ELSEVIER
Volume 39
Issue 1
January 2011
ISSN 0095-4470
Journal of
Phonetics
Editor
Kenneth de Jong
Associate Editors
Ocke-Schwen Bohn
Taehong Cho
Jennifer Hay

Journal of Phonetics 39 (2011) 96–109

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Letter to the Editor

Acoustic reduction in conversational Dutch: A quantitative analysis based on automatically generated segmental transcriptions

Barbara Schuppler [a,*], Mirjam Ernestus [a,b], Odette Scharenborg [a], Lou Boves [a]

[a] *Center for Language and Speech Technology, Radboud University Nijmegen, Erasmusplein 1, 6525 HD Nijmegen, The Netherlands*
[b] *Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands*

ARTICLE INFO

Article history:
Received 2 September 2009
Received in revised form
12 November 2010
Accepted 22 November 2010
Available online 11 January 2011

ABSTRACT

In spontaneous, conversational speech, words are often reduced compared to their citation forms, such that a word like *yesterday* may sound like [ˈjɛʃeɪ]. The present chapter investigates such acoustic reduction. The study of reduction needs large corpora that are transcribed phonetically. The first part of this chapter describes an automatic transcription procedure used to obtain such a large phonetically transcribed corpus of Dutch spontaneous dialogues, which is subsequently used for the investigation of acoustic reduction. First, the orthographic transcriptions were adapted for automatic processing. Next, the phonetic transcription of the corpus was created by means of a forced alignment using a lexicon with multiple pronunciation variants per word. These variants were generated by applying phonological and reduction rules to the canonical phonetic transcriptions of the words. The second part of this chapter reports the results of a quantitative analysis of reduction in the corpus on the basis of the generated transcriptions and gives an inventory of segmental reductions in standard Dutch. Overall, we found that reduction is more pervasive in spontaneous Dutch than previously documented.

1. Introduction

Each speaking style has its own characteristics. In spontaneous speech, words are often reduced compared to their canonical pronunciations, such that a word like *yesterday* may sound like [ˈjɛʃeɪ]. A study on American English showed that whole syllables are absent in 6% of the word tokens and that segments are absent even in every fourth word (Johnson, 2004). Recent linguistic research has investigated reductions of different degrees as well in other languages, from segment shortening and lenitions (e.g., Janse, Nooteboom, & Quené, 2007 for Dutch) to the deletion of segments and syllables (e.g., Adda-Decker, Boula de MareuBooil, Adda, & Lamel, 2005 for French), to the absence of complete words (e.g., Kohler, 1998 for German). The present study contributes to the research on reduction by quantifying how often specific segment deletions and substitutions occur in spontaneous Dutch on the basis of automatically generated segmental transcriptions.

Statistics about segment deletions and substitutions are necessary to improve automatic speech recognition (ASR) systems. Reduced word forms do not match well with their canonical pronunciations, which are often the only ones stored in the pronunciation lexicons of such systems. This mismatch leads to recognition errors. Saraçlar, Nock, and Khudanpur (2000) showed that pronunciation variability correlates with the recognition error rate of ASR systems. They orthographically transcribed conversational speech, which then was read by the same speakers. The word error rate for the original data was more than 50% higher than for the read version. One solution for dealing with spontaneous speech is to add reduced variants to the ASR lexicon. However, this approach has its limits because as the number of pronunciation variants increases, the internal lexical confusability increases as well: For instance, if the pronunciation variant [hɛd] is permitted for the English word *had*, it can be confused with the canonical pronunciation of *head* (Saraçlar et al., 2000). Adding variants can only help in conjunction with accurate estimates of the conditions under which specific reductions are likely to occur.

Research on reductions and the conditions under which specific variants occur is also of importance for psycholinguistic models of speech production and perception. Most models do not account for the pronunciation variation found in spontaneous conversations (e.g., Levelt, Roelofs, & Meyer, 1999; Norris, McQueen, & Cutler, 1995). Information about the conditions that favor the occurrence of specific pronunciation variants is necessary to adapt existing psycholinguistic models so that they can deal with spontaneous speech (Scharenborg & Boves, 2002). Information about the frequency of pronunciation variants is also important for research on the structure of the mental lexicon (e.g., Connine, Ranbom, & Patterson, 2008).

[*] Corresponding author. Tel.: +31 2436 11668; fax: +31 2436 11070.
E-mail address: barbara.schuppler@gmail.com (B. Schuppler).

0095-4470/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wocn.2010.11.006

Reliable estimates of the conditions under which specific pronunciation variants occur require large corpora with suitable phonetic transcriptions. Broadly speaking, there are two ways to obtain segmental transcriptions of speech corpora. Traditionally, transcriptions are produced manually by one or more human transcribers. This method is not restricted to segmental transcriptions, but also gives the possibility to annotate materials on a fine phonetic level (e.g., Mitterer & Ernestus, 2006). Since this approach is time consuming, only a relatively small amount of data can be processed. Moreover, human transcribers are influenced by their expectations, which is especially an issue in the transcription of reduced speech. For instance, Ernestus (2000) reported that her three transcribers disagreed about the presence versus absence of the first vowel of the word *natuurlijk* 'of course' for 58% of the 274 tokens. Also other studies show very high inter-transcriber inconsistencies for the transcription of spontaneous speech (e.g., Kipp, Wesenick, & Schiel, 1997) and the question arises how to deal with this inter-transcriber disagreement (Rietveld, van Hout, & Ernestus, 2004).

A more recently available method is to create phonetic transcriptions by using an automatic speech recognition (ASR) system to determine the most likely pronunciation variant for each word in a spoken corpus (e.g., Binnenpoorte, 2006; Cucchiarini & Binnenpoorte, 2002; Van Bael, Boves, van den Heuvel, & Strik, 2007). With this method large amounts of speech material can be transcribed in a relatively short period of time. Furthermore, ASR systems do not have expectations as humans do. Their choices are tractable because they are limited by the possible pronunciation variants in the lexicon or the rules that can be applied internally to generate such variants. Errors and inaccuracies can still occur, but they are systematic throughout the whole corpus and can therefore be taken into account in the analysis of the transcriptions. However, there are also disadvantages of the automatic approach. First, conventional ASR systems have difficulty processing segments with very short durations. If the presence of such segments is detected at all, our experience showed that almost invariably the boundaries are misplaced. Second, while in principle automatic transcription tools can transcribe phonetic details, systems that can do this reliably are still in their infancy. Finally, it is not only humans that provide more reliable transcriptions for read than for spontaneous speech, automatic transcription tools perform better on read than on spontaneous speech as well (Cucchiarini & Binnenpoorte, 2002).

In this paper, we analyze the frequency of occurrence of reductions in spontaneous speech at the segmental level. For this purpose, we compare the segmental transcriptions of the words in our speech material with their canonical pronunciations. We consider a word as reduced if it is produced with either a lower number of segments (i.e., the absence of segments) or if a phone is produced with less articulatory effort (e.g., a full vowel realized as schwa or a long vowel realized as a short vowel, so called *lenitions*). For this kind of analysis, we need segmental transcriptions of large amounts of spontaneous speech material. In the first part of the chapter, we describe the method with which we automatically transcribed a corpus of spontaneous speech. We used a lexicon with many pronunciation variants for each word, which we generated by means of rules applied to the canonical pronuniciations. Contrary to Cucchiarini and Binnenpoorte (2002) and Van Bael and Boves et al. (2007), whose rules were insensitive to the stress pattern and syllable structure of the word, our rules are sensitive to this information. As a result, we obtained a larger number of probable variants. In addition to segment deletion and lenition rules, we incorporated a wider range of co-articulation and phonological rules in order to improve the coverage of plausible variants.

In the second part of the paper, we focus on the main goal, which is to obtain a better understanding of the conditions under which reductions occur and with which frequencies. With the present study we aim at quantifying rules which have earlier been mentioned in the phonological literature and/or in the phonetic literature based on impressionistic observations (e.g., Ernestus, 2000). With the term 'rules' we refer to the simple mapping from the segmental transcription of the canonical pronunciation of a word to the pronunciation variant that (1) is generated for the lexicon used to automatically transcribe the corpus and (2) is present in the speech material.

In contrast to previous quantitative research on segment deletion in Dutch, which have only given absolute deletion rates of phones (e.g., Kessens, Wester, & Strik, 2000; Van Bael, Baayen, & Strik, 2007; Wester, Kessens, & Strik, 1998), we also analyze consonant and vowel reductions in terms of their frequencies relative to the frequencies with which these reductions could have occurred given the words in the corpus. Moreover, we also investigate the deletion of full vowels and the frequencies of co-articulation and phonological rules and we analyze which segmental contexts favor these rules.

The rest of this paper is organized as follows. In Section 2, we present the corpus of Dutch spontaneous dialogues used in the study. Section 3 is dedicated to the automatic generation of the phonemic transcription of this corpus. In Section 4, we present and discuss the results of the analysis of reductions based on the automatically generated transcriptions. The paper ends with a discussion of the findings.

2. Corpus data

The corpus used in this study is the Ernestus Corpus of Spontaneous Dutch (*ECSD*, Ernestus, 2000), which contains spontaneous conversational Dutch. All conversations in this corpus were produced by healthy, male native speakers of Dutch of similar social and economic background. They lived in the western provinces of the Netherlands and have academic degrees. The speakers were between 21 and 55 years old. They were classified as speakers of standard Dutch by trained phoneticians.

To obtain spontaneous conversations the following set-up was used in the recordings: Pairs of colleagues or friends talked with each other, seated some 1.5 m from each other at a table in a soundproof room. The speakers chose the topics for the first 40 min of the conversations freely. The second part of the recording was a role-play, where the speakers negotiated about the purchase of camping goods. In the role-play the speakers pursued partly conflicting goals that were assigned to each speaker individually before the start of the recording session; no further instructions were given. The experimenter was only present during the first part, but hardly participated in the conversations. This set-up resulted in dialogues with a casual, chatty style. All conversations have a duration of approximately 90 min. In total, 153,200 word tokens and 9035 word types were spoken in 15 h of recordings.

The recordings were made with two Sennheiser MD527 supercardioid microphones, one on each channel, on Sony DAT tapes. The available orthographic transcription was realized in the Praat Long TextGrid format (Boersma, 2001), where different tiers were used for the different speakers. The orthographic transcriptions were manually aligned with the speech signal in chunks, which are stretches of speech that are transcribed as one complete unit. Fig. 1 shows an example from the transcription: while Speaker 1 (transcribed on the first tier, recorded on the left channel) is speaking, Speaker 2 begins to laugh.

