Available online at www.sciencedirect.com

ScienceDirect

Journal of Memory and Language 59 (2008) 133–152

Journal of Memory and Language

www.elsevier.com/locate/jml

How we hear what is hardly there: Mechanisms underlying compensation for /t/-reduction in speech comprehension

Holger Mitterer [a,*], Kiyoko Yoneyama [b], Mirjam Ernestus [a,c]

[a] *Max-Planck-Institut für Psycholinguistik, Postbus 310, NL-6500 AH Nijmegen, The Netherlands*
[b] *Daito Bunka University, Tokyo, Japan*
[c] *Radboud Universiteit Nijmegen, The Netherlands*

Received 5 June 2007; revision received 20 February 2008
Available online 11 April 2008

Abstract

In four experiments, we investigated how listeners compensate for reduced /t/ in Dutch. Mitterer and Ernestus [Mitterer, H., & Ernestus, M. (2006). Listeners recover /t/s that speakers lenite: evidence from /t/-lenition in Dutch. *Journal of Phonetics, 34*, 73–103] showed that listeners are biased to perceive a /t/ more easily after /s/ than after /n/, compensating for the tendency of speakers to reduce word-final /t/ after /s/ in spontaneous conversations. We tested the robustness of this phonological context effect in perception with three very different experimental tasks: an identification task, a discrimination task with native listeners and with non-native listeners who do not have any experience with /t/-reduction, and a passive listening task (using electrophysiological dependent measures). The context effect was generally robust against these experimental manipulations, although we also observed some deviations from the overall pattern. Our combined results show that the context effect in compensation for reduced /t/ results from a complex process involving auditory constraints, phonological learning, and lexical constraints.

Keywords: Speech comprehension; Phonetic reduction; Auditory processing; Phonological learning; MMN; Dutch

---

Introduction

An often overlooked difference between written and spoken language stems from the fact that what is unacceptable in written language is common practice in spoken language: the reduction and even deletion of segments (being either graphemes or phonemes). In formal writing, it is considered an error to write *posgraduate*, but this form is completely natural in spontaneous speech. Abbreviations (e.g., *e.g.* for *for instance*) occur in written language as well, but, in contrast to most reductions in spoken language, they are based on highly conscious conventions.

Different types of reduction have been documented for spoken language. High-frequency words may be strongly reduced: the Dutch word for 'actually' *eigenlijk*/ɛixələk/, for instance, may be pronounced as [ɛik] (Ernestus, 2000). Moreover, words that often co-occur may melt into a single phonological word, as, for instance, Kohler (1990) showed for the German version of the phrase 'have we' *haben wir*/habən wɪr/, which may be pronounced as [hɑmwə]. Another type of reduction occurring frequently in spoken, but not in written, words is the lenition of word-final consonants, which

---

[*] Corresponding author. Fax: +31 24 3521213.
E-mail address: holger.mitterer@mpi.nl (H. Mitterer).

0749-596X/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2008.02.004

may give rise to assimilation to a following segment (e.g., ‘wine bottle’/wɑɪn/ bɔtl/ → [wɑɪmbɔtl]) or even the apparent deletion of a segment, as in ‘perfect memory’ /pəfekt meməɹɪ/ → [pəfekmeməɹɪ], in which the /t/-release is masked by the following labial closing gesture of the /m/ (Browman & Goldstein, 1990). Due to such pronunciations in spontaneous speech, listeners have to recover the intended meaning from a much more variable input during listening than during reading.

How listeners recover the intended meaning from reduced forms has been a major focus of research in psycholinguistics, especially during the last decade, bridging the gap between the fields of spoken-word recognition and speech perception (e.g., Ernestus, Baayen, & Schreuder, 2002; Gaskell & Marslen-Wilson, 1996, 1998, 2001; Gow, 2001, 2002; Gow & Im, 2004; Hura, Lindblom, & Diehl, 1992; Kemps, Ernestus, Schreuder, & Baayen, 2004; Lahiri & Reetz, 2002; Mitterer & Blomert, 2003; Mitterer, Csépe, Blomert, 2006; Mitterer, Csépe, Honbolygo, & Blomert, 2006; Mitterer & Ernestus, 2006). Most studies focused on the perception of assimilated forms and converged on two findings.

First of all, segmental context plays a crucial role in the recognition of assimilated forms, as an assimilated form is only recognized in contexts that actually allow the assimilation to occur. For instance, the assimilated form *wime* of *wine* only occurs if the following word starts with a bilabial consonant (e.g., *bottle*) but not if this word starts with a velar consonant (e.g., *glass*). Gaskell and Marslen-Wilson (1996, 1998) were the first to show that perception mirrors this context effect in production and *wime* is recognized as *wine* only in the bilabial context *…bottle* but not in the velar context *…glass* (see also Gow, 2003; Mitterer & Blomert, 2003).

A second finding with regard to assimilated forms is that phonetic detail plays an important role. Gow (2002, 2003) showed that there are subtle acoustic differences between a bilabial sound resulting from assimilation—as in *gum production* meaning *gun production*—and an intended bilabial—as in *gum production* actually meaning *gum production*: the acoustic evidence for a bilabial nasal is stronger in case of an intended labial than in case of an assimilated labial. Listeners are sensitive to these subtle differences, and only assume the presence of an underlying coronal segment in a homo-organic labial cluster (such as [mb]), if the cues for labiality are weak in the first segment.

Compared to the now extensive literature on assimilation, less work has been dedicated to the perception of forms in which segments are (apparently) missing. Manuel (1992) investigated the perception of English words in which schwa has been severely reduced (*support*/səpɔɹt/ → [spɔɹt]). Production data of this phenomenon indicated that, while the glottal gesture for the schwa is reduced so that no vocal-fold vibration occurs, the oral gesture often remains, leading to subtle acoustic differences between the forms of [spɔɹt] meaning either *support* or *sport*, that listeners exploit to disambiguate realizations such as [spɔɹt].

Mitterer and Ernestus (2006) reached a similar conclusion for word-final /t/ in Dutch, which, according to previous studies (e.g., Ernestus, 2000), is frequently deleted in connected speech. Like Manuel (1992), they first investigated the phonetic detail associated with the reduction of word-final /t/. Their production and corpus studies indicated that “deletion” is not the correct term to describe what often happens to word-final /t/ in connected speech. A supposedly deleted /t/ often leaves behind two residual cues to its underlying presence (see also Browman & Goldstein, 1990): first, the presence of a closure or low-amplitude frication between the preceding and following segment (e.g., between the /s/ and the /b/ in /mɛst bəstɛlt/ *mest besteld* ‘fertilizer ordered’) and, second, a shorter duration of the preceding segment (the /s/ tends to be shorter in /mɛst bəstɛlt/ than in /mɛs bəstɛlt/ *mes besteld* ‘knife ordered’, independent of the realization of the /t/).

Mitterer and Ernestus (2006) also observed that /t/-reduction, like assimilation, is conditioned by segmental context. Confirming the picture arising from studies of /t/-deletion in Germanic languages (Ernestus, 2000; Ernestus, Lahey, Verhees, & Baayen, 2006; Grimson & Cruttenden, 1994; Guy, 1980, 1992; Kohler, 1995), reduction of /t/ is most likely to occur before bilabial consonants and after the alveolar fricative /s/, so that the /t/ is likely to be reduced in /mɛst bəstɛlt/. Even in this context, however, /t/-reduction is optional, and an unreduced segment with a [t]-release may be observed. This underscores the inherent variability of /t/-reduction and its substantial contribution to the invariance problem.

In a series of perception experiments, Mitterer and Ernestus (2006) found both aspects of /t/-reduction—the presence of residual cues and the role of segmental context—to influence listeners’ recovery of reduced word-final /t/. First of all, listeners infer the presence of an underlying /t/ on the basis of residual cues in the acoustic signal. Second, there is an effect of segmental context on the interpretation of these residual cues. A given cue (e.g., a short silence) is more likely to trigger the perception of an underlying /t/ if the preceding segment is /s/ than if it is /n/. This context effect is beneficial for speech comprehension, because /t/ is more likely to be reduced after /s/ than after /n/.

In addition, Mitterer and Ernestus (2006) have shown that lexical processing contributes to compensation for /t/-reduction. Dutch listeners tend to infer the presence of a word-final /t/ more often if this “generates” an existing word (e.g., the English “fros” + “t” → “frost”) than if it does not (“blouse” + “t” → nonword). Importantly, lexical restoration does not suffice to explain all aspects of compensation for /t/-reduction. First of all, the segmental context effect also occurs for nonwords. More importantly, Dutch listeners tend to infer a /t/ more often after the fragment "moeras" than after the fragment "charman", even though a purely lexical-restoration account would predict the opposite, as "moeras" and "charmant" are words in Dutch while "moerast" and "charman" are not. This shows that the segmental context has more leverage than lexical restoration.

The purpose of this paper is to investigate the mechanisms that drive the context effect in compensation for /t/-reduction. Note that this context effect is similar in nature to the context effect found for compensation for assimilation. In both cases, listeners "mirror" in perception what happens in production. If, in production, a given segment is more likely to be altered in Context A than in Context B, listeners are more likely to "perceive" the original, underlying identity of that segment in Context A than in Context B. For the case of assimilation, an [m] is more likely to be perceived as /n/, if it is followed by /b/ than if followed by a non-bilabial segment; and for the case of /t/-reduction, an underlying /t/ is more likely to be perceived as present after /s/ than after /n/. Given the similarity between the context effects in compensation for assimilation and for /t/-reduction, we consider the viability of the mechanisms that have been proposed for compensation for assimilation.

Two accounts for compensation for assimilation make extensive use of the term *parsing*. They assume that the properties of the acoustic signal are parsed in a non-linear fashion and attributed to the underlying segments. Gow (2001, 2002, 2003) assumes that the stream of incoming phonetic feature cues is parsed according to Gestalt laws (cf. Bregman, 1990), while a direct-perception account assumes that the acoustic signal is parsed along gestural lines (Fowler, 1996; Fowler & Smith, 1986). These accounts explain well the context effect in compensation for assimilation: before a bilabial, the labiality of an assimilated, underlyingly alveolar segment (as the [m] in *wime bottle*) can be parsed to this bilabial context, which leads to compensation for assimilation. Because a following velar (as in *wime glas*) is not able to host labiality, an [m] is not interpreted as /n/ before /g/. While parsing is conceptually well able to account for compensation for assimilation, it is difficult to see how parsing can explain the role of context in compensation for /t/-reduction. A reduced /t/ does not carry any cues that belong to its surrounding segments, but rather carries too few cues for its own presence. So rather than parsing cues *from* the reduced segment and assigning them to the context, the listener would have to parse cues from the context *to* the reduced segment. It is difficult to see how /s/ and /n/ could provide *differential* cues to the presence of a following /t/, as they are both alveolar.

