How linguistic and probabilistic properties of a word affect the realization of its final /t/: Studies at the phonemic and sub-phonemic level

Barbara Schuppler\textsuperscript{a,b,*}, Wim A. van Dommelen\textsuperscript{c}, Jacques Koreman\textsuperscript{c}, Mirjam Ernestus\textsuperscript{d,e}

\textsuperscript{a} Center for Language and Speech Technology, Radboud University Nijmegen, The Netherlands
\textsuperscript{b} Signal Processing and Speech Communication Laboratory, Graz University of Technology, Inffeldgasse 16, 8010 Graz, Austria
\textsuperscript{c} Department of Language and Communication Studies, NTNU, Trondheim, Norway
\textsuperscript{d} Center for Language Studies, Radboud University Nijmegen, The Netherlands
\textsuperscript{e} Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

\textbf{Abstract}

This paper investigates the realization of word-final /t/ in conversational standard Dutch. First, based on a large number of word tokens (6747) annotated with broad phonetic transcription by an automatic transcription tool, we show that morphological properties of the words and their position in the utterance’s syntactic structure play a role for the presence versus absence of their final /t/. We also replicate earlier findings on the role of predictability (word frequency and bigram frequency with the following word) and provide a detailed analysis of the role of segmental context. Second, we analyze the detailed acoustic properties of word-final /t/ on the basis of a smaller number of tokens (486) which were annotated manually. Our data show that word and bigram frequency as well as segmental context also predict the presence of sub-phonemic properties. The investigations presented in this paper extend research on the realization of /t/ in spontaneous speech and have potential consequences for psycholinguistic models of speech production and perception as well as for automatic speech recognition systems.

\textbf{1. Introduction}

A frequent phenomenon observed in spontaneous, conversational speech is that words are produced in a reduced way compared to their canonical pronunciations: a phrase like ‘supposed to see’ may sound approximately like [sasasi]. A study on American English shows that whole syllables may be absent in 6\% of the word tokens and that segments may be absent or substituted in every fourth word (Johnson, 2004). In Germanic languages, one phoneme that is frequently reduced is /t/ (e.g., Jurafsky, Bell, Gregory, & Raymond, 2001, for conversational American English, and Goeman, 1999, for dialectal Dutch). Nearly, all studies of reduction of /t/ have restricted themselves to studying the presence versus absence of /t/ and investigated only a small number of possible predictors. The aim of the present paper is to investigate the roles of a wide variety of variables in the reduction of /t/ in conversational standard Dutch on the basis of broad phonetic transcriptions and of annotations in terms of sub-phonemic properties.

This research is theoretically important. Most psycholinguistic models of speech perception do not take into account the pronunciation variation found in spontaneous conversations. They assume that only the canonical pronunciations of the words are stored in the lexicon and do not explicitly provide mechanisms to map reduced pronunciation variants on these canonical pronunciations. The model Shortlist (Norris, 1994), for instance, has a word error rate (WER) of 64.5\% with a lexicon of canonical pronunciations for spontaneous Dutch. If pronunciation variants are added to the lexicon in combination with estimates of their prior probability, then the WER goes down to 48.2\% (Scharenborg & Boves, 2002). Thus, information about the conditions under which segments are likely to be reduced is necessary to adapt existing psycholinguistic models so that they can deal with spontaneous speech. Also, most models of speech production do not take into account that words may be reduced (e.g., Levelt, Roelofs, & Meyer, 1999). As a consequence, these models cannot process natural conversations and are not ecologically valid. Quantitative corpus studies on reductions will show which reduced word forms these models should be able to process and under which conditions.

Quantitative studies on reduction are also necessary to improve automatic speech recognition (ASR) systems. Whereas for read
speech the accuracies obtained are typically in the range of 85–90\% words correctly identified, the recognition accuracies drop to 50–
60\% for spontaneous speech (Ali Raza, Hussain, Sarfraz, Ullah, &
Sarfraz, 2010; Greenberg, 1997; Greenberg & Chang, 2000). Schuppler,
Nock, and Khudanpur (2000) showed that the drop in performance
relates especially with greater pronunciation variability in sponta-
neous speech. They recorded and transcribed conversational
speech, which was then read aloud by the same subjects. The error
rate for the conversational data was more than 50\% higher than for
the read version. This variability can at least partly be captured by
the incorporation of several pronunciation variants for each word in
the recognition lexicon in conjunction with their prior probabilities
and with statistics about the conditions under which these are likely
to occur (e.g., Wakita, Singer, & Sagisaka, 1999; Wester, 2002).

Traditional psycholinguistic models of speech production and
comprehension as well as most ASR systems assume that speech is
represented as a sequence of phones. One restriction of this assump-
tion is that pronunciation variation can only be described in terms of
phone deletions, insertions and substitutions. The acoustic results of
overlapping, asynchronous gestures of the articulators cannot be
captured. More recent psycholinguistic models that can account for
such realizations include Articulatory Phonology (Brownman &
Goldstein, 1992) and exemplar based models (e.g., Goldinger, 1997;
Johnson, 2004). For ASR systems, models are developed based on
acoustic–phonetic features (APFs, e.g., Kirchhoff, Fink, & Sagerer,
2002; Scharenborg, Wan, & Moore, 2007), but progress is slow due to
lack of appropriately labeled material on the APF level (Schuppler,
van Doremalen, Scharenborg, Cranen, & Boves, 2009) and lack of
quantitative phonetic studies on sub-phonemic variation.

The present study provides a detailed analysis of the condi-
tions favoring the acoustic absence of Dutch word-final /t/ and its
sub-phonemic properties. We investigated word-final /t/ for several reasons. First, word-final /t/ is known to be frequently
reduced in Germanic languages. Second, word-final /t/ in Dutch
can function as a grammatical morpheme (e.g., in loopt ‘he
walks’, where it marks the second and third person singular
present tense), and a study of word-final /t/ reduction can there-
fore reveal a role of morphology in the reduction of phones. Third,
in word-final position, /t/ may be followed by different types of
syntactic boundaries and we can therefore investigate whether
their presence plays a role in phone reduction. Finally, an analysis
of /t/ is also interesting from an engineering point of view. Most
ASR systems rely on the assumption that speech is stationary
within a window of 25 ms. Since plosives in conversational
speech may be much shorter and moreover consist of at least
two different phases (constriction and burst\(^1\)), the accurate
detection of plosives requires a higher temporal resolution (e.g.,
Schuppler, van Doremalen, Scharenborg, et al., 2009). In order to
improve automatic plosive detectors, more quantitative phonetic
knowledge about their sub-phonemic properties is necessary.

The present paper consists of two studies that analyze the
realization of word-final /t/ based on a corpus of conversational
standard Dutch. Study I investigates the acoustic presence versus
absence of word-final /t/ on the basis of a large number of tokens
(6747) phonetically annotated by means of an ASR system. Its
main focus is on the roles of morphology and syntax, while this
study also replicates earlier findings on the roles of bigram and
word frequency and segmental context (Section 3.3.7). The auto-
nomatically generated transcriptions treat the signal as if it consists
of beads on a string, with each bead representing a single, clearly
realized phone (Ostendorf, 1999). As a consequence, realizations
resulting from articulatory overlap with neighboring segments
cannot be captured. In order to get a better insight into how

\(^1\) Dutch plosives are not aspirated.
has a longer duration when it is a productive morpheme, we also expect fewer reductions for tokens of word-final /t/ that function as a grammatical morpheme (e.g., in loop-t ‘[he] walk-s’ in which the /t/ indicates the second or the third person singular present tense) than for those that are part of the stems of words (e.g., in kast ‘cubby’).