Table 1 shows a summary of characteristic properties and word types that reflect the casual chatty speaking style of the corpus. First, we note that speakers often (898 times) produced noises other than speech, such as laughter, and that not all produced

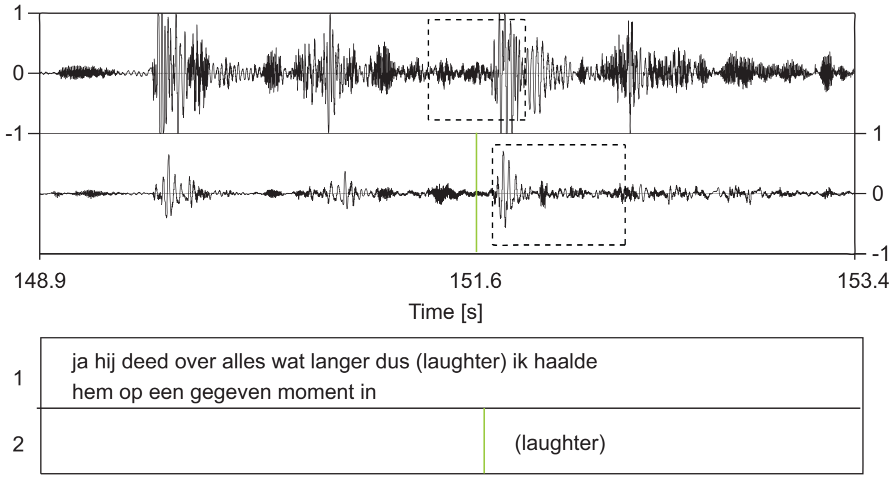

Fig. 1. Example for a chunk of ECSD: Before rechunking. The dashed lines mark stretches of laughter.

Table 1
Properties of the Ernestus Corpus of Spontaneous Dutch showing its spontaneous speech style.

Tokens	Total number
Word tokens	153,200
Word types	9035
Hapax legomena	4879
Laughter and other speaker noises	898
Chunks with unintelligible speech	115
Backchannels (*hm*)	819
Word *ja* 'yes'	6471
Word *maar* 'but'	2451
Word *nou* 'now', 'well'	1572
Word *nee* 'no'	1238
Broken words	376
Explicit disfluencies *eh, ah, uh*	3677
Speaking errors	25
Onomatopoeia	72

speech is intelligible to the transcribers afterwards (115 chunks of a total length of 138 s). Second, we see that among the most frequent word types of the corpus are backchannels (*hm*) and backchannel-like words such as *ja* 'yes', *maar* 'but', *nou* 'now, well' and *nee* 'no'. The table shows the total number of occurrences of these word types. Note that not all of the tokens also function as backchannel. These four word types account for 8.2% of all word tokens. Furthermore, disfluencies are relatively common in spontaneous speech. A high number of broken words and fillers *eh, ah, uh* is another indication of the degree of casualness of a corpus. In *ECSD* we counted a filler rate of 2.4 per 100 words. This number is nearly identical to the number that Bortfeld, Leon, Bloom, Schober, and Brennan (2001) reported for a corpus of American English conversations (overall mean of 2.6 of the fillers *eh, ah, uh* per 100 words). Finally, we also observed speaking errors (e.g., *rugzak* 'backpack' produced as [ˈrʏxslɑk] instead of [ˈrʏxsɑk]) and onomatopoeia, which are words that imitate the source of the sound they are describing, such as 'tring tring' for a telephone. Another characteristic of the corpus is the relatively high proportion of word tokens that occur only once, i.e., *hapax legomena* (54.0%), most probably because all free conversations were about different topics.

Besides the characteristics shown in Table 1, a large amount of overlapping speech is typical for conversational speech. After the rechunking procedure described in Section 3.2.2, 38.1% of the chunks with an average chunk length of 1.95 s contained overlapping speech. This is very similar to what Chino and Tsuboi (1996) report for a corpus of Chinese spontaneous telephone dialogues (40% overlap with an average chunk length of 1.75 s).

3. Creating a broad phonetic transcription automatically

3.1. Introduction: forced alignment

The phonetic transcription was created by means of a forced alignment. Input to the forced alignment procedure are the speech files organized in chunks, the orthographic transcriptions of the chunks, a lexicon containing multiple pronunciation variants of all word types in the corpus, and acoustic models for the phones used to specify the pronunciation variants. An ASR system determines the most likely pronunciation variants for the sequence of words in each speech chunk. Note that the pronunciation variation that can be captured is limited by the set of phonetic symbols for which acoustic models have been trained and which typically represent the phonemes of the language. Therefore, we speak of a *broad phonetic* (or *phonemic*) *transcription*. Furthermore, the variation that can be captured depends on the pronunciation variants incorporated in the lexicon.

The ASR system we used was based on the hidden Markov model Toolkit HTK (Young et al., 2002). The acoustic phone models used for all alignments presented here were 37 32-Gaussian tri-state monophone acoustic models (Hämäläinen, Gubian, ten Bosch, & Boves, 2009) that had been trained on 396,187 word tokens of the Dutch Library of the Blind of the *Spoken Dutch Corpus* (CGN, Oostdijk et al., 2002). The models were trained at a frame shift of 5 ms and a window length of 25 ms, where for each frame 13 MFCCs (i.e., the mel-scaled cepstral coefficients C0–C12) and their first and second order derivatives (39 features) were calculated. We used a shorter frame shift than the default of 10 ms used in earlier studies of segmental reductions (e.g., Adda-Decker et al., 2005; Schuppler et al., 2009; Van Bael & Boves et al., 2007) in order to achieve more accurate positions of the segment boundaries and to be able to identify very short segments. With a frame shift of 5 ms and acoustic models consisting of three emitting states (no skips), segments will be assigned a minimum length of 15 ms. This does not mean that shorter segments cannot be annotated, but that their segment boundaries will be placed within the neighboring segments.

3.2. Adapting the existing orthographic transcription

3.2.1. Adapting the verbatim orthographic transcription

Many available corpora have been recorded before automatic transcription became possible. This was also the case for the *ECSD*, which was collected in the mid-1990s. While the original orthographic transcription was perfectly suitable for manual analysis, adaptations were necessary for allowing automatic processing. We transformed the transcriptions to the standards developed in the CGN project (Oostdijk et al., 2002). First, we annotated audible noises. This includes the annotation of laughter, as well as the annotation of filled pauses, where we limited ourselves to the word types shown in Table 1. We used mark-up symbols to annotate broken words (*), speaking errors (\v), onomatopoeia (\o), and when the speaker was spelling a word (\-). Furthermore, inconsistencies in the spelling of words were corrected and the use of capital letters was limited to proper nouns. Moreover, digits were transcribed as full orthographic words. These adaptations decreased the original size of the lexicon (see Section 3.3).

3.2.2. Rechunking

High quality phonetic transcriptions can only be created by means of a forced alignment for chunks containing *uninterrupted speech* (i.e., speech for which the orthographic transcription provides a sequence of words for which we can predict a sequence of phones corresponding to our acoustic models). In the original version of *ECSD*, only 36.5% of the chunks contained uninterrupted speech. Since we did not want to discard 63.5% of the recordings, we developed a procedure to shorten the chunks automatically, because shorter chunks lead to more speech that can be automatically transcribed. An example is shown in Fig. 1. The speaker of Tier 1 laughs in the middle of his utterance and the second speaker is laughing simultaneously. For laughter acoustic models cannot reliably be trained. Therefore, before rechunking, the complete chunk would be lost, even though effectively only the second half is problematic. After rechunking (Fig. 2), however, the first part of the chunk could be transcribed automatically.

The new chunk boundaries had to be set in positions that are automatically detectable but the resulting chunks need also still be useful for phoneticians and linguists. Therefore, we introduced new chunk boundaries only in silences between words. We aimed at cutting down the length of the chunks to approximately 3 s, which from our own observations is a length at which high quality alignments can be produced with HTK. We first carried out a forced word alignment on the original (long) chunks, which gave us the approximate positions of the word boundaries and the positions of the silences. In addition to the original chunk boundaries, we then put chunk boundaries in the middle of the silences, while leaving the original chunk boundaries intact, and we extracted the orthographic transcription for the new chunks from the ASR-generated word-level transcriptions (Figs. 1 and 2).

The rechunking increased the total duration of chunks with uninterrupted speech by 50.9%, the number of word tokens by 32.3% and the number of word types by 9.2%. Whereas in the original transcriptions only 61.3% of the chunks were shorter than the 3 s suggested for optimal alignment quality, after rechunking, 88.2% of the chunks fulfill this condition. One hundred percent cannot be reached, because speakers sometimes produce longer stretches of speech uninterrupted by silence.

3.3. Building the lexicon

For the forced alignment, we need a lexicon containing the orthographic transcriptions of all word types and their plausible pronunciations. This lexicon was built in three steps. First, a lexicon with the canonical phonemic transcriptions had to be built. In the second step, these canonical transcriptions were used to generate pronunciation variants. In the final step, highly reduced pronunciations for a small number of words were added to the lexicon.

3.3.1. Building a lexicon of canonical transcriptions

The canonical phonemic representations were obtained from the TST-lexicon, which is a Dutch-language lexical database containing 361,163 word tokens. It was compiled by merging lexical resources such as CELEX (Baayen, Piepenbrock, & Gulikers, 1995), RBN (van der Vliet, 2007) and CGN (Oostdijk et al., 2002). This lexicon makes use of a set of 46 phoneme symbols of the speech assessment methods phonetic alphabet (SAMPA) for Dutch (Wells, 1997), which is a machine-readable representation of the IPA symbols. After the orthographic transcriptions were adapted (cf. Section 3.2.1), 8.9% of the word types were still absent in the TST lexicon. The majority of the missing words were compounds, which in Dutch are written as single words. The formation of compounds is highly productive in Dutch and novel compounds abound in spontaneous speech. We manually split up the compounds into their parts, for which subsequently the transcriptions were looked up in the TST-lexicon. If the parts were found, the canonical phonemic transcriptions were concatenated. Then, degemination was applied and stress-marks and syllable-boundaries were hand-checked. For non-compounds that were not present in the

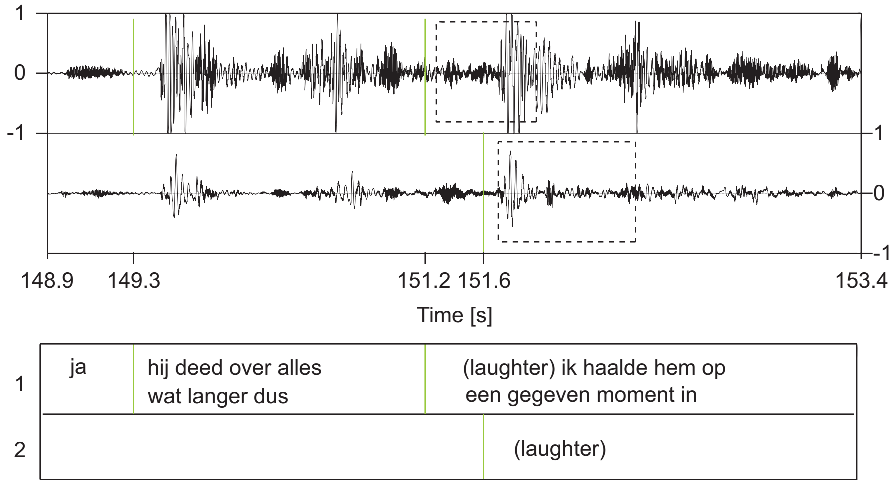

Fig. 2. Example for a chunk of ECSD: After rechunking. The dashed lines mark stretches of laughter.

