

Processing Reduced Word-Forms in Speech Perception Using Probabilistic Knowledge About Speech Production

Holger Mitterer and James M. McQueen
Max Planck Institute for Psycholinguistics

Two experiments examined how Dutch listeners deal with the effects of connected-speech processes, specifically those arising from word-final /t/ reduction (e.g., whether Dutch [tas] is *tas*, *bag*, or a reduced-/t/ version of *tast*, *touch*). Eye movements of Dutch participants were tracked as they looked at arrays containing 4 printed words, each associated with a geometrical shape. Minimal pairs (e.g., *tas/tast*) were either both above (*boven*) or both next to (*naast*) different shapes. Spoken instructions (e.g., “*Klik op het woordje tas boven de ster*,” [Click on the word *bag* above the star]) thus became unambiguous only on their final words. Prior to disambiguation, listeners’ fixations were drawn to /t/-final words more when *boven* than when *naast* followed the ambiguous sequences. This behavior reflects Dutch speech-production data: /t/ is reduced more before /b/ than before /n/. We thus argue that probabilistic knowledge about the effect of following context in speech production is used prelexically in perception to help resolve lexical ambiguities caused by continuous-speech processes.

Keywords: spoken-word recognition, continuous-speech processes, /t/-reduction, eye tracking, Dutch

Continuous speech is peppered with noncanonical pronunciations of words. Continuous-speech processes, such as assimilation, reduction, and deletion, operate during speech production and can substantially change the way words are realized. For example, a speaker of British English may say [waɪmbɔʔl] in a casual request for a wine bottle, but say [waɪn] and [bɔtəl] when asked how these words should be pronounced. It is the listener’s task to recognize spoken words in spite of this variability. A particularly acute problem is when a continuous-speech process creates a lexical ambiguity. We examine that situation here, focusing on the effects of word-final /t/ reduction. If a speaker does not clearly pronounce the final [t] of *duct*, for example, how does the listener work out that they are not talking about a *duck*?

Two answers to this question have been proposed. The first is based on lexical storage. There are in fact three radically different versions of this view. Lahiri and Marslen-Wilson (1991) and Lahiri and Reetz (2002) proposed a model of the mental lexicon in which entries are highly abstract and phonologically underspecified. A lexical entry for a given word only includes the phonological features that are reliably associated with tokens of that word. For instance, the alveolar nasal consonants in English are assumed to be specified only as nasal, with no specification of place of articulation, which may vary due to phonological assimilation. The token [waɪn] can be recognized as an instance of the word *wine* because there is no mismatch between the last nasal segment [m] in the input and the lexical representation for this segment. Sim-

ilarly, lexical entries may be underspecified for the presence of a word-final /t/. The input [dʌk] would thus match a stored lexical representation corresponding to *duct*. The second lexical-storage account is offered by episodic models (e.g., Bybee, 2001; Goldinger, 1998; Hawkins, 2003). According to the episodic view, the entry for a given word in the mental lexicon consists of nothing more—and nothing less—than the detailed and concrete episodic memories of pronunciations of that word that have been encountered previously. Episodic traces of noncanonical forms such as [waɪn] and [dʌk] would thus exist side by side with traces of canonical forms of those words. The third lexical-storage account is intermediate in that it assumes that different pronunciation variants are stored in the lexicon, but are stored as fully specified abstract phonological forms (Connine, 2004; McLennan, Luce, & Charles-Luce, 2003; Ranbom & Connine, 2007). Instead of a single, underspecified representation for *duct*, for example, both /dʌk/ and /dʌkt/ would be stored, but as unitary abstract representations that do not include information about voice quality, speech rate, and so on, as would be the case for episodic traces of each variant.

In spite of these fundamental differences in assumptions about the nature of lexical representations, all three lexical-storage accounts assume that disambiguation of forms created by processes such as /t/ reduction occurs after lexical access. The input [dʌk], for example, would be mapped onto lexical entries corresponding to both *duct* and *duck*, and the task of deciding which word it was would then be no different from that required for the resolution of other types of lexical ambiguity (e.g., selection of the meaning of a polysemous word such as *bank* using contextual information).

The alternative view of how lexical ambiguities created by continuous-speech processes are handled is that processes act prior to lexical access to undo the effects of the speech-production process. The core idea here is that listeners can compensate for coarticulatory influences on the acoustic signal using segmental

We thank Marloes van der Goot for helping to prepare and for running these experiments. We also thank Gareth Gaskell for comments on an earlier version of the manuscript.

Correspondence concerning this article should be addressed to Holger Mitterer or James M. McQueen, Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands. E-mail: holger.mitterer@mpi.nl or james.mcqueen@mpi.nl

context (see, e.g., Fowler, 2005; Fowler, Best, & McRoberts, 1990; Fowler & Brown, 2000; Kingston & Macmillan, 1995; Lindblom & Studdert-Kennedy, 1967; Mann, 1980; Mann & Repp, 1980, 1981; Mann & Soli, 1991; Mitterer, 2006b; Smits, 2001a). Once again, there are multiple theoretical variants of this view. According to Motor Theory (Lieberman, 1996), listeners have innate knowledge of how segments interact in fluent production, and therefore can recover intended speech gestures from the distorted acoustic input. According to Direct Realism (Fowler, 1996), speech gestures are again the object of perception, but on this view speech gestures can be perceived directly and without the need for a priori knowledge. According to this latter account, the term /t/-deletion or even /t/-reduction is not appropriate (Goldstein & Fowler, 2003), because the /t/ gesture is assumed not to be reduced or deleted. Instead, the acoustic traces of the /t/ gestures become less salient because of gestural overlap. (Similarly, place assimilation may also be the wrong term, because a word-final /n/ never “loses” its place of articulation, but instead is coproduced with the gesture of the following consonant.) There is some evidence for residual gestures if segments are apparently deleted (Manuel, 1992) or assimilated (Nolan, 1992), but it remains an empirical issue whether this is the case for all apparent deletions and assimilations (for contrary evidence, see Gow & Im, 2004). Such considerations, however, do not change the fact that the listener is still faced with an enormous amount of variability caused by continuous-speech processes. By using the terms *reduction* and *assimilation*, we do not want to imply that these are phonological, rule-governed, all-or-none processes (cf., Gow, 2002). We use the terms simply to denote the kinds of production variance the listener has to compensate for.

Fundamentally different alternative accounts to the gestural ones are that general auditory mechanisms (Holt & Lotto, 2002; Holt, Lotto, & Kluender, 2000; Lotto & Kluender, 1998; Lotto, Kluender, & Holt, 1997) or general pattern-classification principles (Nearey, 1997; Smits, 2001b) can account for context sensitivity in speech perception. As an example of the latter kind of mechanism, Gaskell (2003) presented a model that uses a context-sensitive compensation mechanism to adjust specifically for the effects of place assimilation in English. The model adjusts the recognition of the segment x on the basis of the segment $x + 1$: If [m] is followed by [b], the input is adjusted prelexically so that the [m]-segment is recoded as being ambiguous between [m] and [n]. After this correction, the input [waim] can be matched onto the lexical representation of *wine*, specified as having a word-final /n/, without a prohibitive amount of mismatch between the input and the stored form. The model is based on the assumption that listeners are attuned to abstracted probabilistic knowledge about the phonological patterns they hear and use that knowledge in perception. Gow (2002, 2003) has proposed yet another account of prelexical compensation for place assimilation, which assumes that language-specific phonetic features are grouped together by a general, Gestalt-like grouping process.

A broad range of similar prelexical compensation mechanisms could be postulated for recovery from the effects of /t/ reduction. According to any of these accounts, phonologically altered forms can be recognized without explicit lexical storage of those alternative pronunciations. Although there are thus very different theoretical views of the nature of the compensation process, the idea in all of these accounts is that noncanonical input (e.g., [dʌk] for

duct) is transformed prior to lexical access in such a way that it then matches the canonical lexical representation ([dʌkt] stored in the mental lexicon. The lexical ambiguity (is it *duck* or *duct*?) still needs to be resolved, of course, but the primary perceptual problem (the apparent mismatch between [dʌk] and [dʌkt]) has been dealt with. Furthermore, in cases where /t/ reduction does not create a lexical ambiguity, the prelexical compensation process is all that is required to undo the effects of the reduction. That is, for /t/-final words without a competitor (such as *frost*, where *fros* is not a word), recognition is straightforward once the apparent mismatch between the /t/-reduced input and the stored form has been dealt with.

Here we investigated whether lexical storage or prelexical compensation better explains how Dutch listeners recover from the effects of word-final /t/ reduction. In particular, we tested whether Dutch listeners used the phonological context following a potential [t] deletion and probabilistic knowledge based on that context to help in lexical disambiguation. We focus on the effects of following context because of their potential to distinguish between the lexical storage and prelexical compensation accounts.

Following context effects have been influential in the theoretical debate surrounding place assimilation. As we have already argued, lexical storage accounts can apparently explain how listeners can recognize tokens such as [waim] as an instance of *wine*, in spite of the change in the place of articulation of the final consonant. But these accounts ignore a critical feature of place assimilation, namely that it is contextually conditioned. The form [waim] can occur before labial consonants, such as in *wine bottle*, but never occurs before a velar consonant, as in *wine glass*. Lexical-storage models predict, however, that [waim] should be recognized as *wine* independently of the segmental context (once the alternate pronunciation is stored, in whatever way, it should be available for recognition). This prediction has not been borne out: A large number of experiments have shown that a word pronounced with an assimilatory change is only recognized in a segmental context that licenses that assimilation (Darcy, Peperkamp, & Dupoux, 2007; Gaskell & Marslen-Wilson, 1996, 1998, 2001; Gow, 2002, 2003; Mitterer & Blomert, 2003; Mitterer, Csépe, & Blomert, 2006; Mitterer, Csépe, Honbolygo, & Blomert, 2006). This behavior is thus better explained in terms of prelexical compensation for pronunciation variation, which of course is based on the use of context.

/t/ reduction offers a critical test of the lexical storage and prelexical compensation accounts. Mitterer and Ernestus (2006) recently investigated pronunciation variation of word-final /t/ in Dutch. They found, first, that /t/ reduction is not a categorical, all-or-none process. That is, traces of /t/ are often preserved in the acoustic signal. /t/ reduction is thus a more appropriate description of this phenomenon than /t/ deletion. For instance, if reduction of /t/ occurs before /b/ so that there is no /t/ release (note that Dutch /t/ is usually released, Cho & McQueen, 2006), an elongated, partly voiceless closure can be observed, which is a residual cue to the presence of an underlying voiceless stop /t/ (see Figure 1D in Mitterer & Ernestus). Moreover, if /t/ is reduced in a word-final consonant cluster (e.g., in *kast*, [cupboard]), the remaining penultimate consonant /s/ is shorter than in words ending underlyingly with this consonant (such as *kas*, ‘greenhouse’). In this regard, /t/ reduction is similar to place assimilation, which has also been shown to be phonetically incomplete (Gow, 2002, 2003). In perception

experiments, Mitterer and Ernestus found that listeners, just as in the case of assimilation, are attuned to these residual cues, so that the presence of an elongated closure or a shorter penultimate consonant led listeners to infer the presence of /t/.

Mitterer and Ernestus (2006) also found that /t/ reduction is similar to place assimilation in another regard: Its occurrence in production is contextually conditioned. Specifically, /t/ reduction occurs particularly often after /s/ and before labial consonants. The latter finding is particularly important for two reasons: it concerns a cross-word dependency, and one that is probabilistic. In the casual-speech subcorpus of the Corpus of Spoken Dutch (<http://lands.let.kun.nl/cgn/ehome.htm>), for example, word-final /t/ was reduced in 86% of following labial contexts (after factoring out effects of preceding context), but in 33% of following alveolar contexts (see Table 3 in Mitterer & Ernestus, 2006). Cross-word dependencies of this type are difficult for lexical-storage accounts to explain (Mitterer, 2006a). A prelexical mechanism, however, could make use of these tendencies so that, for example, utterances of [dʌk] followed by an alveolar would lead to weaker (probably spurious) activation of *duct* than utterances of [dʌk] followed by a labial, where *duct* is more likely to be the word the speaker intended.

In their perceptual experiments, Mitterer and Ernestus (2006) varied the previous and following context of potentially reduced word-final /t/s. They failed to find a following-context effect. In these experiments, Dutch listeners heard target (non)words in a sentence frame and decided whether the targets ended in a /t/ or not (*duct* or *duck*, to use our English example) in a two-alternative forced choice (2AFC) task. A /t/ was heard in targets followed by a word starting with /b/ as often as in targets followed by a word starting with a /k/, although /t/ reduction is more likely before /b/ than before /k/. Nevertheless, listeners took the previous context into account when judging whether targets ended on /t/. The same residual cues led to a percept of /t/ more often if it occurred after /s/ than after /n/, mirroring the production finding that /t/ is more likely to be reduced after /s/ than after /n/. This pattern of data fits well with a lexical-storage account, because the lexical representations of words with word-final /t/ contain the penultimate segment, and, accordingly, the influence of preceding segmental context on the reduction of a word-final /t/ can be coded lexically. The context following the final /t/ of a word is obviously not a part of that word's lexical representation, and, in keeping with the lexical-storage account, cross-word dependencies apparently cannot influence the perception of reduced /t/.

