



# Perceptual adaptation to segmental and syllabic reductions in continuous spoken Dutch



Katja Poellmann<sup>a,b,\*</sup>, Hans Rutger Bosker<sup>a</sup>, James M. McQueen<sup>a,c</sup>, Holger Mitterer<sup>a,1</sup>

<sup>a</sup> Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands

<sup>b</sup> International Max Planck Research School for Language Sciences, P.O. Box 310, 6500 AH Nijmegen, The Netherlands

<sup>c</sup> Behavioural Science Institute and Donders Institute for Brain, Cognition & Behaviour, Radboud University Nijmegen, Postbus 9104, 6500 HE Nijmegen, The Netherlands

## ARTICLE INFO

### Article history:

Received 8 March 2013

Received in revised form

18 June 2014

Accepted 24 June 2014

Available online 26 July 2014

### Keywords:

Segmental reduction

Syllabic reduction

Adaptation

## ABSTRACT

This study investigates if and how listeners adapt to reductions in casual continuous speech. In a perceptual-learning variant of the visual-world paradigm, two groups of Dutch participants were exposed to either segmental (/b/ → [u]) or syllabic (*ver-* → [f:]) reductions in spoken Dutch sentences. In the test phase, both groups heard both kinds of reductions, but now applied to different words. In one of two experiments, the segmental reduction exposure group was better than the syllabic reduction exposure group in recognizing new reduced /b/-words. In both experiments, the syllabic reduction group showed a greater target preference for new reduced *ver*-words. Learning about reductions was thus applied to previously unheard words. This lexical generalization suggests that mechanisms compensating for segmental and syllabic reductions take place at a prelexical level, and hence that lexical access involves an abstractionist mode of processing. Existing abstractionist models need to be revised, however, as they do not include representations of sequences of segments (corresponding e.g. to *ver-*) at the prelexical level.

© 2014 Elsevier Ltd. All rights reserved.

## 1. Introduction

Phonological reductions, that is, “the articulatory weakening or complete deletion of segments and syllables” (Ernestus, 2009, p. 1875), are very common in casual speech (e.g., Johnson, 2004; Patterson, LoCasto, & Connine, 2003). A native speaker of Dutch might for instance articulate the /b/ in *baron* “baron” as a labio-dental approximant ([u]) or shorten the prefix *ver-* in *verlangen* “desire” to [f:]. Despite these deviations from canonical pronunciations, speakers and listeners do not seem to be hampered in their communication. It is thus an intriguing question how listeners overcome such distortions in order to be able to recognize the originally intended words. The present study tests whether listeners can adapt to different types of reduction. Having heard a sufficient number of either segmental (/b/ → [u]) or syllabic (*ver-* → [f:]) reductions, participants might be able to transfer their acquired knowledge about the speaker’s pronunciation habits to previously unheard words.

### 1.1. Compensation for reduction: Multiple mechanisms

In casual speech, a variety of reduced forms can be found ranging from assimilations (e.g., [tœymbaŋk] for Dutch *tuinbank* “garden bench”; see e.g., Mitterer, Csépe, & Blomert, 2006; Mitterer, Csépe, Honbolygo, & Blomert, 2006) to extreme reductions (e.g., [ɛik] instead of [ɛixələk] for Dutch *eigenlijk* “actually”; see e.g., Brouwer, 2010; Ernestus, 2000). Segmental and syllabic reductions lie in between these two endpoints of the reduction continuum. This continuum is determined, in part, by a word’s frequency of occurrence. Usually, the more frequent a lexical item is, the more likely it is to observe strongly reduced forms (e.g., Ernestus, 2000;

\* Corresponding author. Present address: Northeastern University, Speech-Language Pathology & Audiology, 360 Huntington Avenue, 226 FR, Boston, MA 02115, USA. Tel.: +1 617 373 5795.

E-mail addresses: k.poellmann@neu.edu (K. Poellmann), hansrutger.bosker@mpi.nl (H.R. Bosker), j.mcqueen@pwo.ru.nl (J.M. McQueen), holger.mitterer@um.edu.mt (H. Mitterer).

<sup>1</sup> Present address: Department of Cognitive Science, Faculty of Media and Knowledge Sciences, University of Malta, Msida MSD2080, Malta.

Jurafsky, Bell, Gregory, & Raymond, 2001). So, while any word in a sentence can undergo assimilation, only high-frequency words are reduced more severely (deletion of segments or even syllables).

Listeners have different ways of dealing with reduced speech. For example, they can take the (broader) sentence context and the (narrower) phonological context into account. Furthermore, they can extract information about the reduced word itself, such as lexical knowledge, word frequency and frequency of the surface form, from the mental lexicon. Finally, they can make use of fine phonetic detail in the speech signal and even use auditory processes to recognize reduced words.

The sentence context plays an important role in recognizing strongly reduced forms (e.g., Brouwer, Mitterer, & Huettig, 2013; Ernestus, Baayen, & Schreuder, 2002), while the utilization of probabilistic cues from the surrounding context helps listeners to deal with segmental reductions (Mitterer & McQueen, 2009b). In Dutch, when the word-final /t/ is preceded by an /s/ and followed by a bilabial sound (/b/, /m/), the /t/ is very likely to be reduced. Participants were found to be sensitive to these probabilistic facts.

Moreover, listeners use the knowledge stored in the mental lexicon to recover from reductions. For instance, listeners showed a lexical bias when judging the presence or absence of word-final /t/ in Dutch (Mitterer & Ernestus, 2006). Furthermore, word frequency appears to be one of the major cues that listeners use to compensate for extreme reductions like [ɛik] (e.g., Ernestus, 2000). For example, Mitterer and Russell (2013) showed that listeners exploit the correlation of lexical frequency and phonological reduction and assume, for instance, that a reduced prefix is more likely to belong to a high-frequency word than to a low-frequency word. Connine, Ranbom, and Patterson (2008) emphasize the role of lexically stored pronunciation variants in recognizing schwa-deleted words. In syllable judgment and lexical decision tasks, listeners reacted faster to more frequent surface forms, regardless of whether they were schwa-bearing or not.

Fine phonetic detail is another means by which listeners can recognize reduced words. It can hint at the presence of an apparently deleted segment, like a schwa: Listeners are able, for example, to differentiate a reduced form [sp<sup>h</sup>ɔ:t] of “support” from the unreduced form [spɔ:t] “sport” (Manuel, 1992). In the case of assimilated forms like [tœymbɔŋk], the intended word can be recognized by means of fine phonetic detail in the speech signal (Gow, 2002, 2003). For instance, an assimilated labial segment like the [m] in *gardem bench* is acoustically different from the intended [m] in *the same bench*.

Finally, the fact that assimilated forms can be recognized when listening to an unfamiliar language (Mitterer et al., 2006; Mitterer, Csépe, Honbolygo, et al., 2006) suggests that language-independent perceptual processes are involved in compensation for at least some forms of assimilation. Mitterer et al. (2006) argued further that language-universal auditory processes are involved in compensation, since similar effects could be obtained with speech and non-speech materials.

## 1.2. Speaker-dependent properties

There are thus different kinds of information used when compensating for (different kinds of) reductions. Frequency information, contextual cues, fine-phonetic detail, and lexical and probabilistic knowledge all have roles to play in the compensation process, and that process is based on low-level (auditory), prelexical and lexical mechanisms.

Research on how listeners are able to compensate for reduced speech has concentrated so far on these phonetic and language-dependent properties. In contrast, the current study focuses on speaker-dependent properties of reduced speech. This may be an important piece of the puzzle, since reductions depend on the speaker in several ways: First, reductions are optional. There are of course phonological contexts in which they are more likely to occur, but nothing restrains the speaker from articulating words in their full form (Mitterer & Ernestus, 2006). Similarly, reductions are not a direct by-product of a fast speaking rate. There are speakers who speak very fast but clearly (Shockey, 2003; Van Son & Pols, 1990, 1992), while others even produce reductions when reading aloud (Warner & Tucker, 2011). Second, when speakers reduce words, they can do so in many different ways. For instance, the Dutch suffix *-lijk* can be realized at a continuum ranging from the citation form [lɛk] to highly reduced [ə] or [k] (Pluymaekers, Ernestus, & Baayen, 2005). As speakers can choose from a great variety of possible reductions of a given sequence, they vary in the choices they make (Ernestus et al., 2002). Moreover, Keune, Ernestus, van Hout, and Baayen (2005), investigating how common reductions are among different talker groups (male and female speakers of Dutch and Flemish), found that Dutch men reduced most often, while Flemish highly educated women were least likely to do so.

The question then arises whether listeners tune in to the reduction styles of given speakers in order to compensate for the speakers' imperfect articulation. In other words, is there adaptation to specific reduction styles? We asked that question here. In particular, we tested whether adaptation to a specific reduction facilitates the recognition of new reduced words spoken by the same talker. Please note that, in the following, we will use the terms *adaptation* and (*perceptual*) *learning* interchangeably.

## 1.3. Speaker-specific learning

There is ample evidence that listeners tune in to characteristics of particular speakers. For instance, Ladefoged and Broadbent (1957) and Sjerps, Mitterer, and McQueen (2011) found that the same vowel ([ɪ] or [ɛ]) was perceived differently depending on the F1 of the context it appeared in. So listeners do not interpret an incoming sound solely on the basis of its acoustic properties, but take speaker-specific properties (like the available F1 range) into account. Another example of how listeners tune in to a particular talker is provided by Reinisch, Jesse, and McQueen (2011). They showed that the perception of a juncture phoneme with a constant, but ambiguous duration (e.g., [s] in Dutch *eens* (s)peer “once (s)pear” changed depending on the rate at which the preceding context was

spoken. Again, listeners do not only evaluate the raw incoming signal, they compute the incoming information (segment duration in this case) relative to talker-specific information (like speaking rate) available from the preceding context.

Note, however, that these adjustment processes may not be tied to the perception of speaker identity, because evidence suggests that these normalizations occur at an early auditory level of processing (e.g., Reinisch & Sjerps, 2013; Sjerps, et al., 2011, for effects of preceding contexts; however, see Toscano & McMurray, 2012, for different effects of following contexts). In line with this assumption, normalization for speaking rate is speaker-independent, and speaker identity of the precursor has hardly any effects on normalization (Newman & Sawusch, 2009). Even though these processes may not represent speaker-specific adaptation, they still help to reduce variance in the speech signal due to speaker differences.

Another normalization process, however, is more tightly coupled to speaker identity. Norris, McQueen, and Cutler (2003) tested whether listeners could adapt to speaker-specific idiosyncrasies, in their case a speaker who produced either the fricative /f/ or the fricative /s/ as an ambiguous sound ([<sup>s</sup>/f]). In these perceptual learning studies (for a review, see Samuel & Kraljic, 2009), one group of participants was trained to perceive an ambiguous sound [<sup>s</sup>/f] as /s/, another group was trained to perceive the same sound as /f/ based on lexical bias. This kind of exposure leads to a recalibration of the /s/-/f/ contrast which generalizes across words (McQueen, Cutler, & Norris, 2006). Moreover, this perceptual learning effect proved to be talker-specific and stable over time (Eisner & McQueen, 2005, 2006). Investigating whether there is “a return to normal”, Kraljic and Samuel (2005) found that only canonical pronunciation of *both* phonemes (/s/ and /f/ in this case) appeared to be able to reset the phonemic categories to pre-learning parameters. Importantly, these unambiguous instances had to be uttered by the same speaker listeners had been trained on. Learning about segmental idiosyncrasies involving fricatives (i.e., sounds that convey more talker-specific information; Andics, McQueen, & Van Turenhout, 2007) was again found to be talker-specific, while learning about plosives (which contain less talker-specific information; Andics et al., 2007) generalized across speakers (Kraljic & Samuel, 2007). In all these studies, listeners learned to interpret an artificially constructed ambiguous sound as an (odd) pronunciation variant of a given phoneme depending on the use the speaker made of this sound.

Maye, Aslin, and Tanenhaus (2008) showed that listeners can also adapt to vowel differences that depend on sociolinguistic factors. In their study, they tested adaptation to a hypothetical regional dialect of American English. Participants listened to a story twice (in two separate sessions), first to a canonical version and the second time to a manipulated version mimicking a regional accent in which the front vowels were lowered in F1–F2 vowel space. In the following lexical decision task, which was identical for the two sessions, listeners categorized more accented words as existing English words after exposure to the accented story than after exposure to the canonical version of the story. This adaptation effect generalized to words which had not occurred in the story.

Dahan, Drucker, and Scarborough (2008) investigated adaptation to a natural dialect of American English. Participants were exposed to a dialect speaker who naturally raised the vowel /æ/ before /g/ but not before /k/. They could identify *back*-like words (which are unaffected by the dialect shift) more easily after exposure to accented *bag*-like words. This facilitation was probably due to reduced competition between the two word types because the different vowels contributed to the disambiguation. Listeners can thus adapt to unusual pronunciations attributed to the dialect of a given speaker.

In the present study, we asked whether listeners can also adapt to the reduction style of a given speaker. That is, like Dahan et al. (2008), we wanted to test listeners' ability to adjust to naturally-occurring idiosyncrasies of a speaker, but unlike them we were interested in stylistic rather than in geographical variation as the source of the idiosyncrasies. Geographical variation is tied to groups of speakers (e.g. from a given region), while stylistic variation can be unique to particular individuals.

#### 1.4. Design and predictions

In the present experiments, we used a variant of the visual-world eye-tracking paradigm (Mitterer & McQueen, 2009b) that was modified to permit study of perceptual learning. Different groups of participants were exposed to one of two different forms of reduction, and then tested on both forms. One group of participants was exposed to segmental reductions; the other group to syllabic reductions. That is, one group (the segmental reduction group) heard /b/-initial words where the /b/ of an unstressed first syllable was replaced by a sound similar to the Dutch labio-dental approximant [ʋ] (e.g., [ʋikini] instead of [bikini] “bikini”). The other group (the syllabic reduction group) heard *ver*-words where the prefix *ver*- was replaced by a long [f] (e.g., [f:lədə] instead of [vərledə] *verleden* “past”; note that most Dutch speakers already devoice initial /v/ to [f], so that the latter form becomes [fərledə]). In the test phase, both groups heard both kinds of reductions. We reasoned that, if listeners are able to tune in to respective reduction styles, the segmental reduction group should be better at recognizing new reduced /b/-words than the syllabic reduction group, and the opposite should be true for new reduced *ver*-words (the syllabic reduction group should perform better on them than the segmental reduction group).

The main goal of the present study was thus to establish whether listeners can adapt to acoustic reductions and generalize this learning effect to new words. The segmental reduction condition was implemented in such a way as to closely resemble perceptual learning studies about ambiguous fricatives (e.g., McQueen et al., 2006). We expected to find a learning effect for this condition. In addition, we sought in the syllabic reduction condition to discover whether listeners are also able to adapt to reductions that affect more than a single segment. Note that the emphasis of the syllabic reduction condition was on the size of the reduced linguistic unit, not on its morphological status. The purpose was to reduce (in this case: *shorten*, not completely delete) a linguistic unit that is larger than a single segment. The predictions for this novel syllabic reduction condition were less clear. In this respect, the segmental reduction condition also served as a control condition. If there were learning for segmental reductions, but not for syllabic reductions, the latter null effect could not be due to listeners' general inability to adapt to reductions.

In the test phase, all reduced items were words that participants had not encountered in the exposure phase. Finding an adaptation effect would thus be evidence of generalization of learning to new words. Such a finding would contribute to the ongoing debate on theories of speech processing. In some models (e.g., TRACE, McClelland & Elman, 1986; Shortlist B, Norris & McQueen, 2008), lexical representations are abstract: The acoustic input is mapped onto abstract prelexical phonological units and then onto corresponding entries in the mental lexicon. Prelexical processes have to compensate for continuous speech phenomena like reduction so that underlying phonological representations can be accessed. In contrast, in episodic models, the entry for a particular word in the mental lexicon consists of detailed and concrete episodes of a word's pronunciation that the listener has encountered previously (e.g., Bybee, 2001; Goldinger, 1998). Thus, an entry not only consists of various memories of a word's citation form, but also contains episodic traces of reduced forms showing different kinds and/or degrees of reduction. A model storing only acoustically detailed lexical episodes and using only those episodes in word recognition cannot take advantage of any sublexical regularities during the word recognition process.

It has been shown that neither purely abstractionist models nor extreme episodic models can explain all aspects of spoken word recognition alone. For evidence against pure abstraction, see for example Nygaard and Pisoni (1998); for evidence against pure episodic storage, see for example Cutler, Eisner, McQueen, and Norris (2010). The field, therefore, has converged on the idea of hybrid models combining aspects of both theories (e.g., Cutler et al., 2010; Ernestus, 2009; Mitterer, Chen, & Zhou, 2011). In this context, we can ask how important the episodic and abstractionist modes of lexical access are in the processing of reductions. If episodic storage of reduced forms is the only means to recognize them, there should be no generalization of learning about reductions to new reduced words. If, however, abstraction plays an important role in the recognition of reduced forms, such generalization should occur. Prelexical abstraction would enable the learning effect to generalize to previously unheard reduced words (cf. McQueen et al., 2006). Generalization would thus indicate not only that there is abstraction, but also that this is localized at the prelexical processing level, and hence would place constraints on future hybrid models.

