Journal of Phonetics 34 (2006) 73–103

Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch

Holger Mitterer*, Mirjam Ernestus

Max-Planck Institute for Psycholinguistics, P.O. Box 310, NL-6500 AH Nijmegen, The Netherlands

Received 28 June 2004; received in revised form 11 March 2005; accepted 21 March 2005

Abstract

In everyday speech, words may be reduced. Little is known about the consequences of such reductions for spoken word comprehension. This study investigated /t/-lenition in Dutch in two corpus studies and three perceptual experiments. The production studies revealed that /t/-lenition is most likely to occur after [s] and before bilabial consonants. The perception experiments showed that listeners take into account both phonological context, phonetic detail, and the lexical status of the form in the interpretation of codas that may or may not contain a lenited word-final /t/. These results speak against models of word recognition that make hard decisions on a prelexical level.

1. Introduction

Science has divided the recognition of spoken words into two disciplines. Thus, in the Handbook of Psycholinguistics (Gernsbacher, 1994), there are two chapters dealing with the comprehension of speech. One (Kluender, 1994) is dedicated to the perception of acoustic cues for phonological units, either phonological features, phonemes, or syllables. The other (Lively, Pisoni, & Goldinger, 1994) deals with the recognition of words given these phonological units. This divide is mirrored in models of spoken-word recognition. In most models (e.g., Lahiri & Reetz, 2002; Luce, Goldinger, Auer, & Vitevitch, 2000; McClelland & Elman, 1986; Norris,

*Corresponding author. Tel.: +31 24 3521372.
E-mail addresses: holger.mitterer@mpi.nl (H. Mitterer), mirjam.ernestus@mpi.nl (M. Ernestus).

McQueen, Cutler, & Butterfield, 1997), a first stage makes decisions about what phonemes or features are in the acoustic input, and a second stage looks up which words fit with this transformation of the acoustic input. The divide is also evident in the type of work that is done in different scientific fields of inquiry: On the one hand, phonetics investigates the production and perception of phonological units (e.g., Stevens, 1989, 1998) and their interaction in co-articulation (e.g., Farnetani, 1997; Mann, 1980). On the other hand, psycholinguistics classically intended to investigate the structure of the lexicon and to explain frequency and neighborhood effects, which could be well modeled given a "mock-input" of phonemes or features (Baayen, McQueen, Dijkstra, & Schreuder, 2003; Goldinger, Luce, & Pisoni, 1989; Pisoni, Nusbaum, Luce, & Slowiaczek, 1985).

The 'great divide' may to some extend be a consequence of the speech materials that were under investigation in both fields. Mostly, these materials consisted of recordings of speakers reading lists of words or nonwords. The speakers were well aware of the fact that they were recorded and may therefore have aimed for "hyper-articulated" speech (Lindblom, 1990). For this kind of speech, the great divide seems to work well. Words are produced in citation form, and variability is mostly caused by variables such as vocal-tract length and (mild) coarticulation. Phonetic processing has been shown to be able to compensate for these sources of variability (see, e.g., Fowler, Brown, & Mann, 2000; Ladefoged, 1989; Ladefoged & Broadbent, 1957; Lotto & Kluender, 1998; Mann, 1980; Mann & Repp, 1981; Watkins & Makin, 1996) and can extract an invariant phonological code for the same intended word. This code is then delivered to the lexicon. Psycholinguistic models of word recognition account for effects of frequency and phonological neighborhood (Baayen et al., 2003; Goldinger et al., 1989; Pisoni et al., 1985).

However, the 'great divide' is challenged by observations on casual speech: In casual speech, words may be produced in significantly different phonological shapes than in isolation (e.g., Bybee & Scheibman, 1999; Ernestus, 2000; Johnson, 2004). Johnson, for instance, concluded that over 60% of the words in his corpus of casual speech deviate from their citation form on at least one phone, and 28% of the words deviate on two or more phones. Recovering the citation form with such input is a tremendous challenge for phonetic normalization processes. "Wrong" decisions are likely to occur at a phonetic level of processing. In such cases, it is difficult to see how lexical processing could recover the intended words. Some evidence suggests that semantic, syntactic, and pragmatic constraints may compensate for "wrong" phonetic decisions (Cole, 1973; Marslen-Wilson, 1987; Warren, 1970). However, recent evidence suggests that signal-based constraints take precedence over higher-level constraints (Dahan & Tannehaus, 2004; Utman, Blumstein, & Burton, 2000), so that higher-level constraints may not suffice to achieve recognition of reduced forms, which are likely to induce "wrong" decisions on a prelexical level.

Given this state of affairs, processes in speech production that change the phonemic surface form of words present a challenge to our understanding of spoken-word recognition. In this paper, we aim to contribute to an understanding of how reduced forms are recognized.

At least two processes that change the phonemic surface form have already been the focus of some attention: phonological assimilations and extremely reduced forms. Phonological assimilation, such as coronal place assimilation in, for instance, English, changes the place of articulation of word final alveolar nasals and stops. For instance, the phrase 'garden bench' may be pronounced as [gɑrdmbɛntʃ]. Extremely reduced word forms occur in fluent speech for high-frequency words. A well-documented case of such reductions are the Dutch words eigenlijk

/ɛixələk/ 'actually' and natuurlijk /natyrlək/ 'naturally'. Ernestus (2000) showed that these forms may be pronounced as [ɛik] and [tyk] in casual speech. (All phonetic transcription in this article are broad transcriptions.)

The recognition of assimilated words forms has been explained by the assumption that lexical representations are underspecified for features that may change due to assimilation (Lahiri & Marslen-Wilson, 1991; Lahiri & Reetz, 2002). Hence, the English word 'garden' is specified as /gɑrd + (nasal)/. This representation fits with the citation form [gɑrdn] as well as with the assimilated form [gɑrdm], as in 'garden bench,' because both contain [gɑrd] followed by a nasal consonant. Theoretical and empirical arguments can be put forth against underspecification. Obviously, underspecification of lexical representations leads to a substantial increase of competition between candidate words, because an input such as [rʌm] would activate both 'run' and 'rum.' Moreover, a growing body of evidence shows that listeners' sensitivity to segmental context and phonetic detail allows assimilated word forms to be recognized without the burden of lexical underspecification. A form such as [gɑrdm] occurs only if it is followed by a bilabial consonant; hence it may occur in the composite noun [gɑrdmbɛntʃ] 'garden bench' but not in *[gɑrdmtʃeir] 'garden chair'. Listeners make use of this regularity and recognize 'garden' in [gɑrdmbɛntʃ] but not in *[gɑrdmtʃeir] (Gaskell & Marslen-Wilson, 1996, 1998, 2001; Gow, 2002, 2003; Manuel, 1995). Mitterer and Blomert (2003) showed that this context effect is rooted in early, automatic perceptual processes: They measured the MisMatch Negativity (MMN), a component of the auditory evoked potentials (for reviews, see Näätänen & Winkler, 1999; Picton, Alain, Otten, Ritter, & Achim, 2000) for phonologically viable and unviable assimilations. An MMN can be observed when participants are confronted with a train of stimuli while reading or watching a silent movie. The train of stimuli is composed of two stimuli. One is called *standard*, because it accounts for 85–95% of all stimuli in the train of stimuli and a *deviant*, which only account for 15–5% (more complex designs are possible, see Näätänen, Tervaniemi, Sussman, Paavilainen, & Winkler, 2001). The MMN depends on the ability of the listeners to detect a difference between standard and deviant (Näätänen, Schröger, Karakas, Tervaniemi, & Paavilainen, 1993). Mitterer and Blomert found no MMN for a standard-deviant pair that consisted of a citation pronunciation of a Dutch word ([tœynbɑŋk] tuinbank 'garden bench') as standard and a viable assimilation as deviant ([tœymbɑŋk]). However, an MMN was observed when the deviant ([tœymstul]) was an unviable assimilation of the standard ([tœynstul] tuinstoel 'garden chair'). Given that the MMN was observed even though participants were not actively listening, and since the MMN occurred directly after the segment licensing or not licensing the assimilation—[…stul] or […bɑŋk]—, one may conclude that an /m/ that is followed by a /b/ is treated as not differing from an /n/ in early automatic perception.

Besides phonological context, sub-phonemic details also influence the perception of assimilated forms: Gow (2002) showed that assimilated forms, such as [rɑipbɛriz], which the speaker intended as 'right berries,' differ acoustically from [rɑipbɛriz] in which the /p/ is intended ('ripe berries'). In a nutshell, the word-final stop that is the result of assimilation in 'right berries' is a less clear instance of /p/ than the final stop in 'ripe berries'. These sub-phonemic cues affect recognition: If confronted with weak cues for a [p] in [rɑipberiz], listeners access the meaning "right," but "right" is not accessed if there are strong cues for [p] in [rɑipberiz] (see also Mitterer, Csépe, & Blomert, 2003).

Additional evidence shows that assimilation is undone before lexical access is achieved. Mitterer and Blomert (2003) showed that German listeners compensate for assimilation in a Dutch word that is unlikely to be perceived as an existing German word. While this results may still be explained by learning of assimilation rules—German has the same nasal place assimilation as Dutch—, the results of Mitterer et al. (2003) show that Dutch listener compensate for a Hungarian assimilation rule, that has no counterpart in Dutch.

Different results have been obtained for highly reduced forms, another type of casual-speech reduction that has previously been investigated. Highly reduced word-forms are not restored on the basis of the acoustic signal alone. Ernestus, Baayen, and Schreuder (2002) investigated the recognition of extremely reduced forms in Dutch, such as [ifɑl] for /ɪn idər xəfɑl/ in ieder geval 'in any case' and [ɛik] for /ɛixələk/ eigenlijk 'actually.' They showed that listeners need the sentence context in order to understand such forms: In isolation, these forms were hardly recognized as meaning 'in any case' or 'actually.' If presented with a minimal phonetic context, recognition performance improved, but did not reach ceiling level. This was only achieved if a whole sentence was presented. These results show that semantic context is necessary for the recognition of highly reduced word forms. (However, the study does not address the question of whether semantic context alone is sufficient.)

Kemps, Ernestus, Schreuder, and Baayen (2004) showed that listeners recognize extremely reduced forms as being presented in their full form: Listeners were asked to monitor for /l/. They often responded—wrongly—to reduced forms such as [ɛik] (with the unreduced variant /ɛixələk/), but only if the reduced form was presented in a sentence context. Thus, the reduced form [ɛik] appears to activate the full phonological form /ɛixələk/or the orthographic form eigenlijk, which both justify an "l"-response. On the basis of several control experiments, Kemps et al. could determine that this effect is only partly due to an orthographic influence on phoneme monitoring. The reduced form, given the right context, activates above all the phonological full form, which then influences phoneme-monitoring performance. In summary, these two studies show that extremely reduced forms are recognized only if presented in phrases of several words, which provide semantic context (Ernestus et al.), and that recognition involves the activation of the phonological full form (Kemps et al.).

Crucially, phonological assimilations and extremely reduced forms are recognized through quite different mechanisms: Reduced forms are reconstructed using lexical knowledge and sentence context, while assimilation is undone at least partly prelexically (Gaskell & Marslen-Wilson, 1998,2001; Mitterer & Blomert, 2003). This is to some extent not surprising as the two connected-speech processes may be viewed as endpoints on a continuum of articulatory simplifications that occur in casual speech. Extreme reductions seem to be limited to a small number of high-frequency words (at least in Dutch, Ernestus, 2000), while assimilation is a general process that may apply to any word. Moreover, extreme reductions differ drastically and in a partly unpredictable way from the corresponding phonological full forms. In contrast, assimilated forms have a systematic relation to the citation form. In addition, assimilations are incomplete and subtle acoustic cues are consequently helpful in recognizing assimilated forms (Gow, 2002, 2003).

Assimilation and extreme reduction are not the only documented articulatory simplifications that change the phonemic surface form of a word. Reduction of single segments, which may result in "apparent" deletion, has also been studied. Manuel (1992) investigated the perception of words in which schwa has been deleted (support /səpɔːt/→[spɔːt]). Production data of this phenomenon indicated that, while the glottal gesture for the schwa was reduced so that no vocal fold vibration occurred, the oral gesture often remained, leading to subtle acoustic differences between the forms of [spɔːt] meaning either *support* or *sport*. These acoustic differences were picked up by listeners to disambiguate forms of [spɔːt] (see also Racine & Grosjean, 2000).

