitor and looks to the reduced form competitor ($\beta_{\text{Intercept}} = 0.06$, $p_{\text{MCMC}} > .1$), and this pattern was not modulated by Word Form ($p_{\text{MCMC}} > .1$).

We also compared the "canonical forms in casual speech" condition from this experiment with that of Experiment 1 to test whether there were more looks to targets in Experiment 1. The analysis indeed revealed a difference with more looks to targets in Experiment 1 than in Experiment 3 ($\beta_{\text{Experiment}} = -1.30$, $p_{\text{MCMC}} < .0001$), but no other differences (overall competition: $\beta_{\text{Experiment}} = 0.08$, $p_{\text{MCMC}} > .1$; specific competition: $\beta_{\text{Experiment}} = 0.20$, $p_{\text{MCMC}} > .1$). That is, there were more looks to targets in Experiment 1 than in Experiment 3, but the patterns of competition were equivalent.

Experiment 3 is thus consistent with our interpretation of the results of the first two experiments. The competitors that are activated upon hearing a given word not only depend on the sentential context and the phonetic form of that word. The data from the present experiments are strong evidence that competition processes are also influenced by the amount of reduction the listener encounters in a given listening situation. In addition, the data show that the large amount of looks to targets observed in Experiment 1 may be due to the addition of the syllable context condition, in which the target appears short after speech onset. Finally, the results of Experiment 3 rule out that the absence of a preference for the canonical form competitor given a canonical form in the auditory input—as observed in Experiments 1 and 3—was due to repetition of target words. Repetition of target words in the visual display occurred in Experiment 1 but not in Experiment 3, yet the absence of a preference for the canonical form competitor given a canonical form in the input was found in both experiments.

GENERAL DISCUSSION

In three eye-tracking experiments, we examined whether spoken word recognition in casual speech is different from spoken word recognition in laboratory speech. Participants heard spoken sentences while they saw four printed words in a visual display. Sentences originated from a spontaneous speech corpus or from carefully pronounced laboratory speech. Eye movements were measured while participants listened to sentences containing a critical target word—also presented visually on the screen—which was realized in its canonical or in its reduced form.

Experiment 1 examined whether phonological competition is modulated by the exact phonetic form of the target word (canonical vs reduced). The data showed that on either hearing the reduced realization [pjutər] or the canonical realization [kɔmpjutər] of *computer*, listeners directed their attention to a similar degree to the same competitors. We interpreted this finding as indicating that when listening to reduced speech, listeners are more tolerant of acoustic mismatches.

Experiment 2 was designed to further investigate this hypothesis. We compared the recognition of canonical forms in laboratory speech with casual speech. Importantly, in contrast to Experiment 1, we did not include any reduced forms in the experiment. We observed that in such a listening situation there was no influence of speech style on competition processes. Listeners directed significantly more overt attention to the canonical form competitor than the reduced form competitor not only in the laboratory speech condition but also in the casual speech condition of Experiment 2. In the identical condition of Experiment 1, there was no such bias [see Figure 4(a)]. The only difference between the experiments was that in Experiment 1, the canonical forms were intermixed with reduced forms, whereas in Experiment 2, listeners only heard carefully articulated fully pronounced canonical forms. In Experiment 2, participants first listened to a block of laboratory speech before they listened to a block of casual speech. This suggests that participants adjusted to listening to carefully pronounced canonical forms.

The results of Experiment 3 provided further support for the account that speech-intrinsic variation such as reduced speech affects the recognition of clearly articulated words. In Experiment 3, in which the canonical forms of Experiment 2 were intermixed with reduced forms, we replicated the competition pattern of Experiment 1. Once again there was no difference between listeners' fixations to canonical form and reduced form competitors. This shows that in a listening situation with casual speech which includes a great deal of reduced forms, listeners are more tolerant to acoustic mismatches between input and canonical form. As a consequence, medial and offset overlap competitors become stronger candidates in casual speech (than in a listening situation in which listeners are exposed to carefully articulated fully pronounced speech only) because the initial mismatch is less bad.