Morphologically complex words are hypothesized to be more reduced if they are retrieved as wholes from the lexicon instead of being computed from their parts. In line with this hypothesis, Losiewicz (1992) showed that word-final /t/ and /d/ in English are longer in past-tense morphemes of low frequency verbs than of high frequency verbs. Hay (2003) investigated the role of the frequency of a derived form relative to the frequency of its stem. She reports that /t/ in words that are more frequent than their stems (e.g., the word swiftly because the frequency of swiftly is greater than the frequency of swift) tend to be more reduced than /t/ in words which are less frequent than the stems they contain (e.g., the word softly because the frequency of softly is lower than the frequency of soft). She suggests that the relative frequency reflects the decomposability of words: the higher the relative frequency, the more likely words are to be retrieved as whole words from the lexicon. In our study, we investigate whether the reduction in the Dutch inflectional morpheme /t/ can also be predicted by the frequency of the word relative to the frequency of its stem.

The studies presented above all suggest that higher predictability results in higher degrees of reduction. Contradictory are the results by Kuperman, Puyluym, Ernestus, and Baayen (2007). They showed that interfixes in Dutch compounds have longer durations the more probable they are given the compound and its constituents. On the basis of their results, they formulated the Paradigmatic Signal Enhancement Hypothesis, which states that the most likely alternative in a paradigm is realized with greater acoustic salience. Their explanation for this phenomenon is that speakers are more confident when selecting more probable members of the morphological paradigm than when selecting a less probable one. In Dutch, the verb stems in verb stems + /t/ combinations, which we investigate in the present study, also occur as verb forms just by themselves (e.g., loop is also the first person singular present tense). Thus, the frequency of the verb stem + /t/ combination relative to the frequency of the stem shows which of the two forms is the more frequent one in the paradigm. Therefore, the Paradigmatic Signal Enhancement Hypothesis predicts that /t/ tends to be less reduced in highly predictable word forms, which is the opposite of the prediction just formulated above given the results by Hay (2003).

1.3. Syntactic and prosodic properties

Linguistic research has shown that the underlying syntactic structure of the utterance is manifested in the phonetic detail of the words. A well-studied phenomenon is final lengthening, which marks the boundaries of linguistic units, including the boundaries of words (word-final lengthening) and of phrases (phrase-final lengthening, e.g., Beckman & Edwards, 1990; Fuchs, Krivokapic, & Jannedy, 2010). The phonological literature provides evidence that the underlying syntactic structure affects pronunciation via the prosodic structure. The type of prosodic boundary onto which a syntactic boundary maps depends on speech rate and the number of words in the syntactic constituent, among other factors (Nespor & Vogel, 2007).

Prosodic boundaries do not only condition lengthening but also define the application domains of cross-word phonological rules. For example, intervocalic /s/-assimilation in Greek applies across a syntactic boundary (between a noun phrase and a verb phrase) if the constituents on either side are short, but not if they are long (Nespor & Vogel, 2007). To our knowledge, no earlier studies have investigated the role of syntactic structure and of the length of constituents on the phonetic realization of words in large corpora of natural conversations. In our study, we investigate (1) whether tokens of /t/ that are in the middle of a syntactic constituent are more reduced than tokens of /t/ that are at the right edge of a syntactic constituent and (2) whether tokens of /t/ at the right edge of a syntactic constituent tend to be less reduced if the constituent is longer.

1.4. Segmental context

It is well known that sounds show different properties depending on the segmental context they occur in due to co-articulation. Similarly, segmental context can condition the acoustic absence of segments. The acoustic absence of a sound does not necessarily imply that the segment was not articulated. For instance, Browman and Goldstein (1990) measured the movements of the articulators with an X-ray microbeam system, which tracked the positions of lead pellets placed on the articulators. They found that word-final /t/ in word combinations like perfect memory could be absent in the acoustic signal, even though the tongue clearly moved to the alveolar ridge. This gesture was acoustically hidden behind the bilabial gesture.

The role of segmental context for the acoustic absence of /t/ in Dutch has been documented by Ernestus (2000) for casual speech and by Mitterer and Ernestus (2006) for read speech. Both studies showed that /t/ is more often acoustically absent when preceded and followed by consonants than by vowels. Moreover, /t/ is most frequently absent before the voiced bilabial plosive /b/, probably because this plosive may hide the articulatory gestures for /t/, as shown by Browman and Goldstein (1990). In the present study, we provide a quantitative analysis of the effect of segmental context on the reduction of /t/ at both the phonemic and the sub-phonemic level.

2. Corpus data

Our research is based on the 10 spontaneous Dutch dialogues that form the ERNESTUS CORPUS OF SPONTANEOUS DUTCH (ECSD; Ernestus, 2000). Each of these conversations has a duration of approximately 90 min. In total, they contain 153,200 word tokens representing 9035 word types produced in 15 h of speech. Characteristic for this corpus is the high level of spontaneity and the speakers’ homogeneity in geographical and social background. All 20 speakers are male native speakers of Dutch, all from the Western provinces of the Netherlands and all holding academic degrees. The speakers were between 21 and 55 years old. They have been classified as speakers of standard Dutch.

The following set-up was used for the recordings: Two speakers were seated at about 1.5 m from each other at a table in a sound-proof room. They were recorded with two Sennheiser MD527 supercardioid microphones onto Sony DAT. They were free to choose their topics for the first 40 min of the recordings. The second part of the recording was a role play, where they negotiated about the purchase of camping equipment. Both speakers separately received written instructions on the goals they had to reach in the role-play; they were not given any further specific instructions. The experimenter was only present during the first part, but did not take an active part in the conversations. As the speakers were friends talking about everyday issues, the atmosphere during the conversations was relaxed, resulting in a casual, chatty speech style.

The handmade verbatim orthographic transcriptions of the corpus were prepared for automatic processing as described in Schuppler, Ernestus, Scharenborg, and Boves (2011). On the basis
of these orthographic transcriptions, the corpus was enriched with part-of-speech tags (POS tags) and a syntactic annotation, both generated by means of the Alpino-parser (Bouma, van Noord, & Malouf, 2000).

3. Study I

3.1. Material

An ASR system was used to create a broad phonetic transcription for the ECD. Automatic transcriptions have the advantage that they are consistent and can be more easily obtained than manual transcriptions for large data sets. We used the so-called forced alignment for creating the broad phonetic transcriptions. Input for the forced alignment were the speech files, the orthographic transcriptions of these files, a pronunciation lexicon of the words in these transcriptions, and acoustic models for each phone that had been trained beforehand. First, the words from the orthographic transcriptions were looked up in a pronunciation lexicon containing multiple pronunciation variants per word. Then, given the acoustic signal and the acoustic phone models, the ASR system chose the pronunciation variant that matched best with the speech signal.