Another well-studied example of articulatory simplifications is the variation in the production of word-final /t/. In some variants of English, word-final /t/ in postvocalic position is commonly pronounced unreleased or as a glottal stop, even in single-word sentences and in careful speech. Sumner and Samuels (2005) showed in a priming experiment that all variants of /t/ equally and effectively prime a semantically related target.

A different form of word-final /t/ variation is /t/-lenition in Standard Dutch. In Dutch, /t/ is generally released in careful speech. In less formal registers, however, word-final /t/ may be lenited, especially in coda clusters, even up to a degree of complete absence, that is /t/-deletion. Words such as /mɪst/ mist 'mist' may hence be pronounced as [mɪs] (e.g., Ernestus, 2000; Goeman, 1999). This form of /t/-variation occurs in similar form in other Germanic languages as well (Guy, 1980, 1992; Grimson & Cruttenden, 1994, for English and Kohler, 1995, for German). /t/-deletion falls between assimilation and extreme reduction, in terms of the degree of "deformation" of the citation form: The number of features of the citation form present in the variant decreases from assimilations to deletions to extreme reductions. Moreover, Goeman (1999) reported that, for the Dutch dialects, /t/-deletion is not highly frequent in monomorphemic words. Similar to other sentence-level processes, /t/-deletion is optional (see Kaisse, 1985, for a general overview).

To the best of our knowledge, no published studies yet investigated how words with /t/-deletions or lenited /t/s are recognized. Based on previous work on assimilation and extremely reduced forms, the question arises whether listeners overcome the absence of a clear /t/ prelexically or lexically. We addressed this issue in a series of experiments and corpus studies in this paper.

We restricted ourselves to /t/-lenitions in nominal forms, since from the perspective of spoken-word comprehension, /t/-lenition within mono-morphemic words is more of a problem than deletion of morphemic /t/. Deletion of /t/ in the Dutch mono-morphemic word <u>kast</u> /kɑst/ 'cupboard', for example, leads to [kɑs], the Dutch word for 'green house', whereas deletion of the morphemic /t/in the 3rd person singular present tense <u>botst</u> /bɔtst/ 'collides' leads to [bɔts], the 1st person singular present tense form. Obviously, /t/-deletion that results in a completely different word, as in the case of *kast*, is a more serious problem for word comprehension than /t/-deletion that results in a different form of the same lemma.

Lenition of /t/ may only challenge word comprehension if it adds to the variability in the speech signal. If, in a given dialect of Dutch, word-final /t/ is always deleted, then the citation form /kast/ can be unambiguously recovered from the reduced form [kas]. The word *kast* will always be realized as [kas], and there will be no invariance problem. We thus focus on /t/-deletion in mono-morphemic words in Standard Dutch, in which /t/-deletion is not obligatory. In Standard Dutch listeners encounter both [kas] and [kast] for *kast*. How do they process forms such as [kas] or [kast] with a lenited /t/?

In order to answer this question, we first have to establish with what kind of forms the listener is actually confronted. That is, is any word-final /t/ likely to be lenited, or do certain phonological contexts facilitate /t/-lenition? From a survey of phoneme realizations in Dutch, Ernestus (2000) speculated that /t/ is most likely to be deleted after /s/ and before bilabial consonants in casual speech (see also Jongenburger & van Heuven, 1993, for more formal Dutch). Ernestus only performed a qualitative analysis, so that the exact likelihood of the absence of /t/ in different segmental contexts is yet unknown. Furthermore, Ernestus based the presence or absence of /t/ in the acoustic signal on impressionistic phonetic-expert transcriptions. As such, the production data may be contaminated by perceptual processes.

We first investigated in which circumstances /t/ is likely to be lenited in Standard Dutch. We investigated word-final /t/ in read stories and in spontaneous casual speech. These corpus studies confirmed that /t/-lenition is a variable phenomenon that occurs in Standard Dutch. Hence, it constitutes part of the invariance problem in speech perception. We then conducted three perceptual experiments in order to investigate how listeners overcome lenited /t/s.

2. Corpus study 1

In our first corpus study, we investigated the realization of word-final /t/ in read stories in the corpus of spoken Dutch (Corpus Gesproken Nederlands, CGN). We investigated word-final /t/ after /n/ as in kant /kɑnt/ 'side,' after /s/ as in kast /kɑst/ 'cupboard,' and after /x/ as in bocht /bɔxt/ 'turn' (noun). Given previous reports on /t/-deletion in different contexts (e.g., Guy, 1980, for English; and Ernestus, 2000; Jongenburger & van Heuven, 1993 for Dutch), we expected that the likelihood of /t/ lenitions should be lowest after /n/, higher after /x/, and maximal after /s/.

In addition, Jongenburger and van Heuven (1993) and Ernestus (2000) argued that the following context also influences /t/-deletion. Deletion is more likely in consonant clusters, and according to Ernestus, especially before /b/. Therefore, we also took the following segmental context of the /t/ into account: We investigated /t/-lenition in the six different phonological contexts that arise by combining all the three preceding phonological contexts /s/, /n/, and /x/ with the two following contexts Obstruent versus Vowel.

2.1. Method

2.1.1. Materials

We have drawn our materials form the Corpus of Spoken Dutch (http://lands.let.kun.nl/cgn/ehome.htm). This corpus contains almost 9 million words, produced, among others, in face-to-face conversations, read aloud stories, interviews, telephone conversations, meetings, talks and radio programs. All words in the corpus have been orthographically transcribed, and approximately 10% of the corpus has been transcribed phonemically. Part of the corpus has received syntactic tagging.

We selected all words in CELEX (Baayen, Piepenbrock, & Gulikers, 1995) that end in /nt/, /xt/, or /st/, are (potentially) monomorphemic, consist of no more than three syllables, and, according to CELEX, have a token frequency of more than 1 per million and less than 1000 per million. From the subcorpus "library for the blind" of the CGN corpus, which consists of the read aloud stories, we selected all utterances containing these words. We disregarded utterances in which the following phoneme was an alveolar stop /t/ or /d/, since word-final /t/ cannot be well distinguished from the following stop in these utterances. In addition, we disregarded utterances in which the word was directly followed by a clause boundary or a clause-internal pause. Previous work on fricative assimilation (Holst & Nolan, 1995; Nespor & Vogel, 1986) indicated that sandhi-processes are less likely to occur across higher prosodic boundaries. From this initial data set, we then made a random selection of 50 utterances, representing ten to twenty different words, for each of the six phonological context conditions. These 300 utterances were produced by 187 different speakers (86 male and 101 female). Most speakers ($n = 110$) contributed only one token, and no speaker contributed more than 6 tokens. (The complete list of the utterances can be obtained from the first author.)

2.2. Results and discussion

2.2.1. Qualitative analysis

Previous work indicated that sandhi-processes may not only be variable in the sense of being optional but also gradient in the sense that they are not complete (e.g., Gow, 2003; Holst & Nolan, 1995; Nolan, 1992; Zsiga, 1995). Browman and Goldstein (1990, pages 360–365) discussed X-ray data tracking the positions of small lead pellets placed on the lips and the tongue of participants producing the sentence "perfect memory". The /t/ in this phrase was sometimes inaudible, suggesting /t/-deletion. The X-ray data, however, showed that the gesture of the /t/ was always present, but that it sometimes overlapped with the lip closure for the following bilabial nasal. This overlap may be absent, incomplete, or complete, resulting in clear /t/s, lenited /t/s, and deleted /t/s.

We also found lenited /t/s in our data set. According to Stevens (1998, page 348), a voiceless unaspirated stop as realized in careful speech contains a closure, a transient, and frication noise. An example of such a /t/ is depicted in Fig. 1A. Fig. 1B shows a less straightforward example that occurred in our data set. The speaker intended to say kast gegooid /kɑst xəxojt/ 'cupboard thrown.' In the obstruent cluster between the vowels /ɑ/ and /ə/, there are three discernable events, indicated by three local maximum amplitudes separated by two local minima. There is a clear acoustic correlate of the alveolar fricative /s/ after /ɑ/: The spectral center of gravity of the frication maximum—see arrow—directly after /ɑ/ is 4267 Hz, indicating alveolar frication. There is also a clear acoustic correlate of the velar fricative /x/ before /ə/, in the form of energy in the regions of the second, third, and fourth formant and a center of gravity at 1635 Hz. Between these two maxima, there is neither a closure nor a transient signal, which would classically be indicative of a /t/. There is, however, an /s/-like frication, with a center of gravity at 4700 Hz, which is slightly higher than the center of gravity of the preceding /s/. A likely interpretation of this pattern is that this "middle hump" of frication noise is in fact a residue of /t/, without closure and transient but with frication noise. In Fig. 1B, this third event is transcribed as [t̞].

Another pattern that was often encountered is illustrated in Fig. 1C. The speaker intended to say kast weer /kɑst ʋeːr/ 'cupboard again.' The alveolar fricative /s/ after /ɑ/ fades out, and we observe 72 ms of low-amplitude frication noise. This frication is somewhat lower in terms of spectral center of gravity (3120 Hz) than the frication in Fig. 1B, due to the slow introduction of low-frequency components. Though this low-amplitude frication is hardly noticeable, it is most likely a signature of an underlying /t/ (see the arrow in Fig. 1C).

A fourth type of realization that we encountered is shown in Fig. 1D. The speaker intended to say mest besteld /mɛst bəstɛlt/ 'fertilizer ordered.' We observe a rather long stretch of silence, or at

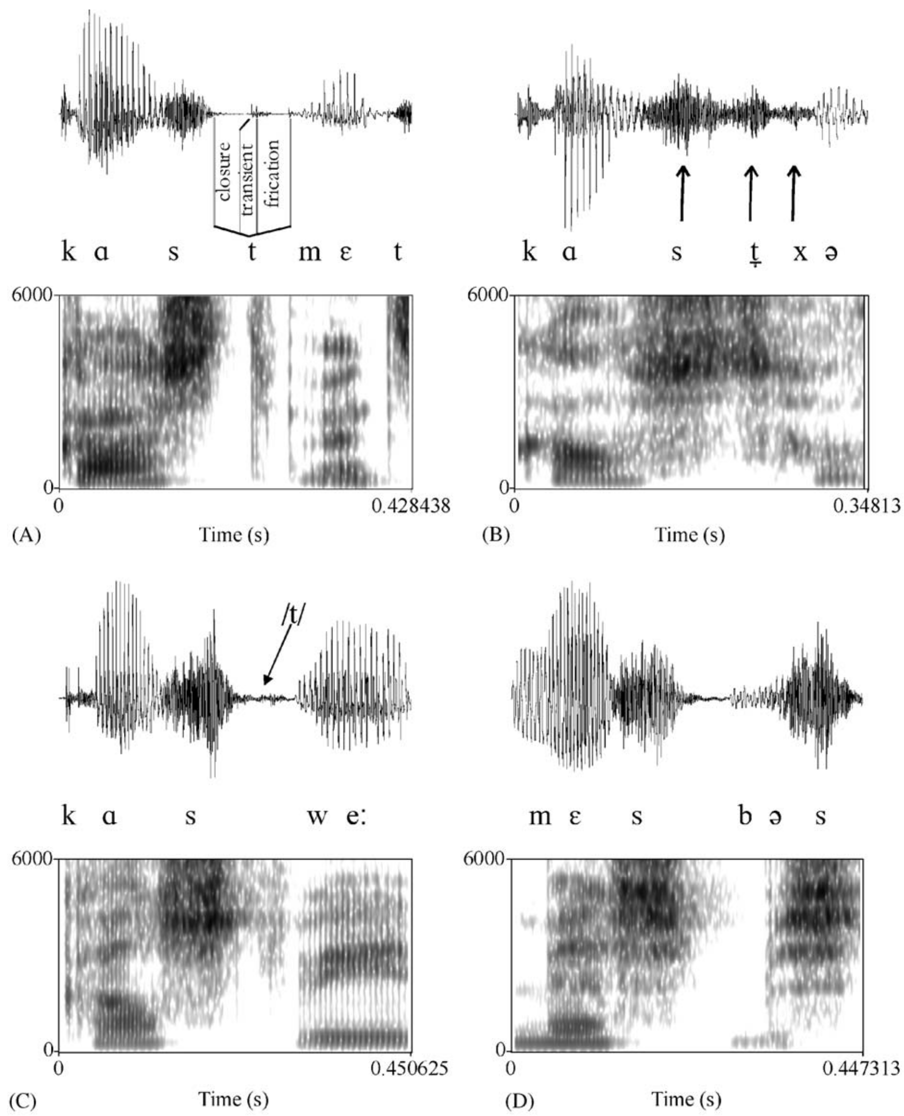


Fig. 1. Realizations of word-final /t/ in four Dutch casual-speech utterances from full production (Fig. 1A) to deletion (Fig. 1D). See text for details.

least low-amplitude frication, of more than 50 ms. This silence is followed by the prevoicing of the /b/. The silence and prevoicing add up to approximately 100 ms. This would be an extraordinary long closure for /b/ in running speech. Therefore, we may assume that the long closure consists of both the closures of the /t/ and the /b/.