However, the finding that reduced speech leads to a reduction of the mismatch criterion could also be interpreted the other way round. That is, laboratory speech could have *strengthened* the mismatch criterion. In Experiment 2, participants listened to a block of laboratory speech prior to a block of casual speech. This first block of carefully articulated speech could have triggered a mechanism that decreased listeners' tolerance of mismatches. Consequently, listeners looked more at the canonical form than the reduced form competitors. The current data seem to show flexibility to how listeners treat mismatches. In the presence of laboratory speech, mismatches may play a larger role than in the presence of strongly reduced speech.

It is important to note that we did not compare cohort with rhyme competitors as in the studies of Allopenna et al. (1998) and McQueen and Viebahn (2007). In our materials, some competitors shared onset overlap *both* with the canonical form (e.g., canonical form competitor *wetboek* "statute book" for *wedstrijd* "match") and the reduced form (e.g., "reduced form" competitor *wesp* "wasp" for *wedstrijd* "match"). Importantly, however, the reduced form competitor always deviated from the canonical form by more segments than the canonical form competitor (see the Appendix). With such an item set, kept constant across all experiments, we found results similar to those of Allopenna et al. and McQueen and Viebahn in the laboratory speech condition and the casual speech condition of Experiment 2. This shows that our weaker manipulation of canonical form versus reduced form competitors was still able to produce qualitatively similar results relative to the manipulation of onset versus offset competitors in these earlier experiments.

Why do speech reductions change the dynamics of spoken word recognition? Interestingly, previous research on assimilation suggests that listeners are also more tolerant of phonological changes leading to mismatches if the context allows the phonological change. Gaskell and Marslen-Wilson (1996), for example, examined how listeners deal with assimilations (e.g., "lean bacon" → "leam bacon"). In a cross-modal priming experiment, they found an effect of priming for unassimilated (e.g., "lean") and assimilated auditory primes (e.g., "leam") presented in isolation. A second experiment presented the assimilated tokens in two contexts: a viable context (e.g., "leam bacon"), allowing for assimilation, or an unviable context (e.g., "leam gammon"). In the viable context, a priming effect was found for both assimilated and unassimilated primes. However, in the unviable context, the assimilated primes showed reduced priming effects as compared to unassimilated primes.

Mitterer and Blomert (2003) also investigated how listeners cope with the variation caused by place assimilation in continuous spoken word recognition. Participants had to indicate whether the Dutch word *tuin* "garden" was pronounced canonically or as [tœym] due to nasal place articulation. These target words were presented in a context which allowed assimilation (*tuinbank* "garden bench") or in a context that did not (*tuinstoel* "garden chair"). In the viable context condition, listeners (incorrectly) perceived the target *tuimbank* as *tuinbank* (see Coenen, Zwitserlood, & Bölte, 2001; Gaskell & Marslen-Wilson, 1998; Gow, 2003; Mitterer, Csépe, Honbolygo, & Blomert, 2006, for similar findings). These results suggest that listeners tolerate variation in the input if the context allows the variation.

Our results indicate another form of mismatch tolerance based on speech-intrinsic factors, but on a much larger timescale. The experiments on assimilation showed that listeners take the immediately following context—in the range of fractions of seconds—into account to licence a mismatch between input and canonical form. Our experiments reveal that a general tolerance for mismatch can also be based on the time range of minutes. If participants listen to a mix of canonical and reduced forms embedded in casual speech sentences (as in Experiments 1 and 3), listeners tolerate onset mismatches to a greater extent than when listeners are first confronted with speech that is carefully produced in a laboratory setting before they listen to casual speech (as in Experiment 2).