However, the gestural parsing account may also explain the observed compensation if the pronunciation may result from gestural overlap. Gestural overlap may be responsible for place assimilation in an /nb/ sequence (*gardenbench* → *gardembench*), and accordingly, perceptual compensation can be explained in gestural terms. The tongue-tip gesture of the /n/ overlaps with the labial gestures of the /b/, resulting in a phoneme that is similar to [m]. If listeners would parse the speech signal along gestural lines, compensation for assimilation occurs by attributing or parsing the evidence for the labial closing gesture to the /b/ (but see Mitterer, Csépe, Blomert, 2006, for problematic empirical data). In contrast, the higher likelihood of /t/-reduction after /s/ than after /n/ cannot easily be explained in gestural terms. In both contexts, the tongue-tip is in (nearly) the correct position for an alveolar closure, and it is difficult to see why an identical amount of articulatory overlap would give rise to more acoustic reduction of /t/ in /st/-clusters than in /nt/-clusters. In conclusion, it is unclear how parsing in terms of features or speech gestures might explain the context effect in compensation for /t/-reduction. Accordingly, we will not pursue these accounts here further.

Am different account of compensation for assimilation is the phonological-inference account as proposed by Gaskell (2003). According to this framework, listeners learn which reductions occur in which segmental contexts, and then apply this knowledge "in reverse" in perception. Importantly, this account also assumes that the speech stream is first categorized probabilistically independently of the phonological context. Only at a later stage, context is taken into account. Applied to compensation for /t/-reduction, the first stage would recognize the preceding context (e.g., as /s/ or /n/), but the subtle cues in the signal for a reduced /t/ would not suffice for an unambiguous recognition of /t/. Then, phonological inference would be applied to raise the activation level of the /t/ in case the preceding segment is /s/ rather than /n/, because the perceiver has learned that /t/ is more likely to be reduced after /s/ than after /n/. Phonological inference is a possible candidate to account for the context effect in compensation for /t/-reduction.

Obviously, a prediction of this model is that language experience is crucial to achieve compensation. In line with this assumption, Beddor, Harnsberger, and Lindemann (2002) showed that the degree of compensation for vowel-to-vowel coarticulation is language-specific: Shona listeners compensate more strongly than do English listeners, even when listening to the same stimuli, which is in accordance with the fact that vowel-to-vowel coarticulation is stronger in Shona than in English. Similarly, Darcy, Peperkamp, and Dupoux (2007) found that listeners compensate more strongly for assimilations that occur in their native languages than for assimilations that do not. We will investigate the validity of the two crucial assumptions of the model—necessity of learning and a two-step categorization—for compensation of /t/-reduction in this paper.

Yet another possibility is that the context effect in compensation for assimilation just results from human general auditory processing. Auditory processes play an important role in theories of speech production that allow perception to influence production (see, e.g., Boersma, 1998; de Boer, 2000; Hume & Johnson, 2001; Hura et al., 1992; Kohler, 1990; Lindblom, 1990; Mitterer, Csépe, Blomert, 2006; Mitterer, Csépe, Honbolygo, et al., 2006; Ohala, 1990; Schwartz, Boë, Valleé, & Abry, 1997; Steriade, 2001). In this view, listeners and speakers have conflicting interests: listeners prefer unreduced word forms, which facilitate word recognition, while speakers like to reduce as much as possible in order to minimize articulatory effort (Lindblom, 1990). The interplay of these conflicting interests lead to a compromise in which only those reductions are introduced by speakers that do not lead to perceptually salient differences between unreduced and reduced forms. The salience of a difference is at least partly determined by language-independent auditory processes. An auditory account therefore predicts that the context effect in the comprehension of reduced forms does not necessarily result from language experience.

Mitterer and colleagues (Mitterer, Csépe, Blomert, 2006; Mitterer, Csépe, Honbolygo, et al., 2006) provided empirical evidence that the context effect in compensation for assimilation is at least partly based on auditory processing independent of language experience (but see also Darcy et al., 2007). In addition, it has been argued that auditory processing also plays an important role in the compensation for other connected speech processes, such as coarticulation (Lotto, Kluender, & Holt, 1997) and /h/-deletion in Turkish (Mielke, 2003). Based on these findings, an auditory source for the context effect in compensation for /t/-reduction is not unlikely. Moreover, the acoustic properties of the segments /n/, /s/, and /t/ also make an auditory account plausible. The segments /t/ and /s/ are spectrally more similar than /t/ and /n/. Both /t/ and /s/ lack a harmonic structure—due to the absence of vocal-fold vibration—and have energy predominantly in higher-frequency bands (>3 kHz, depending on the speaker). The nasal /n/, in contrast, has a harmonic structure and the nasal cavity generates a “zero”, which filters out higher frequencies, so that the acoustic energy is concentrated in lower-frequency bands (<1 kHz). So the /t/ is more salient after /n/ than after /s/. Any variation in the realization of /t/, such as reduction, will hence be less salient after /s/ than after /n/, and this would lead listeners to treat reduced and full forms of /t/ as more similar after /s/ than after /n/. In other words, according to an auditory account, reduction of /t/ would be less salient after /s/ than after /n/. Note that this explanation suggests that the higher likelihood of /t/-reduction after /s/ than after /n/ is in fact an adaptation of the speaker to the listener (Steriade, 2001), while the phonological-inference model assumes that it is the listener who adapts to the speaker.

Both the phonological-inference account and the auditory account hold promise to explain the context effect in compensation for /t/-reduction. In this paper, we will compare these two accounts. (We will consider another possibility, an episodic account, in the General Discussion.) Importantly, the auditory account assumes that the context effect arises in early and automatic auditory processing, and thus should be resistant against experimental manipulation. In contrast, the phonological-inference account assumes that a first categorization of the speech signal is context-independent and the subsequent re-classification depends on language experience. This implies that a phonological-inference account should be preferred over the auditory account, if it is possible to find conditions under which the context effect disappears.

Since the phonological-inference account predicts the absence of a context effect if listeners base their “responses” on the early, context-insensitive categorization of the speech input, we will start by trying to create situations which favor access to these early auditory or phonetic representations. In Experiment 1, we tried to generate such a situation with the same task as used in Mitterer and Ernestus (2006): a two alternative forced choice task (2AFC). In order to maximize the likelihood of finding a context effect, Mitterer and Ernestus varied the context in which a reduced /t/ appeared—the target word as well as the carrier sentence—from trial to trial. Here, we minimized the variation in the context by keeping it constant within an experimental block. This should allow listeners to focus their attention on the acoustic–phonetic details of the reduced /t/s and ignore the phonological context.

Experiment 1

In most phonetic and psycho-acoustic experiments, context is minimal and target words differ minimally. For instance, in the seminal study by Mann (1980) on “compensation for coarticulation,” only two syllables were presented on a trial; the first syllable was always either [al] or [ar], while the second syllable was a member of a [da]-[ga] continuum. In Experiment 1, we tested whether the context effect in compensation for /t/-reduction can also be found when the task has as little variation as Mann’s (1980) study. In this case, listeners might be better able to focus on the relevant part of the acoustic signal, because it is more predictable where in the stimulus the critical information is occurring. Moreover, any acoustic variation might be more salient if presented in constant context.

Methods

Participants

Eighteen (14 female) members of the Max-Planck Institute's subject pool participated in the experiment. All participants were native speakers of Dutch, lived in the Netherlands, and were between 19 and 29 years of age (median: 22). Fourteen were right-handed, four left-handed.

Materials

The materials used in this experiment formed a subset of the synthesized materials used in the experiments by Mitterer and Ernestus (2006). The three sequences [dri], [bla], and [spe] were concatenated with either an [n] or an [s] plus a member of a [t]-∅ series and were followed by another syllable. The [t]-∅ series consisted of five steps, all with a duration of 65 ms. The first step was a signal similar to a canonical [t] with a 35 ms closure and a 30 ms transient-frication sequence. The second step was a 65 ms frication noise, as often found after [s] in /st/ codas. The third step was spectrally the same as the second signal, but its amplitude had been reduced to 20% of the original amplitude (=−14 dB). The fourth step was a 65 ms interval of silence, simulating a closure. For the fifth step, the preceding consonant was elongated by 65 ms and was followed directly by the onset of the next syllable. The three carriers [dri], [bla], and [spe] crossed with the two segmental contexts ([n] or [s]) and the five target signals gave rise to 30 target words. These target words had a duration of 275, 310, and 355 ms for the [dri], [bla], and [spe] carriers, respectively. Note that none of these CCV carriers form existing Dutch words if concatenated with an [n], [nt], [s], or [st] coda.

Three different contexts followed these thirty target words. They were the first syllables of the adverbs *krachtig* [krɑx-təx] 'forcefully,' *prima* [pri-ma] 'nicely,' *moeilijk* [muj-lək] 'with trouble'. These syllables had a duration between 160 and 180 ms. The complete stimuli, presented in Table 1, varied in duration from 435 ms to 535 ms.

Procedure

The experiment was run on a standard PC with the NESU package (Wittenburg, Nagengast, & Baumann, 1998). Participants were wearing headphones and looked at a computer screen with a two-button response box in front of them. They were asked to press the right button if the target word contained a [t] and the left button if this word did not contain a [t].

The trials had the following structure. After 150 ms of blank screen, the orthographic transcriptions of the two response alternatives for the target word (e.g., "dris" and "drist") were presented in the upper left and right corner of the computer screen. After another 450 ms, the target word followed by the context syllable was played. From the onset of the target word, participants had 2.5 s to choose one of the two orthographic representations. After responding, this chosen orthographic representation was moved further to the upper right or left corner while the other one was removed from the screen. In case of a time-out error, a stopwatch was shown to remind participants to respond faster. The feedback signal, which indicated either which response had been registered or that no response had been registered, stayed on the screen for 1 s before the next trial began.

A given participant was presented with only one carrier (e.g., [dri…]) and one following context (e.g., [krɑx]). For a given stimulus, the carrier was combined with either [n] or [s] (the context preceding the target signal) and one of the five different target signals from the [t]-∅ series to generate the target word (see Table 1). This gave rise to ten different stimuli to be presented to a given subject, each of which was presented ten times, leading to 100 trials for every participant. The trials with the same preceding segmental context ([n] or [s]) were blocked, so that for the first 50 trials, the participant had to decide, for instance, whether she heard [drint] or [drin], while in the second part of the experiment, after a short break, she had to decide whether she heard [drist] or [dris], or the other way round, as the order of [n] and [s] was counterbalanced across subjects.

