UNIVERSITEIT VAN AMSTERDAM

Master's Thesis

Segmentation of reduced speech by native and non-native listeners of Dutch

Netta Meijer

Supervisors	Dr. S.J. Andringa & Dr. N.M. Olsthoorn
Second assessor	Prof. dr. J.H. Hulstijn
Date	August 2008
Institute	General Linguistics
	Faculty of Humanities
	Universiteit van Amsterdam

Acknowledgements

I would like to thank all the participants, first and foremost. At times, I was quite desperate when it came to finding an adequate number of willing subjects. Many thanks for the kindness of strangers, acquaintances and friends and their willingness to help.

Thank you to my supervisors, Sible Andringa and Nomi Olsthoorn for their trust in me and in my contribution to their project. I am also very grateful to Jan Hulstijn for introducing me to my supervisors, to the STILIS-project and to second language acquisition in the first place.

Many friends had a positive influence on this thesis and I am eternally grateful for their support. I thank Emmie for the many *scriptiedates* that were not always as productive as we hoped, but nevertheless very encouraging and enjoyable. Sterre, thank you for helping me with the experiment, and for so much more. And many many thanks to *De Acht* for letting me be a part of this exceptional friendship.

Finally, I thank my parents, Liad, and Kerren for their love and support. And for their belief in me, even when I did not fully believe in myself.

Abstract

When perceiving auditory input, listeners are subjected to a continuous speech signal. From that signal they have to retrieve word boundaries in order to extract discrete meaningful units, i.e. words. This segmentation process is not the only problem listeners are faced with. In casual speech, segments are oftentimes reduced – sometimes to an extreme extent – which leads to large variations in the realization of speech sounds.

In this thesis the effect of reduction processes on speech segmentation was studied in both native and non-native listening. The research question was whether there is a difference between native and non-native speakers of Dutch in the segmentation of reduced and unreduced Dutch forms. For this purpose, two segmentation tasks were constructed. Participants were presented with full and reduced auditory input and had to indicate how many, as well as which words they heard. Both speed and accuracy of segmentation were measured. Three types of reductions (i.e. vowel, consonant, extreme reduction) were incorporated in the segmentation tasks. I could thus also address the question whether there is a particular type of reduction that is difficult to segment.

15 native and 14 non-native speakers of Dutch took part in the experiment. Native speakers were faster and more accurate than non-native speakers in segmenting unreduced and reduced speech. This may be explained by the inability of non-natives to effectively use Dutch segmentation cues. They may also be less able to use reconstructive processes, as is indicated by a larger effect of reduction than in the native group. A notable finding was the interaction between nativeness and task type. Non-native speakers were better at counting words than at identifying them. This may represent a phase in the – what seems to be a – gradual process in the acquisition of segmentation skills. Analysis of the three reduction types showed that extreme reductions are the most difficult to segment. There was no difference between vowel and consonant reduction.

This study constitutes a stepping stone toward bridging the gap between studies of segmentation and studies of reduction processes.

Table of contents

1 Introduction

Ask any Dutch speaking person what the utterance [ivɑl] means, and he will most probably be dumbfounded. It is, however, a highly frequent phrase in Dutch. The problem is that it is extremely reduced in comparison with its full canonical equivalent. The consequence is that it can become incomprehensible, especially out of context. In everyday conversations it should be fairly simple to recognize the above form as the Dutch phrase *in ieder geval* [ɪn idər xəvɑl] (English, *in any case*, *anyway*). Reduction has been shown to delay comprehension in native speakers of Dutch (Ernestus & Baayen, 2007). The question may arise as to how non-native speakers are affected by reduction processes. How do these listeners perceive reduced words in the L2 speech stream? In this thesis I examine the effect of reduction and deletion processes on how native and non-native speakers of Dutch segment the incoming speech stream.

The basis of this study was the STILIS-project (Studies in Listening), funded by the Netherlands Organization for Scientific Research (NWO) and conducted at the Amsterdam Center for Language and Communication (ACLC) of the University of Amsterdam by Prof. dr. Jan Hulstijn, Dr. Sible Andringa and Dr. Nomi Olsthoorn.[1] The goal of this project is to explore the componential structure of listening proficiency, more specifically individual differences in comprehending oral language. For this purpose 500 people will be tested, consisting of both native and non-native speakers of Dutch with both high and low levels of education and of different age groups. Their performance will be assessed on the listening subset of the *Staatsexamen Nederlands als tweede taal* 1 and 2 and on the segmentation and comprehension of Dutch utterances. Several other variables (working memory, receptive vocabulary and non-verbal reaction speed) will be included to investigate their potentially moderating role. Findings from this project will be used to develop a theory of second language proficiency.

Within this framework, I have focused my attention on segmentation and in particular on the segmentation of reduced speech. Speech segmentation in general

[1] See also http://www.hum.uva.nl/Stilis.

has been frequently studied, in both native (see, e.g., Cutler, 2002; Cutler & Clifton, 1999; Nakatani & Schaffer, 1978; Van der Lugt, 1999) and non-native listening (see, e.g., Altenberg, 2005; Carroll, 2004). The effects of casual speech phenomena – such as reduced speech – on word recognition have not been studied until recently (Ernestus, Baayen, & Schreuder, 2002). These studies all direct their attention to a specific form of reduction, e.g. vowel reduction (Coussé, Gillis, & Kloots, 2007; Kloots, De Schutter, Gillis, & Swerts, 2003), consonant reduction (Mitterer & Ernestus, 2006), prefix reduction (Ernestus & Baayen, 2007; Pluymaekers, Ernestus, & Baayen, 2005), suffix reduction (Kemps, Ernestus, Schreuder, & Baayen, 2004; Keune, Ernestus, Van Hout, & Baayen, 2005), or extreme reduction (Ernestus et al., 2002; Johnson, 2004).

To my knowledge, segmentation has never been studied in relation to reduction. Research on reduction mostly involved presenting only the reduced word as the stimulus, thus not investigating its recognition within a larger speech stream. In this thesis I will do just this. Moreover, I will incorporate different types of reduction, thereby enabling a possible classification into degree of difficulty; in other words, the degree to which the different types of reduction hinder comprehension. All of this will be studied in the context of native versus non-native listening. The research questions of this thesis are:

1a *Is there a difference between native and non-native speakers of Dutch in the segmentation of reduced and unreduced forms?*

1b *If so, is that a difference in speed or in accuracy, or in both?*

2a *Is there a particular type of reduction that is difficult to segment?*

2b *If so, is that comparable for native and non-native speakers?*

The structure of this thesis is as follows. In Chapter 2 the process of speech segmentation will be introduced, followed by background information on reduction in Chapter 3. In Chapter 4 I will give an overview of reduction in Dutch. Chapter 5 will describe the empirical study, along with the research findings in Chapter 6. I will end with a discussion of the findings (Chapter 7) and the conclusion of my study (Chapter 8).

2 Segmentation

When perceiving a language, listeners are subjected to auditory input from which they have to extract meaningful units, i.e. words. The acoustic signal itself is not inherently discrete, as, for instance, written language is. The writing system consists of distinct units (e.g. letters, words, sentences) that are clearly separated from one another. Spoken language, however, is continuous and consists of connected units. It is the task of the listener to segment the incoming stream of sounds, so that word boundaries become evident. What further hinders the word recognition process is the large variation in the realization of speech sounds, reduction being one of the processes causing this variation. The problem of segmentation and the problem of variation have to be overcome (Van der Lugt, 1999). Only then the listener is able to recognize the input. Here in Chapter 2 I will discuss the first of these problems. The comprehension difficulties caused by reduction will be addressed in Chapter 3.

2.1 Speech recognition

According to Van der Lugt (1999) speech recognition consists of four levels of processing: (1) general sensory processing, (2) prelexical speech processing, (3) lexical processing, and (4) sentential processing. Cutler and Clifton (1999) also discuss these four stages – although using different phrasing – in listening to spoken language in what they call the *blueprint of the listener*. General sensory processing is the purely acoustical processing of the input, thus not specific to speech. Here speech has to be distinguished from irrelevant noise, which may be present in the signal. Prelexical speech processing is the recognition of words. On this level discrete segments have to be recognized within the continuous speech stream. Lexical processing enables us to map the spoken words onto stored lexical knowledge. In this level of processing, words from the mental lexicon that match the input are selected and all information about them (e.g. semantic, phonological, orthographical information) becomes available. In the final stage of the process – i.e. sentential processing – individual words are integrated to form a complete sentence.

This thesis will focus on the prelexical level of processing, the ability to recognize words in the speech stream. Following Davis (2000), the term *segmentation* will refer to the segmentation of connected speech into lexical units. As mentioned above, the first obstacle for prelexical processing is the identification of word boundaries in the continuous speech stream, i.e. segmentation. McQueen and Cutler (2001) mention in their introductory paper on spoken word access processes that these are often "taken for granted":

> They are not the processes by which a listener interprets the sequence of words in an utterance, but rather the processes by which that sequence of words is derived from the acoustic speech signal. (p. 470)

To the layman the process of segmentation may indeed seem straightforward; if you know the words in the input, you do not hear a continuous stream of sound, but rather a string of separable words. But this simplistic view does not explain how infants learn their first language or how L2 learners segment the continuous speech stream. After all, they do not yet have an extensive mental lexicon at their disposal.

Studies of infant listening have shown that children as young as six months old can distinguish between sound contrasts and at ten months of age they are familiar with the probabilities of phoneme sequences and word structure in their L1 (Cutler, 2002). Here the foundation is laid for future segmentation processes. According to Cutler, these experimental findings along with the fact that more than 90% of the language a baby hears between the age of six and nine months is connected speech (Van de Weijer, 1999), suggest that babies are capable of extracting individual words from continuous speech.

As adults, listeners do have a large vocabulary do draw from, but this does not necessarily make segmentation effortless. In fact, several features of language can make the task rather challenging (Cutler, 2002). Languages consist on average of 30 morphemes and short words are favored over long ones. The consequence of having many short words, all comprised of the same limited set of morphemes, is that many words resemble one another and that short words are often embedded within longer words. Along with the continuity of spoken language, it can sometimes be quite difficult – even for adults – to recognize individual words. However, as will become apparent in Section 2.2, listeners are able to employ a number of effective cues to facilitate the segmentation process.

2.2 Segmentation cues

Retrieving word boundaries from the continuous speech stream is aided by the utilization of segmentation cues. In his doctoral dissertation, Davis (2000) distinguishes three main categories of cues. The first category consists of acoustic markers in the speech stream, such as stress pattern and rhythm of speech. Nakatani and Schaffer (1978) investigated whether knowledge of these prosodic cues contributed to the inference of word boundaries. Listeners heard nonsense phrases, which had the same stress pattern and rhythm as actual English adjective-noun sequences. Because nonsense phrases were used, the speech did not possess any semantic or phonological cues as to where word boundaries should be. Listeners correctly segmented the phrases, corresponding to segmenting an adjective from its noun. From this, the researchers concluded that listeners use prosodic cues when segmenting words.

A second class of segmentation cues is the knowledge listeners have of the statistical or distributional structure of linguistic items. According to Davis (2000), this cue can be applied in several domains, such as phonology or metrical stress. The latter domain has been studied by Cutler and Norris (1988), who proposed the *strong syllable segmentation hypothesis*, in which strong syllables are a cue for segmentation. This theory is based on statistical evidence that far more lexical words in the English vocabulary start with strong syllables. From this the researchers deduced that "a recognizer that started lexical access at strong syllables would actually miss very few word beginnings" (Cutler & Norris, 1988, p. 114). In their experiments, listeners had to detect target English words embedded in nonsense bisyllables (i.e. a word spotting task). This was conducted in two contexts, one across two strong (SS) syllables and the other across a strong syllable followed by a weak one (SW). The researchers observed that listeners were slower to detect the word when it was embedded in two strong syllables than when it was embedded in a strong syllable followed by a weak syllable. Cutler and Norris explain this by positing that the second strong syllable in the SS condition triggered segmentation, thus breaking off the target word in two segments and in doing so, delaying the detection of the target word.

The distributional structure in the domain of phonology also provides listeners with cues to find word boundaries. Phonotactic strategies, i.e. the knowledge of possible sound combinations, have been found to be helpful during segmentation (McQueen, 1998). Since certain phoneme combinations never occur within the same syllable, their presence is likely to trigger segmentation. McQueen tested this hypothesis in a Dutch word spotting task, showing that the legality of phoneme combinations is being verified during word recognition. Phonotactic constraints are thus being used as cues for where words are likely to begin or end.