2.2.2. *Quantitative analysis*

The qualitative analysis indicated that the attested realizations are not well described by the inherently dichotomous transcription as "realized" versus "deleted" /t/. In order to perform a quantitative analysis, however, we had to quantify the different realizations according to some criterion. Since the different /t/-realizations do not seem to form a natural continuum, a quantification on a continuous scale would be difficult, if not impossible. We therefore nevertheless dichotomized the realizations: /t/ was assumed to have either strong or weak cues, henceforth to be either present or absent, on the basis of an amplitude criterion. A /t/ was labeled as present if there was a part of the speech signal with an appreciable amplitude, that is, comparable to the surrounding segments, that could be attributed to an underlying /t/. Accordingly, we classified realizations such as Examples 1A and B as containing /t/, whereas Examples 1C and D were transcribed as deleted /t/s. Although such a categorization is rather simplistic and may lead to an unduly data reduction, it is more reliable than a more elaborate classification scheme (cf. Bornstein, 1987). An alternative possibility would be to rely on impressionistic phonetic transcriptions of the realizations. However, as noted above, impressionistic transcriptions may be contaminated by perceptual processes, which do not always faithfully reflect the characteristics of the acoustic signal (see, e.g., Deutsch, 1992; Mitterer & Blomert, 2003).

The 300 utterances in this corpus study were examined by a research assistant unaware of the purposes of the study, who classified the /t/-sound as present or not on the basis of the spectrograms of the stimuli. Table 1 shows the proportion of deleted /t/s in the various contexts. An ANOVA revealed that the proportion of /t/-deletion was affected by both Preceding Context ($F(2,294) = 9.10$, $p < 0.001$) and the Following Context ($F(1,294) = 11.70$, $p < 0.01$). The interaction was not significant ($F < 1$). A post hoc test (HSD) for the effect of Preceding Context showed that /t/ deletion was more likely after /s/ than after /n/ and /x/ ($p < 0.05$), while these latter contexts did not differ significantly. With respect to following context, /t/ was more often absent before obstruents than before vowels.

Table 2 shows the number of deletions and observations for each of the following obstruents. Given the small number of observations in some cells, strong claims are unwarranted. The small number of /t/-deletion before vowels may be due to resyllabification, which makes the /t/ syllable-initial, and hence less likely to be lenited. Nevertheless, it is noteworthy that /t/-deletion in this data set is most frequent preceding the voiced bilabial stop, which is in line with the observations by Ernestus (2000) on casual speech. Ernestus conjectured that lack of evidence for a /t/ before labials may result from the realization of the labial closing gesture before the /t/ release, as it may in English (Browman & Goldstein, 1990).

Note that /t/-lenition is not mandatory in any phonological context in Dutch. For instance, Example 1A, which appeared in the context kast met /kɑst mɛt/ ‘cupboard with’ is, according to

Table 1
The proportion of /t/-deletion in read speech in various phonological contexts (50 observations per cell)

Following context	Preceding context			Mean
	/s/	/n/	/x/	
Obstruent	0.26	0.06	0.10	0.14
Vowel	0.10	0.00	0.00	0.03
Mean	0.18	0.03	0.05	0.08

Table 2
The number of deleted /t/s (left of the forward slash) and total number of observations (right of the forward slash) for each combination of preceding segment and following obstruent

Previous Context	Following context							
	/b/	/x/	/h/	/k/	/p/	/s/	/v/	/z/
/x/	3/5	0/4	0/8	1/2	0/1	0	1/20	0/6
/n/	2/5	0/3	0/4	0/1	0/3	0	1/29	0/4
/s/	5/5	2/6	1/8	0/1	0	0/3	4/26	1/1
∅/N	0.67	0.15	0.06	0.25	0.00	0.00	0.08	0.09

The last row summarizes the proportion of /t/-deletion per type of following obstruent (N = number of observations for each following obstruent).

our findings, very likely to give rise to /t/-lenition. Nevertheless, /t/ is produced in its full form, with visible acoustic correlates of closure, transient, and frication. It is important to note that, paradoxically, from the point of view of perception, infrequent, optional /t/-deletion is more detrimental than mandatory /t/-deletion. Mandatory /t/-deletion is easier to overcome, given its high predictability, while infrequent /t/-deletion adds to the invariance problem.

3. Corpus study 2

As a next step, we investigated /t/-lenition in the subcorpus "casual speech" of the Corpus of Spoken Dutch, whereas the previous corpus study investigated read speech. Given the differences between casual and read speech, we felt compelled to change our sampling method. Repetitions of the same token may lead to successively more reduction (Fowler, 1988; Fowler & Housum, 1987; but see also Hawkins & Warren, 1994). Because of the greater need of speech planning in casual, self-generated, speech than in read speech (Levelt & Indefrey, 2004), this repetition is very likely to induce reduction in casual speech. Therefore, we added repetition of the same token by the same speaker as an independent variable.

3.1. Method

3.1.1. Materials

We selected the same /t/-final words as in Corpus Study 1. In order to evaluate the effect of word repetition, we restricted the data set to words that were produced at least five times by a given speaker. There were 30 word types that had each been uttered five times per speaker (so 150 word tokens). These 30 quintets were produced by 24 different speakers, (11 males and 13 females), with one speaker producing three quintets, four speakers producing two quintets, and 19 speakers one quintet. All utterances were labeled as containing a /t/ or not, along the lines discussed above. Of the 150 utterances, 25 had to be discarded due to background noises, such as the other speaker(s) speaking simultaneously, or because the critical word was followed by a clause boundary. (The complete list of the utterances can be obtained from the first author.)

3.1.2. Procedure

The preceding and following contexts were categorized according to their phonological content. The preceding context was classified as either vowel, a consonant other than /s/, or /s/. For the following context, five categories were constructed. One category included all vowels. Consonants were categorized according to place of articulation in four categories: labial, labiodental, alveolar, and a back category for velar, uvular, and laryngeal place of articulation (chosen on the basis of the results of Corpus Study 1, which had suggested an influence of place of articulation). This leads to a two-factor design with Preceding Context and Following Context as independent variables. Repetition served as a co-variate. The proportion of /t/-deletion was again the dependent variable. Note that a quantification of /t/-lenition on a continuous scale is even more hazardous for spontaneous speech than for read aloud stories, as the background noise in the corpus makes fine-grained distinctions impossible.

3.2. Results and discussion

Table 3 shows the proportion of /t/-deletion in the various cells of the design. First of all, it is important to note that /t/-deletion is much more likely in casual speech (45%) than in read speech (8%). An ANOVA revealed that the covariate Repetition did not affect the proportion of /t/-deletion ($F<1$). There are three possible explanations for this unexpected null-effect. First of all, we did not take into account the time between repetitions. Therefore, we computed the time elapsed between a realization and its predecessor for the second to fifth realizations. This variable did not have a significant contribution ($p = 0.29$) when used as predictor in a logistic regression with /t/-deletion as dependent variable. Secondly, the null-effect might be due to the fact that we did not count realizations of the target within the conversation context by other speakers, which was suggested to be an important variable for reduction of articulatory effort by Bard et al. (2000). Third, the role of repetition may be smaller than suggested in some studies because of the confound of repetition with prosodic factors, such as accent (Hawkins & Warren, 1994).

In contrast, the effects of Preceding ($F(2, 109) = 26.76$, $p<0.001$) and Following Context ($F(4, 109) = 5.56$, $p<0.001$) were significant, as was the interaction between these two factors ($F(8, 109) = 2.67$, $p<0.05$). In order to evaluate the source of this interaction, we tested the effect of preceding context on all five levels of the following context. First, no F ratio could be computed

Table 3
The proportion of /t/ deletion as a function of previous and following context (number of observations in brackets) in Corpus Study 2 (C stands for consonant)

Following context	Preceding context			
	/s/	Consonant	Vowel	Mean
Vowel	0.77 (13)	0.18 (22)	0.00 (8)	0.32 (43)
Bilabial C	0.67 (3)	0.90 (10)	1.00 (2)	0.87 (15)
Back C	0.75 (8)	0.14 (14)	0.00 (2)	0.33 (24)
Alveolar C	1.00 (9)	0.00 (4)	0.00 (1)	0.50 (18)
Labiodental C	0.75 (16)	0.00 (12)	0.00 (3)	0.39 (31)
Mean	0.80 (50)	0.33 (65)	0.12 (16)	0.46 (125)

for the following alveolar consonants, because there was no error variance: /t/ was deleted in all ten cases after /s/, while it was never deleted after a vowel or another consonant than /s/. Second, if the word following the word-final /t/ started with a labial consonant, preceding context did not influence the proportion of /t/-deletion ($F(2, 12) = 1.54$, $p > 0.2$). In the three other following contexts, there was a significant effect of Preceding Context (labiodental: $F(2, 12) = 20.32$, $p < 0.001$; back consonants: $F(2, 21) = 6.92$, $p < 0.01$; vowels: $F(2, 21) = 13.84$, $p < 0.001$). Post hoc tests (HSD, $p < 0.05$) indicated for these three following contexts that /t/-deletion was more likely to occur after /s/ than after vowels or other consonants, while there were no significant differences between the categories Vowel and Other Consonant.

This indicates that /t/-deletion is likely to occur after /s/ independently of the following context. With regard to the following context, /t/-deletion most often occurs before bilabials, that is, before /b/ and (predominantly) before /m/. This effect of bilabials is not just an effect of /b/, since the majority of labials are nasals.

Since all participants provided at least three tokens, we were able to evaluate whether the effects reported above may be caused by only a sub-group of the participants. We tested this hypothesis by repeating the analyses with a multi-level model with participant as a random variable. The results of this analysis revealed no interaction of the main effects with speaker.

We now have obtained a picture of /t/-lenition in nominal forms in standard Dutch. /t/-lenition often gives rise to forms with very little evidence for /t/, actually leading to /t/-deletion. This/t/-deletion is most likely after /s/ and before bilabial consonants. However, even after /s/ and before bilabial consonants, /t/-deletion is optional (see also Fig. 1A).

How are listeners able to recognize words with lenited /t/s? Do listeners rely mainly on phonetic detail and phonological context in order to infer the intended word, as seems to be the case for assimilations, or do listeners also rely on semantic constraints in order to overcome /t/-lenition? We studied these questions in Experiments 1–3.

4. Experiment 1

Experiment 1 addressed the question whether listeners make use of segmental contextual and residual acoustic cues that are associated with /t/-lenition. Listeners decided in a 2AFC task which of two 'words'—with or without a word-final /t/—had been presented in slightly varying carrier sentences. In order to minimize lexical biases (Ganong, 1980; see also McQueen, 1991, 1996) we used only nonwords. The stimuli carried different forms of /t/ in different phonological contexts.