Table 1
Stimuli used in Experiment 1

Carrier	Preceding Context	Target signal	Following Context
		Canonical [t]	
dri [drɪ]	n	[t]-Frication	krach [krɑx]
bla [blɑ]		Weak[t]-frication	pri [prɪ]
spe [spɛ]	s	Silence	moei [muɪ]
		Elongated consonant	

Design

There were three independent variables. The first was the nature of the Target Signal at the end of the target word (canonical [t], [t]-frication, weak [t]-frication, silence, elongated consonant). The second independent variable was the Preceding Context of the Target Signal (/n/ versus /s/). Note that this segment is the penultimate phoneme of the target word. These two independent variables varied within subjects. The third independent variable was the Following Context, the initial consonant of the following adverb (/k/, /p/, or /m/). This was a between-subjects variable. The dependent variable was the percentage of [t]-responses derived from ten repetitions of a given stimulus for a given subject.

Results and discussion

Fig. 1 shows the mean percentages of [t]-responses in all conditions, and Table 2 the ANOVA over these data. There is a clear effect of the preceding context, which is maximal for the more ambiguous members of the [t]-∅ series (see the effect sizes and the confidence intervals in Fig. 2). These results replicate the findings by Mitterer and Ernestus (2006), despite the procedural differences between the current and those earlier experiments.

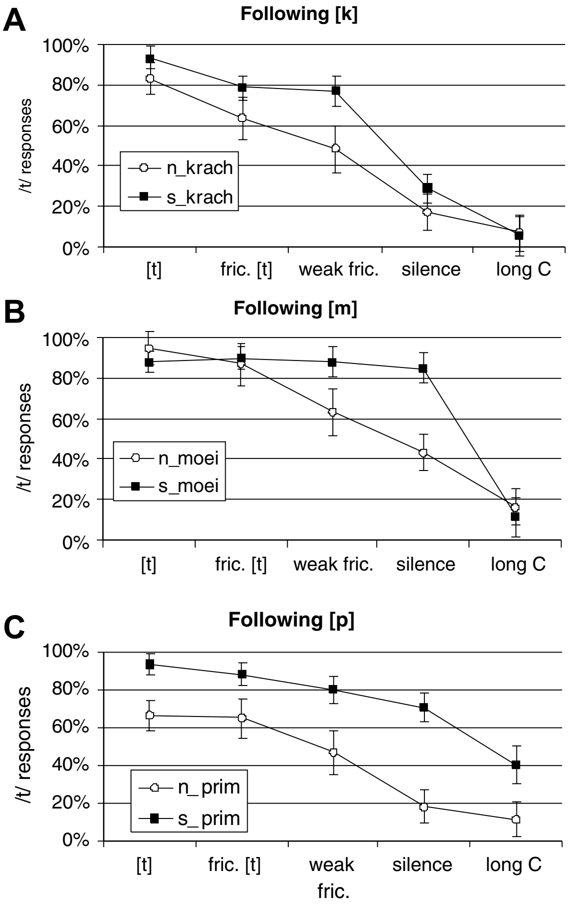


Fig. 1. Percentage of /t/-responses for the five-step /t/-∅ series (*x*-axis of the graphs) as a function of the preceding context (symbols) and following context (different panels). "Fric." stands for Frication and "C" for consonant.

Moreover, we also found an effect of following context that interacted with the target signal. We analyzed this interaction by testing the effect of following context on each level of the factor Target signal using two linearly independent contrasts for following context, one pooling the two bilabial following contexts against the velar following context, and one testing between the two bilabial contexts. Fig. 2 shows that these contrasts were significant only on the fourth level of the Target-Signal factor (silence), where an underlying /t/ was inferred most often before [m] (64.2%), less so before [p] (44.4%), and least of all before [k] (23%). This result mirrors findings in production and corpus studies that /t/ is more likely to be reduced before bilabials than before velars, and that, before a bilabial, reduction often results in just a closure (Ernestus et al., 2002, 2006; Mitterer & Ernestus, 2006). Hence, this context effect in perception is functional, as it undoes reduction occurring in production.

The difference in /t/-responses between [m] and [p] can be explained by the acoustic characteristics of these two bilabials. The oral closure for the stop [p] leads to a silence in the acoustic signal. The oral closure for [m], in contrast, does not lead to silence, neither after [n] nor after [s], due to air escaping via the nasal passage. This suggests that the presence of a silent closure in sequences such as [blɑnm] or [blɑsm] is a stronger cue for the presence of a voiceless stop (/t/) than a silent closure in sequences such as [blɑnp] or [blɑsp].

In summary, the current experiment showed that segmental context effects in the perception of word-final /t/ can also be found if the target signals appear in fixed two-syllable carrier phrases. Even in tasks with little

Table 2
ANOVA summary table for the data obtained in Experiment 1 ($^*p < .05$)

Source	*df*	SS	MS	*F*
Following Context	2	0.844	0.422	5.1^*
Subjects (S)	15	1.246	0.083	
Target signal	4	12.035	3.009	73.7^*
Target Signal × Following Context	8	0.840	0.105	2.6^*
S × Target Signal	60	2.450	0.041	
Preceding Context	1	1.630	1.630	25.1^*
Preceding Context × Following Context	2	0.425	0.213	3.3
S × Preceding Context	15	0.976	0.65	
Target Signal × Preceding Context	4	0.538	0.135	4.1^*
Target Signal × Preceding Context × Following Context	8	0.283	0.035	1.1
S × Target Signal × Preceding Context × Following Context	60	1.951	0.033	

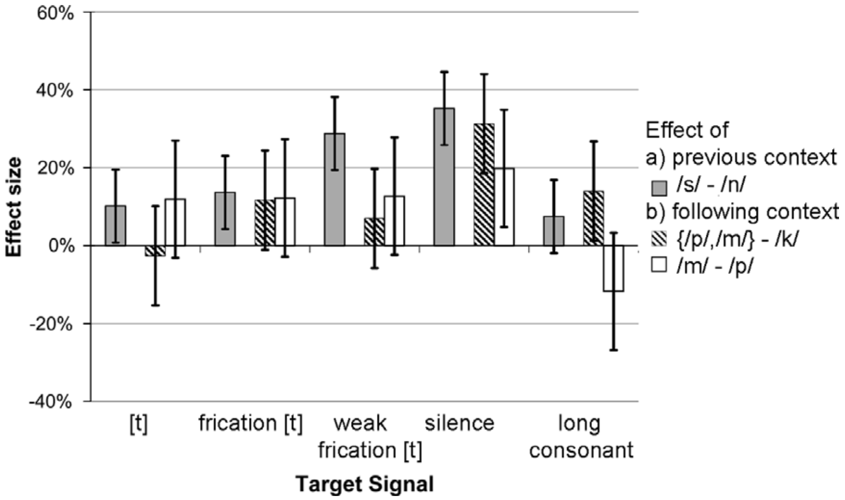


Fig. 2. Effect sizes and confidence intervals for the previous and following context on all levels of the five-step /t/-∅ series.

stimulus variation, listeners are influenced by segmental context and are more likely to infer an underlying /t/ after [s] than after [n]. Hence, the effect of segmental context observed in compensation for /t/-reduction appears stable with regard to task demands.

Experiment 2

Experiment 1 showed that the segmental context effect in compensation for /t/-reduction is rather robust to experimental manipulations in an identification task. In Experiments 2 and 3, we have undertaken another attempt to make the context effect disappear. We used a discrimination task here, following the example set by several other researchers (Beddor & Krakow, 1999; Fitch, Halwes, Erickson, & Liberman, 1980; Kingston & Macmillan, 1995; Macmillan, Kingston, Thorburn, Dickey, & Bartels, 1999), arguing that the discrimination task reveals processing at an earlier level than the identification task. If the context effect occurs at early auditory levels, it should be observed in a discrimination task as well. If, however, the context effect arises at a higher, phonological, processing level, context should not affect discrimination performance.

Obviously, this kind of argument is valid only if discrimination performance is not influenced by identification performance. The relation between identification and discrimination has been the topic of a lively debate in the field of speech perception (see the volume by Harnad, 1987), as correspondences in identification and discrimination performance were one of the basic pillars in the first formulations of Motor Theory (Liberman, 1957). Other researchers (Massaro, 1987) later suggested that these correspondences were due to a ‘‘phonological-recoding’’ strategy in the discrimination task, to reduce the memory load.

Recently, Gerrits and colleagues (Gerrits & Schouten, 2004; Schouten, Gerrits, & van Hessen, 2003) have suggested that discrimination performance is indeed independent of identification performance in certain discrimination tasks, which thereby seem to prevent phonological-recoding strategies. One such task is the four-interval oddity task (4I-oddity), in which four stimuli are presented of which three are identical (the standard stimuli) and one is deviant (the deviant stimulus). The deviant always occurs at either the second or the third position in the sequence of four. The task of the subject is to indicate whether the deviant is the second or the third stimulus. Given the potential dissociation of identification and discrimination performance in this task, we used it in our own experiments.

If compensation for /t/-reduction is a consequence of learning at a later, phonological, level, we predict that discrimination performance may not be influenced by context, but only depend on the acoustic differences between the standard and the deviant. That is, there may be a dissociation between identification and discrimination performance.

In contrast, if the effect of segmental context on the perception of reduced /t/ is a consequence of auditory processes, we expect that discrimination performance is similarly affected by context as identification performance in Experiment 1, because auditory effects are reflected in both discrimination and identification performance: stimuli with reduced /t/ should be more readily discriminated from stimuli with canonical /t/ after [n] than after [s].

The situation is less straightforward for the discrimination of stimuli with reduced forms of [t] from stimuli with a long consonant, which are, independently of the identity of this consonant, perceived as containing no /t/. Clearly, discrimination of reduced /t/ after [s] from a long [s] should be easy, given that these stimuli are already categorized differently in Experiment 1, and different categorization presupposes discriminability. After [n], reduced variants of /t/ are easy to identify, as we argued in the Introduction, because of the acoustic differences between /n/ on the one hand and full and reduced forms of /t/ on the other hand. Reduced /t/s after /n/ are therefore sufficiently different from a canonical /t/, but, because of their acoustic salience, they are also clearly different from a stimulus with just a long [n]. Hence, we do not expect a context effect for pairs with a long consonant.

Methods

Participants

Fifteen (of which 14 were female) members of the Max-Planck institute's subject pool participated in the experiment. All participants spoke Dutch as their native language and lived in the Netherlands. Two were left-handed, the rest right-handed. The participants were between 17 and 26 years of age (median: 23). None of them had participated in Experiment 1.

