Qualitative and quantitative aspects of phonetic variation in Dutch *eigenlijk*

Mirjam Ernestus[a,b] & Rachel Smith[c]
[a]Centre for Language studies, Radboud University, Nijmegen, The Netherlands
[b]Max Planck Institute for Psycholingusitics, Nijmegen, The Netherlands
[c]Glasgow University Laboratory of Phonetics, University of Glasgow, UK

1. Introduction

In informal conversations in many languages, many word tokens are pronounced with fewer segments or with segments that are articulated more weakly than in careful speech (for an introduction to the phenomenon, see Ernestus and Warner 2011). For instance, the word *particular* may be pronounced like [pʰtʰɪkʰə] and *hilarious* like [hlɛrɛs] (Johnson 2004). These short word pronunciation variants are generally referred to as reduced forms, and we adopt this terminology here. Reduced forms typically occur in weak prosodic positions, especially in unaccented positions in the middle of sentences (e.g. Pluymaekers, Ernestus & Baayen. 2005a). This paper contributes to our knowledge of the characteristics of reduced forms by studying in detail one word type in Dutch (i.e. *eigenlijk* 'actually') that is known to show wide variation in its realization (e.g. Ernestus 2000: 141). The results shed light on the variation that a word may show and on how speakers from the same sociolinguistic group may differ in how they reduce. In addition, the results raise questions about the nature of reduced forms, the mental lexicon and psycholinguistic models of speech production and comprehension.

Nearly all previous research on reduced forms focused on the presence versus absence of segments and on the duration of these segments or (parts of) the words as measures of reduction. These studies have shown that many different factors affect the probability that a given word appears in a reduced form. These factors include speech rate (e.g. Raymond, Dautricourt, and Hume 2006; Kohler 1990), the word's phonological neighborhood density (Gahl, Yao, and Johnson 2012), its prosodic position (e.g. Bell et al. 2003), its a-priori-probability (e.g. Pluymaekers, Ernestus and Baayen 2005a; Gahl 2008), its probability in context (e.g. Bell et al. 2003; Pluymaekers, Ernestus, and Baayen 2005b; Bell et al. 2009) and the presence versus absence of a following hesitation (e.g., Bell et al. 2003). The influences of these factors suggest that reduction may result from time pressure: when time pressure is high, for instance because speech rate is high or because the word or the following word was easy to plan and is ready to be articulated, reduction is more likely to occur (e.g. Bell et al. 2009; Gahl et al. 2012).

In addition, several studies focusing on duration and on the presence versus absence of segments suggest that degree of reduction is under the speaker's direct control. For instance, speakers may choose not to reduce at high speech rates (e.g. van Son and Pols 1990, 1992), and degree of reduction correlates with speaker characteristics, including gender (e.g. Guy 1991; Phillips 1994), age (e.g. Guy 1991; Strik, van Doremalen, and Cucchiarini 2008) and socio-economic status (e.g. Labov 2001). Furthermore, speakers of different regiolects of a language may differ in degree of reduction for some words (e.g. Keune et al. 2005). Reduction is therefore not a fully automatic process, but is speaker-dependent and probably at least partly under the speaker's control.

Only a few studies so far have investigated more detailed phonetic characteristics of reduced forms. Such studies have shown that information about a word's identity is often preserved despite reduction or reorganization of the acoustic features or articulatory gestures that would be found in a canonical form. For example, reduced tokens of *support* may lack any evidence of a vowel portion between /s/ and the closure of /p/, yet may maintain aspiration of /p/, which is consistent with a singleton syllable-initial stop, rather than an /sp/ cluster. Thus the laryngeal specification of the stop preserves information that prevents reduced *support* from becoming ambiguous with *sport* (Manuel 1991; Manuel et al. 1992; see also Davidson 2006, and see Aalders & Ernestus, in preparation, for evidence that this also holds in casual speech). Similarly, some reduced forms of French *c'était* 'it was' can lack a voiced vowel between /s/ and /t/, yet retain traces of the vowel in the form of a lowered spectral centre of gravity in the latter part of the /s/ (Torreira and Ernestus 2011). In English *the,* /ð/ can assimilate in manner of articulation to a preceding nasal or lateral (in phrases like *in the, all the*), losing any evidence of frication; yet residues of /ð/ tend to be retained in the form of dentality (as cued by F2 at the nasal or lateral boundaries) and duration (Manuel 1995).

In extreme cases of reduction it may be impossible to linearly segment the speech signal, yet sufficient phonetic residue of a word's form may remain as to make it fully identifiable. Kohler (1999) described such residues as "articulatory prosodies", which "persist as non-linear, suprasegmental features of syllables, reflecting, e.g., nasality or labiality that is no longer tied to specific segmental units'' and may be quite extended in time (p. 89). For example, the German discourse marker *eigentlich* 'actually' is canonically produced as [aɪgŋtliç], but can be reduced to [aɪŋi] or [aĩĩ], with palatality, nasality, and duration serving to convey the word's "phonetic essence" (Niebuhr and Kohler 2011). Perception tests indicate that listeners may be sensitive to such articulatory prosodies, even in the absence of contextual clues (Niebuhr and Kohler 2011), just as they are to other aspects of phonetic detail in reduced speech (Manuel 1991, 1995). Thus reduction may involve significant departures from a word's canonical form, while preserving phonological contrast quite well (Warner and Tucker 2011).

The degree of reduction may be affected by the function that a word performs. Plug (2005) investigated reduction of the Dutch discourse marker *eigenlijk* 'actually' as a function of its pragmatic function. Plug analysed 49 tokens of *eigenlijk* performing two of the word's subfunctions, one being correction or clarification of a statement or assumption in a speaker's own utterance (self-repair), and the other being correction or clarification of something said or assumed by the interlocutor (other-repair). He observed that tokens with the function of self-repair tended to be produced fast and to be highly reduced in terms of their number of syllables and segments. Tokens whose function was other-repair tended to occur at the edges of prosodic phrases and to be produced slower and with more phonetic elaboration.

Like Plug (2005), the present study focusses on the Dutch word *eigenlijk*. As is the case for nearly all words in every language, we have very little *detailed* knowledge about the possible pronunciation variants of the word, and about how frequently these variants occur and under which conditions. The present study examines over 150 tokens of this word, produced by 18 speakers in informal conversations, examining their detailed phonetic characteristics and when particular clusters of characteristics are most likely to occur. Detailed data on the pronunciation variation of this word will form a good testing ground for common assumptions about reduction, including the assumption that reduced forms only occur in prosodically weak positions, and the related assumption that speakers mainly reduce to cope with time pressure.

Our main reason for studying *eigenlijk* is that it is known to show a wide variation in pronunciation, ranging from trisyllablic /'ɛɪxələk/ (see Figure 1 for an example) to monosyllabic variants like /'ɛɪxk/ and /'ɛɪk/ (see e.g. Ernestus 2000; Plug 2005; Pluymaekers, Ernestus and Baayen 2005b). The word shares this variability with many other words also ending in the suffix -/lək/ -*lijk* (e.g. Pluymaekers, Ernestus and Baayen 2005b). The word occurs relatively frequently in informal conversations (e.g. 1922 tokens per million word tokens in the Spoken Dutch Corpus, Oostdijk 2000), which allows us to study both intra- and interspeaker variation on the basis of tokens produced in a relatively short period of time.

As mentioned above, the word *eigenlijk* is a discourse marker that in general signals a contrast between what the speaker is saying and what (s)he or the interlocutor just said or implied, or is assumed to believe (e.g. Plug 2005; van Bergen et al. 2011); see, for instance, sentences (1, 2, 3) from our data set.

(1) *een van de, of eigenlijk de oudste, acht geveild zou worden.*
one of the, or actually the oldest, eight [a type of rowing boat] would be auctioned.

(2) *Nee, tenten hoef ik eigenlijk niet*
No, I actually do not need tents.
(In response to the interlocutor's request whether he would like to buy any tents.)

(3) *Van tandartsen word ik altijd eigenlijk helemaal nooit goed.*
I always actually completely never feel good around dentists
(Following the speaker's remark that he has a dentist appointment next Monday)

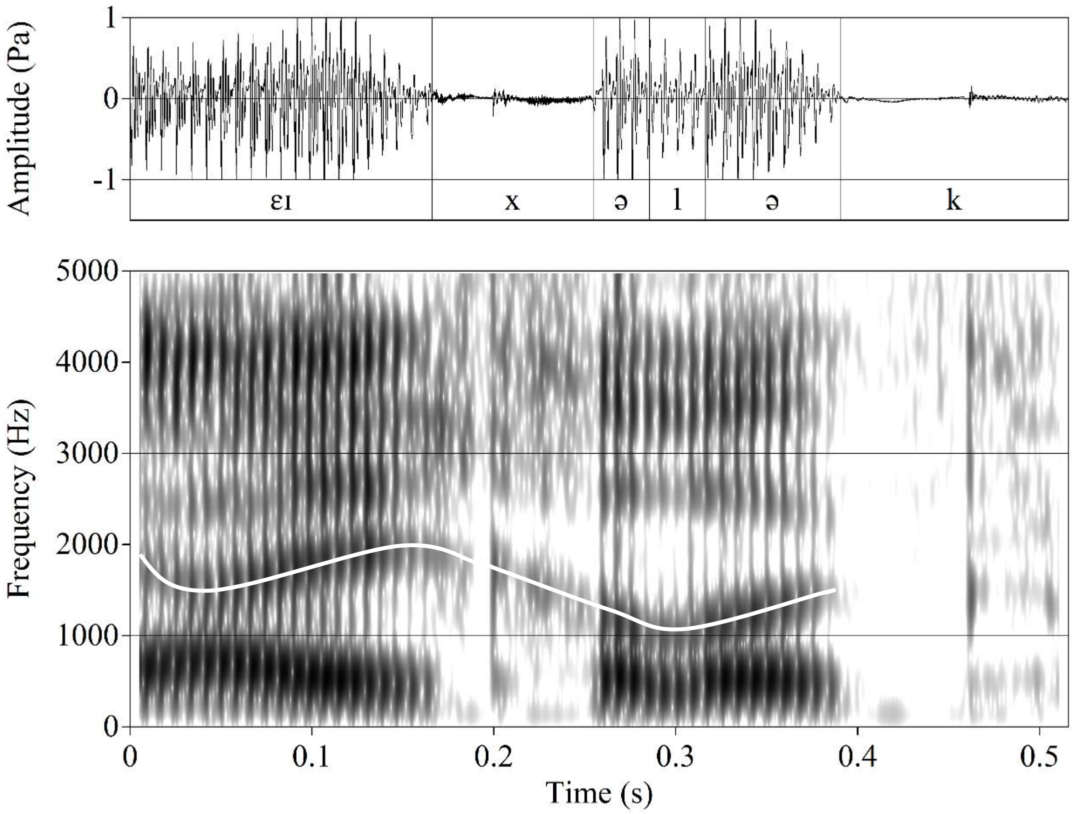


Figure 1: Waveform and spectrogram of an unreduced token of *eigenlijk*, produced as /ɛɪxələk/ by speaker S in the Ernestus Corpus of Spontaneous Dutch (Ernestus 2000). The white line indicates the F2 trajectory.

Within this broad function, several subfunctions can be distinguished. As described above, Plug (2005) investigated phonetic characteristics of 49 tokens representing two of these subfunctions. In the present paper, we will not distinguish between these subfunctions because they are often difficult to distinguish and because a focus on one or several of the subfunctions would severely reduce the number of word tokens that can be analyzed (as in Plug's study).

The Dutch word *eigenlijk* can occur in different positions in the sentence as illustrated in examples (1, 2, 3, 4, 5). Moreover, it can follow and precede different word types, as illustrated in these same examples. Noteworthy is example (3), in which *eigenlijk* is surrounded by other adverbs, as is frequently the case in spontaneous conversations.

(4) *Eigenlijk is dat actief.*
Actually that is active.

(5) *Het gaf heel veel informatie eigenlijk.*
It gave a lot of information actually.

The first part of this study provides an overview of the types of variation that we attest in our data set, and of the frequencies with which specific phonetic characteristics occur. The second part investigates which factors predict certain phonetic properties, including the number of syllables, presence of creaky voice, and the presence versus absence of /l/. Our analyses will include predictors that have been reported before to correlate with reduction degree (e.g. speech rate, the predictability of the preceding and following word, and the presence of sentential accent).