TST-lexicon, including names and foreign words, e.g., *Tatort*, *PhD-student*, *honeymoon*, *correctness*, *come-back* and *Bond-film*, canonical transcriptions were created manually. For all compounds, including those that were present in the TST-lexicon, secondary stress marks were added by hand.

3.3.2. *Generation of pronunciation variants*

In general, pronunciation variants can either be extracted from a large corpus that has already been transcribed manually at the segmental level (*data-driven approach*, e.g., Hämäläinen, ten Bosch, & Boves, 2007; Kessens, Cucchiarini, & Strik, 2003) or they can be generated by applying a set of rules, proposed in the phonological/phonetic literature, to the canonical forms in the lexicon (*knowledge-based approach*, e.g., Van Bael & Boves et al., 2007). The set of variants derived with a data-driven approach depends on the corpus from which the variants are extracted and tends to contain fewer pronunciation variants for most words than a lexicon created with the knowledge-based approach. Not all plausible variants will be present for all word types, especially for words with a low frequency of occurrence. For highly frequent words, however, the data-driven approach yields a good set of pronunciation variants.

Since substantial knowledge about phonological rules (Booij, 1995) and reduction phenomena (Ernestus, 2000) is available for Dutch, we opted for a knowledge-based approach. We applied a large set of rules to the canonical pronunciations of all words in the lexicon and a small set of additional rules to function words only. Finally, we also incorporated a number of highly reduced pronunciation variants described in the literature (Ernestus, 2000).

Tables 2 and 3 list the rules used for generating the pronunciation variants. There are five phonological rules (cf. Table 2), three co-articulation rules (cf. Table 2), and 22 reduction rules (cf. Tables 2 and 3). Phonological rules that apply within words and that are not discussed here are already integrated in the canonical pronunciations from the TST-lexicon. Only the columns 'Type' and 'Order' are relevant for this section; the other columns will be discussed in Section 4. Some of the rules are well-studied for Dutch and have been used before in the automatic generation of phonemic transcriptions (Kessens et al., 2003; Van Bael & Boves et al., 2007); these are: 'schwa-insertion' (1.0), '[n]-deletion after schwa' (1.1), 'regressive assimilation of voice for obstruents before voiced plosives' (1.2), 'devoicing of plosives following voiceless plosives' (1.3), 'devoicing of fricatives in all word-positions' (1.4), '[t]-deletion in word-final position, preceded by consonant' (4.8) and '[r]-deletion after schwa' (4.5). The other rules were formulated on the basis of the research by Ernestus (2000) on voice assimilation and segment reduction in casual Dutch. In the following we describe the application and the ordering of the rules in more detail, because our set of rules has not been used in this form before.

The column 'Type' in Tables 2 and 3 shows the conditions for the application of the rules. Application of rules marked with a 'C' depends on the segmental context of the target segment; 'P'-rules are position dependent (either position in the word or position in the syllable). The segmental context and the position in the word were the only criteria considered for the generation of pronunciation variants in Kessens et al. (2003) and Van Bael and Boves et al. (2007). Our rules also use the syllabic structures and the stress patterns (Type 'S') of the words. The stress pattern is especially relevant for vowel deletions and lenitions (rules 3.2–3.4, 4.13–4.16, and 4.18–4.19), which have been shown in the literature to affect mostly unstressed syllables (e.g., Rietveld & Koopmans-van Beinum, 1987; van Bergem, 1993). Finally, rules marked with a 'W' were only applied to function words and all verb forms of *hebben* 'to have'. For example, the rule 'deletion of word-initial vowels' (4.17) affects function words only.

We chose to use a tree-structured algorithm, which implies that each reduction rule is applied to the canonical representation of a given word and to all its pronunciation variants that are already generated at the moment the rule is applied. The order in which the rules were applied is shown in Tables 2 and 3 in the column 'Order'. The rule 'schwa-insertion' (rule 1.0) has order '0' because it was applied only to the canonical pronunciation of a word and its output was not used as input for the following reduction rules. Rules that were independent of other rules (marked with 'I' in the column 'Type' in Tables 2 and 3), as for example '[n]-deletion after schwa' (rule 1.1), were generally applied at the beginning. For some rules their relative order of application is relevant. For instance, we applied '[r]-deletion after schwa' (rule 4.5) before we applied rules that substituted vowels by schwas (rules 3.3 and 3.4).

An advantage of the tree-structured algorithm compared to the conventional sequential (feeding–bleeding) application of rules, where the input for a rule is only the output of the previous rule, is that the order of rules is less important and that more pronunciation variants are generated. Inevitably, implausible variants are created as well. We believe that the trade-off between missing

Table 2
Word level segment modification rules and their frequencies. Column 'Nb' contains the rule ID number. Column 'Type': The application of the rule depends on segmental context (C), position within the syllable or word (P), word stress (S), word type (W) and/or is independent of other rules (I). Column 'Order': Order in which the rules were applied. Column 'Abs': Absolute number of word tokens to which the rule could be applied. Column 'Tokens': % of word tokens showing the rule compared to the total number of word tokens which could have shown the rule. Column 'Types': % of word types showing the rule compared to the total number of word types which could have shown the rule.

Nb	Segment modification rules	Type	Order	Abs	Tokens	Types
0	Canonical pronunciation			56,262	59.7	40.0
1	**Phonological rules from the literature**					
1.0	Schwa-insertion	C PI	0	106	21.2	37.8
1.1	[n]-deletion after schwa	C PI	2	7304	76.7	88.8
1.2	Regressive assimilation of voice for obstruents before voiced plosives	C	3	123	41.3	42.1
1.3	Devoicing of plosives following voiceless plosives	C P	28	28	18.8	19.2
1.4	Devoicing of fricatives in all word-positions	PI	6	7504	56.7	69.8
2	**Coarticulation**					
2.1	Voicing of intervocalic obstruents	C	4	1011	22.0	31.2
2.2	Devoicing of obstruents in obstruent clusters	C	5	31	50.0	66.7
3	**Lenitions**					
3.1	Word-initial /b/ pronounced as [m]	PI	26	525	22.6	28.2
3.2	Long vowels produced as short	S	8	859	21.7	39.3
3.3	One vowel produced as schwa	SI	16	3280	38.1	52.7
3.4	Two vowels produced as schwa	SI	17	116	37.5	40.2

Table 3
Word level deletion rules and their frequencies. Column 'Nb' contains the rule ID number. Column 'Type': The application of the rule depends on segmental context (C), position within the syllable or word (P), word stress (S), word type (W) and/or is independent of other rules (I). Column 'Order': Order in which the rules were applied. Column 'Abs': Absolute number of word tokens to which the rule could be applied. Column 'Tokens': % of word tokens showing the rule compared to the total number of word tokens which could have shown the rule. Column 'Types': % of word types showing the rule compared to the total number of word types which could have shown the rule.

Nb	Segment deletion rules	Type	Order	Abs	Tokens	Types
4.1	[n]-deletion between vowels and /s/	C	7	501	45.1	53.7
	Absence of consonants following nasals					
4.2	Deletion of bilabial plosives after /m/	C P	9	60	28.4	38.3
4.3	[k]-deletion after /ŋ/ and [s]-deletion after /n/	C	10	384	38.1	40.3
	[**r**]-**deletion**					
4.4	[r]-deletion after low vowels	C P	11	2590	43.5	54.9
4.5	[r]-deletion after schwa	C P	12	2319	53.1	60.6
	[t]-deletion					
4.6	[t]-deletion between /s/ and consonant	CI	21	231	51.8	60.6
4.7	[t]-deletion between consonant and plosive	C PI	22	29	48.3	63.2
4.8	[t]-deletion in word-final position, preceded by consonant	C P	23	2610	43.2	49.3
4.9	[t]-deletion between vowel and plosive	C	24	21	37.5	41.9
	Word specific deletions					
4.10	Suffix -*lijk* [lək] reduced to [k] or [ək]	C WI	20	537	59.5	70.4
4.11	Absence of /h/ in verb forms of *hebben* ('have') and in *het* ('the/it')	P WI	1	1197	68.1	100
4.12	Absence of word-final [x] in *nog* ('yet') and *toch* ('still')	P WI	27	337	35.7	100
	Vowel and schwa deletion					
4.13	Deletion of short vowels between voiceless obstruents	C S	13	26	9.0	14.4
4.14	Deletion of short vowels between /v/ and /n/	C S	14	18	14.9	36.4
4.15	Deletion of short vowels	S	15	1129	14.6	15.7
4.16	Deletion of long vowels	SI		390	11.8	10.1
4.17	Deletion of word-initial vowels in function words	P WI	26	4347	31.0	56.0
4.18	Deletion of one schwa	S	18	11,579	41.0	55.7
4.19	Deletion of two schwas	S	19	922	21.8	29.7
4.20	Extremely reduced words	WI	Nil	1620	44.2	73.9

relevant variants and generating highly improbable ones would not have been better when using two-level rules (Koskenniemi, 1983). The order of the rules was determined by trying many orders and inspecting the generated variants on plausibility and whether important variants known from the literature (Ernestus, 2000) were present.

After applying all reduction rules, we applied degemination to all generated pronunciation variants, since Dutch does not allow sequences of identical segments. Moreover, we only allowed pronunciation variants that did not contain sequences of more than three consonants, except if one of four consonants was a sonorant. Duplicate variants generated in different branches of the algorithm were removed as well.

Finally, we added extremely reduced forms for 23 word types such as the following:

eigenlijk	/'ɛɪɣənlək/	['ɛɪk]	'actually'
bijvoorbeeld	/bɛɪ'vorbelt/	['vɔlt]	'for example'
natuurlijk	/na'tyrlək/	['tyk]	'naturally'

These extremely reduced forms result from multiple segment and syllable deletions and contain only the stressed vowel plus a few consonants, possibly from other syllables (Ernestus, 2000). They are listed in Appendix A.

The average number of pronunciation variants per word type was 24.1. As a final step we converted the 46 SAMPA symbols to the set of 37 symbols that represent the trained acoustic models. Loan vowels from French and English were mapped to vowels of Dutch such that long lax vowels (/ɛː/, /ʏː/, /ɔː/) were shortened and nasal vowels (/ɛ̃/, /ɑ̃/, /ɔ̃/, /ʏ̃/) were considered as oral. /ɟ/ was converted to the sequence /nj/ and /ʒ/ to /zj/. The CGN corpus (Oostdijk et al., 2002) does not contain sufficient speech data to train acoustic models for these sounds.