Finally, identification of lexical items is also considered to prompt segmentation. Van der Lugt (1999) defines this as follows:

> Parsings that leave a residue that cannot be a possible word, for instance a single consonant, are considered much less likely than parsings that leave a bit of speech that could be a word. (p. 8)

This has been introduced as the *Possible Word Constraint* (PWC) (Norris, McQueen, Cutler, & Butterfield, 1997). It has been shown that listeners had more difficulties in identifying the word *apple* in a context like *fapple* – where this would leave the single consonant [f], not being a possible word – than in the context of *vuffapple* – where *vuff* could be an English word. Research has shown that lexical identification is the result of "continuous activation of multiple candidate words, and that there is a process of competition between the activated candidates out of which the eventual winning words emerge" (McQueen & Cutler, 2001, p. 476). The PWC is claimed to regulate this competition process in lexical activation by "reducing the activation of candidate words if their recognition would imply word status for adjacent input which could not be a word" (Norris et al., 1997, p. 191).

As can be seen from the different categories of segmentation cues, the identification of word boundaries can be approached as either a bottom-up or top-down analysis. In a bottom-up approach, comprehension is guided by the smallest elements of the signal; it is driven by the acoustic input itself (e.g. prosodic cues to segmentation). Top-down analysis starts with the listeners' world knowledge and works down to the speech signal (e.g. segmentation through identification of lexical items). Cutler (1996) differentiates between *Explicit Segmentation Models* and *Serendipitous Segmentation Models*. The former use observable features (i.e. bottom-up) in the signal as the basis for segmentation. In the latter, top-down processes, such as lexical selection, make the boundaries emerge implicitly when words are selected from the lexicon. Mattys, White and Melhorn (2005) make the same distinction, using the phrases lexical (word selection) and sublexical (acoustic markers and phonotactic regularities) cues. A vast amount of evidence exists for both accounts of segmentation, but both also seem to have their limitations (ibid., p. 477). Sublexical cues do not explain facilitation through lexical and sentential contributions (i.e. semantic context), while lexically driven segmentation produces an excessive amount of activity – when for instance shorter words are embedded in longer ones, all potential word candidates are activated in the mental lexicon, even though their activity is oftentimes redundant. Lexical cues are also ineffective when words have to be segmented, which are not represented in the mental lexicon. Fortunately, the various heuristics used in segmentation are not mutually exclusive and listeners can utilize all the cues available to them.

Mattys et al. (2005) show how the combined operation of both lexical and sublexical cues can be represented. In a cross-modal identity-priming task, native speakers of English were presented with English sentences over headphones and subsequently a string of letters appeared on a computer monitor in front of them. Participants were instructed to decide as quickly as possible whether the string of letters was a word using a two-button response box (lexical decision task). When the visual stimulus is associated with the auditory prime, priming effects are expected to emerge. Priming effects between the orally and visually presented stimuli were then used as a measure for segmentation; such effects mean that the auditory prime has been accurately segmented. From a series of experiments using the cross-modal identity priming paradigm Mattys et al. concluded that segmentation cues are hierarchically integrated, with lexical cues at the top of the hierarchical structure.

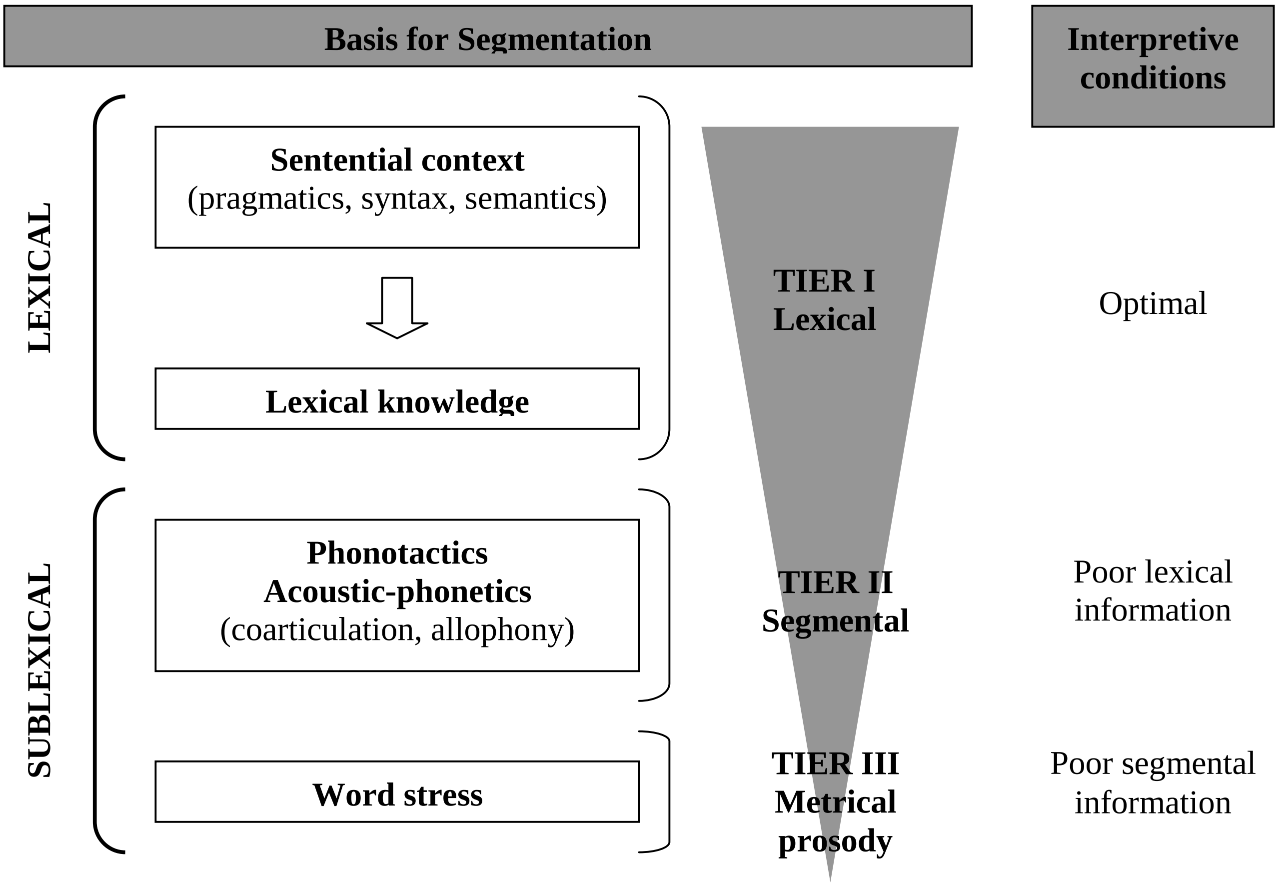

Figure 1 Hierarchy of segmentation strategies (Mattys et al., 2005)

The hierarchical structure of segmentation cues can be seen in Figure 1. Mattys et al. found a ranking order, in which lexical and semantic cues take precedence over sub-lexical:

> When all cues are optimally available, speech segmentation is lexically driven (Tier I), even in the presence of discrepant sublexical cues. Although we do not directly test the concept in the present experiments, we expect that, in

> many instances, the semantic and syntactic content of an utterance contribute to lexically driven segmentation by favoring those words most likely given a particular context. Sublexical cues are called on when lexical information is unavailable, impoverished, or ambiguous. A further distinction is made between segmental information (Tier II) and metrical prosody (Tier III), with the former outweighing the latter. Metrical prosody, or word stress, appears to best induce segmentation when word boundaries cannot be inferred using segmental cues. (Mattys et al., 2005, p. 487)

In conclusion, an integrated approach of both lexical and sublexical cues – to which listeners assign different weights, i.e. sublexical cues being outweighed by lexical cues – seems to be the approach native listeners employ during lexical segmentation.

2.3 Native versus non-native listening

In the previous section I discussed speech segmentation in the context of native listening. But how does the way native speakers segment speech compare to how non-native speakers set about the segmentation process? Cutler (2002) reasons that it is not unwarranted to attribute the same basic processes of native listening to non-native listening as well. She does point out, however, that this assumes that the non-native listener is equipped with the necessary vocabulary. And this assumption does not hold for the bulk of non-native listeners. Field (2003) acknowledges the difficult task of segmentation for non-native listeners with limited or weak L2 knowledge. According to him, L2 learners scan continuous speech for matches between sounds in the input and words known to them. In this process, word boundaries are often violated in order to achieve matches. An example of this is given, where a learner extracts the word *sister* from the sentence *went to assist a passenger*. The danger Field sees in this matching strategy is that learners might go on to construct a mental model of what they hear, based on an inaccurate segmentation and consequently may even reshape the input to comply with their mental model.

According to Cutler (2002), the insufficient L2 vocabulary is not the only factor affecting lexical segmentation. Non-native listeners appear to apply native phoneme sequence probabilities inappropriately to L2 input. This may result from activation of the native vocabulary by the non-native input, which means that the knowledge of the native language is interfering with non-native listening.

A study by Altenberg (2005) investigated how L2 learners compare to native listeners in using acoustic-phonetic cues (e.g. aspiration, glottal stop) to segment natural speech into words. The main question was whether non-native listeners are able to acquire and use the cues of their second language. In her study, Altenberg presented Spanish L2 learners of English and native speakers of English with stimulus phrases such as *keeps parking* versus *keep sparking*, in which only acoustic-phonetic features provide the necessary segmentation cue. Participants had to indicate which of the two phrases they heard. L2 learners were significantly worse at using the acoustic-phonetic cues than native speakers. Another finding was that the aspiration stimuli were harder for the non-natives to segment than stimuli where a glottal stop was the essential cue. According to Altenberg a possible explanation for this finding is transfer. In Spanish there are no occurrences of aspiration, while glottal stops do occur. Cues available in the native language are thus also believed to be applicable in the L2, while cues not available in the native language not.

Sanders, Neville and Woldorff (2002) investigated the use of other segmentation cues (i.e. lexical, syntactic and stress-pattern cues) by native and non-native speakers of English. They presented participants with three types of sentences, varying in the amount of lexical, syntactic, and stress-pattern information available to the listener. All information was available in the *semantic* sentences (for example, *The child stopped crying when a balloon was given to her*), whereas lexical information was reduced in the *syntactic* sentences (for example, *The ferp trepped plawing when a barreal was kaffen to her*), and both lexical and syntactic information were missing in the *acoustic* sentences (for example, *Sa ferp trepp plawel ron I barreal hof kaffem gi wem*). Participants carried out a phoneme monitoring task, where they had to press a button when they heard the target phoneme in the carrier sentence. They had to specify their response by pressing different buttons, according to the place of the target phoneme within a word (i.e. beginning, middle, or end). Identification of the location of a phoneme within a word indicates where participants place word boundaries. With this design Sanders et al. could study which types of information are employed in the process of speech segmentation. As for semantic and lexical information, this study suggests that both native and non-native speakers benefit similarly from these types of information. This is in contrast to syntactic information, which was only used by the native group. This finding indicates an "inability of non-native speakers to process syntactic information in a native-like manner" (Sanders et al., 2002, p. 526). Finally, the results pertaining to stress-pattern effects indicate that the non-native group did use this type of information, but to a smaller extent when lexical and semantic information were present. This may be seen as comparable to the hierarchical structure of segmentation cues we saw for native listeners.

The comprehension of auditory input may seem effortless, but the continuous speech stream requires listeners to first segment the input in distinct units. Native speakers are able to utilize segmentation cues in order to detect word boundaries; they may use prosodic cues, cues regarding the distributional structure of lexical items, phonotactic strategies, or the identification of lexical items. Non-native speakers appear not to be able to use acoustic-phonetic cues and syntactic information as efficiently as native speakers. Studies discussed in this chapter tested their claims with carefully recorded speech, but how do listeners segment speech that is casually produced and full of variation? How do native and non-native listeners approach conversational reduced speech segmentation? And are non-native speakers at all capable of segmenting reduced forms? In the following chapter I will discuss these questions.

3 Reduction

Now that the problem of segmentation has been elucidated, I will turn to reductions to the speech signal. Field (2003) discusses the accommodotory phonological processes causing variation in the speech signal, saying that they particularly operate at word beginnings and endings. Especially in word beginnings and endings, listeners need unambiguous information in order to successfully identify word boundaries. The fact that variation typically occurs at word boundaries may prove to hinder speech recognition. Examples of accommodotory phonological processes are assimilation and reduction. According to Field these two processes differ from one another in the consistency of the pattern in which they occur. Assimilation takes place in an anticipatory manner, adjusting certain sounds to those that follow. For instance, in Dutch the [n] may be pronounced as [m] if followed by a sound with a labial place of articulation. The Dutch *tuin* [tœyn] (English, *garden*) may be assimilated into [tœym] in the context of *tuinbank* (English, *garden bench*) (Mitterer & Blomert, 2003). Reduction, however, is not as consistent in its pattern. Reduction can be seen in the reduction of full vowels to schwas, for instance when the Dutch word *moment* [moment] (English, *moment*) becomes [mәment] (Coussé et al., 2007). Speech segments[2] can also be entirely absent in the signal. A syllable is deleted in the case when [moment] becomes [ment] (ibid.). In future, when referring to reduction both processes (i.e. reduction and deletion) are implied. The challenge of dealing with its unpredictability in realization is the reason I focus on reduction in this thesis.