Materials

The same six target words as in Experiment 1 were used, that is, 'spes,' 'spen,' 'dris,' 'drin,' 'blas', and 'blan' followed by one of the five signals forming the /t/-∅ series. Each target word was again followed by the first syllable of each of the three context words 'krachtig' *forcefully*, 'moeizaam' *with trouble*, or 'prima' *nicely*. These 90 stimuli had a duration of 450–540 ms.

Procedure

Participants were seated in front of a computer screen and a two-button response box. They first read the instructions on the screen. Participants were instructed that they would hear a series of four stimuli on every trial, in which either the second or the third stimulus differed from the other stimuli. They were asked to indicate which stimulus they thought was the deviant. If they thought that the second stimulus was deviant, they were to press the left button. If they thought that the third stimulus was deviant, they were to press the right button. Participants were explicitly instructed that stimuli that should be spelled identically may still differ in how they sound.

A trial started with 250 ms of blank screen. Then the two response alternatives were presented: the digit '2' in the upper left quadrant of the screen and the digit '3' in the upper right quadrant of the screen, corresponding to the response key allocation. After another 250 ms, the sequence of four sounds started, with an interstimulus interval of 900 ms. From the offset of the fourth speech sound, participants had 2.5 s to respond. In case of a response, feedback indicated whether the choice was correct or not. If no response was given, a feedback screen asked participants to respond faster.

The experiment started with four trials with pure tones of 300 and 400 Hz as standard and deviant in order to familiarize the participants with the task. Then, the 324 trials (see the Design section for how we arrived at this number) with speech materials were presented, with a short break after every 50 trials. The standard and deviant of a trial differed only in the realization of the /t/. Between trials, the carrier ([blɑ…], [drɪ…], or [spɛ…]), the preceding context ([n] or [s]), and the following context ([muɪ], [pri], or [krɑ]) varied.

Design

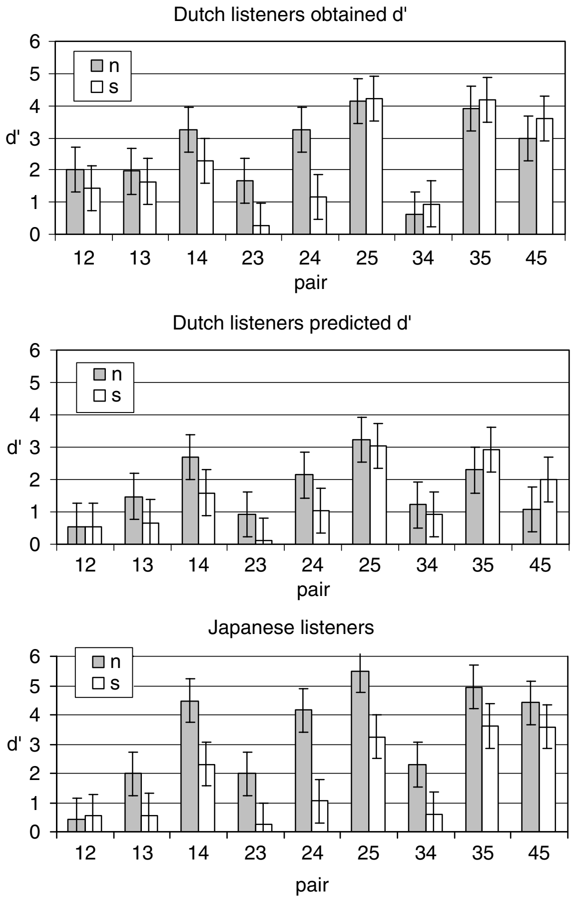
There were two independent variables in this experiment. The first independent variable was Pair, indicating the realization of /t/ in the standard and the deviant. We will refer to pairs by assigning numbers to the different steps of the /t/-∅ series (1 = canonical /t/, 2 = frication /t/, 3 = weak frication /t/, 4 = silence, 5 = long coda consonant). Given that the extremes of this series were well-separated in both contexts in Experiment 1, we did not use the four-step pair [1–5]. This left us with nine pairs: two three-step ([1–4], [2–5]), three two-step ([1–3], [2–4], and [3–5]) and four one-step ([1–2], [2–3], [3–4], and [4–5]) pairs. In all trials, the less reduced form was the standard and the more reduced form the deviant. The second independent variable was Preceding Context, with two levels, [n] and [s].

Each participant completed 18 trials for each of the 18 (9 pairs * 2 contexts) cells of this design. For a given cell (e.g., pair [1–2] in the [n]-context), these 18 trials arose from presenting the relevant pair twice with each of the nine combinations of the three carriers ([blɑ…], [drɪ…], [spɛ…]) and the three following contexts ([muɪ], [pri], [krɑ]). The dependent variable d' was calculated from these 18 trials.

Results and discussion

The mean d' values and their confidence intervals are displayed in the upper panel of Fig. 3, while Table 3 provides a summary of the ANOVA. The figure shows that the predicted context effect was significant only for a subset of the pairs. The full /t/, for instance, was more difficult to discriminate from the reduced form with just a silence (pair [1–4]) in /s/- than in /n/-context, but there was no context effect in the other pairs with full /t/ (pairs [1–2] and [1–3]).

The context effects observed in the discrimination task are partially in agreement with the assumption that


Dutch listeners obtained d'

Dutch listeners predicted d'

Japanese listeners

Fig. 3. Mean discrimination performance for canonical and reduced /t/s after [s] and after [n] in Experiments 2 and 3. The upper panel shows the results obtained from Dutch listeners, the middle panel the predicted performance for Dutch listeners on the basis of the identification results of Experiment 1, and the lower panel the results from Japanese listeners.

Table 3
ANOVA summary table for the data obtained in Experiment 2 ($^*p < .05$)

Source	*df*	SS	MS	*F*
Subjects (S)	14	241.0	17.2	
Context	1	11.1	11.1	3.85
S × Context	14	40.5	2.9	
Pair	8	363.0	45.4	18.35*
S × Pair	112	277.0	2.5	
Context × Pair	8	44.2	5.5	2.85*
S × Context × Pair	112	217.1	1.9	

there is an auditory basis to the context effect in compensation for /t/-reduction. As predicted, the discrimination of pairs [1–4], [2–3], and [2–4] is more difficult after [s] than after [n]. A context effect was nevertheless also predicted for pairs [1–2], [1–3], and [3–4]. The absence of a context effect for pair [3–4] can be attributed to a floor effect, because discrimination is very difficult after [n] already. The absence of a context effect for pairs [1–2] and [1–3] is, however, not easily explained. It shows that auditory processing alone cannot account for the context effect as observed in an identification task. This view is supported by an earlier finding (Mitterer & Ernestus, 2006) that higher-level constraints, such as lexical status, influence compensation for /t/-reduction more strongly than compensation for phonological assimilation, a process with a strong auditory component.

There are two possibilities to explain the current results within the framework of phonological inference. First of all, the listeners may have used a strategy of phonological recoding, at least to some degree, which resulted in the context effect for some pairs. The 4I-oddity task appears to discourage a phonological-recoding strategy for “short” stimuli, such as a single vowel or a stop-vowel syllable (Gerrits & Schouten, 2004; Schouten et al., 2003). Our stimuli are, however, much longer than that, leading to a longer trial time, which may promote a strategy of phonological recoding in order to reduce memory load. Moreover, the reactions in a discrimination task—and most other behavioral tasks—are relatively late with respect to the early perceptual processes at the focus of this study. This may leave ample time for phonological recoding to influence these overt responses.

The fact that the discrimination results do not seem to follow the identification results in all respects, however, is not in line with a phonological-recoding account of the current data. To test this dissociation of identification and discrimination more stringently, we transformed the identification scores from Experiment 1 to discrimination scores (following Macmillan & Creelman, 1991, pp. 211–213) and statistically compared these predicted discrimination scores from Experiment 1 with the obtained discrimination scores from Experiment 2. If participants applied a phonological-recoding strategy, the predicted and obtained discrimination scores should be similar. In contrast, the analysis (see Table 4 and the comparison between the upper and the middle panel of Fig. 3) revealed a main effect of Task—with obtained discrimination performance being better than predicted ($d' = 1.58$, observed $d' = 2.50$, CI = 0.84)—and also an interaction of Task with Pair. This indicates that phonological recoding may play only a minor role in the 4I-oddity task (confirming the results of Gerrits & Schouten, 2004). This conclusion depends, however, on the validity of the conversion of identification scores to d' by z-transformation. One of the underlying assumptions is that the differences between stimuli are uni-dimensional. This is the case for the often used F2 continua ranging from [ba] to [da], but it is not for our /t/-∅ series.

Table 4
ANOVA summary table for the comparison of Identiciation (Experiment 1) and Discrimination (Experiment 2) performance by Dutch Listeners ($^{*}p < .05$)

Source	*df*	SS	MS	*F*
Task	1	124.11	124.11	9.72^{*}
Task × S	31	396.00	12.77	
Context	1	18.51	18.51	4.60^{*}
Context × Task	1	0.39	0.39	0.10
Context × S	31	124.66	4.02	
Pair	8	556.73	69.59	28.58^{*}
Pair × Task	8	51.39	6.42	2.64^{*}
Pair × Task × S	248	604.12	2.44	
Context × Pair	8	72.75	9.09	6.03^{*}
Context × Pair × Task	8	10.45	1.31	0.87
Context × Pair × Task × S	248	374.26	1.51	

There is yet another possible explanation for the results of Experiment 2 within the framework of phonological inference. The poor discrimination performance for canonical and reduced /t/ after [s] may be due to an *acquired similarity*: Dutch listeners may have learned that reduced /t/ has to be treated similarly to canonical /t/ after [s], and this acquired knowledge may interfere with their ability to distinguish such realizations. An interpretation of the current results in terms of acquired similarity is supported empirically by Guenther, Husain, Cohen, and Shinn-Cunningham (1999). They showed that, if listeners receive categorization training for a variety of stimuli, the ability to discriminate between stimuli belonging to the same category decreases significantly. Further support for acquired similarity comes from the countless demonstrations that language experience leads to a decreased ability to discriminate between speech sounds which are allophones in one's native language (for a review, see Cutler & Broersma, 2005).

We tested the possible role of *acquired similarity* in Experiment 3, which is an exact replication of Experiment 2, except that the participants in Experiment 3 were monolingual native speakers of Japanese. Japanese does not allow consonant clusters in syllable coda, and this precludes monolingual Japanese listeners from acquiring experience with /t/-reduction in /st/- and /nt/-coda clusters. If the context effect observed in the previous experiments is at least partly due to general auditory processing, Japanese participants should also find it harder to differentiate between different reduced forms of /t/ after /s/ than after /n/. If phonological learning contributes to the context effect, we expect the context effect to be smaller or even absent for Japanese in comparison to Dutch listeners.