Table 4
Material used for the validation of the automatically generated transcriptions and summary of the absolute and relative numbers of deviations to the reference transcription (IFA Corpus).

Speech material (IFA)		**Deviations**	Absolute	Relative (%)
Total speech duration	1867 s	Insertions	528	2.6
Number of utterances	693	Deletions	1369	6.8
Number of phones	20,021	Substitutions	927	4.6
Mean chunk duration	1.29 s	Total operations	2824	14.0

3.4. Validation of the phonemic transcription

3.4.1. Material and procedure

Since manual transcriptions of the *ECSD* that could serve as a reference transcription for the evaluation of the quality of our transcription procedure did not exist in sufficient quantity, we evaluated the quality of our transcription procedure by using part of the spontaneous speech of the IFA Corpus (Son, Binnenpoorte, van den Heuvel, & Pols, 2001), which was produced by seven speakers from both genders. A summary of the material used for the validation is shown in Table 4. The IFA corpus comes with a labeling that was created in two steps. First, an automatically generated transcription was built by means of a forced alignment with a lexicon with canonical transcriptions of the words. In the second step, this alignment was corrected by human transcribers. Therefore, the reference transcriptions may be biased towards canonical forms.

We carried out a forced alignment for the IFA corpus with the same procedure as for the *ECSD*: We used the same speech recognition toolkit, the same acoustic models (i.e., with the same frame shift and window length trained on the same speech material) and the same procedure for generating the pronunciation variants (Section 3.3.2). Then, the hand-corrected reference transcription was compared with our automatically generated transcription using the ADAPT-tool (Elffers, Van Bael, & Strik, 2005). This tool first searches the optimal alignment of the two strings of phones (i.e., reference transcription and automatic transcription) for each utterance separately. Then, the number of phone insertions, deletions and substitutions are calculated for all chunks.

3.4.2. Results

Table 4 shows the difference between the automatically generated transcription and the reference transcription quantified by the number of phone insertions, deletions and substitutions relative to the total number of segments in the IFA corpus. Overall, we observed a 14.0% discrepancy. A comparison of that percentage with values found in the literature shows that our transcription is as reliable as a human transcription: Disagreements between human transcribers may vary between 5.6% and 21.2%, depending on the degree of spontaneity of the speech (Kipp, Wesenick, & Schiel, 1996, 1997). Moreover, the discrepancy is small compared to other discrepancies between human-made and automatically generated transcriptions reported in the literature. For instance, Cucchiarini and Binnenpoorte (2002) report a deviation of 12.5% for read speech and of 24.3% for spontaneous speech. The higher agreement between the reference transcription and our automatic transcriptions can be explained by our set of reduction rules which is tailored to the spontaneous, casual speaking style of our corpus.

It is well-known that for certain sounds it is especially difficult for human transcribers to decide whether they are absent or present. For example, in Kuipers and van Donselaar (1997) three phonetically trained transcribers disagreed in 10% of cases on the presence versus absence of schwa in read Dutch sentences, which is nearly twice as high as the overall disagreement between manual transcriptions of read speech (5.6%, Kipp et al., 1997). For schwas in our transcriptions of the IFA corpus, we observed a 24.1% discrepancy with the reference transcription, which is not much higher than the overall disagreement between human transcriptions of spontaneous speech (21.2%, Kipp et al., 1997). Furthermore, it has been shown that the decision whether consonants are voiced is difficult for human transcribers. Ernestus (2000) reported that three phoneticians disagreed on the voicing of intervocalic plosives in 15% of the cases. Our automatic transcriptions deviated in 8.7% of the cases from the reference on the voicing of plosives. This high degree of agreement is remarkable, if only because obstruent voicing is also cued by phonetic characteristics of the neighboring segments and by the durations of the segments themselves. Monophone HMM models are not capable of encoding this kind of linguistic information.

Since overall the observed discrepancies between the automatically generated and the manual reference transcriptions are in the range of discrepancies between human labelers, we conclude that our transcriptions form a reliable data source for studies on pronunciation variation on the segmental level in spontaneous Dutch.

4. Analysis of phonological, co-articulation and reduction rules

In the analysis of the frequencies of phonological, co-articulation and reduction rules we excluded interjections, disfluencies, response tokens (e.g., *hm, aha*) and broken words, because these words do not have unambiguous canonical representations. This leaves 94,241 word tokens, representing 6839 word types, for analysis.

As mentioned above we distinguished three types of rules: phonological, co-articulation and reduction rules that modify segments and that delete segments. During the generation of the pronunciation variants in the lexicon, the rules that contributed to their creation were logged and the number of word types to which a given rule applied was counted. In the forced alignment, the ASR systems chose the best matching pronunciation variant on the basis of the speech signal. From these chosen variants we computed how often the rules were actually applied in terms of numbers of word tokens (column 'Abs' in Tables 2 and 3), also relative to the total number of word tokens (shown in the column 'Tokens') and types in the corpus (shown in the column 'Types') to which the rule could have been applied. The relative token frequency of a rule shows how important this rule is in the corpus. The relative type frequency of the rule shows whether a rule is specific for a small number of words or rather word-type independent.

The rules formulated in Tables 2 and 3 are not sensitive to the words preceding or following the target word. However, we know that pronunciation variation, especially for the word-initial and final segments, may be induced by segmental context in neighboring words. Therefore, we conducted separate analyses for segments at the word boundaries. For instance, the rule 'Absence of /h/' (rule 4.11, Table 3) was applied to the forms of the verb *hebben* 'have' and to *het* 'the/it' and we investigated in which preceding segmental context this rule applied especially frequently.

Obviously, segmental context also has an impact word-internally. For this reason, the rules that delete consonants were applied only in specific contexts. For instance, rules 4.6–4.9 in Table 3 distinguish between four different contexts for [t] deletion. In contrast, all rules concerning vowel lenitions (3.2–3.4 in Table 2) and deletions (4.13–4.19 Table 3) were applied to all unstressed syllables, irrespectively of segmental context, mainly because not enough knowledge was available to formulate context dependent rules. Obtaining a better understanding of the conditions under that influence reductions and the frequency of their occurrence is of course the second goal of this chapter. We investigated the impact of word-internal segmental context on vowel lenitions and deletions in separate analyses.

Overall, 40.3% of the word tokens in the analyzed speech material were not produced in their canonical form and 60.0% of the word types occur at least once with one of the non-canonical pronunciation variants. More hapax legomena occur in a non-canonical variant (71.2% of the types) than words that occur more often (41.1% of the word types). An explanation for the different behavior of the hapaxes is that most hapaxes are long compounds for which extremely high numbers of pronunciation variants have been generated (three times as many as the average), so that the probability that one of these variants is chosen in the alignment is very high.

Our results seem to differ from the observations by Johnson (2004), who reported for a corpus of conversational American English that more than 60% of the word tokens deviated at least in one segment from their citation forms. That his number of deviations was larger than ours can partly be explained by the larger set of phonetic symbols that he used to transcribe the speech material (59 symbols of the ARPABET versus 37 acoustic phone models used to transcribe *ECSD*).

In the following subsections, we discuss the results obtained for all rules, which we present in the order in which they appear in Tables 2 and 3. For each rule, we first give information from the literature, show an example, and then we discuss the quantitative results from our study, comparing them with quantitative results from other studies, if available.

4.1. Results and discussion: segment modification rules

4.1.1. Phonological rules from the literature

Rules 1.0–1.4 in Table 2 show rules that have been described in the literature on the phonology of Dutch (e.g., Booij, 1995). Schwa may be inserted in word-final consonant clusters consisting of a liquid and a final consonant other than /n/, /t/ or /d/ (rule 1.0), such that for instance *melk* 'milk' /'mɛlk/ may be pronounced as ['mɛlək]. In absolute numbers, we observed schwa-insertion 106 times, that is in 21.2% of the word tokens in which schwa-insertion could have occurred. Swerts, Kloots, Gillis, and De Schutter (2001) reported that schwa-insertion occurs more frequently (28.1% of the tokens) in isolated words carefully spoken by teachers of the Dutch language of the same regional background as the speakers of our material. Schwa-insertion therefore appears to be slightly less pervasive in connected spontaneous speech than in words spoken in isolation.

The deletion of [n] after schwa in word and syllable final position (rule 1.1) has been described as obligatory, in particular for speakers of the western part of the Netherlands (Booij, 1995), which is the regional background of the speakers of the *ECSD*. For example, the word *lopen* 'to walk' /'lopən/ would be pronounced as ['lopə]. We observed [n]-deletion in 76.7% of the word tokens and 88.8% of the word types. We therefore would not consider this rule as obligatory, but compared to all other rules investigated in this study, it is the most frequent one. Previous studies on [n]-deletion in Dutch have reported that [n] was absent in approximately 40% of the word tokens (Kessens et al., 2000; Wester et al., 1998). An explanation of this latter much lower frequency is that these studies were based on a corpus of careful speech (Strik, Russel, van den Heuvel, Cucchiarini, & Boves, 1997). Another explanation could be differences in the regional background of the speakers.

The phonological literature states that regressive voice assimilation (rule 1.2) is obligatory within prosodic words and compounds in Dutch (Booij, 1995). For example, the word *voetbal* 'football' /'vutbɑl/ would be pronounced as /'vudbɑl/. Since this rule was already incorporated in the canonical pronunciations of the TST lexicon, it could only be applied to those compounds that we added to the lexicon ourselves. Those compounds are of low frequencies and specific for the topics of the conversations in the *ECSD*. We found that only 42.1% (123 tokens) showed regressive voice assimilation. This percentage is similar to the one reported by Ernestus, Lahey, Verhees, and Baayen (2006) for read speech (43%). Apparently, within-word regressive voice assimilation is as frequent in casual speech in compounds of low frequencies as in more formal speech styles, and this rule appears to be optional, rather than obligatory.

According to the phonological literature, word-internal progressive assimilation of voice has been stated to be limited to fricatives (as for instance in *opvallend* 'notable' where /pv/ is produced as [pf]) (Booij, 1995). Our data show that progressive voice assimilation (rule 1.3) also occurs in plosive clusters, namely in 18.8% of the tokens where this rule could apply. For instance, the word *postbank* 'postbank' with the canonical pronunciation /'pɔstbɑŋk/ was produced as ['pɔstpɑŋk]. This percentage is slightly higher than the one reported by Ernestus et al. (2006) for Dutch word-internal plosive clusters in read speech (11%).