We also investigate the influence of a new predictor, the rhythm of the sentence. Because the word *eigenlijk* can be preceded and followed by a wide variety of words, the numbers of directly preceding and following unstressed syllables can vary as well. Speakers of Germanic languages prefer alternating patterns of stressed and unstressed syllables (e.g. Kelly and Bock 1988 and references therein). It is therefore possible that speakers of Dutch opt for a variant of *eigenlijk* with a number of unstressed syllables that optimizes the rhythm of the phrase. For instance, they may prefer a stressed monosyllabic variant if the word is followed by several unstressed syllables, and a di- or tri-syllabic variant, ending in one or two unstressed syllables, respectively, when the word is followed by a word with initial stress.

The next sections describe the corpus and our selection of the tokens (Section 2), the annotation system (Section 3), and provide a qualitative description of the attested variation (Section 4). We then present the results of our statistical modeling of some of the tokens' characteristics (section 5). We conclude the paper with a general discussion of these results (sections 6 and 7).

2. Materials

We extracted the tokens from the Ernestus Corpus of Spontaneous Dutch (Ernestus 2000). This corpus, recorded in the nineties, contains high quality recordings of ten conversations, each 90 minutes long, between pairs of friends or direct colleagues. A DAT-recorder recorded the speech by each interlocutor on a different track of a tape, by means of unidirectional microphones placed on a table between the speakers. The speakers are all male, highly educated and lived their whole lives in the Western part of the Netherlands. They speak a 'western' variant of Standard Dutch, which implies, among other things, that they do not distinguish between the voiced and voiceless velar fricative.

The conversation during the first part of each recording was elicited by a third person, who knew at least one of the speakers well. The speakers discussed topics as diverse as television quizzes, how they chose their professions, their experiences with dentists, and their opinions of euthanasia. In the second part of the recording, the speakers participated in a role-play in which one speaker sold tents, sleeping bags and backpacks to the other speaker, who pretended to be a shop owner. This second part also contained conversations covering a wide range of other topics since the speakers were encouraged to do so before and after the negotiations. The speech in the corpus sounds natural and casual, as is evidenced, among other things, by ratings of six other native speakers, the high frequency of discourse markers (including *eigenlijk*), and the amount of gossip in the corpus.

The 20 speakers in the corpus produced in total 339 *eigenlijk* tokens. The number of tokens per speaker ranges from 6 to 45 (see Table 1). We randomly selected 159 tokens, taking into account the following constraints and preferences. First, we only incorporated tokens that were produced fluently, without laughing and without much background noise (including the interlocutor's speech), so that detailed phonetic analysis is possible. Secondly, we wished to have minimally five tokens per speaker, so that we could study intraspeaker variation, and maximally eleven tokens, so that the data set would not be dominated by just a few speakers. Third, we preferred tokens produced in the free conversations over tokens produced in the role play and we discarded tokens that were produced in the first ten minutes of a recording because the speaker might not yet have been speaking very naturally. Many tokens did not meet all these requirements and preferences and, as a consequence, we lost Speaker G. We also decided not to incorporate Speaker K because this speaker often stumbled over his words. Table 1 shows the resulting number of tokens per speaker.

Table 1. The number of tokens produced by each speaker, the number incorporated in our analyses, and the number of studied tokens that were monosyllabic (see the section on Individual speaker differences).

Speaker ID	Total tokens	Tokens studied	Monosyllabic tokens (percentage of the tokens studied)
A	6	5	5 (100%)
B	10	9	2 (22%)
E	12	8	1 (13%)
F	26	10	0 (0%)
G	8	-	-
H	20	8	4 (50%)
I	15	9	1 (11%)
J	10	9	1 (11%)
K	28	-	-
L	45	11	0 (0%)
M	20	9	7 (78%)
N	13	7	2 (29%)
O	9	9	8 (89%)
P	12	11	2 (20%)
Q	26	10	1 (10%)
R	15	7	5 (71%)
S	17	10	3 (30%)
T	15	9	3 (67%)
U	22	9	7 (78%)
V	10	9	5 (56%)

3. Labeling procedure

A phonemic transcription of the entire intonational phrase containing *eigenlijk* was made by the first author, and checked by the second author. For the token of *eigenlijk,* an allophonic transcription was also made, specifying voicing of /k/ and /x/, but no other detail. The number of syllables in *eigenlijk* was identified. Then, labelling of prosody and of segmental detail were carried out by the two authors independently. All cases where the transcribers used different labels were resolved by joint listening, as were all cases where they placed labels more than 20 ms apart, which was not often the case. There was no obvious bias towards either transcriber's labelling.

For the prosodic labelling, the boundaries of the intonational phrase containing *eigenlijk* were annotated, and all syllables within this phrase were labelled as primary-accented, secondary-accented, stressed, or unstressed, that is, we distinguished four levels of prosodic strength, defined as follows.

Primary-accented: the most prominent pitch-accented syllable in the phrase. All phrases included minimally one primary-accented syllable; only six included two primary-accented syllables, and most of these were produced by the same speaker and contained equally prominent accents on *eigenlijk* and another word.

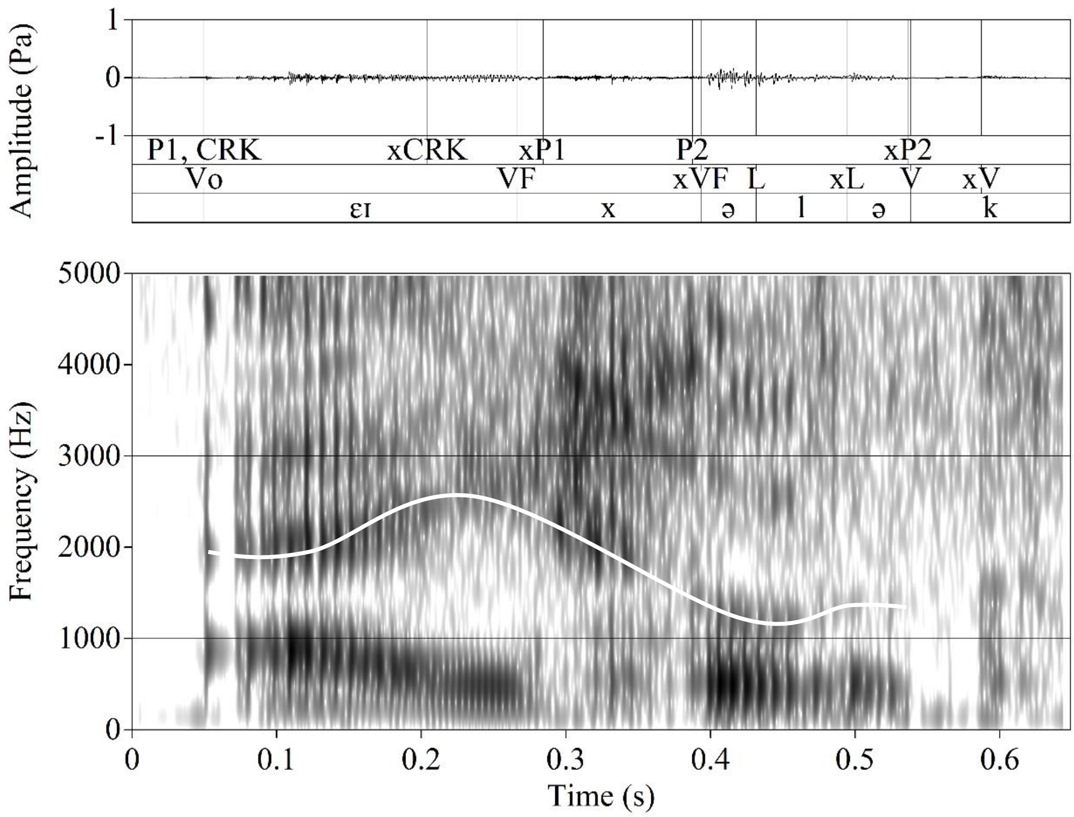
Secondary-accented: lexically-stressed syllables that were produced with a pitch movement or, rarely, a substantial increase in loudness in the absence of a pitch accent.

Stressed syllables: lexically-stressed syllables that were produced without a pitch movement. Unaccented function words were labelled as stressed unaccented if they had a full vowel and no evidence of segmental reduction, or as unstressed otherwise.

Unstressed: syllables lacking lexical stress. Filled pauses were always labelled as unstressed.

For the labelling of segmental detail, we defined a set of articulatory events in the larynx and supraglottal tract which are present in an unreduced token of *eigenlijk*, including the onsets and offsets of periodicity and the onsets and offsets of creaky voice. These events are described in detail in the following paragraph and illustrated in Figure 2. If an event was identifiable in the waveform and spectrogram, it was labelled. If it was absent or unidentifiable because of reduction, then that label was omitted. By labelling events rather than segments, we aimed to achieve maximum comparability across tokens that had different degrees of reduction, while avoiding parsing the signal exhaustively into phoneme-sized segments, which can be very challenging for reduced speech. We made one exception: if a velar stop was present, we marked its offset in the signal, even if the stop was unreleased and directly followed by another stop. In these cases we placed the boundary for the stop in the middle of the long closure formed by the two stops. We thus maximized the number of velar stops whose durations we could analyse (see below). However, utterance-final unreleased stops were excluded from this analysis, as we had no principled way to estimate their durations.

As Figure 2 shows, on the *larynx* tier we labelled the onsets and offsets of periodicity (numbered as P1 and xP1 for the first portion of periodicity, P2 and xP2 for the next, etc). We also labelled the onset (CRK) and offset (xCRK) of creaky voice, if present during /ɛɪ/. Our criterion for identifying creaky voice was irregularity of periods. On the *upper articulators* tier we labelled the following acoustic events: the onset of the stressed /ɛɪ/ vowel (Vo); the onset and offset of velar frication (VF, xVF respectively), of lateral quality (L, xL) and of velar closure (V, xV). These labels allowed us to calculate the duration of the whole word token (defined as extending from the onset of the stressed vowel to the last labelled event) and durations of individual segments, including the unstressed vowels, if present.



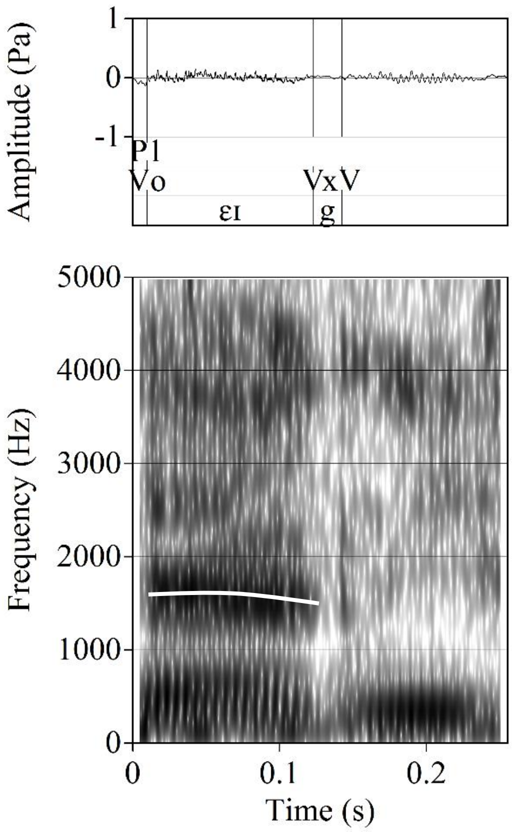


Figure 2: Labelling of segmental detail of two tokens of *eigenlijk.* Top, an unreduced token produced as /ɛɪxələk/ by speaker H. Bottom, a reduced token produced as [ɛɪg] (phonemically /ɛɪk/) by speaker M (the following word, *niet* 'not', is also shown). In each case, the top tier indicates laryngeal events, the second tier indicates events involving the upper articulators, and the third tier shows the allophonic transcription (see text for details). The white line indicates the F2 trajectory.

4. General description of variation within the word

The Appendix lists the variation that we discuss in this and the following section for which we can provide frequency data.