The ASR system we used was the Hidden Markov Model speech recognition toolkit HTK (Young et al., 2002). The pronunciation lexicon contained canonical phonemic representations and several pronunciation variants for each word type. These variants were generated by means of a set of 32 phonological, co-articulation and reduction rules applied to the canonical pronunciations of the words. These rules were formulated on the basis of observations from earlier studies on spontaneous, casual Dutch (Ernestus, 2000) and included one rule that deleted /t/ in word final position independent of any other criteria. The rules created on average 27 pronunciation variants per word type. A detailed description of the automatic transcription procedure can be found in Schuppler et al. (2011).

The acoustic models were 37 32-Gaussian tri-state monophone acoustic models that had been trained on the 396,187 word tokens in the read speech component ‘Library for the blind’ incorporated in the Spoken Dutch Corpus (Oostdijk et al., 2002). The models were trained at a frame shift of 5 ms and a window length of 25 ms (Hämäläinen, Gubian, ten Bosch, & Boves, 2009). We used acoustic models of a shorter frame shift than the default of 10 ms used in earlier studies (e.g., Adda-Decker, Boula de Mareeuw, Adda, & Lamel, 2005; Schuppler, van Dommelen, Koreman, & Ernestus, 2009; Van Bael, 2007) in order to obtain more accurate phonetic transcriptions and positions of the segment boundaries. Since we used a frame shift of 5 ms and the acoustic models minimally consist of three emitting states (no skips), annotated segments have durations that are multiples of 5 ms and a minimum duration of 15 ms. This does not mean that shorter segments cannot be annotated at all, but that their boundaries are placed in the neighboring segments.

The resulting transcriptions reached a good labeling agreement2 with manual transcriptions, as observed in an earlier study (Schuppler et al., 2011). That study is based on the same set of tokens and their manual transcriptions which is also used in Study II of this paper. We showed that the automatic transcriptions are in good agreement with both the manually made transcriptions and with the perceptual presence of /t/ in the tokens.

The tokens for the study were chosen in such a way that the word following the target token was part of the same utterance—given the punctuation of the orthographic transcription—and was neither one of the fillers eh, ah, uh nor a broken word. Furthermore, we excluded utterances that could not be assigned a syntactic annotation and/or POS tag with high certainty. We also excluded the highly frequent words dat ‘this’, het ‘it’, and niet ‘not’ because they are represented by a much higher number of tokens (2725, 2188 and 954, respectively) than the other words (average number of tokens: 22.6) and therefore show idiosyncratic behavior (Ernestus, 2000). This leaves 6747 word tokens representing 556 word types for the analysis.

For this study, we consider /t/ as present when a word token was transcribed with a /t/ in the broad phonetic transcription (i.e., has been classified as present by the ASR system). As we will see in Section 4, /t/s classified as present can vary in their detailed acoustic realizations (see also Figs. 1 and 2, which show different realizations of /t/ that were all classified as present by the ASR system).

3.2. Analysis method

Overall, we observed that 36.8% of all tokens of word-final /t/ were classified as absent, ranging from 19.7% to 53.5% for the 20 different speakers. To investigate the conditions favoring the presence versus absence of word-final /t/, we used the statistical modeling technique of mixed effects logistic regression with a binomial logit link function and contrast coding (Jaeger, 2008). All models presented in this section contain the random variables Speaker, Word, and Following Word, because they all were statistically significant predictors (for all random variables: p < 0.0001).

We first present a control model, which shows the roles of prosodic variables unrelated to syntax, and phonetic variables capturing rough differences in segmental context. To this model, we separately added the probabilistic word predictability variables, the morphological variables, the syntactic variables, and the variables that capture details of the segmental context. Furthermore, we tested the interactions between these variables. From all models, we removed predictors and interactions that were not statistically significant and subsequently we only present the significant effects.

3.3. Results and discussion

3.3.1. Control model

Variables. The independent variables of the control model were the prosodic variables Syllabic Stress, which indicates whether the word-final syllable is stressed, and Number of Syllables in the word, whose range is shown in Table 1. These measures were included because it has been shown that stressed syllables tend to be longer than unstressed syllables (e.g., Ladefoged, 1982) and that segments tend to be longer in shorter words than in longer words (e.g., Nooteboom, 1972). Further, since previous research has shown that more /t/s are acoustically absent if preceded or followed by consonants than by vowels (e.g., Ernestus, 2000; Mitterer & Ernestus, 2006), we added the independent variables Previous Segment and Following Segment with the values ‘silence’ (only for Following Segment), ‘consonant’ and ‘vowel’. The values of all these measures were determined on the basis of the canonical transcriptions of the words. Finally, several studies have shown that function words tend to be more reduced than content words (e.g., Bell et al., 2009; Johnson, 2004) and we therefore also added the independent variable Word Class, with the values ‘function word’ and ‘content word’, as indicated by the POS tags of the words. There were 1617 function words representing 15 word types and 5130 content words representing 539 word types. The control model was calculated for the complete data set (N=6747).

---

2 With good labeling agreement, we refer to an agreement at least in the rage of agreement between human transcribers for the same speech style.
Results. Table 2 shows the results for the control model (M0). The prosodic variable Number of Syllables is significant: /l/ is significantly more often acoustically present in longer words, as defined by the (canonical) Number of Syllables. This tendency is exactly opposite to the findings of earlier studies (e.g., Nooteboom, 1972; Torreira & Ernestus, 2009). One reason could be that longer words tend to be less frequent. We will come back to this possibility in the following section. Furthermore, both Previous and Following Segment are significant: /l/ is less often absent after vowels (31.9%) than after consonants (42.3%) and it is less often absent before vowels (24.9%) and silence (17.5%) than before consonants (45.1%).

3.3.2. Word and bigram frequency

Variables. We added the probabilistic variables Word Frequency and Bigram Frequency to the control model (M0), where we defined Bigram Frequency as the frequency of the word combination consisting of the target word and the following word. We extracted both frequency measures from the Spoken Dutch Corpus (Oostdijk et al., 2002), taking into account the part of speech tag of the target word and the following word, and applied a logarithmic transformation. Table 1 shows the ranges of the two variables. Since the two measures are correlated ($r = 0.47$, $p < 0.0001$), we first orthogonalized Word Frequency and Bigram Frequency by replacing Word Frequency by the residuals of a linear regression model predicting Word Frequency as a function of Bigram Frequency.

Results. Both measures showed significant effects (residuals: $\beta = -0.07$, $z = -4.35$, $p < 0.0001$ and Bigram Frequency $\beta = -0.42$, $z = -4.62$, $p < 0.0001$), but the $\beta$-value of the residuals of Word Frequency was much smaller than the $\beta$-value of Bigram Frequency. This does not necessarily mean, however, that Bigram Frequency is the more important predictor, since part of the predictive power of Word Frequency has been removed in the orthogonalization procedure. We therefore also orthogonalized Word Frequency and Bigram Frequency the other way around. For this purpose, we built a linear regression model predicting Bigram Frequency as a function of Word Frequency and added the residuals in addition to Word Frequency to the control model. In the resulting model (M1 in Table 2), the $\beta$-value of Word Frequency was still smaller than the $\beta$-value of the residuals of
Bigram Frequency. Since the Bigram and Word Frequency have different ranges (see Table 1), it is possible that the difference in their \( \beta \) values does not actually reflect a difference in effect size. Therefore, we calculated the range dependent effect size as

\[
\text{Max}_{\text{value}} \beta \text{value} - \text{Min}_{\text{value}} \beta \text{ value}.
\]

The effect sizes computed with this formula are 8.88 for Word Frequency and 9.45 for Bigram Frequency, which also indicates that Bigram Frequency had a greater effect than Word Frequency.