Our experimental set-up, however, may also provide data that are problematic for current versions of abstractionist models. This is because we test whether listeners can adapt to regular syllabic reductions, such as the Dutch prefix *ver-* being shortened to a long [f] (e.g., [f:ledə] for *verleden*). In current abstractionist models, the acoustic input is mapped at the prelexical level onto abstract phonological representations. These are taken to be features (Gaskell & Marslen-Wilson, 1997), phonemes (Norris & McQueen, 2008) or features and phonemes (McClelland & Elman, 1986). None of these models, however, stipulates prelexical representations that are larger than a segment and that would be able to encode at a prelexical level that [f] is a possible pronunciation of the three-segment string /vər/. Such representations, however, would be necessary to explain generalization from one word to another for a reduction such as /vər/ → [f].

### 1.5. Corpus analyses

Before testing whether listeners can adapt to reductions, we needed to know whether Dutch speakers actually produce the kind of /b/- and *ver*-reductions we wanted to investigate. Pluymaekers et al. (2005) examined reduced affixes in spoken Dutch, amongst which the prefix *ver-*. They first established how many different word types starting with *ver-* occurred in the Corpus of Spoken Dutch (Oostdijk, 2000). Here, the notion “word type” not only includes words belonging to different lemmas but also different word forms of one lemma (e.g., the present and past tense of a given verb). Only one randomly chosen token per word type was considered for analysis and thus not all instances of *ver*-words that occurred in the corpus. The prefix *ver-* was reduced to [f] in 23 out of 140 word types (i.e., in 16.43%). There were no such data available on word-initial /b/-reductions resulting in a labio-dental approximant. So, following Pluymaekers et al. (2005), we conducted a similar study for the /b/-reductions making use of the same corpus. Out of 111 word types of /b/-initial words, where the /b/ was followed by a full vowel and where the first syllable was unstressed, 15 were pronounced with a labio-dental approximant [ʋ]. This was the most frequent non-canonical variant (13.51%). Its frequency is similar to the percentage of *ver*-to-[f]-reductions. Word-initial /b/s in unstressed position may also be reduced to approximants in other languages. Warner and Tucker (2011), investigating phonetic variability of stops in spontaneous and careful speech, noted for American English that expected voiced stops (/b, g/) may be realized as approximants or vowel-like sounds. All stops examined in their study occurred between vowels and/or sonorants. Consonants were more reduced between two unstressed syllables (also across word boundaries) than after a stressed syllable. In the current study, the experimental sentences were chosen to fulfill all of these conditions.

## 2. Experiment 1

The aim of Experiment 1 was to apply a learning paradigm consisting of exposure and test phases to investigate whether listeners are able to adapt to talker-specific reduction styles. In the exposure phase, one group of participants (the segmental reduction group) was exposed to /b/-reductions, while another group (the syllabic reduction group) was exposed to *ver*-reductions. In the test phase, both groups were tested on new reduced words of both reduction types. If participants in the segmental reduction group can tune in to /b/-reductions, they should perform better than the syllabic reduction group on new reduced /b/-words. The same holds for participants in the syllabic reduction group: If they can adapt to *ver*-reductions, they should outperform participants in the segmental reduction group on new reduced *ver*-words.



Note that although we investigated adaptation to talker-specific reduction styles, this study does not warrant claims that the learning process is necessarily talker-specific since we did not test generalization of learning across talkers. Rather, we exposed one group of listeners to a talker who reduces only word-initial /b/s and another group to the same talker who reduces only the prefix *ver-*. We thus use the term “talker-specific” in the sense that, from a listener’s point of view, the reduction type encountered in the exposure phase was a feature of the exposure talker.

## 2.1. Method

### 2.1.1. Participants

Fifty-six participants from the Max Planck Institute subject pool, all native speakers of Dutch, were paid to take part. The segmental reduction group consisted of five males and 23 females, aged 18–30 years (mean: 21 years). The syllabic reduction group consisted of three males and 25 females, aged 18–23 years (mean: 20 years). None reported any hearing disorders and all had normal or corrected-to-normal vision.

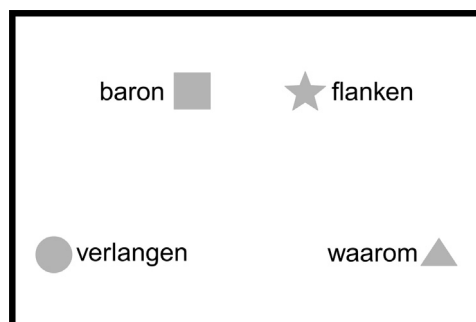
### 2.1.2. Design

In the exposure phase, one group of participants (the segmental reduction group) was exposed to four critical types of items (see Table 1): They heard reduced /b/-words (e.g., /bikini/ *bikini* produced as [b<sup>h</sup>ikini]), canonical /u/-words (e.g., /uustənɛɪ/ *woestenij* produced as [uustənɛɪ]), canonical *ver*-words (e.g., /vərledə/ *verleden* produced as [fərledə]) and quasi-canonical /f/-words (e.g., /flɔsə/ *flossen* produced as [f:ɔsə]). The other group (the syllabic reduction group) was exposed to the reversed design: They heard canonical /b/-words, quasi-canonical /u/-words (e.g., /uustənɛɪ/ *woestenij* produced as [b<sup>h</sup>uustənɛɪ]), reduced *ver*-words (e.g., /vərledə/ *verleden* produced as [f:ledə]) and canonical /f/-words. This procedure was chosen to enhance the contrast between the two groups, and to enhance the perceptual contrasts within each group, as in the perceptual learning paradigm on which the experiment was based (Norris et al., 2003). Listeners in the segmental reduction group were led to interpret the ambiguous sounds [b<sup>h</sup>] as [b] (hearing them at the beginnings of /b/-words) and were thus implicitly encouraged to disregard the cues indicating a syllabic reduction (hearing the long [f]s in /f/-words). Conversely, the syllabic reduction group learned that a long [f] is a variant of *ver*-that the speaker produces, and was again implicitly encouraged to ignore the cues indicating a /b/-reduction (hearing the [b<sup>h</sup>]s at the beginnings of /u/-words).

Note that, due to this design, we cannot attribute a potential learning effect solely to the exposure to one reduction type (e.g., /b/-reductions for the segmental reduction group), as the contrast group (e.g., the syllabic reduction group) might adapt to the manipulation of the contrast sound (e.g., the /u/-manipulation). As we compare the performance of the two groups, it is not clear how large the effects are that each group contributes. However, as the reductions under investigation are real-world phenomena, whereas the manipulation of the quasi-canonical stimuli is not, it seems likely that it is easier for listeners to adapt to the reductions than to the quasi-canonical stimuli.

**Table 1**  
Experimental design and types of stimuli.

	/b/-reduction group (segmental)		<i>ver</i> -reduction group (syllabic)	
Exposure phase	Reduced /b/-words	[b <sup>h</sup> ikini]	Canonical /b/-words	[b]ikini
	Canonical /u/-words	[u]oestenij	Quasi-canonical /u/-words	[b <sup>h</sup> u]oestenij
	Canonical <i>ver</i> -words	[fər]leden	Reduced <i>ver</i> -words	[f:]leden
	Quasi-canonical /f/-words	[f:]lossen	Canonical /f/-words	[f]lossen
Test phase		[b <sup>h</sup> u]aron		[b <sup>h</sup> u]aron
		[b <sup>h</sup> u]aarom		[b <sup>h</sup> u]aarom
	All items reduced	[f:]langen	All items reduced	[f:]langen
		[f:]lanken		[f:]lanken



**Fig. 1.** Example display of a test trial.

In the test phase, both groups heard only reduced forms of both types, in new /b/-, /v/-, ver- and /f/-words (see Table 1). If participants did not adapt to the reduction type they were exposed to, there should not be a group difference in the test phase. If, however, participants did adapt to the reduction type they were exposed to, they should outperform the other group on new items of the same reduction type.

### 2.1.3. Materials

The critical words were embedded in a spoken carrier sentence like “Klik op het woordje XXX daar links van het rooie vierkantje” (“Click on the word XXX there to the left of the red rectangle”). In this sentence, the word preceding the target word (i.e., the word which participants should click on) ended in a sonorant voiced segment (the schwa of *woordje* [ʊo:rtʃə]). This is the phonological context for the occurrence of /b/-reductions (Warner & Tucker, 2011). Note that ver-reductions occur in a variety of contexts. Participants hearing a sentence with the reduced word [b̥] *Jaron* “baron” would see a display like the one in Fig. 1. In the exposure and test trials, there were always a /b/-word, a /v/-word, a ver-word and an /f/-word on the screen. The use of printed-word eye-tracking (as opposed to eye tracking with a display of pictures) made it possible to test a larger number of (not necessarily picturable) words.

For the test phase of the segmental reduction condition, 24 /b/-initial words were selected which did not carry initial stress. Their /v/-initial competitors were matched on following vowel and stress pattern (e.g., *boerenKOOL* “kale” and *woesteNIJ* “wasteland” both showing a weak-weak-strong pattern). For the test phase of the syllabic reduction condition, 24 words beginning with ver- and followed by a liquid were selected as targets. Their competitors were /f/- or /v/-initial words followed by an identical combination of liquid and vowel, so that a target-competitor pair would overlap in at least three segments in their reduced forms (e.g., the first syllable(s) of *verliezer* “loser” and *vlieger* “kite” would both be pronounced as [fli]). The target-competitor pairs were also matched on number of syllables in their reduced forms (i.e., *verliezer* is counted as having two syllables when reduced and thus matches *vlieger*). However, due to the onset requirements of the stimuli as well as their stress pattern constraints, it was not possible to match the stimuli on frequency of occurrence (see Appendix A, Table A1). Furthermore, an additional 24 /v/-initial and 24 /f/- or /v/-initial target words as well as their /b/- and ver-initial competitors were selected in order to balance the responses to /b/- and /v/-words as well as the responses to ver- and /f/-words.

For the exposure phase, four new sets of 24 /b/-, /v/-, ver-, and /f/- or /v/-initial words and competitors were selected. The target exposure words did not overlap with their competitors to the same extent as the test words did.

### 2.1.4. Stimulus construction

Digital recordings of the stimuli were made by a male native speaker of Dutch (the second author) in a sound-proof booth, sampling at 44.1 kHz. This speaker naturally devoices instances of [v] to [f], so that a ver-to-[f]-reduction was plausible. As a consequence, it was possible to use /v/-initial words (which would be devoiced by this speaker) and /f/-initial words as competitors for the ver-words. In order to make the occurrence of reductions more plausible, we tried to elicit a casual speaking style by asking the speaker to produce sentences according to the following template: “Klik op het woordje... daar links/rechts van het gele/groene/rooie/zwarze rondje/sterretje/driehoekje/vierkantje” (“Click on the word... there to the left/right of the yellow/green/red/black circle/star/triangle/rectangle”). The sentences were prompted by a visual display on a computer screen with a printed target word and the corresponding shape positioned to the left or to the right of the word. This recording method produced some hesitations in the later

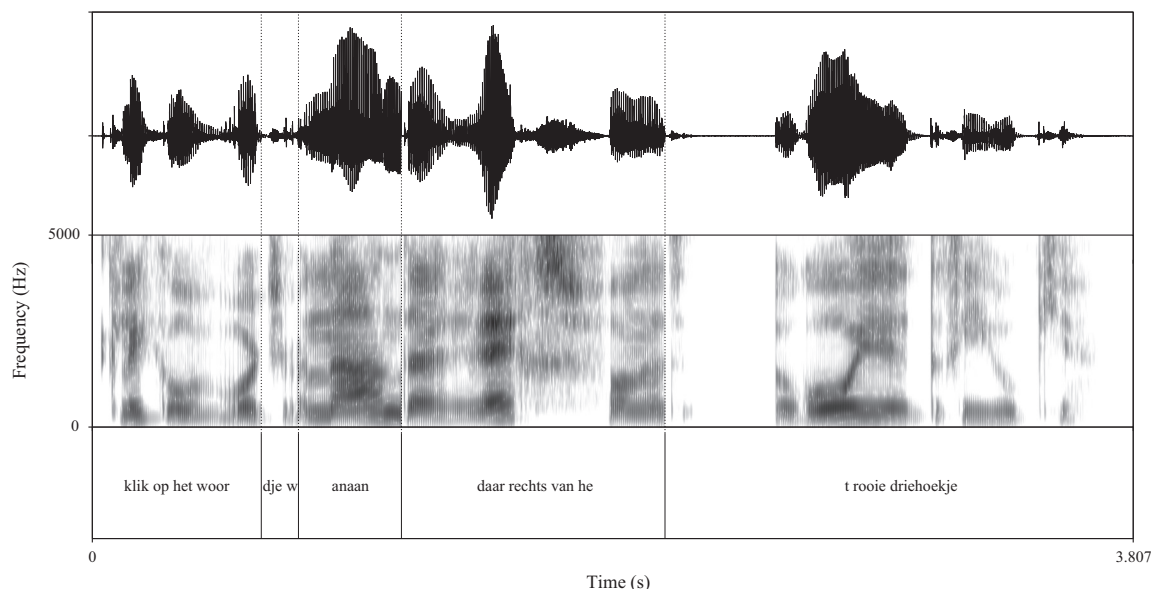


Fig. 2. Splicing points in an experimental sentence containing a segmental reduction in Experiment 1.

part of the template (i.e., after the target word), which were retained as they made the recordings sound more casual. Also the colloquial *daar* “there” was included in the sentence for this purpose. The aim was thus not to have any other reductions occurring in the carrier sentence, but to provide other indicators of casual speech. The sentence accent was always placed on *rechts* or *links*. As for the target words, the speaker produced each /b/- and *ver*-word both canonically and in reduced form, that is, he consciously replaced the initial [b]s by short [u]s and the initial *ver*-s by long [f]s.

The experimental sentences were created by cross-splicing. Splicing points were chosen at major positive-going zero-crossings before the stops in *woordje*, *daar* and *het* (see Fig. 2). The first fragment *Klik op het woor-* was the same for all sentences. Likewise, only one version of the sentence parts *daar links van het*, *daar rechts van het*, and of each color-shape combination was used.

To create the reduced /b/-words, ambiguous sounds [b<sub>u</sub>] midway between a /b/ and a /u/ were selected by means of a pretest (see below). These ambiguous sounds and their respective *dje*'s were then spliced between the first sentence fragment (*Klik op het woor-*) and the rest of the /b/-word respectively (see Fig. 2). For the reduced *ver*-words, the recorded versions containing a long [f:] were used. That is, the reduced *ver*-words did not contain an additional splicing point within the target. The canonical versions of the /u/-initial and the /f/-initial words were cross-spliced to create what we called the quasi-canonical forms. For the /u/-initial words, the same procedure as for the reduced /b/-words was used. That is, if there were splicing artefacts, they occurred in both the reduced /b/-words and the quasi-canonical /u/-words. For the /f/-initial words, one of the *dje* [f:] sequences (i.e., the consciously reduced *ver*- and its original *dje*) was spliced between the first sentence fragment (*Klik op het woor-*) and the rest of the original /f/-word. As the splicing point in the /f/-initial words was located at the end of a fricative and the spliced-in segment ended in a fricative as well, the transition did not give the splicing away. The remainder of the /f/-word fitted well with a somewhat longer [f] than the one of the original recording. It thus did not matter that /f/-initial words were manipulated while reduced *ver*-initial words were not.

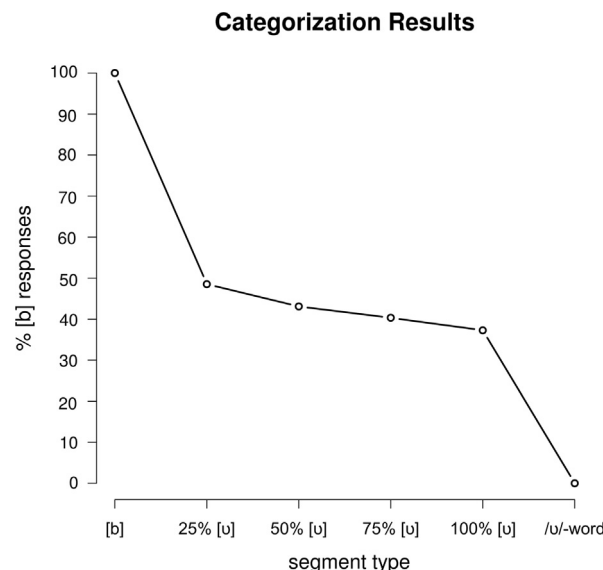
In order to prevent coarticulation features in the cross-spliced materials from interfering, attention was paid to the fact that the source words of the reduced segments (the various [b<sub>u</sub>]s and [f:]s) overlapped in at least one following segment with the target words they were spliced onto (e.g., the ambiguous [b<sub>u</sub>] that stemmed from the consciously reduced *baron* was also used for *waarom* and *barak*). All /b/- and /u/-initial target words of the test phase therefore started with an ambiguous [b<sub>u</sub>]. Similarly, all *ver*- and /f/-initial test targets started with a long [f:]. In the exposure phase however, the segmental reduction group heard manipulated /b/-words (containing [b<sub>u</sub>]s) and /f/-words (containing [f:]s) only, while the syllabic reduction group heard manipulated *ver*-words (containing [f:]s) and /u/-words (containing [b<sub>u</sub>]s) only.