We first focus on the variation in the number of syllables. Figure 3 shows the number of mono-, di- and trisyllabic tokens for the four prosodic statuses that we distinguished. We see that the majority of tokens are disyllabic, and are thus one syllable shorter than the full form. Moreover, we find that many disyllabic and some monosyllabic tokens are accented (8 monosyllabic tokens are primary-accented and 16 secondary-accented). The word *eigenlijk* thus does not follow the well-known pattern that accented word tokens show little reduction (e.g. Bell et al. 2003). The variation in the number of syllables does not just result from the prosodic status of the word.

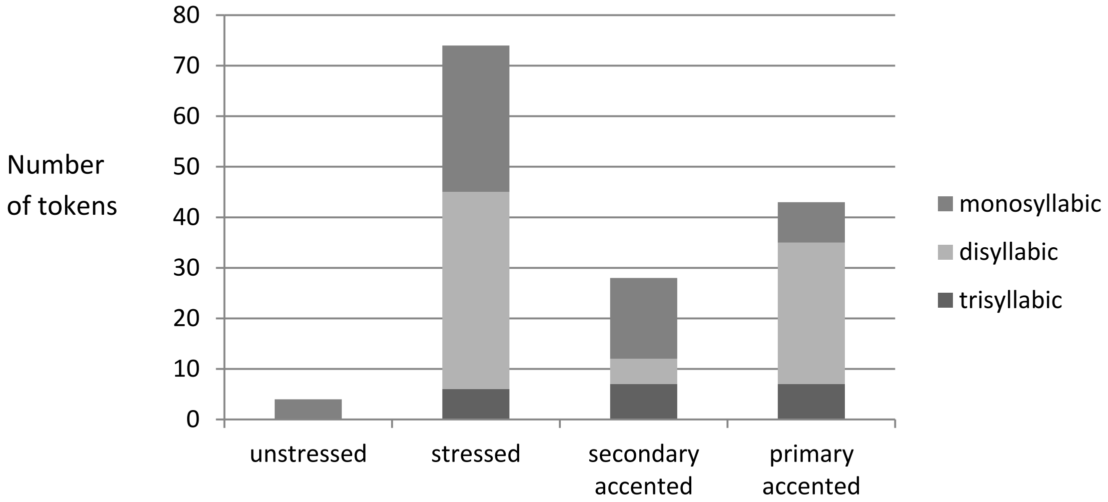


Figure 3: Number of trisyllabic, disyllabic and monosyllabic tokens that are unstressed, stressed, secondary accented or primary accented.

With regard to duration, tokens were on average longer when they had a greater number of syllables (3 syllables: 386 ms; 2 syllables: 310 ms; 1 syllable: 197 ms), but Figure 4 shows that the durational ranges for tri-, di- and monosyllabic tokens overlapped considerably, and we observed trisyllabic tokens as short as 257 ms.

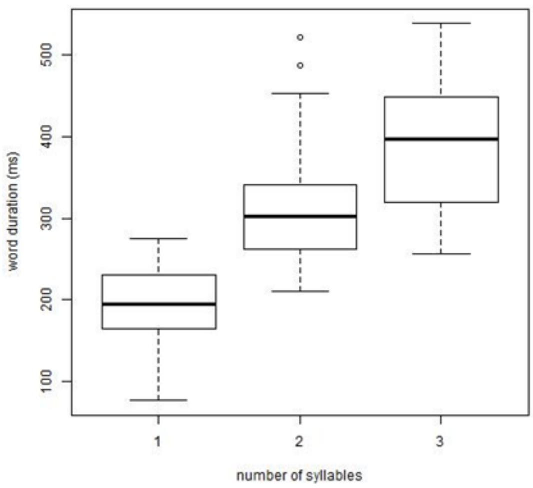


Figure 4: Boxplot of the duration of *eigenlijk* (ms), according to its number of syllables. The bottom and top of each box indicate the first and third quartiles; the band inside the box represents the second quartile (the median), while the

whiskers extend to the minimum and maximum values that are maximally 1.5 interquartiles from the box. The two small circles are outliers.

Table 2 lists the phonemic transcriptions of the tokens, and allophonic transcriptions specified for the voicing of /k/ and /x/ but showing no other detail, along with the frequency of each transcription. The most frequent phonemic form is disyllabic /ɛɪxlək/ (57 tokens). The monosyllabic form /ɛɪk/ (22 tokens) comes in second position, and the full form /ɛɪxələk/ (20 tokens) in third. Also common are monosyllabic forms /ɛɪxk/ (16 tokens) and /ɛɪx/ (13 tokens) and the disyllabic /ɛɪxək/ (14 tokens).

Importantly, we did not see a clear correlation between the structure of a token and how clearly its segments were articulated. Tokens that are highly reduced in terms of their number of syllables and segments, could nonetheless exhibit clearly articulated and tightly coordinated segments, and vice versa, as discussed below.

Table 2 makes clear that all forms in our dataset contain, minimally, a full front vowel, and at least one obstruent articulated at the back of the mouth. These appear to be essential phonetic components of *eigenlijk*. The vowel is typically a closing diphthong, but varies in degree of diphthongization, and can look and sound quite monophthongal (e.g. Figure 2, right panel; Figure 9). As regards the obstruent(s), 114 tokens (72%) were transcribed as containing both a fricative and a stop, 19 (12%) only a fricative and 26 (16%) only a stop. The obstruents are typically voiceless, but are voiced in 18% of cases. The fricative's place of articulation is normally velar or uvular. Eighty-seven tokens (55%) were transcribed as containing /l/, while several more contain a residual trace of /l/, as discussed further below.

Table 2. Tokens of *eigenlijk* found in the corpus. Phonemic and allophonic transcriptions are shown. The allophonic transcriptions specify the voicing of /k/ and /x/, but no other detail. Note that [ɛɪxl] was once perceived as disyllabic and once as monosyllabic.

Token structure	**N**	**Transcription**	
		Phonemic	**Allophonic**
Trisyllabic	**20**		
Vowel + fricative + schwa + lateral + schwa + stop	20	/ɛɪxələk/ 20	[ɛɪxələk] 17, [ɛɪxələg] 2, [ɛɪɣələk] 1
Disyllabic	**82**		
Vowel + fricative + lateral + schwa + stop	60	/ɛɪxlək/ 57 /ɛɪxləŋ/ 1 /ɛxlək/ 2	[ɛɪxlək] 42, [ɛɪxləg] 7, [ɛɪɣlək] 8, [ɛɪɣləg] 1 [ɛɪxləŋ] 1 [ɛxlək] 1, [ɛxləg] 1
Vowel + fricative + schwa + stop	15	/ɛɪxək/ 14 /ɛxək/ 1	[ɛɪxək] 6, [ɛɪxəg] 2, [ɛɪɣək] 5, [ɛɪɣəg] 1 [ɛxəg] 1
Vowel + fricative + lateral	1	/ɛɪxl/ 1	[ɛɪxl] 1
Vowel + fricative + schwa + lateral	1	/ɛɪxəl/ 1	[ɛɪxəl] 1
Vowel + lateral + schwa + stop	4	/ɛɪlək/ 4	[ɛɪlək] 3, [ɛɪləg] 1
Monosyllabic	**57**		
Vowel + fricative	16	/ɛɪx/ 13 /ɛx/ 3	[ɛɪx] 11, [ɛɪɣ] 2 [ɛx] 2, [ɛɣ] 1
Vowel + stop	22	/ɛɪk/ 22	[ɛɪk] 13, [ɛɪg] 9
Vowel + fricative + stop	18	/ɛɪxk/ 16 /ɛxk/ 2	[ɛɪxk] 15, [ɛɪɣk] 1 [ɛxk] 2
Vowel + fricative + lateral	1	/ɛɪxl/ 1	[ɛɪxl] 1
Total	**159**		

Trisyllabic tokens were produced as the canonical form /ɛɪxələk/. Figures 1 and 2 (left panel) show typical trisyllabic tokens. Note the formant dynamics, in particular how F2 rises through the first diphthong to reach a maximum at the start of /x/, then falls to reach its minimum during /l/, before rising again into /k/. The first schwa is typically shorter than the second (mean durations 19 versus 46 ms, respectively).

Despite containing all or almost all of the segments expected in the canonical form, some of the trisyllabic tokens did display elements of reduction. Some were quiet and/or breathy, particularly if phrase-final. Others were rapidly articulated: Figure 5 shows a trisyllabic token that is very short indeed (268 ms) and whose formant dynamics follow a less extreme version of the pattern described above. Among the types of reduction found in trisyllabic tokens were also monophthongisation of the vowel, incomplete closure of the stop, or devoicing of the first schwa.

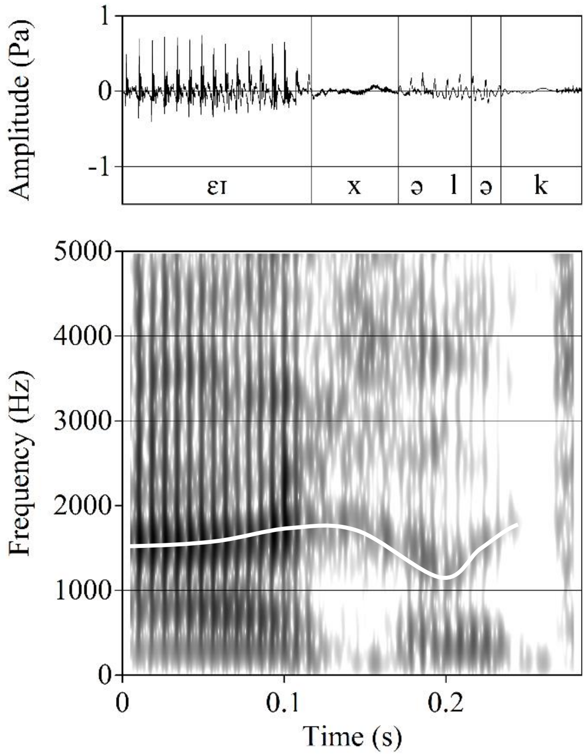


Figure 5: Trisyllabic token of *eigenlijk* produced as /ɛɪxələk/ by speaker E. Note the short duration and the less extreme excursions of F2 (indicated by the white line) compared to the trisyllabic tokens in Figures 1 and 2. The white line indicates the F2 trajectory.

Disyllabic tokens took a wide range of forms. The most common disyllabic form was /ɛɪxlək/, very similar to trisyllabic tokens, but with no discernible schwa before /l/. Such schwa loss in unstressed syllables before /l/ and /r/ is very common also in English (e.g. Gimson & Cruttenden 1994).

Disyllabic tokens evinced a range of interesting phonetic behavior at the juncture between the stem and suffix. First, there was variation in the extent to which the expected laryngeal events were produced, and in how they were aligned with respect to events involving the upper articulators. For example, we encountered numerous cases of progressive voice assimilation of /l/ to the velar fricative, that is, cases where /l/ and on occasion the word's entire second syllable was devoiced (e.g. Figure 6). We also found cases of regressive voice assimilation, that is, where /x/ was voiced by assimilating to /l/, sometimes resulting in a token that was voiced in its entirety (e.g. Figure 7). Voice assimilation involving /l/ has not been described for Dutch so far. When voicing of the fricative occurred, it was sometimes accompanied by weakening of the degree of stricture, and/or loss of place cues, such that the frication sounded glottal rather than velar or uvular. In four disyllabic tokens, phonemically /ɛɪlək/, we found no evidence of velar frication at all, but simply a lateral approximant at the syllable boundary.