If this analysis is correct, it would suggest that different types of continuous-speech phenomena are dealt with in perception in different ways: through lexical storage for /t/ reduction, but through prelexical compensation for place assimilation. This conclusion, however, is based on the absence of an effect of following context in Mitterer and Ernestus (2006). The experimental technique they used was a simple 2AFC task. Listeners indicated whether a word or nonword contained a word-final /t/ or not. This task was adequate, because the experiments revealed a clear effect of preceding segmental context. Moreover, a lexical effect like the Ganong (1980) effect was observed: A /t/ was more often perceived when a final /t/ was consistent with a word than when it was consistent with a nonword. Nevertheless, this forced-choice task necessarily focuses listeners' attention on the acoustic-phonetic properties of the targets and not on the interpretation of the targets

in their sentence contexts. This may have masked an effect of following context.

In Experiment 1, therefore, we tested the role of following context in recovery from /t/ reduction using a task that encouraged listeners to focus on utterance interpretation and hence away from the detailed acoustic-phonetic properties of the stimuli. We used a new variant of the visual-world paradigm (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). In this paradigm, eye movements to displays of visual objects are measured during the presentation of spoken utterances (often instructions to manipulate the objects in the display). It is possible to track listeners' interpretations of those utterances as they unfold in time, including, critically, sentences that are temporarily phonologically ambiguous (Allopenna, Magnuson, & Tanenhaus, 1998; Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Salverda, Dahan, & McQueen, 2003; Shatzman & McQueen, 2006). For example, as Dutch listeners heard *panda* [panda], they looked at pictures of a panda and of a pan (Salverda et al., 2003), suggesting that they were, at least temporarily, considering both of these words as hypotheses for what it was they were hearing. It should therefore be possible to track eye-movements to pictures of, for example, *kas*, 'greenhouse' and *kast*, 'cupboard', as Dutch listeners hear instructions such as "*Klik op de kas boven/naast de ster*" [Click on the greenhouse above/next to the star]. /t/ reduction should occur more often when the following context is *boven* (beginning with the labial /b/) than when the following context is *naast* (beginning with the alveolar /n/). If Dutch listeners are sensitive to this tendency, they should look more at the *kast* in the former case. That is, they should temporarily have a greater bias to interpret [kas] as a /t/-reduced token of *kast* in the *boven* context. Such a finding would indicate that Dutch listeners use prelexical compensation rather than lexical storage to recover from /t/ reduction.

A problem with this idea, however, is that many of the minimal pairs of /t/-reduced and unreduced words in Dutch, such as *tast/tas* [touch/bag], consist of at least one word that is not easily picturable (e.g., *tast* [sense of touch]). Even if items are easily picturable, they may not be easily identifiable for the average participant (e.g., such as *gist* [yeast] or *dors* [young cod]). Furthermore, not all pairs of picturable nouns are matched in gender, making it hard to match instructions (e.g., *de mest/het mes*, [the fertilizer/the knife]). Not enough items could be found that matched these constraints. Recently, however, a printed-word variant of the eye-tracking paradigm has been developed (Huettig & McQueen, 2007; McQueen & Viebahn, 2007). Pictures in the display are replaced with printed words. Phonological effects using this form of the paradigm closely match results from similar designs using pictures (e.g., McQueen & Viebahn observed similar patterns of eye movements to those found by Allopenna et al., 1998). If the display can consist of printed words, then restrictions on target picturability evaporate. We therefore used the printed-word variant of the paradigm here.

It was necessary, however, to extend the visual-world paradigm further. We capitalized on the fact that one can present instructions in this paradigm that are unambiguous globally, but ambiguous locally. Participants thus had no difficulty following instructions that were nevertheless temporarily ambiguous. Consider the example display in Figure 1, presented with the instruction *Klik op het woordje tas boven de ster* [Click on the word *tas* above the star]. We will identify the word to be clicked on as the *target*, a word with a similar or identical form as the *competitor*, and the

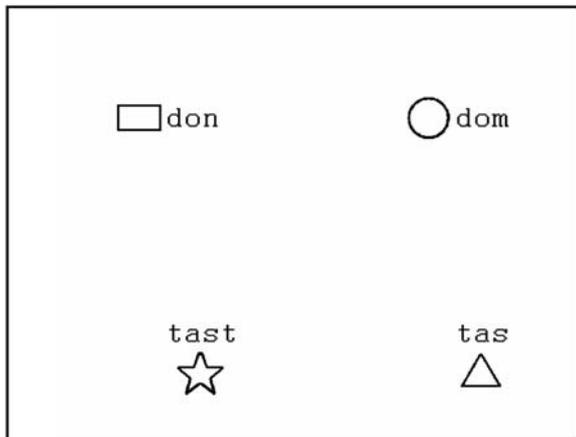


Figure 1. A stimulus array accompanying the instruction “click on the word *tas* above the star.”

other words as *distractors*, so that, in Figure 1, the target is *tast*, the competitor is *tas*, and the distractors are *don* and *dom*. The instruction ‘Click on the word *tas* above . . .’ is ambiguous, because *tas* may refer to *tas* or *tast* (if the word-final /t/ has been reduced) and because both *tas* and *tast* are above a geometrical shape. Despite this temporary ambiguity, the complete prepositional phrase “above the star” uniquely specifies the word the participants had to click on. Participants therefore did not have to focus their attention on the acoustic–phonetic properties of the instructions. We predicted that these conditions would increase the likelihood of finding an effect of following context on lexical interpretation.

In the critical *stop consonant* condition (henceforth, the *stop* condition) of Experiment 1, therefore, we tracked listeners eye-movements to displays such as that in Figure 1. We manipulated following context by changing the instructions (and the displays). Specifically, we also looked at the situation where both target and competitor were not above, but next to their respective geometrical shapes. The instruction, for example, was “. . . *tas* naast de ster” [. . . *tas* next to the star’]. If following context influences perceptual recovery from /t/ reduction, the competitor *tast* should be more eligible as a target word when followed by the preposition *boven* than when followed by *naast*, and listeners should thus look more at *tast* in the former case. This was the main hypothesis tested in Experiment 1.

In addition, we attempted to replicate the finding of Mitterer and Ernestus (2006) that listeners are sensitive to subphonemic cues that are likely to be associated with a reduced /t/. Therefore, we presented utterances of words such as *tast* with more or less acoustic evidence for the presence of a word-final /t/. There was never a /t/-release, but the preceding consonant was either short or long followed by a long or short silence, respectively. The first form (short consonant and long silence) should lead listeners to infer the presence of an underlying /t/ and hence look to the orthographic target *tast* more than the second form (long consonant and short silence) should.

Our use of combinations of printed words and geometrical shapes allowed us to test the effect of following context on the perception of reduced /t/ in a controlled way, with instructions that

were appropriate to the situation the participants faced. This version of the visual-world paradigm had not been used before, however, so we also included conditions that evaluated the suitability of this extension of the paradigm. These conditions consisted of *semantic* trials. Target words were accompanied by identical competitors (e.g., the word *tast* appeared on the screen twice), and the disambiguating information was supplied either by the positional preposition (*position* trials; e.g., one of the words was above and the other next to a star) or the shape itself (*shape* trials; e.g., one *tast* was above a star and the other above a circle). If participants use information as soon as it is available, they should look faster towards the target on position trials than on shape trials, simply because of the word order in the instructions. Such a pattern would show that the extended visual-world paradigm is able to reflect fast and immediate use of information as speech unfolds.

Semantic aspects of language are more salient than phonological aspects (cf. Morais, Cary, Alegria, & Bertelson, 1979). The semantic trials thus had the added benefit of distracting participants from the phonological manipulation in the stop trials, thus discouraging participants from adopting a strategy of predicting the targets on the basis of the targets’ phonological form. There are other strategies, however, that participants might use to predict the targets. For example, given that we had to present minimal pairs of targets and competitors in the displays, participants could have learned that the targets were always one of the two similar words on the display. The distractors in the stop trials therefore always consisted of a word-final /n/–/m/ minimal pair (see Figure 1). The presence of a phonological neighbor in the display thus did not give away a word as a potential target. Nevertheless, if all trials had been of this type, participants could still have learned that only members of the /Ct#/-/C#/# pairs (with C standing for consonant and # for the word boundary) were potential targets. We therefore included *nasal* trials, in which a member of the /n/–/m/ minimal pair was the target and the members of the /Ct#/-/C#/# pair acted as distractors.

The nasal trials also allowed us to investigate an established context effect: compensation for nasal place assimilation. Dutch words ending in /n/ may be pronounced with an assimilated [m] if followed by /b/, and, in perception, [m] is likely to be perceived as /n/ if followed by /b/ (Gaskell & Marslen-Wilson, 1996, 1998, 2001; Mitterer & Blomert, 2003). In order to replicate this effect in the current design, we generated ambiguous tokens of Dutch minimal pair words ending in either /n/ or /m/, such as *boon/boom* (‘bean/tree’). If these tokens are followed by *boven*, they should tend to be interpreted as the /n/-word, because of the context effect exerted by the labial /b/. So we expected more looks to the /n/-words if the preposition was *boven* than when it was *naast*. This would replicate a picture-based eye-movement study by Gow and McMurray (2007), in which listeners looked more at a picture of a cat when hearing “cap box” than when hearing “cap drawing.”

In summary, the semantic and nasal trials in Experiment 1 provided control conditions for the stop trials, plus additional tests. The semantic trials measured the rapidity of uptake of disambiguating information when participants looked at displays consisting of combinations of printed words and geometrical shapes. The nasal trials examined how listeners recover from the effects of place assimilation. The critical stop trials examined perceptual recovery from /t/ reduction by manipulating the acoustic form of

the target words, and, more importantly, the context following the targets. We predicted that if Dutch listeners are sensitive to the probabilistic tendencies of Dutch speakers, they would be more likely to interpret [tas] as a token of *tast* before *boven* than before *naast*. This use of following context would indicate that, as in the more widely studied case of place assimilation, recovery from /t/ reduction is based on a prelexical compensation mechanism rather than on lexical storage.

Experiment 1

Method

Participants

Forty members of the Max Planck Institute subject panel participated in the experiment. All participants had learned Dutch as their native language and none reported hearing problems.

Materials and Stimulus Construction

Visual stimuli were presented on a computer screen, positioned approximately 60 cm in front of the participants. Each display consisted of four printed words and four geometrical shapes. Figure 1 shows an example. The printed words were generated in Courier font (48-point font size); and the geometrical shapes (rectangle, star, triangle, and circle) were fitted in a 70×70 pixel square with at least 5 pixels margin on a 15-in. screen with an 800×600 resolution. The center of the printed words coincided with the center of the one of the quadrants of the screen, independent of word length. Geometrical shapes appearing next to a word were always positioned on the right of the words, so that shapes both below and next to words stood closer to the center of the screen than did the printed words as often as the shapes stood further from the center of the screen than did the printed words. The positions of the shapes were adjusted to the length of the words, so that the distance between shapes next to a word and the word itself did not vary with word length.

In the lexical database CELEX (Baayen, Piepenbrock, & Gulikers, 1995), there are 52 minimal pairs of the type Ct#–C# (such as *rijst* ‘rice’ and *reis* ‘journey’), and 32 minimal pairs of the type Vn#–Vm# (such as *boon* ‘bean’ and *boom* ‘tree’; with V standing for vowel). From this sample, 32 Ct#–C# pairs and 16 Vn#–Vm#

pairs were selected by excluding pairs with an unusual orthography–phonology relationship (e.g., the English loan *team* /tim/, which should be pronounced as /te.am/ if Dutch orthography-to-phonology rules would apply), excluding Ct#–C# pairs in which word-final /t/ arose by word-final devoicing (e.g., *wild* [wilt] ‘wild’, with the inflected form *wilde* [wildə]), and, finally, balancing the frequency differences between the pairs (Appendix A lists the items and their lexical frequencies). These stimuli were used as visual targets, competitors, and distractors, in combination with line-drawings of a star, a rectangle, a circle, and a triangle in the stop, nasal, and semantic trials, as described in the *Procedure and Design* section.

Each of the 48 pairs of words was recorded twice by a female native speaker of Dutch in a sentence frame as indicated in Table 1. The instructions used in the experiment were derived from Table 1 and consisted of cross-spliced materials. If cross-splicing occurred in voiced segments, the splice points were at major positive going zero-crossings. Each experimental sentence consisted of five cross-spliced parts, as indicated in Figure 2. The first part was the same for all sentences and led up to the closure for the /d/ in [wordjə]. The second part was a token of the syllable [djə]. This had to be a different token for each initial consonant of the following target word to let the sentences sound natural; additionally, different tokens of the [djə] syllable had to be used for every /h/-initial word, because the strong vowel coloring of the [h] made transitions sound unnatural (e.g., if [djə] spliced from a following [hyls] was combined with the onset of the word [hal], the result sounded unnatural). The third cross-spliced part was the target word. The fourth part contained positional prepositions *boven* and *naast* as well as the following definite article *de*, but excluding the formant transitions from the schwa to the following consonant. This transition was part of the final part, including one of the four shape names.