The present findings also fit well with recent data about the influence of extrinsic factors on spoken word recognition. Huettig and McQueen (2009) investigated listener flexibility by comparing the dynamics of the spoken word recognition process in clear speech and speech disrupted by radio noise. In Huettig and McQueen's Experiment 1, Dutch participants listened to clearly articulated spoken Dutch sentences which each included a critical word, while their eye movements to four visual objects were measured. There were two critical conditions. In the first, the objects included a cohort competitor (e.g., *parachute*) with the same onset as the critical spoken word (e.g., *paraplu*, "umbrella") and three unrelated distractors. In the second condition, a rhyme competitor (e.g., *hamer*, "hammer") of the critical word (e.g., *kamer*, "room") was present in the display, again with three distractors. Their Experiment 2 was identical to their Experiment 1, except that phonemes in the spoken sentences were replaced with radio-signal noises (as in AM radio listening conditions). Importantly (as in our present study) the critical words (and the immediately surrounding words) were not changed. Huettig and McQueen observed a significant experiment by competitor-type interaction. In Experiment 1 (no noise), participants fixated both kinds of competitors more than unrelated distractors, but there were more and earlier looks to cohort competitors than to rhyme competitors (as in the Allopenna study). In Experiment 2 (with radio noise), participants still fixated cohort competitors more than rhyme competitors but the early cohort effect was reduced and the rhyme effect was stronger and occurred earlier.

Their results suggest that speech-extrinsic factors such as AM radio noise also change the dynamics of spoken word recognition. Thus, the well-attested finding of stronger reliance on word onset overlap in speech recognition appears to be due in part to the use of carefully articulated fully pronounced and noise-free speech in most experiments. When onset information becomes less reliable, either because of speech-intrinsic factors such as reduced speech or speech-extrinsic factors such as noise, listeners appear to depend on it less. A core feature of the speech recognition system thus appears to be its flexibility.

We conclude that the dynamics of spoken word recognition are influenced by the speech style. When listening to strongly reduced speech changes, listeners penalize acoustic mismatches less strongly than when listening to fully pronounced laboratory speech. Our data demonstrate that speech-intrinsic variation such as reduced speech modulates phonological competition. Flexibility to adjust to speech-intrinsic (and speech-extrinsic) factors is a key feature of the spoken word recognition system.

REFERENCES

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*, 419–439.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

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

Boersma, P. (2001). PRAAT, a system for doing phonetic by computer. *Glot International*, *5*(9/10), 341–345.

Bradlow, A. R., Nygaard, L. C., & Pisoni, D. B. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception & Psychophysics*, *61*(2), 206–219.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*, 335–359.

Coenen, E., Zwitserlood, P., & Bölte, J. (2001). Variation and assimilation in German: Consequences of assimilation for word recognition and lexical representation. *Language and Cognitive Processes*, *16*, 535–564.

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*, 193–210.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language. A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *6*, 84–107.

Cutler, A. (1998). The recognition of spoken words with variable representation. (Ed.), *Proceedings of the ESCA Workshop on Sound Pattern of Casual Speech* (pp. 83–92). Aix-en-Provence, France: ESCA.

Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review*, *12*, 453–459.

Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language*, *59*, 294–311.

Dixon, P. (2008). Models of accuracy in repeated-measures design. *Journal of Memory and Language*, *59*, 447–456.

Ernestus, M. (2000). *Voice assimilation and segment reduction in casual Dutch: A corpus-based study of the phonology–phonetics interface*. Utrecht, The Netherlands: LOT.

Ernestus, M., Baayen, H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language*, *81*, 162–173.

Gaskell, M. G., & Marslen-Wilson, W. D. (1996). Phonological variation and interference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(1), 144–158.

Gaskell, M. G., & Marslen-Wilson, W. D. (1998). Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(2), 380–396.

Goldinger, S. D., Luce, P. A., & Pisoni, D. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language*, *28*(5), 501–518.

Gow, D. W. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception and Psychophysics*, *65*(4), 575–590.