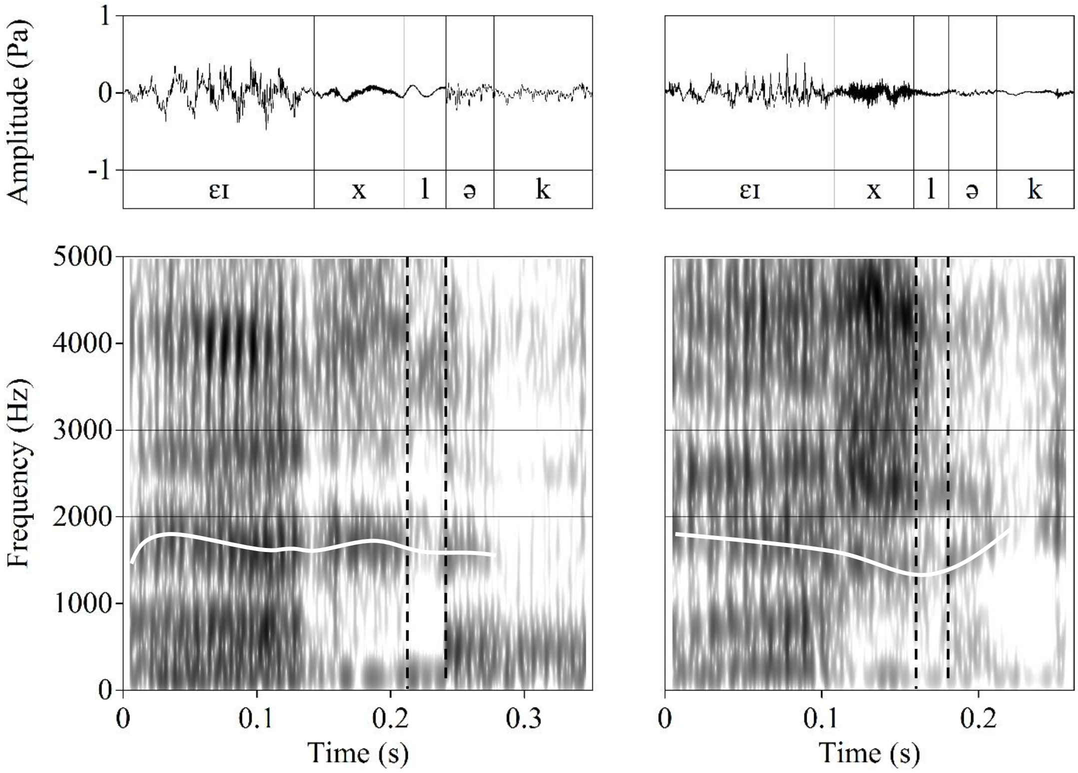


Figure 6: Disyllabic tokens of *eigenlijk* produced as /ɛɪxlək/, illustrating devoicing in the second syllable. Left: Token produced by speaker N, with a devoiced /l/. Right: Token produced by speaker V, where the second syllable is devoiced, but preserves the formant dynamics consistent with a /lə/ sequence. Dashed lines on spectrograms indicate the boundaries of /l/. White lines indicate the F2 trajectory.

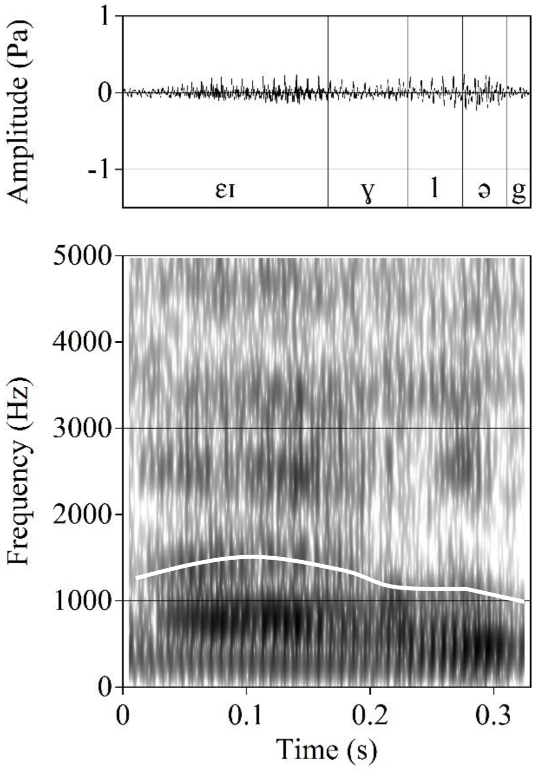


Figure 7: Fully voiced disyllabic token of *eigenlijk* produced as [ɛɪɣləg] (phonemically /ɛɪxlək/) by speaker T*.* The white line indicates the F2 trajectory.

Second, in disyllabic tokens we found variation in the alignment of events involving the upper articulators. In several tokens, most of them spoken by speaker F, the /x/ and /l/ appear strongly coarticulated. This speaker seems to produce lateral frication (e.g. Figure 8, left panel) as a solution to the problem of producing two very different articulations in swift succession. A different strategy is adopted by speaker S (e.g. Figure 8, right panel), who appears to start backing the tongue body in preparation for the /l/ already from the start of the frication, such that F2 reaches its minimum in the middle of the fricative portion. Finally, among the tokens transcribed phonemically as /ɛɪxək/, we also observed cases where an /l/ was not unambiguously present, but nevertheless left some residual trace in the signal, in the form of an F2 dip (e.g. Figure 9).

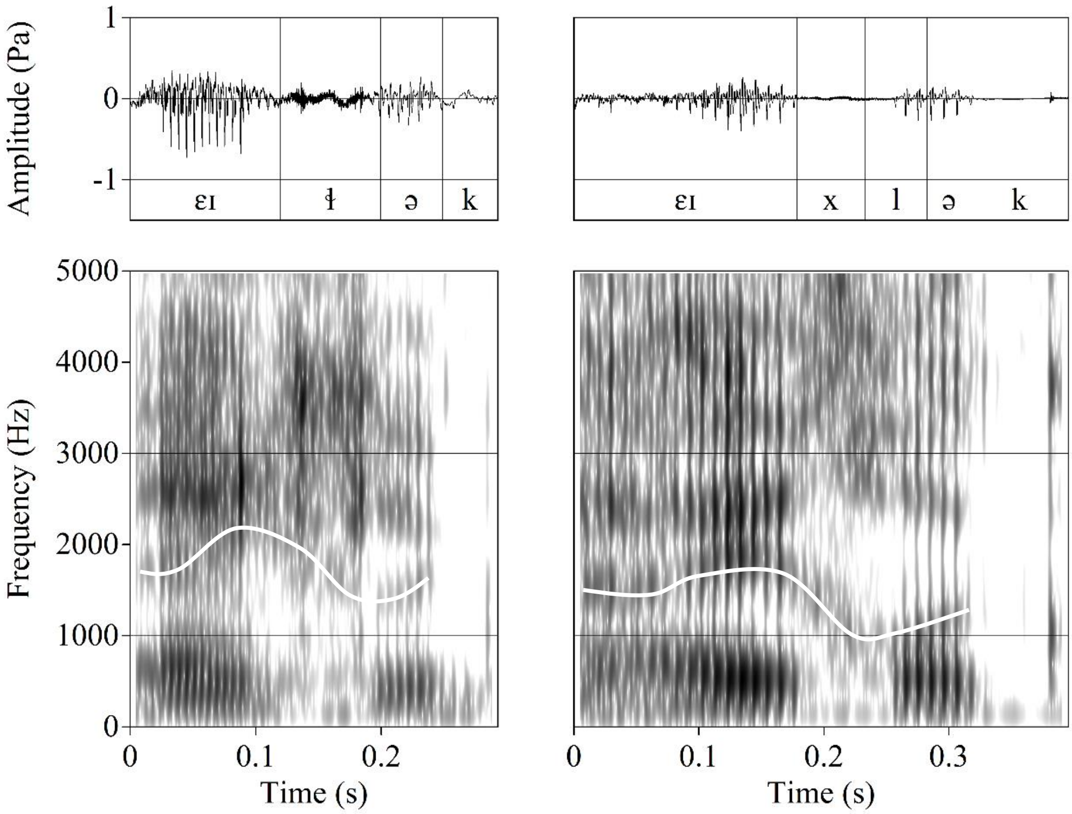


Figure 8: Disyllabic tokens of *eigenlijk* in which strong coarticulation between velar frication and laterality is audible. White lines indicate F2 trajectories. Left panel: token produced as /ɛɪxlək/ by speaker F, i.e. with an apparent lateral fricative [ɬ] in place of /xl/. Right panel: token produced as /ɛɪxlək/ by speaker S, in which F2 (indicated by the white line) falls rather steeply from the start of the frication.

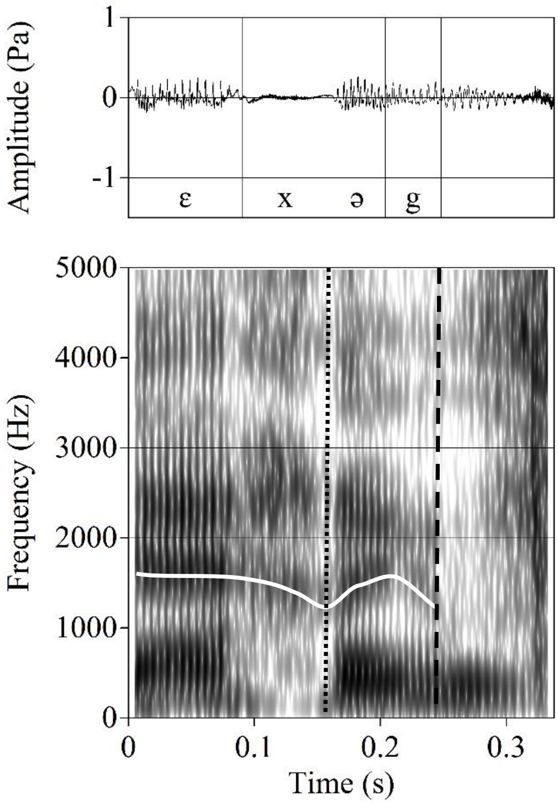


Figure 9: Disyllabic token of *eigenlijk*, produced as [ɛxəg] (phonemically /ɛxək/) by speaker I in the context *dat doe ik eigenlijk nooit* 'I actually never do that'. This token was not heard as containing a definite /l/, but the spectrogram indicates a residual trace of /l/, manifest as an F2 dip around 0.15 seconds (the white line indicates F2, and the black dotted line the F2 minimum). Note also the assimilation of the final stop to the following /n/ in terms of voicing and nasality. The black dashed line indicates the start of *nooit* (produced with laughter).

Monosyllabic tokens also took a number of forms. Some ended in a sequence of fricative followed by stop (phonemically /ɛɪxk/). The obstruent cluster /xk/ is not a legitimate syllable coda in Dutch. It was often produced with rather long duration relative to the vowel (e.g. Figure 10). A small number of tokens, with particularly long obstruent clusters, were difficult to classify in terms of their number of syllables: despite having only one syllable peak, they sounded almost disyllabic (see Aoyagi in press for a possible theoretical account of this finding).

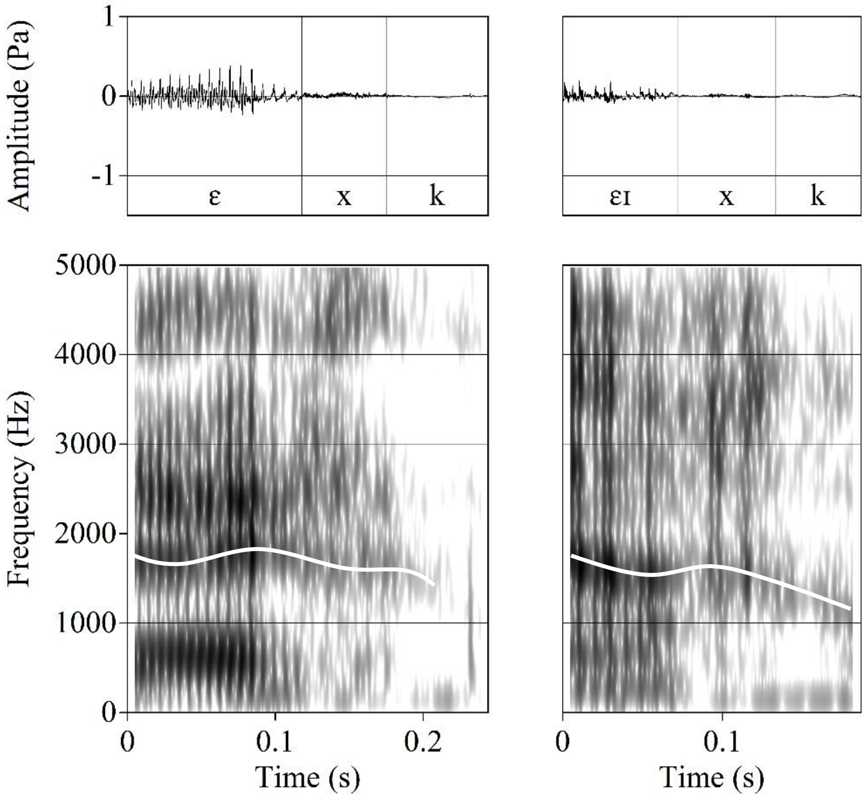


Figure 10: Monosyllabic tokens of *eigenlijk*, produced as /εxk/ by speaker I (left panel) and as /εɪxk/ by speaker O (right panel). Note the long duration of the obstruent portion compared to the vowel in both tokens. White lines indicate F2 trajectories.