3.1 Careful versus casual speech

Spoken language can be either carefully articulated or casual. In careful speech – often observable in formal addresses and public speaking – all segments are clearly vocalized. Casual, spontaneous conversations on the other hand are typically lacking attention to careful pronunciation, which can result in reduction to the speech signal. Johnson (2004) found reductions to be frequently occurring in conversational American English. Segments are reduced in up to 25% of content words and in about 40% of function words. Deletion of complete syllables occurs in

[2] To avoid any confusion, when referring to speech segment in the sense of segment reduction, I mean any possible fragment (e.g. phonemes, morphemes, syllables) of a lexical form that may be reduced in the realization of that form. The use of the term *segment* in this sense follows from the literature and does not pertain to the sense of being the result of the segmentation process, i.e. an identified meaningful unit extracted from the speech signal.

6% of content words and 4,5% of function words. These reductions also appear in varying forms; Keating (1997) found an average of 14 distinct pronunciations of high-frequency words. Exact numbers on the occurrence of reduction in Dutch have not been reported. However, as will become apparent in Chapter 4, there are numerous observations of reduction in Dutch as well. It is the task of the listeners to recognize the reduced forms as variations of the full forms.

Differences between careful and casual speech arise from the compromise between ease of articulation and ease of perception. Ease of perception is given priority when people talk carefully. They want to be as intelligible as possible, at the expense of effortless articulation. When talking casually on the other hand, more importance is given to ease of articulation, at the cost of ease of perception; speakers put less effort in articulation of their speech, thereby risking incomprehension on the side of the listener. In a similar way, Lindblom (1990) proposes in his *H&H theory* that speakers adapt their production according to communicative and situational demands, thus varying the output along a continuum of Hyperspeech (i.e. careful articulation) and Hypospeech (i.e. reduced speech).

Ernestus (2000) emphasizes that although reduction of articulatory effort – and the reduced speech associated with it – is a natural tendency, it will never occur at the cost of possible obstruction in communication. As a consequence, reduction tends to affect only those segments whose reduction is not likely to cause misunderstanding. According to Ernestus, these are typically segments that are less acoustically salient, less relevant for recognition, and of high frequency. Especially frequency of occurrence seems to influence pronunciation to a large extent, i.e. more prominently occurring words tend to be subjected more to reduction than words occurring less frequently. A different study (Pluymaekers et al., 2005) also investigated the effect of frequency on acoustic reduction in Dutch. Measurements were made of the duration of four Dutch affixes and the realization of these affixes was found to be shorter in high-frequency words. Given that high-frequency words are recognized more easily than low-frequency words, they can be reduced, whilst still maintaining their intelligibility. Ernestus also gives a possible alternative explanation for the high reduction rate in high-frequency words; reduced forms of these high-frequency items may be stored in the mental lexicon, making them directly accessible during the recognition process. I will come back to the issue of how reduced forms are represented in the mental lexicon in Section 3.2.

Not only does the frequency of a word determine whether it is likely to be reduced. The probability of a word given its surrounding context has an effect on reduction as well (Jurafsky, Bell, Gregory, & Raymond, 2001). According to the *Probabilistic Reduction Hypothesis* (PRH), words are reduced when their "neighboring words, syntactic and lexical structure, semantic expectations, and discourse factors" (ibid., p. 229) render them highly probable. Jurafsky et al. directed their attention primarily on the context of neighboring words. They compiled a dataset of 38.000 transcribed words, all instances of the ten most frequent English function words from the *Switchboard* corpus[3]. They then coded vowel reduction and duration for each function word, which they used to test their hypothesis. The same was also done for content words. Results supported the PRH, i.e. words that are more probable given the surrounding words, are more likely to be reduced. The effect of the surrounding context was higher for function words than for content words, which was attributed by the researchers to "the smaller number of observations in the content words dataset or to the general lower frequency of content words" (ibid., p. 245).

Still, Jurafsky et al. (2001) needed to invalidate a potential confounding variable, namely the lexicalization of word combinations. Many of the word pairs in the dataset could be seen as lexicalizations (e.g. *sort of*, *kind of*). If these combinations are indeed lexicalized units (i.e. single items), then the effect of predictability on reduction says nothing about the relation between two separately accessed words. Rather, it relies solely on an orthographic definition of a word, in which, for instance, *sort* and *of* are only seen as separate words because of the white space between them. Jurafsky et al. conducted a further experiment to discover whether higher predictability leads to more reduction, even in word combinations that are not lexicalized. They found that this was indeed the case; even predictability in word pairs that are unlikely to be lexically combined, exhibits a reduction effect. This

[3] The *Switchboard* corpus comprises of 2430 spontaneous telephone conversations in American English, adding up to 240 hours of recorded speech (Godfrey, Holliman, & McDaniel, 1992).

effect, however, was smaller than the effect found earlier with potentially lexicalized combinations. Thus, lexicalization does appear to influence the reduction process.

3.2 Perception of reduced forms

Reduced words can deviate greatly from their full equivalents. In the previous sections I have established that reduction occurs mostly in high-frequency words and that this reduction can take on quite extreme forms. Notwithstanding this large occurrence of – often extremely – reduced forms, there seems to be little difficulty in the perception of these forms in everyday life. Or, as McQueen and Cutler state, "the system seems to tolerate natural variation" (2001, p.478). However, to this date not much is known about how exactly listeners comprehend acoustically reduced speech (Ernestus & Baayen, 2007).

Mitterer and Ernestus (2006) investigated both production and perception of a certain type of reduction in Dutch, i.e. word-final /t/-lenition. Most relevant in this discussion are the perception experiments, which showed that native listeners are able to use phonetic and contextual cues to assess the probability of reduction. Listeners more often reported hearing a word-final /t/ when phonetic and contextual cues suggested a high probability of an underlying /t/ at the end of the word. Apart from the effects of this bottom-up information, the effects of top-down processes were also studied. The researchers set up another experiment to find out if lexical knowledge influenced how listeners perceive a lenited /t/. Listeners appeared to hear a /t/ more often if its presence gave rise to an existing word, than if its presence brought about a non-word. Both data-driven and concept-driven processing thus aid listeners in the comprehension of reduced forms. Still, comprehension of such forms is slower as compared to the comprehension of unreduced forms (Ernestus & Baayen, 2007). This was concluded by Ernestus and Baayen after subjecting 47 native speakers of Dutch to a listening experiment. In the experiment Dutch prefixed words were recorded either with all segments clearly articulated, or with a reduced prefix. Participants listened to these realizations and indicated whether they heard a real or a non-word. Words with segment deletion resulted in delay of word recognition, as compared to their unreduced equivalents.

Context has been showed to have an effect on recognition of reduced forms as well (Ernestus et al., 2002). In this experiment native listeners of Dutch were presented with word forms in several degrees of reduction and context and were given the task to write down what they heard. Word forms were selected from a corpus of spontaneous casual conversations. The degrees of reduction were (1) High, (2) Medium, and (3) Low. For instance, the Dutch word *daarom* [darɔm] (English, *therefore*) had respectively the three realizations [dam], [darəm], and [darɔm]. The contexts in which the word form was presented were (1) the complete sentence in which the word forms were realized (Full Context), (2) together with the adjacent vowels and intervening consonants (Limited Context), and (3) no context (Isolation Context). Results showed that listeners are able to recognize highly reduced word forms very accurately, but only when those are presented in their full context. When the context was limited, the extent of recognition was dependent on the degree of reduction; the more reduced a word form was, the more difficult it was to recognize. Apparently, the phonetic cues in the limited context were not sufficient for recognition. Listeners seem to need the semantic and syntactic information available in the full context to accurately reconstruct the reduced words.

Ernestus et al. (2002) posit that reconstructive processes are at play when listeners perceive reduced forms. This notion relates to how forms are accessed in the mental lexicon. Two distinct theories have been postulated regarding this matter (ibid.); the direct lexical access theory and the reconstruction theory. The first one claims that all possible phonetic forms of a word (i.e. full form and reduced variants) are stored in the lexicon, making it possible to directly access these lexical items. According to the second hypothesis only one form (i.e. the full form) is represented in the mental lexicon. For the reduced variants to be recognized, they first have to be reconstructed to match the stored full form. The fact that in Ernestus et al. degree of reduction was found to affect recognition refutes the first theory. Were all the reduced forms to have their own representation in the mental lexicon, then these representations would be equally well accessible, contrary to the findings.

Further evidence against the direct lexical access theory comes from phoneme-monitoring studies. One such study (Kemps et al., 2004) presented native speakers of Dutch with full realizations of Dutch words containing the phoneme [l] and their reduced equivalents that did not contain the target phoneme. Participants had to decide whether they did or did not hear an [l]. As in Ernestus et al. (2002), degree of context was varied and it was found that in the presence of context many participants gave false positive responses; they reported hearing the target phoneme when it was actually not present in the acoustic signal. Kemps et al. concluded from this that "the conscious percept is based on the activated canonical representation, not so much on (a pre-lexical representation of) the acoustic signal itself" (p. 125). Given enough contextual cues, listeners are capable of reconstructing the reduced form to its full form. The activation of the full form makes them think they heard the full realization, when, in fact, they heard the reduced variant. Participants did not report hearing the target phoneme in the isolation context, which means no reconstruction had taken place in that context (i.e. if reconstruction had taken place, the canonical full form would have been activated and participants would report hearing the target phoneme). The reduced form did not trigger reconstruction into the full form, because no cues were available to initiate the reconstruction process. When no reconstruction takes place, the full form does not get activated and as a consequence, listeners may not recognize what is being said.

3.3 Native versus non-native listening

The literature reviewed in Section 2.3 has made clear that non-native speakers are presented with a difficult task when listening to the L2 speech stream. But how do the non-native listeners deal with the task of segmenting speech that is lacking of certain speech sounds? Non-native speakers are far less proficient in the L2 than native speakers, plainly because it is not their dominant language. Therefore, many words may be unfamiliar to them. But reduction may even cause familiar words to seem unknown, given the deviation between the full form and its reduced equivalent.

Native speakers usually have no trouble comprehending reduced speech, provided that it is perceived within a rich context. Difficulties may occur in native listening, but those are primarily due to signal degradation or hearing loss, i.e. signal access problems (Bradlow & Bent, 2002). Bradlow and Bent believe that non-native speakers are hindered in the speech perception process due to a different problem, i.e. code access problems. They studied the effect of clear speech – as opposed to conversational speech – on speech perception in both native and non-native listeners. Clear speech was elicited with the instruction to speak as though speaking to a listener with a hearing loss or to a non-native speaker. Conversational speech was elicited by instructing speakers to talk at a normal pace and with no particular attention to clarity. Clear speech production involves two types of enhancements. On the one hand, it involves signal enhancements, which make the signal more acoustically salient (e.g. slower speaking rate, wider pitch range). And on the other hand, it pertains to enhancements to the linguistic code, such as adhering to pronunciation norms. The goal of the study was to shed light on whether clear speech has a facilitative effect for non-native listeners. Subjects were either native or non-native speakers of English. Both groups were believed to benefit from the signal enhancements of clear speech, but only the native speakers were thought to be aided by the code enhancements. This hypothesis is based on Bradlow and Bent's belief that "only a listener who is already sensitive to the important dimensions of contrast that define the sound system of the target language or who is familiar with the detailed pronunciation patterns of native speakers is likely to benefit from the code enhancements of clear speech" (ibid., p. 273). Participants heard sentences that varied in speaking style (conversational versus clear) and their task was to write down what they had heard. Performance of the two groups confirmed the researcher's expectations; the non-native speakers showed a smaller clear speech effect than the native speakers. While native speakers benefit from both acoustic and code enhancements, non-native speakers are only aided in the perception process by the acoustic enhancements in clear speech.

Bradlow and Bent (2002) investigated the effects of signal access and linguistic code enhancements. Reduction can pertain to pronunciation norms, and thus to the linguistic code investigated by Bradlow and Bent. However, reduction in itself was not specifically addressed in the study. The question remains as to how listeners – be it native or non-native – segment reduced speech. Studies of native listening have shown that comprehension of reduced forms is slower than that of unreduced forms and that native listeners struggle when reduced forms are presented in isolation. Evidence was also found for the reconstruction theory of reduction. But how do all these findings relate to the way non-native listeners perceive reduced forms. To my knowledge, non-native segmentation of reduced forms has not yet been studied. In this thesis, I hope to provide insight into this matter.