For the /b/-reductions, an ambiguous sound midway between /b/ and /u/ was chosen in order to stay as closely as possible to the original perceptual learning paradigm of Norris et al. (2003). The *ver*-reductions under investigation, however, were less ambiguous in nature. Therefore, the remaining fricative was lengthened in order to provide listeners with fine phonetic detail which they could use for learning.

### 2.1.5. Pretest

In order to find ambiguous sounds [b<sub>u</sub>] between /b/ and /u/ that would serve as reduced [b]s, a pretest was run. Five native Dutch employees of the Max Planck Institute for Psycholinguistics, Nijmegen, participated.

A /b/-/u/-continuum was created by excising each first segment of the 48 consciously reduced /b/-initial words (i.e., the /b/-initial words that were recorded with a [u] instead of a [b]) and shortening its original duration to 25%, 50%, and 75% with PSOLA (using Praat 5.1.26, Boersma & Weenink, 2010). The shortened intermediate continuum steps were added in case the consciously reduced [b]s sounded too approximant-like. Five versions of every /b/-initial word (e.g., *banaan* “banana”; one version pronounced canonically



**Fig. 3.** Pretest: Percentages of [b] responses for /b/-initial words (e.g., *banaan* “banana”) whose first segment was a natural [b], 25%, 50%, 75% and 100% of a [u] replacing the original [b] (e.g., *wanaan*) and canonical /u/-initial words (e.g., *waanzinnig* “insane”).

with [b], one pronounced with [u] – *wanaan* – plus the three shortened versions of [u]) as well as the 48 canonically pronounced /u/-initial words (e.g., *waaninnig* “insane”) were presented to the participants in a random order which was different for each participant. Participants heard sentences like “Klik op het woordje banaan” (“Click on the word banana”) and had to decide whether the last word of the sentence was pronounced with a [b] or with a [u], regardless of the spelling of the word. Participants were told explicitly that /b/-initial words could be mispronounced with an initial /u/. They were not put under time pressure and were encouraged to make their decision carefully. After the sentence was played, they were asked to click with the mouse on a field labeled “b” or “w” (the orthographic form of /u/ in Dutch) depending on what they had heard.

Fig. 3 shows the results of the pretest. Quite unsurprisingly, the canonical /b/- and /u/-words got 100% and 0% of [b] responses respectively. However, the different replacements of the [b] in the /b/-words were all judged quite similarly with the votes lying within a range from 49% to 37% of [b] responses. Although the value of the 25% [u] stimuli was closest to the 50% criterion (i.e., the perfect value of ambiguity), we decided to use the 50% [u] stimuli for the main experiment, as their values were comparable to those of the 25% [u] stimuli but the items were less strongly manipulated and hence sounded more natural.

### 2.1.6. Procedure

Participants were seated in a sound-attenuated booth at a comfortable viewing distance from the computer screen. Eye movements were monitored using an SR Research EyeLink II, sampling at 500 Hz. The auditory stimuli were presented to the participants over headphones. Prior to the experiment, participants received written instructions in which they were shown an example sentence and an example display. The instructions read that they should click on the word that occurred in the sentence. Participants were told to pay attention to the whole sentence, as it could be that a word appeared twice on the screen (filler items). They were told that, in these cases, the disambiguating information about the color, form or position of the geometrical shape relative to the target word would be provided in the last part of the sentence.

During the experiment, participants heard sentences like “Klik op het woordje [f:]langen daar rechts van het rooie rondje” (“Click on the word desire there to the right of the red circle”) and were asked to follow these instructions. Fig. 1 shows an example of a display; the example is taken from a trial in the test phase. In the exposure and test trials, there were always a *ver*-word, an /f/-word, a /b/-word and a /u/-word on the screen which were associated with one of four different shapes in the same color (see Fig. 1). Each center of the printed words coincided with the center of one quadrant of the screen, independently from word length. The corresponding shapes were either placed to the left or to the right of the printed words. The shapes did not have an absolute position on the screen, rather the distance to the printed words was kept constant. Targets and competitors were positioned on the same side of their corresponding geometrical shape sharing the same color.

The main difference between exposure and test displays lay in the degree of similarity between target and competitor words. In the exposure phase, in order to support learning, they did not share any overlapping segments following the reduced onset segment (s) (e.g., *bikini* “bikini” and *wagon* “wagon” or *verleden* “past” and *vogelkooi* “birdcage”). In the test phase, in order to provide a temporary ambiguity against which a learning effect could be measured, they overlapped at least until the first vowel following the reduced onset segment(s) (e.g., *baron* “baron” and *waarom* “why” or *verliezer* “loser” and *vlieger* “kite”).

Furthermore, there were 48 filler trials interspersed among the 96 exposure and 96 test trials. Fillers consisted of two different words that appeared twice on the screen. The identical words differed either in their position (left or right) relative to the same shape in the same color (i.e., it was a side trial), or they were placed on the same side of the same shape but the color of the shape was different (a color trial), or they were positioned on the same side of different shapes in the same color (a shape trial). The filler trials were supposed to draw participants’ attention to the information given in the second half of the experimental sentences (i.e., the information about position, color and shape of the object associated with the word they should click on) so that participants would not concentrate too much on the reduction phenomena. While participants did not need the shape/color/side information in the exposure and test trials to decide which word they should click on, this information was essential in the filler trials. Mitterer and McQueen (2009b) found that the inclusion of such trials made phonological effects more consistent.

At the beginning of each trial, a fixation cross appeared in the center of the screen for 500 ms. Then the four printed words (in a 28-point Arial font) and the four associated shapes were presented. After 1500 ms, the auditory stimulus was played. As soon as participants had listened to the entire sentence and had clicked with the mouse on the screen, the following trial was initiated. Every 10 trials, a drift correction was carried out. Participants had the opportunity to take a break after every 50th stimulus. The experiment started with four practice trials. The randomized order of presentation (exposure and filler trials, then test and filler trials) was different for each participant. An experimental session took approximately 35 minutes.

## 2.2. Results

### 2.2.1. Fillers

Filler trials were not considered for statistical analysis. Nevertheless, we plot the eye movement data of the fillers in Fig. 4 to show how participants used the unfolding information over time. On the y-axis, the distance in pixels between participants’ fixation points and the centers of the three word types on the screen (target, competitor and two distractors) are plotted (100 pixels correspond to 3.2 cm in Experiment 1; average word length was 165 pixels). If participants look to the target word, the distance between the fixation point and the center of the target word should become smaller. In the instruction sentence, the disambiguating information became available first in the side trials then in the color trials and finally in the shape trials. This is reflected in when participants started to prefer the target over the competitor: A preference for the target over the competitor arose first in side trials, then in color trials, and



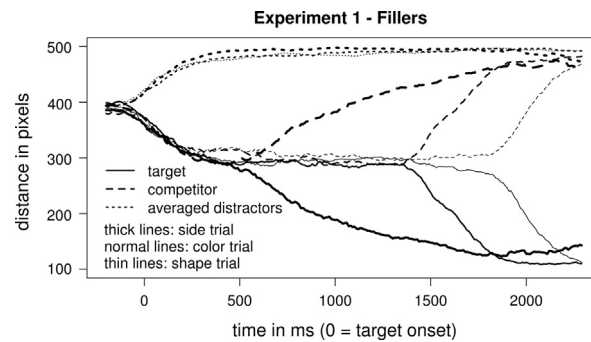


Fig. 4. Distances between fixations and targets, competitors and averaged distractors in pixels for the filler trials in Experiment 1 (e.g., *Klik op het woordje konijn daar rechts van het groene driehoekje* "Click on the word rabbit there to the right of the green triangle"). See text for details about the definitions of side trial, color trial and shape trial.

Table 2

Descriptive statistics of RTs for correct trials for both groups in the exposure and test phases in Experiment 1.

RT in ms	Segmental reduction group		Syllabic reduction group	
	Exposure	Test	Exposure	Test
mean	2278	2362	2104	2207
sd	804	859	803	800
min	495	551	326	342
max	9177	8385	5471	7110

finally in shape trials. In a side trial (represented by thick lines), participants started to look to the target at around 500 ms after target onset. In a color trial (represented by lines of medium thickness), participants had a preference for the target about 1400 ms after target onset. Finally, participants started to look to the target only about 1800 ms after target onset if the information about the shape (thin lines) disambiguated the target word. Fig. 4 thus shows that participants were using information as soon as it was available.

### 2.2.2. Exclusion criteria

The dependent variables were mouse clicks and eye movements. Trials in which participants did not click within a circle with a radius of 180 pixels around one of the centers of the four words on the screen were not considered further. Due to this criterion, 0.3% of all trials (36 trials) were excluded. For the eye-tracking data, we analyzed the data from the better eye of the participants (i.e., the eye that showed less error in the validation of the calibration of the eye-tracker). Only trials in which participants clicked on the target word were included in the eye movement analysis. This led to the exclusion of 81 trials (0.7%) in total. Out of these 81 trials, 76 were trials of the test phase in which participants clicked incorrectly on the competitor (1.4% of the test trials). Moreover, 194 trials (1.8%) in which participants did not look on the screen at some point in time were discarded.

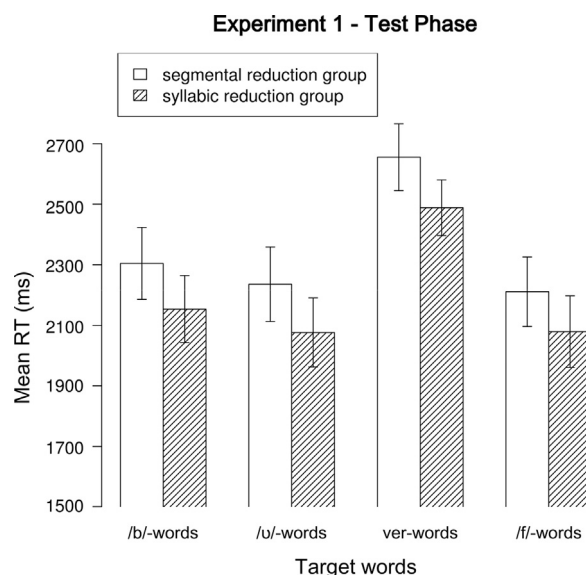
Participants in both groups performed the task very accurately. The percentages of correct responses (clicks on the target word) equaled or exceeded 96% in all conditions for both exposure and test phases. Descriptive statistics of reaction times (RTs) for correct trials are displayed in Table 2 for both groups in the exposure and test phases. The standard deviation (sd) values indicate a high amount of between-subject variability. Some participants clicked as soon as possible, others often waited for the instruction sentence to finish. In general, participants in the syllabic reduction group responded faster than participants in the segmental reduction group (see also Figs. 5 and B1). The large difference between minimum and maximum RTs prompted us to exclude outliers. To achieve this, a linear mixed-effects model containing only participants and items as random effects and Trial Number as fixed effect was run. Then the residuals of this (atheoretical) model were computed. Based on a visual inspection of a residual plot, 33 trials in the exposure phase (with residuals higher than 1800 ms or lower than −1300 ms) and 21 trials in the test phase (with residuals either below −1500 ms or above 2300 ms) were excluded.

### 2.2.3. Statistical testing

We used linear mixed-effects models on both the click responses<sup>2</sup> and the eye movement data.<sup>3</sup> Analyses were conducted for the segmental reduction condition (/b/-words and /u/-words) and the syllabic reduction condition (ver-words and /f/-words) separately. Participants and Items were entered in the models as random factors, including random slopes. Group (i.e., training group: exposure to segmental or syllabic reductions) and Target Word (/b/- vs. /u/-target; ver- vs. /f/-target) served as fixed effects. Both fixed effects were coded as a numerical contrast (−0.5 and 0.5). In all following analyses, the segmental reduction group is coded as −0.5 and the syllabic reduction group as 0.5. Likewise, /b/-targets are always coded as −0.5 and /u/-targets as 0.5 in the segmental reduction

<sup>2</sup>  $\text{lmer}(\log(\text{rt}) \sim \text{TargetWord} * \text{Group} + \text{scale}(\text{TrialNumber}, \text{scale} = \text{F}) + (1|\text{Subject}) + (0 + \text{TargetWord}|\text{Subject}) + (1|\text{Item}) + (0 + \text{Group}|\text{Item}), \text{data} = \text{testRt}, \text{subset} = (\text{trialType} == "test\_b")$

<sup>3</sup>  $\text{lmer}(\log(\text{compDistance}) - \log(\text{targetDistance}) \sim \text{TargetWord} * \text{Group} + (1|\text{Subject}) + (0 + \text{TargetWord}|\text{Subject}) + (1|\text{Item}) + (0 + \text{Group}|\text{Item}), \text{data} = \text{testEM}, \text{subset} = (\text{trialType} == "test\_b")$



**Fig. 5.** Mean RTs in ms and SE of both groups for each type of target word in the test phase of Experiment 1. All /b/- and *ver*-words were reduced. Also the /u/- and /f/-words were slightly manipulated: The [u]s of /u/-words were replaced by [ʊ]s and the [f]s of /f/-words were replaced by [f̥]s (for more details see main text).

condition. In the syllabic reduction condition, the *ver*-targets are coded as  $-0.5$  and the /f/-targets as  $0.5$ . Task performance, as measured for example by RT, often improves over the course of an experiment. Therefore, Trial Number was added as another fixed effect with values centered around zero in the model for the click data. Thus, we tested whether log-transformed RTs and log-transformed target preference (as determined by the difference between the distance of the fixation to the competitor and its distance to the target) were influenced by the above mentioned fixed effects. Note that a small target distance (i.e., the participant looks at the target) likely co-occurs with a large competitor distance (i.e., the participant does not look at the competitor). This means that the target preference (the difference between the competitor distance and the target distance) is positive. The eye-movement analyses tested whether there is a group difference in the size of the target preference. For the eye-tracking analyses, we used sliding 200 ms time windows from 200–1500 ms after target onset starting at every 100 ms (Barr, 2008). If several subsequent time windows showed significant effects, another model was then run on the data for the larger window that included as many of the 200-ms windows as were separately significant. The reported *p*-values reflect the results of the analyses of these larger time-windows. To estimate *p*-values, Markov chain Monte Carlo simulations were used. Note that this is possible when separately specifying random intercepts and random slopes (see models in footnotes) for binary factors in the slope term. Note also that formulating the model in this way does not require the random intercepts and random slopes to be uncorrelated. The correlations are just not estimated as a model parameter.

Note also that in the exposure phase, Group (exposure to segmental or syllabic reductions) and Target Word (e.g. /b/- vs. /u/-word) are necessarily confounded, as only the segmental reduction group was exposed to reduced /b/-words and only the syllabic reduction group was exposed to reduced *ver*-words. Furthermore, an effect of reduction should manifest itself either as a main effect of Group or as an interaction of Group and Target Word. In the exposure phase, the syllabic reduction group, for example, should be slower to respond to *ver*-words than the segmental reduction group, as the former hears them in their reduced forms whereas the latter hears them in their canonical forms (see, e.g., Pitt, 2009, on the RT advantage for canonical over reduced forms). The test phase will show whether we can find an adaptation effect, that is, an effect of exposure. This effect should manifest itself again as either a main effect of Group or as an interaction of Group and Target Word, but this time the effect should go in the opposite direction: The syllabic reduction group should respond faster to the *ver*-words than the segmental reduction group, as both groups now hear reduced *ver*-words and the syllabic reduction group has the advantage of being already familiar with *ver*-reductions while the segmental reduction group might struggle with the reduced sounds that they only previously heard in the contrast (i.e., /f/-) words. The reverse holds for segmental reductions.

Similar results are expected for the eye movement data, with a greater preference for the target word when it was heard in its canonical form in the exposure phase and a greater preference for the target word when it was produced with the more familiar reduction type in the test phase.

#### 2.2.4. Exposure phase

Only an overview of the results will be given here. For detailed analyses, see Appendix B. Both groups reacted equally fast to /b/-words, regardless of whether they were reduced or not. The RT data showed an effect of reduction, however, for the syllabic reduction condition. That is, participants in the syllabic reduction group were slowed down more strongly by reduced *ver*-words compared to participants in the segmental reduction group who heard canonical *ver*-words. The eye-tracking data revealed significant effects of reduction for both conditions: Participants in the segmental reduction group (who heard reduced /b/-words)

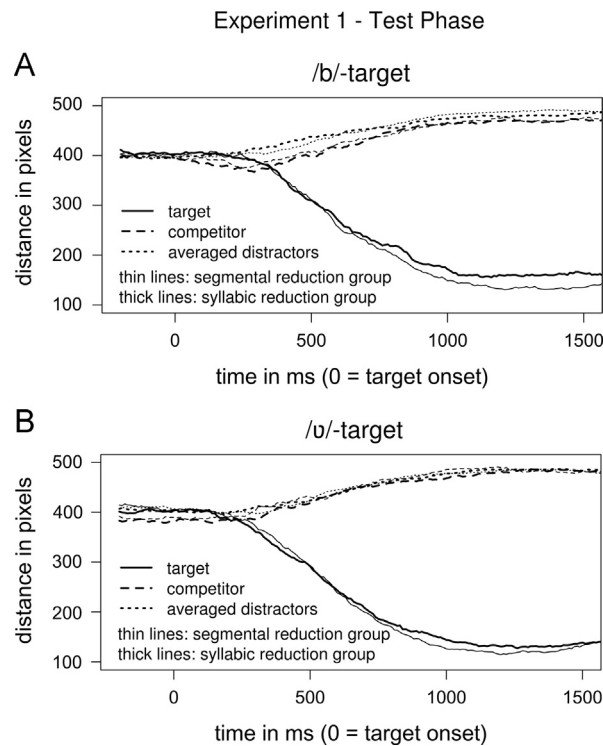


Fig. 6. Distances between fixations and targets, competitors and averaged distractors in pixels for the /b/- and /u/-targets in the test phase of Experiment 1.

showed a smaller target preference for /b/-words than participants in the syllabic reduction group (who heard canonical /b/-words). The reverse was true for the syllabic reduction condition: Participants in the syllabic reduction group (who heard reduced *ver*-words) had a smaller target preference for *ver*-words than participants in the segmental reduction group (who heard canonical *ver*-words).