The target words were spliced from the onset of the first consonant to the offset of the last consonant, but the offset portions differed across trial types. For targets ending in a nasal that appeared in semantic trials, unedited tokens of the spoken target words were used. For the nasal trials, however, ambiguous tokens of the final consonants were made. The final nasal consonants in each pair of targets (e.g., *boon/boom*) were excised and cut to a duration of about 70 ms. The glottal cycles were then marked for

Table 1
Sentence Frame for the Instructions

Instruction	Preceding context	Target	Following context	Geometrical shape
Dutch	klik op het woordje	<i>Target</i>	boven de naast de	ster rechthoek cirkel driehoek
Phonetic form	klik ɔp ət wordjə	C(C)V(V)(C)Ct C(C)V(V)(C)C C(C)V(V) n C(C)V(V) m	bovə də nast də	stɛr REXthuk sɪrkəl drihuk
Gloss	click on the little word	<i>Target</i>	above the next to the	star rectangle circle triangle

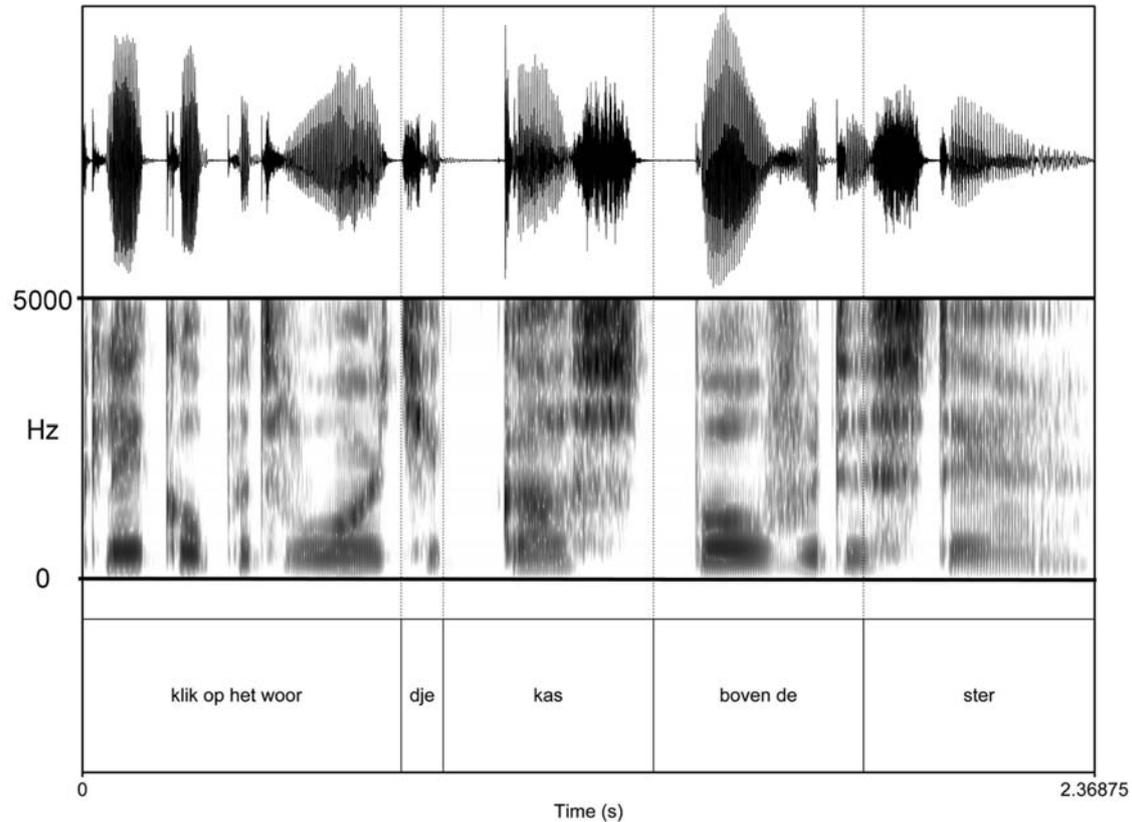


Figure 2. An experimental sentence with splicing points.

both consonants, and a continuum of new nasals was generated by mixing together, cycle by cycle, the waveforms of each pair of consonants (the shorter of any pair of cycles was zero-padded) and then concatenating the cycles. The different waveforms were mixed in proportions ranging from 0.1/0.9 to 0.9/0.1 in steps of 0.1, giving rise to nine sounds per continuum. Each of these nine consonants was then spliced onto the natural initial portion of the /n/-final word (e.g., the *boo-* from *boon*). Four native speakers of Dutch then indicated which member of each nine-step continuum was the most ambiguous token. These judgments never differed by more than two steps. The tokens nearest to the mean of the four judgments for each continuum were used in the main experiment.

The target words for the stop trials were generated from tokens in which the speaker intended to say the word with final /t/. From these tokens, two versions were created, one with more and one with less evidence for an underlying /t/. For 19 of the 32 words, a natural utterance with a reduced /t/ (i.e., without a /t/-release but with a visible closure) was used (see Appendix A; cf. Mitterer & Ernestus, 2006). For the other 13 tokens (those the speaker did not spontaneously reduce; see Appendix A), the /t/ release was replaced by a 50-ms closure signal. These 32 tokens served as the +/t/ bias stimuli: the stimuli with good evidence for /t/. Tokens with less evidence for /t/ (the -/t/ bias stimuli) were generated by shortening the closure by 25 ms and increasing the duration of the previous consonant by 25 ms with PSOLA (using Praat 4.4, Boersma & Weenink, 2006).

Both following contexts—*boven* and *naast*—required two tokens. One token of each word was used following targets ending in nasals: The version of *boven* had a prevoiced /b/ and the version of *naast* had an /n/ that immediately had its maximal amplitude. The other tokens of each word were used following the closure of the /t/-words: The version of *boven* contained a /b/ with an unvoiced closure (100 ms), and the token of *naast* had an /n/ that was unvoiced for 20 ms. The tokens and their durations were prototypical for this speaker.

The splice end-point for the tokens of *boven de* and *naast de* was the middle of the schwa in the following [də]. The formant transitions of the schwa into the following consonant were part of the spliced files for the shapes. In each case three glottal cycles with a formant transition appropriate to the first consonant of the shape word were taken from the recorded utterances along with the shape word itself (e.g., the sound file for the shape ‘circle’ started with three cycles with an alveolar formant transition and then the waveform for the word [sɪrkel]).

Procedure and Design

The procedure was tailored in such a way that the presentation of a visual display contained no cues to what the target was. Therefore, we had to make sure that each quadrant of the screen and each of the four shapes were associated with the target on the same number of trials. Moreover, the distractors on a given trial had to be not recog-

nizable as such. To achieve this, we used the same items as targets, competitors, and distractors on different trials.

For semantic trials, the target and the competitor were identical words, to be distinguished either by the positional relation with their associated shapes or by the type of shapes. Distractors on these trials were two tokens of another printed word, so that the appearance of two identical words on the screen did not give that pair away as necessarily containing the target. In addition, distractors were accompanied by the same object on position trials, but in different spatial relations, just as the target and its competitor were. For example, if the target was above a star and its competitor was next to a star, then both of the distractors could have been associated with a circle (one above and one next to it). Similarly, on shape trials, two identical distractors appeared both above or both next to different objects, just as targets and competitors did. For example, if the target was above a star and the competitor was above a circle, one distractor could have been next to a triangle and the other next to a rectangle.

For the nasal and stop trials—the trials with minimal pairs, such as *kast* and *kas* and *don* and *dom*—four different words from two minimal pairs appeared on the screen, one differing with regard to the presence of word-final /t/ and one differing in word-final /n/-/m/ (see Figure 1). Because the word-final /t/ minimal pairs served as targets and competitors on stop trials and as distractors on the nasal trials, and vice versa for the /n/-/m/ pairs, any of the four words on a given trial could have been a target. Furthermore, on each stop and nasal trial, targets and competitors appeared in the same spatial relation with regard to their shapes, and both distractors appeared in the same spatial relation to their shapes (see Figure 1).

Participants were tested individually in a sound-damped booth. They were told that they would hear instructions, presented over headphones, directing them to use the computer's mouse to click on one of the words in displays that they would see on the computer screen. Each participant first completed six practice trials before eye-tracking cameras were mounted. The practice trials used the minimal nasal and stop pairs that had been rejected from the main set of materials and unedited spoken instructions. There were four trials in which four different words comprising two minimal pairs were presented on a screen plus two trials with two words appearing twice on the screen, in which either the preposition or the shape disambiguated the printed word to be clicked on. After the practice session, the eye-tracker (an SMI Eyelink system, sampling at 250 Hz) was mounted and calibrated. After that, 144 experimental trials were presented. The presentation of the auditory stimuli (spoken instructions) and visual stimuli (printed words and shapes) was controlled with the NESU programming package running on a standard PC.

Each participant received a different randomization of trials, always starting with a semantic trial. In order to generate a trial list for each participant, an abstract trial list was created first with 48 semantic trials, 32 nasal trials, and 64 stop trials. This abstract list specified in which quadrant of the display the target word appeared, which shape was paired with the target, and what the spatial relation between the two was. The semantic trials contained 24 position trials in which the position of the word relative to the printed word disambiguated the instruction and 24 shape trials in which the object itself disambiguated the instruction. For example, the instruction "click on the word *zoon* above the star" is disambiguated by the position if the display contains the word *zoon*

above and next to a star, but is disambiguated by the shape itself if the display contains the word *zoon* once above a star and once above a circle. Each combination of target, shape, and quadrant was used at least once for each condition and never more than twice. Each shape and quadrant was associated with the target six times in the 24 shape and 24 position trials.

The 64 stop trials contained four major conditions with 16 trials each. These four conditions arose by presenting stimuli with a +/t/-bias or a -/t/-bias in either a labial context or an alveolar context (i.e., *boven* or *naast*). Each combination of target, shape, and quadrant was used once in each of the four conditions. Similarly, the 32 nasal trials contained two major conditions in which an ambiguous target word was followed by either *boven* or *naast*.

Items were allocated to this abstract trial list in a randomized fashion that was different for each participant. For the semantic trials, each of the 48 minimal pairs was used once as a target and once as a distractor. For each trial, one member of one of the 48 pairs was selected at random, but without replacement, to serve as a target (and as the identical competitor) and one member of another pair was chosen in the same way to serve as the two distractors. Target position, the shape going with the target, and the spatial arrangement of the two were determined by the trial structure. The positions of the competitor and the two distractors, as well as their accompanying objects, were then chosen randomly, within the remaining degrees of freedom. So in a shape trial, for example, the competitor had to be in the same relative position to the object as the target.

Overall, each minimal pair was used three times as a target, twice in its respective phonological condition, and once in a semantic trial. In order to prevent target repetitions across adjacent trials, each minimal pair was used once in every third of the experiment as a target. Moreover, each Ct#-C# pair was used once as a distractor on a nasal trial; and each n#-m# pair was used four times as a distractor on stop trials. This design ensured that on every trial, two pairs of similar or identical words appeared on the screen, and the trial structure contained no cues to what the likely target might be. After the allocation of items to the abstract trial list, each third of the list was randomized, with the constraint that target repetition at the adjoining ends and beginnings of these thirds was prevented. Moreover, drift-correction trials for eye-tracking were added after every sixth trial.

In summary, the design entails three types of trials with different independent variables. For the semantic trials, the independent variable is the disambiguating information (position vs. shape). For the stop trials, there are two independent variables, form (\pm /t/-bias) and following context (bilabial vs. alveolar, *boven* vs. *naast*). For the nasal trials, the only independent variable was following context. There were three dependent variables. Two dependent variables were identical for all types of trials: accuracy and reaction time of the mouse clicks. The third dependent variable was derived from the eye-movement data: the Euclidean distances between the current fixation and the centers of gravity of the four words on the screen, which coincided with the center points of the four quadrants. On the semantic trials, the dependent variable was the distance to the target; on the stop trials, it was the distance to the words with /t/; and, on the nasal trials, it was the distance to the /n/-final word. If at any point there was a saccade or a blink, the last fixation position was used to determine the Euclidean distances. Usually, eye-tracking studies use fixation proportion as a dependent variable (Allopenna et al., 1998; Salverda

et al., 2003; Shatzman & McQueen, 2006; Tanenhaus et al., 1995). Given that the printed word and its associated shape formed a Gestalt on the screen (see Figure 1), it is difficult to determine when a fixation actually is on that Gestalt. Using a distance measure frees us from the need to define some arbitrary boundary of the printed word-shape Gestalt. Moreover, if a saccade undershoots or if there is a small deviation between actual and recorded fixation position due to drift, a distance measure still correctly reflects that the fixation is closest to one of the four Gestalts on the screen, whereas a fixation measure would yield a fixation on none of the presented objects.

Euclidean distances were sampled at intervals of 25 ms, with small deviations depending on the duration of the different components of the instructions. The parts containing *klik op het woordje* were on average 900 ms (range 843 to 950 ms). The first 750 ms (i.e., up to the first splice point, see Figure 2) were identical for all trials and eye positions were sampled at 30 equal 25-ms intervals in this window. The part of the utterance with the target word—counting from the coarticulated last syllable of *woordje*—had durations ranging from 388 to 714 ms, with a mean of 538 ms. Eye-position was sampled for 22 equally spaced points in these intervals, leading to average intervals of 24.5 ms. The positional prepositions plus the following definite article had an average duration of 468 ms (446 to 491 ms), and eye positions were sampled at 19 intervals with an average duration of 24.6 ms. From the onset of the word for the object, eye positions were again sampled every 25 ms for another 1.5 s.

Results

Fifteen experimental sessions could not be finished due to malfunctions of the experimental software. For the eye-tracking data, we discarded all trials that failed a maximum-distance criterion, namely those trials in which eye-position was tracked poorly, with estimated fixations outside the screen area. This excluded nearly all trials for one participant, for which eye-tracking calibration was poor. The data from this participant were therefore excluded from any further analysis, leaving 24 participants for the final analysis, for which 193 trials were excluded (5.5% of the trials) by the maximum-distance criterion.