3.4 Hypotheses

On account of the literature it is possible to formulate hypotheses regarding the research questions. The research questions of this thesis are the following:

1a *Is there a difference between native and non-native speakers of Dutch in the segmentation of reduced and unreduced forms?*

1b *If so, is that a difference in speed or in accuracy, or in both?*

2 *Is there a particular type of reduction that is difficult to segment?*

2b *If so, is that comparable for native and non-native speakers?*

Non-native speakers (NNS) of Dutch are believed to experience more difficulties than native speakers (NS) in segmenting both reduced and unreduced forms (Altenberg, 2005; Sanders et al., 2002). Previous findings – with only unreduced stimuli – suggest that this difficulty manifests itself in accuracy, but whether it also manifests itself in speed and whether the same holds for reduced forms, is not known. Overall, NNS should perform slower and less accurately than NS, due to their inability to effectively use segmentation cues, and due to their lack of knowledge of Dutch reduction conventions.

As for the second research question, no previous studies have investigated segmentation in a diverse set of reduction categories. This question will thus be addressed in an exploratory manner. It is reasonable to assume that single segment reduction – i.e. vowel or consonant reduction – will present fewer problems than extreme reduction, where multiple segments have been deleted. However, reductions are believed to be employed only if they do not hinder comprehension (Ernestus, 2000). Deleting multiple segments heightens the possibility of incomprehension. Thus for extremely reduced forms to be permitted, they have to be highly recognizable. If the reduced form thereby becomes the default form at the cost of the unreduced canonical form, then extreme reductions may perhaps prove to be easier to comprehend.

I have constructed two segmentation tasks, including unreduced and (three categories of) reduced Dutch forms. Performance on these tasks by native and non-native speakers of Dutch was studied in order to test the hypotheses. Before turning to the actual research, I will give an overview of three reduction categories in Dutch.

4 Categories of reduction

A study on segmentation of reduced forms requires knowledge of how words are reduced in conversational speech. I have mentioned the dissertation of Mirjam Ernestus (2000), which includes a large survey of reductions in Dutch. I will describe the generalizations made by Ernestus in the following sections, distinguishing three categories of reduced forms (i.e. consonant, vowel, and extreme reduction). The choice for these particular categories follows from the classification made by Ernestus and all of the examples in this chapter are also taken from her dissertation.

4.1 Consonant reduction

Ernestus (2000) presents a rough generalization of consonants, which she found to be frequently absent in casual speech. I will mention these findings concisely, for a more detailed discussion I refer to Chapter 6 in Ernestus' dissertation.

Absence of [t]

In Dutch the [t] tends to not be realized in the middle of consonant clusters, especially after [s] (Example 1a), and before bilabial stops (Example 1b). It is also frequently absent in the word *niet* (English, *not*) (Example 1c) and at the end of certain verb-stems (Example 1d).

(1) Absence of [t][4]

a.	*juist gehoord*	/jŒyst xəˈhord/	[jŒysx□ˈhort]	‘just heard’
b.	*lijkt me*	/l□ikt mə/	[l□ikm□]	‘seems to me’
c.	*niet echt*	/nit □xt/	[ni□□xt]	‘not really’
d.	*vind ik*	/vɪnd ɪk/	[fɪnɪk]	‘am of the opinion I’

Absence of [r]

When in coda position the [r] is particularly missing after schwa (Example 2a) and low vowels (Example 2b). In onset position the reduction of [r] is limited to the word *precies* (English, *precisely*) (Example 2c).

(2) Absence of [r]

a.	*anders*	/ɑnd□rs/	[ɑnd□s]	‘different’
b.	*daarna*	/darna/	[dana]	‘after that’
c.	*precies*	/pr□sis/	[p□sis]	‘precisely’

Absence of [n]

The [n] is mainly reduced after schwa (Example 3a), after full vowels – either in word-medial (Example 3b) or in word-final position (Example 3c).

(3) Absence of [n]

a.	*eten*	/et□n/	[et□]	‘to eat’
b.	*tenminste*	/t□nmɪnst□/	[t□mɪst□]	‘at least’

[4] The layout of the examples is replicated from Ernestus (2000). The Dutch orthographic form is given in italics, the phonological underlying form enclosed in slashes, the phonetic realization between square brackets, and the English (literal) translation between single quotation marks.

c.	*gaan ze*	/xan/ *ze*	[xa]*ze*	'go they'

Absence of [h]

Absence of [h] typically occurs in forms of the verb *hebben* (English, *to have*) that follow a consonant and are not accented (Example 4).

(4)	Absence of [h]			
	ik heb	/ɪk h□b/	[ɪk□p]	'I have'

Absence of [x]

The segment [x] is often missing in the combination of the words *nog* (English, *still*) and *toch* (English, *yet*) with labial consonants (Example 5a) or an [n] (Example 5b).

(5)	Absence of [x]			
a.	*nog wat*	/nɔx □ɑt/	[nɔ□ɑt]	'still something'
b.	*toch niet*	/tɔx nit/	[tɔnit]	'yet not'

Absence of [k]

The [k] may be absent in [ŋk] clusters, like in *denk* (English, *think*) (Example 6).

(6)	Absence of [k]			
	denk	/d□ŋk/	[d□ŋ]	'think'

Absence of [d]

When the [d] is located within a [nd] cluster, followed by schwa, then it is oftentimes not realized (Example 7).

(7)	Absence of [d]			
	anders	/ɑndərs/	[ɑnərs]	'different'

Absence of [l]

Non-realization of the segment [l] tends to be limited to the word *als* (English, *if*) (Example 8).

(8) Absence of [l]

als /ɑls/ [ɑs] ‘if’

Absence of [f]

This last of the consonants Ernestus (2000) described as frequently occurring in Dutch is the [f]. As with some of the other segments, the non-realization of the [f] is also restricted to particular lexical items, in this case *zelfde* (English, same) and *zelfs* (English, *even*) (Example 9).

(9) Absence of [l]

zelfde /z□lfdə/ [z□fdə] ‘same’

4.2 Vowel reduction

Ernestus (2000) as well as other studies (Coussé et al., 2007; Kloots et al., 2003) mention vowel reduction being an umbrella term for vowel shortening, reduction to schwa and vowel deletion. Ernestus focuses primarily on reduction of full vowels to schwas and vowel deletion.

Reduction of full vowels to schwas

Several factors seem to influence the reduction of vowels to schwas (Ernestus, 2000). Stressed vowels, vowels in onsetless syllables, vowels in closed syllables, and vowels in word-final syllables are seldom realized as schwas. I have already mentioned that reduction in general is highly influenced by frequency of occurrence, this is also the case with vowel reductions; vowels in high-frequency words tend to be reduced to schwas more often than vowels in low-frequency words. There appears to be a hierarchy in the likelihood of a certain vowel to be reduced to schwa: /y, u, ø/ > /i/ > /o, ɔ/ > /a, ɑ/ > /e, ɪ/; the lower a vowel's position in the hierarchy, the more often it is reduced to schwa (Booij, 1995). According to Booij, diphthongs are not likely to be realized as schwas. Ernestus did, however, find examples of this in her dataset (Example 10a-b). Another example is the Dutch possessive form *mijn* [m□in] (English, *my*) that has been found to be often realized with a schwa, albeit only if it is not stressed (Gillis, Kloots, & Swerts, 2004).

(10) Reduction of diphthongs to schwa

a.	*altijd*	/ɑlt□it/	[ɑltət]	'always'
b.	*bij mij*	/b□i m□i/	[bə m□i]	'with me'

Absence of vowels

Ernestus (2000) posits the following regarding the complete absence of vowels:

> Unstressed vowels which are present in formal Dutch are absent in casual Dutch particularly in three types of contexts: in vowel hiatus position, between obstruents and liquids, and adjacent to continuants. The absence of vowels often coincides with the absence of adjacent consonants. (p. 129)

Vowel hiatus occurs when one vowel is separated from another by a word boundary. One of the two vowels is deleted as a consequence (Example 11a). The absence of a schwa may occur between obstruents and liquids (Example 11b). Or, it may occur adjacent to continuants (Example 11c).

(11) Absence of vowels

a.	*sta ik*	/sta ɪk/	[stak]	'stand I'
b.	*makkelijk*	/mɑkələk/	[mɑklək]	'easy'
c.	*vorige*	/vɔrɪxə/	[vɔrxə]	'previous'

4.3 Extreme reduction

In extremely reduced forms several segments are absent. Ernestus (2000) observes that particular segments are typically not subject to reduction – i.e. initial and final segments, and the onset and nuclei in the stressed syllables – and thus form the basis of the extreme reduction (Example 12a-f).

(12) Extremely reduced forms

a.	*volgend*	/vɔlxənd/	[fɔlnt]	'next'
b.	*allemaal*	/ɑləmal/	[ɑməl]	'all'
c.	*daarom*	/darɔm/	[dam]	'therefore'
d.	*ongeveer*	/ɔnxəver/	[ɔfer]	'approximately'
e.	*natuurlijk*	/natyrlək/	[tyk]	'of course'
f.	*eigenlijk*	/☐ixənlək/	[☐☐ik]	'in fact'

These generalizations in Dutch reduction show that reduction can be rather subtle, as in the reduction of full vowels to schwas, but also quite extreme (like the extremely reduced forms in Example 12). The three reduction categories were used in the segmentation task that will be described in the next chapter.

5 Method

In this chapter I will describe the empirical study that was designed to assess segmentation of reduced speech. Differences in segmentation between native and non-native speakers of Dutch were investigated, as well as the nature of possible differences (i.e. speed and accuracy). A second aspect of the study was the comparison of three categories of reduction discussed in the previous chapter. These categories were incorporated into the segmentation tasks to find out whether there is a particular type of reduction that is difficult to segment.

5.1 Participants

30 participants took part in the experiment, most of which were students at the University of Amsterdam. 15 were native speakers (NS) and 15 were non-native speakers (NNS) of Dutch. The NNS did not all have the same first language and they varied in the amount of exposure to the Dutch language (see Appendix I for background information of all participants). There was considerable variety in the number of years the participants had been learning and speaking Dutch; the estimated level of proficiency varied from beginner to near-native.

Because the effects of age and level of education[5] on segmentation are yet to be investigated, I have decided to keep these variables stable. The target population consisted of highly educated normal-hearing adults up to the age of 35 (age range 21-34). The age limit was an arbitrary boundary, limiting the population to younger adults.

5.2 Materials

5.2.1 Items

The segmentation task (see Appendix II for the complete test) consisted of 60 items (3 practice and 57 target items), each containing a reduced Dutch form. These forms were gathered from previous studies (Coussé et al., 2007; Ernestus, 2000; Ernestus & Baayen, 2007; Ernestus et al., 2002; Kemps et al., 2004; Keune et al., 2005; Kloots et al., 2003; Pluymaekers et al., 2005) and were divided into three categories: vowel reduction, consonant reduction, and extreme reduction. In (13) examples are given of items from the segmentation task, differentiated by type of reduction. Example (13a) shows cases of vowel reduction, Example (13b) of consonant reduction, and Example (13c) of extreme reduction.

(13) Items from segmentation task

a. Vowel reduction

[5] The effects of individual differences in age and level of education will be investigated in the earlier mentioned STILIS-project.

	hij is lang	/h□i □ɪs lɑŋ/	[h□islɑŋ]	'he is tall'
	ga ik ook	/xa □ɪk □ok/	[xakok]	'I also go'
b.	Consonant reduction			
	dat vind ik	/dɑt fɪnt ɪk/	[dɑfɪnɪk]	'I find that'
	toch wel	/tɔx □□l/	[tɔ□□l]	'anyway'
c.	Extreme reduction			
	was inderdaad heel	/□ɑs ɪndərdat hel/	[□ɑsɪdathel]	'was indeed very'
	of het mogelijk is	/ɔf hət moxələk ɪs/	[ɔfətmokɪs]	'if it is possible'

The context in which the items were presented, varied from 2 to 4 words per item and was obtained by searching the *Corpus Gesproken Nederlands*[6] (CGN, Spoken Dutch Corpus) for realizations of the 60 reduced forms. The longer the items, the more information needs to be held and processed in working memory before a response can be given. To limit the load on working memory, the maximum amount of words in one item was restricted to 4.

All of the words in the task were controlled for frequency of use; only high-frequency words were selected. This was done to ensure understanding of the items by all participants, including the non-native speakers. Given that reaction times were measured, the items were also controlled for location of the reduced form. This form was always presented at the second to last position in the item. Some items had more than one reduced word, but in all items it was seen to that the last word was unreduced, as well as always being a single-syllable word. In doing so, comparability was maximized in the elapsed time between the presentation of the reduced form and the response.