The levels of /t/-lenition and the different phonological contexts investigated in this experiment were dictated by the results of our earlier corpus studies. Apart from a full /t/ with closure, transient and frication, we used (1) a "frication-only" /t/, (2) a "weak-frication" /t/, and (3) a stretch of near silence, indicating a closure. In addition, for reasons indicated below, we also included (4) a long preceding consonant, as a signal unlikely to arise from a /Ct/ coda. These realizations do not form a natural continuum, but we feel that they represent tokens on a scale from least to most lenited /t/s. For convenience, we will refer to this scale as a five-step /Ct/-/C/ series.

In addition, our corpus studies indicated that both previous and following phonological context influences the likelihood of /t/-lenition. Lenition of /t/ is most common in Dutch after /s/.

Therefore, we had /n/ and /s/ as the preceding context, which differ strongly in the likelihood of subsequent /t/-lenition. With regard to the following phonological context, the deletion of a word-final /t/ in Dutch is most likely before bilabial consonants. Therefore, we used three words with initial Dutch bilabial phonemes /b/, /p/, and /m/ for the following context. These were contrasted with words with initial alveolar /n/ or velar /k/ as examples of following contexts that do not induce /t/-lenition to the degree that bilabial consonants do. If listeners take advantage of contextual and residual acoustic cues, we expect the following: /t/ should be reported more often in contexts that induce /t/-lenition, that is, after /s/ and before words starting with bilabial consonants.

4.1. Method

4.1.1. Participants

Ten female and four male members of the Max Planck Institute's subject pool were paid to participate in the experiment. They were aged 17–26 years (mean: 22). All participants were native speakers of Dutch and had no or a minimal knowledge of phonetics. No participant reported any (history of) hearing problems.

4.1.2. Stimuli

A male native speaker of Dutch recorded the sentences schematized in Table 4 several times. The target words—all Dutch nonwords—were pronounced with and without [t] according to the script given to the speaker. Formants and bandwidths were measured in the natural utterances at least at the beginning and at the end of each segment, and for vowels also at the midpoint. These formant and bandwidth values were used as templates for synthesis via a Klatt (1980) synthesizer. When the formant measurements yielded unstable results, values were chosen in accordance with the recommendations by Klatt (1980). Moreover, nasalization for a nasal was carried over into the vowel up to the midpoint of the vowel and it was anticipated for postvocalic nasals from the midpoint of the preceding vowel. Between specified points, values were interpolated linearly (for inter-point intervals <50 ms) or in a sigmoid fashion (for longer intervals). If necessary, values were altered to prevent clicks and other transients in the synthesized signal, which may occur if

Table 4
Sentence frame for the stimuli in Experiment 1

Subject + inflected verb form		Target word	Adverb + verb particle	
		spes[t… ∅] /spɛs[t… ∅]/	krachtig /ˈkrɑxtəx/	
		spen[t… ∅] /spɛn[t… ∅]/	nauwelijks /ˈnɑuʋələks/	
Wim/ʋɪm/	sprak	dris[t… ∅] /drɪs[t… ∅]/	bevend /ˈbeːvənt/	uit /œyt/
Jan /jɑn/	/sprɑk/	drin[t… ∅] /drɪn[t… ∅]/	prima /ˈprimaː/	
		blas[t… ∅] /blɑs[t… ∅]/	moeizaam /ˈmujzaːm/	
		blan[t… ∅] /blɑn[t… ∅]/		

The target word ends in one of the sounds of the [t.. ∅] series. English translations are Wim, Jan (Common Dutch Christian names), sprak…uit 'pronounced' (3rd singular past tense), krachtig 'forcefully,' nauwelijks 'hardly,' bevend 'shakingly,' prima 'nicely,' moeizaam 'with difficulty.' All targets, whether pronounced with or without /t/, are nonwords in Dutch.

control parameters change too rapidly. Amplitude values were iterated to imitate the amplitude envelope of the natural utterances. (The complete parameter set can be obtained from the first author.)

We synthesized five signals for the coda of the target word (see Fig. 2). First of all, we created a signal akin to a full [t] with a 25 ms closure and a 45 ms transient-frication sequence. In order to prevent an unnatural flat line in the signal, the closure was realized with 20 dB amplitude of fraction (AF) and a 30 dB amplitude bypass (AB) of the parallel branch of the synthesizer. At release, AB was set to zero, and AF increased to 40 dB, and decreased again to 15 dB at the end of 65 ms. For this and all other target signals, amplitude changes were loglinear in dB to achieve a linear amplitude envelope. The transient-frication signal was dominated by the fifth and six formants (55 dB) starting at 5700 Hz (bandwidth of 500 Hz) and 7500 Hz (700 Hz), respectively. These formants fell to 5200 and 6850 Hz at 65 ms. The second, third, and fourth formant stayed

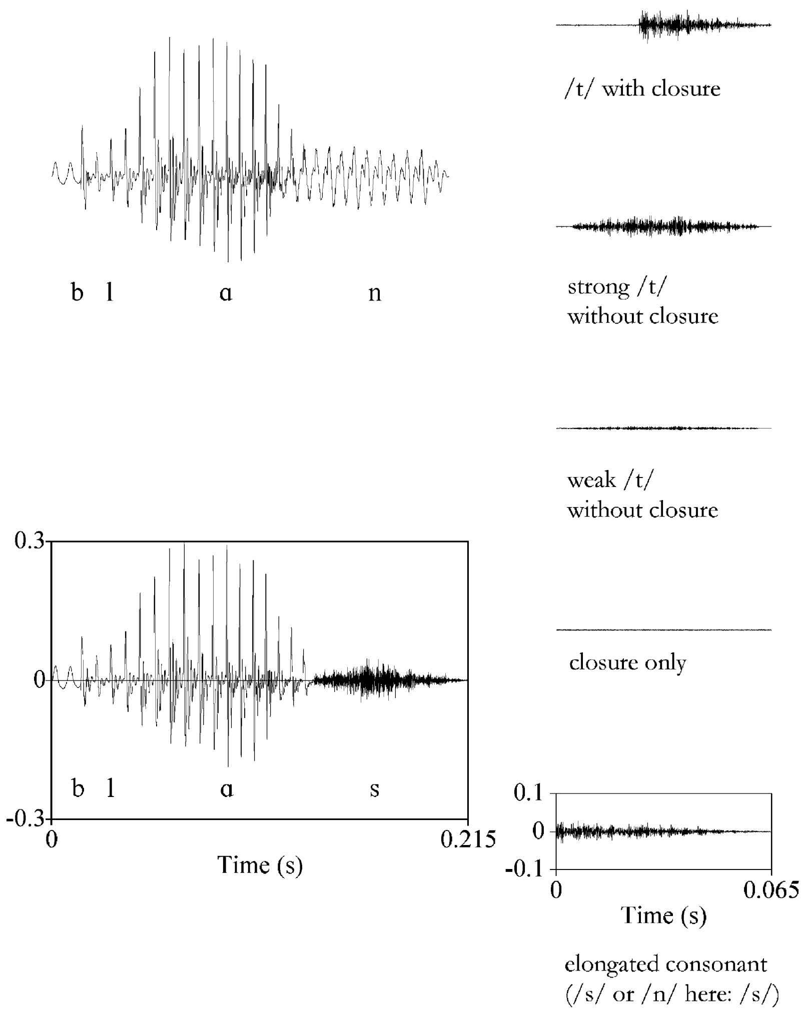


Fig. 2. The different target signals presented in Experiment 2. See method section for details.

constant throughout the transient-frication sequence at 1434 Hz (200), 2212 Hz (500) and 3840 Hz (600), respectively, and their amplitude increased from 20 to 25 dB in order to mimic the increasing low-amplitude frication in the model [t].

The second coda signal was a 65 ms frication noise, as often found in /st/ codas. The amplitude of frication started and ended at 15 dB with a 40 dB maximum at 30 ms. The settings for the second, third, and fourth formant were identical to the settings in the full [t]. The fifth and sixth formants started at 5130 Hz (bandwidth: 500 Hz) and 6750 Hz (700 Hz) and fell to 4620 Hz and 6080 Hz at the end of the signal (As = 55 dB). As Fig. 2 shows, the complete target then had a 130 ms frication with two amplitude maxima, rather like the natural utterance depicted in Fig. 1B. The weak-frication signal was derived from this signal by reducing the just described fricative noise by a factor of 5 (= 14 dB). The closure-only signal was synthesized with the same setting as for the closure of the full /t/.

In order to approximate a series in which /t/ was progressively deleted (henceforth:/t/-$\varnothing$ series), we elongating the (pen-)ultimate consonant (/n/ or /s/) by the duration of the other target signals, that is, 65 ms. The resulting stimulus in case of a preceding /s/ differed from the frication-[t] condition by having only one amplitude maximum in the 130 ms of frication. This signal is likely to prompt the perception of the absence of final /t/, due to the following reasons. Klatt (1974) showed that, in English, [s] is shorter in onset clusters (as in stake) than in simple onsets (as in sake), in line with the general trend of phonemes being shorter in larger units (e.g., Lehiste, 1970; Salverda, Dahan, & McQueen, 2003). In a small-scale study, we replicated this effect for Dutch final clusters. Two speakers produced short sentence frames (e.g., Zeg __ nog eens 'say __ once more') with the Dutch words kas 'green house' and kast 'cupboard', as well as the nonwords plis and plist. Only one out 120 cases showed evidence for /t/-lenition, probably due to the fact that the target word was in focus position, which may have encouraged a citation pronunciation. As expected, [s] was shorter in clusters (mean = 98 ms) than in simple codas (mean 148 ms; $t(70) = 7.02$; $p < 0.001$). An evaluation of tokens in Corpus Study 1 showed that /t/-deletion did not give rise to a significant lengthening of a preceding /s/ (durational difference: + 7 ms, $t(98) = 1.03$, $p > 0.2$). We can therefore conclude that a long [s] occurs most likely in underlying /s/ codas and not in /st/ codas, even in case of /t/-deletion. Accordingly, a long [s] can be considered to signal an underlying simple /s/ coda and not an /st/ coda in which /t/-deletion has occurred.

There were thus five coda signals for the target words: full /t/, strong frication /t/, weak-frication /t/, closure only, and long consonant. The signals were concatenated with the six initial parts of the target words (see Table 4), yielding 30 stimuli, for which participants had to decide whether the speaker produced a /C/- or a /Ct/-coda. Fig. 2 shows the initial parts of two target words and the five different signals of /t/. The complete target words had a duration ranging from 270 ms ("blan…") to 350 ms ("spes…"). Thus, the speaking rate was 3-to-5 syllables per second.

The 30 target words were presented in ten carrier sentences, resulting from combining the two possible initial parts of the sentences ('Wim sprak…' and 'Jan sprak…') with the five possible final parts (five adverbs followed by the verb particle 'uit'). This gives rise to 300 different sentences (see Table 4).

4.1.3. Procedure

Experiments were run on a standard PC running the NESU package. Participants wore headphones and faced a computer screen with a two-button response box in front of them. They were asked to press the right button if the sentence they heard contained the word presented in the upper right corner on the computer screen, and to press the left button if the sentence contained the word presented visually on the upper left corner. Each participant heard the 300 sentences in a different random order.

The experiment started with four practice trials in order to familiarize the listeners with the procedure. These practice trials contained four examples of the natural sentences used as templates for re-synthesis.

The practice and the experimental trials had the same structure. After 150 ms of blank screen, one response alternative (e.g. "drist") was presented in the upper right corner of the computer screen while the other (e.g. "dris") was presented in the upper left corner. The response allocation for /Ct/ and /C/ codas was the same throughout the experiment. After another 450 ms, the sentence was played. From the onset of the target word, participants had a 2.5 s interval in which to respond. After responding, the visual display of the chosen alternative was moved further to the upper right or left corner while the other alternative was removed from the screen. In case of a time-out error, a stopwatch was shown to remind participants to respond faster. All three feedback signals—no response, [C] response, [Ct] response—stayed on the screen for 1 s before the next trial began. Each of the 300 sentences was presented once to each participant. Participants had the opportunity to take a short break after each 50th trial.

4.1.4. Design

There were three independent variables. The first was the nature of the Coda Signal at the end of the target word (full /t/, /t/-frication, weak /t/-frication, closure, elongated consonant). The second independent variable was the Preceding Context of the Target Signal (/n/ versus /s/). Note that this segment is the (pen)ultimate phoneme of the target word. The third independent variable was the Following Context (/k/, /n/, /p/, /b/, or /m/), which is the initial consonant of the following adverb. The dependent variable was the percentage of /t/-responses in each cell of this design.