Other monosyllabic tokens contain a single voiceless obstruent, either /x/ or /k/. The formant dynamics of the vowel are quite variable in these tokens. Some tokens show a flat or falling F2 during the vowel (e.g. Figure 11). Others have a typically diphthongal vowel ending in a clear velar pinch (convergence of F2 and F3) at the transition into the obstruent (e.g. Figures 12 and 13). We are not sure to what to attribute the difference between these two patterns, but it may relate to the place of articulation of the obstruent: the diphthongal vowels may precede consonants with a velar place, and the vowels with flat or falling F2 may precede post-velar or uvular consonants (for comments on how a velar vs. uvular place distinction can affect vowel formants, see Gordon, Barthmaier and Sands 2002).

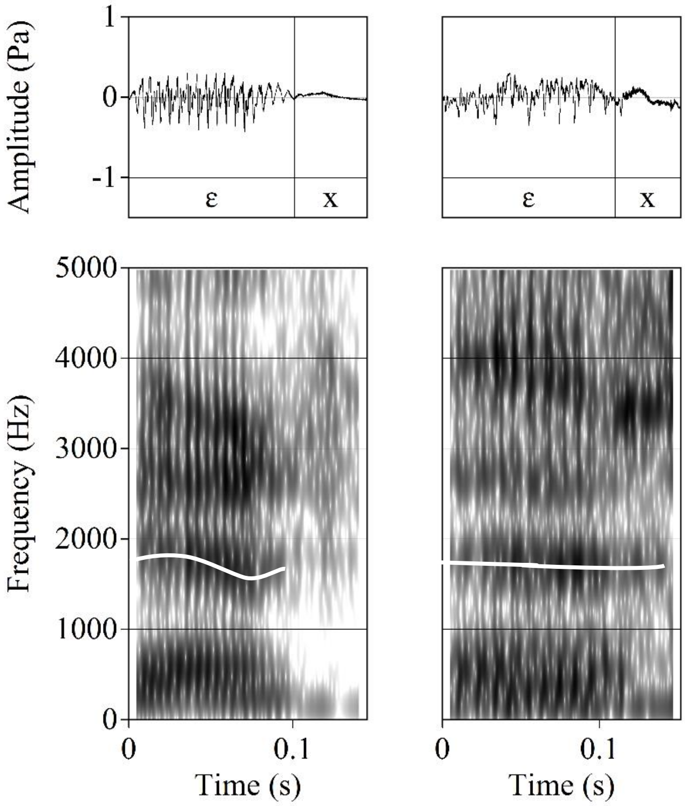


Figure 11: Monosyllabic tokens of *eigenlijk,* produced as /εx/ by speakers P (left panel) and N (right panel). In each case F2 is indicated by a white line; note the flat or falling F2 at the transition into the obstruent. White lines indicate F2 trajectories.

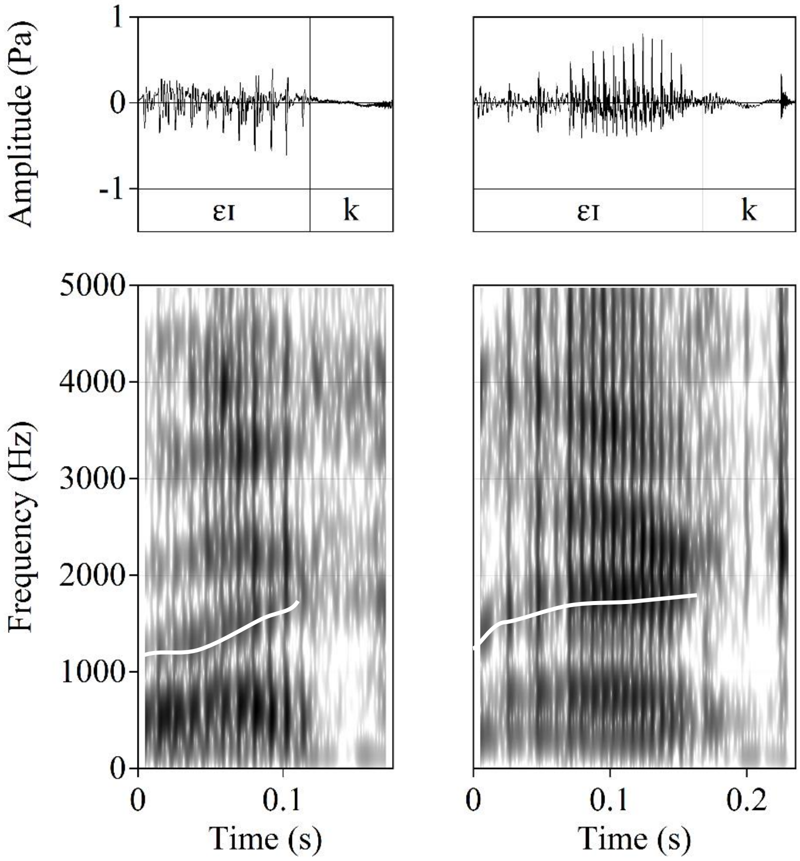


Figure 12: Monosyllabic tokens of *eigenlijk,* produced as /ɛɪk/ by speakers S (left panel) and R (right panel). F2 is indicated in each case by a white line; note its rise and the convergence of F2 and F3 (velar pinch) at the transition into the obstruent. White lines indicate F2 trajectories.

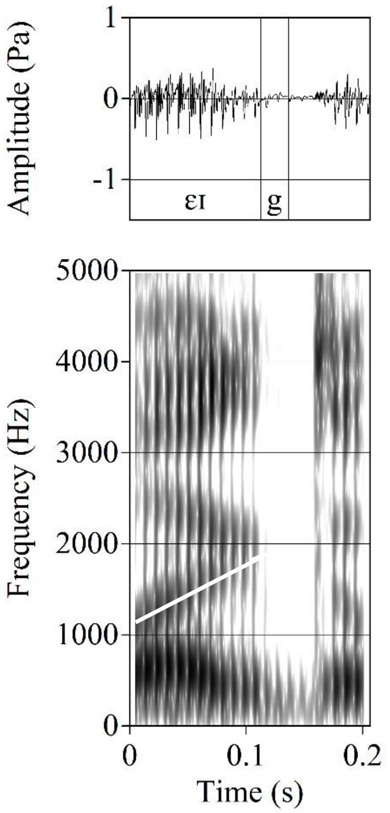


Figure 13: Monosyllabic token of *eigenlijk,* produced by speaker S as [ɛɪg] (phonemically /ɛɪk/) in the context *eigenlijk door* 'actually by'. Note the velar pinch at the transition into the stop. The white line indicates the F2 trajectory.The final stop is also assimilated in voice to the following /d/, and is unreleased.

Finally, creaky voice was common in the initial full vowel /ɛɪ/. This vowel often carries stress or accent (see e.g. Figure 3), and according to van Jongenburger and van Heuven (1991), the word may therefore be expected to be preceded by a glottal stop or similar phonetic events (i.e., glottalisation/creaky voice), especially after a vowel. Indeed, half of our tokens showed creaky voice at the start of the vowel (83 cases, or 52% of the data set), and less frequently other types of non-modal voice quality, such as harshness or breathiness.

4.2 Voice assimilation

The *eigenlijk* tokens, whether tri-, di- or mono-syllabic, show unexpected patterns of voice assimilation. Dutch is typically assumed to have only two processes of voice assimilation affecting sequences of obstruents. The first process voices obstruents preceding /b/ and /d/, while the second devoices /v/ and /z/ after voiceless obstruents (e.g. Booij 1995: 58, 59). We observed examples of these processes: see e.g. Figure 13 above. Yet we also found cases of assimilation not described in the literature. Thirteen tokens showed voicing of the word-final velar obstruent before nasal-initial words such as *niet* 'not', *nooit* 'never', and *nog* 'yet' (e.g. Figure 2, right panel, which illustrates an assimilated final stop in a monosyllabic token of *eigenlijk,* in the set phrase *ik weet het eigenlijk niet* 'I actually don’t know'; Figure 9; Figure 14) and we also observed voicing before /ʋ/ in one case. Typically for tokens of this kind, the stop is weak and very short, as little as 20 ms, relative to a vowel lasting 110-160 ms.

Furthermore, both /x/ and /k/ were sometimes voiced when followed by a vowel, whereas intervocalic voice assimilation at prosodic word boundaries is assumed to be restricted to fricatives (Booij 1995: 147). Finally, a handful of tokens of *eigenlijk* were followed by devoiced nasal stops, probably resulting from progressive voice assimilation induced by /k/ (e.g. Figure 15). All in all, the tokens of *eigenlijk* in our data set show more voice assimilation than would be expected on the basis of the existing literature.

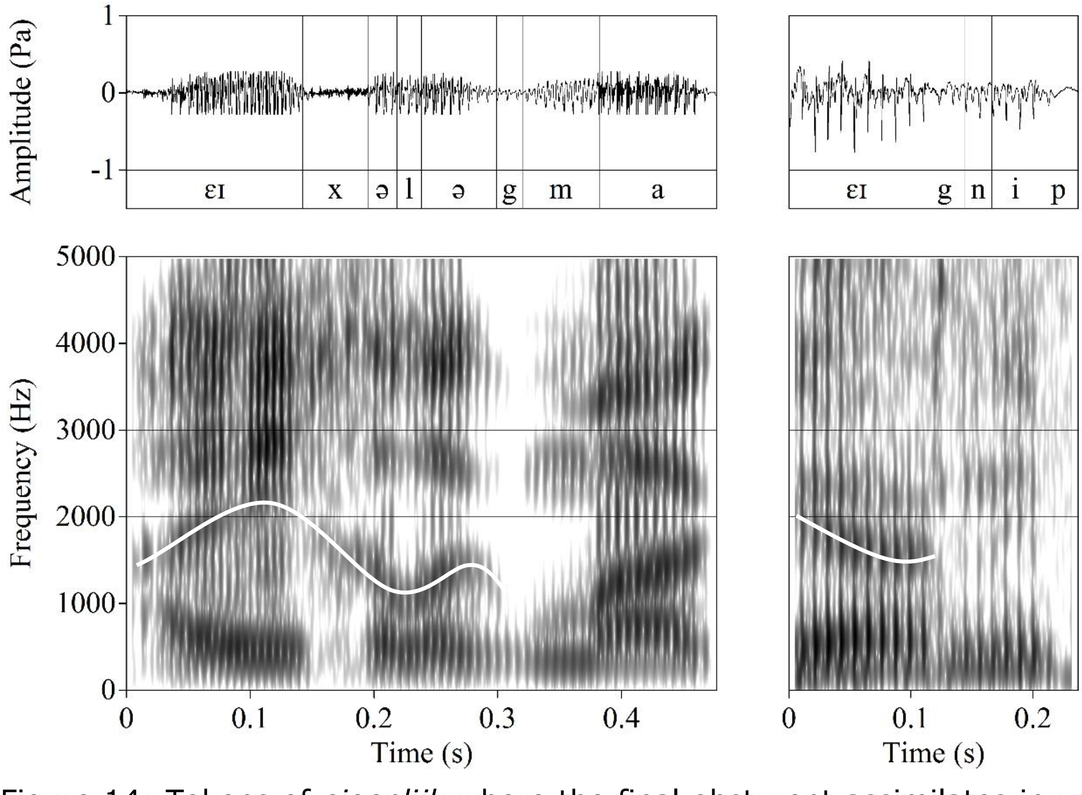


Figure 14: Tokens of *eigenlijk* where the final obstruent assimilates in voice to a following nasal segment. Left panel: [ɛɪxələg] (phonemically /ɛɪxələk/), in the context *eigenlijk maak ik…* 'actually I make...', by speaker I. Right panel: [ɛɪg] (phonemically /ɛɪk/), in the context *ik weet het eigenlijk niet precies* 'I actually do not know exactly', by speaker U. White lines indicate F2 trajectories.