A female native speaker of Dutch produced the selected stretches of speech, which were recorded in a recording studio at the institute of phonetic science of the University of Amsterdam. Two versions of each item were recorded; one was a full realization with pronunciation of all segments, and the other version was reduced.

[6] The *Corpus Gesproken Nederlands* (CGN) is a databank containing a vast amount of spoken Dutch, recorded in both the Netherlands and Flanders. The corpus consists of nearly 9 million words, adding up to 800 hours of speech (http://lands.let.kun.nl/cgn/ehome.htm).

Attention was paid to reduce only vowels in the vowel reduction category and only consonants in the consonant reduction category. In all reduction categories, the items were realized according to the phonetic transcriptions in the literature. Both reduced and unreduced versions were produced with the same intonation pattern.

5.2.2 Tasks

The 60 items made up the segmentation task, which consisted of two subtests; a word-count task and an identification task. In the word-count task participants heard an utterance and had to respond as quickly as possible by typing the number of words they had heard on a keyboard. This format was chosen because it seems to enable both speed and accuracy measurement (i.e. *word-count speed* and *word-count accuracy*). A possible objection to the word-count format may be that it appeals to processes that normally do not operate when listening to oral language. It may require explicit access to the phonological loop because an utterance needs to be held in working memory long enough to count the individual words.

The above objection may make counting the number of words one has heard a debatable measure of segmentation. That is why a second subtest was included, viz. the identification task. Participants heard the same stretches of speech, but were now asked to identify the words they had heard by typing them on a keyboard according to their orthographic form. The process of identifying seems more natural and automatized than counting the words, but it is only functional as an accuracy measure, not a speed measure. Two types of accuracy are measures in this task; the task assesses how many words are correctly identified (*identification count*, comparable to the *word-count accuracy* measure), as well as the correct identification of the words (*identification accuracy*).

To test if working memory capacity indeed plays a role in the word-count subtest, a working memory task was included to the study. In this task the capacity of the phonological loop was measured by presenting participants with two series of non-words. The participants then had to decide whether the two series were identical or not. The strings were grouped in series of three, according to the number of non-words in the string. The first series contained strings of two non-words, which increased by one non-word each time a new series began. Greater demand was placed on the phonological loop by increasing the number of non-words in the series. The more words one can hold in working memory, the larger its capacity is. The working memory task consisted of two parts. In the first part the non-words in a series were distinct in sound, so that they were easily distinguishable from one another. In the second part the non-words were similar in sound, which is likely to cause mixing up of the words. The second part is thus believed to make a greater demand on the phonological loop.

5.3 Design

Two versions of each item were recorded; a full realization with pronunciation of all segments, and a reduced version. The two versions of each item were randomly distributed over two lists (1 and 2), each list making up a different stimulus set. Both lists contained all 60 items, but differed in the version of each item (i.e. either reduced or full realization). Appendix II shows which items were (un)reduced in each of the two lists. Participants were randomly assigned to one of the lists, an equal number of participants hearing each list. These lists were used in two tasks – i.e. word-count task and identification task – to assess the segmentation skills of the participants. The dependent variables were segmentation speed and segmentation accuracy. The word-count task measured both speed (*word-count speed*) and accuracy (*word-count accuracy*), while the identification task measured only accuracy (*identification accuracy* and *identification count*) [7].

For the purpose of answering the first research question – about segmentation of reduced and unreduced speech by native and non-native speakers – the factors nativeness, reduction, and task type were examined. The data were structured, so

[7] *Identification count* is included in the measure of *identification accuracy*, as identifying which words they hear, automatically forces participants to segment the speech stream in the according number of words. Taken together with the comparable measure of *word-count accuracy*, the *identification count* was not included in analyses other than the correlation analysis.

that these three factors became within-item[8] factors; for each of the 60 items the scores were given of the different levels of the three factors. Main effects of all three factors – as well as interaction effects – could be determined this way. This provided the necessary information about possible differences between native and non-native speakers in segmentation speed and/or accuracy of (un)reduced forms.

For the second research question regarding the three reduction categories, the data set was ordered according to a within-subject design. The within-subject factors were reduction category and task type. Mean scores were thus obtained for each subject on each task and for each reduction category. The between-subject factor was nativeness. This allowed studying segmentation skills – both speed and accuracy – of native and non-native speakers on the three types of reduction.

In the Materials section, concerns were raised as to the role of working memory capacity on the word-count subtest. That is why working memory was included in the design as a moderation variable.

5.4 Procedure

The test battery was administered individually in a quiet room in the university building. Participants were seated in front of a laptop, wearing headphones. The first task they were given was the word-count task. They heard the stretches of speech over the headphones and responded as fast as they could *how many* words they had heard by pressing the corresponding digit on the keyboard. Responses of all the tasks – including reaction times – were measured and stored in E-Prime, a software program for conducting scientific experiments. The order of the items was randomized for each participant. After the word-count task, the same items were presented, but now the participants had to identify *which* words they had heard; the words were typed on the keyboard and stored in E-Prime. Items were again presented in a random order. After the two segmentation tasks, the working-memory task was administered. Participants heard two strings of non-words and

[8] The dataset was ordered with items as within-variable, hence the use of the term *within-item*. This is comparable to the *within-subject* design, where the same subjects are tested in a repeated manner.

had to decide whether the two were identical or not, by pressing respectively 'Z' and 'M' on the keyboard. The experiment was set up such that it automatically stopped when at least two out of three strings in a series were responded to incorrectly.

After completing the tests, participants were given a questionnaire to fill in. This was done in order to obtain demographic data, as well as linguistic background information (see also Appendix III).

6 Results

6.1 Scoring

The segmentation task originally consisted of 60 items. However, item analysis showed that perhaps not all items were adequately constructed. The criterion for excluding particular items from analysis was a non-perfect accuracy score of native speakers on the identification of unreduced items. Native speakers should be able to correctly identify spoken words in their full canonical form. If that is not the case, then there is probably something wrong with the item. For this reason twelve items were not included in the data analysis. These twelve items are presented separately in Appendix II.

The tasks were too difficult for one participant, as demonstrated by a very poor performance in all tasks. Her level of Dutch proficiency was assumed to be too low and her data was therefore not included in the analysis.

6.2 Correlational analysis

An independent samples *t*-test was carried out to ascertain the comparability of the two lists participants heard during the segmentation tasks. Correlation analysis was carried out to determine the relationship between the different measures of segmentation, as well as with the working memory task.

In performing the segmentation tasks, participants heard one of two lists, which differed in the realization of each item (i.e. either reduced or full realization).

An independent *t*-test was performed to make sure the two lists were comparable. Mean and standard deviation are given in Table 1 for both lists on each measure of segmentation. There were no significant differences between the two lists on all measures of segmentation, i.e. word-count speed ($t(27) = -,75$, $p > ,05$), word-count accuracy ($t(27) = -1,09$, $p > ,05$), identification accuracy ($t(27) = -,67$, $p > ,05$), identification count ($t(27) = -,92$, $p > ,05$). Given that there was no difference between the two lists, no distinction was made between the two in further analyses.

Table 1 Mean and SD of all segmentation tasks on the two lists

	List 1 (N=15)		List 2 (N=14)	
	Mean	SD	Mean	SD
Word-count speed	2397,38	644,70	2574,33	617,38
Word-count accuracy	,88	,09	,91	,06
Identification accuracy	,83	,19	,86	,07
Identification count	,93	,08	,95	,04

Results of the correlation analysis can be seen in Table 2. All four segmentation measures were included, as well as the two parts of the working memory task. There was a significant positive relationship between the two accuracy measures, i.e. word-count accuracy and identification accuracy ($r = ,80$, $p < ,05$). However, only identification accuracy correlated with the speed measure ($r = -,47$, $p < ,05$). This was a negative relationship, so the faster the speed measure, the higher the accuracy measure and vice versa; being fast in segmentation speed correlated with being good in segmentation accuracy. Identification count (i.e. how many words were identified) correlated with all other segmentation measures; a negative correlation with word-count speed ($r = -,39$, $p < ,05$) and a positive correlation with word-count accuracy ($r = ,68$, $p < ,05$) and identification accuracy ($r = ,77$, $p < ,05$). Finally, the working memory task did not show a significant relationship with any of the segmentation measures.

Table 2 Correlation matrix of all (sub)tasks with Spearman's rho correlation coefficient (N=29)

		Word-count speed	Word-count accuracy	Identification accuracy	Identification count	Working memory (similar)	Working memory (distinct)
Word-count speed	Correlation Coefficient	1,00	-,26	-,47 (**)	-,39 (*)	,14	-,22
	Sig. (2-tailed)	.	,176	,010	,036	,480	,251
Word-count accuracy	Correlation Coefficient	-,26	1,00	,80(**)	,68(**)	,25	,10
	Sig. (2-tailed)	,176	.	,000	,000	,201	,601
Identification accuracy	Correlation Coefficient	-,47(**)	,80(**)	1,00	,77(**)	,01	,06
	Sig. (2-tailed)	,010	,000	.	,000	,956	,772
Identification count	Correlation Coefficient	-,39(*)	,68(**)	,77(**)	1,00	-,18	,08
	Sig. (2-tailed)	,036	,000	,000	.	,340	,665
Working memory (similar)	Correlation Coefficient	,14	,25	,01	-,18	1,00	,24
	Sig. (2-tailed)	,480	,201	,956	,340	.	,213
Working memory (distinct)	Correlation Coefficient	-,22	,10	,06	,08	,24	1,00
	Sig. (2-tailed)	,251	,601	,772	,665	,213	.

****. Correlation is significant at the ,01 level. *. Correlation is significant at the ,05 level.**

6.3 Research question 1

The first research question asked whether there is a difference between native and non-native speakers of Dutch in the segmentation of reduced and unreduced forms, and if so, whether this difference pertained to speed, accuracy, or both. In this factorial design the independent variables comprised of nativeness (2 levels: native, non-native), reduction (2 levels: unreduced, reduced) and task type (2 levels: word-count, identification). To obtain answers to this research question, two *general linear* *models* (GLMs) for repeated measures were carried out on the data set – one for the speed measure and one for the two accuracy measures – enabling the testing of significance of the different factors. First, the results for the speed measure (i.e. reaction times in the word-count task) are presented. In Table 3 mean reaction times are given of segmentation speed on the word-count task for both native speakers and non-native speakers.

Table 3 Mean reaction time (in ms) and SD of segmentation speed (word-count task) (N=45)

	Unreduced		Reduced	
	Mean	SD	Mean	SD
NS	2111,41	629,96	2312,02	662,78
NNS	2588,02	627,99	2960,09	753,74

In segmentation speed, a significant main effect of nativeness was found ($F_{(1,44)}$ = 106,03, $p < ,05$). NS obtained faster reaction times on the word-count task than NNS. There was also a significant main effect of reduction ($F_{(1,44)} = 19,23$, $p < ,05$); participants reacted faster on unreduced items than on reduced items. No interaction effect was found; native and non-native speakers reacted similarly to reduction in segmentation speed.

Segmentation accuracy was measured with both word-count and identification task. Performances on both tasks are presented in Table 4.

Table 4 Mean accuracy and SD of segmentation accuracy (N=45)

	Word-count accuracy				Identification accuracy			
	Unreduced		Reduced		Unreduced		Reduced	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
NS	,96	,07	,91	,13	1,00	,00	,83	,23
NNS	,92	,12	,78	,23	,91	,11	,64	,27

As with segmentation speed, main effects of nativeness ($F_{(1,44)}$ = 64,05, p < ,05) and reduction ($F_{(1, 44)}$ = 44,16, p < ,05) were also found for accuracy. NS gave more accurate responses than NNS and performance on the unreduced items was more accurate than on the reduced items. There was also a main effect of task type ($F_{(1, 44)}$ = 6,39, p < ,05), which indicated that subject were more accurate on the word-count task than on the identification task. As opposed to segmentation speed, an interaction effect between nativeness and reduction was found in segmentation accuracy ($F_{(1, 44)}$ = 11,11, p < ,05).