### 2.2.5. Test phase

The click responses for the test phase are displayed in Fig. 5 in terms of mean RTs for both groups and each type of target word. In the test phase, both groups heard reduced /b/- and reduced *ver*-words, and the stimuli in the “competitor” conditions were manipulated: The [ʊ]s in /u/-words were replaced by [b/ʊ]s and the [f]s in /f/-words were replaced by [f:]s. Although the syllabic reduction group was numerically faster in responding than the segmental reduction group, this difference was again not significant, as there was no main effect of Group in either of the two conditions. The main effect of Trial Number was not significant either, indicating that participants did not react faster over the course of the test phase.

In the segmental reduction condition, neither the main effect of Target Word nor the main effect of Group was significant. That is, both groups responded to quasi-canonical /u/-words and to reduced /b/-words equally fast. Comparing the RTs for reduced *ver*-words and quasi-canonical /f/-words in the syllabic reduction condition, it can be seen that both groups took longer to respond to the reduced *ver*-words. This difference was significant (i.e., there was a main effect of Target Word:  $b_{\text{Target Word}} = -0.20$ ,  $SE = 0.03$ ,  $t = -6.8$ ,  $p_{\text{MCMC}} < 0.001$ ). The interaction of Target Word and Group was not significant in either condition. Thus, we did not observe an effect of exposure in the click responses of the test phase for the two conditions.

Figs. 6 and 7 display the corresponding eye-tracking results of the test phase. In the segmental reduction condition (Fig. 6A and B), participants in both groups behaved very similarly until approximately 750 ms after target onset, as the target lines are close to each other until that point in time. Then the target lines of the two groups diverge. There was, however, no main effect of Group in any time window indicating that one group looked more to the targets than the other. However, the main effect of Target Word was significant in the full 200–1500 ms time window ( $b_{\text{Target Word}} = 0.17$ ,  $SE = 0.07$ ,  $t = 2.6$ ,  $p_{\text{MCMC}} < 0.05$ ), indicating that both groups had a greater target preference for the /u/-targets (mapped onto 0.5, which results in a positive value when multiplied with the positive  $b$ -value) than for the /b/-targets. The interaction between Target Word and Group was not significant. That is, we did not find an effect of exposure in the segmental reduction condition.

In the syllabic reduction condition (Fig. 7A and B), the two groups differ in their looking behavior to the *ver*-targets from around 500–800 ms after target onset, whereas they seem to fixate the /f/-targets equally well. The main effect of Group was not significant in any time window. However, the main effect of Target Word was significant from 300 ms after target onset onwards (i.e., to 1500 ms;  $b_{\text{Target Word}} = 0.49$ ,  $SE = 0.06$ ,  $t = 8.5$ ,  $p_{\text{MCMC}} < 0.001$ ). That is, participants in both groups had a greater target preference for the /f/- than for the *ver*-targets. Importantly, the interaction between Group and Target Word was significant in the time window from 1100–1500 ms after target onset ( $b_{\text{Group} \times \text{Target Word}} = -0.25$ ,  $SE = 0.10$ ,  $t = -2.5$ ,  $p_{\text{MCMC}} < 0.05$ ). That is, we did find an effect of exposure in the syllabic reduction condition.

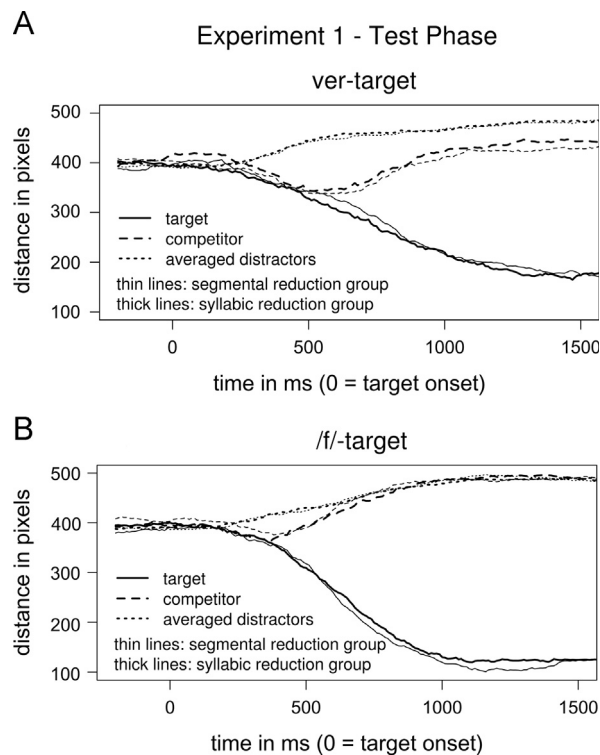


Fig. 7. Distances between fixations and targets, competitors and averaged distractors in pixels for the *ver*- and */f/*-targets in the test phase of Experiment 1.

### 2.3. Discussion

In Experiment 1, we found that both kinds of reduction, segmental and syllabic, initially hinder word recognition. The effects of reduction observed in the exposure phase indicate that reduced */b/*-words were harder to recognize than canonical */b/*-words. Similarly, reduced *ver*-words were harder to recognize than canonical *ver*-words. The finding that listeners struggle to process reduced variants also holds for the test phase where reduced */b/*- and *ver*-words were harder to recognize than quasi-canonical */u/*- and */f/*-words. This is consistent with other studies observing reduction costs in perception (Brouwer, Mitterer, & Huettig, 2012; Ernestus et al., 2002).

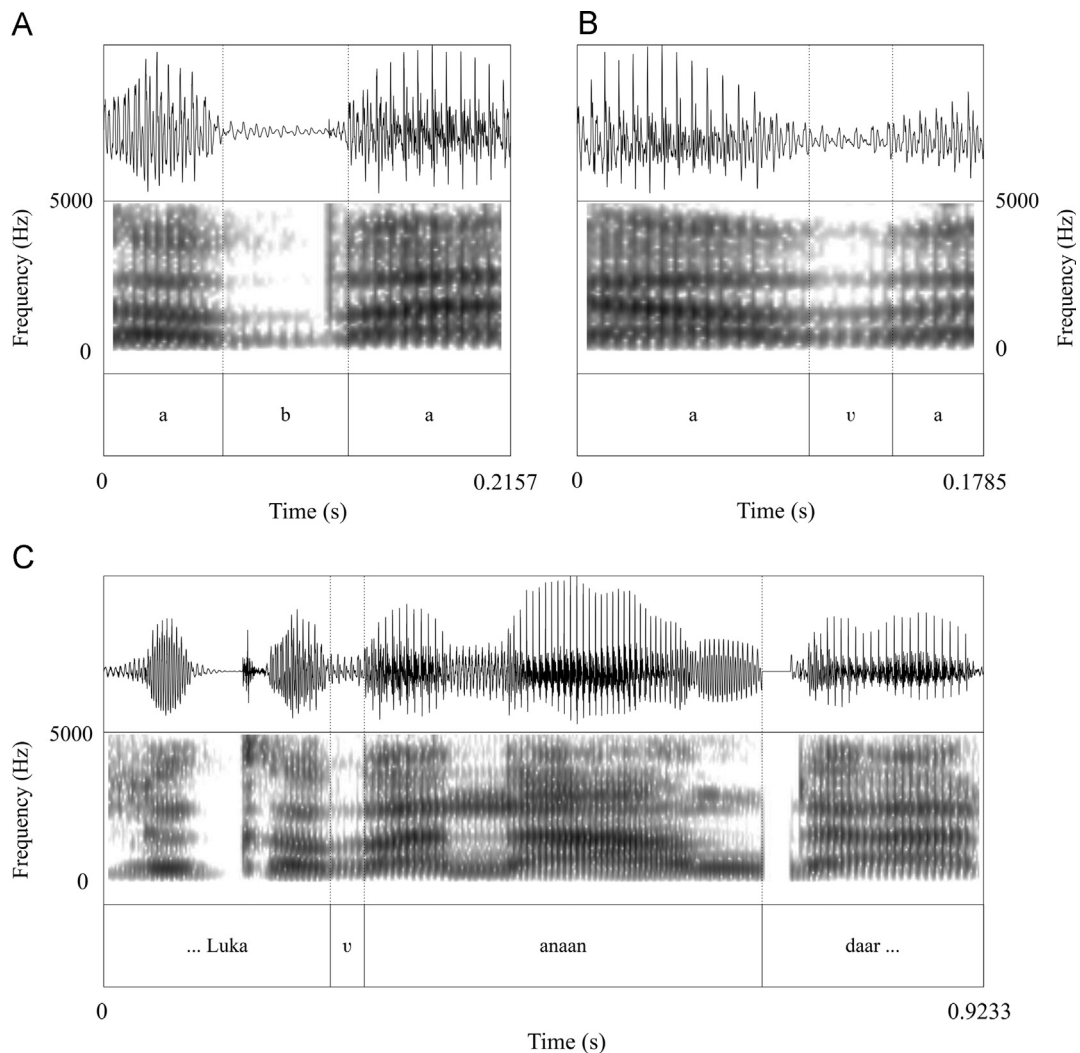
In the test phase, we observed an adaptation effect for the syllabic reduction group. The eye-tracking data show that the syllabic reduction group had an advantage over the segmental reduction group when they heard reduced *ver*-words: The syllabic reduction group, who had already been exposed to reduced *ver*-words, showed a greater target preference than the segmental reduction group who had heard the reduced sounds only in */f/*-words. Thus, listeners are able to adapt to syllabic reductions. But we did not observe a similar adaptation effect in the segmental reduction group: Participants in this group did not have a greater target preference for */b/*-words than participants in the syllabic reduction group. This null effect of learning for the segmental reductions is quite surprising given the learning effect for the syllabic reductions. It should be easier to adapt to segmental reductions, as they do not deviate as much from canonical pronunciations as syllabic reductions do. Furthermore, most models of spoken-word recognition have prelexical representations of individual segments, but not of larger units; according to such accounts learning about segments ought to be easier than learning about syllables.

The reason for the null effect in the segmental reduction condition might lie in our materials. To be more precise, the reason might lie in the articulation of our speaker: According to the corpus study we conducted before Experiment 1, a */b/-to-[u]-reduction* was most likely to occur after a vowel (40% of the cases) or a nasal (27% of the cases), that is, after a sonorant voiced segment. A closer look at the Experiment 1 stimuli revealed, however, that our speaker systematically devoiced not only all the */v/*s but also all the schwas at the end of *woordje*. As the unvoiced schwa is phonetically not a sonorant voiced segment, the */b/*s of the following */b/*-targets did not appear in an appropriate context for reduction to occur. Experiment 2 was run to address this issue and to replicate the adaptation effect for syllabic reductions.

## 3. Experiment 2

In Experiment 2 we wanted to test whether the null effect for the segmental reduction condition in Experiment 1 was due to an inappropriate reduction context. The same critical target words and design were used, but the words were embedded in a new carrier sentence template: “*Nu zegt Luka... daar links/rechts van het rondje/sterretje/driehoekje/vierkantje*” (literally: “Now says Luka ... there to the left/right of the circle/star/triangle/rectangle”). The color adjective was omitted to shorten the experiment. Importantly, the





**Fig. 8.** Example for cross-splicing of an experimental sentence containing a /b/-reduction (*banaan* "banana") in Experiment 2. (A) shows the waveform of a natural *a.ba*-sequence as it occurs in "Luka banaan". (B) shows the waveform of a naturally reduced *a.ba*-sequence as it occurs in *cabaret*. (C) displays the partial waveform of an experimental sentence containing a spliced-in naturally reduced [b] and shows all splicing points.

beginning of the sentence was changed so that the pre-target word ended in a full vowel (*Luka*); full vowels are less prone to reduction than the pre-target schwa used in Experiment 1. After an open jaw position, as is necessary for /a/, achieving a closure for the next consonant represents an effort for the speaker. Thus, the /b/-targets appeared now not only in an appropriate phonological context (i.e., in intervocalic position) but also in a phonetic context which should make reductions more likely.

### 3.1. Method

#### 3.1.1. Participants

Fifty-two Dutch native speakers were recruited from the Max Planck Institute subject pool and paid for their participation. The segmental reduction group consisted of three males and 23 females, aged 18–25 years (mean: 21 years). The syllabic reduction group consisted of nine males and 17 females, aged 18–28 years (mean: 21 years). The participants reported normal hearing and normal or corrected-to-normal vision. None of the participants had taken part in the previous experiment.

#### 3.1.2. Stimulus construction

The new sentences were recorded digitally by a female native speaker of Dutch who also devoiced [v] to [f] naturally. The sentence accent remained on the side disambiguation word (*links/rechts*). The recording procedure was virtually identical to the one described in Experiment 1. However, knowing from earlier recordings that this speaker reduces word-internal /b/s naturally to labiodental approximants, we did not ask her to produce reductions consciously. Instead, we let her additionally record words containing the sequence *a.bV*, where the *a* occurs in the coda of one syllable, the *b* in the onset of the next, unstressed syllable followed by one of the vowels occurring in the original /b/-target words (e.g., *cabaret*; the underlined sequence corresponded for instance to "... *Luka banaan* ..." in one of the experimental sentences). Without instructions, the speaker reduced these word-internal [b]s naturally. These naturally reduced [b]s were then extracted and spliced onto the beginnings of the corresponding /b/- and /v/-initial target words (see

Fig. 8). The pitch of each reduced [b] (a labio-dental approximant) was matched to the surrounding context using the PSOLA algorithm in PRAAT. Intensity was changed where necessary so that the difference in intensity between the reduced [b] and the following vowel corresponded to the difference measured in the source word (where the reduction occurred naturally, e.g., *cabaret*). If there was no Dutch word containing a given sequence *a.bV* word-internally, so that there was no naturally reduced [b] for that vowel, either the reduced [b] preceding a neighboring vowel (in the vowel space) or the reduced [b] preceding schwa was used. Contrary to Experiment 1, the naturally reduced [b]s used for reduced /b/-words and quasi-canonical /u/-words in Experiment 2 were not pretested for ambiguity.

As the speaker was not asked to produce reductions consciously, we had to create the *ver*-reductions digitally. Therefore, the longest [f] of a set of natural /f/-words sharing the first vowel was selected and lengthened by 20%. This long [f] was then spliced onto all /f/- and *ver*-words containing the corresponding vowel in second position.

Contrary to Experiment 1, every target word had its own token of the first part of the carrier sentence (*Nu zegt Luka*). For the last part of the carrier sentence indicating the position of the geometrical shape (e.g., *daar rechts van het driehoekje*), however, only one version for each combination of position and shape was used. If there was a sudden change in amplitude at the transition between the preceding vowel and the spliced-in reduced [b] or the spliced-in reduced [b] and the following vowel, it was smoothed by adding an amplitude decrease to the corresponding vowel.

### 3.1.3. Design and procedure

The design was the same as in Experiment 1. The procedure also corresponded to the earlier one, with only small changes. Eye movements of one eye were now recorded with an SR Research EyeLink 1000, sampling at 1 kHz. The right eye was tracked for all participants. As the color adjective was omitted in the carrier sentence, so were the color trials among the fillers. Consequently, a reduced number of 30 fillers were interspersed between the 96 exposure and 96 test trials. An experimental session now took approximately 30 min.

## 3.2. Results

### 3.2.1. Exclusion criteria and statistical testing

Thirty trials (0.3%) were excluded due to imprecise mouse clicks (i.e., not within a circle of 180 pixels around one of the centers of the four words on the screen). In Experiment 2, 100 pixels correspond to 4 cm due to a different screen size; average word length was again 165 pixels. Only trials in which participants clicked on the target word were included in the eye movement analyses. In trials where participants did not fixate the screen, the respective *x*- and/or *y*-value was set to the corresponding edge of the screen. Outliers were excluded based on RT criteria following the procedure used in Experiment 1. Thus, in the exposure phase, trials whose residuals were higher than 1500 ms or lower than −1000 ms were removed (13 trials). In the test phase, 10 trials were excluded as their residuals were either below −1200 ms or above 1800 ms. These values were chosen based on visual inspection of a residual plot.