4.2. Results and discussion

The mean percentages of /t/-responses for all cells of the design are shown in Fig. 3A–E, while Fig. 3F shows the mean percentages of /t/-responses aggregated over Following Context. As is immediately evident, Preceding Context /s/ elicited more /t/-responses than preceding /n/. While the likelihood of a /t/-response decreased continuously for successive reductions of /t/ in the preceding-/n/ context, the likelihood of a /t/-response after /s/ remained comparatively level for the first three reduced realizations of /t/—frication only, weak frication, and closure only.

This was confirmed by an ANOVA with Coda Signal, Preceding Context, and Following Context as predictors on the arcsine transform of the percentages of /t/-responses. This ANOVA yielded a significant effect of Preceding Context ($F(1,13) = 71.91 p < 0.001$), Following Context ($F(3,38) = 3.06$, $p < 0.05$), and Coda Signal ($F(4,32) = 97.01$, $p < 0.001$). Where appropriate, degrees of freedom have been corrected using the Greenhouse–Geisser method. In addition, there was a significant interaction of Preceding Context and Target Signal ($F(4,46) = 14.51$, $p < 0.001$, see below for further testing of this interaction). All other interactions failed to reach significance [Preceding Context by Following Context ($F(4,46) = 2.16$, $p > 0.05$); Following Context by

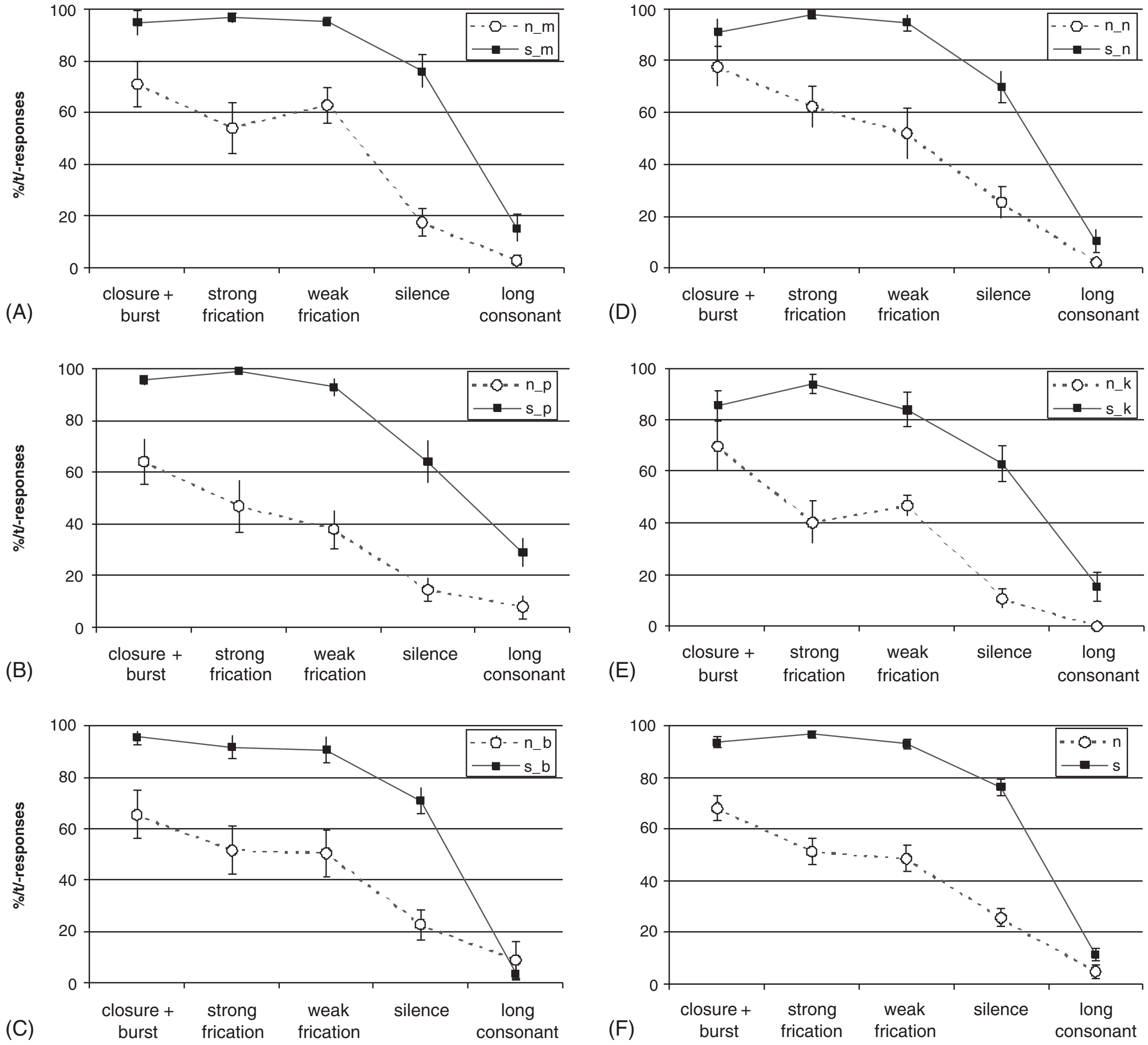


Fig. 3. Percentages of /Ct/ identifications in Experiment 2. The Figs. 3A to 3E each show the data for one following context indicated in the legend. Fig. 3F shows the results collapsed over the five following contexts. The open circles indicate the results obtained for codas with /n/ and the filled squares for codas with /s/. Errors bars indicate the standard error of the mean.

Target Signal ($F<1$); Preceding Context by Following Context by Target Signal ($F(6, 83) = 1.59$, $p>0.1$)]. Post hoc tests (HSD) for the factor Following Context yielded no significant difference between any two means. A least-significant difference criterion revealed significantly more /t/-responses before nasals (/n/ and /m/, 58.4% and 58.6% respectively) than before /k/ (51.0%), and more /t/-responses before /m/ than before /p/ (54.2%). This weak pattern does not show any clear association with the place of articulation of the following segment.

A surprising aspect of the current data is that listeners did not reach 100% /nt/-judgments when the /n/ was followed by a full /t/. Overall, these stimuli were perceived as /nt/ in about 70% of trials (see Fig. 3F). A likely interpretation of this result is that /t/s following /n/ may be perceived as epenthetic stops, which may occur after nasals in Dutch (see Warner & Weber, 2001).

In order to determine the nature of the interaction between Preceding Context and Coda Signal, we examined the effect of Preceding Context on all levels of our /t/-$\varnothing$ series. All five *t*-tests were significant ($p_{max} < 0.005$). Hence, the interaction must be caused by a difference in size of the effect of Preceding Context on the different levels of the series. To examine this, we computed the differences between the percentages of /t/ responses with /n/ and /s/ as Preceding Context for all five levels of the Coda Signal, after applying an arcsine transformation. These five difference values were subjected to a one-factorial ANOVA ($F(4, 49) = 14.51$, $p < 0.001$). Post hoc tests (HSD) showed that the differences were larger for a strong frication /t/, a weak-frication /t/, or a closure than for a full /t/ or a long consonant /n/ or /s/. That is, the effect of the preceding context was larger in the cases of more ambiguous signals—the intermediate conditions of our Coda Signal series—than in cases of clear signals—the endpoints of our Coda-Signal series. This mirrors other findings on context effects in speech perception, which have also shown maximal context effects for ambiguous signals (Allen & Miller, 2001; Gow, 2003; Fowler et al., 2000; Liberman, 1996; Lotto & Kluender, 1998; Mann, 1980; Massaro & Cohen, 1983; Nittrouer & Studdert-Kennedy, 1987; Sinnott & Saporita, 2000).

This experiment shows that listeners take into account the preceding phonological context in their interpretation of codas. Listeners tend to report /t/ more often in contexts where it is often severely lenited. That is, speakers lenite /t/ after /s/ more often than after /n/, and listeners 'hear' /t/ more often after /s/ than after /n/. However, an alternative interpretation of this result is that listeners were influenced by statistical regularities of Dutch, that is, the high number of /st/ responses compared to the /nt/ responses might be due to the preponderance of /st/ and simple /n/-codas in Dutch. In order to examine this possibility, we calculated the ratio of /C/- to /Ct/-words in the Dutch CELEX lexical database for /s/ and /n/. This showed that, contrary to the preferences of the listeners in the present experiment, /s/-codas are more likely than /st/-codas (/s#/-to-/st#/ ratio: lemmas, 3.81; wordforms, 3.22), while there is no clear bias for either coda type for /n/ (/n#/-to-/nt#/ ratio: lemmas, 1.04; wordforms, 0.63).

In summary, this experiment clearly shows that listeners take into account the preceding phonological context and sub-phonemic detail in their interpretation of codas. Listeners 'hear' /t/ more often after /s/ than after /n/. This effect of the preceding context was modulated by sub-phonemic detail: The signals with a short preceding consonant plus a closure triggered about 50% more /t/-responses in the /s/-context than in the /n/-context. In contrast, if there was no closure but instead a long preceding consonant, this difference was reduced to 12%.

5. Experiment 2

In the previous experiment, the effects of bottom-up information were investigated. In Experiment 2, we investigated the impact of higher-level knowledge. That is, are listeners more likely to infer the presence of a /t/ if this would give rise to a more meaningful interpretation of the utterance? If this is the case, we would expect listeners to report a word-final /t/ more often if this interpretation of the signal would turn the stimulus into an existing word. That is, they would report /t/ more often for forms such as [ɔʀkɛs…], which are existing Dutch words with, but not without, final /t/ (orkest /ɔʀkɛst/ 'orchestra') than for forms such as [muʀɑs…], which are existing words without, but not with, /t/ (moeras /muʀɑs/ 'swamp').

It is well known that listeners have a tendency to interpret an ambiguous signal in such a way that the input forms a word (Ganong, 1980). The same fricative noise is more likely to be perceived as an /s/ in the form /krɪsmɑ/ + fricative, in line with the word 'christmas', than in the form /fulɪ/ + fricative, which becomes the word 'foolish' if the fricative is interpreted as /ʃ/ (see, e.g., McQueen, 1996, for a review). Such lexical context effects are, however, not mandatory (McQueen, 1991) and seem to have less of an impact on phonetic perception than phonetic context effects (Allen & Miller, 2001).

In line with findings that lexical effects are not compulsory, Mitterer and Blomert (2003) found that compensation for nasal-place assimilation in Dutch was independent of the lexical status of the target word. In Experiment 2, we tested whether a lexical bias arises in the interpretation of lenited /t/s. We used the same task as with which Mitterer and Blomert failed to find a lexical effect for compensation for place assimilation, that is a 2AFC task.

5.1. Method

5.1.1. Participants

Thirteen female and three male members of the Max-Planck-institute's subject pool were paid to participate in the experiment. They were aged 19–25 years (mean: 21). All participants were native speakers of Dutch. No participant reported any (history of) hearing problems. One participant had already participated in Experiment 1. The results of this participant followed the general trend of the other participants.

5.1.2. Stimuli

The target stimuli used in this experiment were the nonwords /ɔʀ-ˈkɛs/ and /ʃɑʀ-ˈmɑn/, as well as the Dutch words /mu-ˈʀɑs/ and /kaː-ˈnɔn/ followed by one of the five /t/-realizations described in Experiment 1 (full /t/, strong frication /t/, weak-frication /t/, closure-only, long consonant). The targets /ɔʀ-ˈkɛs/ and /ʃɑʀ-ˈmɑn/ are existing Dutch words if followed by /t/ (orkest /ɔʀ-ˈkɛst/ 'orchestra' and charmant /ʃɑʀ-ˈmɑnt/ 'charming'). The two targets /mu-ˈʀɑs/ 'swamp' and /kaː-ˈnɔn/ 'canon' are existing Dutch words but no real words can be formed from these by adding a following final /t/ (that is, there are no real words /mu-ˈʀɑst/ or /kaː-ˈnɔnt/). These four words have similar lemma frequencies in the CELEX lexical database (Baayen et al., 1995): charmant 11.6 per million, kanon 11.0 per million, orkest 11.5 per million, and moeras 11.4 per million. The four words are phonotactically legal independently of whether they are realized with or without /t/. The change from one-syllable words in Experiment 1 into two-syllables words here was introduced both because this allowed a better matching of word frequencies and because longer words give rise to stronger lexical effects (see Norris, McQueen, & Cutler, 2003).