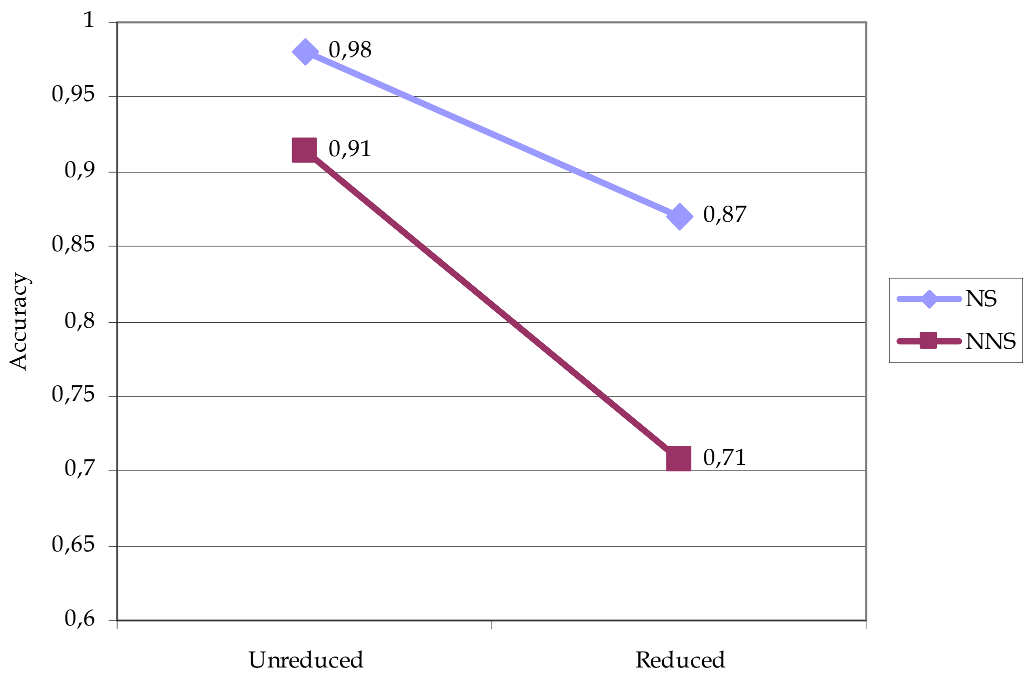

Figure 2 Interaction effect between nativeness and reduction (mean scores of both word-count accuracy and identification accuracy)

In segmentation accuracy, non-native speakers seemed to be more sensitive to reduction than native speakers; the effect of reduction was larger in the group of NNS than in that of NS, as can be seen in Figure 2. An interaction effect was also found between task type and reduction ($F_{(1, 44)}$ = 11,83, p < ,05). In Figure 3 it can be seen that the identification task seemed to be more sensitive to differences in reduction than the word-count task.

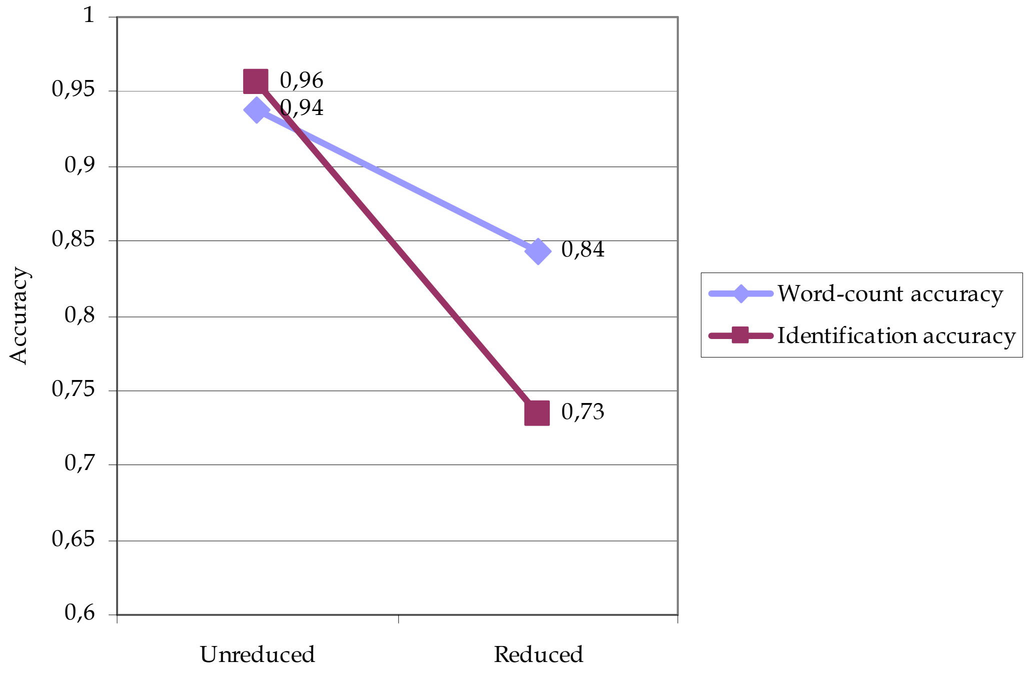

Figure 3 Interaction effect between task type and reduction

There was also a significant interaction effect between nativeness and task type ($F_{(1, 44)} = 5{,}10$, $p <$,05). The NNS were more sensitive to task type than the NS; there was a larger deviation between the two tasks within the group of NNS than in that of NS (see also Figure 4). The non-native speakers were significantly better at counting words than at identifying them. Further analysis was done to see if this interaction effect was also visible when only the reduced items were considered. Figure 5 shows that this was not the case ($F_{(1, 44)} = 2{,}432$, $p =$,126); NS and NNS are equally sensitive to task type when only the reduced items are included in the analysis.

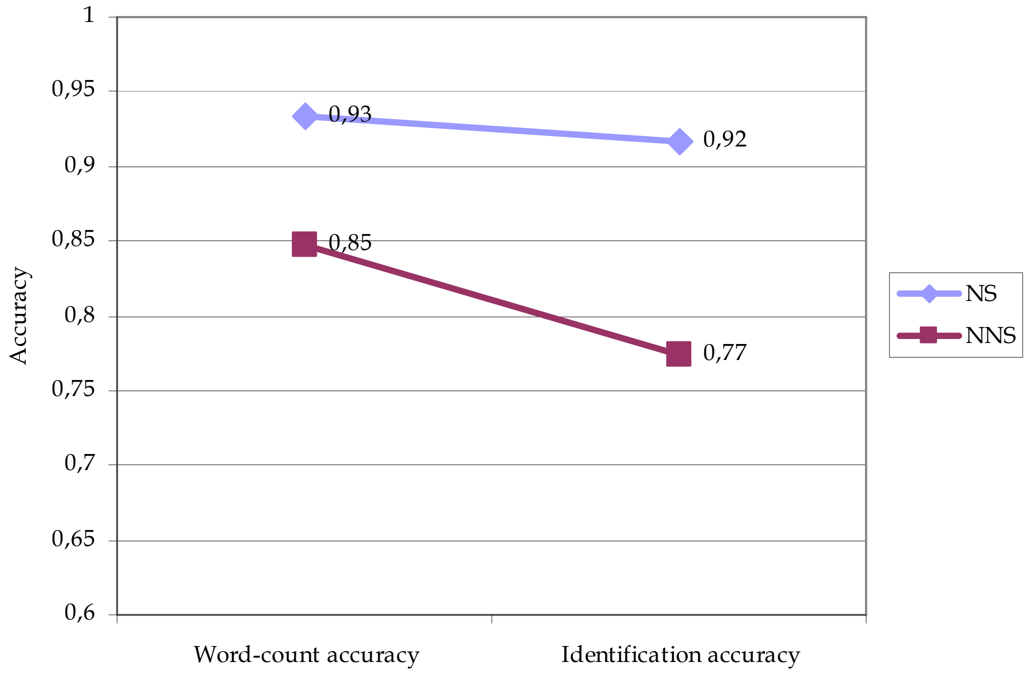

Figure 4 Interaction effect between nativeness and task type (both reduced and unreduced items)

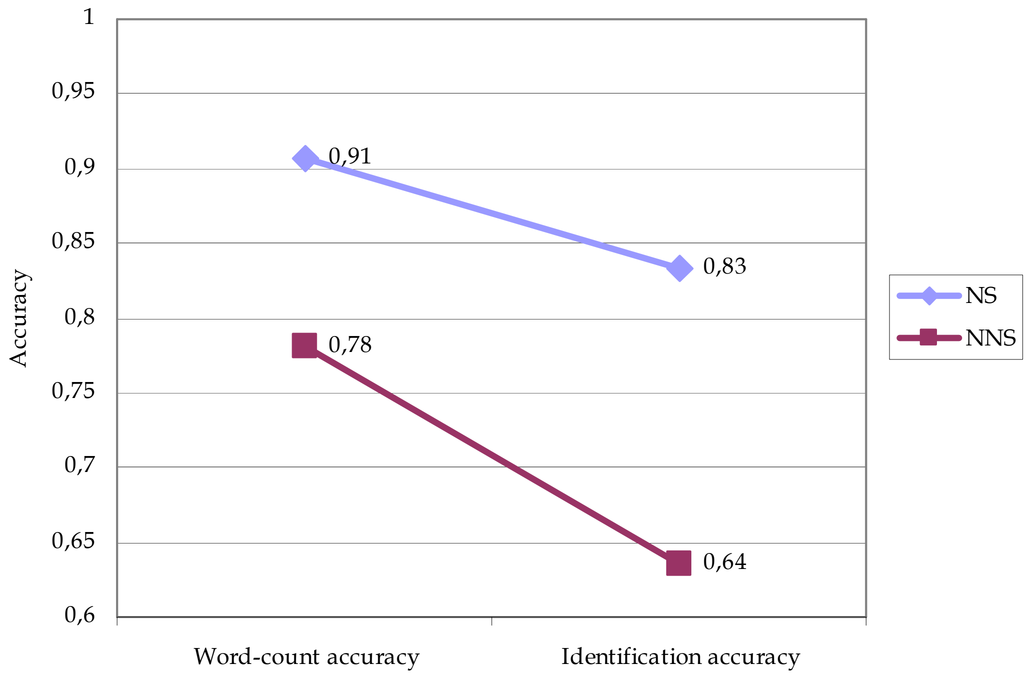

Figure 5 No interaction effect between nativeness and task type (only reduced items)

6.4 Research question 2

The second research question was whether there is a particular type of reduction that is difficult to segment for native and non-native speakers. The factorial design included nativeness (2 levels: native, non-native), task type (2 levels: word-count, identification), and reduction category (3 levels: vowel, consonant, extreme reduction) as independent variables. As with the first research question two GLMs (*general linear models*) were performed, first on the speed measure and subsequently on the accuracy measures. Given that the reduction categories only applied to reduced items, unreduced items have been left out of the analysis. Results relating to segmentation speed will be presented first, followed by those relating to segmentation accuracy. Table 5 shows mean reaction times for each of the three reduction types.

Table 5 Mean reaction time (in ms) and SD of segmentation speed for each reduction type (N=29)

	NS		NNS	
	Mean	SD	Mean	SD
Vowel	2090,92	331,41	2591,20	632,89
Consonant	2000,42	335,91	2490,88	663,32
Extreme	2586,22	541,17	3218,56	1000,30

A significant main effect was found for nativeness ($F_{(1,27)}$ = 6,45, p < ,05). The finding from the first research question that NS are faster at segmenting than NNS was replicated here. There was also a significant main effect for reduction type ($F_{(2,54)}$ = 43,58, p < ,05). According to the results (see also Figure 6), responses to extreme reductions were the slowest. Pairwise comparisons revealed that this type of reduction differed significantly in reaction time from consonant and vowel reduction. There was no significant difference between consonant and vowel reduction. This applied to both native and non-native listeners. No interaction effect was found between nativeness and reduction type; NS and NNS appeared to be equally sensitive to the different reduction types in segmentation speed. The hierarchy from the fastest to the slowest reduction type to segment, was thus the same for both NS and NNS, and looked as follows: consonant reduction / vowel reduction > extreme reduction.

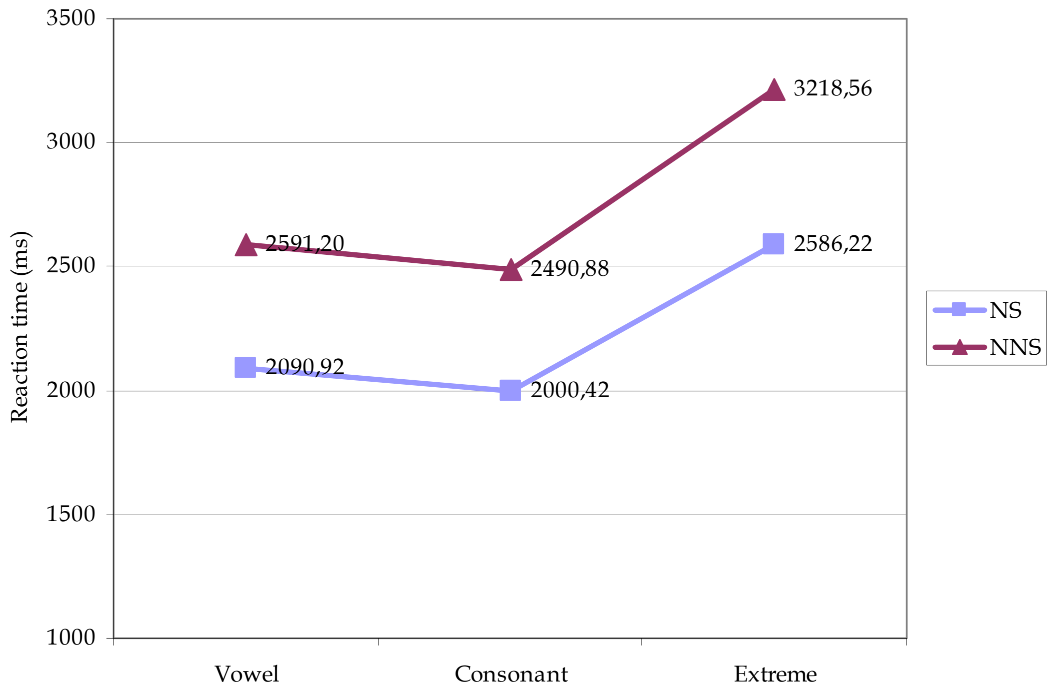

Figure 6 Mean reaction time (in ms) for each reduction type

The same analysis was done with the accuracy data. The descriptive statistics of the two accuracy measures is presented in Table 6.