Participants in both groups performed the task accurately. The percentages of correct responses (clicks on the target word) equaled or exceeded 92% in all conditions for both exposure and test phases. Descriptive statistics of RTs for correct trials are displayed in Table 3 for both groups in the exposure and test phases. This time, participants in the segmental reduction group responded faster than participants in the syllabic reduction group (see also Figs. 9 and C1). The overall lower mean RTs in this experiment are probably due to the fact that the female speaker of Experiment 2 spoke faster than the male speaker of Experiment 1. Moreover, the sentences did not contain information about the color of the geometrical shape anymore and were thus shorter than in Experiment 1. Both facts led to faster RTs for those subjects that clicked only once the sentence was finished.

The same models and the same procedure as in Experiment 1 were used to analyze the data. Groups and target words were contrast coded in the same way as before.

### 3.2.2. Exposure phase

Only a summary of the exposure phase results follows. For detailed analyses, see Appendix C. In the RT and eye-tracking data, effects of reduction became evident: The segmental reduction group (who heard reduced /b/-words) showed a smaller target preference for /b/-words than the syllabic reduction group (who heard canonical /b/-words). Vice versa, the syllabic reduction group (who was exposed to reduced *ver*-words) took longer to respond to and showed a smaller target preference for *ver*-words than the segmental reduction group (who heard canonical *ver*-words).

**Table 3**  
Descriptive statistics of RTs for correct trials for both groups in the exposure and test phases in Experiment 2.

RT in ms	Segmental reduction group		Syllabic reduction group	
	Exposure	Test	Exposure	Test
mean	1493	1641	1633	1753
sd	493	526	426	510
min	500	571	383	458
max	5574	3853	3753	7767

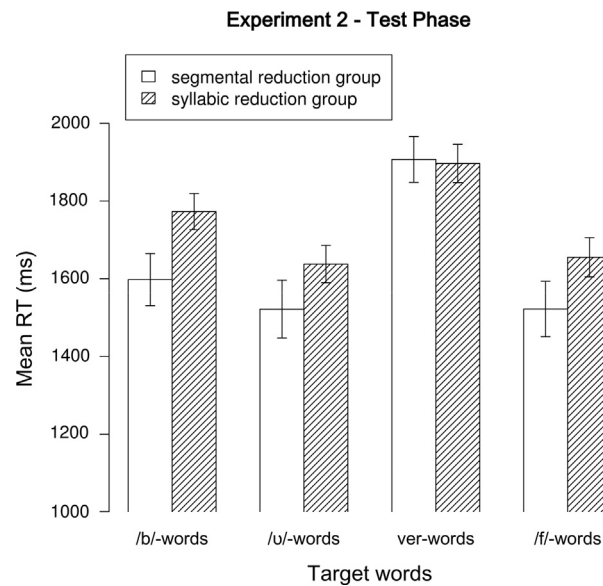


Fig. 9. Mean RTs in ms and SE of both groups for each type of target word in the test phase of Experiment 2. All /b/- and *ver*-words were reduced. All /u/- and /f/-words were presented in quasi-canonical pronunciation.

### 3.2.3. Test phase

Fig. 9 displays the mean RTs of the click responses for both groups. The segmental reduction group responded faster than the syllabic reduction group for every word type except for the (reduced) *ver*-words. Compared to the other word types, it is also the *ver*-words that both groups took much longer to react to. Moreover, the syllabic reduction group was also strongly slowed down by the now reduced /b/-words. Statistical analyses showed that in the segmental reduction condition (/b/- and /u/-targets), the main effects of Group and Target Word were significant ( $b_{\text{Group}}=0.11$ ,  $SE=0.05$ ,  $t=2.1$ ,  $p_{\text{MCMC}} < 0.01$ ;  $b_{\text{Target Word}}=-0.07$ ,  $SE=0.02$ ,  $t=-3.6$ ,  $p_{\text{MCMC}} < 0.001$ ). Thus, in responding to reduced /b/- and quasi-canonical /u/-words, the segmental reduction group (mapped onto -0.5) was significantly faster than the syllabic reduction group. Furthermore, both groups reacted faster to /u/-words (mapped onto 0.5) than to /b/-words. The interaction between Group and Target Word was not significant. The main effect of Group shows that the segmental reduction group dealt better with the reduced /b/-words and the quasi-canonical /u/-words. Apparently, they managed to adapt to reduced /b/s without assuming that each approximant must be an underlying stop. The main effect of Trial Number was not significant. That is, participants did not get faster in responding to /b/- and /u/-words.

In the syllabic reduction condition (*ver*- and /f/-words), the main effect of Group was not significant. So, no group was faster than the other when reacting to both *ver*- and /f/-words. In contrast, the main effect of Target Word was significant ( $b_{\text{Target Word}}=-0.19$ ,  $SE=0.03$ ,  $t=-7.5$ ,  $p_{\text{MCMC}} < 0.001$ ), indicating that both groups responded faster to quasi-canonical /f/- than to reduced *ver*-words. The interaction between Group and Target Word was also significant ( $b_{\text{Group} \times \text{Target Word}}=0.09$ ,  $SE=0.03$ ,  $t=3.1$ ,  $p_{\text{MCMC}} < 0.001$ ). This interaction is driven by the segmental reduction group, as participants in this group were slowed down more strongly by *ver*-reductions which they had not encountered in the exposure phase than the syllabic reduction group (see the four rightmost bars in Fig. 9). Put another way, the overall disadvantage of the syllabic reduction group is reduced for the items on which they have been trained (i.e., the reduced *ver*-words). The main effect of Trial Number was not significant, indicating that participants did not react faster to *ver*- and /f/-words during the test phase.

That is, in the syllabic reduction condition, the overall advantage of the segmental reduction group is modulated in a way that indicates that experience with a reduction type speeds up RT.

The corresponding eye-tracking data for the test phase are displayed in Figs. 10 and 11. In the segmental reduction condition (/b/- and /u/-words), both groups first fixate the competitor (dashed lines) when hearing a reduced /b/-word, before looking to the target at around 450 ms after target onset (see Fig. 10A). Later on, the segmental reduction group (thin lines) looked closer to the /b/-targets. When hearing a quasi-canonical /u/-word, the two groups behaved very similarly (the target lines overlap, see Fig. 10B). Statistical analyses revealed a marginally significant main effect of Group in the time window from 600–800 ms ( $b_{\text{Group}}=-0.15$ ,  $SE=0.08$ ,  $t=-1.9$ ,  $p_{\text{MCMC}}=0.055$ ). That is, the segmental reduction group (mapped onto -0.5) showed a greater target preference for both reduced /b/-words and quasi-reduced /u/-words than the syllabic reduction group. Moreover, there was a main effect of Target Word in all time windows ( $b_{\text{Target Word}}=0.25$ ,  $SE=0.05$ ,  $t=4.9$ ,  $p_{\text{MCMC}} < 0.001$ ) indicating that both groups looked more to the /u/-words than to the /b/-words. The interaction between Group and Target Word was significant from 800 ms to 1300 ms after target onset ( $b_{\text{Group} \times \text{Target Word}}=0.30$ ,  $SE=0.12$ ,  $t=2.5$ ,  $p_{\text{MCMC}} < 0.01$ ). Thus, in contrast to Experiment 1, we found an effect of exposure in the segmental reduction condition. The segmental reduction group looked closer to the reduced /b/-targets presumably because they had more experience with /b/-reductions than the syllabic reduction group, who in fact had a small advantage when /u/-words were the targets.

In the syllabic reduction condition (*ver*- and /f/-words), both groups looked first to the competitor (dashed lines) when hearing a reduced *ver*-word, but later on, the syllabic reduction group (thick lines) looked closer to the *ver*-targets (see Fig. 11A). When an /f/-word was the target, it was the segmental reduction group who looked closer to the targets (see Fig. 11B). Statistical analyses did

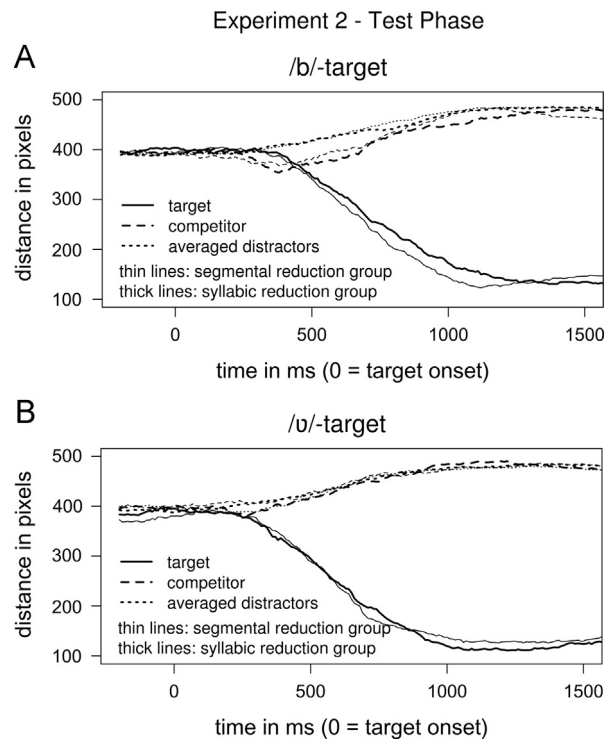


Fig. 10. Distances between fixations and targets, competitors and averaged distractors in pixels for the /b/- and /v/-targets in the test phase of Experiment 2.

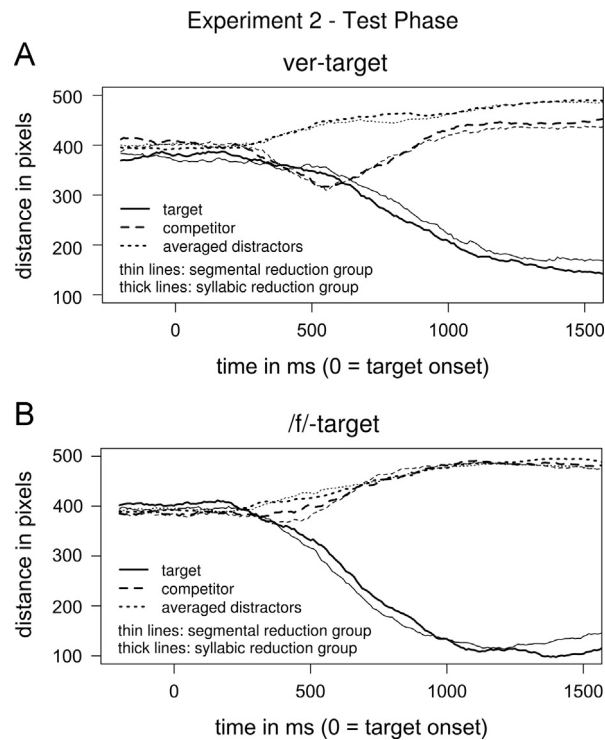


Fig. 11. Distances between fixations and targets, competitors and averaged distractors in pixels for the ver- and /f/-targets in the test phase of Experiment 2.

not reveal a main effect of Group. That is, no group looked more to the targets (when averaged over both conditions) than the other. In contrast, there was a main effect of Target Word from 300 ms to 1500 ms after target onset ( $b_{\text{Target Word}} = 0.45$ ,  $SE = 0.07$ ,  $t = 6.5$ ,  $p_{\text{MCMC}} < 0.001$ ) indicating that both groups had a greater target preference for the /f/-words (mapped onto 0.5) than for the ver-words. The interaction between Group and Target Word was significant in a time window from 600 ms to 900 ms after target onset ( $b_{\text{Group} \times \text{Target Word}} = -0.28$ ,  $SE = 0.13$ ,  $t = -2.1$ ,  $p_{\text{MCMC}} < 0.05$ ). Thus, we replicated the effect of exposure for the syllabic reduction condition found in Experiment 1.



### 3.3. Discussion

Experiment 2 was conducted to investigate whether the null effect of exposure for the segmental reduction condition found in Experiment 1 was due to an inappropriate phonetic context for the *b*-to-[*u*]-reduction. There were three minor changes between experiments. First, the sentences were shortened by omitting the color indication of the geometrical shapes. Therefore, the number of filler sentences was also reduced. Second, a different speaker, who reduced word-internal intervocalic [*b*]s naturally to [*u*]s, was recorded. It seems unlikely that the change of speaker (gender) should matter when it comes to adaptation to a single individual. Third, small changes were made in the way both sets of materials were made. For the segmental reductions, instead of the duration-manipulated stimuli used in Experiment 1, naturally occurring */b*-reductions were spliced onto the */b*- and */u*-stimuli. For the syllabic reductions, instead of the consciously produced *ver*-reductions used in Experiment 1, duration-manipulated material was used to create the *ver*- and */f*-stimuli. The major change in procedure concerned the carrier sentence. In Experiment 1, attention was paid to the fact that the word preceding the targets ended in a vowel (the schwa of *woordje*). However, as the speaker in Experiment 1 devoiced this schwa, a carrier sentence was chosen that ended in a full open vowel (*Luka...* instead of *woordje...*) to ensure that an appropriate phonetic context for the occurrence of */b*-reductions was given.

In the exposure phase, the data show inhibitory effects of reduction on word recognition, similar to the findings from Experiment 1. The target preference for */b*-words was larger for the syllabic reduction group, who heard them in the canonical form, than for the segmental reduction group, who heard them in the reduced form. The same is observed for the *ver*-words. Here, the segmental reduction group, who heard these words in the canonical form, showed a larger target preference than the syllabic reduction group, who heard these words in the reduced form. These reduction costs are also reflected in the RT data. That is, the new implementation of the reductions again hindered word recognition.

In the test phase, we again observed an adaptation effect for the syllabic reduction condition: The interaction between Group and Target Word, which reflects an effect of exposure, was significant not only in the eye-tracking data but also in the RT data. That is, the syllabic reduction group could make use of their prior exposure to reduced *ver*-words and recognized new reduced *ver*-words faster and had a greater target preference than the segmental reduction group for whom the *ver*-reductions were new. Thus, we replicated the adaptation effect for syllabic reductions found in Experiment 1.

Importantly, we also observed an adaptation effect for the segmental reduction condition: The analyses of the RT and the eye-tracking data revealed an effect of exposure. That is, the segmental reduction group, who had already heard */b*-reductions in the exposure phase, responded faster to new reduced */b*-words and showed a greater target preference than the syllabic reduction group who had heard the reduced sounds only in */u*-words. This effect was larger and stretched out over a longer period of time than the replicated adaptation effect for syllabic reductions. While we cannot say exactly why there was no effect in this condition in Experiment 1, the results of Experiment 2 suggest that learning is more likely for phonetically more realistic stimuli. Furthermore, the fact that a change in the carrier sentence (i.e., the word preceding the reduced */b*-target ending in a full vowel instead of a devoiced schwa) led to the observation of an adaptation effect for */b*-reductions shows how context-sensitive listeners are. In an abstractionist account, this context-sensitivity might be implemented at the prelexical level with respect to the representation of a reduced allophone of */b*. The existence of such a reduced allophone would be based on prior experience with this reduction type in intervocalic contexts and this contextual conditioning could also be represented. Listeners would therefore be more likely to adapt to reduced */b* when hearing it in its natural context. The suggestion that there are context-specific allophones is compatible with recent findings suggesting that the targets of recalibration may not be abstract phonemes (Mitterer, Scharenborg, & McQueen, 2013; Reinisch, Wozny, Mitterer, & Holt, 2014).

Figs. 10A and 11A show that both groups first look to the competitor (dashed lines) when hearing reduced */b*- and *ver*-words. This suggests that listeners first interpret the incoming speech according to the acoustical information they get as [*u*] or [*f*] respectively. Only later, they re-evaluate their categorization as the rest of the word becomes available (for the untrained group) or learning intervenes (for the trained group).

It is important to note that, in both experiments, the learning effects could be due not only to exposure to one or other reduction, but also to exposure to the quasi-canonical stimuli. The segmental reduction group might recognize reduced */b*-words better than the syllabic reduction group because of what the segmental reduction group heard during exposure (reduced */b*-words), or because of what the syllabic reduction group heard (quasi-canonical */u*-words), or both, and likewise for the reduced *ver*-words. This design was chosen to maximize the contrast between the two exposure groups, following the perceptual-learning design of Norris et al. (2003). Note, however, that both groups received evidence in the exposure phase that the speaker also pronounced clear */u*s (in *woordje*) and clear */f*s (in *van* and *vierkantje*). It thus seems unlikely that the groups' adaptation was based solely on the quasi-canonical forms (because they got inconsistent input regarding the speaker's pronunciation of */u*s and */f*s). The idea that the quasi-canonical forms were not essential for learning is consistent with the demonstration by Norris et al. (2003) that contrast (e.g. between ambiguous */f* and unambiguous */s* in their design) is not necessary for perceptual learning – exposure to the ambiguous sounds alone was enough for learning to take place. It therefore seems most likely that the effects observed here are due primarily to learning about the reduced forms and not to learning about the quasi-canonical forms. Note that we have also observed adaptation to reductions if the contrast sounds (corresponding to the quasi-canonical stimuli in the present study) were not manipulated (Poellmann, Mitterer, & McQueen, 2014).