Previous studies have shown that fricatives are often devoiced (rule 1.4) in Dutch spoken in the Netherlands (e.g., Van de Velde, Gerritsen, & van Hout, 1996; Van den Broeke & van Heuven, 1979). Our data support these studies. For instance, word *zwemmen* 'to swim' /zwɛmən/ was produced as ['swɛmə]. We saw that voiced fricatives not preceded by an obstruent were produced as voiceless in 56.7% of the word tokens. Table 5 shows the frequencies of devoicing separately for the three different fricatives. It shows that /ɣ/ is more often devoiced than /v/, and that /v/ and /z/ are devoiced equally frequently. These findings are in line with those reported by Van de Velde and van Hout (2001) for a corpus of read speech produced by teachers of Dutch. They observed that /ɣ/ was devoiced in 50% and both /v/ and /z/ in 35% of the tokens by speakers from the same region as the speakers of *ECSD*. A possible reason for why our frequencies are higher is that our classification of voice is binary whereas the transcribers in the study by Van de Velde and van Hout (2001) had the choice between voiced, partially voiced and voiceless.

4.1.2. Co-articulation

Rules 2.1 and 2.2 in Table 2 concern co-articulation of voicing within words. Intervocalic obstruents may be voiced (rule 2.1), such that a word like *lopen* 'to walk' /'lopən/ may be produced as ['lobə]. We observed that intervocalic obstruents were voiced 1011 times, that is in 22.0% of the word tokens with an invervocalic obstruent. The plosives /p/, /t/ and /k/ were slightly more often voiced (22.4%) than the fricatives /s/, /x/ and /f/ (18.0%). Obstruents were devoiced in obstruent clusters (rule 2.2) in only 31 tokens, but these tokens represent 50.0% of the possible tokens and 66.7% of the possible word types. For instance, the word *budget* 'budget' with the canonical pronunciation /bʏd'zjɛt/ was produced as [bʏt'ʃɛt]. The plosives /p/ and /t/ were much less often devoiced (39.1%) than the fricatives /s/, /x/ and /f/ (83.3%).

4.1.3. Lenitions

Rules 3.1 to 3.4 in Table 2 refer to the realization of /b/ as [m] (3.1) and to vowel lenition (3.2–3.4). Our data show that word-initial /b/ is produced as [m] in 22.5% (525) of the word tokens and in 28.2% of the word types starting with /b/. An analysis of the affected word types showed that 49.0% of the tokens represent only four highly frequent types (see Table 6). Furthermore, we saw that, including these four word types, this rule (3.1) is more frequently applied for tokens following a vowel.

In Dutch, phonologically short and long vowels are not only different in their duration but also in their quality. Long vowels are typically realized as tense and short vowels as lax. Nooteboom (1979) observed that in casual speech tense vowels sometimes sound similar to their lax counterparts, such that a word like *bijvoorbeeld* 'for example' /bɛɪ'vorb**e**lt/ sounds like [bɛɪ'vob**ɪ**lt]. Our data show that this is the case in unstressed syllables in 21.7% of the

Table 5
Devoicing of fricatives in all word-positions, excluding fricatives preceded by an obstruent. Column 'Tokens': % shows the proportion of the relevant word tokens in which the rule applied. Column 'Types': % shows the proportion of word types to which the rule applied at least once.

Fricative	Tokens	Types
Total	56.7	69.8
/ɣ/	71.2	76.0
/v/	54.4	50.9
/z/	54.8	51.2

Table 6
Word-initial /b/ pronounced as [m]. Without 1-4: Data set without the four word types *bedoel, bij, ben* and *beetje*. Column '%': % of tokens of word-initial /b/s pronounced as [m] for all segmental contexts, and when preceded by a vowel respectively ('after vowel').

Word type	**All contexts**		**After vowel**	
	Tokens	%	Tokens	%
Total	2326	25.5	581	31.3
Without 1–4	1186	21.2	305	18.7
1 *bedoel* 'mean'	143	16.3	0	0
2 *bij* 'at, with'	425	22.8	108	18.5
3 *ben* 'am'	317	32.2	38	42.1
4 *beetje* 'a bit'	255	49.4	128	68.8

relevant word tokens. Furthermore, we observed that low long vowels are produced as short vowels more than twice as often as high long vowels (see Table 7).

Ernestus (2000) observed that all types of vowels can be reduced to schwa in unstressed syllables (rules 3.2–3.4). For instance, the word *contract* 'contract' /kɔn'trɑkt/ may be pronounced as [kən'trɑkt]. Our data show that in 38.1% of the word tokens that contain at least one short or long unstressed vowel in the canonical pronunciation, a vowel was reduced to schwa. Reduction of more than one vowel to schwa in one word token (rule 3.4) is less frequent: this occurred in only 116 tokens. This low number can be explained by the low number of words that have more than one unstressed full vowel (329 tokens, 254 types). Table 7 shows how often different types of vowels are reduced to schwa. Overall, the quantitative findings on vowel lenition support the impressionistic observations by Ernestus (2000) based on the same speech corpus that we are using.

Table 8 shows how often full vowels were realized as schwa in the different preceding and following segmental contexts, distinguishing between no neighboring segment within the word and consonants of different manner and place of articulation. We distinguished only four places of articulation, since we merged dental and alveolar consonants and velar and uvular consonants. Vowels were most frequently realized as schwa when preceded by fricatives (4.3%) and liquids (4.8%) and when followed by a full vowel (12.0%) or a glide (7.1%). With regard to place of articulation, vowels were least frequently realized as schwa when preceded by a consonant of bilabial place of articulation (1.7%). Additional analyses will be presented in our discussion of the contextual influences on vowel deletions (Section 4.2.5).

Table 7
Vowels and their realizations. Column 'Total': Total number of vowel tokens in the corpus. Column 'Short': % of tokens that were produced as short vowels. Column 'Schwa': % of tokens that were produced as schwa. Column 'Absent': % of tokens that were absent.

Type of vowel	Total	Short	Schwa	Absent (%)
Low long vowel	21,650	3.5%	4.5%	6.6
Low short vowel	22,009	90.8%	3.1%	6.1
High long vowel	19,178	1.0%	3.2%	3.8
High short vowel	30,377	87.1%	1.6%	11.3
Diphthong	7504	Nil	Nil	0.8
Schwa	30,519	Nil	55.3%	44.7

4.2. Results and discussion: segment deletion rules

4.2.1. Absence of [n]

Ernestus (2000) reported that word-medial /n/ after full vowels and before consonants may be absent in casual Dutch. In our analysis, we saw that /n/ was absent in nearly half of the word types and word tokens where it followed a vowel and preceded /s/ (rule 4.1). For example, *mensen* 'people' was pronounced as ['mɛsə] instead of the canonical form /'mɛnsən/. It is quite possible that where in the segmental transcriptions an [n] is absent, the preceding vowel was nasalized, so that remnants of the nasal segment remained present (e.g., Ernestus, 2000). The numbers given here for the 'Absence of [n]' therefore rather reflect the absence of the nasal closure than the absence of the feature nasality.

4.2.2. Absence of consonants following nasals

Rule 4.2 in Table 3 deletes bilabial plosives after /m/. Since words with this segment-sequence are rather rare in the corpus, this reduction rule only affected 60 tokens representing 44 word types, which is one third of the word tokens and one third of the word types with /mp/ and /mb/. For example, the word *olympische* 'olympic', for which the canonical pronunciation is /o'lɪmpisə/, was produced as [o'lɪmpisə].

Ernestus (2000) reported the absence of /k/ after /ŋ/ and /s/ after /n/ (rule 4.3). Our results confirm this observation. Overall, 38.1% of the instances of /k/ and /s/ following /ŋ/ and /n/, respectively, were absent. For example, the word *denk* 'think' /'dɛŋk/, which occurs 229 times, was pronounced as ['dɛŋ] 134 times. Another example is *volgens* /'volɣəns/, which was produced without /s/ in 49 of the 98 tokens. The deletion of [s] has been reported before for conversational Dutch (Van Bael & Baayen et al., 2007), however without information about the context in which [s]-deletion occurs. Deletion of /d/ and /t/ after nasals is captured by the [t]-deletion rules (4.7 and 4.8 in Table 3) and discussed together with [t]-deletions in other segmental contexts later in this section.

4.2.3. Absence of [r]

A well studied segmental reduction in Dutch is the absence of [r] after vowels (rule 4.4 and 4.5 in Table 3), such that a word like *anders* 'different' with the canonical form /'ɑndərs/ may be pronounced as /'ɑndəs/. The absence of [r] has been reported both for carefully produced speech (van den Heuvel & Cucchiarini, 2001;

Table 8
Absolute and relative numbers of full vowels and schwas realized as schwa (% Schwa) or being absent (% Absent) in the different preceding and following contexts. Ons = at the onset of a word. End = at the end of a word. Place of articulation: Bil= bilabial, LaDe= labiodental, De/Al= dental and alveolar, Ve/Uv= velar and uvular.

			Manner of articulation					Place of articulation			
			Plosive	Fricative	Nasal	Glide	Liquid	Bil	LaDe	De/Al	Ve/Uv
Preceding context	Ons	Vowel									
Total # full vowels	22,873	104	22,001	19,613	11,089	10,496	7034	28,798	11,625	29,432	20,231
% Schwa	1.4	1.9	3.2	4.3	2.9	1.7	4.8	1.2	2.9	3.3	3.5
% Absent	20.1	5.8	2.1	2.9	6.3	1.5	7.0	1.7	2.3	3.5	14.3
Total # schwa	2383	2	11,620	7481	1981	3565	3487	2578	2419	13,181	9950
% Absent	17.8	100	39.5	50.2	45.8	59.8	52.2	43.8	42.5	41.9	55.5
Following context	End	Vowel									
Total # full vowels	11,346	50	26,960	12,915	21,597	1212	19,134	6331	3576	46,534	25,377
% Schwa	2.0	12.0	1.6	3.4	3.6	7.1	3.4	4.1	3.5	3.0	2.8
% Absent	9.5	10.0	1.2	5.2	8.0	19.8	8.5	10.0	10.7	6.2	11.4
Total # schwa	8952	53	2240	2691	10,616	640	5327	826	929	13,194	6559
% Absent	47.9	60.4	46.9	44.4	36.9	75.2	52.1	47.2	67.9	42.1	41.6

Wester et al., 1998) and spontaneous speech (Ernestus, 2000; Van Bael & Baayen et al., 2007). Our study showed that [r] was absent after low vowels in 43.5% and after schwa even in 53.1% of the word tokens. The 54.8% of all word types with a postvocalic /r/ occurred at least once without [r]. van den Heuvel and Cucchiarini (2001) reported similar frequencies for [r] deletion after schwa in Dutch (56%) as we found. Their study was based on a corpus of spontaneous human-machine interactions (Strik et al., 1997). Lower deletion rates have been reported for careful speech: 29% deletion for tokens where [r] preceded a consonant and followed schwa, long vowels or unstressed short vowels (Wester et al., 1998).