Table 6 Mean accuracy and SD of segmentation accuracy for each reduction type (N=29)

	Word-count task				Identification task			
	NS		NNS		NS		NNS	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Vowel	,94	,12	,88	,11	,95	,06	,81	,16
Consonant	,94	,07	,87	,07	,93	,04	,77	,15
Extreme	,92	,07	,79	,09	,87	,08	,73	,24

It was confirmed once again that NS were more accurate than NNS ($F_{(1,27)}$ = 13,01, p < ,05) and that the word-count task was done more accurately than the identification task ($F_{(1,27)}$ = 6,12, p < ,05). As with the speed measure, a significant main effect for reduction type was also found with the accuracy measures ($F_{(2,54)}$ = 12,24, p < ,05). Again, extreme reductions appeared to be the most difficult of the three reduction types (see also Figure 7). Pairwise comparisons showed that the difference between extreme reduction and consonant reduction was significant for NS and approaching significance for NNS. In neither groups a significant difference was found between vowel and consonant reduction. Based on the accuracy scores, the hierarchy would look like this: vowel reduction / consonant reduction > extreme reduction. No interaction effects were found among the three variables of nativeness, reduction type and task type. This included the interaction effect between nativeness and task type that was reported for the first research question; it was not visible in this second research question.

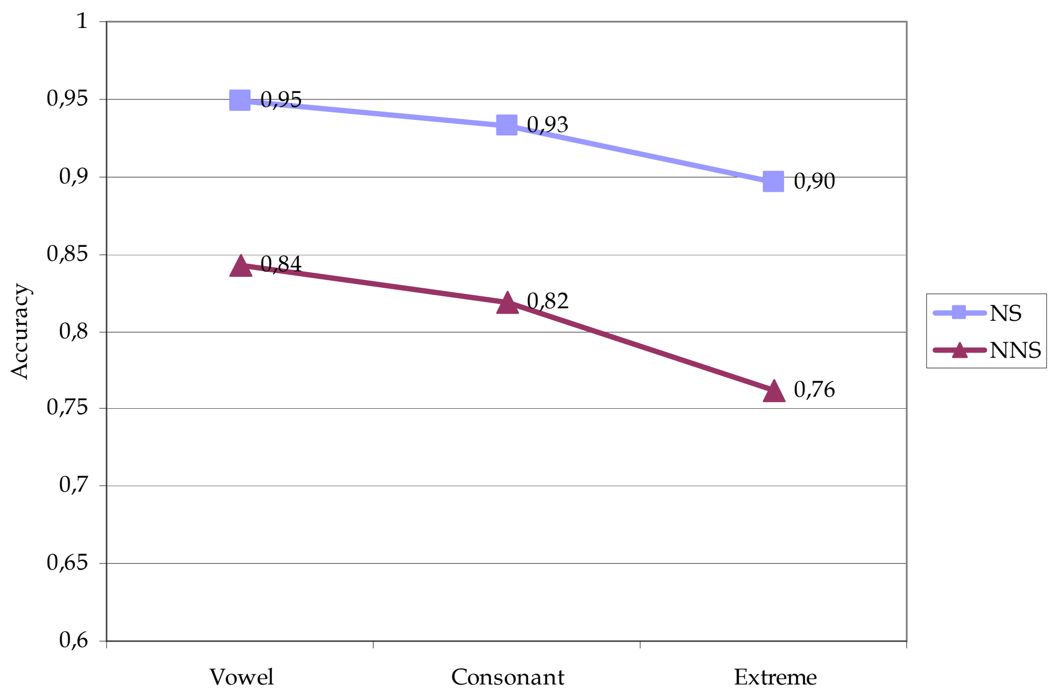

Figure 7 Mean segmentation accuracy for each reduction type calculated over both accuracy measures

7 Discussion

This experimental study used two segmentation tasks to assess the difference between native and non-native speakers of Dutch in the segmentation of reduced and unreduced Dutch forms. The segmentation task consisted of two subtasks; one subtask (i.e. word-count task) measured speed and accuracy of the number of words participants heard, and the identification subtask measured accuracy of which words were heard. Correlational analyses showed a strong relationship between both accuracy measures. This suggests that both subtasks measure the same construct, which is assumed to be segmentation. Word-count accuracy does not correlate with word-count speed, but this may be explained as follows. Responses on the word-count task were given under time pressure, both responses on word-count speed and on word-count accuracy. In the word-count accuracy measure, there may have been a trade-off between accuracy and time pressure. Some responses may have been given very fast because participants were very certain of their response. But other responses may have been compromised by the time pressure of the task; wanting to respond as quickly as possible, may also have led to inaccuracy. This trade-off situation may explain the relatively small correlation coefficient. The fourth measure in this study – i.e. identification count – correlates to all other measures. The predictive power makes this variable a strong measure, provided we can be certain this measure does indeed measure segmentation.

The analysis furthermore revealed no correlation between working memory and any of the segmentation measures. On the one hand, this may be a positive outcome; if there is no relationship between the segmentation tasks and working memory, then the tasks may be assumed to not make too much an appeal on working memory. There would thus be no reason to reject the tasks on the basis of being a burden on the phonological loop. On the other hand, it is highly surprising that no correlation at all is found with working memory. Especially the word-count task was expected to make an appeal on the phonological loop. A possible explanation for the lack of correlation between the working memory task and all other tasks, may involve the distinction between recall and recognition. In memory retrieval, recall is the act of actively bringing information from memory. Recognition, however, only bears on recognizing information stored in memory. The working memory task in this study called upon recognition; participants had to recognize if two series of non-words were identical or not. They did not have to actively recall the non-words. The segmentation tasks did call upon recall skills, as participants had to recall the (number of) words back from their working memory. This discrepancy in the nature of memory retrieval may explain why no correlation was found.

Results showed clear advantage of native speakers over non-native speakers in both segmentation speed and accuracy, thus confirming the hypotheses. Non-native speakers of Dutch are slower and less accurate in segmenting unreduced and reduced Dutch forms. This may be explained by a possible inability to effectively use segmentation cues or a lack of knowledge of Dutch reduction conventions. Furthermore, and also according to expectation, unreduced items yield faster and more accurate performance than reduced items. But the effect of reduction in segmentation accuracy appears to be larger in the non-native group. Reconstructive processes have been found to be at play when native listeners perceive reduced forms (Ernestus et al., 2002; Kemps et al., 2004). Apparently, non-native listeners are less able to make use of these reconstructive processes, as they are more sensitive to reduction than native listeners.

The main effect of task type revealed that counting words is easier – i.e. results in more accurate responses – than identifying words. However, a critical note needs to be placed regarding the two accuracy measures; that is, they are very different in nature. For one, they measure different things. While the word-count task merely requires counting the words, the identification task necessitates their lexical identification. Secondly, word-count responses were given under time pressure, as reaction times were measured. In the identification task participants were free to take as long as they needed to complete the task. And finally, the nature of the responses is distinct. The word-count task was multiple-choice (i.e. 1, 2, 3, 4, or 5 words), while the identification task was not bound by a given number of response possibilities. In the word-count task chance may have played a role, but this was not the case in the identification task. Comparing the two may consequently be illegitimate given all these differences. That being said, the interaction effect between nativeness and task type may still be quite a notable finding. Non-native speakers appear to be significantly less accurate in the identification task than in the word-count task; they are better at counting words, than at identifying them. This finding may be relevant for the acquisition of segmentation skills. The results may represent a phase, where listeners are able to recognize distinct units in the speech stream, yet not completely able to correctly identify them. Their bottom-up processing seems to enable segmentation into distinct units, but sufficient lexical knowledge (i.e. top-down processing) may still be lacking to accurately map lexical items onto the segmented units. This may in turn suggest a gradual process in the acquisition of L2 segmentation skills, starting with non-native listeners not being able to segment the L2 speech stream at all. In a following stage they would be able to use sublexical cues to recognize distinct units in the speech. And in the final stage they would also be capable of correctly identifying the words. This final stage may only be reached once enough lexical knowledge is acquired to enable top-down processes – as opposed to only bottom-up – and reconstructive processes. A possible problem with this line of reasoning is the fact that all lexical items in the segmentation tasks were controlled for frequency of use. Only highly frequent words were selected to ensure understanding of the items by all participants, including the non-native speakers. This is in contradiction with the thought that the non-natives lack in lexical knowledge to accurately map lexical items onto the segmented units. Perhaps some of the selected items were complex on a level other than word frequency, for which the non-native speakers' knowledge was not adequate. The interaction effect between reduction and task can be explained by the hierarchy of segmentation strategies (Mattys et al., 2005). According to this hierarchy, speech segmentation is lexically driven when all cues are optimally available, which is the case in the unreduced forms. However, reduction results in lexical information being impoverished. It is therefore understandable that in the reduced condition performance on the identification task – which requires lexical knowledge for the accurate identification of the words – will decrease more than when lexical information is not necessarily needed for good performance (i.e. in the word-count task). And this is exactly what the interaction effect shows; the identification task is more sensitive to differences in reduction than the word-count task.

The second research question pertained to reduction categories. Only reduced items were included in the analysis, given that reduction types only apply to items that are in fact reduced. The different data set may explain why an interaction effect between nativeness and task type was found for the first research question, but not for the second. Taking together both reduced and unreduced items (as in research question 1, see also Figure 4), it appeared that accuracy scores of NS and NNS varied more on the identification task than on the word-count task. But when only reduced items were considered (as in the second research question, see also Figure 5) there was no interaction effect between nativeness and task type; accuracy scores of NS and NNS varied similarly on both tasks. The larger deviation in accuracy scores between NS and NNS in the identification task thus disappeared when only reduced items were taken into consideration. A possible explanation may be that accurate identification – more so than accurate counting – requires accurate comprehension. When the items were reduced, comprehension may have been compromised, for both NS and NNS. As a consequence their scores deviated less from one another than when both reduced and unreduced items were included in the analysis.

The extreme reduction proved to be most difficult to segment, both in speed and in accuracy. No differences were found between the other two reduction types. This applies to both native and non-native listeners. The finding that extreme reduction is more difficult to segment than consonant and vowel reduction is not surprising. The extremity of the reduction namely lies in the featuring of both vowel and consonant reduction. The extremely reduced forms – as the name also suggests – have more missing segments than either vowel or segments reduction alone and are thus harder to comprehend. This finding may not be striking, but it does show that although highly frequent, the reduced forms have not replaced the unreduced equivalent as the default form. This was conceivable because reductions are only permitted if they do not hinder comprehension. Deleting multiple speech segments – as is the case in extreme reduction – heightens the risk of incomprehension. For extreme reductions to be permissible, they have to be highly recognizable, thereby maybe capable of surpassing the unreduced form in recognizability. Findings from the study did not support this hypothesis. There appeared to be no difference in the segmentation of either vowel or consonant reduction. Both seemed to be equally important for the identification of word boundaries.

8 Conclusion

In this thesis I have attempted to examine the effect of reduction processes on how native and non-native listeners of Dutch segment oral language. Segmenting auditory input is aided by segmentation cues, both lexical and sublexical. Non-native listeners are less able in using these cues as efficiently as native listeners. In conversational speech variations may arise in the speech signal as a result of reduction processes. I have constructed two segmentation tasks, including both unreduced and reduced Dutch forms, to examine the difference between native and non-native speakers of Dutch in the segmentation of reduced and unreduced forms. By incorporating three types of reduction I could also look into the possible classification of these reduction types into degree of difficulty.

Performances of 29 participants – 15 native and 14 non-native speakers of Dutch – were analyzed. Native speakers were both faster and more accurate than non-native speakers in segmenting unreduced and reduced speech. A possible explanation for this finding is the inability of non-native speakers to effectively use Dutch segmentation cues or the lack of knowledge of Dutch reduction conventions. The effect of reduction is larger in the group of non-natives, which may suggest that they are less able to make use of reconstructive processes, as compared to the group of native speakers.