Task Performance

Clicks were counted as being sufficiently close to the target object if they were within 120 pixels in the width-dimension and 90 pixels in the height-dimension. (These tolerances were determined by a visual inspection of the deviation distribution.) For all types of trials, we analyzed the error pattern and, for correct trials only, the Reaction Times—RTs; we used $\log_e(\text{RT})$ in the statistical analyses. Significance of effects was in all cases assessed with linear mixed-effects models with item and subject as random factors and in which higher-order interactions were successively removed if not significant. Trial number was included in these analyses. Trial number reduced error variance significantly in the semantic and stop trials (RTs became faster toward the end of the experiment). This was not the case for the nasal trials.

Semantic trials. Table 2 shows the accuracy and the mean RTs for trials in which the target was disambiguated earlier in the sentence by the positional preposition or later in the sentence by the shape itself. Accuracy did not differ significantly between

Table 2
Task Performance in the Semantic Trials (Experiment 1)

Condition	% correct	Reaction time, in ms
Shape	99.5	2,109
Position	99.8	2,023

conditions ($p > .1$), whereas RTs did, $t(1,122) = 4.6$, $p < .001$, $d = 0.23$ ¹: Participants responded faster when the disambiguation occurred earlier.

Nasal trials. Table 3 shows the mean error rates and the mean RTs for trials in which the target ended on either /n/ or /m/ followed by either the preposition *boven* or *naast*. There were no significant effects of the independent variables on either accuracy or latency ($p_{\min} > 0.1$).

Stop trials. Figure 3 shows the mean error rates and the mean RTs for trials in which the target ended on either /C/ or /Ct/ followed by either the preposition *boven* or *naast*. There were no significant effects on the accuracy rates ($p_{\min} = 0.07$). However, there was a significant interaction of target, form, and following context for the RTs, $t(1,456) = 2.0$, $p < .05$, $d = 0.35$. As Figure 3b shows, this is due to the fact that participants reacted faster if the target was the /C#/#/ word, but only if the following context was *naast* and the targets had a $-t/$ bias. In all other cases, reactions were faster to /Ct#/#/ targets.

Eye Movements

Semantic trials. Figure 4 represents the averaged distances of the fixation positions to the different stimuli. At the beginning of the sentences, participants did not have a preference for any of the stimuli. During the presentation of the targets, participants started to look towards the target and the competitor at the cost of the distractors. During the presentation of the positional preposition, participants started to look away from the competitor and towards the target only in the position condition, in which the preposition distinguished the target from the competitor (remember that in these trials targets and competitors were the same word). In the shape condition, however, targets became preferred over competitors only after the presentation of the shape name (i.e., again at the point where targets and competitors were distinguished).

Because the distances to target and competitor are not independent from one another, the statistical analysis tested for differences in the distance to target between conditions. In order to reduce the data, we aggregated the distance-to-target measure for the time window 300 to 700 ms after onset of the preposition. In this window, the distance to the target was significantly smaller in the position condition than in the shape condition, $t(1,003) = 6.1$, $p < .001$, $d = 0.35$. To narrow down the time frame in which the preposition had a significant effect, we conducted a series of

¹ There is no consensus yet on how to calculate effect sizes analogous to eta-squared in a linear-mixed-effect model. Hence we report Cohen's d , which was calculated by dividing the estimated effect size by the pooled standard deviation.

Table 3
Task Performance in the Nasal Trials (Experiment 1)

Preposition	Target ended on	% correct	Reaction time, in ms
boven	/n/	98.4	2,089
	/m/	98.4	2,111
naast	/n/	98.4	2,159
	/m/	99.0	2,161

analyses in which we compared target distance in the shape condition with target distance in the position condition for each bin between bin 60 and bin 100. Figure 5 shows the results of these analyses, revealing a significant effect of condition between bins 71 and 92. We therefore used a window from bin 72 to bin 91 to evaluate the effects of the independent variables for the other type of trials, as in this “window of ambiguity” the information about shape identity does not yet direct gaze unambiguously to the target.

Note also that the distance functions in Figure 4 begin to converge after bin 100 in the position condition and after bin 105 in the shape condition. At these time points (approximately 1,750 to 1,875 ms after target onset), about a third of the mouse clicks have already occurred, and the convergence thus reflects movements to random parts of the displays after the task instruction has been completed. Similar convergence patterns can be seen in the other trials (and in Experiment 2); all such patterns can be ignored.

Nasal trials. Figure 6 shows the general patterns of eye movements for the nasal trials as well as the stop trials in terms of the distance between the fixation position and the different types of stimuli with regard to their function in the trial, as in Figure 4. The general pattern is similar to that for the semantic trials. The hypothesis for the nasal trials, however, was that fixations would be closer to the /n/-final words if the target word occurred in a labial context than if the target occurred in an alveolar context, independent of whether the /n/-word happened to be the target or the competitor on a given trial. Figure 7 therefore plots the distance between fixation position and the /n/- and /m/-final words, as well as to the distractors for both context conditions. As inspection of the figure suggests, there was no effect of condition on the distance to the /n/-final words ($t^2 < 1$) in the designated window (bins 72 to 91).

Stop trials. As for the nasal trials, Figure 8 plots the distance between fixation position and the different types of stimuli (/Ct#/-words, /C#/-words) independent of their status of target and competitor (see the lower panel of Figure 6 for the general patterns of fixation to target, competitors, and distractors). Both acoustic form, $t(1,403) = 4.2, p < .001, d = 0.22$, and following context, $t(1,396) = 3.6, p < .001, d = 0.18$, influenced the distance to the /Ct#/-word in the designated window (bins 72 to 91), whereas there was no interaction between these factors, $t(1,423) = 1.6, p > .1$. The mean distance to the /Ct#/- words was 27 pixels smaller if the following context was labial (“boven”) and 32 pixels smaller if the coda consonant had a +/t/ bias.

It is also interesting to note that the overall pattern reveals a preference to look towards the /Ct#/- words rather than to the /C#/- words in the window of ambiguity (compare the upper and lower panels of Figure 8). Furthermore, after presentation of the target word and the following positional preposition, there were further changes in the distance to both types of words except the condition with a -/t/ bias in alveolar context. In this condition, the distances

to the /Ct#/- and the /C#/- words were stable over time from about bin 60 onwards and comparable (approximately 250 pixels for both word types).

Discussion

The main purpose of Experiment 1 was to investigate whether listeners use following phonological context in compensation for /t/ reduction. The results showed that they do: A word was more likely to be interpreted as containing a /t/ if the following context started with a labial. This effect is functional for speech perception, because speech production data have shown that a following labial makes it more likely that a word-final /t/ will be reduced (Mitterer & Ernestus, 2006). There was also an effect of acoustic form: Primes with a short consonant and a longer silence were more likely to be interpreted as containing a /t/ than forms with a longer consonant and a shorter silence, replicating the perceptual results of Mitterer and Ernestus.

The influence of acoustic form and following context appeared to be different across dependent measures (the two factors interacted in the click latencies but were additive in the eye-tracking data). But a more detailed comparison shows that there is no real conflict here. As shown in Figure 3, the

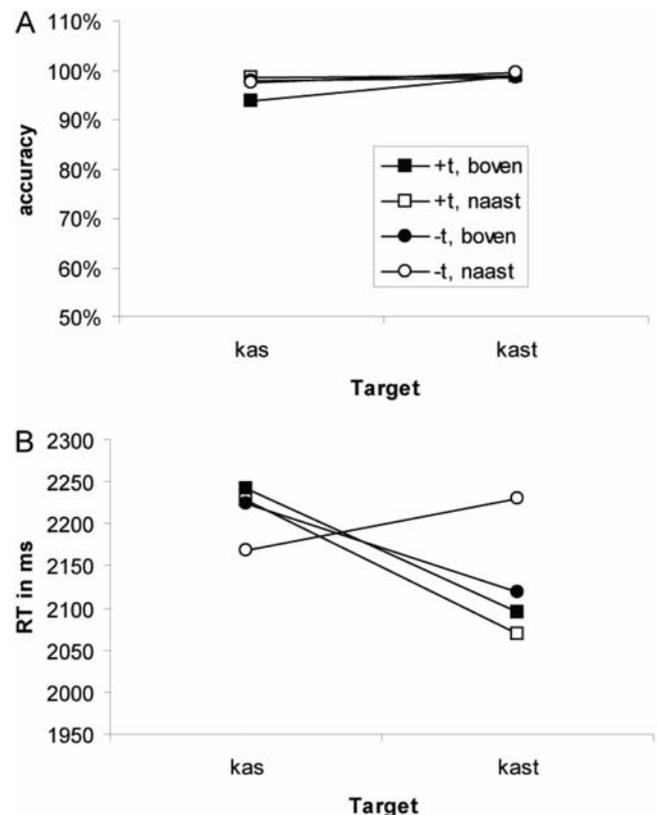


Figure 3. Experiment 1: Accuracy (Panel A) and mean Reaction Times (RTs, in ms; Panel B) for trials in which the target ended on either /Ct/ or /C/ depending on the type of the target word (/C/, e.g., *kas*, or /Ct/, e.g., *kast*), the acoustic form of the coda (favoring /Ct/, +t, or favoring /C/, -t) and the following context (*boven* or *naast*).

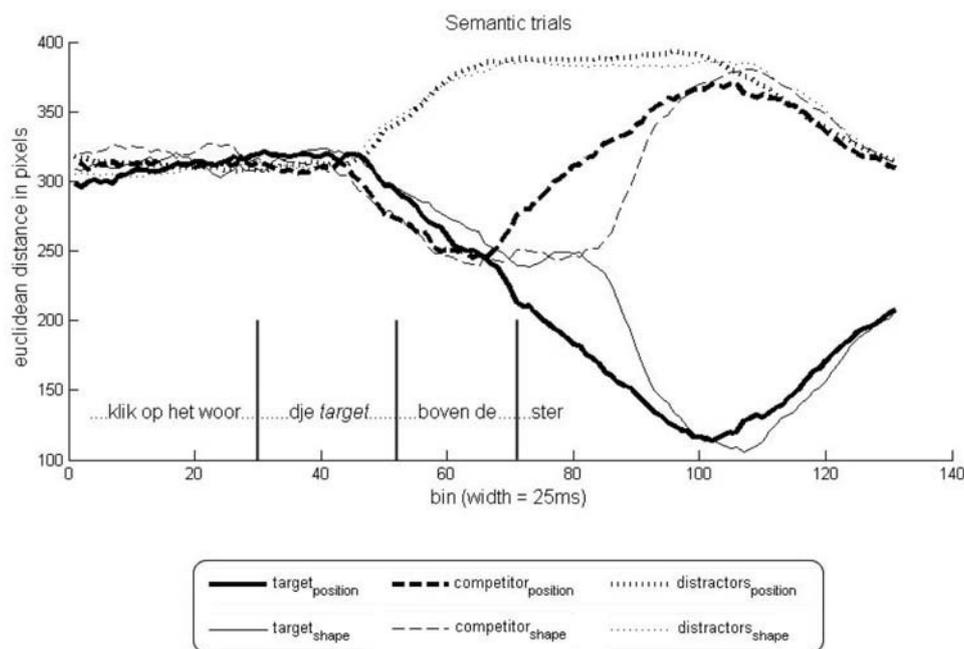


Figure 4. Experiment 1: Distance of the fixations to the different stimuli in the semantic trials. The thick lines represent the data from the position condition, the thin lines the data for the shape condition. The continuous lines represent the distance to the target, the dashed lines the distance to the competitor, and the dotted lines the averaged distance to the distractors. The vertical lines indicate the splicing points in the sentences.

condition with a *-t/* bias and an alveolar following context behaved differently from the other three conditions (specifically, there was no RT bias favoring either target in this condition, but in the other conditions responses to */t/-*final targets were faster). As shown in Figure 8, the same condition was again the exception (i.e., the viewing distance for only this condition was comparable for both types of target from bin 60 onwards, whereas for the other three conditions there was a bias favoring */t/-*final targets). It thus appears that according to both measures there was a */Ct#* target bias for all stimuli except

those with the strongest evidence against a */t/-*final interpretation, for which there was no overall bias.

The results from the semantic trials revealed that participants were using information immediately as it came in, even at a semantic level. The eye movements to the target occurred earlier if the positional preposition gave away the target than if the object gave it away. In the nasal trials, however, we did not find the expected effect of following context. This is especially surprising as an earlier eye-tracking study found an effect of following context on the perception of assimilated segments

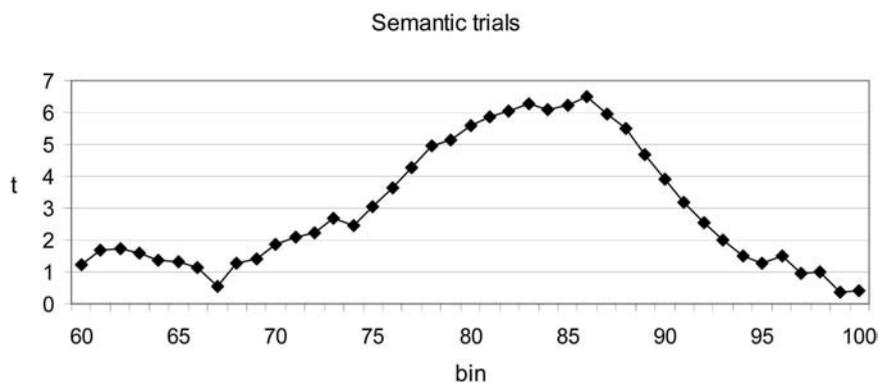


Figure 5. Experiment 1: Results of bin-by-bin analyses comparing the shape and position conditions in the semantic trials. The *t* values indicate that fixations were closer to the target from bin 71 to bin 92, $t_{crit} (df > 120) = 1.98$.