The stimuli were synthesized using the same model speaker as for Experiment 1, and embedded in the same carrier sentences as in Experiment 1. In order to allow a direct comparison between the different targets, the same synthesis parameters were used for the (pen)ultimate segment /n/ or /s/, respectively. Hence, the /n/ was acoustically identical in /ʃɑʀ-ˈmɑn/ and /kaː-ˈnɔn/, and the /s/

was acoustically identical in /ɔʀ-ˈkɛs/ and /mu-ˈrɑs/. With these four target words, 200 sentences were created using the two possible lead-ins, the five signals of the target word, representing our /t/-∅ series, and the five adverbs followed by the verb particle 'uit' (see Table 4).

5.1.3. Procedure

The procedure was similar to Experiment 1. After a short practice of four trials, the experimental trials started. The trial set-up was the same as in Experiment 1.

5.1.4. Design

The design resulted in three independent variables: Coda Signal (five levels, cf. Experiment 1), Preceding Context (/n/ or /s/), and Lexical Status (word with or without /t/). The dependent variable was the percentage of /t/-responses calculated from 10 trials per experimental cell.

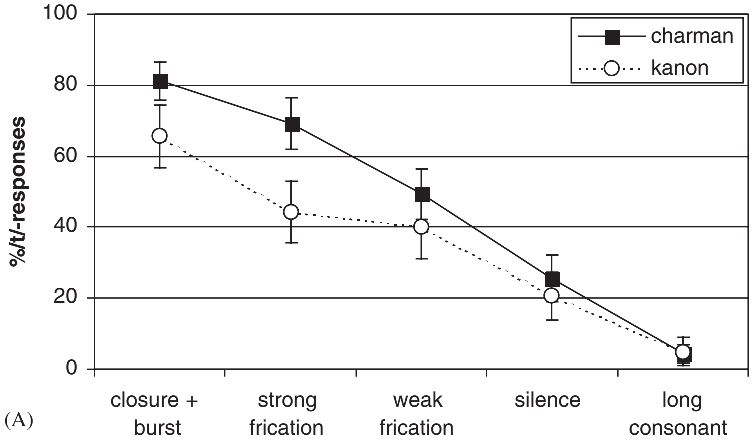
5.2. Results and discussion

Fig. 4 shows the mean percentages of /t/-responses. The individual percentages of /t/-responses were arcsine transformed and submitted to a repeated-measure ANOVA with Lexical Status, Coda Signal, and Preceding Context as predictors. The analysis revealed that listeners were more likely to report the presence of a /t/ if this interpretation gives rise to a word (Lexical status: $F(1,15) = 11.45$, $p<0.005$), if the coda contains an /s/ (Preceding Context: $F(1,15) = 22.59$, $p<0.001$), and if the Coda Signal carries more information for the presence of a /t/ (Coda Signal, $F(2,32) = 96.42$, $p<0.001$). These simple effects were modulated by significant interactions of Lexical Status and Coda Signal ($F(3,48) = 3.17$, $p<0.05$), and Preceding Context and Coda Signal ($F(3,43) = 14.01$, $p<0.001$). In addition, the three-way interaction between Coda Signal, Lexical Status, and Preceding Context ($F(3,52) = 5.32$, $p<0.005$) was also significant (see below).

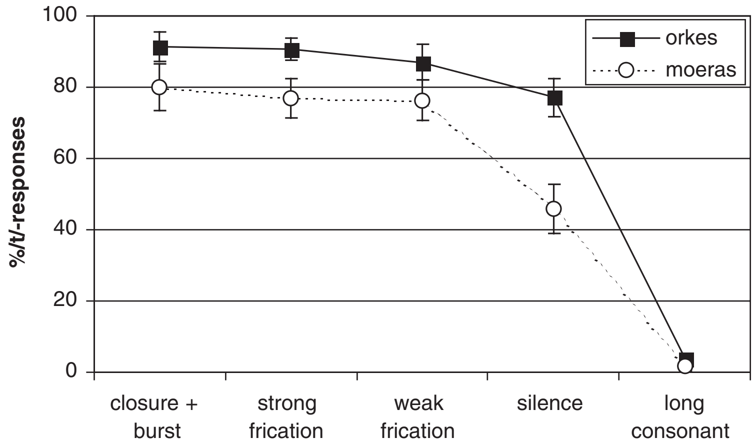
It has to be noted that this experiment replicated the effects of Preceding Context and Coda Signal observed in the previous experiment. Listeners were again more likely to report a /t/ after /s/ than after /n/. As in Experiment 1, the effect of preceding phonological context depended on sub-phonemic details: Both Figs. 3 and 4, for Experiments 1 and 2, respectively, show that the difference in the instances of reported /t/ between /n/ and /s/ codas is larger in the more ambiguous steps of our /C/-/Ct/ series.

In order to investigate the nature of the three-way interaction, we tested the effect of Lexical Status for all cells of the design that arise by crossing Preceding Context and Coda Signal. After /n/, Lexical Status was significant for the full /t/ ($t(15) = 2.24$, $p<0.05$) and the strong frication /t/ ($t(15) = 5.38$, $p<0.001$). Lexical status did not affect responses significantly for any of the weak-frication /t/, ($t(15) = 1.68$, $p>0.1$), the closure-only, or the long-consonant condition (ts<1). After /s/, the effect of Lexical status was significant for frication-/t/ ($t(15) = 3.51$, $p<0.005$) and closure only ($t(15) = 3.39$, $p<0.005$). For the full /t/ and the weak-frication /t/, there was only a trend ($t(15) = 1.83$, $p<0.1$ and $t(15) = 1.6$, $p<0.1$, respectively). There was no significant effect in the long-consonant condition ($t(15) = 1.40$, $p>0.1$).

These analyses in combination with Fig. 3 show that the lexical effect is largest for the frication-/t/ preceded by /n/ and the closure-only Coda Signal preceded by /s/. That is, after both /n/ and /s/, the lexical effect is thus maximal if the Coda Signal does not clearly favor an



(A)



(B)

Fig. 4. Percentages of /Ct/ identifications in Experiment 3. Panel A shows the results obtained for targets with coda /n/, while Panel A shows the results obtained for targets with coda /s/. The open circles indicate results for targets that are words without word-final /t/ and the filled squares for targets that are words with word-final /t/. Errors bars indicate the standard error of the mean.

interpretation as /C/ or /Ct/. This indicates that top-down information mainly influences the interpretation of the target signal if the bottom-up information is ambiguous (cf. Massaro, 1998; Norris, McQueen, & Cutler, 2000). It has to be noted, however, that the lexical effect is stronger for the strong-frication target signal than for the weak-frication target signal, although the strong-frication signal is, if anything, less ambiguous overall than the weak-frication signal.

The main conclusion of this experiment is, however, that even in a 2AFC task, lexical status influences the interpretation of a lenited /t/. This result fits a recent result by Janse, Nooteboom and Quené (submitted). They found that the perception of a word with either a single /s/ or an /st/ coda was strongly influenced by the presence of a lexical competitor, which also shows lexical effects.

6. Experiment 3

In the previous two perception experiments, two patterns give rise to further questions. First of all, listeners were more likely to report an underlying /t/ when the signal consisted of a short /s/ followed by silence than in case of a long /s/. The question is whether this difference is due to the difference in the duration of the /s/ or to the presence of a silence. Previous research on the perception of /s/ versus /st/ onset clusters suggests that the presence of silence may be the strongest cue for listeners to perceive a /st/-cluster (e.g., Best, Morrongiello, & Robson, 1981; Repp, Liberman, Eccardt, & Pesetsky, 1978). A second pattern of observation that needs additional investigation is that listeners often did not report /t/ for a full /t/ preceded by /n/. We argued that listeners might perceive stops after nasals as epenthetic and not as underlying. An alternative possibility is that the /t/s in these synthesized signals were not sufficiently clear.

Both questions can be addressed by a simple modification of Experiment 1, that is, by presenting the stimuli without the following context. The lenited forms short /s/ plus silence and long /s/ then only differ in terms of the length of the /s/, as both are followed by silence until the start of the next trial. If indeed the presence of an inserted silence in the speech stream is the cue for listeners to perceive an underlying /t/, we should not find any difference between these two lenition forms of /t/ before a pause. If, however, the duration of /s/ is the crucial cue, listeners should report /t/ in the short-/s/ condition, but not in the long-/s/ condition. With respect to epenthetic stops, we assume that, since they mainly occur phrase-internally (Warner & Weber, 2001), a phrase-final /t/ is not likely to be interpreted as an epenthetic stop . Hence, if we did not reach 100% of /t/-responses after /n/ in Experiment 1 and 2 because the /t/s were interpreted as epenthetic, we expect that the percentage of /nt/ response should get close to 100% if /t/s are presented without following context.

6.1. Method

6.1.1. Participants

A total of 14 female and six male members of the Max-Planck-institute's subject pool were paid to participate in the experiment. They were aged 17–26 years (mean: 21). All participants were native speakers of Dutch. No participant reported any (history of) hearing problems. Five participant had already participated in Experiment 2. The results of these participants followed the general trend of the other participants.

6.1.2. Materials

The material used in this experiment consisted of the synthesized sentences from Experiment 1 (see Table 4), except that the sentences stopped after the target word (e.g., *Jan sprak blas*).

6.1.3. Procedure

The procedure was the same as in Experiment 1.

6.1.4. Design

There were two independent variables. The first was the nature of the Target Signal at the end of the target word (full /t/, /t/-frication, weak /t/-frication, closure, elongated consonant). The second independent variable was the Preceding Context of the Target Signal (/n/ vs. /s/). Using two different lead in sentences (*Wim sprak* or *Jan sprak*) and three different carrier pseudowords, there were six unique trials in every cell of this design. Each of these unique trials was presented twice, giving rise to twelve trials per cell. The dependent variable was the percentage of /t/-responses calculated over the 12 trials in each cell.

6.2. Results

Fig. 5 shows the results as percentages of /t/-responses. As Fig. 5 shows, we observed similar effects as in our previous experiments, in which the target items were embedded in longer sentences. More /t/-responses were given after /s/ than after /n/, as is borne out by an ANOVA with Preceding Context and Target Signal as predictors. Both main effects were significant (Preceding Context: $F(1, 19) = 8.74$, $p < 0.01$; Target Signal: $F(4, 76) = 69.82$, $p < 0.001$), as was their interaction ($F(4, 76) = 21.92$, $p < 0.001$). In order to investigate the nature of this interaction, we evaluated the effect of Preceding Context on each level of the /t/-$\varnothing$ series. This procedure revealed that the Preceding Context influenced the responses only for the weak-frication coda signal ($t(19) = 3.81$, $p < 0.005$) and the closure-only signal ($t(19 = 5.84$, $p < 0.001$; all other ts < 1).