If part of the observed learning effects is attributable to adaptation to the quasi-canonical forms, this would mean that listeners can adapt easily to variant forms that do not occur naturally in the language. This, in turn, would suggest a rather general mechanism for perceptual learning. The contribution of the quasi-canonical stimuli, however, seems to be minimal given the results of Experiment 1.

There, no learning was observed for the segmental reductions. The larger part of the learning effect in Experiment 2 is therefore likely to be caused by the reduced forms rather than the quasi-canonical forms. Nevertheless, adaptation to a wide range of non-canonical variation in speech has been found for genuine and arbitrary foreign accents (e.g., Weber, Di Betta, & McQueen, *in press*) and for speaker idiosyncrasies (e.g., McQueen, et al., 2006). However, adaptation to such variation did not occur if a native speaker produced similar mispronunciations to a non-native speaker (Eisner, Melinger, & Weber, 2013) or if the variant pronunciation was attributable to incidental factors (like a speaker holding a pen in his or her mouth; Kraljic, Samuel, & Brennan, 2008). These findings suggest that listeners do not adapt blindly to any variation that might occur in speech.

#### 4. General discussion

Previous studies investigating mechanisms compensating for reduced speech have focused on phonetic and language-dependent characteristics. The current study is the first to investigate a learning-based compensation mechanism which takes the speaker into account. Reduced pronunciation variants are a phenomenon of stylistic variation that depend on individual speakers but are not inherent to them. That is, it is the speaker's choice whether he or she reduces certain words or parts thereof (Mitterer & Ernestus, 2006). In the exposure phase, each group of participants heard the speaker produce a certain type of reduction (either segmental, i.e., /b/ → [u], or syllabic, i.e., *ver-* → [f], reductions). Listeners also got evidence that the respective other reduction type was less likely for this speaker – evidence they obtained through hearing the reduced sounds in the contrast words (e.g., [f:] in /f/-words). In the test phase, the speaker produced two types of reduction. In this test phase, the syllabic reduction group was better than the segmental reduction group in recognizing new reduced *ver*-words. This effect was found in both experiments. In Experiment 1, we did not find a similar advantage for the segmental reduction group. Participants in this group were not better at recognizing new reduced /b/-words than the syllabic reduction group. In Experiment 2, therefore, we changed the carrier sentence so that the potentially reduced segment was preceded by a full vowel. This change in phonological and phonetic context made /b/-reductions more likely, and an adaptation effect was then found for the segmental reduction group. Participants in this group were now better at recognizing new reduced /b/-words than the syllabic reduction group. Listeners are thus able to adapt to specific reduction styles of a speaker and this helps them to compensate for speech that has been reduced in a way consistent with prior exposure.

The current study thus extends experimental evidence of perceptual learning for naturally-occurring variation. Mitterer and McQueen (2009a), for example, showed that listeners can adapt to an unfamiliar regional accent in a second language. They exposed Dutch participants to either Scottish or Australian English videos with Dutch, English or no subtitles. The subsequent test phase showed that the lexical information provided by the English subtitles enhanced adaptation effects, whereas the information provided by the Dutch subtitles, which was inconsistent with the spoken English word forms, hindered adaptation. It is unclear, however, which specific phenomena the listeners tuned in to. Dahan et al. (2008) showed that listeners can tune in to regional variation of a given speaker's native language. In their study, listeners adapted to a regional phenomenon (the raising of the vowel /æ/ before /g/, but not before /k/) which facilitates word recognition by reducing ambiguity in the initial part of words like *bag* and *back*. The current study is the first to show a similar adaptation effect for stylistic variation. Listeners attuned to segmental and syllabic reductions which, importantly, increased ambiguity between existing words (e.g., *baron* pronounced as [uaron] can easily be confused with the Dutch word [uaron] *waarom* “why”). This increased ambiguity was found not only for extreme cases such as the quasi-minimal pairs used in the test phase, but also in the exposure phase where reduced forms hindered word recognition even in target–“competitor” pairs of hardly overlapping words (e.g., the target *bikini* pronounced as [vikini] and canonical [vʊstənɛɪ] *woestenij* as “competitor”). Hence, the current study is also the first to show that perceptual learning not only applies to artificially constructed ambiguous sounds that were not modeled on real-life mispronunciations (e.g., McQueen, et al., 2006; Norris et al., 2003) and to naturally-occurring pronunciation deviants that are unambiguous (Maye et al., 2008), but also to ambiguous sounds that occur in natural continuous speech.

Importantly, perceptual learning about segmental and syllabic reductions was applied to previously unheard words in the test phase. Listeners could use the experience they gained with one type of reduction in the exposure phase to recognize new reduced words of the same reduction type in the test phase. McQueen et al. (2006) observed generalization of learning about ambiguous fricatives ([ʃ/ʒ]) to new words. They argued that, for learning about a sublexical unit to be applied to previously unheard words, an abstraction process concerning those units must take place at a prelexical level. Mitterer et al. (2011) conducted a similar perceptual learning experiment about an ambiguous Mandarin tone contour and tested the effect not only for new but also for repeated words. The learning effect was slightly larger for previously heard than for previously unheard words. This suggests that there are small additional effects of episodic learning compared to large effects of prelexical abstraction.

Based on the finding that learning about reductions generalized to new words, we would expect to observe the adaptation effect early in the eye-tracking data. However, we observed quite late effects, starting at the earliest around 600 ms after target onset. In contrast, competitor preferences, although not significant, were observed much earlier (as early as 200–250 ms after target onset). This seems to suggest that the bottom-up information in the speech signal has a strong influence on speech recognition (e.g., Marslen-Wilson, 1987) and learning needs time to override the signal-driven interpretation (i.e. “what-you-hear-is-what-you-get”).

However, Mitterer and Reinisch (2013), who investigated the time-course of perceptual learning about ambiguous fricatives, found early learning effects in their eye-tracking data. They concluded that perceptual learning is indeed perceptual rather than post-perceptual once it is complete. Mitterer and Reinisch (2013) observed these early learning effects using an eye-tracking paradigm with a classical exposure-test design. It thus seems unlikely that the late effects found in the present study are caused by the

combination of a printed-word eye-tracking task and a learning paradigm. Furthermore, the participants in the study by Mitterer and Reinisch (2013) had to make conscious metalinguistic judgments about the speech sounds they heard by clicking on printed words whose spoken realizations deviated from the canonical pronunciations. As this task did not affect the timing of the effects in Mitterer and Reinisch (2013), it seems unlikely that it should have done so in the experiments presented here.

If perceptual learning should and does occur early in general and methodological issues have been ruled out as possible causes, what might then be the reason for the late learning effects observed in this study? One point in which the present study differs from the study by Mitterer and Reinisch (2013) and other classical perceptual learning studies (e.g., McQueen et al., 2006) regards the mapping rules listeners must generate or adjust to correctly recognize the deviant sound(s). In the classical perceptual learning paradigm, listeners have to learn to include a previously unknown ambiguous sound into one of two possible phoneme categories. They must thus generate a new mapping rule for that sound, and that sound is supposed to be truly ambiguous. Ideally, it is judged as belonging to one of the two categories in 50% of the cases. Apart from the pretested [p<sub>u</sub>] sounds in Experiment 1 (for which no learning was found), none of the reduced sounds used in this study were truly ambiguous. The naturally occurring [p<sub>u</sub>]s used in Experiment 2 as the replacement for /b/ were likely more /u/-like than /b/-like. The long [f:]s replacing *ver*-s in Experiments 1 and 2 were clearly more /f/-like than *ver*-like. Listeners thus learned that a particular speaker was likely to pronounce a /b/ as a [u] and hence that an existing sound ([u]) mapped onto two categories for that speaker (i.e., /u/ and /b/). Their perception of a [u] might therefore have shifted from judging it as /u/ in most cases to judging it as /u/ in 80% and as /b/ in 20% of the cases. With this kind of learning, the initial signal-driven hypotheses still strongly favor the canonical form, and only when later-arriving segments rule that form out can the learning take effect. Therefore, as soon as listeners receive evidence that a particular sound can map onto more than one category, the learning process likely needs more time to take effect.

The generalization of learning about reductions across words is evidence for an abstractionist mode of lexical access and therefore also for abstractionist models of the mental lexicon. It suggests that compensation mechanisms for both segmental and syllabic reductions involve a phonologically abstract prelexical level. In the case of segmental reduction, prelexical mapping from an ambiguous reduced [b] to the segmental representation of a /b/ should work quite similarly to the mapping of ambiguous [ʃ/ɹ] to either /s/ or /f/. However, the mapping from a reduced *ver*- (i.e., from the single segment [f:] in our case) to the syllable *ver*- requires a larger prelexical representation than is specified in current abstractionist models (e.g., Gaskell & Marslen-Wilson, 1997; McClelland & Elman, 1986; Norris & McQueen, 2008). One can think of several possibilities regarding the nature of this additional prelexical representation. First of all, given the morphological status of *ver*-, this representation might be morphemic. This would be in line with models of word recognition which assume that morphemes are segmented at the prelexical level (Taft, Hambly, & Kinoshita, 1986). Prelexical morphemic representations are a necessary consequence of prelexical morphological parsing. The input [f:] could thus be mapped onto a representation of the morpheme *ver*-.

Another possibility, which in the case of *ver*- cannot be teased apart from the morphemic account, is that prelexical representations are syllabic. This is certainly a more extreme claim as then every syllable (and there are approximately 21,800 syllables in the Dutch language as calculated on the basis of the CELEX database (Baayen, Piepenbrock, & van Rijn, 1993)) would need its own representation.

A third option assumes that there is a more general prelexical process through which frequently-occurring sequences of segments are bound together. Prelexical processes may be sensitive to transitional probabilities between speech sounds (Pitt & McQueen, 1998). In a series of experiments, Vitevitch and Luce (1999) dissociated lexical and prelexical effects in spoken word recognition by investigating probabilistic phonotactics and neighborhood activation. Their findings support the idea that the prelexical level is sensitive to sequential probabilities. In our case, *ver*- has a very high transitional probability. Prelexical mechanisms would thus have to capture the fact that such high probability sequences can also be realized in a reduced form.

Finally, the prelexical representation might be a combination of segmental representations and an abstract representation of the syllable as a prosodic unit. During speech decoding, a Prosody Analyzer (Cho, McQueen, & Cox, 2007) may use the supra-segmental information available in the speech stream to compute the prosodic structure of the utterance. This structure serves then to unify sequences of segments. At the lowest level in the prosodic hierarchy, the syllable is bound to a combination of segments. In this process, it might be possible to match only one segment ([f:] in the case of *ver*-) to a syllable. Note that such mappings are already necessary for syllabic sonorants, such as in the word *bottle* [bɒ:tɫ], in which the final [ɫ] is syllabic. Dutch listeners are also confronted with syllabic consonants. In Groningen, for example, speakers may pronounce a verb like *laten* /latən/ “to let” as [latɫ].

Future research is required to disentangle these four possibilities concerning what we have called syllabic reductions. For example, it remains to be determined whether learning about supra-segmental units only takes place when those units are phonologically consistent and/or when they have morphological status, as is the case with the prefix tested here. The current results, however, already suggest that prelexical representations or processes are required that go beyond those assumed in current abstractionist models.

Our data show that there is abstraction in spoken-word recognition, and hence suggest, in keeping with other studies (e.g., McQueen et al., 2006; Mitterer et al., 2011), that there is more to this process than the storage of episodes of actual pronunciations. Variation phenomena such as reductions are often the *raison d'être* for episodic models. However, the present study suggests that naturally-occurring systematic variation, at least the segmental and syllabic reductions we examined, is dealt with by a process of phonological abstraction. Cutler et al. (2010) ran simulations with a purely episodic model (MINERVA-2; Hintzman, 1986) to test whether an exemplar-based theory could account for lexical generalization of perceptual learning about ambiguous fricatives. That was not the case. The current study presents similar kinds of learning processes. For the segmental reductions, the reduced element

is a sound in between a bilabial voiced stop and a labio-dental approximant. Learning for this ambiguous sound is likely to be very similar to the ambiguous sound [ʃ̥/ʃ] (between /s/ and /f/) studied by Cutler et al. It is thus unlikely that a purely episodic model could account for the lexical generalization for reduced segments found here, and thus also for the generalization about reduced syllables. Our data thus add to the evidence speaking against any model of the mental lexicon that does not provide a mechanism for generalization of learning to previously unheard words and suggest in particular that that mechanism concerns both segmental and supra-segmental structures.

Consider also, in contrast to classical abstractionist models, the variant proposed by [Ranbom and Connine \(2007\)](#): Their frequency-based account contains gradient and multiple phonological representations for each entry in the lexicon. Citation forms and pronunciation variants of words are stored, graded in strength according to experienced frequency. Thus, no special processes are required to compensate for variation such as reduction in the speech signal. This model, however, has problems accounting for our findings, as there is no mechanism for generalization.

Another abstractionist model that has difficulties dealing with our data is the featurally underspecified lexicon (FUL) model ([Lahiri & Reetz, 2002, 2010](#)). In this model, the entries in the mental lexicon are based on phonological features. The first segment of the Dutch word *baron*, for example, is specified in the lexicon as [labial] and [voice]. This segment is not explicitly specified as [plosive], as stops are assumed to be the default for manner of articulation. The features that are extracted for the first segment of the reduced speech input [ʋarɔn] are categorized as [labial], [continuant] and [voice]. The extracted feature [continuant] does not mismatch the underspecified representation in the lexicon. That is, according to the FUL model, it does not matter that /b/s are reduced to [ʋ]s. Therefore, there is no need for listeners to adapt to these /b/-reductions and adaptation should thus not occur. The fact that we find an adaptation effect for /b/-reductions contradicts the assumption of the FUL model.

Our data thus support fully specified abstractionist models which provide a mechanism for generalization and which include a representation of sequences larger than segments at the prelexical level. It is important to emphasize here, however, that abstractionist models relying completely on prelexical compensation for reductions cannot be the full story. Listeners will often encounter reductions, and especially reduced prefixes, from formerly unknown speakers. These need to be recognized as well, and there is evidence that storage of variant forms can contribute to that. [Mitterer and Russell \(2013\)](#) found that processing costs for reduced prefixes occur mainly for low-frequency words. This can easily be explained by an episodic model, because prefix reductions occur less often in low-frequency words ([Pluymaekers et al., 2005](#)). Such episodic storage, however, does not make prelexical reconstruction superfluous. It has been repeatedly found that, compared to reduced forms, canonical forms have an advantage in speech comprehension (e.g., [Ernestus, 2009](#); [Pitt, Dilley, & Tat, 2011](#)). Adapting to a given speaker's reduction style via prelexical reconstruction would allow listeners to access canonical forms, which would in turn facilitate recognition.

Taken together, the available data are therefore consistent with a hybrid model of speech perception, one that includes both abstractionist and episodic modes of processing, and inconsistent with either theoretical extreme. Prelexical abstraction offers an explanation for generalization of learning to previously unheard words, and hence a plausible account of how listeners recover from specific styles of reduction.

## 5. Conclusion

The present study provides the first evidence for prelexical perceptual learning about reductions, a phenomenon of stylistic variation which is common in everyday casual speech. We have shown that adapting to reduced forms involves more than just storing episodes of those forms, because learning generalized to new words containing the same reductions. We have therefore argued that abstract prelexical representations mediate the mapping of reduced forms onto the lexicon, and that these representations are the locus of the perceptual learning effect. We have also argued that models of spoken-word recognition need to be revised with respect to the nature of these representations. The demonstration that learning about syllabic reductions generalized to new words requires prelexical multi-segment representations.

## Acknowledgments

This work was funded by the DFG (Deutsche Forschungsgemeinschaft “German Research Foundation”) under the priority program 1234 “Phonological and phonetic competence: Between grammar, signal processing, and neural activity”. It is part of the first author's Ph.D. project. The second author obtained his M.Phil. in Linguistics at Leiden University on a preliminary version of Experiment 1 under the supervision of Holger Mitterer, James M. McQueen and Vincent J.J.P. van Heuven. This work was presented at the Nijmegen workshop on “Production and Comprehension of Conversational Speech” in December 2011.

## Appendix A. Critical test stimuli

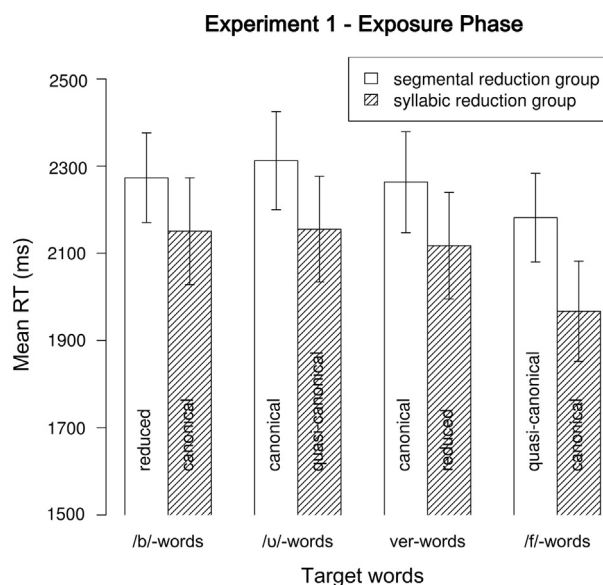
See [Table A1](#).