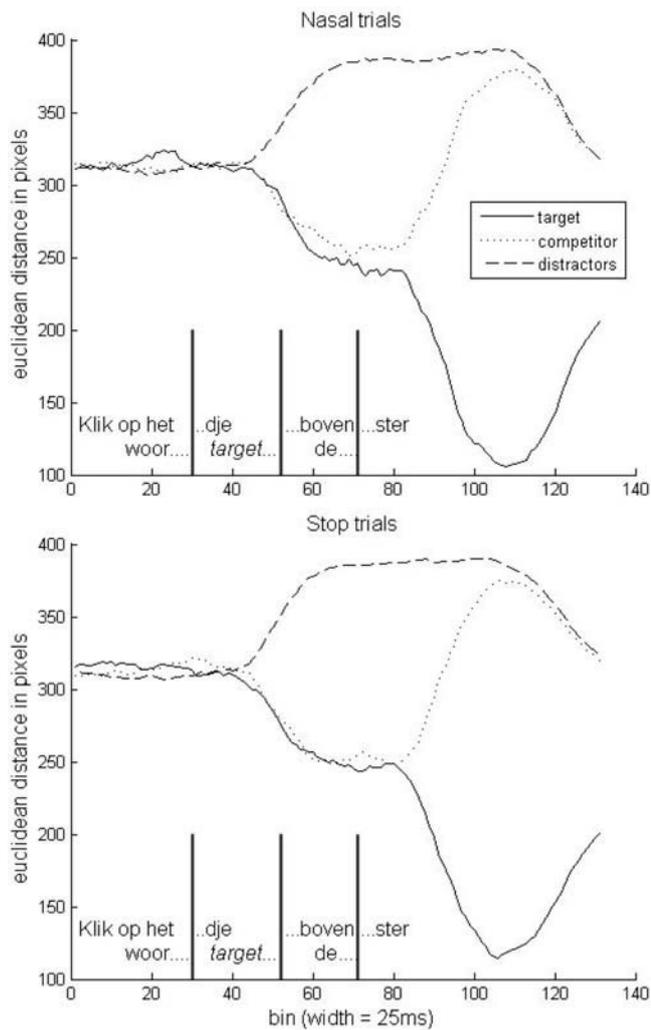


Figure 6. Experiment 1: Distance between fixation position and the different types of stimuli based on their function in the trial in the nasal and stop trials. The continuous lines represent the distance to the target, the dashed lines the distance to the competitor, and the dotted lines the averaged distance to the distractors. The vertical lines indicate the splicing points in the sentences.

(Gow & McMurray, 2007). One possible explanation for this failure to replicate may lie in the way the current stimuli were generated by mixing glottal cycles from a labial and an alveolar nasal. Because glottal cycles hardly ever have exactly the same duration, it was necessary to zero-pad the shorter of the two cycles. This adds a noise-like characteristic to the stimuli, which may have led participants to perceive the nasal as slightly masked by noise and to adopt the strategy of completely ignoring any place information in the nasals. This explanation gains credibility from the fact that previous studies with cross-modal priming, identification tasks, and electrophysiological recordings have shown that the effect of following context on the perception of assimilated segments depends crucially on the phonetic implementation of those segments (Gow, 2002; Mitterer, 2003; Mitterer, Csépe, & Blomert, 2006).

Experiment 2

The aim of Experiment 2 was to examine why an effect of following context was found in the stop trials in Experiment 1 but not in Mitterer and Ernestus (2006). We argued, in motivating Experiment 1, that effects of following context may only be found in more natural listening situations, where listeners are attending to the meaning of spoken utterances, and hence that the reason for the absence of an effect in the earlier study might have been that the 2AFC task used there focused listeners' attention too much on the acoustic-phonetic detail in the stimuli. One reason the semantic trials in particular were included in Experiment 1 was to focus listeners' attention on the semantic level of processing and away from the acoustic-phonetic detail in the stop trials. In Experiment 2, therefore, the semantic and nasal trials were removed. The question was whether this would influence the following-context effect in the stop trials. If so, this would strengthen our claim that effects of following context depend on the kind of information that listeners are encouraged to attend to.

Method

Participants

Twenty-six members of the Max Planck Institute subject panel participated in the experiment. All participants had learned Dutch

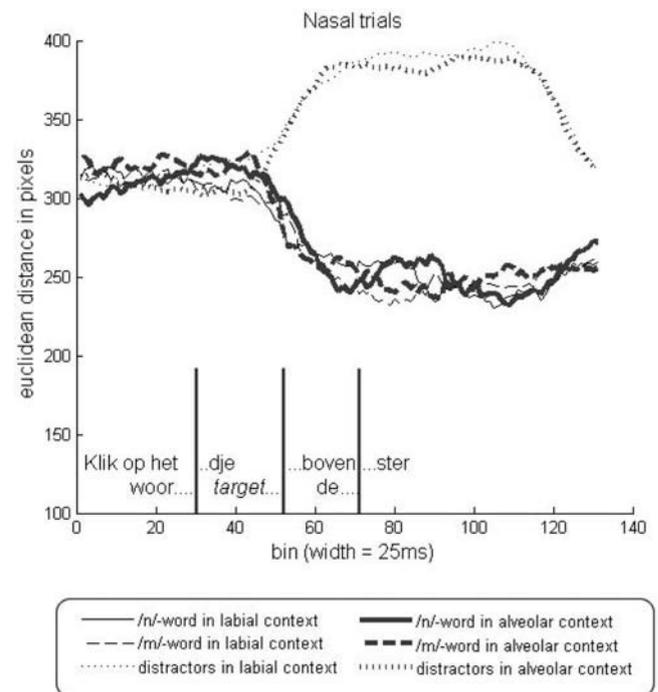


Figure 7. Experiment 1: Distance between fixation position and the different types of stimuli in the nasal trials based on their phonological properties. The thin lines show the distances in the labial-context condition ("boven") and the thick lines show the distances in the alveolar-context condition ("naast"). The continuous lines indicate the distance of the fixation to the /n/-words, the dashed lines to the /m/-words, and the dotted lines to the distractors. Note that the distance to the /n/- and /m/-words comprises cases in which these words were targets as well as competitors.

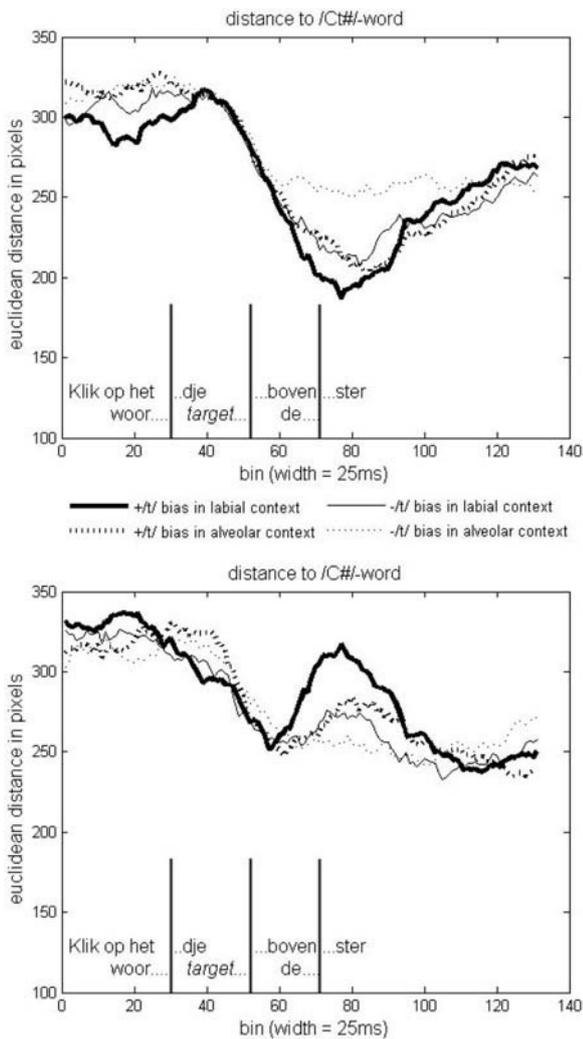


Figure 8. Experiment 1: Distance between fixation position and the different types of stimuli in the stop trials based on their phonological properties. The thick lines indicate the distances for stimuli in which the acoustic form induced a +t/-bias, the thin lines those in which the form induced a -t/-bias. The continuous lines show the distances in the labial-context condition (“boven”) and the dotted lines show the distances in the alveolar-context condition (“naast”).

as their native language. None reported any hearing problems, and none had taken part in Experiment 1.

Stimuli, Procedure, and Design

The stimuli were a subset of those used in Experiment 1. All trials that did not belong to the group of stop trials were filtered out of the trial order files used in Experiment 1. The experimental sessions then consisted of 64 trials with four conditions with 16 trials each. These four conditions arise by presenting stimuli with a +t/-bias or a -t/-bias in a labial context or an alveolar context (i.e., *boven* or *naast*). The design for these trials was thus exactly the same as before, but the design of the experiment as a whole was much simpler,

because the semantic and nasal trials were excluded. Experimental procedure and data analysis were identical.

Results

Trials in which eye-position was tracked poorly, with estimated fixations outside the screen area, were again excluded. This led to the rejection of 46 trials (2.6%), with no more than six trials rejected per participant (<10%).

Task Performance

As in Experiment 1, clicks were counted as being sufficiently close to the target object if they were within 120 pixels in the width-dimension and 90 pixels in the height-dimension. Significance of error and RT effects was again assessed with linear mixed-effects models with items and subjects as random factors, and higher-order interactions were again successively removed if not significant. Trial number was used in these analyses to reduce error variance. It did so marginally significantly in the error analysis and significantly in the RT analysis.

Figure 9 shows the mean error rates and the mean RTs depending on target type (either /C/ or /Ct/), following context (alveolar or labial), and type of coda signal ($\pm t$ -bias). There were no significant effects on the accuracy rates ($p_{\min} > 0.2$). Analysis of \log_e RTs revealed an effect of target type, $t(1,614) = 3.5$, $p <$

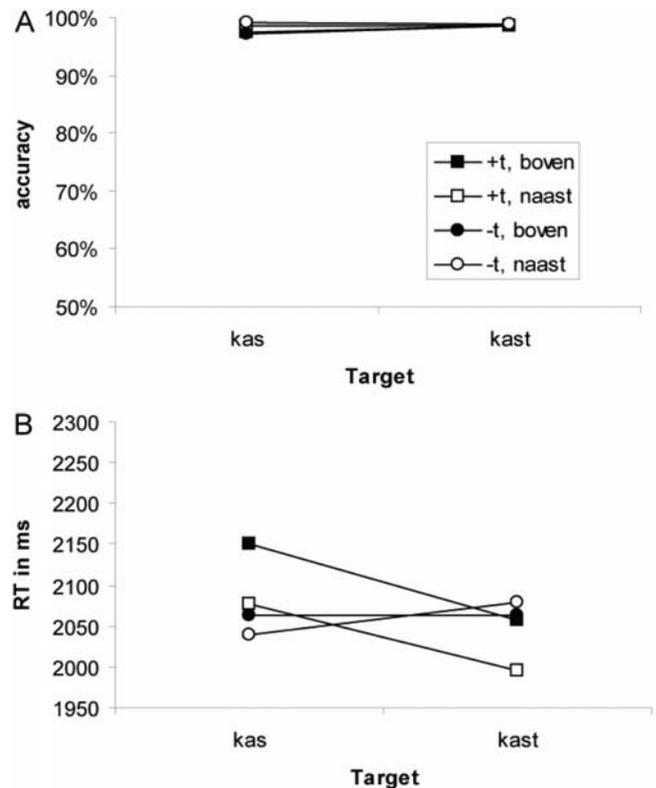


Figure 9. Experiment 2: Accuracy (Panel A) and mean Reaction Times (RTs, in ms; Panel B), depending on the type of the target word (/C/, e.g., *kas*, or /Ct/, e.g., *kast*), the acoustic form of the coda (favoring /Ct/, +t, or favoring /C/, -t), and the following context (*boven* or *naast*).

.001, $d = 0.26$, with faster reactions for /Ct/ targets; an effect of following context, $t(1,621) = 2.1$, $p < .05$, $d = 0.10$, with slower reaction times for the labial context overall, independent of target type; and an interaction of \pm /t/-bias with target type, $t(1,621) = 2.7$, $p < .01$, $d = 0.28$, with faster reactions to /C/-targets if the form had a -/t/-bias. None of the critical interactions involving target type and following context were significant, indicating that following context did not influence strongly the speed with which the target words were interpreted as containing a /t/ or not.