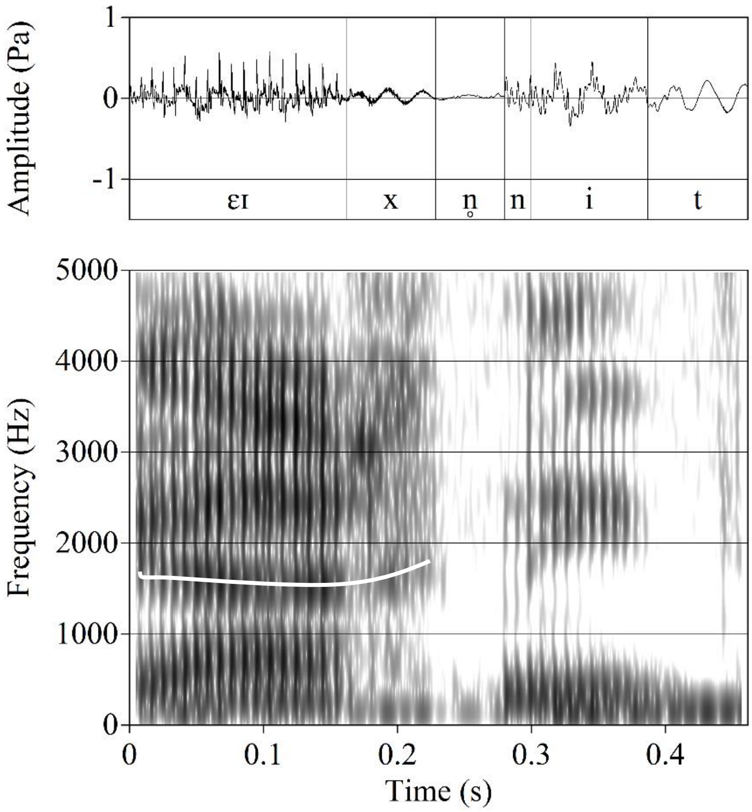


Figure 15: Token of *eigenlijk* produced as /ɛɪx/ by speaker O in the context *eigenlijk niet* 'actually not'*.* The token is followed by a partially devoiced /n/ which has assimilated in voice to the final segment of *eigenlijk*. The white line indicates the F2 trajectory.

4.3 Individual speaker differences

Although the speakers form a homogeneous group (they are all adult speakers coming from the same region and from the same socio-economic class), they clearly show individual differences. Speakers vary in their propensity to produce tokens with different numbers of syllables. This is illustrated in Table 1, which lists for each speaker the percentage of monosyllabic tokens. Speaker A produced only monosyllabic tokens, speakers F and L produced no monosyllabic tokens, and all other speakers produced a mixture of monosyllabic and di-/tri-syllabic tokens. There is also clear variation in the frequency of trisyllabic tokens: half of these were produced by only three speakers (E, J, and S). Part of this variation may be accounted for by individual differences in speech rate (see the next section).

In addition, there may be a role for the position of the word in the prosodic phrase. The speakers varied in terms of where in the prosodic phrase their tokens of *eigenlijk* tended to occur. Speakers E, F, and J produced over half of their tokens at phrase edges, whereas all others produced the majority of their tokens phrase-medially. This variation in prosodic position probably partly explains the interspeaker variation in reduction degree because word tokens at prosodic boundaries tend to be less reduced (e.g. Bell et al. 2003).

At the level of phonetic detail, while most of the patterns observed were common to more than one speaker, there were certain production strategies that appeared to be specific to one or just a few speakers. For instance, speakers F and L were the only ones in the dataset who produced overlapping velar frication and /l/-quality. They contrast with speaker N, for instance, who, often devoiced the /l/. Furthermore, impressionistically, some speakers were very variable in their realisations of *eigenlijk* whereas others were more consistent. Further research has to investigate to what extent physiological differences between the speakers may explain these individual reduction patterns.

5. Analysis of the conditions under which some properties appear

5.1 Predictors

We tested whether several properties of the tokens may be conditioned by the following five types of predictors. First, we investigated the influence of the predictability of the preceding and following word, since both have been shown to correlate with the duration and the number of segments of *eigenlijk* (Pluymaekers, Ernestus, and Baayen 2005b). We defined these predictabilities as the logarithms of the numbers of occurrences of these words plus one in the Spoken Dutch Corpus (ranges for both words: 0 - 12.16).

Secondly, we studied the influence of a temporal measure. As mentioned in the Introduction, many studies have shown that a higher speech rate generally favors a higher reduction degree. Pluymaekers, Ernestus, and Baayen (2005b) showed that this also holds for *eigenlijk*. We defined speech rate as the logarithm of the number of syllables per second in the citation forms of the words in the labeled phrase (mean: 7.7 syllables/second; range: 1.3 - 20.5 syllables/second). We used the number of syllables of the citation form because it is an important predictor of perceived rate (Koreman 2006) though the realized number of syllables also plays a role. We applied a logarithmic transformation because an increase of one syllable per second is likely to have a bigger impact if the rate is one syllable per second than if it is seven syllables per second.

Thirdly, we included prosodic measures: the level of accent on *eigenlijk* and its position in the phrase. We started with regression models in which both predictors had four levels (primary accented, secondary accented, stressed, and unstressed; isolation, phrase-initial position, phrase-final position, phrase-medial position), but models with these predictors did not converge. We therefore simplified these predictors to two-level predictors (accented or not; in phrase-medial position or not).

Further, we included as prosodic measures the number of unstressed syllables preceding *eigenlijk* and the number of unstressed syllables following *eigenlijk* (both ranges: 0 - 3). They provide information about the rhythm of the phrase. We compared these two continuous measures with two predictors that merely indicate whether the preceding/following syllable is unstressed. The tables with the statistical results show the models with these categorical variables. If these models differ significantly from the models with the continuous variables, this is mentioned in the text.

Fourthly, we investigated the roles of the types of preceding and following segments, since through co-articulation they may directly affect the realization of the neighboring segments. These variables, however, never played statistically significant roles.

Finally, we also tested whether the number of syllables could predict the other properties of the tokens. Because we feel uncomfortable modelling one dependent variable with another one, we only report the results of these analyses in the text (i.e. without details in tables). For the same reasons, we did not incorporate in the main analysis vowel duration as a predictor for creaky voice.

5.2 Statistical analyses

We analyzed the continuous dependent variables with linear mixed effects modeling, and the Boolean dependent variables with generalized linear mixed effects modeling with the logit link function, as implemented in the statistical package R (R Core team, 2014). We tested for random intercepts for speaker, preceding word and following word. We did not include random slopes because preliminary testing suggested that models including them did not converge or seemed to overfit the data.

Our final models, reported below, only include statistically significant predictors. Fixed predictors were considered significant if their absolute t-values or z-values were greater than 1.96 (which approximates an alpha level of 0.05). Random intercepts were considered significant if the model with the random intercept outperformed the model without that random intercept as indicated by likelihood ratio tests (again we adopted an alpha level of 0.05). For continuous dependent variables, the final models are only based on those data points that differed less than 2.5 standard deviations from the values predicted by the model.

5.3 Results

We first studied which variables predict the number of syllables of *eigenlijk*, by analyzing two dependent Boolean variables: whether the word contains one syllable or whether the word contains three syllables. Trisyllabic tokens are relatively rare in the dataset (only 20 tokens) and it is therefore not surprising that there are fewer predictors of the occurrence of trisyllabic than of monosyllabic tokens (see Table 3).

We found that monosyllabic tokens are more likely at a higher speech rate and when the preceding word is of a higher frequency of occurrence. In addition, polysyllabic tokens are followed by stressed syllables in 71% of cases, whereas monosyllabic tokens are approximately equally often followed by unstressed and stressed syllables (51% versus 49%). This suggests a relatively strong tendency for *eigenlijk* to be monosyllabic when followed by an unstressed syllable. The categorical variable, which only provides information about whether the next syllable is stressed, results in a model that is as good as the model with the continuous variable indicating the exact number of following unstressed syllables (the two variables results in models with similar Akaike information criterion values: 152 versus 151, Akaike 1974).

Trisyllabic tokens mostly occurred at lower speech rates and at phrase boundaries. Out of the 20 trisyllabic tokens, only five occurred in phrase-medial position (20%), whereas no less than 105 out of the 139 mono- and disyllabic tokens (76%) occurred phrase-medially.

As expected given our previous observations on differences between speakers (see section 4.3), speaker is an important random effect. In addition, the probability of a monosyllabic token was conditioned by the identity of the following word. We also found random effects of speaker and following word in the analyses presented below, and we will no longer mention them separately.

Table 3: Statistical results for the number of syllables. A positive coefficient implies that the predictor increases the probability of one/three syllable(s).

Fixed effects	Exactly one syllable		Exactly three syllables	
	Coefficient	z-value	Coefficient	z-value
Intercept	-11.963	-3.48	3.597	1.81
Speech rate	4.192	2.81	-2.524	-2.53
Preceding word frequency	0.358	2.37	-	-
No following unstressed syllables	-1.996	-2.00	-	-
Phrase-medial position	-	-	-2.311	-3.49
Random effect	Variance	SD	Variance	SD
Speaker	6.965	2.64	1.712	1.31
Following word	4.646	2.16	-	-

We then analyzed the duration of the whole word token, and of the most frequently occurring parts. Table 4 shows the statistical results for the duration of the token as a whole and that of its first, full, vowel. As expected, both units are shorter at a higher speech rate and if non-accented (mean duration of non-accented tokens: 267 ms; mean duration of accented tokens: 294 ms). The complete token is also shorter when in phrase-medial position (mean duration: 255 ms) than at phrase boundaries (mean duration: 337 ms).

The duration of the whole token is in addition affected by the rhythm of the phrase: tokens tend to be shorter (mean of 259 ms versus a mean of 293 ms) if they are followed by (any number of) unstressed syllables. This result is in line with the results for the probability of a monosyllabic variant, presented above. This effect of the rhythm of the phrase only surfaces in the analysis of token duration if we test the categorical variable. A variable that exactly indicates how many unstressed syllables are following is not predictive of the duration of the token, and results in a model with a slightly higher Akaike Information Criterion value (1777 versus 1775).

The number of syllables is a good predictor for the durations of both the complete token and the full vowel. The statistical results for vowel duration hardly change if the number of syllables is incorporated as an additional predictor. This is different for the token duration. The effect of the rhythm of the phrase on the duration of the token is no longer significant, which is not surprising because the number of syllables of *eigenlijk* and the rhythm of the phrase are correlated (see above). In addition, the duration of the token is no longer predicted by whether the token is accented or not. This may be more surprising because accentedness does not predict number of syllables in our analyses presented above. Possibly, these analyses were not sufficiently sensitive since we investigated two Boolean dependent variables (monosyllabic or not; trisyllabic or not).

Table 4: Statistical results for word and full vowel durations

Predictor	Word duration		Vowel duration	
	Coefficient	t-value	Coefficient	t-value
Intercept	525.115	12.68	229.666	10.47
Speech rate	-112.823	-5.66	-41.141	-3.89
Phrase-medial position	-35.455	-2.86	-	-
Non-accented	-24.350	-2.67	-18.650	-3.79
No following unstressed syllable	24.603	2.37	-	-
Random effect	Variance	SD	Variance	SD
Speaker	512.9	22.65	109.7	10.48
Following word	1606.0	40.08	-	-

Table 5 shows the statistical results for the duration of the velar fricative, if it was present (133 tokens). The fricative was shorter at higher speech rates and when in phrase-medial position (mean duration: 61 ms) rather than at a phrase boundary (mean duration: 68 ms). These variables also predicted token or vowel durations, and in the same directions (see above). The token's number of syllables does not predict the duration of the fricative.

If we test the effect of the rhythm of the phrase on the duration of the fricative with the categorical variables, we found no effects (as indicated in Table 5). We find an effect, in contrast, if we test the continuous variable indicating the exact number of following unstressed syllables (coefficient: 6.705; t-value: 2.58). Surprisingly, the fricative tends to be longer if *eigenlijk* is followed by a higher number of following unstressed syllables. The absence of an effect of the categorical variable and the unexpected direction of the effect of the continuous variable (which is also opposite to what we found for the probability of a monosyllabic form and for the duration of the complete token) raises the question whether this effect on the fricative duration may be a Type 1 error or just arises because the number of following unstressed syllables happens to be correlated with some other relevant predictor.