All these analyses allow us to conclude that it is especially Bigram Frequency, and not Word Frequency, that predicts the acoustic presence of word-final /t/. Both the effects of Bigram Frequency and Word Frequency show that word-final /t/ is more often acoustically absent in units of higher frequencies. This finding is in line with several earlier corpus studies (e.g., Bell et al., 2009; Pluymakers et al., 2005a; Torreira & Ernestus, 2009) that support the Probabilistic Reduction Hypothesis (Jurafsky et al., 2001), which states that more predictable linguistic units tend to receive shorter and weaker pronunciations.

The independent variable Number of Syllables of the control model (M0) correlated with the probabilistic measures Bigram Frequency (\( r = -0.31 \)) and Word Frequency (\( r = -0.32 \)). Therefore, we orthogonalized Number of Syllables, Word and Bigram Frequency by replacing Number of Syllables by the residuals of the linear regression model which predicts Number of Syllables as a function of Word Frequency and the residuals of Bigram Frequency. The residuals of the orthogonalization model have the same effect in M1 as Number of Syllables had in the control model (M0), namely that /t/ tends to be more often present in longer words. Future studies have to further investigate the possible sources of this unexpected effect.

### 3.3.3. Morphological properties

**Variables.** In order to investigate whether morphological properties of the words influence the acoustic absence versus presence of word-final /t/, we built models for content words only, since content words can end in the suffix +t. First, we added the variable Morphological Status to the model M1, which indicated whether the word-final /t/ forms a suffix or is part of the stem. The independent variable Word Class was excluded, since only one value was left (i.e., ‘content word’). Morphological status and Bigram Frequency were correlated, hence we orthogonalized these two variables by replacing Morphological Status by the residuals of a general linear regression model predicting Morphological Status as a function of Bigram Frequency.

**Results.** In this model, Morphological Status did not show an effect on the presence or absence of /t/ (and hence Table 2 does not show this model).

### 3.3.4. Morphological complexity (in phonemically identical word pairs)

**Data.** In a next step, we restricted our data set to phonemically identical word pairs consisting of words with an identical canonical phonemic pronunciation but differing in whether the final /t/ also represents a morpheme on its own (\( N = 366, 18 \) word types). For instance, the words vind ‘[i] find’ and vindt ‘[he] finds’ share the canonical pronunciation [\( \text{vindt} \)], but only in vindt the /t/ also carries grammatical meaning.

**Results.** The resulting model (see M2 in Table 2) is very similar to Model M1 of the complete data set. Importantly, however, the residuals of Morphological Status appeared now to be significant in the expected direction: /t/ is less likely to be absent if it also has a morphological function than if it is only part of the stem. Since the word pair vind and vindt covers nearly half of the tokens on which M2 is based, and additionally vind is four times as frequent as vindt, we excluded this word pair from the data and re-ran the model (\( N = 152, 16 \) word types). We found again an effect of the Residuals of the Morphological Status in the expected direction (\( \beta = -0.72, z = -2.57, p < 0.01 \)).

### 3.3.5. Frequency of the word relative to the frequency of its stem

**Variables.** As discussed in Section 1.2, English adverbs that are more frequent than their stems tend to show higher degrees of reduction (Hay, 2003). In contrast, interfixes in Dutch compounds tend to be longer the more probable they are given the compound’s constituents (Kuperman et al., 2007). We investigated whether the likelihood of the presence of the suffix +t as reflected by the log ratio of the frequency of the word and the frequency of its stem influenced its acoustic realization. We built a model for all word tokens ending in the suffix +t (\( N = 2110 \)).

**Results.** The significant predictors of the resulting model (M3) are shown in Table 2. The frequency ratio appeared to be a significant predictor: /t/ is more likely to be present in words with higher relative frequencies (i.e., word frequency relative to the frequency of its stem). This finding supports the Paradigmatic Signal Enhancement Hypothesis and thus suggests that this hypothesis also holds for inflectional morphemes.

There are two possible reasons why our results are in line with the results of Kuperman et al. (2007), rather than with the results of Hay (2003). First, whereas Hay (2003) investigated the reduction of a stem-final segment before a suffix, Kuperman et al. (2007)
investigated the reduction of the affix itself, as we did. Second, whereas adverbs always end in the suffix +en, there are three Dutch prefixes speakers have to choose from when building a compound. Similarly, in our study, speakers had to choose between several forms of the inflectional paradigm (+er, +t, or +en). Our results thus indicate that the informational load carried by the /t/ is reflected in its acoustic realization.

Our results for phonemically identical word pairs and the effect of relative frequency show that morphological structure affects the phonetic realization of words. Interestingly, Warner, Good, Jongman, and Sereno (2006) provided evidence that morphological structure only affects segmental duration if it is reflected in the words’ orthographic representations. In contrast to their study, which was based on read speech, our study is based on conversational speech. Future studies are necessary to draw conclusions about whether orthography plays similarly a strong role in spontaneous conversational speech as in read speech.

3.3.6. Syntactic structure

Variables. We added two independent variables capturing syntactic structure to model M1 (complete data set N = 6747). The first variable is Same Constituent, which has two values: either the target word and the following word belong to the same syntactic constituent, such as a noun phrase or an adverbial phrase, or they do not. The second variable is Constituent Length expressed in the Number of Syllables. Its range is shown in Table 1.

Results. Table 2 shows the results for this model (M4). Since Constituent Length interacted significantly with Same Constituent, we carried out separate analysis for /t/ tokens at the right edge of a syntactic constituent and /t/ tokens in the middle of a constituent. A constituent analysis revealed that the effect of Constituent Length was only significant for constituent final /t/ tokens: word-final /t/ is more likely to be present at the end of longer constituents. It is probable that the found effect of Constituent Length on phrase-final /t/ reflects prosodic final lengthening. Whereas it is unlikely that a prosodic boundary is placed between short syntactic constituents, such a boundary is more likely after long syntactic constituents, and a prosodic boundary often leads to stronger articulation of the preceding segment (Beckman & Edwards, 1990; Nespor & Vogel, 2007).

3.3.7. Segmental context

In the control model (M0), we only distinguished between vocalic and consonantal context and silence. In order to investigate the effects of the segmental context on the acoustic presence versus absence of /t/ in more detail, we built separate models, one for the subgroup of tokens where /t/ is preceded by a consonant, and one where it is followed by a consonant. Importantly, this data separation is possible since an initial analysis showed no significant interactions between the following and preceding context and because there were no collinearities between these variables. This data separation allows us to investigate effects of place and manner of articulation of neighboring consonants. Table 4 gives an overview of how often /t/ was absent in these different segmental contexts.