Eye Movements

Figure 10 plots the overall distance between targets, competitors, and distractors and the fixation position. Although the overall pattern is similar to Experiment 1, there seems to be an early preference for the target and competitor over the distractors before target onset. Note that this is *not* an effect of precognition, because the participants in this experiment could predict that the potential target belonged to the /Ct#/-/C#/ minimal pair. We tested this effect in bins 20 to 40 (i.e., before the onset of the target) by aggregating the distance to the target and competitor and subtracting those from the distance to the distractors. These values were tested against zero in a mixed-effect model. Neither trial number nor the intercept, reflecting an overall bias, were significant ($p_{\min} > 0.2$). With trial number removed, there was a significant effect of intercept, $t(30) = 3.2$, $p < .01$, $d = 0.08$, indicating a preference of the /Ct#/-/C#/ minimal pairs over the /n/-/m/ minimal pairs.

Figure 11 plots the distance between fixation position and the different types of stimuli (/Ct#/-words, /C#/ -words) independently of their status as target or competitor. As in Experiment 1, we aggregated the distances to the /Ct#/-words in the window of ambiguity (bins 72 to 91). A linear mixed-effect model analysis on these measures indicated that both the acoustic form, $t(1,585) =$

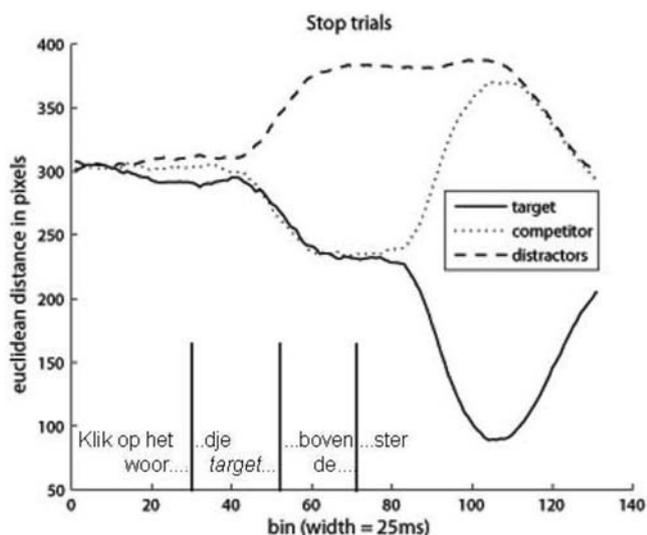


Figure 10. Experiment 2: Distance between fixation position and the different types of stimuli based on their function in the trial. Note that target and competitors were members of /Ct#/-/C#/ minimal pairs and distractors were members of /n/-/m/ minimal pairs.

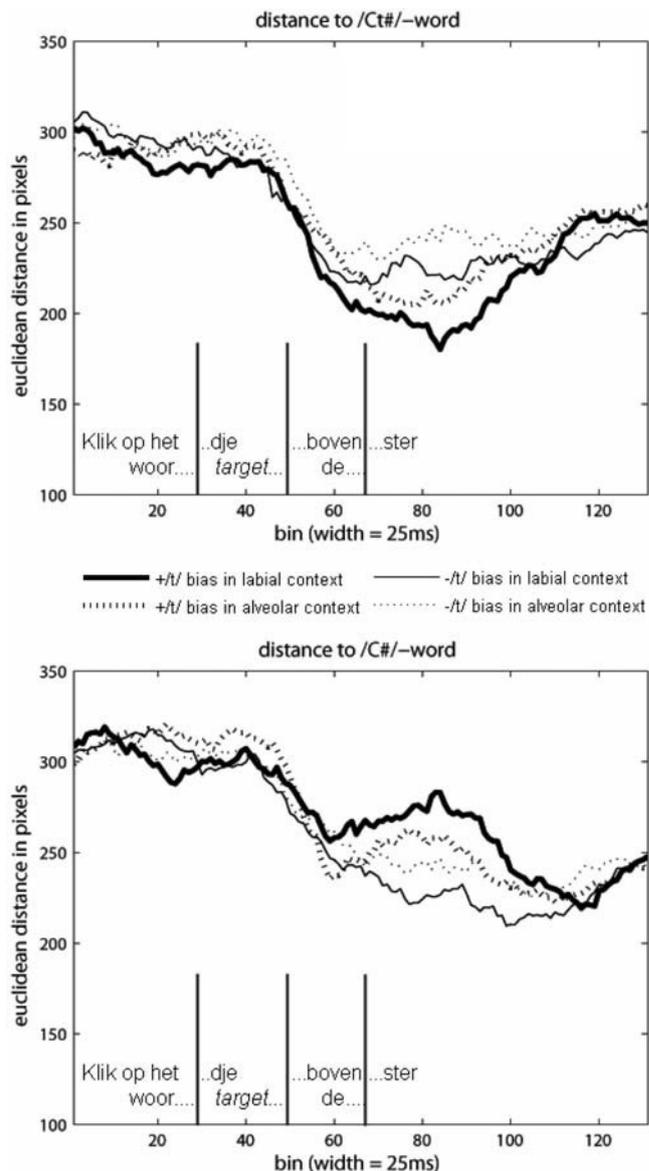


Figure 11. Experiment 2: Distance between fixation position and the different types of stimuli in the stop trials based on their phonological properties. The thick lines indicate the distances for stimuli in which the acoustic form induced a +/t/ bias, the thin lines those in which the form induced a -/t/ bias. The continuous lines show the distances in the labial-context condition ("boven") and the dotted lines show the distances in the alveolar-context condition ("naast").

4.8, $p < .001$, $d = 0.24$, as well as the following context, $t(1,580) = 2.4$, $p < .05$, $d = 0.12$, influenced the distance to the /Ct#/-word in the designated window, whereas there was no interaction between the factors ($t^2 < 1$). The mean distance to the /Ct#/-words was 17 pixels smaller if the following context was labial (boven) and 33 pixels smaller if the coda consonant had a +/t/ bias. Interestingly, there was also an effect of trial number. Over the course of the experiment, participants had a tendency to look more to the /Ct#/-words, $t(1,574) = -2.9$, $p < .01$, $d = 0.24$. This

indicates that the overall tendency to perceive the stimuli as containing a /t/ increased over the course of the experiment.

We also conducted a pooled analysis of Experiment 1 and 2, again with distance to the /t/-word as the dependent variable. In this analysis, the effects of acoustic form and following context were significant ($p < .001$) and independent of the factor experiment (interactions with Experiment, $F_s < 1$). There was only one significant interaction involving the factor experiment, that with trial number ($p < .01$). Participants tended to look more at /t#/-words over the course of Experiment 2 but not over the course of Experiment 1. As previously mentioned, and as can be seen in Figures 8 and 11, participants had a preference for the /t#/-words in the window of ambiguity, which apparently got larger in the experiment that focused on the /Ct#/-/C#/ minimal pairs.

Given that we found an effect of following context in these data sets, we also investigated whether there was also an effect of preceding segmental context. Although the scarcity of minimal pairs and the focus on the following context in the current design make it far from perfect for this purpose, we can still test whether the 21 pairs with preceding /s/ (such as *kas-kast*) generate more looks to the /Ct#/ words than the 11 other pairs with another preceding consonant (/l/, /r/, /x/, or /f/). This was the case. The mean distance to the /Ct#/ words for the items with /s/ as penultimate consonant was 14 pixels smaller in the window of ambiguity than for the other items, $t(3,036) = 2.4$, $p < .05$, $d = 0.10$, and independent of experiment ($F < 1$). Hence, we find an effect of both following and preceding context on the distance of the fixation position to the /Ct#/ word.

Discussion

This experiment replicated both critical effects on recovery from /t/ reduction found in Experiment 1. Both acoustic form and following context influenced the interpretation of a word that might or might not have had a word-final /t/. The effect of trial number nevertheless indicates that participants clearly learned about the likely targets over the course of the experiment, unlike in Experiment 1. This highlights that participants were affected by the semantic and nasal trials used in Experiment 1. Despite these strategic effects in Experiment 2, including the early looks to the /Ct#/-/C#/ minimal pairs, the results from the window of ambiguity suggest that the filler trials in Experiment 1 were not essential.

Nevertheless, we investigated how stable the effects were by drawing random samples of 16 subjects and running the analysis on 1,000 such samples. The effect of acoustic form was stable: It was significant for 95% of the samples from Experiment 1 and all samples from Experiment 2. The effect of following context, however, was significant for 97% of the samples from Experiment 1 but only 40% of the samples from Experiment 2. This indicates that the effect of following context was much less stable in the second experiment.

These analyses thus suggest that although the effect of following context is replicable, it does depend on the task situation. When there are no semantic and nasal trials, participants notice the acoustic-phonetic manipulation in the stop trials more and, thus, use following context less. Moreover, they are more likely to develop strategies and, thus, over the course of the experiment, tend to look more to the potential targets during the precursor sentence and more towards the type of words they hear most (i.e.,

those ending on /t/). In a more extreme case, where task demands actively encourage listeners to focus on the acoustic-phonetic properties of the stimuli (i.e., the 2AFC task used by Mitterer & Ernestus, 2006), no effect of following context was found. Focusing listeners' attention on phonetic detail hence seems to weaken the effect of following context in compensation for /t/ reduction.

General Discussion

The purpose of these experiments was, first, to examine a new variation of the visual-world paradigm and, second, to apply it to investigate whether following segmental context influences the perception of words with reduced /t/. Previous experiments on spoken-word recognition with the visual world paradigm used pictures of objects or shapes (e.g., Allopenna et al., 1998; Dahan et al., 2001) or printed words (Huettig & McQueen, 2007; McQueen & Viebahn, 2007). We used both pictures and words here and combined them in a novel way. Targets were identified by combinations of printed words, shapes, and their spatial relation, in instructions such as "Click on the word *kast* above the star." Despite this added complexity, eye-movements reflected the immediate use of acoustic-phonetic information. Given the previous instruction, for example, listeners took longer to look towards the target if two instances of the printed word *kast* appeared on the screen above their associated shapes than if one of the printed words *kast* was above its shape and the other was next to its shape. Mouse-click latencies (as a measure of explicit task performance) indicated a similar disadvantage in the former type of trial. This mirrors the temporal unfolding of the instruction sentence ("... *kast* above the star"), in which the information about the position of the word precedes the identity of the geometric shape.

With this paradigm it is possible to exert precise control over the phonological context in which a word appears without unnatural or, given the visual displays, odd sentence constructions. Although we failed to replicate the effect of following context on the perception of assimilated segments with this design—probably due to the nature of our stimuli (Gow, 2002; Mitterer, 2003; Mitterer, Csépe, Honbolygo, et al., 2006)—we succeeded in showing in both experiments that following context has an influence on the perception of reduced /t/. Listeners were presented with minimal pairs differing in the presence or absence of a word-final /t/, such as *kast* (/Ct#/) and *kas* (/C#/), as printed words on the screen. At the same time, an ambiguous pronunciation was presented, which either favored a /Ct#/ or a /C#/ interpretation. The stimuli with a +/t/ acoustic bias contained a shorter penultimate consonant and a longer closure than the stimuli with -/t/ acoustic bias (following properties of normal spoken Dutch; Mitterer & Ernestus, 2006). This manipulation was successful: Distances between fixation position and the /Ct#/ word were smaller if the form had a +/t/ bias. More importantly, listeners also looked more toward the /Ct#/ words if the following context was labial. Behavioral measures also indicated an influence of following context on compensation for /t/ reduction: Mouse clicks on /C#/ words were faster than clicks on /Ct#/ words if both the following context and the phonetic form biased perception towards the /C#/ interpretation. An additional analysis showed that previous context also influenced eye movements, so that listeners looked more towards the /Ct#/ words if the preceding consonant C was an /s/, in line with earlier findings (Mitterer & Ernestus, 2006). Therefore, the current data

indicate the efficacy of all three cues to the presence or absence of a word-final /t/. Acoustic cues, previous context, and following context all influence compensation for /t/ reduction.

A slightly problematic aspect of the data is that we did not observe smaller distances to the /C#/ words than to the /Ct#/ words in any of the conditions in both experiments. It thus appears that, overall, stimuli were slightly biased toward a +/t/ interpretation. One possible explanation for this pattern is that all stimuli were generated from utterances with an underlying word-final /t/. Although we manipulated the temporal properties of these pronunciations, there is the possibility that their spectral properties carry additional cues to the underlying /t/. Such spectral properties would obviously not have been obliterated by manipulation of the temporal cues. To evaluate this possibility, the spectral properties of the consonants before a word-final /t/ were compared with the spectral properties of the same consonants when they were themselves word-final. Figure 12 shows how the spectral centers of gravity of these segments changes over time. There is a great deal of commonality between the different consonants. The presence of a /t/ has little influence on the spectral properties of the beginning

of the preceding segment, but the tails of all five segments have a higher center of gravity if an underlying /t/ follows—possibly reflecting the smaller front cavity as the tongue approaches an alveolar closure. It is likely that these spectral cues contributed to the overall bias towards /Ct#/ words.

This bias, however, was orthogonal to the other effects that were found. Replicating Mitterer and Ernestus (2006), the current results indicate that listeners treat the presence of a long penultimate consonant as evidence against an underlying /t/, whereas a long closure duration is treated as evidence for the presence of an underlying /t/. It is interesting to note that Mitterer and Ernestus obtained their results with synthesized speech, whereas the current results were obtained with edited natural speech. The convergence of results indicates that the use of synthetic speech in a perception experiment is a way to achieve stimulus control without necessarily sacrificing the experiment's ecological validity.