Each of the other segments (i.e. the segments in the suffix -/lək/) was too often absent to allow for an analysis of its duration. We therefore analyzed the duration of the suffix as a whole, in the 140 tokens in which at least one of its segments was present and in which the final segment was not an unreleased stop followed by silence (see section 3). The results are presented in the last two columns of Table 5. The suffix is shorter at higher speech rates and if in phrase-medial position rather than at phrase boundaries (mean durations: 83 ms versus 135 ms). The effect of speech rate disappears if the number of syllables is taken into account. As expected, the suffix is longer if the token contains more syllables and the suffix is thus present and does not only consist of a velar consonant.

Table 5: Statistical results for the durations of the velar fricative and the suffix -/lək/

Predictor	Fricative duration		Suffix duration	
	Coefficient	t-value	Coefficient	t-value
Intercept	100.916	8.18	192.672	7.98
Speech rate	-17.372	-2.86	-33.653	-2.83
Phrase-medial position	-7.867	-2.34	-39.728	-5.18
Random effect	Variance	SD	Variance	SD
Speaker	18.01	4.24	159.2	12.62
Following word	-	-	474.6	21.79

From all segments, only /l/ was both present in many tokens (88) and absent in many tokens (71). This was therefore the only segment for which we could analyse which variables predicted its presence. The results are presented in Table 6. The segment /l/ was less often present at higher speech rates and when located in phrase-medial position (47% of tokens) rather than at a phrase boundary (84%). Moreover, /l/ was more often realized if the *eigenlijk* token was not followed by unstressed syllables (63%) than if one or more unstressed syllables followed (39%). This latter effect of rhythm only surfaces if we test the categorical rather than the continuous variable indicating the exact number of following unstressed syllables.

All effects disappear if the number of syllables in the spoken *eigenlijk* token is taken into account. The strong effect of the number of syllables is unsurprising: /l/ is seldom present in monosyllabic tokens, quite often present in disyllabic tokens, and always present in trisyllabic tokens (see Table 2). The number of syllables is predicted by the same variables as the presence of /l/ (see Table 3).

Table 6: Statistical results for the presence of /l/ and of creaky voice

Predictor	/l/		Creaky voice	
	Coefficient	z-value	Coefficient	z-value
Intercept	6.1744	-2.76	3.098	1.87
Speech rate	-2.6348	-2.49	-1.715	-2.16
Phrase-medial position	-1.8897	-3.29	-0.920	-2.10
No following unstressed syllable	1.1174	2.22		
Following word frequency	-	-	0.128	2.23
Random effect	Variance	SD	Variance	SD
Speaker	2.746	1.66	0.968	0.98

Finally, we analyzed creaky voice, a variable that has not been analyzed so far as a measure of degree of reduction. Creaky voice was present in 83 tokens. We modeled the probability that creaky voice was present as well as the duration of creaky voice if present. The results for the presence of creaky voice are presented in the last two columns of Table 6. Creaky voice was less often present at higher speech rates and when the token was in phrase-medial position (in 52% of phrase-medial tokens) rather than in phrase-initial or -final position (in 63% of these tokens). Furthermore, there was a correlation with the frequency of the following word: a more predictable following word appears to increase the likelihood of creaky voice.

The presence of creaky voice cannot be predicted by the number of syllables of the token of *eigenlijk*. In contrast, an additional analysis established that the presence of creaky voice can be predicted by the duration of the vowel. (Recall that we did not include vowel duration in the main analysis for creaky voice because we did not want to model one dependent variable with another one; see 5.1 above.) If vowel duration is incorporated as a predictor in the model for the presence of creaky voice, the effect of the following word frequency is still statistically significant, while those of speech rate and of phrase medial position are only marginally significant ($ps < 0.07$).

The duration of creaky voice, if present, could not be predicted by any of our independent variables. We could only find a high correlation with vowel duration (coefficient: 0.228; $t = 2.02$), showing that the longer the vowel, the longer the part with creaky voice tends to be. In addition, we found a random effect of speaker (variance: 281; S.D. = 16.76).

6. General Discussion

This chapter has presented a detailed analysis of the properties of 159 tokens of the Dutch discourse marker *eigenlijk*, which occur in the Ernestus Corpus of Spontaneous Dutch (Ernestus 2000). Our aim was to document the wide variation in the pronunciation of the word and to analyze which properties of the context predict this variation. Previous research suggests that the exact meaning of a token has an influence on its detailed phonetic characteristics (Plug 2005). We did not distinguish between the different meanings because this would have resulted in too small a data set for statistical analyses. Moreover, the meaning of the word only seems to favor some pronunciation variants rather than completely excluding others. As such, it would have only been one of our many predictors. It would be worthwhile investigating the role of the word's exact meaning in a larger data set.

The qualitative analysis of the tokens in our dataset supported previous studies (e.g. Ernestus 2000) in demonstrating a wide range of variation in the production of the word, ranging from trisyllabic tokens closely resembling the word’s citation form, through to phonetically minimal monosyllabic tokens consisting merely of a vowel followed by a single obstruent consonant. The disyllabic forms occur most frequently (52% of tokens), followed by the monosyllabic variants (36%), while trisyllabic forms are relatively rare (13%).

Reduction of *eigenlijk* may be manifest in a number of ways, in the number of syllables that a token contains, or in its duration, or the clarity of its articulation. We expected these properties to correlate with one another. In fact, we found surprisingly little clear correlation between the different indices of reduction. Although the number of syllables in a token did increase along with duration, the durational ranges found in mono-, di- and tri-syllabic tokens clearly overlapped. Moreover, some tokens that had only one syllable nonetheless had clearly articulated and tightly coordinated segments, while some di- and trisyllabic tokens appeared to be articulated rather laxly. These observations underscore that reduction is not a simple or automatic consequence of speaking under time pressure.

We found some support for the concept of “articulatory prosodies” in the sense of non-linear features of syllables that are no longer tied to specific segmental units (Niebuhr and Kohler, 2011). In particular, the /l/ of *eigenlijk* is not always present, but when a definite lateral articulation is absent, acoustic and auditory traces of /l/ often remain. Also, the weak syllables of the word may be absent yet leave an acoustic residue in the signal, for instance, as extra duration of the consonants in monosyllabic tokens produced as /ɛɪxk/. Following Niebuhr and Kohler’s (2011) reasoning further, can we specify a “phonetic essence” of the word *eigenlijk?* Apparently, the only essential components are a front, usually diphthongal vowel and at least one back (velar or uvular) obstruent. Even these essential parts allow for variation: the vowel can lose its diphthongal quality and can become somewhat more backed; the obstruent can be either stop or fricative, and given an appropriate conditioning context, it can lose its voicelessness.

We observed unexpected patterns of voice assimilation within the word and at word boundaries that are not described in the literature. We found both regressive and progressive voice assimilation of /l/ within the word, and of the final voiceless obstruents of *eigenlijk* to following nasal consonants. Local (2003) proposes that different patterns of assimilation occur for function words (e.g. *I’m*) compared to content words (e.g. *lime*). Our data suggest that the same may be true for discourse markers, and perhaps for frequent sequences such as *eigenlijk niet 'actually not'.* Further work is needed to show whether the observed assimilation patterns are indeed specific for *eigenlijk*.

Also unexpectedly, the dataset showed no clear relationship between a token's prosodic status and its reduction in terms of either duration or number of syllables. We observed a surprisingly large number of accented tokens that were produced with only one syllable (eight primary-accented and 16 secondary-accented monosyllabic tokens). This clearly shows that reduction of *eigenlijk* is not a phenomenon restricted to prosodically weak positions. A token may be heavily reduced in duration or number of syllables, yet may still constitute the most prominent word in its local context. Future studies have to show which other (types of) words can also be drastically reduced in prosodically strong positions.

The dataset contains tokens from 18 speakers from a rather homogeneous group (all highly educated men raised in the Western part of the Netherlands). Nevertheless, there are substantial differences between speakers in their reduction degree. Some speakers are clearly more likely to produce monosyllabic forms of *eigenlijk* than others. Speakers also differ in how they solve the problem of quickly producing a velar/uvular fricative followed by a lateral, for instance by complete coarticulation of the two sounds, resulting in a lateral fricative, or by weakening of the degree of stricture for the fricative. Finally, in our analyses investigating which variables predict the duration and presence versus absence of segments, speaker was always a significant random effect.

These individual patterns confirm and extend the results by Hanique, Ernestus and Boves (2015). By means of computational modeling of automatically generated segmental transcriptions of the speech in the entire corpus, these researchers showed that the speakers in our data set can be better distinguished from each other if not only the word types and combinations of word types that they produced are taken into account but also how these speakers reduced phones and combinations of phones. Our study contributes to the results by Hanique and colleagues by documenting interspeaker variation in the pronunciation at the segmental level of one entire word (instead of a single phone or a sequence of three phones), which forms part of the information the computer modelling of Hanqiue and colleagues was based on. On top of this, our results show that the speakers differ at the subsegmental level, which was not taken into account by Hanique and colleagues.

In the second part of the chapter, we investigated which variables may predict the number of syllables of a token, the durations of a token and its parts, and the presence of /l/ and creaky voice. In the Introduction to this chapter, we hypothesized that the rhythm of the sentence may have an effect on reduction degree of *eigenlijk*. Speakers of Germanic languages prefer sentences with approximately equally long intervals between stressed syllables. In order to minimize the sequence of unstressed syllables, they may realize tokens of *eigenlijk* followed by unstressed syllables as (stressed) monosyllabic variants. Conversely, in order to avoid stress clashes, they may prefer a di- or tri-syllabic variant, ending in one or two unstressed syllables, respectively, when the word token is followed by a word with initial stress. Our data set provides support for this hypothesis. Tokens of *eigenlijk* appear to be more often monosyllabic, to be shorter in duration, and to be less likely to contain /l/s when followed by unstressed syllables. To our knowledge, effects of rhythm on speech reduction have not been documented before.

Interestingly, the categorical variable that just indicates whether the token of *eigenlijk* was followed by either a stressed or an unstressed syllable outperformed a continuous variable indicating the exact number of following unstressed syllables for token duration and for presence of /l/. (For the probability of a monosyllabic form, the categorical and the continuous variables are equally predictive). This suggests that only the prosodic status of the immediately following syllable is relevant. Possibly, when producing *eigenlijk,* speakers have only taken decisions about the prosodic status of the next syllable. Another possible explanation is that speakers cannot or are not inclined to reduce *eigenlijk* even more when it is followed by more than one unstressed syllable.

Unexpectedly, it is the exact number of following unstressed syllables rather than the presence of a following unstressed syllable that predicts the duration of the fricative. Moreover, this effect goes in the non-hypothesized direction: this fricative was longer if it was followed by more unstressed syllables. We do not know how to explain this unexpected result.

Of the fixed predictors that have been shown before to correlate with reduction degree, two emerged as statistically significant in (nearly) all analyses: tokens were more reduced at higher speech rates and in phrase-medial position. These results replicate previous findings (e.g. Bell et al. 2003; Kohler 1990, 2009; Raymond, Dautricourt, and Hume 2006). We found these effects also for the probability of creaky voice, that is, creaky voice was less likely to occur at higher speech rates and phrase-medially.

Two analyses showed effects of the predictability of a neighboring word. The word *eigenlijk* was more likely to be monosyllabic if it was preceded by a more frequent word. This effect is in line with earlier findings by Pluymaekers, Ernestus and Baayen (2005b) for this same word. In addition, we found an effect of the frequency of occurrence of the following word on the likelihood of creaky voice in the full vowel: creaky voice was more often present when the word token was followed by words of higher frequencies.

These results with respect to creaky voice are interesting. We see that creaky voice is more often absent under the same conditions where segments tend to be absent and shorter (i.e. at higher speech rates and in phrase-medial position). This suggests that the absence of creaky voice results from the same mechanisms that also reduce segments. This hypothesis, however, does not fit with our observation that creaky voice tends to be more often present if the following word is of a higher frequency: creaky voice behaves differently in this respect from segments, which tend to be more, rather than less, reduced before high frequency words. We propose that these conflicting results are the consequence of the ambiguous character of creaky voice. On the one hand, creaky voice at the start of a vowel-initial word may function as a phonetic cue to prosodic strength (cf. Jongenburger & van Heuven 1991), which tends to be reduced in the same conditions where segments are reduced. On the other hand, separate from this function in marking vowel-initial word onsets, the presence of creaky voice may in some cases result from reduced articulatory effort in voicing (cf. Gobl and Ní Chasaide 2003). In that case, we expect the presence of creaky voice in those contexts where segments tend to be reduced, including before a high frequency word. Further research is needed to test this hypothesis.