4.2.4. Absence of [t]

The absence of [t] is well-documented for Germanic languages (e.g., Goeman, 1999; Losiewicz, 1992; Mitterer & Ernestus, 2006). However, quantitative studies are limited to English (Dilley & Pitt, 2007; Jurafsky, Bell, Gregory, & Raymond, 2001) and German (Kohler, 2001). We analyzed the absence of [t] in four different contexts (rules 4.6–4.9 in Table 3). The number of word tokens (260) where /t/ occurs in the middle of word–medial consonant clusters (rules 4.6 and 4.7) is small and these tokens mainly result from compounding. The [t] was absent in nearly half of these tokens, such that a word like *standaardprijs* 'standard price' was pronounced as ['stɑndər'prɛɪs]. Between vowels and plosives, as for example in *voetbal* 'soccer', and in word-final position after a consonant, as for example in *gezicht* 'face' /xəzɪxt/, the [t] was absent in one third of the word tokens. Wester et al. (1998) found that only 19% of the [t]s in consonant clusters and at the end of word–final consonant clusters were absent in a corpus of carefully produced speech. Possibly, as for the absence of [r], speech register affects the frequency of this reduction rule.

In 474 of the 2610 tokens of word-final [t]-deletion the following word started with a /t/ or /d/, and therefore these cases represent cross-word degemination (Booij, 1995). If we exclude these degemination cases, 2.3% of all word tokens in the corpus were affected by word-final [t]-deletion.

Overall, 11.9% of all /t/s in all contexts and word positions were transcribed as being absent. This frequency is as high as the frequencies reported by Van Bael and Baayen et al. (2007). They reported that 11.5% of all [t]'s were absent in a corpus of spontaneous telephone conversations.

4.2.5. Absence of full vowels

The literature suggests that vowel deletion mainly affects unstressed short vowels and schwas preceded by a syllable onset (e.g., Ernestus, 2000; Rietveld & Koopmans-van Beinum, 1987; van Bergem, 1993). Our data set shows that deletion of unstressed vowels is indeed very frequent in conversational Dutch (rules 4.13–4.20 in Table 3). Unstressed short vowels were absent in approximately 15% of the word tokens (rules 4.13–4.15), while unstressed long vowels were absent in 11.8% of the word tokens for the absence of schwa. Furthermore, we observed that in 31.0% of the tokens of vowel–initial function words, this word–initial vowel was absent. What is more, the data showed that in polysyllabic function words also word–initial vowels that carry word stress may be absent, such that for example *enkel* 'only' /'ɛŋ-kəl/ is pronounced as ['ŋkəl].

Table 7 provides an overview of how many tokens of the different types of vowels were absent. Of the full vowels, high short vowels were absent most frequently (11.3%) and diphthongs least frequently (0.8%). Overall, these vowel deletion rates are higher than previously reported for a corpus of Dutch telephone dialogues by Van Bael and Baayen et al. (2007). The deletion rate for diphthongs, however, is in the same range.

Table 8 shows how often vowels were absent in the different preceding and following segmental contexts. With regard to the manner of articulation of the neighboring consonants, vowels were least often absent after glides (1.5%), plosives (2.1%) and fricatives (2.9%) and before plosives (1.2%) and fricatives (5.3%), while they were most frequently absent before glides (19.8%). In word-onset position, vowels were deleted in 20.1% of the tokens. These tokens were part of function words or of words for which extremely reduced variants were incorporated in the pronunciation lexicon. With regard to the place of articulation of the surrounding consonants, vowels were absent most frequently before a velar or uvular (14.3%) and least frequently after a dental or alveolar (6.2%).

Table 8 allows us to compare the contexts that are prone to vowel lenition, discussed in Section 4.1.3, and vowel deletion. If vowel deletion were the natural end point of gradual vowel lenition, the contexts in which vowels are absent least and most frequently would be the same as the contexts in which vowels are realized as schwa least and most frequently. Our data do not support this hypothesis. For instance, whereas vowels are most often absent after velar and uvular consonants, vowels are not especially often realized as schwa in this context compared to how often they are realized as schwa in other preceding contexts.

4.2.6. Absence of schwa

Schwas were absent in 11,579 tokens, that is, in 12.3% of all word tokens in the corpus. A single schwa was absent in 41.0% of the word tokens containing at least one unstressed schwa. Two schwas were absent in 21.8% of the word tokens containing two or more schwas. For instance, *bodemverontreiniging* /'bo.dəm.vər.'ɔnt.rɛɪ.nə.ɣɪŋ/ 'ground pollution' was produced as ['bodm.və.'ɔnt.rɛɪn.ɣɪŋ]. Overall 44.7% of the schwas were absent (see Table 7). These schwa deletion rates are similar to what Dalby (1986) reported for extremely fast spoken American English (overall 43% schwa deletion) and to what Wester et al. (1998) reported for a corpus of carefully spoken Dutch (also 43%). This leads us to conclude that the absence of schwa is less dependent on speech style than the previously discussed absence of [t] and [r].

As previously described for vowel deletions, we carried out additional analyses on the frequency of schwa deletion in the different preceding and following segmental contexts (see Table 8). With regard to manner of articulation, schwas are most frequently absent after (59.8%) and before (75.2%) glides. They are absent least often after plosives (39.5%) and before nasals (36.9%). With regard to place of articulation of the neighboring consonants, schwas tend to be more often absent after velars and uvulars (55.5%) and before labio-dentals (67.9%).

Table 8 also allows us to compare whether full vowels and schwas tend to be absent in the same segmental contexts, since the full vowel and schwa deletion rules were identical (with the exception of rule 4.17, which only applied to full vowels at the beginning of function words). In general, full vowels and schwas are deleted in the same contexts. For example, both full vowels and schwas are most often absent after velar and uvular consonants. However, there are also differences. Whereas vowel deletion is least frequent after glides, schwa deletion is most frequent in this context.

4.2.7. Absence of syllabic nuclei

The literature on reduction frequently reports deletion rates for syllables, where syllable deletion is defined as reduction in the number of syllabic nuclei of a word (Johnson, 2004; Van Bael & Baayen et al., 2007). Since in standard Dutch no sounds other than vowels can be syllable bearing,[1] all rules that lead to vowel deletion (rules 4.10 and 4.14–4.20) also lead to syllable deletion. Note that syllable deletion is not the same as the deletion of all segments of one syllable. For example *natuurlijk* 'naturally' has three syllables in its citation form /na.'tyr.lək/ and both pronunciations ['tyk] and ['ntyk] count as variants with two deleted syllables, even though in the second case the [n] remained of the first syllable [na].

Our data show that 19.0% of all word tokens underwent syllable deletion. 87.9% of these had only one syllable deleted. A high percentage (44.6%) of these single syllable deletions occur in monosyllabic function words, such that the function word *het* /'hɛt/ 'the/it' was pronounced as [t] or the conjunction *en* /'ɛn/ 'and' as [n]. Excluding such monosyllabic words, 9.1% of all syllabic nuclei (131,298) were absent. These frequencies are higher than those reported by Johnson (2004) for American English (6% for content words and 4.5% for function words) and by Van Bael and Baayen et al. (2007) for spontaneous Dutch (5.5% overall syllable deletion).

4.2.8. *Word-specific reductions*

Rule 4.10 reduces the suffix *-lijk* /lək/ to [ək] or [k], as for example in *hopelijk* 'hopefully'. This rule was applied in more than half of the tokens, which may be surprising since Pluymaekers, Ernestus, and Baayen (2005) perceived an /l/ in nearly 90% of tokens in the spontaneous speech part of the CGN (Oostdijk et al., 2002). One explanation could be that the preference of the ASR system for pronunciation variants without [l] may be related to the short duration of the affix as a whole. Also, [ə] and [l] may be perceptually encoded in the same (short) stretch of the signal.

Word-initial /h/ (rule 4.11) of the verb *hebben* ('have') and the function word *het* 'the/it' was absent in 68.1% of the tokens. In one third of the deletion cases, the previous word ended in a plosive. In such a context it is likely that the release friction of the plosive and the [h]-friction are coarticulated. For carefully spoken Dutch, [h]-deletion has only been reported for *het*. Our data showed that in casual Dutch also in forms of *hebben* the absence of [h] is frequent (64.3%).

Ernestus (2000) was the first to document the absence of [x] (rule 4.12) in the function words *nog* /'nɔx/ 'still' and *toch* /'tɔx/ 'yet' in casual Dutch. Our observations support her findings: [x] was absent in 35.7% of the tokens in the speech material.

4.2.9. *Extremely reduced words*

For 23 word types in our corpus we had added extremely reduced pronunciation variants based on the observations by Ernestus (2000). These extremely reduced pronunciations are the result of several segmental deletions applying simultaneously (some of these deletions were not incorporated in our set of rules, because they apply only in a very limited number of word types). Appendix A provides a list with all these word types, their canonical and extremely reduced pronunciations and the frequencies with which the variants occur in the corpus. The results for all words together are shown as rule 4.20 in Table 3. For 17 of these word types we found that the extremely reduced form was produced by the speakers. Overall, 44.2% of the tokens were produced as the extremely reduced variant. For some word types, the frequency was especially high. For instance, the word *gewoon* /xə'won/ 'simply' was produced as ['xon] in more than half of the tokens. Furthermore, some of the words were never produced with their canonical pronunciation (*natuurlijk* 'of course', *mogelijk* 'possible' and *bijvoorbeeld* 'for example').

5. General discussion and conclusions

This paper described the automatic generation of segmental transcriptions for a corpus of spontaneous Dutch and presented a quantitative analysis of phonological, co-articulation, lenition and reduction rules on the basis of these transcriptions.

In the first part, we showed how we automatically created a segmental transcription for Ernestus Corpus of Spontaneous Dutch. One important step in preparing the orthographic transcription was the automatic rechunking of the acoustic signal and the corresponding orthographic transcription, which increased the amount of speech that could be transcribed automatically by 50.9%. Subsequently, the acoustic signal was automatically transcribed by means of a forced alignment procedure, which had as its input the acoustic signal, orthographic transcriptions, a lexicon containing on average 24.1 pronunciation variants for each word in the corpus, and acoustic phone models. The ASR system chose the most probable pronunciation variant for each word token. The acoustic phone models were trained at a frame shift of 5 ms instead of the default of 10 ms (e.g., Adda-Decker et al., 2005; Schuppler et al., 2009; Van Bael & Boves et al., 2007) in order to improve the annotation of very short segments, which are especially frequent in spontaneous speech.

The lexicon with pronunciation variants had one major improvement over previous lexica used for the automatic creation of phonetic transcriptions for Dutch (Cucchiarini & Binnenpoorte, 2002; Van Bael & Boves et al., 2007): Our lexicon contained pronunciation variants generated by, among others, vowel deletion rules that referred to the stress patterns and syllable structures of the words (3.2–3.4 in Table 2 and 4.13–4.19 Table 3). The incorporation of vowel deletion rules without constraining them on the basis of the word's prosodic characteristics would generate a large amount of implausible pronunciation variants. The fact that these rules were very applied frequently (cf. Table 7) shows the need for incorporating vowel deletion rules. In addition to the set of rules that was applied to all words, we constrained certain rules to a limited number of word types only, and we added extremely reduced pronunciation variants for some words that have been found in a previous study on casual Dutch (Ernestus, 2000).