These findings were to be expected and may be stating the obvious. A more notable finding was the interaction between nativeness and task type. Non-native speakers were better at counting words than at identifying them. This suggests that they are able to segment the auditory input into distinct units (i.e. counting the words), but are less skilled in mapping lexical items onto the segmented units (i.e. identifying the words). This may represent a phase in the – what seems to be a –

gradual process in the acquisition of segmentation skills. Time and practical limitation have not allowed for a larger sample size. Further research would benefit from testing more participants. If distinction can be made in several proficiency levels within the sample, then gradation of the segmentation process may be examined more extensively. Then it would perhaps be possible to map all the consecutive stages in the acquisition of segmentation skills.

Of the three reduction types that were incorporated in the segmentation task extreme reduction proved the most difficult to segment; responses on extremely reduced items were the slowest and least accurate. Vowel reduction and consonant reduction yielded the same degree of difficulty. The proposed hierarchy in reduction types is the following: consonant reduction / vowel reduction > extreme reduction.

I have tried to be as methodologically correct in constructing the segmentation task and carrying out this research. There are, nonetheless, a number of possible limitations. Constructing a segmentation task – or any other experimental task – requires an elaborate validation study and without the necessary validation the conclusions of this research remain uncertain. It should be noted, however, that extensive validation was not feasible because the combined aspects of segmentation and reduction have not been studied before. Therefore, there were no previous measures to validate the currently used tasks with. This study thus constitutes a stepping stone toward bridging the gap between studies of segmentation and studies of reduction processes. Future research could benefit from this first step in the testing of segmentation and reduction.

Bibliography

Altenberg, E. P. (2005). The perception of word boundaries in a second language. *Second Language Research, 21*(4), 325-358.

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

Bradlow, A. R., & Bent, T. (2002). The clear speech effect for non-native listeners. *The Journal of the Acoustical Society of America, 112*(1), 272-284.

Carroll, S. E. (2004). Segmentation: Learning how to 'hear words' in the L2 speech stream. *Transactions of the Philological Society, 102*(2), 227-254.

Coussé, E., Gillis, S., & Kloots, H. (2007). Verkort, verdoft, verdwenen: Vocaalreductie in het Corpus Gesproken Nederlands. *Nederlandse Taalkunde, 12*(2), 109-138.

Cutler, A. (1996). Prosody and the word boundary problem. In J. Morgan & K. Demuth (Eds.), *Signal so syntax: Bootstrapping from speech to grammar in early acquisition*. Mahwah, NJ: Erlbaum.

Cutler, A. (2002). Native listeners. *European Review, 10*(1), 27-41.

Cutler, A., & Clifton, C., Jr. (1999). Comprehending spoken language: A blueprint of the listener. In C. M. Brown & P. Hagoort (Eds.), *The neurocognition of language*. Oxford: Oxford University Press.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology, 14*(1), 113-121.

Davis, M. H. (2000). *Lexical segmentation in spoken word recognition*. London: Doctoral dissertation, Birkbeck College, University of London.

Ernestus, M. (2000). *Voice assimilation and segment reduction in casual Dutch : A corpus-based study of the phonology-phonetics interface*. Utrecht: Doctoral dissertation, Landelijk Onderzoeksschool Taalwetenschappen.

Ernestus, M., & Baayen, R. H. (2007). *The comprehension of acoustically reduced morphologically complex words: The roles of deletion, duration, and frequency of occurrence.* Paper presented at the 16th International Congress of Phonetic Sciences, Saarbrücken, Germany.

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

Field, J. (2003). Promoting perception: Lexical segmentation in L2 listening. *ELT Journal, 57*(4), 325-334.

Gillis, S., Kloots, H., & Swerts, M. (2004). De uitspraak van het possessivum 'mijn' in spontaan Standaardnederlands. In J. de Caluwé, G. de Schutter, M. Devos & J. van Keymeulen (Eds.), *Taeldeman, man van de taal, schatbewaarder van de taal*. Gent: Academia Press.

Godfrey, J. J., Holliman, E. C., & McDaniel, J. (1992). *SWITCHBOARD: Telephone speech corpus for research and development.* Paper presented at the IEEE International Conference on Acoustics, Speech, and Signal Processing, San Fransisco.

Johnson, K. (2004). Massive reduction in conversational American English. In K. Yoneyama & K. Meakawa (Eds.), *Spontaneous speech: Data and analysis*. Tokyo: The National Institute for Japanese Language.

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

Keating, P. A. (1997). *Word-level phonetic variation in large speech corpora*. Paper presented at the The word as a phonetic unit.

Kemps, R., Ernestus, M., Schreuder, R., & Baayen, R. H. (2004). Processing reduced word forms: The suffix restoration effect. *Brain and Language, 90*, 117-127.

Keune, K., Ernestus, M., Van Hout, R., & Baayen, R. H. (2005). Variation in Dutch: From written MOGELIJK to spoken MOK. *Corpus Linguistics and Linguistic Theory, 1*(2), 183-223.

Kloots, H., De Schutter, G., Gillis, S., & Swerts, M. (2003). Vocaalreductie in spontaan gesproken Standaardnederlands: Een verkennende studie. In T. Koole, J. Nortier & B. Tahitu (Eds.), *Artikelen van de vierde sociolinguïstische conferentie*. Delft: Eburon.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modeling*. Dordrecht: Kluwer.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General, 134*(4), 477-500.

McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language, 39*(1), 21-46.

McQueen, J. M., & Cutler, A. (2001). Spoken word access processes: An introduction. *Language and Cognitive Processes, 16*(5/6), 469-490.

Mitterer, H., & Blomert, L. (2003). Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception & Psychophysics, 65*(6), 956-969.

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

Nakatani, L. H., & Schaffer, J. A. (1978). Hearing "words" without words: Prosodic cues for word perception. *The Journal of the Acoustical Society of America, 63*(1), 234-245.

Norris, D., McQueen, J. M., Cutler, A., & Butterfield, S. (1997). The Possible-Word Constraint in the segmentation of continuous speech. *Cognitive Psychology, 34*, 191-243.

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

Sanders, L. D., Neville, H. J., & Woldorff, M. G. (2002). Speech segmentation by native and non-native speakers: The use of lexical, syntactic, and stress-pattern cues. *Journal of Speech, Language, and Hearing Research, 45*(3), 519-530.

Van de Weijer, J. (1999). *Language input for word discovery*. Nijmegen: Doctoral dissertation, Katholieke Universiteit Nijmegen.

Van der Lugt, A. (1999). *From speech to words*. Nijmegen: Doctoral dissertation, Katholieke Universiteit Nijmegen.

Studies in listening proficiency (STILIS). Retrieved June 6, 2008, from http://www.hum.uva.nl/Stilis

Corpus Gesproken Nederlands. Retrieved August 15, 2008, from http://lands.let.kun.nl/cgn/ehome.htm

Appendix I: Background information of all participants

Non-native speakers										
Participant number	Sex	Age	First language	Exposure to Dutch (in years)	Self-rating of proficiency in Dutch					
					Speaking	Writing	Reading	Listening	Grammar	Vocabulary
02	F	27	Hebrew	16	5	5	5	5	5	5
04	F	29	Multani	25	5	5	5	5	5	5
06	M	25	Hebrew	5	4	3	5	4	5	4
08	F	33	German	6	4	3	4	4	4	4
10	F	33	Spanish	3	4	3	5	3	2	3
12	F	34	Turkish	1	3	1	4	3	1	4
14	M	22	Italian	4	4	2	3	4	2	3
16	F	27	German	2	4	3	5	4	3	4
18	F	33	English	4	4	4	5	4	3	4
20	F	23	English	5	4	3	3	5	4	2
22	M	32	Berber	1	2	2	2	2	2	2
24	F	29	Hebrew	17	3	3	4	3	3	3
26	F	34	English	5	3	4	5	5	3	4
28	F	26	English	2	2	2	4	3	2	3
30	F	28	Turkish	10	4	3	5	5	4	4

Native speakers		
Participant number	Sex	Age
03	F	22
05	F	25
07	F	21
09	F	29
11	M	25
13	F	24
15	F	24
17	M	23
19	F	22
21	F	25
23	F	22
25	F	24
27	F	25
29	F	21
31	F	21

Appendix II: Items segmentation task

Item number	Item	List 1	List 2	Context (number of words)	Type reduction
1	bij mij	reduced	full	2	vowel
3	het hek	reduced	full	2	vowel
5	hij is lang	reduced	full	3	vowel
8	ga ik ook	reduced	full	3	vowel
9	ik was ontzettend blij	reduced	full	4	vowel
10	verloren heeft	reduced	full	2	vowel
11	wat er gezegd is	full	reduced	4	vowel
12	heeft namelijk veel	full	reduced	3	vowel
13	mijn plan	full	reduced	2	vowel
14	bedoel ik niet	full	reduced	3	vowel
15	laat ze altijd gaan	full	reduced	4	vowel
16	erg makkelijk voor	full	reduced	3	vowel
17	vorige nog	full	reduced	2	vowel
21	en vast zit	reduced	full	3	consonant
22	je moet eerst gaan	reduced	full	4	consonant
23	denkt men	reduced	full	2	consonant
24	die lijkt me	reduced	full	3	consonant
25	moet je anders doen	reduced	full	4	consonant
26	over de	reduced	full	2	consonant
27	in je leven wil	reduced	full	4	consonant
28	ik heb dat	reduced	full	3	consonant
29	toch wel	reduced	full	2	consonant
30	zit nog wel	reduced	full	3	consonant
33	dat weet	full	reduced	2	consonant
34	het zit precies op	full	reduced	4	consonant
35	zeg maar	full	reduced	2	consonant
36	dan gaan ze	full	reduced	3	consonant
37	wil je daarna met	full	reduced	4	consonant
38	dat vind ik	full	reduced	3	consonant
40	moet je	full	reduced	2	consonant

41	gegeven moment was	reduced	full	3	extreme
42	ik kan volgende keer	reduced	full	4	extreme
43	volgens mij	reduced	full	2	extreme
44	hoeft dan allemaal niet	reduced	full	4	extreme
47	of het mogelijk is	reduced	full	4	extreme
49	is ongeveer wat	reduced	full	3	extreme
50	ik weet eigenlijk ook	reduced	full	4	extreme
51	is natuurlijk al	full	reduced	3	extreme
52	is niet duidelijk wie	full	reduced	4	extreme
54	komt redelijk vaak	full	reduced	3	extreme
55	behoorlijk veel	full	reduced	2	extreme
56	te persoonlijk om	full	reduced	3	extreme
57	bepaalde tijd	full	reduced	2	extreme
58	ik had gewoon iets	full	reduced	4	extreme
60	in ieder geval fijn	full	reduced	4	Extreme

Excluded items					
2	het bijvoorbeeld met	reduced	full	3	vowel
4	dan sta ik weer	reduced	full	4	vowel
6	bij de politie in	reduced	full	4	vowel
7	me ook	reduced	full	2	vowel
18	toen hij later het	full	reduced	4	vowel
19	beter een	full	reduced	2	vowel
31	als in	full	reduced	2	consonant
39	meen je niet echt	full	reduced	4	consonant
45	weet helemaal van	reduced	full	3	extreme
46	daarom ook	reduced	full	2	extreme
53	dadelijk geen	full	reduced	2	extreme
59	was inderdaad heel	full	reduced	3	extreme

Appendix III: Questionnaire

Vragenlijst

1 Wat is uw sekse?

O Vrouw

O Man

2 Wat is uw geboortedatum en –jaar?

____

3 In welk land bent u geboren?

____

4 Wat is uw moedertaal?

____

5 Indien uw moedertaal *geen* Nederlands is:

a. Op welke leeftijd bent u begonnen met het leren van het Nederlands?

____

b. Waar heb je het Nederlands geleerd?

O Zelfstudie

O Cursus, namelijk (instituut + niveau):________________________

O Anders, namelijk: ______________________________________

c. Hoe goed vindt u uw Nederlands op dit moment? Omcirkel op een schaal van 1 tot 5 wat voor u van toepassing is:

Spreken	heel slecht	1	2	3	4	5	heel goed
Schrijven	heel slecht	1	2	3	4	5	heel goed
Lezen	heel slecht	1	2	3	4	5	heel goed
Luisteren	heel slecht	1	2	3	4	5	heel goed
Grammatica	heel slecht	1	2	3	4	5	heel goed
Woordenschat	heel slecht	1	2	3	4	5	heel goed

6 Bent u doof aan één oor of hebt u op dit moment last van gehoorproblemen? Zo ja, hoe ernstig zijn deze problemen?

____