This experiment also investigates the role of phonological recoding in the discrimination task we used in Experiment 2. Japanese listeners cannot recode coda clusters in their native phonology, because complex codas do not occur in Japanese.[1] Therefore, if they show the same pattern as the Dutch listeners, this would make it unlikely that the context effect in discrimination performance observed in this experiment, and Experiment 2, is due to phonological recoding of the stimuli.

Experiment 3

Methods

Participants

Fifteen monolingual speakers of Japanese participated in the experiment.

Materials, procedure, and design

Materials, procedure, and design were the same as in the Experiment 2.

Results and discussion

The lower panel of Fig. 3 shows the mean d'-values and their confidence intervals for the Japanese participants, while Table 5 presents a summary of the ANOVA. The confidence intervals show that there is a significant context effect for most pairs, in the same direction as for the Dutch listeners, demonstrating that Japanese as well as Dutch listeners have more difficulties spotting variation in the realization of word-final /t/ after /s/ than after /n/.

The performance of the Japanese listeners indicates how the "untrained" auditory system treats consonant clusters and their reductions. Apparently, the properties of the auditory system allow compensation for /t/-reduction to occur more easily after /s/ than after /n/, because the allophonic variation of word-final /t/ is less salient after /s/ than after /n/. As we explained in the Introduc-

---

[1] If Japanese listeners would nevertheless try to phonologically recode the stimuli using the syllable structure prescribed by their native phonology, the standard [blɑstmuj] would be perceived as similar to [burɑsutomui], due to the insertion of epenthetic vowels (Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999). The silence in the reduced form [blɑs_muj] is most similar to a devoiced vowel, as they frequently occur in Japanese after /s/, leading to a percept like [burasumui]. These percepts should be perfectly discriminable, as they differ by one syllable. Yet the stimulus pair [blɑstmuj]-[blɑs_muj] leads to more errors than the pair [blɑntmuj]-[blɑn_muj]. The difference between the phonological recoding for the stimuli with /n/ as a preceding context ([burantomui]-[buranmuj]) would also consist of one syllable. Hence, phonological recoding should give rise to context-independent good discrimination performance.

Table 5
ANOVA summary table for the data obtained in Experiment 4 ($^{*}p < .05$)

Source	*df*	SS	MS	*F*
Subjects (S)	14	237.2	16.9	
Context	1	174.2	174.2	137.1*
S × Context	14	17.8	1.3	
Pair	8	549.7	68.7	31.8*
S × Pair	112	241.8	2.2	
Context × Pair	8	50.4	6.3	2.85*
S × Context × Pair	112	222.2	2.0	

tion, the spectral similarity of /t/ and /s/ might be the cause of this.

To allow for a better comparison of the Dutch and Japanese data, we performed a combined analysis of both data sets (see Table 6). In the Dutch data (Experiment 2), the crucial evidence was that the effect of context was especially large for some pairs involving reduced /t/s. This new analysis shows that this interaction of Pair by Context is independent of Native Language. This shows that both listener groups found discrimination of certain reduced forms especially difficult after /s/.

Nevertheless, Native Language did influence performance, and Fig. 4 displays contrasts and their confidence intervals for the significant interactions of Native Language with Context and with Pair. The interaction of Context by Native Language is due to the fact that, overall, Japanese listeners are more strongly influenced by context. The Interaction of Pair by Native Language—independent of context—is due to the fact that the Dutch listeners outperformed the Japanese listeners on pair [1–2]. (Though the confidence interval for Pair [4–5] does not include 0, a *t*-test comparing the *d*′s of Dutch and Japanese listeners for this is not significant.) These effects of Native Language show that there is a role for language learning in the perception of reduced /t/.

Nevertheless, one detail of the current results—the better performance of Dutch participant on the pair [1–2]—is especially informative with regard to the *acquired-similarity hypothesis*, which stated that the context effects observed in Experiment 2 were due learning a similarity relation between full and reduced /t/. According to this hypothesis, the Dutch listeners with experience with “allophonic” variants of /t/ should have more problems to distinguish such variants than Japanese listeners, who are not familiar with /t/ reduction. The opposite result was obtained, however, as Dutch listeners were better than Japanese listeners in discriminating the frication variant of the /t/ from the canonical /t/.

We also argued that the present results would be informative with regard to the role of phonological-recoding in Experiment 2. Since Japanese show similar behavior as the Dutch participants,[2] even though they cannot phonologically recode the current stimuli, it is likely that phonological recoding plays only a minor role in this 4I-oddity task.

It may, however, still be argued that other strategies than phonological recoding may play a role in this task. With our relatively long stimuli, leading to a long trial time, reactions occur more than a second after the perception of the reduced /t/. These long reaction times may promote the use of all kinds of strategies to reduce memory load. Therefore, the final experiment makes use of a dependent variable that occurs much earlier after stimulus onset than the behavioral responses in the first three experiments.

Experiment 4

This fourth and final experiment has a similar rationale as the first three experiments: investigating the possibility that the context effect in compensation for /t/-reduction is *not* robust over experimental manipulation and tasks, as predicted by a phonological-inference account. The previous experiments have shown that the context effect is present in both identification and discrimination tasks. With mean response times of 980 ms in Experiment 1 and of about 4 s—measured from the onset of the second stimulus—in the 4I-oddity task, it is difficult to ascertain that the observed context effect is at least partly caused by early perceptual processes, as predicted by the auditory account. To have a dependent variable temporally closer to the early perceptual processes, we used early auditory evoked potentials in Experiment 4. This experiment also allows us to test whether task affordances contribute to the context effect, as the participant do not perform any task in this experiment.

The auditory evoked potential used here is the MisMatch Negativity (MMN). The MMN arises in so-called oddball series, in which an often presented stimulus—the standard—contrasts with a seldom presented stimulus—the deviant (Näätänen, 1992, 1995; Schröger, 1998). Participants listen passively to an oddball series

---

[2] A comparison of identification and discrimination (as between Experiment 1 and Experiment 2) is not possible for monolingual Japanese listeners. The constraints of Japanese phonology prohibit us to perform an identification task with Japanese listeners. In Experiment 1, Dutch listeners decided which of two orthographically presented nonwords they heard (e.g., “dris” or “drist”). It would be difficult to indicate to monolingual Japanese participants which are the two alternatives, since both Japanese writing systems (a logographic and a syllabic one) do not allow complex codas. The fact that listeners need to be monolingual also excludes the use of an alphabetic script.

Table 6
ANOVA summary table for the comparison of Dutch and Japanese listeners' performance on the discrimination task ($^{*}p < .05$)

Source	*df*	SS	MS	*F*
Native Language	1	0.352	0.352	0.89
Native Language × Subject (S)	28	468.23	17.08	
Context × Native Language	1	48.62	48.62	23.37*
Context × Native Language × S	28	58.26	2.08	
Pair × Native Language	8	61.13	7.77	3.35*
Pair × Native Language × S	224	518.75	2.32	
Context × Pair	8	68.78	8.60	4.39*
Context × Pair × S	224	439.29	1.96	
Context × Pair × Native Language	8	25.86	3.23	1.65
Context × Pair × Native Language × S	224	439.29	1.96	

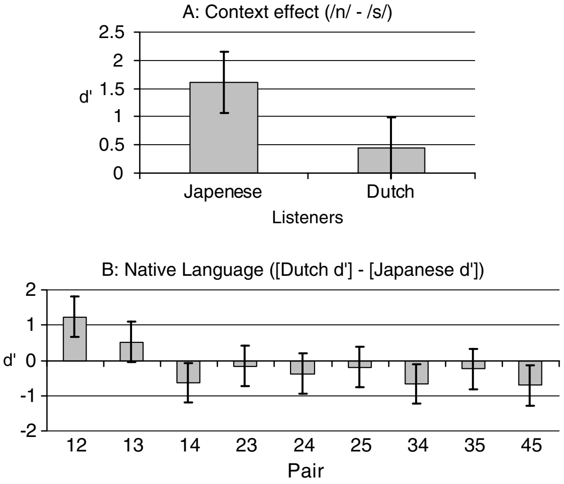


Fig. 4. Effect sizes and confidence intervals for the interactions of Native Language with Context (A) and with Pair (B) in Experiment 4.

and the electrophysiological activity evoked by the two types of stimuli is measured. Typically, the deviant yields a more negative ERP at fronto-central electrodes, which is often accompanied by a polarity inversion at the mastoid electrodes. This MisMatch Negativity arises 100–200 ms after the onset of the acoustic mismatch between standard and deviant.

Mitterer and Blomert (2003) used the MisMatch Negativity (MMN) to investigate the context effect in compensation for assimilation. They showed that the MMN to the Dutch standard-deviant pair [tœynbɑk] 'garden bench' versus [tœymbɑk], in which the deviant is a possible alternative pronunciation of the standard, was smaller than the MMN to the pair [tœynstul] 'garden chair' versus [tœymstul], in which the deviant cannot arise from the standard by assimilation. This shows that the context effect in compensation for assimilation arises in early perceptual processes. If this is also true for the context effect in compensation for /t/-reduction, a deviant that lacks a /t/ that is present in the standard stimulus should elicit a smaller MMN if the /t/ is preceded by /s/ than if preceded by /n/.

In Experiment 4, we focused on the effect of preceding segmental context on the perception of an inserted silence, the fourth level of our [t]-∅ series in the previous experiments. In the results of Experiment 1, the context effect is maximal for the stimuli with the following context /m/. Therefore, we used these stimuli here. In this set, the silence triggered the perception of an underlying /t/ after /s/ as often as the canonical /t/ (see Fig. 1B). If this "illusory" /t/ arises in early perceptual processes, the perceptual distance between the stimulus [blɑs_muj]—the underscore indicates the presence of a silence in the stimulus—and the corresponding stimulus with a canonical /t/ [blɑstmuj] should be smaller than the perceptual distance between [blɑn_muj], in which no "illusory" /t/ is perceived, and the corresponding [blɑntmuj].

Note that it has been shown previously that a silence in the speech stream may be sufficient to trigger the perception of a /t/ (Repp, Liberman, Eccardt, & Pesetsky, 1978). Importantly, the current experiment is targeted at the context effect that arises in the perception of such silences. Why are silences more likely to give rise to a /t/-percept if the preceding context is /s/ than if it is /n/?