Variables. Place of Articulation could either be ‘homorganic’ or ‘heterorganic’ with the place of articulation of the /t/, which in Dutch is articulated at the alveolar ridge. The two independent variables Place and Manner of Articulation replace the predictor Previous Segment in M1, which has only one value left for this data set (i.e., ‘consonant’). Manner of Articulation had the values ‘plosive’, ‘fricative’, ‘nasal’, ‘glide’ and ‘liquid’.

Results: preceding consonant. First, we investigated the role of the consonant preceding /t/, Table 3 shows a statistical summary for the resulting model (M5). We observed that /t/s are more likely to be absent if preceded by a fricative (52.6%). To find out whether there were also significant differences between plosives (41.3%), nasals (45.3%), glides (30.0%) and liquids (27.5%), we ran the same model again, but excluding in subsequent steps fricatives, glides and liquids. We found significant differences between glides and liquids (β = 0.80, z = 2.07, p < 0.01), between glides and plosives (β = 0.92, z = −3.37, p < 0.001), between liquids and nasals (β = −0.59, z = −3.17, p < 0.001) and between liquids and plosives (β = −0.71, z = −8.78, p < 0.001). Not surprisingly, the percentages of absent /t/s after the most vowel-like consonants (i.e., glides and liquids) were similar to the percentage of absent /t/s after vowels (31.9%).

Results: following consonant. For the subgroup of /t/ tokens followed by a consonant, we built a model (M6 in Table 3) with the independent variables present in M1 (with the exclusion of Following Segment), and the Place of Articulation and Manner of Articulation of the following consonant. We observed that /t/ is absent least often before liquids (19.7%) and most often before homorganic plosives (31.8%). In order to find out whether the differences between fricatives (43.6%), nasals (41.8%) and glides (36.3%) were also significant, we ran the same model again, but excluding in subsequent steps fricatives, glides and liquids. We found significant differences between glides and liquids (β = 0.81, z = 2.05, p < 0.01), glides and plosives (β = −0.94, z = −3.39, p < 0.001), liquids and nasals (β = −1.05, z = −2.85, p < 0.001) and between nasals and plosives (β = −0.77, z = −3.71, p < 0.0001). Furthermore, significantly more /t/s were absent before a homorganic (51.8%) than before a heterorganic (40.2%) consonant. Plosives that are homorganic with /t/ are /t/ and /d/. Hence, this effect of place of articulation may be a mere proof of (voicing assimilation followed by) degemination (since Dutch does not allow geminate consonants). We therefore excluded all /t/ tokens followed by /t/ or /d/ and re-ran the model. The results were very similar to those of the previous model (Place-of-Articulation: Homorganic: β = −0.49, z = −4.21, p < 0.001; Manner of Articulation: plosive: β = −0.70, z = −4.90, p < 0.0001). We thus conclude that /t/s are less often present before homorganic than before heterorganic consonants and before plosives than before other consonants. Possibly, /t/ are more often absent before heterorganic plosives due to gestural overlap (Brownman & Goldstein, 1992).

<table>
<thead>
<tr>
<th>Predictor</th>
<th>β</th>
<th>z-Value</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>M5: Preceding context</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept = fricative</td>
<td>0.01</td>
<td>0.022</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>Word Frequency</td>
<td>−0.09</td>
<td>−2.91</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Residuals Bigram Frequency</td>
<td>−0.19</td>
<td>−2.96</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Previous Segment ‘glide’</td>
<td>1.37</td>
<td>3.03</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Previous Segment ‘liquid’</td>
<td>1.33</td>
<td>7.52</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Previous Segment ‘nasal’</td>
<td>0.82</td>
<td>4.43</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Previous Segment ‘plosive’</td>
<td>0.55</td>
<td>2.71</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Following Segment ‘vowel’</td>
<td>1.39</td>
<td>8.48</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Following Segment ‘silence’</td>
<td>0.93</td>
<td>5.85</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>M6: Following context</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept = fricative</td>
<td>0.66</td>
<td>3.08</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Word Frequency</td>
<td>−0.08</td>
<td>−4.30</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Residuals Bigram Frequency</td>
<td>−0.40</td>
<td>−3.80</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Previous Segment ‘vowel’</td>
<td>0.74</td>
<td>5.95</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Following Segment ‘glide’</td>
<td>0.38</td>
<td>1.79</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>Following Segment ‘liquid’</td>
<td>1.16</td>
<td>3.39</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Following Segment ‘nasal’</td>
<td>0.21</td>
<td>1.17</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>Following Segment ‘plosive’</td>
<td>−0.56</td>
<td>−3.48</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Following Place ‘homorganic’</td>
<td>−0.38</td>
<td>−2.71</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>
Table 4
Study I: absolute and relative numbers of absent [t]s in the different preceding and following contexts. Hom., homorganic place of articulation with [t]. Het., heterorganic place of articulation with [t].

<table>
<thead>
<tr>
<th>Preceding context</th>
<th>Vowel</th>
<th>Consonant</th>
<th>Manner of articulation</th>
<th>Place of articulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absent/total</td>
<td>1138/3570</td>
<td>1345/3177</td>
<td>167/404</td>
<td>426/810</td>
</tr>
<tr>
<td>% absent</td>
<td>31.5%</td>
<td>42.3%</td>
<td>41.3%</td>
<td>52.6%</td>
</tr>
<tr>
<td>Following context</td>
<td>494/1987</td>
<td>1890/4194</td>
<td>683/1235</td>
<td>659/1511</td>
</tr>
<tr>
<td>% absent</td>
<td>24.9%</td>
<td>45.1%</td>
<td>55.3%</td>
<td>43.6%</td>
</tr>
</tbody>
</table>

3.4. Summary

The first study of this paper investigated which linguistic and probabilistic properties predict the acoustic absence versus presence of word-final /t/ on the basis of 6747 tokens from a Dutch Corpus of spontaneous dialogues. First, we replicated earlier findings on effects of word frequency and contextual predictability (e.g., Bell et al., 2009; Jurafsky et al., 2001; Pluymaekers et al., 2005a; Torreira & Ernestus, 2009); /t/ tends to be absent more often in words of higher frequencies and in word combinations (bigram with the following word) of higher frequencies. In addition, we documented a role for the morphological properties of a word. On the basis of phonemically identical word pairs, we showed that /t/ tends to be less often absent if it also functions as a grammatical morpheme than if it is only part of the stem of the words. Further, the frequency of a word relative to the frequency of its stem predicts the absence versus presence of /t/; /t/ is more likely to be acoustically present in words with higher relative frequencies. This finding is in line with the Paradigmatic Signal Enhancement Hypothesis (Kuperman et al., 2007), and thus suggests that the hypothesis holds for inflectional paradigms as well as derivational paradigms. Moreover, we investigated the role of the syntactic properties of the utterance. Our data showed that /t/ is less likely to be absent at the end of longer syntactic constituents. Since prosodic boundaries are more likely at the end of longer constituents, this finding probably results from prosodic final lengthening. Finally, we observed that segmental context plays an important role in the realization of /t/. In line with previous reports, we found that /t/ is mainly absent in consonant clusters (Ernestus, 2000; Mitterer & Ernestus, 2006).