While the general pattern of results is not different from the previous experiments, a difference emerged at the /t/-end of the series where the effect of Preceding Context was completely absent. In our earlier experiment, participants did not reach a very high percentage of /t/-responses after /n/, even if a full /t/ was presented. We deemed it likely that these /t/s were perceived as epenthetic stops, and the current results support this interpretation. Also, note that the short /s/ plus silence condition elicited a large percentage of /t/-responses after /s/ in both Experiment 1 and this experiment. In Experiment 1, the silence was part of the speech stream as it was followed by another consonant, and may have acted as a cue for /t/ (see, e.g., Best et al., 1981; Repp et al., 1978) In this experiment, the silence was not followed by another consonant, and therefore not an integral part of the speech stream. Accordingly, the main difference between the short /s/ plus

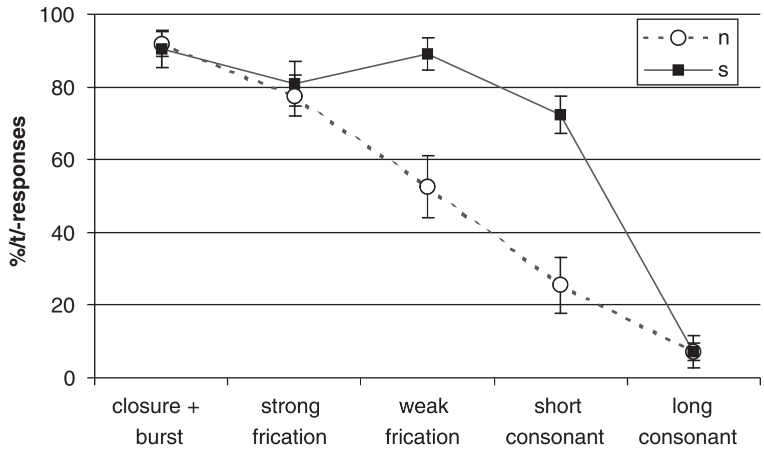


Fig. 5. Percentage of [Ct] identifications in Experiment 3. The open circles indicate the results obtained for codas with [n] and the filled square for codas with [s]. Errors bars indicate the standard error of the mean.

silence condition and the long /s/ condition is the duration of the /s/ in this experiment. This difference appeared to be sufficient to prompt the perception of an underlying /t/ for the short /s/ but not the long /s/.

7. General discussion

This study investigated the production and perception of lenited word-final /t/ in Dutch. Two corpus studies confirmed earlier results obtained for the English language and incidental observations on Dutch: Word-final /t/ is most likely to be lenited after /s/ and preceding bilabial consonants. In addition, the likelihood of /t/-lenition was shown to depend on speech register: /t/-lenition was most likely in spontaneous speech, less likely in read stories, and hardly ever occurred in the sentence-reading task described in the Materials section of Experiment 1, which may have encouraged a citation pronunciation.

It should be noted that the last register, short sentences elicited in a reading task, constitutes the vast majority of the material used in spoken-word recognition research. Accordingly, spontaneous-speech phenomena, such as /t/-lenition, are seriously underrepresented in current research on spoken-word recognition. The current paper is an attempt to start filling this lacuna.

The corpus studies also indicated that /t/-lenition is not obligatory in any phonological context: Even in the phonological context in which /t/ is most likely to be lenited (after /s/ and before bilabial consonants), it is sometimes realized in its full form (with closure and release). The variation in the realization of /t/ is a challenge for word recognition: All words may or may not be pronounced with full, lenited or even absent word-final /t/s, which implies that final /t/ is not an invariant property of words. Fortunately, sub-phonemic cues may signal the presence of an underlying /t/: A fricative is shorter in a /Ct/ coda than in a /C/ coda.

The lenition of /t/ does not necessarily lead to /t/-deletion. We observed several forms of reduction of the canonical realization of /t/ in careful speech: /t/ was sometimes produced as a fricative, without any visible closure or transient release burst. In some cases, this frication had a relatively low amplitude. Moreover, we also observed cases in which there was only a closure, but no visible /t/-release. Obviously, /t/-deletion is not a phonological rule-governed process, but is more likely to arise from a weakening of the /t/-gesture or from gestural overlap with adjacent segments (Browman & Goldstein, 1990, 1992). Evidence that assimilation (Holst & Nolan, 1995; Zsiga, 1994, 1995) and schwa-deletion (Manuel, 1992) are gradient has been reported previously and so is not unexpected for /t/-deletion.

Given these findings from the corpus studies, we ran three perceptual experiments. The first of these experiments examined whether listeners make use of the cues that phonetic detail and phonological context supply for the probability for an underlying /t/. This appeared to be the case. Listeners are more likely to infer the presence of a /t/ after /s/ than after /n/, compensating, as it were, for the higher likelihood of /t/-lenition after /s/ than after /n/. Moreover, a coda consisting of a long /n/ or /s/ discouraged listeners from inferring the presence of a /t/.

A second experiment examined the influence of lexical status on the perception of lenited /t/. Listeners were more likely to infer a /t/ if assuming a /t/ gave rise to an existing word. Hence, a /t/ was more often assumed to be present after the nonword /ɔʀkɛs/, which becomes the word orkest /ɔʀkɛst/ 'orchestra' by adding /t/, than after the word moeras /muˈʀɑs/ 'swamp', which becomes a nonword by adding /t/. This lexical bias was modified by the combination of preceding context and phonetic detail. In general, the lexical bias was maximal, if combination of preceding context and phonetic detail gave rise to an ambiguous interpretation of the coda. This replicated the general pattern of higher-level context effects in speech perception (e.g., Massaro, 1998).

A third perceptual experiment clarified some of the earlier results. In this experiment, the target words were presented without any following context. This allowed us to determine whether the high percentage of /t/-responses in Experiments 1 and 2 to a short /s/ and a following silence was due to the duration of the /s/ or to the presence of an inserted silence in the speech stream. Moreover, we hypothesized that in Experiments 1 and 2 listeners failed to report /t/ after /n/, even in signals in which there was considerable acoustic evidence for the presence of an underlying /t/ (i.e., full /t/, strong frication), because they interpreted the stops as epenthetic. Epenthetic stops are infrequent utterance-finally, and we thus predicted close to 100% of /t/-identification for full /t/ after /n/ utterance-finally. Experiment 3, in combination with Experiment 2, showed that listeners interpreted short [s] equally often as /t/ with and without following context. Hence, the length of the [s] is the major cue to the identity of the coda. Furthermore, listeners reported /t/ almost always for a full /t/ after [n], if no following context was used. This suggests that the comparably low percentage of /t/-responses after /n/ in Experiment 1 and 2 may be interpreted as a consequence of the interpretation of the stops as epenthetic.

The results of our perceptual experiments constitute a new example of a sequential context effect in speech perception. Previous results mainly indicated that the perception of the *identity* of a segment is affected by its adjacent segments (Lindblom & Studdert-Kennedy, 1967; Mann, 1980; Mann & Repp, 1981; Miller & Liberman, 1979; Gaskell & Marslen-Wilson, 1996). Our present results add a new twist to such context effects by showing that adjacent segments also influence the perception of the *presence* of a segment (see Repp, 1983; Repp et al., 1978, for similar effects). It is still unclear how to account for this context effect. Future research will have to investigate whether it is the consequence of co-variant learning (Gaskell, 2003), hierarchical categorization of phoneme sequences (Smits, 2001a, b), the perception of residual speech gestures (Fowler, 1996), or a consequence of auditory processing (Lotto & Kluender, 1998).

The interaction of this context effect with lexicality, as observed in Experiment 2, indicates that /t/-lenition, in terms of perception, is halfway between phonological assimilations and extreme reductions. Since a lexical bias on a similar task as in the current study was not observed for phonologically assimilated forms (Mitterer & Blomert, 2003), compensation for phonological assimilation seems to depend more strongly on ‘bottom-up’ information than compensation for /t/-lenition.

This leads us to conclude that compensation for /t/-lenition depends more on higher-level knowledge than compensation for assimilation. This is not surprising, given that /t/-lenition seems to lead to stronger deviation from the full pronunciation of a word than assimilation (see also Gow, 2002). The bottom-up signal seems to provide fewer acoustic cues for the canonical form in case of /t/-lenition than in case of assimilation. As a consequence, lexical knowledge plays a greater role in compensation for /t/-lenition than in compensation for assimilation. In this sense, the perception of /t/-lenition bears some resemblance to the perception of extremely reduced forms. Ernestus et al. (2002) found that extremely reduced word forms cannot be recognized on the basis of phonological context alone. Instead, a complete sentence context has to be available.

The present results constrain the possible architectures of models of spoken-word recognition. One possibility for solving the invariance problem is to specify only those attributes of a word that are invariantly present in the input (Lahiri & Reetz, 2002). Given that word-final /t/ may be deleted after /s/ and after /n/, lexical representations should be underspecified for words ending in /nt/ as well as for words ending in /st/. Accordingly, Dutch listeners should infer a /t/ as often at the end of /ɔʀkɛs/ as at the end of /ʃɑʀmɑn/, because these surface form both match an underspecified representation of the words *orkest* and *charmant*. However, contrary to this prediction, we found that /t/ was more often inferred after /ɔʀkɛs/ than after /ʃɑʀmɑn/.

Moreover, the gradient nature of /t/-lenition disfavors any model that makes hard decisions at a prelexical processing stage, since some of the residual cues for /t/ are not entirely reliable. For instance, our results seem to suggest that any short pause, as well as short /s/, in the acoustic signal would have to be probabilistically coded as signaling the presence of a /t/. Such a criterion would give rise to an enormous amount of false alarms at, for instance, prosodic boundaries' short pauses. Hence, it is difficult to reconcile a hard decision with the variability of /t/ in spontaneous speech.

Furthermore, it is important to note that ambiguous forms of /t/ are accepted as underlying /t/s depending on the phonological context. Models making hard decisions on a prelexical stage may accommodate this phonological-context effect by probabilistic binary coding: For instance, it may be assumed that after /s/, a /t/ is assumed to be present in a certain percentage of cases, while preceding /n/ does not enlarge the likelihood of /t/.

Even though both main effects of phonetic detail and phonological context may be incorporated into models making hard decisions on a prelexical stage, such models, in their simplest forms, would predict these main effects to be additive (cf. Massaro, 1998, pages 58–59), which is not consistent with the results from the present experiments. After a long coda consonant, /t/ was hardly ever inferred in our experiments, irrespective of the phonological context in which /t/ occurred. In contrast, if there is a closure, a /t/ was inferred in about 25% of the cases after /n/ and in about 65% of the cases after /s/. In order to explain the context effects and their interaction with phonetic detail within models making hard decisions on a prelexical level, we have to assume that these decisions are postponed until all segments that may affect the identification of a given segment have been perceived. This requires that decisions about individual segments be deferred for at least 200 ms (e.g., Nooteboom, 1979; Nooteboom & Doodeman, 1980; Sawusch & Newman, 2000). However, Phillips (2001) presented electrophysiological evidence that at least preliminary phonetic categorizations for stops are available around 150 ms after stimulus onset.

The interaction of phonological context and phonetic detail is better explained by prelexical representations consisting of continuously activated prelexical segments. With such an architecture, preliminary phonetic categorizations can be immediately achieved without losing phonetic detail information as implied by hard decisions (cf. Davis, Marslen-Wilson, Gaskell, 2002; Gow & Gordon, 1995; Salverda et al., 2003).

Categorical decisions on a prelexical stage are also absent in exemplar-based models, which can also account for our data (Goldinger, 1998; Pierrehumbert 2003). In such models, phonetic detail is part of the cognitive representation of lexical items, as the lexicon consists of episodic traces of words. Hence, phonetic detail is bound to influence lexical activation. The fact that effects of phonetic detail not only occur for words (Experiment 2) but also for nonwords (Experiment 1) may result from lexical analogy (cf. Kemps, Ernestus, Schreuder, & Baayen, in press). Episodic models can also explain the prevalence of preceding over following context effects. Obviously, preceding context will be invariant for final segments of a given word, while the following context varies over instances.

In summary, the current data show that /t/-lenition is not only acoustically but also in terms of processing similar to both phonological assimilations and extremely reduced forms. Listeners are attuned to the phonological contexts and phonetic details associated with /t/-lenition. In this respect, compensation for /t/-lenition is similar to compensation for phonological assimilation. However, lexical biases influence compensation for /t/-lenition more strongly than compensation for assimilation, indicating that bottom-up information may not be sufficient to disambiguate all instances of /t/-lenition. In this respect, the perception of /t/-lenition is similar to the perception of reduced forms.

Acknowledgements

The authors wish to thank Anne Cutler, Jonathan Harrington, Sarah Hawkins, and two anomymous reviewers for their comments on previous versions of this manuscript.

References

Allen, J. S., & Miller, J. L. (2001). Contextual influences on the internal structure of phonetic categories: A distinction between lexical status and speaking rate. *Perception & Psychophysics*, *63*, 798–810.