In this experiment, the stimuli with a full [t] were always used as standards. Besides the deviants with a short silence, we used a second kind of deviant to test an additional prediction (for the viability of multiple deviants, see Näätänen, Pakarinena, Rinnea, & Takegata, 2004). In Experiment 1, the stimuli with a long consonant, [blɑsmuj] and [blɑnmuj], were both predominantly perceived as containing no /t/. Since consonants tend to be long in simple codas (i.e.,/…VC#/), this result was expected. If the evidence for the absence of /t/ is indeed equally strong for the two stimuli and not affected by the type of long consonant (/n/ versus /s/), these stimuli should lead to equally strong Mismatch Negativities when compared to the standards [blɑstmuj] and [blɑntmuj].

Methods

Participants

Thirteen (of which 10 female) members of the Max-Planck institute's subject pool participated in the experiment. All participants spoke Dutch as their native language and lived in the Netherlands; one was left-handed, the others were right-handed. The participants were 20–26 years of age (median: 23). None of them had participated in the previous experiments.

Materials

The materials used in this experiment were taken from the synthesized materials used in the previous experiments, more specifically, the stimuli with the following context /m/. The stimuli with a canonical /t/—[blɑntmuj] and [blɑstmuj]—served as standards. Two types of deviants were used, one type contained a short [n] or [s] followed by a silence ([blɑn_muj] and [blɑs_muj]), the other type contained a long [n] or [s] ([blɑnmuj] and [blɑsmuj]). This gave rise to three stimuli with [n] preceding a possible /t/ ([blɑntmuj], [blɑn_muj], and [blɑnmuj]) and three stimuli with [s] preceding a possible /t/ ([blɑstmuj], [blɑs_muj], and [blɑsmuj]). From both triples, we generated sound trains of 805 stimuli (SOA: 1 s), with the full /t/ stimulus ([blɑntmuj] or [blɑstmuj]) accounting for 76% of the stimuli and the two deviants accounting each for 12% of the stimuli (96 stimuli per sound train), randomized individually for each participant. Each sound train started with five standards that were not used for the ERP averaging. A deviant was followed by at least one standard.

Procedure

After the electrodes had been mounted and impedances had been checked, participants were seated in a comfortable chair. While watching a silent video, they then heard four 15-min long sound trains of 805 stimuli over speakers. Every sound train, with a given random order of stimuli, was used only once in the experiment. Every participant heard two sound trains in which the standard and deviants contained /n/ before the target signal, and two sound trains in which the preceding segment was /s/. In half of the recording sessions, the sound trains with the /n/-context constituted the first and third block of the experiment and the sound trains with the /s/-context the second and fourth block. In the other half of the recording sessions, this order was reversed.

Electrophysiological recording and data reduction

The nose-referenced electro-encephalogram (EEG) (0.1–100 Hz, sampling rate 250 Hz) was recorded at eleven electrodes: at both mastoids and a 3 × 3 square of electrode locations using three frontal (F3, Fz, F4), three central (C3, Cz, C4) and three parietal sites (P3, Pz, P4). Blinks and vertical eye movements were monitored with electrodes placed at the sub- and supra-orbital ridge of the right eye. Lateral eye movements were monitored by a bipolar montage using two electrodes placed on the right and left external canthus. All electrode impedances (EEG and EOG) were kept below 5 kΩ. Using a 32 channel SynAmp amplifier and the SCAN program of the Neuroscan software package (Neurosoft Inc.), the brain's electric activity was recorded in continuous mode.

In order to generate ERPs, the acquired EEG was sliced into epochs from 52 ms before stimulus onset to 948 ms after stimulus onset, band-pass filtered from 1 to 30 Hz, and baseline-corrected from −50 ms to the point at which standard and deviant started to differ (205 ms). Artefacts from vertical eye-movements were reduced using linear regression. After de-correlation, samples were rejected if the voltage on any channel excluding the vertical eye channel exceeded a value of |75| μV. Epochs were then averaged for each stimulus type and participant. These individual ERP averages were obtained from maximally 192 epochs—the number of tokens of each deviant presented to a given participant over the complete experiment—for each stimulus, by using all deviant stimuli and the same number of standard stimuli randomly drawn from all standard stimuli, excluding standards directly following deviants.

For data analysis, mean amplitudes at Fz were determined from a 50 ms window with the peak of the grand-average in the difference waves between standard and deviant as anchor. This grand-average peak was determined by visual inspection of Fz and mastoid electrodes.

Design

There were two independent variables. The first one was the nature of the stimulus functioning as the Deviant (weak Deviant: silence; strong Deviant: elongated consonant). The second independent variable was the Preceding Context (/n/ versus /s/) of the full or reduced /t/. The dependent variable was the mean amplitude of the difference waves (deviant–standard) at Fz, re-referenced against the linked mastoids, in a 50 ms window around the peak of the Mismatch Negativity (cf. Schröger, 1998).

Results and discussion

The data of one subject were discarded because less than 50% of the trials were within the limits of artefact rejection. Fig. 5 shows the grand averages over all eleven electrodes for the remaining 12 participants with a high-frequency cut-off of 15 Hz for display purposes. The data show a typical pattern for auditory evoked potentials with larger amplitudes at the frontal and central electrodes than at the parietal electrodes and an inversion of the pattern at the mastoid electrodes (i.e., at Fz the first peak is negative, at the mastoids the first peak

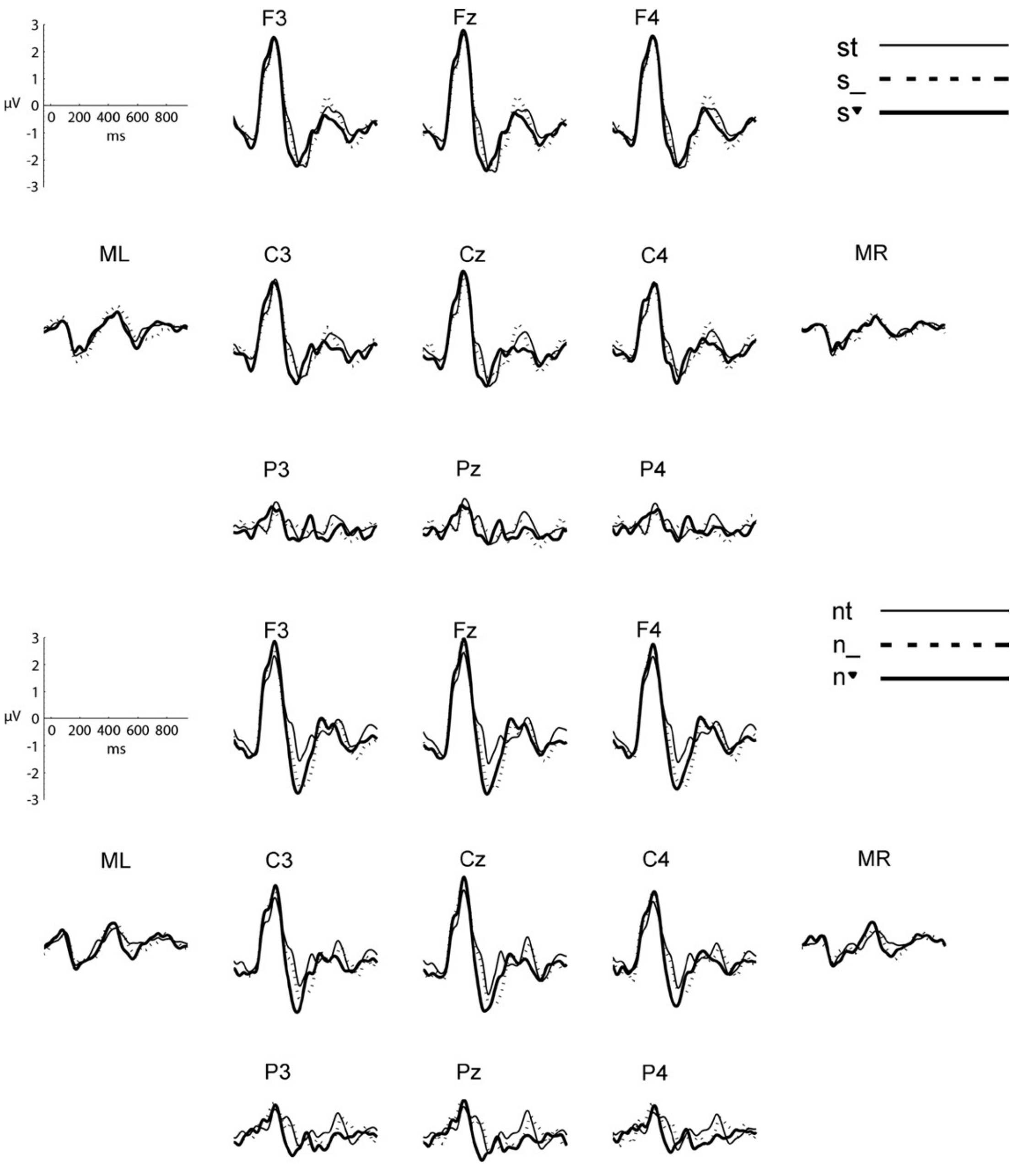


Fig. 5. Grand-average ERPs for the six stimuli of Experiment 3 for all electrodes.

is positive). In the /s/-context, there are no immediately visible differences between the ERP of the standard (thin solid line) and the deviant with a closure (dotted line) and the deviant with the long consonant (thick solid line). But in the /n/-context, the deviants have a more negative ERP than the standard in the window 300–500 ms. To better visualize the differences between standards and deviants, Fig. 6 shows the difference waves at Fz. All four deviants—the two deviants in the two different contexts—show a Mismatch Negativity around 310 ms after stimulus onset. Fig. 6 also reveals a second peak in the difference waves for the /n/-context around 420 ms. We will call these two Mismatch Negativity windows early and late Mismatch Negativity.

These latencies may seem rather late, given that the usual latency of the Mismatch Negativity is between 100 and 300 ms (Schröger, 1998). Such latencies, however, are attested for sounds which differ right from

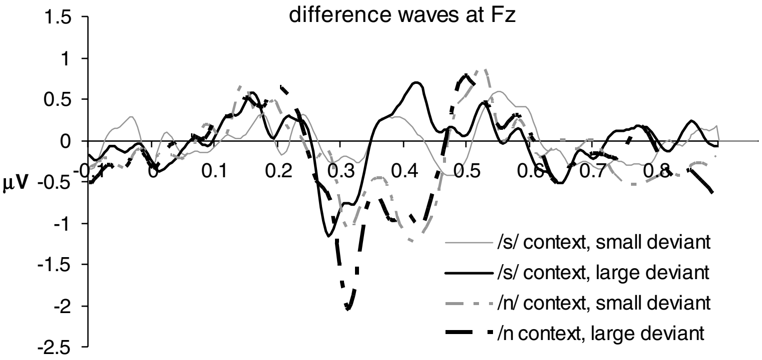


Fig. 6. Difference waves (Deviant–Standard) in Experiment 3 at Fz.