Table A1

Target and competitor words of the test phase for the syllabic and segmental reduction condition with their word frequency per million according to SUBTLEX-NL (Keuleers, Brysbaert, & New, 2010).

Target	English Translation	Word Frequency	Competitor	English Translation	Word Frequency
<i>Test phase: segmental reductions</i>					
baatzuchtig	selfish	0.00	waaninnig	insane	8.35
bagage	luggage	17.91	waartegen	against what	2.65
balans	balance	9.70	waarheen	where to	31.72
banaan	banana	5.33	waarnaast	beside which	0.02
barak	barrack	3.54	waartoe	for which	3.96
baron	baron	7.68	waarom	why	2141.72
basaal	basic	0.07	waarna	whereupon	3.27
bavianen	baboons	0.94	waardeloosheid	worthlessness	0.02
balkondeur	balcony door	0.07	waxinekaars	tea light	0.00
baldadig	wanton	0.05	walhall	Walhall	0.66
Berlijn	Berlin	16.81	welaan	well then	0.09
bermuda	bermuda shorts	1.90	welzeker	certainly	0.18
benzine	petrol	24.65	wellevend	courteous	0.00
berinnen	female bears	0.05	weerspiegelen	to reflect	0.34
binnendoorweg	short cut	0.05	wisselvallig	changeable	0.37
binnenskamers	indoors	0.41	wispelturig	fickle	1.12
biljoenen	trillions	0.64	wilskrachtig	strong-minded	0.05
biljetten	bills	3.73	Wilhelmus	name of national anthem	0.00
biscuitblik	biscuit tin	0.00	wiskundig	mathematical	1.12
bijvoeglijk	adjectival	0.25	wijdmazig	wide-meshed	0.00
bijdehand	quick-witted	4.78	weinigzeggend	uninformative	0.02
boosaardigheid	malice	0.57	woonwerk-verkeer	commuter traffic	0.00
boerenjongen	country boy	0.57	woensdagavond	Wednesday evening	1.37
boerenkool	kale	0.14	woestenij	wasteland	1.60
<b>Average</b>		<b>4.16</b>	<b>Average</b>		<b>91.61</b>
<i>Test phase: syllabic reductions</i>					
verleuteren	waste one's time talking	0.00	vleugel	wing	8.83
verlustigen	to amuse	0.09	vluchtig	cursory	1.58
verlangzamen	to slow down	0.02	fladderen	to flutter	0.64
verlaging	reduction	0.16	vlagen	bursts	0.69
verlangen	(to) desire	15.66	flanken	flanks	1.10
verlelijken	to uglify	0.00	vieselijk	caral	0.27
verlening	grant	0.07	vlegel	flail	0.82
verleppen	to wither	0.02	flatje	little flat	1.39
verliefd	in love	107.16	vlieg	fly	29.07
verlies	loss	49.07	vlies	fleece	1.94
verliezer	loser	6.54	vlieger	kite	3.80
verlichting	lighting	4.28	vlinders	butterflies	3.22
verleider	seducer	0.62	vleien	to flatter	1.49
verloofd	engaged	14.38	vloot	fleet	13.17
verlof	leave	11.02	vlot	raft	6.43
verlokkelijk	tempting	0.21	vlokken	flakes	0.32
verlossen	to release from	4.53	flossen	to floss	0.85
verloedering	degradation	0.43	vloeibaar	liquid	2.42
verregenen	to rain off	0.00	vreemdeling	stranger	10.47
verrekenen	to settle	0.09	vreselijk	terrible	112.35
verrekken	to strain	1.60	vrekken	misers	0.14
verrichten	to perform	4.85	frictie	friction	0.11
verruilen	to swap	0.18	fruit	fruit	12.94
verroeren	to stir	0.71	vroege	formerly	124.45
<b>Average</b>		<b>9.24</b>	<b>Average</b>		<b>14.10</b>



**Fig. B1.** Mean RTs in ms and SE of both groups for each type of target word in the exposure phase of Experiment 1. The term “quasi-canonical” indicates that the stimuli of the contrast conditions were slightly manipulated: The [u]s of /u/-words were replaced by [ʊ]s and the [f]s of /f/-words were replaced by [f̥]s (for more details see main text).

## Appendix B. Detailed analysis of the exposure phase results of Experiment 1

Fig. B1 shows the mean RTs in the exposure phase for both groups and each type of target word sampled in two conditions (segmental and syllabic reduction condition). The syllabic reduction group responded faster than the segmental reduction group in both conditions. There was, however, no main effect of Group in either condition. In contrast, the main effect of Trial Number was significant not only in the segmental reduction condition ( $b_{\text{Trial Number}} = -0.003$ ,  $SE = 0.0002$ ,  $t = -15.6$ ,  $p_{\text{MCMC}} < 0.001$ ) but also in the syllabic reduction condition ( $b_{\text{Trial Number}} = -0.003$ ,  $SE = 0.0002$ ,  $t = -14.3$ ,  $p_{\text{MCMC}} < 0.001$ ) indicating that participants reacted faster over the course of the exposure phase.

For the segmental reduction condition, Fig. B1 shows that the syllabic reduction group responded to both /b/- and /u/-words equally fast, whereas the segmental reduction group was somewhat faster to respond to (reduced) /b/- than to (canonical) /u/-words. In this case, the main effect of Target Word was not significant. That is, participants, irrespective of group, did not react faster to either /b/- or /u/-words. Also the interaction between Target Word and Group was not significant. Thus, we did not observe an effect of reduction here: The segmental reduction group did not take significantly longer to respond to reduced /b/-words than the syllabic reduction group who heard them in their canonical form.

In the syllabic reduction condition, both groups needed more time to respond to ver- than to /f/-words, but this difference was bigger for the syllabic reduction group: The main effect of Target Word was significant ( $b_{\text{Target Word}} = -0.05$ ,  $SE = 0.01$ ,  $t = -4.1$ ,  $p_{\text{MCMC}} < 0.001$ ). Note that, because the ver-targets were mapped onto -0.5, a negative  $b$ -value indicates that both groups took longer to respond to the ver-targets (the multiplication of a negative  $b$ -value with -0.5 results in a positive estimate which needs to be added to the intercept). This main effect was moderated by a significant interaction between Target Word and Group ( $b_{\text{Group} \times \text{Target Word}} = 0.06$ ,  $SE = 0.03$ ,  $t = 2.4$ ,  $p_{\text{MCMC}} < 0.05$ ). That is, we found an effect of reduction, as the syllabic reduction group was slowed down more strongly by the ver-words, which they heard in their reduced form, than the segmental reduction group, who heard them in their canonical form.

The corresponding eye-tracking data for the exposure phase are displayed in Figs. B2 and B3. Figs. B2A and B show the results for the /b/- and /u/-target words (segmental reduction condition). While the mean fixation of the syllabic reduction group (represented by the thick lines) is closer to the /b/-words than the mean fixation of the segmental reduction group (represented by the thin lines, see Fig. B2A), the two groups do not differ in their looking behavior to the /u/-words (the two target lines overlap, see Fig. B2B). Statistical analyses showed that there was a main effect of Target Word in the time window from 700–1500 ms after target onset ( $b_{\text{Target Word}} = -0.11$ ,  $SE = 0.05$ ,  $t = -2.25$ ,  $p_{\text{MCMC}} < 0.05$ ). That is, both groups had a greater target preference for the /b/-words (mapped onto -0.5) than for the /u/-words. Moreover, the main effect of Group was significant in an early time window from 200–400 ms after target onset ( $b_{\text{Group}} = 0.15$ ,  $SE = 0.07$ ,  $t = 2.1$ ,  $p_{\text{MCMC}} < 0.05$ ). The positive  $b$ -value indicates that the syllabic reduction group (mapped onto 0.5) showed a greater target preference for the /b/- and /u/-targets than the segmental reduction group. The interaction between Group and Target Word was significant from 400–700 ms after target onset ( $b_{\text{Group} \times \text{Target Word}} = -0.24$ ,  $SE = 0.13$ ,  $t = -1.8$ ,  $p_{\text{MCMC}} < 0.05$ ). Thus, there was an effect of reduction in the segmental exposure phase, as participants in the segmental reduction group (hearing reduced versions; thin lines) had a smaller target preference for the /b/-words than the syllabic reduction group (hearing canonical versions; thick lines), while both groups behaved in the same way for the /u/-targets.

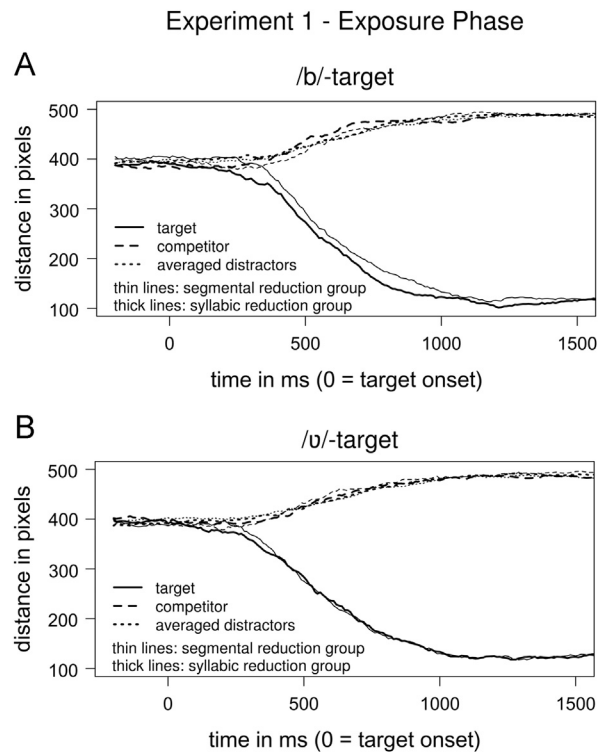


Fig. B2. Distances between fixations and targets, competitors and averaged distractors in pixels for the /b/- and /u/-targets in the exposure phase of Experiment 1.

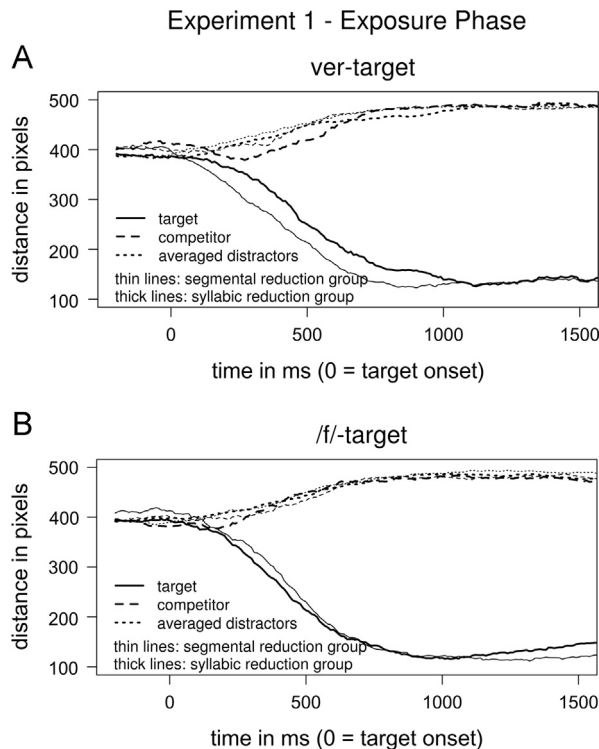
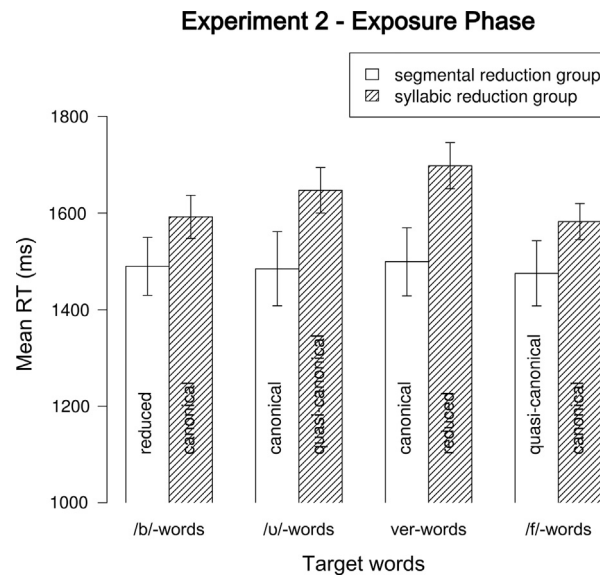


Fig. B3. Distances between fixations and targets, competitors and averaged distractors in pixels for the ver- and /f/-targets in the exposure phase of Experiment 1.

The eye-tracking results for the syllabic reduction condition of the exposure phase are shown in Figs. B3A and B. There is a clear difference in the looking behavior of the two groups when the target word is a *ver*-word (see the distance between the target lines in Fig. B3A). The segmental reduction group (thin lines) looked closer to the targets than the syllabic reduction group (thick lines) did. For the /f/-words, both groups behaved very similarly, with the segmental reduction group looking slightly less to the targets in the first 500 ms after target onset. Running the linear mixed-effects model revealed a late main effect of Target Word in the time window from 1100–1500 ms after target onset ( $b_{\text{Target Word}}=0.15$ ,  $SE=0.07$ ,  $t=2.2$ ,  $p_{\text{MCMC}}<0.05$ ). That is, both groups of participants had a



**Fig. C1.** Mean RTs in ms and SE of both groups for each type of target word in the exposure phase of Experiment 2. The term “quasi-canonical” indicates that the stimuli of the contrast conditions were slightly manipulated: The [u]s of /u/-words were replaced by naturally occurring /b/-reductions and the [f]s of /f/-words were replaced by [f:]s.

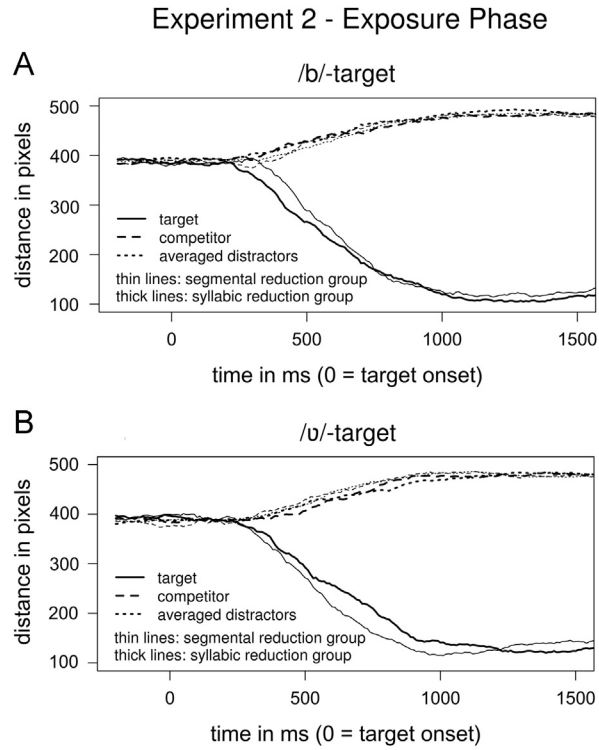
greater target preference for /f/-words (mapped onto 0.5) than for *ver*-words. Furthermore, a significant main effect of Group was found in the early time window from 200–400 ms after target onset ( $b_{\text{Group}} = -0.14$ ,  $SE = 0.07$ ,  $t = -2.0$ ,  $p_{\text{MCMC}} < 0.05$ ). As the  $b$ -value is negative, it is the segmental reduction group (mapped onto -0.5) that showed a greater target preference for the *ver*- and /f/-words than the syllabic reduction group. The interaction between Group and Target Word was significant in a time window from 200–900 ms ( $b_{\text{Group} \times \text{Target Word}} = 0.36$ ,  $SE = 0.09$ ,  $t = 4.2$ ,  $p_{\text{MCMC}} < 0.001$ ) and marginally significant in the latest time window from 1300–1500 ms ( $b_{\text{Group} \times \text{Target Word}} = -0.22$ ,  $SE = 0.11$ ,  $t = -1.9$ ,  $p_{\text{MCMC}} = 0.053$ ). So we also found an effect of reduction in the syllabic exposure phase, as participants in the segmental reduction group (thin lines) looked more to the *ver*-words (hearing them in their canonical form) than participants in the syllabic reduction group (thick lines) who were exposed to reduced versions. The two groups did not differ much when looking to the /f/-words.