4. Study II

Study II is a detailed phonetic analysis of part of the material from Study I. The automatically generated broad phonetic transcriptions used in Study I treat the signal as if it consists of beads on a string, with each bead representing a single, clearly realized phone (Ostendorf, 1999). As a consequence, pronunciation variation could only be captured as phone substitution, insertion or deletion. However, phonetic reality is more complex. Especially, speech of an informal speaking style, like our material, may show periodicity or only contain periodicity of rapidly decreasing amplitude. All burst segmentations and with energy in only part of the spectrum. In addition, bursts were labeled as strong or weak, one short impulse, separated from friction noise of the next segment by short duration and relatively strong intensity. We classified a burst as ‘multiple burst’ if there was two or more release impulses that are distinct from the friction noise of the next segment by short duration and relatively strong intensity. We classified a burst as ‘single burst’ if there was one short impulse, separated from friction noise of the next segment. In addition, bursts were labeled as strong or weak, where weak bursts were characterized by extremely short durations and with energy in only part of the spectrum. All burst labels were based on the bursts’ acoustic representations in the spectrograms.
Table 5
Counts of acoustic observations. Nas. F., nasal friction; Nas., nasal constriction; Smooth Start /t/, smoothly starting friction of /t/; Smooth Start F., smoothly starting friction of the following consonant; Simultaneous, Smooth friction of any kind is starting immediately after the burst; Alveolar F., alveolar friction. In brackets: For Constriction and Alveolar F., numbers in brackets indicate the number of these tokens that were voiced in the constriction; for Burst, the numbers in brackets indicate the number of these bursts that were labeled as strong.

<table>
<thead>
<tr>
<th>Property</th>
<th>Present</th>
<th>Absent</th>
<th>Closure</th>
<th>Frict.</th>
<th>Nasal frict.</th>
<th>Nasal Status</th>
<th>Canonical /t/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constriction</td>
<td>392 (77)</td>
<td>94</td>
<td>244 (44)</td>
<td>113 (19)</td>
<td>27 (9)</td>
<td>8 (5)</td>
<td>Closure</td>
</tr>
<tr>
<td>Burst</td>
<td>254 (115)</td>
<td>232</td>
<td>102 (51)</td>
<td>152(63)</td>
<td></td>
<td></td>
<td>One (strong)</td>
</tr>
<tr>
<td>Smooth Start /t/</td>
<td>368</td>
<td>24</td>
<td>209</td>
<td></td>
<td></td>
<td></td>
<td>Absent</td>
</tr>
<tr>
<td>Smooth Start F.</td>
<td>239</td>
<td>114</td>
<td>108</td>
<td></td>
<td></td>
<td></td>
<td>Absent</td>
</tr>
<tr>
<td>Alveolar Frict.</td>
<td>391 (28)</td>
<td>95</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Present</td>
</tr>
</tbody>
</table>

We observed that acoustically reduced /t/‘s abruptly followed by friction of the next segment are perceived as strong /t/‘s (compared to /t/‘s where the friction of the next segment starts smoothly). Therefore, the two labelers also scored each /t/ token for whether following friction in the signal started gradually (see Fig. 1, right panel) instead of abruptly (see Fig. 2, left panel). If the friction was part of the /t/, information on its start was captured by the variable Smooth Start /t/ and if it was part of the following consonant this information was captured by Smooth Start of Following consonant. Both types of friction were classified as simultaneous if they started immediately after the /t/ burst. An example of simultaneous friction can be seen in Fig. 1, left panel, one of non-simultaneous friction in Fig. 2, right panel.

Finally, the place of articulation of the friction was determined auditorily. The variable Alveolar Friction was scored as present for tokens with alveolar friction and as absent for tokens with friction of other places of articulation.

4.2. Analysis method

To investigate which variables predict the presence of each of these sub-phonemic acoustic properties, we built statistical models using mixed effects logistic regression with Speaker, Word and Following Word as crossed random variables. As in Study I, we first built control models, using the same set of independent variables, to which we separately added the probabilistic word predictability variables, the morphological variables, the syntactic variables and the variables that capture details of the segmental context. All independent variables were determined as described for Study I. As in Study I, for this subset of the material, the logged Word Frequency (min=0, max=12.1, mean=6.2) and Bigram Frequency (min=0, max=2.4, mean=1.2) are correlated ($r = 0.44$, $p < 0.0001$). As in Study I, we orthogonalized Word Frequency and Bigram Frequency by replacing Bigram Frequency with the residuals of the linear regression model predicting Bigram Frequency as a function of Word Frequency. Predictors and interactions that were not statistically significant were removed from the models and in the following only the significant effects will be presented. If Word Frequency was not significant, but the residuals of Bigram Frequency were, we removed Word Frequency and re-ran the model with the independent variable Bigram Frequency (instead of its residuals).

4.3. Results

Table 5 presents the counts of all acoustic observations. For each sub-phonemic property, the columns Present and Absent show the number of tokens in which this property was present and absent, respectively. For those tokens that were realized with the sub-phonemic property, the columns on the right further specify their realizations. This table shows that there is great variation on the sub-phonemic level, even for tokens coming from a limited number of segmental contexts. Only 11.7% of the tokens show canonical realizations of /t/ (i.e., a voiceless closure, one strong burst, produced at an alveolar place of articulation). There are nearly as many tokens of /t/ without bursts (232) as with bursts (254) and out of the 254 tokens with bursts only 152 have a single burst. Further, only 200 of the 486 tokens were realized with the canonical voiceless closure and in 95 tokens the place of articulation was auditorily assimilated to that of the following consonant. Complete absence of any /t/ properties occurred in 8.0% of the tokens.

The statistical analysis of our data did not show significant effects for the morphological and syntactic variables for any of the sub-phonemic properties, even though the data showed similar trends as in Study I. Most probably the numbers of tokens (137 for the testing of morphological properties and 486 tokens for the testing of syntactic properties) were too small. We also did not find significant predictors for the presence of the sub-phonemic property Smooth-Start /t/, which is not surprising since this property was present in more than 90% of the tokens (see Tables 6 and 7). The results for the other sub-phonemic properties will be presented below. For each property, we first describe the model that predicts the presence versus absence of the property for the complete data set. Next, we present models that investigate details of the property based on the subsets of tokens possessing that property. In a last step, we investigate the effects of place and manner of articulation of neighboring consonants on the sub of /t/-tokens in consonantal segmental context.

4.3.1. Constriction

The first statistical model (M7 in Table 8) investigated the presence of a constriction. In total, 80.7% of the tokens of word-final /t/ were realized with some form of constriction. Bigram Frequency appeared a significant predictor: Constructions are more often present in word combinations of lower frequencies. Furthermore, the independent variables Previous Segment and Following Segment showed significant effects. More constructions are found after (84.1%) and before (91.9%) vowels than after nasals (68.5%) and before consonants (76.4%). A detailed overview

3 In the model for the sub-phonemic property Constriction, Morphological Status was marginally significant ($\beta = -0.489741, z = -1.761, p < 0.1$). However, we do not include marginally significant variables in the final models.
of the presence of a constriction in the different contexts is given in Tables 6 and 7.