Table 7
ANOVA summary table for the data obtained in Experiment 4 ($^{*}p < .05$)

Source	*df*	Early MMN (285–335 ms)			Late MMN (375–425 ms)		
		SS	MS	*F*	SS	MS	*F*
Subjects (S)	11	30.35	2.75		13.2	1.2	
Context	1	11.85	11.85	23.38*	20.05	20.05	37.6*
Context × S	11	5.57	0.51		5.87	0.54	
Deviant	1	4.37	4.37	15.90*	1.96	1.96	2.01
S × Deviant	11	3.21	0.30		10.68	0.97	
Context × Deviant	1	0.02	0.02	0.02	0.04	0.04	0.09
S × Context × Deviant	11	8.67	0.79		4.62	0.42	

the onset. In this study, standard and deviant were identical for the first 200 ms after onset. If this overlap is subtracted, both peaks in the difference wave (see Fig. 6) are in the usual latency range for a Mismatch Negativity.

In order to statistically compare the Mismatch Negativities, we calculated the mean amplitude for 50 ms windows around the grand-average peak for the electrode Fz (cf. Schröger, 1998). The results of these analyses are displayed in Table 7. Fig. 7 shows the contrasts and confidence intervals for the two independent variables (weak versus strong deviant, context /n/ versus context /s/) and their interaction.

These results show a context effect: the Mismatch Negativities are smaller after /s/ than after /n/. This is in line with the predictions of an auditory account. A second prediction—orthogonal to the question of the underlying mechanisms—was that the Mismatch Negativity to the strong deviants, with the long simple consonants, would not be context sensitive. However, there was no interaction of context and deviancy. Apparently, segmental context not only influences the interpretation of reduced cues for an underlying /t/ but also the complete absence of cues as in the long consonant condition. This finding is in line with a recent result obtained by Janse, Nooteboom, and Quéne (2007), who found in a cross-modal priming experiment that the complete absence of word-final /t/ after [s] did not impede the recognition of words underlyingly ending in /st/.

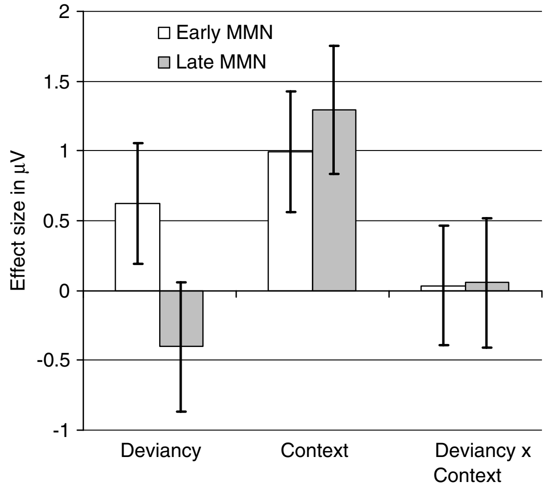


Fig. 7. Contrasts for the early and late MMN obtained in Experiment 3.

Importantly, the current results show that the context effect in compensation for /t/-reduction occurs even if participants listen only passively and do not perform a categorization or discrimination task. In this respect, the electrophysiological results converge with the results of the first three behavioral experiments in showing that the context effect forms an inherent part of the perceptual process. Moreover, Experiment 4 showed that the context effect arises—in terms of timing—rather early, already 130 ms after the onset of the auditory deviance, in the early Mismatch Negativity window.

Nevertheless, the Mismatch Negativity results alone do not allows us to argue strongly that the context effect arises partly at an auditory level, because previous studies have shown that both auditory and phonological deviance contribute to the Mismatch Negativity (Näätänen, 2001; Näätänen et al., 1997; Phillips et al., 2000; Winkler et al., 1999). Therefore, we cannot exclude the possibility that the Mismatch Negativity is already influenced by phonological inference.

The presence of two Mismatch Negativity windows is, however, suggestive. Winkler and colleagues (1999) observed a similar biphasic pattern and interpreted the early Mismatch Negativity as an auditory and the later as a phonological Mismatch Negativity. Such an interpretation is also plausible for the present results, as there was a main effect of deviancy in the earlier, but not the later time window. That is, the stronger deviant with a long consonant elicited a larger Mismatch Negativity than the deviant with the short silence in the early, but not in the late time window. The early Mismatch Negativity reflects the larger auditory difference between the standard and the strong deviant on the one hand and the standard and the weak deviant on the other. The later Mismatch Negativity abstracts over the acoustic details of the stimuli pairs. Given this interpretation of the two Mismatch Negativity windows, the phonological-inference account would have been strongly supported if there had been an effect of context only in the later but not in the early Mismatch Negativity window. However, this was clearly not the case, which suggests a role of auditory processing in compensation for reduced /t/.

General discussion

Mitterer and Ernestus (2006) reported that Dutch listeners restore severely reduced /t/ using segmental context. These listeners infer /t/ more readily after [s]—a context that facilitates /t/-reduction—than after [n]. The purpose of this study was to determine the possible loci of this context effect. Two possible accounts were tested: a phonological-inference account and an auditory account. As a mean to distinguish the accounts, we tested the robustness of the context effect. According to a phonological-inference account, it should be possible to find evidence for context-insensitive processing of reduced /t/. An auditory account, in contrast, predicts that the context effect is an inherent part of the perceptual process and as such robust against variations in task and dependent variable.

The first three experiments employed behavioral paradigms. In Experiment 1, we used the same 2AFC-task as Mitterer and Ernestus (2006) but reduced the amount of variation in the stimulus material. This should allow listeners to focus their attention on the acoustic–phonetic details of the target sounds. Nevertheless, they still recovered /t/ more often after [s] than after [n]. The opposite result would have ruled out an auditory account.

Experiment 2 and Experiment 3 used a 4I-oddity discrimination task. In Experiment 2, we tested Dutch participants. If their performance would be context-independent in this discrimination task, while being context-dependent in identification tasks (see Experiment 1 and the experiments in Mitterer & Ernestus, 2006), the auditory account would be difficult to maintain. However, also the 4I-oddity discrimination task showed a clear context effect. Reduction of /t/ was more difficult to spot after [s] than after [n]. Interestingly, however, this was not the case for all levels of reduction. This means that an auditory account is not sufficient to explains all aspects of compensation for /t/-reduction.

In Experiment 3, we investigated whether this also holds for listeners who do not have any experience with reduction of /t/ in coda clusters. This allows us to evaluate the role of language experience, which according to the phonological-inference account should be an important factor. This experiment was a replication of Experiment 2, but now with Japanese monolingual listeners. Despite the lack of experience with context-dependent /t/-reduction in Dutch or any other language, the Japanese listeners also had more trouble in discriminating different forms of /t/ after [s] than after [n]. The data pattern of Dutch and Japanese participants were, however, not identical, which also shows that an auditory account is not sufficient to explains all aspects of compensation for /t/-reduction.

In Experiment 4, we investigated how early, at least in terms of timing, the context effect arises by using an electrophysiological measure. Moreover, this experiment allowed us to investigate whether the context effect is independent of any strategies that might operate in an off-line task with relative long response latencies and whether it also occurs if participants listen only passively to the stimuli. We found that the reduction of /t/ led to a smaller MisMatch Negativity—an electrophysiological measure of change detection—after [s] than after [n]. This is especially noteworthy as the observed MisMatch Negativity was biphasic, and the context effect was present in the early, possibly auditory, phase (around 130 ms after the acoustic change) as well as in the late phase (100 ms later), a phase which has been claimed to reflect phonological processes (cf. Winkler et al., 1999).

In contrast to our behavioral results, the Mismatch Negativities revealed that the effect of context also arises when *all* cues for the presence of an underlying /t/ are absent. This suggests that the perceptual system is more tolerant towards the absence of /t/ after [s] than our earlier results indicated. A similar conclusion was reached by Janse et al. (2007), who found that, in Dutch, auditory primes with and without final /t/ (e.g., *kas* 'green house' and *kast* 'cupboard'), prime lexical decisions to visually presented /st/-final words (*kast*) equally well. Apparently, the categorical-perception task used by Mitterer and Ernestus (2006) and Experiment 1 partially underestimated the context effect.

In summary, the combined results of all four experiments indicate a contribution of auditory processing to the context effect in compensation for /t/-reduction. In addition, Experiments 2 and 3 also produced evidence for a role of language experience in the perception of reduced /t/ in two respects. First, experience with reduced /t/ appeared to decrease the overall context effect. Although this shows a role for language learning in the perception of reduced /t/, it does not support a phonological-learning account, because the effect was opposite to what such an account would predict. Note that this result actually strengthens the claim that in "untrained" auditory processing, variation of /t/ is less salient after /s/ than after /n/.

Secondly, Dutch listeners were more efficient than Japanese listeners in distinguishing a full /t/ from a /t/ realized as a fricative. One speculative explanation for this latter result is that Dutch listeners may need to distinguish such forms, because the amount of phonetic reduction may be important in discourse (see, e.g., Plug, 2005). Since consonant lenition is grammaticalized in some languages (Shockey, 2003), it seems not far fetched to assume that reduction may have pragmatic purposes in other languages. This opens a new field of investigation in which phoneticians and experts on dialogue may fruitfully collaborate, in order to reveal the pragmatic functions of phonetic reduction.

Additionally, as mentioned above, a comparison of Experiments 1 and 2 also shows that auditory processing does not sufficiently account for all data. The context effect arose only for a subset of reductions in the discrimination task of Experiment 2. Moreover, the compensation process for /t/-reduction also shows a lexical effect (Mitterer & Ernestus, 2006). The combined results therefore show that compensation for /t/-reduction is based on a multitude of perceptual and cognitive processes.

In this paper, we based ourselves on previous work on compensation for assimilation in trying to account for the context effect in compensation for /t/-reduction. This meant that we had a bias towards "processing" accounts of pronunciation variation, which assume that the input is processed in order to be aligned with a more or less abstract, canonical, lexical representation. However, "storage" accounts offer an alternative by assuming that the different pronunciation variants are stored explicitly, removing the need for a transformation of the input. The pronunciation variation may be stored on a pre-lexical level in the form of allophones (see, e.g., Sumner & Samuel, 2005, for the perception of word-final /t/ variants after vowels in English) or on a lexical level, leading to an episodic model of the lexicon (see, e.g., Bybee, 2001). A simple allophonic model can obviously not account for the context effect in the current data, because the activation of allophones of word-final /t/ would, without additional mechanisms, be independent of the segmental context. However, if the allophonic units are larger than segments and consist of syllable parts, such as the coda, the context effect can arise due to different stored exemplars of "allo-codas" of /st/ and /nt/. Similarly, episodic models of the lexicon with word-sized units can easily account for the context effect in existing words, because pronunciations with reduced /t/ are more frequent for /st/-final than /nt/-final words. The context effect for nonwords may result from lexical analogy (Ernestus & Baayen, 2006).