Huettig, F., & Altmann, G. T. M. (2004). The online processing of ambiguous and unambiguous words in context: Evidence from head-mounted eye-tracking. In M. Carreiras & C. Clifton (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERP and beyond* (pp. 187–207). New York: Psychology Press.

Huettig, F., & Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, *96*, B23–B32.

Huettig, F., & Altmann, G. T. M. (2007). Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness. *Visual Cognition*, *15*(8), 985–1018.

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language*, *57*, 460–482.

Huettig, F., & McQueen, J. M. (2009). *AM radio noise changes the dynamics of spoken word recognition*. Paper presented at the AMLaP 2009 conference in Barcelona, Spain.

Johnson, K. (2004). Massive reduction in conversational American English. In K. Yoneyama & K. Maekawa (Eds.). *Casual speech: Data and analysis. Proceedings of the 1st Session of the 10th International Symposium*, Tokyo, Japan (pp. 29–54).

Kemps, R., Ernestus, M., Schreuder, R., & Baayen, H. (2004). Processing reduced word forms: The suffix restoration effect. *Brain and Language*, *90*, 117–127.

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

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

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

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1–86.

McQueen, J. M., & Cutler, A. (2001). Spoken word access processes: An introduction. *Language and Cognitive Processes*, *16*(5/6), 469–490.

McQueen, J. M., & Viebahn, M. (2007). Tracking recognition of spoken words by tracking looks to printed words. *The Quarterly Journal of Experimental Psychology*, *60*(5), 661–671.

Mehta, G., & Cutler, A. (1988). Detection of target phonemes in casual and read speech. *Language and Speech*, *31*(2), 135–156.

Mitterer, H., & Blomert, L. (2003). Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception and Psychophysics*, *65*(6), 956–969.

Mitterer, H., Csépe, V., Honbolygo, F, & Blomert, L. (2006). The recognition of phonologically assimilated words does not depend on specific language experience. *Cognitive Science*, *30*, 451–479.

Norris, D. G. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*, 189–234.

Oostdijk, N. (2000). The Spoken Dutch Corpus Project. *The ELRA Newsletter*, *5*, 4–8.

Phaf, R. H., Van der Heijden, A. H. C., & Hudson, P. T. W. (1990). SLAM: A connectionist model for attention in visual selection tasks. *Cognitive Psychology*, *22*, 273–341.

Samuel, A. (1996). Phoneme restoration. *Language and Cognitive Processes*, *11*, 647–653.

Smiljanic, R., & Bradlow, A. R. (2008a). Temporal organization of English clear and plain speech. *Journal of the Acoustical Society of America*, *124*(5), 3171–3182.

Tabossi, P. (1988). Accessing lexical ambiguity in different types of sentential context. *Journal of Memory and Language*, *27*, 324–340.

Tabossi, P., Colombo, L., & Job, R. (1987). Accessing lexical ambiguity: Effects of context and dominance. *Psychological Research*, *49*, 161–167.

Tabossi, P., & Zardon, F. (1993). Processing ambiguous words in context. *Journal of Memory and Language*, *32*, 359–372.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632–1634.

Van der Heijden, A. H. C., Hagenaar, R., & Bloem, W. (1984). Two stages in postcategorical filtering and selection. *Memory and Cognition*, *12*, 458–469.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, *167*, 392–393.

Weber, A., Melinger, A., & Lara Tapia, L. (2007). The mapping of phonetic information to lexical representation in Spanish: Evidence from eye movements. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007)* (pp. 1941–1944). Dudweiler: Pirrot.

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition*, *32*, 25–64.