Unlike previous studies (e.g. Bell et al. 2003; Pluymaekers, Ernestus, and Baayen 2005b; Bell et al. 2009), we did not find an effect of the frequency of the following word on the presence versus absence of segments or on phone, affix, or token durations. A possible explanation is that the previous studies did not incorporate following word as a random variable. This hypothesis is supported by an analysis showing that the likelihood of a monosyllabic token is predicted by the frequency of the following word, if following word is not incorporated as a random effect as well. The random effect of the following word was significant in half of our analyses.

These observations may give some clues as to what may be stored in the mental lexicon. On the one hand, the high frequency of highly reduced forms and the existence of consistent phonetic patterns in their production encourages the conclusion that multiple variants are stored as separate pronunciation targets. On the other hand, such variants need to retain some link to the word's canonical form in order to account for the presence of articulatory residues in the phonetic detail of the forms produced. Put differently, a reduced token of *eigenlijk*, even if broadly transcribable as [ɛɪk], probably often differs from the same phoneme string produced in the content word *eik* 'oak', as other authors have demonstrated for word pairs like (reduced) *support* versus *sport* (e.g. Manuel 1991; Manuel et al. 1992). To substantiate these suggestions requires further acoustic analysis—in particular of spectral properties which we could not address in this paper.

Furthermore, our data show that models of speech production have to take the rhythm in the phrase into account when explaining the reduction degree of *eigenlijk*. This strongly suggests that the assumption that degree of reduction is only determined by how much time a speaker needs to plan and produce the word or the following word is too simplistic: the speaker also appears to be concerned with at least the rhythm of the phrase.

From a perceptual point of view, the data emphasize that the speech comprehension system has to deal with a great deal of variation when it comes to recognizing *eigenlijk*. One aspect that may help the listener is that several core properties of *eigenlijk* are high in acoustic salience: the formant dynamics characteristic of the diphthong /ɛɪ/, the compact mid-frequency spectral prominences that characterize velar and uvular obstruents, and the strident nature of uvular fricatives. These landmarks may guide the listener, enabling the detection of finer details that cue the word's identity, such as traces of /l/.

Perceptual experimentation is needed to test the role of the various gross and subtle acoustic characteristics that we have identified.

7. Rethinking reduction

The research reported on in this chapter has generated data that may cast doubt on some of the common assumptions about the phenomenon of speech reduction. We showed that the variation in the pronunciation of discourse markers may be substantial. The unreduced variant may be less common than reduced variants (in our case only 20% of the tokens are unreduced). This raises the question which form of such a word should be considered as canonical: the full form, which is represented in orthography, or the most frequently occurring reduced form.

Furthermore, our data show that, in contrast to what is generally assumed, highly reduced discourse markers can occur in prosodically strong positions. Reduction is therefore not for all words restricted to unaccented positions. The occurrence of highly reduced forms in accented positions underlines that reduced forms of at least some words are not special and can occur without restrictions.

The pronunciation variation displayed by *eigenlijk* is conditioned, among other factors, by the rhythm of the phrase, and shows large differences between speakers. Moreover, a form may be reduced in one aspect, but not in another. This strongly suggests that reduction is not a fully automatic process that arises when speakers are under time pressure. Speakers clearly have a choice whether to reduce and how to reduce, and they make this choice, among others, on the basis of the rhythm of the phrase, while adhering to their own speech habits.

Finally, we found that every pronunciation variant of *eigenlijk* appears to include two landmarks that may be considered to be the main characteristics of the word (the full vowel and a velar/uvular consonant). These landmarks raise questions about speech processing. Are these landmarks indicated in the mental lexicon? How do listeners use these landmarks during word recognition?

We conclude that this corpus study has shown that many aspects of the phenomenon of speech reduction are not yet well understood. We call for more detailed qualitative and quantitative analyses of many tokens of individual words produced in casual speech because these studies substantially extend our knowledge about speech reduction and about speech production and perception.

Acknowledgements

We would like to thank Harald Baayen for facilitating the second author's stay at the Max Planck Institute for Psycholinguistics, and Brechtje Post for her advice on the prosodic analysis of our phrases. Furthermore, we would like to thank Sarah Hawkins, Ellen Aalders and three anonymous reviewers for their helpful comments on an earlier version of the paper. This research was partly funded by a vici grant from the Netherlands Organization for Scientific Research.

References

Aalders, Ellen, Mirjam Ernestus. In preparation. Is there a link between schwa reduction and the realization of following /p, t/ in casual speech?

Akaike, Htrotugu. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19, 716–723.

Aoyagi, Maki. In press. Japanese vowel devoicing and voicing control in the C/D model. *Journal of the Phonetic society of Japan*.

Bell, Alan, Jason. M. Brenier, Michelle Gregory, Cynthia Girand & Dan Jurafsky. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60. 92-111.

Bell, Alan, Daniel Jurafsky, Eric Fosler-Lussier, Cynthia Girand, Michelle Gregory & Daniel Gidea. 2003. Effects of disfluencies, predictability and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America* 113(2). 1001-1024.

Bergen, Geertje van, Rik van Gijn, Lotte Hogeweg & Sander Lestrade. 2011. Discourse marking and the subtle art of mind-reading: The case of Dutch eigenlijk. *Journal of Pragmatics* 43. 3877-3892.

Booij, Geert. 1995. *The phonology of Dutch*. Oxford: Clarendon Press.

Cruttenden, Alan. 1994. *Gimson's Pronunciation of English* (5th edition). London: Edward Arnold.

Davidson, Lisa. 2006. Schwa elision in fast speech: segmental deletion or gestural overlap? *Phonetica* 63. 79-112.

Ernestus, Mirjam. 2000. *Voice assimilation and segment reduction in casual Dutch, a corpus-based study of the phonology-phonetics interface*. Utrecht: Netherlands National Graduate School of Linguistics.

Ernestus, Mirjam & Natasha Warner. 2011. An introduction to reduced pronunciation variants. *Journal of Phonetics* 39. 253-260.

Gahl, Susanne. 2008. "Thyme" and "Time" are not homophones. Word durations in spontaneous speech. *Language* 84. 474-496.

Gahl, Susanne, Yao Yao & Keith Johnson. 2012. Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language* 66. 789-806.

Gobl, Christer & Ní Chasaide, Ailbhe. 2003. The role of voice quality in communicating emotion, mood and attitude. *Speech Communication* 40. 189-212.

Gordon, Matthew, Paul Barthmaier & Kathy Sands. 2002. A cross-linguistic acoustic study of fricatives. *Journal of the International Phonetic Association* 32. 141-174.

Guy, Gregory R. 1991. Explanation in variable phonology: An exponential model of morphological constraints. *Language Variation and Change* 3. 1-32.

Hanique, Iris, Mirjam Ernestus & Lou Boves. 2015. Choice and pronunciation of words: Individual differences within a homogeneous group of speakers. *Corpus Linguistics and Linguistic Theory* 11. 161-185.

Johnson, Keith. 2004. Massive reduction in conversational American English. In Kiyoko Yoneyama & Kikuo Maekawa (eds.), *Spontaneous speech: Data and*

analysis. Proceedings of the 1st session of the 10th international symposium, 29-54*.* Tokyo: The National International Institute for Japanese Language.
Jongenburger, Willy & Vincent van Heuven. 1991. The distribution of (word-inital) glottal stop in Dutch. In Ans van Kemenade & Frank Drijkoningen (eds.), *Linguistics in the Netherlands*, 101-110. Amsterdam: John Benjamins Publishing Company.
Kelly, Michael H. & J. Kathryn Bock. 1988. Stress in time. *Journal of Experimental Psychology: Human Perception and Performance* 14. 389-403.
Keune, Karen, Mirjam Ernestus, Roeland van Hout & R. Harald Baayen. 2005. Social, geographical, and register variation in Dutch: From written mogelijk to spoken mok. *Corpus Linguistics and Linguistic Theory* 1. 183-223.
Kohler, Klaus J. 1990. Segmental reduction in connected speech in German: Phonological facts and phonetic explanations. In William J. Hardcastle & Alain Marchal (eds.), *Speech production and speech modelling*, 21-33. Dordrecht: Kluwer Academic Publishers.
Kohler, Klaus J. 1999. Articulatory prosodies in German reduced speech. *Proceedings of the XIVth International Congress of Phonetic Sciences* 1. 89-92.
Koreman, Jacques. 2006. Perceived speech rate: the effects of articulation rate and speaking style in spontaneous speech. *Journal of the Acoustical Society of America* 119*.* 582-596.
Labov, William L. 2001. *Principles of linguistic change: Social factors*. Oxford: Blackwell Publishers.
Local, John. 2003. Variable domains and variable relevance: Interpreting phonetic exponents. *Journal of Phonetics* 31(3). 321-339.
Manuel, Sharon Y. 1991. Recovery of deleted schwa. *Perilus: Papers from the Symposium on Current Phonetic Research Paradigms for Speech Motor Control*, University of Stockholm Institute of Linguistics. 115-118.
Manuel, Sharon Y. 1995. Speakers nasalize /ə/ after /n/, but listeners still hear /ə/. *Journal of Phonetics* 23(4). 453-476.
Manuel, Sharon Y., Stefanie Shattuck-Hufnagel, Marie K. Huffman, Kenneth N. Stevens, Rolf Carlson & Sheri Hunnicutt. 1992. Studies of vowel and consonant reduction. *Proceedings of the Second International Conference on Spoken Language Processing*. 943-946.
Niebuhr, Oliver & Klaus J. Kohler. 2011. Perception of phonetic detail in the identification of highly reduced words. *Journal of Phonetics* 39(3). 319-329.
Oostdijk, Nelleke. 2000. The Spoken Dutch Corpus project. The ELRA newsletter 5(2). 4-8.
Phillips, Betty S. (1994). Southern English glide deletion revisited. *American Speech* 69. 15-127.
Plug, Leendert. 2005. From words to actions: the phonetics of eigenlijk in two communicative contexts. *Phonetica* 62. 131-145.
Pluymaekers, Mark, Mirjam Ernestus & R. Harald Baayen. 2005a. Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America* 118. 2561 - 2569.

Pluymaekers, Mark, Mirjam Ernestus & R. Harald Baayen. 2005b. Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica* 62. 146-159.

R Core Team. 2014. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing.

Raymond, William D., Robin Dautricourt & Elizabeth Hume. 2006. Word-internal /t,d/ deletion in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors*. Language Variation and Change* 18. 55-97.

Son, Rob J. J. H. van & Louis C. W. Pols. 1990. Formant frequencies of Dutch vowels in a text, read at normal and fast rate. *Journal of the Acoustical Society of America* 88. 1683-1693.

Son, Rob J. J. H. van & Louis C. W. Pols. 1992. Formant movements of Dutch vowels in a text, read at normal and fast rate. *Journal of the Acoustical Society of America* 92. 121-127.

Strik, Helme, Joost van Doremalen & Catia Cucchiarini. 2008. Pronunciation reduction: how it relates to speech style, gender, and age. *Interspeech*, 1477-1480.

Torreira, Francisco & Mirjam Ernestus. 2011. Vowel elision in casual French: the case of vowel /e/ in the word c'était. *Journal of Phonetics* 39, 50-58.

Warner, Natasha & Benjamin V. Tucker. 2011. Phonetic variability of stops and flaps in spontaneous and careful speech. *The Journal of the Acoustic Society of America* 130(3). 1606-1617.

Appendix

Variation dicussed in sections 4 and 5 for which we can provide the number of tokens. The first two columns provide information about segmental variation, the second two columns about subsegmental variation

Segmental variation	Number of tokens (out of 159)	Subsegmental variation	Number of tokens (out of 159)
Initial vowel is monoph-tongal	6	Creaky voice for first vowel	83
First schwa is absent	139	Voicing of velar stop	25
Velar fricative is absent	22	At least partly devoiced /l/	21
Second schwa is absent	58	At least partly devoiced schwa	7
Velar stop is absent	19		