However, the current data underscore that both types of models are incomplete, because they rely completely on learning. To store pronunciation variation, either at a pre-lexical or lexical level, one has to be exposed to it first. This was not the case for the Japanese participants in Experiment 3, yet their discrimination performance revealed a context effect as well. Hence, storage accounts need to take into account additional processes. This paper shows that at least some of these processes are auditory in nature and that they play an important role in the perception of reduced forms.

Consider, for instance, how such models would account for the perception of reduced /st/ codas if not the /t/, but the /s/ was reduced. Both types of models would cope with such reduced forms just as easily as they cope with actually observed reduced forms. In episodic models of the lexicon with word-sized units, the reduced form [pɔt] would be stored and linked to the full form [pɔst], while, in "allo-coda" models, [t] would be stored as an alternative form of the /st/-coda. These models thus allow phonetic alternations to be arbitrary. This is also the case for the phonological-inference account, which learns arbitrary patterns of covariance. However, it appears that, across languages, pronunciation variation is not arbitrary and a full form such as /pɔst/ is more likely to undergo /t/-reduction than /s/-reduction. This pattern could result from articulatory constraints, but our results indicate that there is also a perceptual basis to this pattern. Any theory accounting for the perception and production of reduced forms in connected speech should take functional constraints imposed by perception into account.

In conclusion, this article has shown that compensation for reduced /t/ involves a complex process resulting from perceptual and lexical constraints as well as phonological learning. Accounting for the comprehension of reduced forms thereby forms a challenge for theories of speech comprehension, as it involves several levels of processing working in concert.

Acknowledgments

We thank Arthur Samuel and an anonymous reviewer for comments on an earlier version of this manuscript. Natasha L. Warner from the Department of Linguistics of the University of Arizona helped us with the likely adaptation of our stimuli to Japanese phonology. We thank Marloes van der Goot, Nina Davids, and Jet Sueters for their help in running the experiments.

References

Beddor, P. S., Harnsberger, J. D., & Lindemann, S. (2002). Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. *Journal of Phonetics, 30*, 591–627.

Beddor, P. S., & Krakow, R. A. (1999). Perception of coarticulatory nasalization by speakers of English and Thai: Evidence for partial compensation. *Journal of the Acoustical Society of America, 106*, 2868–2887.

Boersma, P. (1998). *Functional phonology formalizing the interactions between articulatory and perceptual drives*. The Hague: Holland Academic Graphics.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.

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

Bybee, J. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.

Cutler, A., & Broersma, M. (2005). Phonetic precision in listening. In W. J. Hardcastle & J. Mackenzie Beck (Eds.), *A figure of speech: A Festschrift for John Laver* (pp. 63–91). London: Lawrence Earlbaum.

Darcy, I., Peperkamp, S., & Dupoux, E. (2007). Bilinguals play by the rules: Perceptual compensation for assimilation in late L2-learners. In J. Cole & J. Hualde (Eds.). *Laboratory Phonology* (Vol. 9, pp. 411–442). Berlin: Mouton de Gruyter.

de Boer, B. (2000). Self-organisation in vowel systems. *Journal of Phonetics, 28*, 441–465.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion?. *Journal of Experimental Psychology: Human Perception and Performance 25*, 1568–1578.

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

Ernestus, M., & Baayen, H. R. (2006). The functionality of incomplete neutralization in Dutch: The case of past-tense formation. In L. M. Goldstein, D. H. Whalen, & C. T. Best (Eds.). *Laboratory phonology* (Vol. 8, pp. 27–49). Berlin, Germany: Mouton de Gruyter.

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

Ernestus, M., Lahey, M., Verhees, F., & Baayen, H. R. (2006). Lexical frequency and voice assimilation. *Journal of the Acoustical Society of America, 120*, 1040–1051.

Fitch, H. L., Halwes, T. E., Erickson, D. M., & Liberman, A. M. (1980). Perceptual equivalent of two acoustic cues for stop-consonant manner. *Perception & Psychophysics, 27*, 343–350.

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

Fowler, C. A., & Smith, M. (1986). Speech perception as "vector analysis: An approach to the problems of segmentation and invariance. In J. Perkell & D. Klatt (Eds.), *Invariance and variability of speech processes* (pp. 123–136). Hillsdale, NJ: Lawrence Earlbaum Associates.

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

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

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

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

Gerrits, E., & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics, 66*, 363–376.

Gow, D. W. (2001). Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language, 45*, 133–159.

Gow, D. W. (2002). Does English coronal place assimilation create lexical ambiguity. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 163–179.

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

Gow, D. W., & Im, A. M. (2004). A cross-linguistic examination of assimilation context effects. *Journal of Memory and Language, 51*, 279–296.

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

Guenther, F. H., Husain, F. T., Cohen, M. A., & Shinn-Cunningham, B. G. (1999). Effects of categorization and discrimination training on auditory perceptual space. *Journal of the Acoustical Society of America, 106*, 2900–2912.

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

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

Harnad, S. (1987). *Categorical perception: The groundwork of cognition*. Cambridge: Cambridge University Press.

Hume, E., & Johnson, K. (2001). *The role of speech perception in phonology*. New York, NJ: Academic Press.

Hura, S. L., Lindblom, B., & Diehl, R. (1992). On the role of perception in shaping phonological assimilation rules. *Language and Speech, 35*, 59–72.

Janse, E., Nooteboom, S., & Quéne, H. (2007). Coping with gradient forms of /t/-deletion and lexical ambiguity in spoken word recognition. *Language and Cognitive Processes, 22*, 161–200.

Kemps, R., Ernestus, M., Schreuder, R., & Baayen, H. R. (2004). Processing reduced word forms: The suffix restoration effect. *Brain and Language, 90*, 17–127.

Kingston, J., & Macmillan, N. A. (1995). Integrality of nasalization and f1 in vowels in isolation and before oral and nasal consonants—a detection-theoretic application of the Garner paradigm. *Journal of the Acoustical Society of America, 97*, 1261–1285.

Kohler, K. J. (1990). Segmental reduction in connected speech in German: Phonological facts and phonetic explanations. In W. J. H. A. Marchal (Ed.), *Speech production and speech modelling* (pp. 69–92). Dordrecht: Kluwer.

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

Lahiri, A., & Reetz, H. (2002). Underspecified recognition. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology: Vol. 7* (pp. 637–676). Berlin: Mouton de Gruyter.

Liberman, A. M. (1957). Some results of research on speech perception. *Journal of the Acoustical Society of America, 29*, 117–123.

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, The Netherlands: Kluwer.

Lotto, A. J., Kluender, K. R., & Holt, L. L. (1997). Perceptual compensation for coarticulation by Japanese quail (Coturnix coturnix japonica). *Journal of the Acoustical Society of America, 102*, 1134–1140.

Macmillan, N. A., & Creelman, D. (1991). *Detection theory: A user's guide*. Oxford: Blackwell.

Macmillan, N. A., Kingston, J., Thorburn, R., Dickey, L. W., & Bartels, C. (1999). Basic sensitivity and phonetic labeling measure distinct sensory and decision-rule interactions. *Journal of the Acoustical Society of America, 106*, 2913–2932.

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

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

Massaro, D. W. (1987). Categorical partition: A fuzzy-logic model of categorization behavior. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 254–283). Cambridge, UK: Cambridge University Press.

Mielke, J. (2003). The interplay of speech perception and phonology experimental evidence from Turkish. *Phonetica, 60*, 208–223.

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. (2006). The role of perceptual integration in the perception of assimilated word forms. *Quarterly Journal of Experimental Psychology, 59*, 1305–1334.

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

Mitterer, H., & Ernestus, M. (2006). Listeners recover /t/s that speakers lenite: Evidence from /t/-lenition in Dutch. *Journal of Phonetics, 34*, 73–103.

Näätänen, R. (1992). *Attention and brain function*. Hillsdale, NJ: Erlbaum.

Näätänen, R. (1995). The mismatch negativity: A powerful tool for cognitive neuroscience. *Ear and Hearing, 16*, 6–18.

Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology, 38*, 1–21.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., et al. (1997). Language specific phoneme representation revealed by electric and magnetic brain responses. *Nature, 385*, 432–434.

Näätänen, R., Pakarinena, S., Rinnea, T., & Takegata, R. (2004). The mismatch negativity (MMN): Towards the optimal paradigm. *Clinical Neurophysiology, 115*.

Ohala, J. J. (1990). The phonetics and phonology of aspects of assimilation. In *Papers in laboratory phonology I: Between the grammar and the physics of speech* (pp. 258–275). Cambridge, UK: Cambridge University Press.

Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., et al. (2000). Auditory cortex accesses phonological categories: An MEG mismatch study. *Journal of Cognitive Neuroscience, 12*, 1038–1055.

Plug, L. (2005). From words to actions: The phonetics of eigenlijk in two communicative contexts. *Phonetica, 62*, 131–145.

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 Psychology: Human Perception and Performance, 4*, 621–637.

Schouten, M. E. H., Gerrits, E., & van Hessen, A. (2003). The end of categorical perception as we know it. *Speech Communication, 41*, 71–80.

Schröger, E. (1998). Measurement and interpretation of the mismatch negativity. *Behavior Research Methods, Instruments & Computers, 30*, 131–145.

Schwartz, J.-L., Boë, L. J., Valleé, N., & Abry, C. (1997). Major trends in vowel system inventories. *Journal of Phonetics, 25*, 233–253.

Shockey, L. (2003). *Sound patterns of spoken English*. Cambridge, MA: Blackwell.

Steriade, D. (2001). Directional asymmetries in place assimilation: A perceptual account. In E. Hume & K. Johnson

(Eds.), *The role of speech perception in phonology* (pp. 219–250). New York, NJ: Academic Press.

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

Winkler, I., Lehtokoski, A., Alku, P., Vaiiio, M., Czigler, I., Csépe, V., et al. (1999). Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Cognitive Brain Research, 7*, 357–369.

Wittenburg, P., Nagengast, J., & Baumann, H. (1998). NESU—The Nijmegen experiment setup. In A. Trapp, N. Hammond, & C. Manning (Eds.), *CIP98 conference proceedings* (pp. 92–93). York, England: CTI Centre for Psychology.