Baayen, R. H., McQueen, J. M., Dijkstra, T., & Schreuder, R. (2003). Frequency effects in regular inflectional morphology: Revisiting Dutch plurals. In R. H. Baayen, & R. Schreuder (Eds.), *Morphological structure in language processing* (pp. 355–390). Berlin: Mouton de Gruyter.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database (release 2) [CD-ROM]*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania [Distributor].

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, *42*, 1–22.

Best, C. T., Morrongiello, B., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics*, *29*, 191–211.

Bornstein, M. H. (1987). Perceptual categories in vision and audition. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 287–300). Cambridge, MA: Cambridge University Press.

Browman, C. P., & Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston, & M. E. Beckman (Eds.), *Papers in laboratory phonology 1: Between the grammar and physics of speech* (pp. 341–376). Cambridge: Cambridge University Press.

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, *49*, 155–180.

Bybee, J., & Scheibman, J. (1999). The effect of usage of degrees of constituency: The reduction of don't in English. *Lingusitics*, *37*, 575–596.

Cole, R. A. (1973). Listening for mispronunciation: A measure of what we hear during speech. *Perception & Psychophysics*, *13*, 153–156.

Dahan, D., & Tannehaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: Immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 498–513.

Davis, M. H., Marslen-Wilson, W. D., & Gaskell, G. M. (2002). Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology, Human Perception and Performance*, *28*, 218–244.

Deutsch, D. (1992). Paradoxes of musical pitch. *Scientific American*, *267*, 88–95.

Ernestus, M. (2000). *Voice assimilation and segment reduction in Dutch [dissertation]*. Utrecht, The Netherlands: LOT.

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

Farnetani, E. (1997). Coarticulation and connected speech processes. In W. J. Hardcastle (Ed.), *Handbook of phonetic sciences*. Oxford, UK: Blackwell.

Fowler, C. A. (1988). Differential shortening of repeated content words produced in various communicative contexts. *Language and Speech*, *31*, 307–319.

Fowler, C. A. (1996). Listeners do hear sounds, not tongues. *Journal of the Acoustical Society of America*, *99*, 1730–1741.

Fowler, C. A., Brown, J. M., & Mann, V. A. (2000). Contrast effects do not underlie effects of preceding liquids on stop-consonant identification by humans. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 877–888.

Fowler, C. A., & Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, *25*, 489–504.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology*, *6*, 110–125.

Gaskell, M. G. (2003). Modelling regressive and progressive effects of assimilation in speech perception. *Journal of Phonetics*, *31*, 447–463.

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

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

Gaskell, M. G., & Marslen-Wilson, W. D. (2001). Lexical ambiguity resolution and spoken word recognition: Bridging the gap. *Journal of Memory and Language*, *44*, 325–349.

Gernsbacher, M. A. (1994). *Handbook of psycholinguistics*. San Diego, CA: Academic Press.

Goeman, A. (1999). *T-deletie in nederlandse dialecten. kwantitatieve analyse van structurele, ruimtelijke en temporele variatie [T-deletion in Dutch dialects. Quantitative analysis of structural geographical and temporal variation]*. The Hague: Holland Academic Graphics/Thesus.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*, 251–279.

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

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., & Gordon, P. C. (1995). Lexical and prelexical influences on word segmentation. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 344–359.

Grimson, A., & Cruttenden, A. (1994). *Gimson's pronunciation of English* (5th ed). London: Edward Arnold.

Guy, G. R. (1980). Variation in the group and the individual. In W. Laboc (Ed.), *Locating language in time and space* (pp. 1–36). New York: Academic Press.

Guy, G. R. (1992). Contestual condition in variable lexical phonology. *Language Variation and Change*, *3*, 223–229.

Hawkins, S., & Warren, P. (1994). Phonetic influences conversational speech. *Journal of Phonetics*, *22*, 493–511.

Holst, T., & Nolan, F. (1995). The influence of syntactic structure on [s] to [ʃ] assimilation. In B. Connell, & A. Arvanti (Eds.), *Papers in laboratory phonology IV* (pp. 315–333). Cambridge, UK: Cambridge University Press.

Janse, E., Nooteboom, S., & Quené, H. (submitted). Coping with gradient forms of [t]-deletion and lexical ambiguity in spoken-word recognition.

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

Jongenburger, W., & van Heuven, V. J. (1993). Sandhi processes in natural and synthetic speech. In V. J. van Heuven, & L. C. W. Pols (Eds.), *Analysis and synthesis of speech, strategic research towards high-quality text-to-speech generation* (pp. 261–276). Berlin: Mouton de Gruyter.

Kaisse, E. (1985). *Connected speech: The interaction of syntax and phonology*. Orlando: Academic Press.

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

Kemps, R., Ernestus, M., Schreuder, R., & Baayen, R.H. (in press). Prosodic cues for morphological complexity: The case of Dutch noun plurals.

Klatt, D. H. (1974). The duration of [s] in English words. *Journal of Speech, Language, and Hearing Research*, *17*, 51–63.

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, *67*, 971–995.

Kluender, K. R. (1994). Speech perception as a tractable problem in cognitive science. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 173–217). San Diego, CA: Academic Press.

Kohler, K. (1995). *Einführung in die Phonetik des Deutschen [Introduction to German Phonetics]*. Regensburg: Erich Schmidt Verlag.

Ladefoged, P. (1989). A note on 'Information conveyed by vowels'. *Journal of the Acoustical Society of America*, *85*, 2223–2224.

Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America*, *29*, 98–104.

Lahiri, A., & Marslen-Wilson, W. (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*, Vol. 7 (pp. 637–676). Berlin: Mouton de Gruyter.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

Levelt, W. J. M., & Indefrey, P. (2004). The spatial and temporal signature of word production components. *Cognition*, *90*, 101–144.

Liberman, A. M. (1996). *Speech: A special code*. Cambridge, MA: MIT Press.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H & H theory. In W. J. Hardcastle, & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403–439). Dordrecht: Kluwer.

Lindblom, B., & Studdert-Kennedy, M. (1967). On the role of formant transitions in vowel recognition. *Journal of the Acoustical Society of America*, *42*, 830–843.

Lively, S. E., Pisoni, D. B., & Goldinger, S. D. (1994). Spoken word recognition: Research and theory. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 265–301). San Diego, CA: Academic Press.

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.

Luce, P. A., Goldinger, S. D., Auer, E. T., & Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception & Psychophysics*, *62*, 615–625.

Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception and Psychophysics*, *28*, 407–412.

Mann, V. A., & Repp, B. H. (1981). Influence of preceding fricative on stop-consonant perception. *Journal of the Acoustic Society of America*, *69*, 548–558.

Manuel, S. Y. (1992). *Recovery of "deleted" schwa. Perilius: Papers form the Symposium on current phonetic research paradigms for speech motor control*. Stockholm: University of Stockholm (pp. 115–118).

Manuel, S. Y. (1995). Speakers nasalize /ð/ after /n/, but listeners still hear /ð/. *Journal of Phonetics*, *23*, 453–476.

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

Massaro, D. (1998). *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge, MA: MIT Press.

Massaro, D. W., & Cohen, M. M. (1983). Integration of visual and auditory information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *9*, 753–771.

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

McQueen, J. (1996). Phonetic categorization. *Language and Cognitive Processes*, *11*, 655–664.

McQueen, J. M. (1991). The influence of the lexicon on phonetic categorization: Stimulus quality in word-final ambiguity. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 433–443.

Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics*, *25*, 457–465.

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

Mitterer, H., Csépe, V., & Blomert, L. (2003). Compensation for phonological assimilation in perception: Evidence from Hungarian liquid assimilation. *Proceedings of the 15th International Congress of Phonetic Sciences*, 2321–2324.

Näätänen, R., Schröger, E., Karakas, S., Tervaniemi, M., & Paavilainen, P. (1993). Development of a memory trace for a complex sound in the human brain. *Neuroreport*, *4*, 503–506.

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., & Winkler, I. (2001). Primitive intelligence in the auditory cortex. *Trends in Neuroscience*, *24*, 283–289.

Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin*, *125*, 826–859.

Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht, The Netherlands: Foris.

Nittrouer, S., & Studdert-Kennedy, M. (1987). The role of coarticulatory effects in the perception of fricatives by children and adults. *Journal of Speech and Hearing Research*, *30*, 319–329.

Nolan, F. (1992). The descriptive role of segments: Evidence from assimilations. In G. J. Docherty & R. D. Ladd (Eds.), *Gestures*, *segment*, *prosody* (Papers in Laboratory Phonology, Vol. 2, pp. 261–280). Cambridge: Cambridge University Press.

Nooteboom, S. (1979). Perceptual adjustment to speech rate: A case of backward perceptual normalization. In . *Anniversaries in phonetics: studia gratulatoria dedicated to Henrik Mol* (pp. 255–269). Amsterdam, The Netherlands: Institute of Phonetic Sciences, University of Amsterdam.

Nooteboom, S., & Doodeman, G. J. N. (1980). Production and perception of vowel length in spoken sentences. *Journal of the Acoustical Society of America*, *67*, 276–287.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, *23*, 299–324.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*, 204–238.

Norris, D., McQueen, J. M., Cutler, A., & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology*, *34*, 191–243.

Pierrehumbert, J. B. (2003). Probabilistic phonology: Discrimation and robustness. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probability theory in linguistics* (pp. 177–228). Cambridge, MA: MIT Press.

Phillips, C. (2001). Levels of representation in the electrophysiology of speech perception. *Cognitive Science*, *25*, 711–731.

Picton, T. W., Alain, C., Otten, L., Ritter, W. A., & Achim, A. (2000). Mismatch negativity: Different water in the same river. *Audiology and Neuro-Otology*, *5*, 111–139.

Pisoni, D. B., Nusbaum, H. C., Luce, P. A., & Slowiaczek, L. M. (1985). Speech perception, recognition and the structure of the lexicon. *Speech Communication*, *4*, 75–95.

Racine, I., & Grosjean, F. (2000). Influence de l'effacement du schwa sur la reconnaissance des mots en parole continue [The influence of schwa deletion on the recognition of words in continuous speech]. *L'Année psychologique*, *100*, 393–417.

Repp, B. H. (1983). Bidirectional context effects in the perception of VC–CV sequences. *Perception & Psychophysics*, *33*, 147–155.

Repp, B. H., Liberman, A. M., Eccardt, T., & Pesetsky, D. (1978). Perceptual integration of acoustic cues for stop, fricative, and affricate manner. *Journal of Experimental Psychologly: Human Perception and Performance*, *4*, 621–637.

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resoluation of lexical embedding in speech comprehension. *Cognition*, *90*, 51–89.

Sawusch, J. R., & Newman, R. S. (2000). Perceptual normalization for speaking rate II: Effects of signal discontinuities. *Perception & Psychophysics*, *62*, 285–300.

Sinnott, J. M., & Saporita, T. A. (2000). Differences in American English, Spanish, and monkey perception of the say-stay trading relation. *Perception & Psychophysics*, *62*, 1312–1319.

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.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, *17*, 3–46.

Stevens, K. N. (1998). *Acoustic phonetics*. Cambridge, MA: MIT Press.

Sumner, M., & Samuel, A. (2005). Perception and representation of lawful variation: The case of final /t/. *Journal of Memory and Language*, *52*, 322–338.

Utman, J. A., Blumstein, S. E., & Burton, M. W. (2000). Effects of subphonetic and syllable structure variation on word recognition. *Perception & Psychophysics*, *62*, 1297–1311.

Warner, N., & Weber, A. (2001). Perception of epenthetic stops. *Journal of Phonetics*, *29*, 53–87.

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

Watkins, A. J., & Makin, S. J. (1996). Effects of spectral contrast on perceptual compensation for spectral-envelope distortion. *Journal of the Acoustical Society of America*, *99*, 3749–3757.

Zsiga, E. C. (1994). Gestural overlap in consonant sequences. *Journal of Phonetics*, *22*, 121–140.

Zsiga, E. C. (1995). An acoustic and electropalatographic study of lexical and post-lexical palatalisation in American English. In B. Connell, & A. Arvanti (Eds.), *Papers in laboratory phonology IV* (pp. 282–302). Cambridge, UK: Cambridge University Press.