For those tokens with a constriction (*N*=392), we investigated which factors predict a complete closure. We observed that significantly more constrictions are realized as complete closure if the Preceding Segment is a vowel (66.7%, $\beta = -0.76$, $z = -2.40$, $p < 0.01$) than if it is a nasal (43.2%) and if the Following Segment is a vowel (75.8%, $\beta = -0.68$, $z = -2.34$, $p < 0.01$) than if it is a consonant (57.3%). Separate analyses of tokens followed by a consonant showed an interaction between Place and Manner of articulation ($\beta = -1.47$, $z = -2.59$, $p < 0.01$). Constrictions are less often complete before homorganic fricatives (51.0%) than before heterorganic fricatives (70.4%, $\beta = -0.48$, $z = 1.69$, $p < 0.05$) and more frequently incomplete before heterorganic plosives (76.7%, $\beta = -0.78$, $z = -1.95$, $p < 0.05$) than before homorganic plosives (80.0%).

### 4.3.2. Burst

In total, 52.3% of the tokens were realized with a burst, which consisted of a single burst in 31.3% of the tokens and of multiple bursts in 21.0%. Model M8 in Table 8 shows the predictors for the presence of a burst. As for the constriction, Bigram Frequency is a significant predictor: Bursts are less probable in word combinations of a higher frequency. Also significant was the frequency of the word: Bursts are less probable in words of a higher frequency. Finally, segmental context also showed a significant effect: Bursts occur more often after nasals (64.8%) than after vowels (48.7%), and before vowels (65.2%) than before consonants (33.3%) before fricatives and 54.5% before plosives. Tables 6 and 7 show a summary for the different contexts. We based another model on the subset of tokens that were realized with a burst in order to investigate the predictors for bursts realized with a single versus with multiple bursts, but we did not find any significant predictors.

We then investigated effects of place and manner of articulation of a directly following consonant (*N*=166). We only found significant effects for the presence of one versus multiple bursts: There was an interaction of Manner and Place of articulation ($\beta = -3.67$, $z = -3.60$, $p < 0.0001$). Multiple bursts are more likely than a single burst if word-final /t/ is followed by a homorganic (52.8% multiple burst) than by a heterorganic fricative (24.6%), but less likely if /t/ is followed by a homorganic (29.2%) than by a heterorganic plosive (77.8%).

### 4.3.3. Voicing

Model M9 investigated which variables predict closure voicing of the word-final /t/. It shows that /t/ closure is significantly more often observed after a nasal (23.1%) than after a vowel (16.7%), and significantly more often before a vowel (27.4%) than before a consonant (17.5% before plosives and 13.0% before fricatives). Moreover, voiced realizations are more likely in word combinations of higher frequencies (consistent with the increased likelihood of the burst being absent).

Since voicing may result from regressive voice assimilation with the following voiced plosive (e.g., Booij, 1995), we carried out a separate analysis of all tokens followed by a plosive (*N*=120). We observed that /t/ is indeed significantly ($\beta = 1.95$, $z = 3.00$, $p < 0.0001$) more often realized with periodicity in the closure before voiced plosives (24.2%) than before unvoiced plosives (8.0%).

### 4.3.4. Smooth start of friction

For tokens followed by an obstruent (*N*=349), we investigated the start of the friction of that obstruent. Model M10 in Table 8 shows that this friction is less likely to start smoothly if the /t/ follows a vowel (54.7%, cf. Table 6) than if it follows a nasal (97.2%). Since an abrupt start results in a perceptually strong /t/, we conclude that, as for the other sub-phonemic properties, less reduced realizations of /t/ are more likely after vowels than after consonants. Separate analyses showed that a smooth start of the

---

4 These results do not include the prior probability of there not being a burst present before a plosive.
friction of the following consonant is significantly ($\beta = -17.71$, $z = -3.58$, $p < 0.0001$) less likely before plosives (7.5%) than before fricatives (98.6%).

4.3.5. Alveolar friction

In order to investigate the predictors for the presence of Alveolar Friction (model M11 in Table 8), we excluded tokens of /t/ followed by /s/, /z/, /t/ and /d/, since friction in these tokens is always alveolar. In total, 76.2% of the remaining tokens of word-final /t/ were produced with alveolar friction. Our analysis showed that, similar to the other sub-phonemic properties, alveolar friction occurs more often if the word-final /t/ is followed by a vowel (88.1%, $\beta = 1.73$, $z = 2.57$, $p < 0.01$) than if it is followed by a consonant (67.7%, cf. Table 7), and if the /t/ is part of bigrams of lower frequencies.

4.4. Discussion

Study II was based on 486 tokens of word-final /t/ that were manually annotated for a set of sub-phonemic phonetic properties. The tokens showed a wide variety of realizations. In only 11.7% of the tokens, /t/ was realized canonically and in 8.0% of the tokens it was completely absent. The bigram frequency with the following word proved to be a significant predictor of acoustic features: Constrictions, bursts and alveolar friction are more likely to be absent in word combinations of higher frequencies, while voicing is more likely in word combinations of higher frequencies. These results give further evidence that gestural overlap, strong co-articulation and processes of lenition are more pervasive in highly predictable contexts. Our results, therefore, support the generalization that words in more frequent contexts tend to be more reduced (e.g., Bell et al., 2009; Jurafsky et al., 2001; Torreira & Ernestus, 2009).

For most sub-phonemic properties, we observed that /t/ is reduced more in consonantal than in vocalic context: /t/ is less often realized with a constriction, with a burst and with alveolar friction before a consonant than before a vowel. Moreover, the friction of the obstructant following /t/ is more likely to start abruptly if /t/ is preceded by a vowel than a nasal. In contrast to this tendency, bursts are longer and more often present after nasals than after vowels. One explanation could be that an /nt/ cluster implies a long oral closure during which air pressure is built up in the vocal tract, while there is some nasal friction due to air escaping through a small velar opening. The release of this closure results in a strong burst. In order to be able to draw conclusions about whether bursts are stronger after /n/ than in any other preceding context, further studies based on /t/ tokens from all consonantal contexts are necessary.

Detailed analyses of the constriction showed that fewer constrictions were complete after /n/ and before consonants. This result is not surprising, since constrictions with nasality and friction originate from co-articulation with neighboring nasals and obstruents. Separate analyses of tokens followed by a consonant showed that constrictions are less likely to be complete before a homorganic than before a heterorganic fricative.

Detailed analysis of the burst showed that multiple bursts are more likely if word-final /t/ is followed by a homorganic than by a heterorganic fricative. This may indicate that multiple bursts are caused by the Bernoulli effect: Since the release of the closure is slow before homorganic fricatives, whose articulatory targets are only slightly removed from the target for the /t/, the tongue tip may be ‘sucked back’ against the alveolar ridge due to under-pressure caused by the fast airstream through the small constriction. Before heterorganic fricatives, in contrast, the opening at the alveolar ridge can be expected to be faster and therefore multiple burst are less probable.

Finally, we have observed that 18.1% of all tokens of word-final /t/ were realized with periodicity in the close. Our analysis showed that /t/ closures are voiced more often after nasals than after vowels. Since a nasal already is a voiced closure it is likely that also the subsequent /t/ is realized with periodicity in the close, compared to /t/-closures that are realized after a vowel. Furthermore, the data showed that /t/ closures are voiced more often before vowels than before consonants. Separate analyses of those tokens followed by a plosive showed that /t/ closures are less often realized as voiced before phonologically unvoiced plosives than before phonologically voiced plosives.