Most importantly, however, the current project uncovered an effect of following segmental context on compensation for /t/ reduction. Mirroring the production data (Mitterer & Ernestus, 2006), Dutch listeners are more likely to infer the presence of an

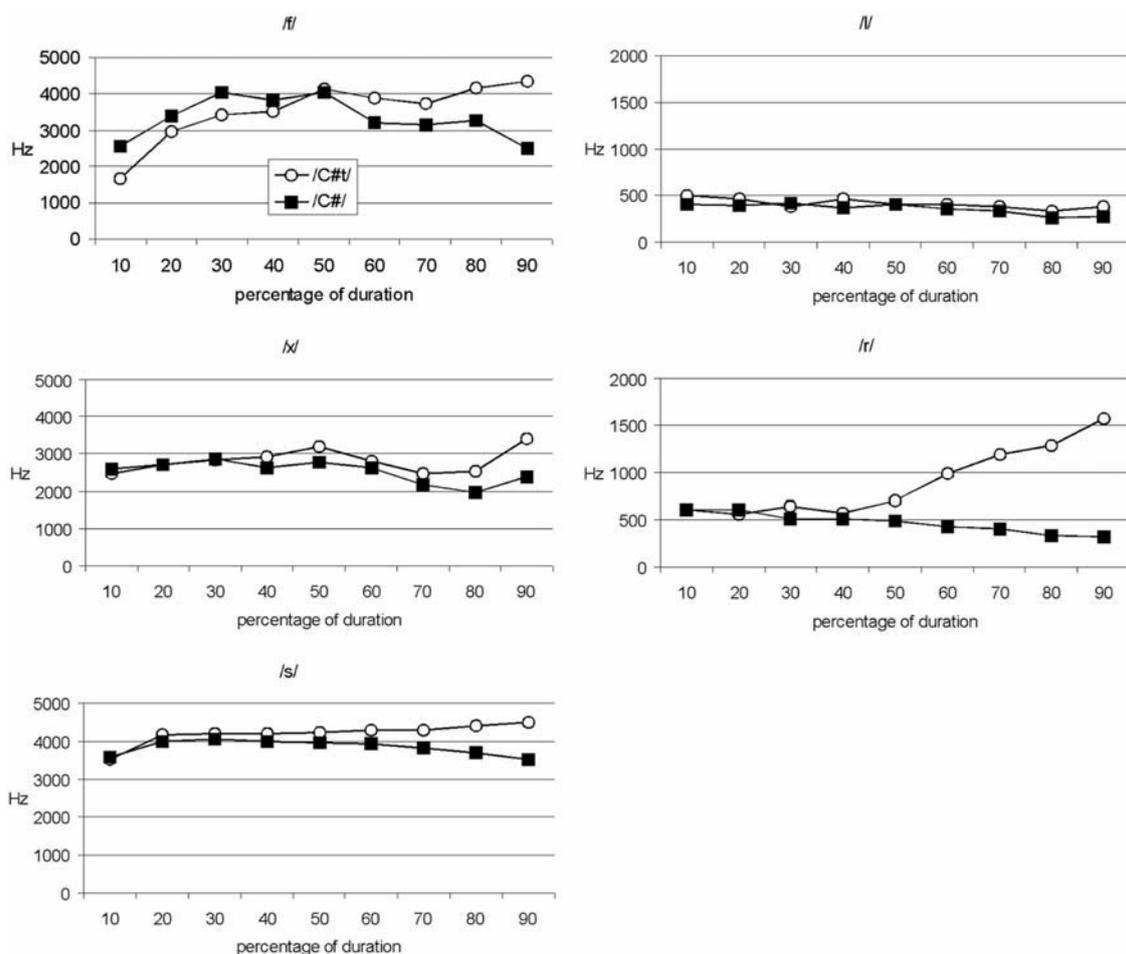


Figure 12. Spectral center of gravity for the penultimate consonant in the stop trials depending on the presence of an underlying word-final /t/, time-normalized for the duration of the segment, and the type of consonant (different panels). Note the different frequency scales for obstruent consonants on the left and sonorant consonants on the right.

underlying /t/ in a following labial context than in a following alveolar context. This makes it unlikely that the recognition of words with reduced word-final /t/ is due to the lexical storage of the reduced form. This is because the influence of the varying following context, which is not part of the lexical representation, is difficult to code lexically. The current data hence speak instead for a prelexical mechanism that compensates for the effects of /t/ reduction using acoustic cues, and both preceding and following phonological context.

We now consider the nature and locus of this prelexical compensation process. Here it is interesting to note that the effect arose in the current experiments but not in the 2AFC identification experiments of Mitterer and Ernestus (2006). This would seem to indicate that the effect of following context only arises in tasks that focus the participants' attention on meaning but not in tasks that focus attention on form. There are, however, other possibilities to explain this dissociation over tasks. First of all, the eye-tracking task might just be more sensitive than the 2AFC task. Secondly, the stimuli were different in the two studies. The latter objection is weakened by the fact that two out of three effects in compensation for /t/-reduction were found with both tasks. Both phonetic detail and previous context influenced both how close the looks were to the /t/-final words in the current experiments and how many /t/-responses were given in the 2AFC task. This suggests that the stimuli were comparable.

The fact that an effect of preceding phonological context was found in the eye-tracking and the 2AFC data also allows us to address the issue of task sensitivity. We transformed the eye-tracking data into binary perceptual identifications by comparing the mean distance to the /Ct#/ words in the critical time window with the mean distance to the /C#/ words. If the mean distance to the /Ct#/ word was lower than to the /C#/ word, we coded this as a /t/-identification. Based on this calculation, there were 4% more /t/-choices with the preceding context /s/ than with the other preceding contexts. In the 2AFC data of Mitterer and Ernestus (2006), however, there were 50% more /t/-responses with preceding /s/ than with preceding /n/. Even though this comparison is far from perfect, it suggests that, with respect to the preceding context effect, 2AFC is in fact more sensitive than eye tracking. It thus seems unlikely that the failure to find an effect of following context in the 2AFC task was caused by insufficient task sensitivity.

Thus, although it remains a valid point that the two experiments are different in several respects and, hence, that a more direct comparison of tasks and stimuli would be valuable, the available evidence nevertheless suggests that the difference in results reflects a difference in the level of processing that the two tasks tend to tap into. The 2AFC task may primarily measure early perceptual processes, whereas eye-tracking may tap more into higher-level processes. This distinction could be explained in the framework of the phonological inference model proposed in Gaskell (2003). According to this model, the speech stream is first phonologically recoded in a context-independent way, then a regressive inference is made on the basis of this initial phonological classification, and finally lexical access is attempted. Although this model was initially proposed to account for compensation for assimilation, it may also be applied to compensation for /t/ reduction. In this framework, the failure to find a following-context effect in the identification task can be explained by the assumption that this task may tend to probe the initial context-independent classifica-

tion stage, whereas the eye-tracking data may be more likely to be influenced by the later inference process.

If the difference in results between the eye-tracking and 2AFC tasks holds up to further experimental scrutiny, it would disfavor an auditory account of the context effect. Mitterer and colleagues (Mitterer, Csépe, & Blomert, 2006; Mitterer, Csépe, Honbolygo, et al., 2006) have argued that compensation for assimilation may (partly) be caused by auditory processes, on the grounds that it was difficult to find a task in which no context effect occurred. This was not due to lack of trying; they tested for context effects with a phonetic-identification task (in which feedback discouraged the application of phonological inference), discrimination tasks, electrophysiological measurements, and cross-linguistic comparisons, and by using nonspeech analogues. There was an effect of context in all tasks, which led Mitterer and colleagues to argue that the effect arises at an early (auditory) level of processing. By the same reasoning, the absence of an effect in the 2AFC task of Mitterer and Ernestus (2006) challenges an auditory account for the current following-context effect.

Similarly, it is difficult to explain the task dissociation in terms of a gestural account. According to gestural theories, the listener can infer (according to Motor Theory, see Liberman, 1996) or directly perceive (according to Direct Realism, see Fowler, 1996) that there was an alveolar closure for /t/, the release of which was masked by the oral closure for the following /b/. In Motor Theory, the failure to find the context effect in the identification task could be explained if it were assumed that the task reflects perception of reduced /t/ as an auditory signal rather than as speech. But speech perception is assumed to be preemptive in Motor Theory (Whalen & Liberman, 1987), which makes it difficult to argue that speech could be perceived as nonspeech. According to Motor Theory, there ought therefore to be following-context effects in all tasks. Direct Realism predicts the same thing. This is because it is assumed that there is a single level of perception, at which listeners perceive sound-producing gestures. Context effects should thus arise irrespective of the task listeners are asked to perform.

We used the fact that the effect of following context arose in the current experiments but not in those of Mitterer and Ernestus (2006) as evidence for the phonological-inference model (Gaskell, 2003). If one accepts this argument, however, it follows that the failure to find a dissociation in the case of compensation for assimilation is problematic for this model. If the dissociation of effects over tasks in compensation for /t/-reduction is evidence for the model, the absence of a dissociation in compensation for assimilation must be evidence against it. In more general terms, it appears that whatever account one assumes for compensation for assimilation, a different account is needed for compensation for /t/-reduction.

Acceptance of this tentative conclusion also has wider implications for the study of the perception of pronunciation variation. Although reduction of word-final /t/ and place assimilation seem to be rather similar processes, the data suggest that perceptual compensation for these two processes is achieved by different mechanisms. It would thus be premature to generalize the current argument for prelexical compensation and against lexical storage to other forms of pronunciation variation. Some other variation types may well be dealt with via prelexical compensation processes rather than lexical storage (including recovery from vowel epenthesis in Dutch, Donselaar, Kuijpers, & Cutler, 1999; and

recovery from variation in the realization of unreduced word-final /t/ in English, Sumner & Samuel, 2005). But for at least three more types of pronunciation variation, lexical storage seems to be the more promising solution. First, it is difficult to see how a purely prelexical mechanism could account for the reconstruction of extremely reduced forms, such as the Dutch [ɪfal] for /in ider gəval/, in *ieder geval*, [in any case] (Ernestus, 2000). Ernestus and colleagues (Ernestus, Baayen, & Schreuder, 2002; Kemps, Ernestus, Schreuder, & Baayen, 2004) showed that listeners can recognize forms such as [ɪfal] in a sentence context. A purely prelexical mechanism would in this case have to restore seven deleted phonemes on the basis of an input of only four phonemes, which seems impossible. Hence, lexical storage of such strongly reduced forms seems unavoidable.

Second, Hawkins (2003) provides the example of *I do not know*, in which different pronunciations have different discourse implications. In such cases, prelexical compensation could have disastrous consequences, as the listener might then be unable to distinguish different forms, leading to inappropriate responses in the discourse context. Third, evidence from variation concerning flapping in American English (both for stops as in, e.g., pretty, pronounced as [prɪtɪ] or [prɪrɪ], Connine, 2004; McLennan et al., 2003; and for nasals, as in, e.g., gentle, pronounced as [dʒɛntl] or [dʒɛrɪ], Ranbom & Connine, 2007) suggests that multiple pronunciation variants may be stored in the lexicon, perhaps as a function of the frequency of occurrence of those variants (Connine, 2004; Ranbom & Connine, 2007).

It thus seems clear that there can be no unitary account for the way in which listeners recover from the effects of continuous-speech processes. Research is required to examine each case in detail. For instance, in the case of schwa deletion (e.g., whether *president* is pronounced with two or three syllables), it is not yet clear what the locus and nature of the recovery process is (Locasto & Connine, 2002; Utman, Blumstein, & Burton, 2000). On the one hand, schwa deletion in a word such as *president* is likely to be independent of the word's segmental context. This favors a lexical storage account, because prelexical mechanisms might have little contextual leverage on which to operate. On the other hand, Manuel (1992) has shown that, as with many other types of phonological simplifications, schwa deletion is phonetically incomplete and, in fact, is thus better termed schwa *reduction*. Acoustic cues to an underlying schwa could be used by a prelexical compensation mechanism. Considerable work is thus still required to establish what kinds of mechanisms listeners use to recover not only from schwa deletion but also for many other continuous-speech processes in different languages. On the bright side, however, the current article contributes a new method to this enterprise.

A final question is how compensation mechanisms come to be able to help listeners overcome /t/ reduction. Our answer is that the prelexical level must acquire probabilistic knowledge about speech production. In order to learn that /t/ is more likely to be reduced if followed by /b/ than by /n/, the recognition system needs to learn to identify phonetic sequences that are likely to contain reductions. These sequences may be learned largely from the acoustic-phonetic signatures of reduced forms themselves but, perhaps, also through lexically guided learning (Norris, McQueen, & Cutler, 2003), that is, learning which occurs when words with an underlying /t/ have been recognized on the basis of reduced-/t/ input.

What is most striking, however, is that this learning reflects knowledge about the frequency of occurrence of these potential reduction sequences in different contexts. The speech processor must thus keep track of cross-word sequential dependencies and not just within-word transitional probabilities (Lugt, 1999; Pitt & McQueen, 1998; Vitevitch & Luce, 1999) and word-specific production frequencies (Connine, 2004; Ranbom & Connine, 2007). The way that listeners recover from lexical ambiguities, such as the *duck/duct* example we focused on here, and, more generally, from the effects of /t/ reduction (e.g., when *frost* is pronounced as *fros*), is thus through the use of a prelexical compensation mechanism that is sensitive to probabilistic knowledge about where and when speakers drop their /t/s in continuous speech.