Appendix: Experimental items: canonical and reduced realizations with their ''canonical form'' and ''reduced form'' competitors, respectively

Target word	*Canonical form*	*''Canonical form'' competitor*	*Reduced form*	*''Reduced form'' competitor*
afspraak "appoinment"	[ɑfspraːk]	[ɑfspreːkə]	[ɑspʋaː]	[ɑspiraːtsi]
apparaat "apparatus"	[ɑpaːraːt]	[ɑpəritif]	[ɔpraːt]	[ɔpraːpə]
beneden "downwards"	[bəneːdə]	[bənaːdeːlə]	[məneːə]	[məneːr]
bijvoorbeeld "for example"	[bəvoːrbelt]	[bəvoːrɛχtə]	[vɔlt]	[vɔlt]
computer	[kɔmpjutər]	[kɔmpətɛnt]	[pjutər]	[putsə]
concert	[kɔnsɛrt]	[kɔnsɛrt]	[kɔzɛr]	[kɔstbaːr]
concurrent "competitor"	[kɔnkyrɛnt]	[kɔnkur]	[kɔ̃krɛnt]	[kɔnkreːt]
constant "constant"	[kɔnstɑnt]	[kɔnsɛntraːtsi]	[kɔzən]	[kozɛin]
cultuur "culture"	[kʏltyr]	[kʏltʏs]	[kɔmtym]	[kɔmst]
december "December"	[deːsɛmbər]	[deːkaːn]	[eːsɛmər]	[eːtaːʒə]
dinsdag "Tuesday"	[dɪnsdɑχ]	[dɪŋən]	[dɪzɑ]	[dɪzɑjn]
directeur "director"	[dɪrɛktør]	[dɪriχeːrə]	[dɪktø]	[dɪktaːtɔr]
kweekschool "school"	[kʋeːksχɔl]	[kʋeːkə]	[kʋeːsχɔl]	[kʋeːstə]
maandag "Monday"	[maːndɑχ]	[maːnd]	[maːnz]	[maːnzaːt]
ogenblik "moment"	[oːχənblɪk]	[oːχkɑs]	[blɪk]	[blɪk]
oktober "October"	[ɔktoːbər]	[ɔktoːpʏs]	[toːʋər]	[toːvərə]
overheid "goverment"	[oːvərhɛit]	[ovərhɛmt]	[oːvərɛi]	[oːvərɛint]
parlement "parliament"	[pɑrləmɛnt]	[pɑrkeːrə]	[pɑləmɛn]	[pɑlɛt]
plaatsen "to place"	[plaːtsə]	[plaːtsnaːm]	[plaːs]	[plaːseːbo]
positie "position"	[poːζitsi]	[poːζeːrə]	[psitsi]	[psiχə]
prestatie "performance"	[prəstaːtsi]	[prɛstiʒə]	[pəstaːsi]	[pɛsimɪst]
principe "principle"	[prɪnsipə]	[prɪns]	[pəsipə]	[pərson]
publiek "audience"	[pyblik]	[pybliseːrə]	[ʋlik]	[ʋil]
redelijk "reasinable"	[reːdələk]	[reːdərɛi]	[reːlək]	[reːlikʋi]
rekenen "to count"	[reːkənə]	[reːks]	[reːχən]	[reːχən]
rotzooi "garbage"	[rɔtsoːj]	[rɔts]	[rɔsɪ]	[rɔsɪχ]
standaard "default"	[stɑndaːrt]	[stɑndplaːts]	[stɑ̃əd]	[stɑŋ]
standpunt "point of view"	[stɑntpʏnt]	[stɑntfɑstəχ]	[stɑmpʏ]	[stɑmpɔt]
station "station"	[stɑtʃɔn]	[staːtʏs]	[saːʃɔn]	[saːtɛin]
tandarts "dentist"	[tɑndɑrts]	[tɑndpɑstaː]	[tɑz]	[tɑs]
wedstrijd "match"	[ʋɛtstrɛit]	[ʋɛtbuk]	[ʋɛs]	[ʋɛsp]
winter	[ʋɪntər]	[ʋɪntstil]	[ʋɪndə]	[ʋɪndə]