5. General discussion and concluding remarks

This paper investigated the role of linguistic and probabilistic properties of words in the reduction of word-final /t/ on the basis of a Dutch Corpus of spontaneous dialogues. Study I was based on automatically generated broad phonetic transcriptions of 6747 tokens, which allowed us to investigate the acoustic absence versus presence of /t/. Study II was based on a manual annotation of a number of sub-phonemic properties for a small subset of the tokens used in Study I (486), which allowed us to investigate the details of the acoustic realization of /t/.

Study I documented a role of morphology in the absence versus presence of /t/. The first piece of evidence comes from phonemically identical word pairs of which one member is a verb form such as vind-‘[he] finds’, where /t/ marks the second or third person singular present tense, and one member ends in /t/ that is only part of the stem (e.g., vind-‘[s] find’). Such word pairs show that /t/ tends to be less often absent if it functions as a grammatical morpheme than if it does not. Second, the role of morphology appears from the effect of the frequency of a morphologically complex word relative to the frequency of its stem (e.g., the frequency of loopt relative to the frequency of loop). We found that ‘morphological’ /t/ is more likely to be present the higher the frequency of the word relative to the frequency of its stem. Since in Dutch the stem (i.e., loop for the earlier example) corresponds to the verb form in first person singular present tense, our results also mean that /t/ is more likely to be present the higher the frequency of a verb in third or second person singular relative to the verb in first person singular present tense. Our findings, therefore, support the Paradigmatic Signal Enhancement Hypothesis (Kuperman et al., 2007), which states that the most likely alternative in a paradigm is realized with greater acoustic salience. These results are contrary to what Hay (2003) found for English adverbs. Our explanation for these opposite observations is that whereas adverbs always end in the suffix +-ly, the speakers of our Dutch material had to choose between several forms of the inflectional paradigm (suffixes +∅, +t, or +en). The morphological /t/ in our data may therefore also carry more information than the /t/ in the study by Hay (2003).

Study I also documented a role of syntax in the absence versus presence of /t/. Word-final /t/ is less likely to be acoustically absent if the target word is situated at the end of a long syntactic constituent than at the end of a short constituent. This finding probably results from prosodic final lengthening because stronger prosodic boundaries are more likely after longer syntactic constituents. Unfortunately, the ECSD does not (yet) come with prosodic annotation and further analysis is therefore left for future research.

Both Study I on the absence versus presence of word-final /t/ and Study II on the acoustic sub-phonemic properties of /t/ showed that /t/ tends to be more reduced in word combinations
of higher frequencies: /t/ tends to be absent more often and it tends to be realized without complete constrictions, without strong bursts and with voicing in the closure. These results are in line with earlier findings on effects of contextual predictability (e.g., Bell et al., 2009; Jurafsky et al., 2001; Pluymaekers et al., 2005a; Torreira & Ernestus, 2009) and support the general assumption that highly predictable words tend to be more reduced.

Also, segmental context played a role at both the phonemic and the sub-phonemic level: /t/ is more likely to be acoustically absent, and so are its sub-phonemic properties, if the /t/ is preceded or followed by phonological consonants than by phonological vowels. These observations are in line with previous reports by Ernestus (2000) on casual Dutch and by Mitterer and Ernestus (2006) on read Dutch.

Theoretically, the absence of /t/ can result from its absence in the selected pronunciation variants in the mental lexicon, from categorical phonological deletion rules or from gradient phonetic reduction processes. Our data suggests that the acoustic absence of /t/ is the natural endpoint of gradient reduction, since both the absence versus presence of /t/ and of its detailed acoustic properties are predicted by the same phonetic and probabilistic predictors. The absence of word-final /t/ in Dutch thus seems to result from similar mechanisms as for instance the absence of schwa in English (Davidson, 2006).

When listening to the data, we had the impression that in a sequence of /t/ + non-homorganic fricative (e.g., /tf/), the /t/ was clearly perceptible as /t/, even though its realization was silence followed by an abrupt rise in amplitude at the start of frication, presumably due to the abrupt release of the closure overlapping with the non-homorganic fricative articulation. This impression was supported by systematic analyses in Schuppler, Ernestus, van Dommelen, and Koreman (2010). This suggests that future studies on reduction in casual speech not only have to consider sub-phonemic properties of the segment itself but also of the surrounding segments, especially since listeners have been shown to use these properties when listening to carefully pronounced speech. For instance, Heinrich and Hawkins (2009) cross-spliced minimal word pairs such as millor versus mirror into a sentence that either matched or did not match the original word. The matched condition provided better intelligibility for /t/ words for young and elderly listeners in adverse noise conditions than the non-matched condition. This indicates that listeners may rely on sub-phonemic properties that occur several segments away from the relevant segment. These findings show the necessity to take the sub-phonemic properties of the neighboring segments into account in the analysis of reduction and in the development of models of speech production and perception.

Study II was a first step in the investigation of probabilistic and linguistic properties conditioning sub-phonemic properties. In order to extend this study to /t/ in word-initial and word-medial position and to other segments, large amounts of annotated data are required. Manually transcribing such properties is extremely time consuming and error prone, since the relevant acoustic properties are both gradient and complex. Preferably therefore, these transcriptions should be generated by means of automatic procedures. On the basis of the results presented in Study II, it seems feasible to develop a two-stage transcription procedure for the annotation of /t/. In the first stage, a broad phonetic transcription, like the one used in Study I, is created, which shows the speech intervals that could contain /t/. In the subsequent stage, classifiers trained on a small amount of manually labeled data could be applied to provide a full description of the sub-phonemic properties. This procedure may also make automatic transcription tools more sensitive to reduced variants of /t/ that are still well perceptible for human listeners (see Schuppler et al., 2010).

Similarly, our findings may inform work on psycholinguistic models of speech production and perception. Production models have to incorporate the types of (sub-phonemic) reduction documented in this paper and to account for why these reduced forms are predicted by bigram frequency, several segmental contexts, and above all, several morphological and syntactic structures. Research in the comprehension domain has to show how listeners use the detailed phonetic information available in the speech signal. Mitterer, Yoneyama, and Ernestus (2008) showed that listeners make use of perceptual mechanisms, learning strategies and lexical knowledge when perceiving different types of reduced /t/. Detailed studies have to show whether the contribution of these processes is similar for the different variants and whether these processes are influenced by the likelihood of reduction given the linguistic and probabilistic properties documented in this paper.

In conclusion, this paper presented an analysis of the conditions favoring the reduction of word-final /t/, based on a corpus of conversational Standard Dutch. We showed that syntactic and morphological structure predict the acoustic presence versus absence of word-final /t/ and that bigram frequency and segmental context affect both the presence of /t/ and its sub-phonemic qualities.

Acknowledgements
This research was supported by the Marie Curie Research Training Network “Sound to Sense”. Mirjam Ernestus was supported by a European Young Investigator award from the European Science Foundation. We would like to thank Geertjan van Noord for his help with creating the syntactic annotations of the corpus.

Appendix A. Supplementary material
Supplementary data associated with this article can be found in the online version of http://dx.doi.org/10.1016/j.wocn.2012.05.004.

References