## Appendix C. Detailed analysis of the exposure phase results of Experiment 2

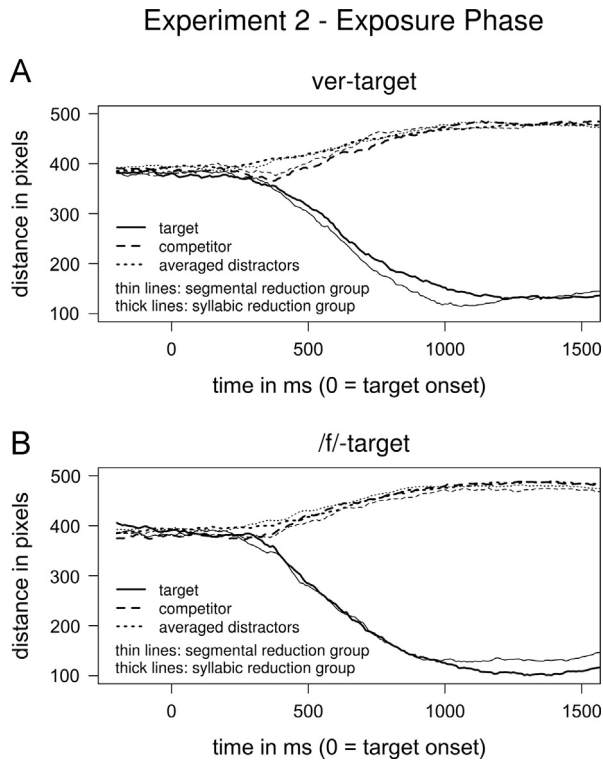
The mean RTs of the click responses for both groups sampled in two conditions are shown in Fig. C1. As can be seen, the segmental reduction group responded numerically faster than the syllabic reduction group for every word type. While the segmental reduction group responded to every word type equally fast, the syllabic reduction group seems to be slowed down by quasi-canonical /u/- and reduced *ver*-words (see higher bars for these words). Statistical analyses revealed a significant main effect of Group in both conditions ( $b_{\text{Group}} = 0.10$ ,  $SE = 0.05$ ,  $t = 1.9$ ,  $p_{\text{MCMC}} < 0.05$  for the segmental reduction condition and  $b_{\text{Group}} = 0.11$ ,  $SE = 0.05$ ,  $t = 2.1$ ,  $p_{\text{MCMC}} < 0.01$  for the syllabic reduction condition). That is, the syllabic reduction group responded overall more slowly than the segmental reduction group (the syllabic reduction group was mapped on 0.5 in the contrast coding). The main effect of Trial Number was significant in both conditions ( $b_{\text{Trial Number}} = -0.002$ ,  $SE = 0.0002$ ,  $t = -10.3$ ,  $p_{\text{MCMC}} < 0.001$  for the segmental reduction condition and  $b_{\text{Trial Number}} = -0.001$ ,  $SE = 0.0002$ ,  $t = -8.6$ ,  $p_{\text{MCMC}} < 0.001$  for the syllabic reduction condition), indicating that participants reacted faster over the course of the exposure phase. In the segmental reduction condition (/b/- and /u/-words), the main effect of Target Word was not significant. That is, both groups did not respond faster to one word than to the other. In contrast, the interaction between Group and Target Word was marginally significant ( $b_{\text{Group} \times \text{Target Word}} = 0.04$ ,  $SE = 0.02$ ,  $t = 1.9$ ,  $p_{\text{MCMC}} = 0.053$ ). This effect was driven mostly by the syllabic reduction group, who took longer to respond to quasi-canonical /u/-words. In the syllabic reduction condition (*ver*- and /f/-words), the main effect of Target Word as well as the interaction between Group and Target Word was significant ( $b_{\text{Target Word}} = -0.04$ ,  $SE = 0.01$ ,  $t = -3.4$ ,  $p_{\text{MCMC}} < 0.001$  and  $b_{\text{Group} \times \text{Target Word}} = -0.06$ ,  $SE = 0.02$ ,  $t = -2.8$ ,  $p_{\text{MCMC}} < 0.01$ ). The main effect of Target Word indicates that both groups responded faster to /f/-words (mapped onto 0.5) than to *ver*-words. The interaction reflects an effect of reduction, driven by the syllabic reduction group who was strongly slowed down by the reduced *ver*-words.

The corresponding eye-tracking data for the exposure phase are displayed in Figs. C2 and C3. Figs. C2A and B show the results for the /b/- and /u/-targets, that is, the segmental reduction condition. It can be seen that for the /b/-words the target lines representing the two groups (thin lines for the segmental reduction group, thick lines for the syllabic reduction group) differ early on (from around 300 ms after target onset, see Fig. C2A), whereas they differ somewhat later (from around 500 ms after target onset, see Fig. C2B) for the /u/-words. Importantly, the position of the lines switched: For /b/-targets, the syllabic reduction group (hearing canonical /b/-words) looked closer to the target, whereas for the /u/-words, the segmental reduction group showed a greater target preference (hearing canonical /u/-words). Statistical analyses showed that the main effect of Target Word was significant in the time window from





**Fig. C2.** Distances between fixations and targets, competitors and averaged distractors in pixels for the /b/- and /u/-targets in the exposure phase of Experiment 2.



**Fig. C3.** Distances between fixations and targets, competitors and averaged distractors in pixels for the ver- and /f/-targets in the exposure phase of Experiment 2.

1000–1500 ms after target onset ( $b_{\text{Target Word}} = -0.18$ ,  $SE = 0.06$ ,  $t = -3.1$ ,  $p_{\text{MCMC}} < 0.01$ ). In this time window, participants in both groups had a preference for the /b/-targets over the /u/-targets (as the /b/-targets are mapped onto  $-0.5$  and the  $b$ -value is negative). The main effect of Group was not significant in either time window. That is, no group had a greater target preference than the other. The interaction between Group and Target Word, however, was significant from 200–900 ms after target onset ( $b_{\text{Group} \times \text{Target Word}} = -0.27$ ,  $SE = 0.07$ ,  $t = -3.8$ ,  $p_{\text{MCMC}} < 0.001$ ). Thus, we found an effect of reduction that went into the expected direction: The

segmental reduction group who was exposed to reduced /b/-words did not look as closely to the /b/-targets as the syllabic reduction group. The opposite was true for the /u/-targets.

The eye-tracking results for the *ver*- and /f/-targets (syllabic reduction condition) are shown in Figs. C3A and B. The target lines differ quite late from around 600–1200 ms after target onset for the *ver*-words (see Fig. C3A) and from around 1000 ms onwards for the /f/-words (see Fig. C3B). However, the differences in mean fixations of the two groups correspond again to expectations: The segmental reduction group (represented by the thin lines and hearing canonical *ver*-words) looked closer to the *ver*-targets than the syllabic reduction group did (hearing reduced *ver*-words). For the /f/-targets, it is the syllabic reduction group (hearing canonical /f/-words) who had a greater target preference than the segmental reduction group (hearing quasi-canonical /f/-words with a long [f:] at the beginning). Statistical analyses did not reveal a main effect of Group in either time window. That is, the groups did not differ in their target preference for *ver*- and /f/-targets respectively. The main effect of Target Word, however, was significant in the time window from 300–1500 ms after target onset ( $b_{\text{Target Word}}=0.10$ ,  $SE=0.04$ ,  $t=2.8$ ,  $p_{\text{MCMC}} < 0.01$ ). This indicates that both groups looked closer to the /f/-targets than to the *ver*-targets (as the  $b$ -value is positive and the /f/-targets are mapped onto 0.5). Also the interaction of Group and Target Word was significant in a time window from 700–1500 ms after target onset ( $b_{\text{Group} \times \text{Target Word}}=0.25$ ,  $SE=0.08$ ,  $t=3.2$ ,  $p_{\text{MCMC}} < 0.01$ ). Thus, also in the syllabic reduction condition, we found an effect of reduction going in the expected direction with closer looks to the *ver*-words if they were heard in unreduced form by the segmental reduction group.

## Appendix D. Supporting information

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.wocn.2014.06.004>.

## References

- Andics, A., McQueen, J. M., & Van Turenhout, M. (2007). Phonetic content influences voice discriminability. In *Proceedings of the 16th international congress of phonetic sciences (ICPhS 2007)* (pp. 1829–1832). Dudweiler: Pirrot.
- Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database [CD-ROM]*. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.
- Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, 59(4), 457–474.
- Boersma, P., & Weenink, D. (2010). *Praat: Doing phonetics by computer (Version 5.1.26)*. Amsterdam: Institute of Phonetic Sciences.
- Brouwer, S. (2010). *Processing strongly reduced forms in casual speech* (Ph.D. thesis). Nijmegen: Radboud University.
- Brouwer, S., Mitterer, H., & Huettig, F. (2012). Can hearing puter activate pupil? Phonological competition and the processing of reduced spoken words in spontaneous conversations. *Quarterly Journal of Experimental Psychology*, 65(11), 2193–2220.
- Brouwer, S., Mitterer, H., & Huettig, F. (2013). Discourse context and the recognition of reduced and canonical spoken words. *Applied Psycholinguistics*, 34, 519–539.
- Bybee, J. (2001). *Phonology and language use*. Cambridge, England: Cambridge University Press.
- Cho, T., McQueen, J. M., & Cox, E. A. (2007). Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics*, 35(2), 210–243.
- Connine, C. M., Ranbom, L. J., & Patterson, D. J. (2008). Processing variant forms in spoken word recognition: The role of variant frequency. *Perception & Psychophysics*, 70(3), 403–411.
- Cutler, A., Eisner, F., McQueen, J. M., & Norris, D. (2010). How abstract phonemic categories are necessary for coping with speaker-related variation. In: C. Fougerson, B. Kühnert, M. D'Imperio, & N. Vallée (Eds.), *Laboratory phonology*, Vol. 10 (pp. 91–111). Berlin: de Gruyter.
- Dahan, D., Drucker, S. J., & Scarborough, R. A. (2008). Talker adaptation in speech perception: Adjusting the signal or the representations? *Cognition*, 108(3), 710–718.
- Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics*, 67(2), 224–238.
- Eisner, F., & McQueen, J. M. (2006). Perceptual learning in speech: Stability over time (L). *Journal of the Acoustical Society of America*, 119(4), 1950–1953.
- Eisner, F., Melinger, A., & Weber, A. (2013). Constraints on the transfer of perceptual learning in accented speech. *Frontiers in Psychology*, 4, 148.
- Ernestus, M. (2000). *Voice assimilation and segment reduction in casual Dutch. A corpus-based study of the phonology-phonetics interface* (Ph.D. thesis). Utrecht: LOT.
- Ernestus, M. (2009). The roles of reconstruction and lexical storage in the comprehension of regular pronunciation variants. In *Proceedings of the 10th annual conference of the international speech communication association (Interspeech 2009)* (pp. 1875–1878). Causal Productions Pty Ltd.
- Ernestus, M., Baayen, H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language*, 81(1–3), 162–173.
- 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.
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105(2), 251–279.
- Gow, D. W. (2002). Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance*, 28(1), 163–179.
- Gow, D. W. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception & Psychophysics*, 65(4), 575–590.
- Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, 93(4), 411–428.
- Johnson, K. (2004). Massive reduction in conversational American English. In: K. Yoneyama, & K. Maekawa (Eds.), *Spontaneous speech: Data and analysis. Proceedings of the 1st session of the 10th international symposium* (pp. 29–54). Tokyo, Japan: The National International Institute for Japanese Language.
- Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In: J. Bybee, & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 229–254). Amsterdam: John Benjamins.
- Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, 42(3), 643–650.
- Keune, K., Ernestus, M., van Hout, R., & Baayen, H. R. (2005). Social, geographical, and register variation in Dutch: From written MOGELIJK to spoken MOK. *Corpus Linguistics and Linguistic Theory*, 1, 183–223.
- Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, 51(2), 141–178.
- Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, 56(1), 1–15.
- Kraljic, T., Samuel, A. G., & Brennan, S. E. (2008). First impressions and last resorts: How listeners adjust to speaker variability. *Psychological Science*, 19, 332–338.
- Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America*, 29(1), 98–104.
- Lahiri, A., & Reetz, H. (2002). Underspecified recognition. In: C. Gussenhoven, & N. Warner (Eds.), *Laboratory phonology*, Vol. 10 (pp. 637–675). Berlin: Mouton de Gruyter.
- Lahiri, A., & Reetz, H. (2010). Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics*, 38(1), 44–59.
- Manuel, S. Y. (1992). Recovery of "deleted" schwa. In: O. Engstrand, & C. Kylander (Eds.), *Perilus: Papers from the Symposium on current phonetic research paradigms for speech motor control*, Vol. XIV (pp. 115–118). Stockholm: University of Stockholm.
- Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, 25(1–2), 71–102.
- Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science*, 32(3), 543–562.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18(1), 1–86.
- McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, 30(6), 1113–1126.
- Mitterer, H., Chen, Y., & Zhou, X. (2011). Phonological abstraction in processing lexical-tone variation: Evidence from a learning paradigm. *Cognitive Science*, 35(1), 184–197.
- Mitterer, H., Csépe, V., & Blomert, L. (2006). The role of perceptual integration in the recognition of assimilated word forms. *Quarterly Journal of Experimental Psychology*, 59(8), 1395–1424.

- 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(3), 451–479.
- Mitterer, H., & Ernestus, M. (2006). Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch. *Journal of Phonetics*, 34(1), 73–103.
- Mitterer, H., & McQueen, J. M. (2009a). Foreign subtitles help but native-language subtitles harm foreign speech perception. *Plos One*, 4(11), A146–A150.
- Mitterer, H., & McQueen, J. M. (2009b). Processing reduced word-forms in speech perception using probabilistic knowledge about speech production. *Journal of Experimental Psychology: Human Perception and Performance*, 35(1), 244–263.
- Mitterer, H., & Reinisch, E. (2013). No delays in application of perceptual learning in speech recognition: Evidence from eye tracking. *Journal of Memory and Language*, 69, 527–545.
- Mitterer, H., & Russell, K. (2013). How phonological reductions sometimes help the listener. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(3), 977–984.
- Mitterer, H., Scharenborg, O., & McQueen, J. M. (2013). Phonological abstraction without phonemes in speech perception. *Cognition*, 129, 356–361.
- Newman, R. S., & Sawusch, J. R. (2009). Perceptual normalization for speaking rate III: Effects of the rate of one voice on perception of another. *Journal of Phonetics*, 37(1), 46–65.
- Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, 115(2), 357–395.
- Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47(2), 204–238.
- Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, 60(3), 355–376.
- Oostdijk, N. (2000). The spoken dutch corpus project. *ELRA Newsletter*, 5, 4–8.
- Patterson, D., LoCasto, P. C., & Connine, C. M. (2003). Corpora analyses of frequency of schwa deletion in conversational American English. *Phonetica*, 60(1), 45–69.
- Pitt, M. A. (2009). The strength and time course of lexical activation of pronunciation variants. *Journal of Experimental Psychology: Human Perception and Performance*, 35(3), 896–910.
- Pitt, M. A., Dilley, L., & Tat, M. (2011). Exploring the role of exposure frequency in recognizing pronunciation variants. *Journal of Phonetics*, 39(3), 304–311.
- Pitt, M. A., & McQueen, J. M. (1998). Is compensation for coarticulation mediated by the lexicon?. *Journal of Memory and Language*, 39(3), 347–370.
- Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005). Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America*, 118(4), 2561–2569.
- Poellmann, K., Mitterer, H., & McQueen, J. M. (2014). Use what you can: Storage, abstraction processes and perceptual adjustments help listeners recognize reduced forms. *Frontiers in Psychology*, 5, 437.
- Ranbom, L. J., & Connine, C. M. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*, 57(2), 273–298.
- Reinisch, E., Jesse, A., & McQueen, J. M. (2011). Speaking rate from proximal and distal contexts is used during word segmentation. *Journal of Experimental Psychology: Human Perception and Performance*, 37(3), 978–996.
- Reinisch, E., & Sjerps, M. J. (2013). The uptake of spectral and temporal cues in vowel perception is rapidly influenced by context. *Journal of Phonetics*, 41(2), 101–116.
- Reinisch, E., Wozny, D. R., Mitterer, H., & Holt, L. L. (2014). Phonetic category recalibration: What are the categories?. *Journal of Phonetics*, 45, 91–105.
- Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Attention Perception & Psychophysics*, 71(6), 1207–1218.
- Shockey, L. (2003). *Sound patterns of spoken English*. Cambridge: Blackwell.
- Sjerps, M. J., Mitterer, H., & McQueen, J. M. (2011). Listening to different speakers: On the time-course of perceptual compensation for vocal-tract characteristics. *Neuropsychologia*, 49(14), 3831–3846.
- Taft, M., Hambly, G., & Kinoshita, S. (1986). Visual and auditory recognition of prefixed words. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, 38(3), 351–365.
- Toscano, J. C., & McMurray, B. (2012). Cue-integration and context effects in speech: Evidence against speaking-rate normalization. *Attention Perception & Psychophysics*, 74(6), 1284–1301.
- Van Son, R. J. J. H., & Pols, L. C. W. (1990). Formant frequencies of Dutch vowels in a text, read at normal and fast rate. *Journal of the Acoustical Society of America*, 88(4), 1683–1693.
- Van Son, R. J. J. H., & Pols, L. C. W. (1992). Formant movements of Dutch vowels in a text, read at normal and fast rate. *Journal of the Acoustical Society of America*, 92(1), 121–127.
- Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, 40(3), 374–408.
- Weber, A., Di Betta, A., & McQueen, J. M. (2014). Treack or trit: Adaptation to genuine and arbitrary foreign accents by monolingual and bilingual listeners. *Journal of Phonetics* (in press).
- Warner, N., & Tucker, B. V. (2011). Phonetic variability of stops and flaps in spontaneous and careful speech. *Journal of the Acoustical Society of America*, 130(3), 1606–1617.