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 /mæst bəstɛlt/ *mest besteld* 'fertilizer ordered') and, second, a shorter duration of the preceding segment (the /s/ tends to be shorter in /mæst bəstɛlt/ than in /mæs 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 /mæst 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 seg-