With this procedure we obtained high quality segmental transcriptions. We validated the automatically generated transcriptions against the manually transcribed spontaneous speech of the IFA corpus. Overall, we observed a disagreement for 14.1% of the segment labels in the reference transcription, which includes segment deletions, insertions and substitutions. This compares favorably to disagreements between independently working human transcribers for the same speech style (e.g., Kipp et al., 1997). Also, the types of disagreement were similar to those between human transcribers for the presence/absence of /ə/ and presence/absence of voicing. In interpreting the (dis)agreement scores it must be kept in mind that the manual and the automatic transcriptions used only 39 phoneme-like labels.

In the second part of the paper, we presented the results of the analyses of phonological, co-articulation, lenition and deletion rules on the basis of the created transcriptions. Since only chunks of uninterrupted speech could be transcribed automatically (88.2% of the chunks), our results can probably only be generalized to other stretches of uninterrupted speech. For the analysis, we compared the pronunciation variants with which the word tokens were transcribed with their canonical transcriptions. A token was considered as reduced if the pronunciation variant selected contains either a lower number of segments (i.e., the absence of segments) or a phone corresponding with less articulatory effort (e.g., lenitions). Contrary to human transcribers, our automatic transcriber can only select pronunciation variants that are present in the pre-defined lexicon. Consequently, we could only derive

[1] There is one region in the East of the Netherlands where sonorants can function as syllabic nuclei. None of the speakers in our corpus originate from this region or has lived there.

frequencies of those segmental reductions that were captured by the rules for the generation of the pronunciation variants. As a consequence, we could not gain knowledge about reductions that we did not envision from the start of the research. However, the high agreement between the manual transcription of the IFA corpus and our automatic transcriptions suggest that only few pronunciation variants were missing in our lexicon. The second limitation of our transcription procedure is that the automatic segmentation always yielded segments that are at least 15 ms long, because of the 5 ms frame shift and the three-state HMMs used in the aligner. As a consequence, for short word tokens, as for instance those produced at high articulation rates, reduced pronunciation variants obtain higher likelihood scores than the canonical variants, even if all segments of the citation forms may be perceived by a native speaker.

Our analysis showed that reduction is highly frequent in casual Dutch and overall more frequent than may have been expected from earlier research on less spontaneous speech material (e.g., van den Heuvel & Cucchiarini, 2001; Wester et al., 1998). In total, only 59.7% of the word tokens were produced canonically while 9.1% of all syllabic nuclei were absent and 19.0% of all word tokens (mono and polysyllabic) were realized with fewer syllabic nuclei than their citation forms suggest. This rate of syllable deletion is higher than previously reported for spontaneous speech: Van Bael and Baayen et al. (2007) reported that 5.46% of all syllables were deleted in a corpus of spontaneous Dutch and Johnson (2004) reported that 7.6% of content words and 5.0% of function words were realized with fewer syllables than their citation forms. As in the present study, these studies considered a syllable as absent if the nucleus was deleted, even if parts of the onset or coda were still present. One reason why syllable deletion is more frequent in our Dutch material than reported by Johnson (2004) for English is that in English sonorants also can be syllable bearing (Johnson, 2004 considered nasals, laterals and rhotics as syllabic nuclei), whereas in standard Dutch only vowels can function as syllabic nuclei.

Vowel lenition and deletion rules were applied to all unstressed syllables, irrespective of segmental context (rules 3.2–3.4 and 4.13–4.19). We carried out additional analyses on how frequent these rules occurred in the different segmental contexts (cf. Table 8) and found that all rules applied very often before glides. Importantly, vowel lenition and deletion differ in their frequencies for the other segmental contexts. This suggests that the absence of vowels is not only the result of gradual lenition but may also result from categorical deletion.

Another focus of our analysis was on voice assimilation. The data showed that assimilation is not obligatory in casual speech (approximately 30%), despite the suggestions made in the phonological literature (e.g., Booij, 1995). Furthermore, we showed that progressive voice assimilation is not limited to fricatives and /d/-initial function words as previously assumed (e.g., Booij, 1995), but that it also occurs in plosive clusters within content words (18.8%, in line with the observations by Ernestus et al., 2006). Similarly frequent is the voicing of intervocalic obstruents (22.0%). Finally, more than half of the phonologically voiced fricatives (i.e., /z,v,ɣ/) were realized voiceless (i.e., [s, f, x]). Due to this high degree of variation in the feature voice, the question arises whether voicing is a reliable cue for listeners at all. Ernestus and Mak (2004) showed with an auditory lexical decision experiment that Dutch listeners rely less upon voice than upon manner and place of articulation for fricative-initial words. They ascribed this result to the fact that there are many more rules affecting voice than manner or place of articulation in Dutch. Our findings support their interpretation of the outcomes of their experiments. Moreover, our data show that speakers do not consistently apply these voicing rules. As a consequence, they inconsistently realize phonologically voiced fricatives as voiceless and vice versa. This example nicely demonstrates how psycholinguistic experiments in controlled conditions and corpus studies based on spontaneous speech can mutually support each other.

Our quantitative analysis of phonological rules showed that '[n]-deletion after schwa' and 'devoicing of fricative' affect a high absolute number of word tokens in the corpus (7304 and 7504 respectively), partly because words ending in /ən/ and containing voiced fricatives are very frequent in Dutch. As a consequence, these rules are especially interesting for pronunciation variation modeling for ASR systems. However, to our knowledge, only '[n]-deletion after schwa' has so far been applied for these purposes (e.g., Hoste, Daelemans, & Gillis, 2004; Van Bael, 2007).

By comparing our results with the results from studies based on carefully produced Dutch, we can draw conclusions about which rules are typical for certain speech styles. We found that the phonological rules 'devoicing of fricatives' as well as 'schwa-deletion' are rather speech style independent, whereas [t] and [r]-deletions appear much more frequent in our conversational corpus than in corpora of carefully produced Dutch. Automatic speech recognition systems could profit from information about the style of speech they are applied to. For example, the lexicon could be adapted to the given speech style such that only those rules are applied to generate pronunciation variants that have high probabilities given the speech style, which, as mentioned above, reduces the size of the lexicon and thus internal confusability.

One limitation of our analyses is that it is based on segmental transcriptions. As a consequence we can only capture the coarse picture of how much and what kind of reduction and pronunciation variation can be found in casual speech. For instance, Schuppler, van Dommelen, Koreman, and Ernestus (2009) showed that whereas 74.6% of word-final /t/ were auditorily present in a subset of this corpus of spontaneous Dutch, only 11.2% were produced with all canonical sub-segmental cues for [t]. Moreover, information about gestures spreading beyond segment boundaries cannot be integrated in segmental transcriptions. For example, the word *mensen* 'people', with the canonical pronunciation /'mɛnsən/, was transcribed as ['mɛsən], which counts as [n]-deletion in our analysis, although it is quite possible that the first vowel was nasalized, so that remnants of the nasal segment remained. Research on the sub-segmental level will provide more detailed information about how speakers produce speech and the type of speech listeners have to cope with.

This paper shows how work in ASR and phonetics can benefit from each other. On the one hand, our analysis of reductions is based on a speech corpus that could be automatically transcribed thanks to the availability of an ASR system. At the same time the ASR system will profit from incorporating the statistics about pronunciation variants that can be derived from very large corpora. The results of our investigation are also relevant for psycholinguistics. They provide information about the type of speech listeners and speakers are processing in everyday life. Psycholinguistic models will have to take this information into account. For instance, they have to consider that voicing is a highly variable property of obstruents in Dutch, and consequently not a very reliable cue for word recognition. Furthermore, our analyses show which are the frequent and interesting pronunciation variants and which deserve further detailed phonetic investigations (see e.g., Schuppler et al., 2009). Finally, the results provide information on how to tune ASR systems to the type of speech and the speaker. As mentioned above, inclusion of variants in the lexicon can only improve ASR systems if the conditions are specified under which specific reductions are likely to occur (e.g., speech style, phonetic context, frequency of the words, word class). The study of casual speech thus is necessarily interdisciplinary in nature.

Acknowledgements

This research was supported by the Marie Curie Research Training Network 'Sound to Sense'. Mirjam Ernestus was supported by a EURYI-award from the European Science Foundation and Odette Scharenborg by a Veni-grant from the Netherlands Organisation for Scientific Research (NWO). We would like to thank Annika Hämäläinen for her help with the acoustic models and Eric Sanders for sharing his experience with transcribing large spoken language corpora. We would also like to thank the anonymous reviewers for their constructive criticism and suggestions on earlier versions of the paper.

Appendix A. List of extremely reduced word types

Table 9 presents the list of extremely reduced forms result from multiple segment and syllable deletions and contain only the stressed vowel plus a few consonants, possibly from other syllables.

Table 9
Extremely reduced word types in the Ernestus Corpus of Spontaneous Dutch. Column 'Total': Total number of occurrence of the considered word type, produced by the 20 speakers. Column 'Canonical': canonical pronunciation of the word and the frequency of this variant relative to all generated pronunciation variants (% all PV). Column 'Extreme': extremely reduced pronunciation of the word and the frequency of this variant relative to all generated pronunciation variants.

Word type	Total	Canonical	% all PV	Extreme	% all PV
allemaal 'all of them'	166	'ɑləmal	2.4	'aməl	62.7
als 'if'	632	'ɑls	5.9	'ɑs	63.3
anders 'otherwise'	64	'ɑndərs	6.25	'ɑs	0.0
bepaalde 'certain'	30	bə'paldə	20.0	palə	36.7
bijvoorbeeld 'for example'	45	bɛɪ'vorbelt	0.0	'vɔlt	46.7
computer 'computer'	7	kɔm'pjutər	14.3	'pjutər	14.3
constant 'constant'	2	kən'stɑnt	100	kən'sən	0.0
eigenlijk 'actually'	237	'ɛɪɣələk	4.6	'ɛɪk	13.1
gaan'go'	220	'xan	59.5	'xə	40.5
gewoon 'simply'	415	xə'won	10.1	'xon	70.1
helemaal 'completely'	170	helə'mal	4.1	heməl	20.0
maandag 'Monday'	7	'mandɑx	57.1	'manz	0.0
mogelijk 'possible'	16	'moɣələk	0.0	'mok	6.3
natuurlijk 'of course'	331	na'tyrlək	0.0	'tyk	33.2
niet 'not'	1230	nit	54.1	ni	45.9
oktober 'October'	3	ɔk'tobər	66.7	'towər	0.0
ongeveer 'approximately'	30	ɔŋɣə'ver	6.7	ɔ'fer	16.7
precies 'exactly'	82	prə'sis	18.3	'psis	24.4
publiek 'public'	3	py'blik	33.3	'blik	33.3
tandarts 'dentist'	17	'tɑndɑrts	5.9	'tɑs	0.0
volgend 'following'	21	'vɔlɣənt	4.8	'folnt	52.4
wedstrijd 'match'	10	'wɛtstrɛɪt	70.0	'wɛs	30.0
zelfs'even'	26	'zɛlfs	19.2	'zɛls	42.3

References

Adda-Decker, M., Boula de MareuBooil, P., Adda, G., & Lamel, L. (2005). Investigating syllabic structures and their variation in spontaneous French. *Speech Communication*, *46*, 119–139.