References

- Alloppenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, 38, 419–439.
- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database* (Release 2) [CD-ROM]. Philadelphia: University of Pennsylvania, Linguistic Data Consortium.
- Boersma, P., & Weenink, D. (2006). Praat: Doing phonetics by computer (Version 4.4) [Computer program]. Amsterdam: Institute of Phonetic Sciences.
- Bybee, J. (2001). *Phonology and language use*. Cambridge, England: Cambridge University Press.
- Cho, T., & McQueen, J. M. (2006). Phonological versus phonetic cues in native and nonnative listening: Korean and Dutch listeners' perception of Dutch and English consonants. *Journal of the Acoustical Society of America*, 119, 3085–3096.
- Connine, C. M. (2004). It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin and Review*, 11, 1084–1089.
- Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, 16, 507–534.
- Darcy, I., Peperkamp, S., & Dupoux, E. (2007). Bilinguals play by the rules: Perceptual compensation for assimilation in late L2-learners. In J. Cole & J. Hualde (Eds.), *Laboratory phonology: 9* (pp. 411–442). Berlin, Germany: de Gruyter.
- Donselaar, W. v., Kuijpers, C., & Cutler, A. (1999). Facilitatory effects of vowel epenthesis on word processing in Dutch. *Journal of Memory and Language*, 41, 59–77.
- Ernestus, M. (2000). *Voice assimilation and segment reduction in casual Dutch. A corpus-based study of the phonology-phonetics interface*. Dissertation. Utrecht, The Netherlands: Landelijke Onderzoekschool Taalwetenschap (LOT; the Netherlands Graduate School of Linguistics).
- Ernestus, M., Baayen, H. R., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language*, 81, 162–173.
- Fowler, C. A. (1996). Listeners do hear sounds, not tongues. *Journal of the Acoustical Society of America*, 99, 1730–1741.
- Fowler, C. A. (2005). Parsing coarticulated speech in perception: Effects of coarticulation resistance. *Journal of Phonetics*, 33, 199–213.
- Fowler, C. A., Best, C. T., & McRoberts, G. W. (1990). Young infants' perception of liquid coarticulatory influences on following stop consonants. *Perception & Psychophysics*, 48, 559–570.
- Fowler, C. A., & Brown, J. M. (2000). Perceptual parsing of acoustic consequences of velum lowering from information for vowels. *Perception & Psychophysics*, 62, 21–32.

- Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 110–125.
- Gaskell, G. M. (2003). Modelling regressive and progressive effects of assimilation in speech perception. *Journal of Phonetics*, 31, 447–463.
- Gaskell, G. M., & Marslen-Wilson, W. D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 144–158.
- Gaskell, G. M., & Marslen-Wilson, W. D. (1998). Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 380–396.
- Gaskell, G. M., & Marslen-Wilson, W. D. (2001). Lexical ambiguity resolution and spoken word recognition: Bridging the gap. *Journal of Memory and Language*, 44, 325–349.
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105, 251–279.
- Goldstein, L., & Fowler, C. A. (2003). Articulatory phonology: A phonology for public language use. In N. O. Schiller & A. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 159–207). Berlin, Germany: Mouton de Gruyter.
- Gow, D. W. (2002). Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance*, 28, 163–179.
- Gow, D. W. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception & Psychophysics*, 65, 575–590.
- Gow, D. W., & Im, A. M. (2004). A cross-linguistic examination of assimilation context effects. *Journal of Memory and Language*, 51, 279–296.
- Gow, D. W., & McMurray, B. (2007). Word recognition and phonology: The case of English coronal place assimilation. In J. Cole & J. Hualde (Eds.), *Laboratory phonology: 9* (pp. 173–200). Berlin, Germany: de Gruyter.
- Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, 31, 373–405.
- Holt, L. L., & Lotto, A. J. (2002). Behavioral examinations of the level of auditory processing of speech context effects. *Hearing Research*, 167, 156–169.
- Holt, L. L., Lotto, A. L., & Kluender, K. R. (2000). Neighboring spectral content influences vowel identification. *Journal of the Acoustical Society of America*, 108, 710–722.
- Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic, and shape information in language-mediated visual search. *Journal of Memory and Language* 57, 460–482.
- Kemps, R., Ernestus, M., Schreuder, R., & Baayen, H. R. (2004). Processing reduced word forms: The suffix restoration effect. *Brain and Language*, 90, 17–127.
- Kingston, J., & Macmillan, N. A. (1995). Integrality of nasalization and F1 in vowels in isolation and before oral and nasal consonants: A detection-theoretic application of the Garner paradigm. *Journal of the Acoustical Society of America*, 97, 1261–1285.
- Lahiri, A., & Marslen-Wilson, W. D. (1991). The mental representation of lexical form: A phonological approach to the lexicon. *Cognition*, 38, 245–294.
- Lahiri, A., & Reetz, H. (2002). Underspecified recognition. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (pp. 637–676). Berlin, Germany: Mouton de Gruyter.
- Liberman, A. M. (1996). *Speech: A special code*. Cambridge, MA: MIT Press.
- Lindblom, B., & Studdert-Kennedy, M. (1967). On the role of formant transitions in vowel recognition. *Journal of the Acoustical Society of America*, 43, 830–843.
- Locasto, P. C., & Connine, C. M. (2002). Rule-governed missing information in spoken word recognition: Schwa vowel deletion. *Perception & Psychophysics*, 64, 208–219.
- Lotto, A. J., & Kluender, K. R. (1998). General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *Perception & Psychophysics*, 60, 602–619.
- Lotto, A. J., Kluender, K. R., & Holt, L. L. (1997). Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*). *Journal of the Acoustical Society of America*, 102, 1134–1140.
- Lugt, A. van der (1999). The use of sequential probabilities in the segmentation of speech. *Perception & Psychophysics*, 63, 811–823.
- Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics*, 28, 407–412.
- Mann, V. A., & Repp, B. H. (1980). Influence of vocalic context on perception of the [j]-[s] distinction. *Perception & Psychophysics*, 28, 213–228.
- Mann, V. A., & Repp, B. H. (1981). Influence of preceding fricative on stop-consonant perception. *Journal of the Acoustical Society of America*, 69, 548–558.
- Mann, V. A., & Soli, S. D. (1991). Perceptual order and the effect of vocalic context of fricative perception. *Perception & Psychophysics*, 49, 399–411.
- Manuel, S. Y. (1992). Recovery of “deleted” schwa. In *Perilus: Papers from the Symposium on Current Phonetic Research Paradigms for Speech Motor Control* (pp. 115–118). Stockholm: University of Stockholm.
- McLennan, C. T., Luce, P. A., & Charles-Luce, J. (2003). Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 539–553.
- McQueen, J. M., & Viebahn, M. (2007). Tracking recognition of spoken words by tracking looks to printed words. *Quarterly Journal of Experimental Psychology*, 60, 661–671.
- Mitterer, H. (2003). *Understanding “gardem bench”: Studies on the perception of assimilation word forms* [dissertation]. Maastricht, The Netherlands: Universiteit Maastricht.
- Mitterer, H. (2006a). Is vowel normalization independent of lexical processing? *Phonetica*, 63, 209–229.
- Mitterer, H. (2006b). On the causes of compensation for coarticulation: Evidence for phonological mediation. *Perception & Psychophysics*, 68, 1227–1240.
- Mitterer, H., & Blomert, L. (2003). Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception & Psychophysics*, 65(6), 956–969.
- Mitterer, H., Csépe, V., & Blomert, L. (2006). The role of perceptual integration in the perception of assimilated word forms. *Quarterly Journal of Experimental Psychology*, 59, 1305–1334.
- Mitterer, H., Csépe, V., Honbolygo, F., & Blomert, L. (2006). The recognition of assimilated word forms does not depend on specific language experience. *Cognitive Science*, 30, 451–479.
- Mitterer, H., & Ernestus, M. (2006). Listeners recover /t/s that speakers lenite: Evidence from /t/-lenition in Dutch. *Journal of Phonetics*, 34, 73–103.
- Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition*, 7, 323–331.
- Nearey, T. M. (1997). Speech perception as pattern recognition. *Journal of the Acoustical Society of America*, 101, 3241–3254.
- Nolan, F. (1992). The descriptive role of segments: Evidence from assimilation. In G. Docherty & D. R. Ladd (Eds.), *Laboratory Phonology II* (pp. 261–280). Cambridge, England: Cambridge University Press.
- Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47, 204–238.
- Pitt, M. A., & McQueen, J. M. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language*, 39, 347–370.
- Ranbom, J. L., & Connine, C. M. (2007). Lexical representation of pho-

- nological variation in spoken word recognition. *Journal of Memory and Language*, 57, 273–298.
- Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, 90, 51–89.
- Shatzman, K. B., & McQueen, J. M. (2006). Prosodic knowledge affects the recognition of newly acquired words. *Psychological Science*, 17, 372–377.
- Smits, R. (2001a). Evidence for hierarchical categorization of coarticulated phonemes. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 1145–1162.
- Smits, R. (2001b). Hierarchical categorization of coarticulated phonemes: A theoretical analysis. *Perception & Psychophysics*, 63, 1109–1139.
- Sumner, M., & Samuel, A. G. (2005). Perception and representation of regular variation: The case of final /t/. *Journal of Memory and Language*, 52, 322–338.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632–1634.
- Utman, J. A., Blumstein, S. E., & Burton, M. W. (2000). Effects of subphonetic and syllable structure variation on word recognition. *Perception and Psychophysics*, 62, 1297–1311.
- Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and spoken word recognition. *Journal of Memory and Language*, 40, 374–408.
- Whalen, D. H., & Liberman, A. M. (1987). Speech perception takes precedence over nonspeech perception. *Science*, 237, 169–171.

Appendix A

/Ct#/-C#/ Minimal Pairs

Ct# words	Gloss	Log fr/10 ⁶	No. of /t/ - reductions (2, maximum)	C# words	Gloss	Log fr/10 ⁶	Delta fr
barst	crack	0.95	0	bars	perch	0.63	0.32
bast	bast	0.46	2	bas	bass	0.32	0.14
best	best	2.04	1	bes	berry	0.91	1.13
brnst	rutting season	0	0	brns	bronze	0.46	-0.46
bult	bump	0.63	2	bul	certificate	0.3	0.33
buurt	neighborhood	2.04	2	buur	neighbor	1.44	0.6
gast	guest	1.76	1	gas	gas	1.39	0.37
gift	gift	0.79	2	gif	poison	0.71	0.08
gist	yeast	0.43	1	gis	guess	0	0.43
haast	hurry	1.98	0	haas	hare	0.97	1.01
halt	stop	1.02	2	hal	hall	1.48	-0.46
hecht	firm	1.12	0	heg	hedge	0.77	0.35
hoest	cough	0.43	2	hoes	cover	0.39	0.04
hulst	casing	0.41	0	huls	shell	0.68	-0.27
kaft	cover	0.33	2	kaf	chaff	0.03	0.3
kast	cupboard	1.68	2	kas	greenhouse	1.1	0.58
kerst	Christmas	0.57	1	kers	cherry	0.7	-0.13
kroost	offspring	0.5	0	kroos	duckweed	0	0.5
kust	coast	1.72	2	kus	kiss	1.27	0.45
last	burden	1.86	0	las	joint	0	1.86
lust	lust	1.35	0	lus	loop	0.75	0.6
mest	fertilizer	0.71	0	mes	knife	1.62	-0.91
mist	mist	1.06	2	mis	miss, error	1.63	-0.57
rijst	rice	0.76	1	reis	journey	1.99	-1.23
roest	rust	0.29	1	roes	high (N.)	0.92	-0.63
rust	rest	1.94	1	Rus	Russian	1.62	0.32
schort	apron	0.8	1	schor	coarse	1.03	-0.23
start	start	1.04	2	star	rigid	1.08	-0.04
tast	sense of touch	0.58	0	tas	bag	1.59	-1.01
vaart	pace	1.39	0	vaar	barren	0.31	1.08
vlucht	flight	1.76	0	vlug	quick	1.99	-0.23
wicht	little child	0.33	0	wig	wedge	0.27	0.06
Mean		1.022813				0.885938	0.136875
SD		0.617713				0.591176	0.657828

/n#/-m#/ Minimal Pairs

/n/-word	Gloss	Log fr/10 ⁶	/m/-word	Gloss	Log fr/10 ⁶	Delta fr
bon	receipt	0.96	bom	bomb	1.36	-0.4
boon	bean	1.19	boom	tree	2.14	-0.95
bron	source	1.81	brom	buzz	0	1.81
don	Don	1.1	dom	stupid	1.71	-0.61
duin	dune	1.2	duim	thumb	1.49	-0.29
kan	jug	0.36	kam	comb	0.91	-0.55

Appendix A (*continued*)

/n/-word	Gloss	Log fr/10 ⁶	/m/-word	Gloss	Log fr/10 ⁶	Delta fr
kin	chin	1.52	kim	rim	1.26	0.26
kraan	crane	1.18	kraam	market stall	0.65	0.53
kruin	(tree)crown	0.96	kruim	crumb	0	0.96
leen	loan	0.51	leem	loam	0.39	0.12
lijn	line	2.02	lijm	glue	0.83	1.19
loon	wage	1.43	loom	sluggish	0.92	0.51
ruin	rune	0.15	ruim	spacious	2.05	-1.9
run	run	0	rum	rum	0.94	-0.94
schuin	slant	1.51	schuim	foam	1	0.51
zoon	son	2.28	zoom	hem	0.56	1.72
Mean		1.13625			1.013125	0.123125
SD		0.645351			0.636446	1.017433