Baayen, R., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database* (*release*2). PA: Linguistic Data Consortium, University of Pennsylvania.

Binnenpoorte, D. (2006). *Phonetic transcriptions of large speech corpora*. Ph.D. thesis, Radboud Universiteit Nijmegen, Nijmegen, The Netherlands.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*(9/10), 314–345.

Booij, G. (1995). *The phonology of Dutch*. Oxford University Press.

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role and gender. *Language and Speech*, *44*(2), 123–147.

Chino, T., & Tsuboi, H. (1996). A new discourse structure model for spontaneous spoken dialogue. In *Proceedings of ICSLP* (pp. 1021–1024).

Connine, C. M., Ranbom, L. J., & Patterson, D. J. (2008). Processing variant forms in spoken word recognition: The role of variant frequency. *Perception and Psychophysics*, *70*(3), 403–411.

Cucchiarini, C., & Binnenpoorte, D. (2002). Validation and improvement of automatic phonetic transcriptions. In *Proceedings of ISCLP* (pp. 313–316). Denver, USA.

Dalby, J. M. (1986). Phonetic structure of fast speech in American English. In *Phonetics and Phonology*. Bloomington: Indiana University Linguistic.

Dilley, L. C., & Pitt, M. A. (2007). A study of regressive place assimilation in spontaneous speech and its implications for spoken word recognition. *Journal of the Acoustical Society of America*, *122*(4), 2340–2353.

Elffers, B., Van Bael, C., & Strik, H. (2005). *ADAPT: Algorithm for dynamic alignment of phonetic transcriptions*. Internal Report, Department of Language and Speech, Radboud University Nijmegen.

Ernestus, M. (2000). *Voice assimilation and segment reduction in casual Dutch. A corpus-based study of the phonology–phonetics interface*. Ph.D. thesis, LOT.

Ernestus, M., Lahey, M., Verhees, F., & Baayen, R. H. (2006). Lexical frequency and voice assimilation. *Journal of the Acoustical Society of America*, *120*(2), 1040–1051.

Ernestus, M., & Mak, W. M. (2004). Distinctive phonological features differ in relevance for both spoken and written word recognition. *Brain and Language*, *90*, 378–392.

Goeman, T. (1999). *T-deletie in Nederlandse Dialecten. Kwantitatieve analyse van structurele, ruimtelijke en temporele variatie*. Ph.D. thesis, HAG.

Hämäläinen, A., Gubian, M., ten Bosch, L., & Boves, L. (2009). Analysis of acoustic reduction using spectral similarity measures. *Journal of the Acoustical Society of America*, *126*(6), 3227–3235.

Hämäläinen, A., ten Bosch, L., & Boves, L. (2007). Modelling pronunciation variation using multi-path HMMs for syllables. In *Proceedings of ICASSP* (pp. IV781–IV784).

Hoste, V., Daelemans, W., & Gillis, S. (2004). Using rule-indication techniques to model pronunciation variation in Dutch. *Computer Speech and Language*, *18*, 1–23.

Janse, E., Nooteboom, S. G., & Quené, H. (2007). Coping with gradient forms of /t/-delition and lexical ambiguity in spoken word recognition. *Language and Cognitive Process*, *22*(2), 161–200.

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

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In: Bybee, J., & Hopper, P. (Eds.), *Frequency and the emergence of linguistic structure* (pp. 229–254). John Benjamins.

Kessens, J. M., Cucchiarini, C., & Strik, H. (2003). A data-driven method for modeling pronunciation variation. *Speech Communication*, *40*, 517–534.

Kessens, J. M., Wester, M., & Strik, H. (2000). Automatic detection and verification of Dutch phonological rules. In *PHONUS 5: Proceedings of the workshop on phonetics and phonology in ASR* (pp. 117–128). Saarbruecken, Germany.

Kipp, A., Wesenick, M., & Schiel, F. (1996). Automatic detection and segmentation of pronunciation variants in German speech corpora. In *Proceedings of ICSLP* (pp. 106–109).

Kipp, A., Wesenick, M., & Schiel, F. (1997). Pronunciation modeling applied to automatic segmentation of spontaneous speech. In *Proceedings of eurospeech* (pp. 1023–1026).

Kohler, K. J. (1998). The disappearance of words in connected speech. *ZAS Working Papers in Linguistics*, *11*, 21–34.

Kohler, K. J. (2001). Articulatory dynamics of vowels and consonants in speech communication. *Journal of the International Phonetic Association*, *31*, 1–16.

Koskenniemi, K. (1983). Two-level morphology: A general computational model for word-form recognition and production. Publication no. 11, University of Helsinki, Department of General Linguistics.

Kuipers, C., & van Donselaar, W. (1997). The influence of rhythmic context on schwa epenthesis and schwa deletion in Dutch. *Language and Speech*, *41*, 87–108.

Levelt, W., Roelofs, A., & Meyer, A. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1–37.

Losiewicz, B. (1992). *The effect of frequency on linguistic morphology*. Ph.D. thesis, University of Texas.

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

Nooteboom, S. (1979). Perceptual adjustment to speech rate: A case of backward perceptual normalization. In H. D. van Witsen (Ed.), *Anniversaries in phonetics: Studia gratulatoria dedicated to Hendrik Mol* (pp. 255–269). Amsterdam: Institute of Phonetic Sciences.

Norris, D., McQueen, J. M., & Cutler, A. (1995). Competition and segmentation in spoken-word recognition. *Journal of Experimental Phonology: Learning, Memory, and Cognition*, *21*(5), 1209–1228.

Oostdijk, N., Goedetier, W., van Eynde, F., Boves, L., Martens, J. -P., Moortgat, M., et al. (2002). Experiences from the spoken Dutch corpus project. In *Proceedings of LREC* (pp. 340–347).

Pluymaekers, M., Ernestus, M., & Baayen, H. R. (2005). Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America*, *118*(4), 2561–2569.

Rietveld, A. C. M., & Koopmans-van Beinum, F. J. (1987). Vowel reduction and stress. *Speech Communication*, *6*, 217–229.

Rietveld, T., van Hout, R., & Ernestus, M. (2004). Pitfalls in corpus research. *Computers and the Humanities*, *38*(4), 343–362.

Saraçlar, M., Nock, H., & Khudanpur, S. (2000). Pronunciation modelling by sharing gaussian densities across phonetic models. *Computer Speech and Language*, *14*, 137–160.

Scharenborg, O., & Boves, L. (2002). Pronunciation variation modelling in a model of human word recognition. In *Proceedings of workshop on pronunciation modeling and lexicon adaptation* (pp. 65–70), Estes Park, USA.

Schuppler, B., van Dommelen, W., Koreman, J., & Ernestus, M. (2009). Word-final [t]-deletion: An analysis on the segmental and sub-segmental level. In *Proceedings of interspeech* (pp. 2275–2278).

Son, R. V., Binnenpoorte, D., van den Heuvel, H., & Pols, L. (2001). The IFA corpus: A phonemically segmented Dutch 'open source' speech database. In *Proceedings of eurospeech* (pp. 2051–2054).

Strik, H., Russel, A., van den Heuvel, H., Cucchiarini, C., & Boves, L. (1997). A spoken dialog system for the Dutch public transport information service. *International Journal of Speech Technology*, *2*(2), 121–131.

Swerts, M., Kloots, H., Gillis, S., & De Schutter, G. (2001). Factors affecting schwa-insertion in final consonant clusters in standard Dutch. In *Proceedings of eurospeech* (pp. 75–78).

Van Bael, C. (2007). *Validation, automatic generation and use of broad phonetic transcriptions*. Ph.D. thesis, Radboud Universiteit Nijmegen, Nijmegen.

Van Bael, C., Baayen, R. H., & Strik, H. (2007). Segment deletion in spontaneous speech: A corpus study using mixed effects models with crossed random effects. In *Proceedings of interspeech* (pp. 2741–2744), Antwerp, Belgium.

Van Bael, C., Boves, L., van den Heuvel, H., & Strik, H. (2007). Automatic phonetic transcription of large speech corpora. *Computer Speech and Language*, *21*, 652–668.

van Bergem, D. (1993). Acoustic vowel reduction as a function of sentence accent, word stress and word class. *Speech Communication*, *12*, 1–23.

Van de Velde, H., Gerritsen, M., & van Hout, R. (1996). The devoicing of fricatives in standard Dutch. A real time study based on radio recordings. *Language Variation and Change*, *8*(2), 149–175.

Van de Velde, H., & van Hout, R. (2001). The devoicing of fricatives in a reading task. In T. van der Wouden, & H. Broekhuis (Eds.), *Linguistics in the Netherlands* (pp. 219–229). John Benjamins.

Van den Broeke, M., & van Heuven, V. (1979). One or two velar fricatives in Dutch. In H. V.D. Witsen (Ed.), *Anniversaries in phonetics: Studia gratulatoria dedicated to Hendrik Mal* (pp. 51–67). Amsterdam: Institute of Phonetic Sciences.

van den Heuvel, H., & Cucchiarini, C. (2001). /r/-deletion in Dutch: Rumors or reality? In H. Van de Velde, & R. van Hout (Eds.), *r-atics: Sociolinguistic, phonetic and phonological characteristics of* /r/ (Vol. 4, pp. 185–198). Etudes and Travaux–ILVP/ULB, Brussels.

van der Vliet, H. (2007). The Referentiebestand Nederlands as a multi-purpose lexical database. *International Journal of Lexicography*, *20*(3), 239–257.

Wells, J. (1997). SAMPA computer readable phonetic alphabet. In D. Gibbon, R. Moore, R. Winski (Eds.), *Handbook of standards and resources for spoken language systems* (chapter IV), B. Mouton de Gruyter.

Wester, M., Kessens, J. M., & Strik, H. (1998). Two automatic approaches for analyzing connected speech processes in Dutch. In *Proceedings of ICSLP* (p. 0373).

Young, S., Evermann, G., Kershaw, D., Moore, G., Odell, J., Ollason, D., et al. (2002). *The HTK book* (*version*3.2). Technical Report, Cambridge University, Engineering Department.