Perception of the tone contrast in East Limburgian dialects
Perception of the tone contrast in East Limburgian dialects

Een wetenschappelijke proeve op het gebied van de Letteren

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. mr. S.C.J.J. Kortmann,
volgens besluit van het College van Decanen
in het openbaar te verdedigen op maandag 1 december 2008
om 15:30 uur precies

door
Rachel Agnès Fournier
geboren op 3 mei 1971
te Sion, Zwitserland
Promotor: Prof. dr. C.H.M Gussenhoven
Copromotor: Prof. dr. M.G.J. Swerts (Universiteit van Tilburg)

Manuscriptcommissie:

Prof. dr. J. Peters (voorzitter; Radboud Universiteit Nijmegen en Universität Oldenburg, Duitsland)
Prof. dr. M. Grice (Universität zu Köln, Duitsland)
Prof. dr. M. van Oostendorp (Meertens Instituut)
Acknowledgements

Even the brightest scientists need, from time to time, a little help from their friends, colleagues and supervisors. Of course I needed far more than that. Luckily enough, there were a lot of people around to keep my engine running, and I am glad that I finally have a chance to thank them.

First of all, my gratitude goes to my supervisor Carlos Gussenhoven for his constant presence, support and most valuable input. I can’t remember a moment in the last years when he didn’t answer my emails almost instantly, no matter the circumstances and the degree of silliness of my questions. He also trusted me with an experiment that I wouldn’t have dreamed of before starting my PhD, and especially during this stage of the project, I learned a lot by watching his adventurous mind at work. Many thanks to Marc Swerts for jumping in as a supervisor at a moment where I could really use his amazing talent of detecting flaws in texts or experimental designs, and giving just the right hints to reparse them efficiently. I am indebted to Peter Hagoort for providing me access to the Donders facilities, and always finding the time for meetings. Ole Jensen showed me around in the MEG world and also proved that you can do serious work without losing your sense of humor. Thank you Ole, but also Bram and Paul, for the good mood and the precious help! I also wish to thank all the native and non-native speakers who participated in the Roermond and Venlo experiments, in particular Lianne Verheggen, Yo van Knippenberg and Nicole Verberkt who actively helped finding subjects and/or scrupulously checked the correctness of the corpora.

I am grateful to Jörg Peters whose sense of detail proved extremely useful, especially near the end of the project, when his own time schedule wasn’t supposed to allow it. Roeland van Hout was prompt to compensate for administrative delays by providing me with a temporary job and a highly interesting workplace amongst piles of science-in-progress. This was a very nice prelude, thanks a lot! I also wish to acknowledge Yiya Chen who cheered me up at critical moments. At one of these moments, Aditi Lahiri gave me the opportunity to spend very constructive days in a fairy-tale castle. The magic worked, thank you for that. Febe de Wet and Ambra Neri, what we created during our weekly lunches must have been magic too since it survived so much time and distance. Many thanks for all your hearty messages, and Ambra, for being my paranimf? Judith Hanssen, thank you for accepting this role too, and for proof-reading the most critical parts of my thesis. I am happy to have stayed around long enough to get to know you better.
The Arnhems Slavisch Koor provided me with the most pleasant taste of Dutch gezelligheid. Singing with you is a treat, and the breaks are, Monday after Monday (since there is always something to celebrate), a sweet blessing for sore throats and hearts. Thank you all for your warm welcome, it is a great honour to be part of such a manifold and tolerant company. Thank you too, Jan and Lucy Moerland, Janine and Peter, for making space for me in your loving family. I truly appreciate your patience and respect. In the country of 1001 memories, where the honey flows on the mountain and gives perfect excuses to meet again, I feel very lucky to have a large and effusive family (quite a manifold tribe too!), including the Weber extension. A thousand thanks and hugs to Marcelle Fournier for being an incredibly energetic and generous mother and grandmother, and to Jérôme and Claire Fournier as well as Roland Sauter for their discrete but heartwarming unconditional support. Perry, this thesis is dedicated to you. Not only did you carefully read and comment my first chapter without complaining, but you also have been putting up with my moods and bad habits for years and still manage to keep your irresistible smile. I'm not surprised at all that you became such an amazing father too. Ivo, sweet little Ivo, maybe one day I will be able to put into words the overwhelming feelings that you revealed in me. For now, let me just take you on my knees and kiss your silky cheeks.
# Table of Contents

ACKNOWLEDGEMENTS 5

TABLE OF CONTENTS 7

PREFACE 11

1 INTRODUCTION 13
1.1 The tone contrast in East Limburgian ................................................................. 13
1.1.1 General description ..................................................................................... 13
1.1.2 Autosegmental Metrical theory .................................................................. 17
1.1.3 The Roermond grammar ............................................................................. 20
1.1.3.1 Building blocks and phonological adjustments ................................... 20
1.1.3.2 Contour inventory ............................................................................... 26
1.1.4 The Venlo grammar .................................................................................... 29
1.1.4.1 Building blocks and phonological adjustments ................................... 29
1.1.4.2 Contour inventory and an alternative account..................................... 34
1.2 Research on tone perception ............................................................................. 37
1.2.1 Tone in context ........................................................................................... 37
1.2.1.1 Tonal coarticulation and contrast neutralization ................................. 37
1.2.1.2 The tone-intonation interaction ........................................................... 39
1.2.2 Acoustic cues .............................................................................................. 40
1.2.3 Language experience .................................................................................. 42
1.2.4 Categorical perception and (neuro)cognition .............................................. 44

2 PERCEIVING WORD PROSODIC CONTRASTS AS A FUNCTION OF SENTENCE PROSODY IN TWO DUTCH LIMBURGIAN DIALECTS 47
2.1 Introduction ........................................................................................................ 47
2.2 Method ............................................................................................................... 50
2.2.1 Target words and carrier sentences............................................................. 50
2.2.2 Experimental tapes ...................................................................................... 52
2.2.3 Judges and procedure .................................................................................. 53
2.3 Acoustic analysis ................................................................................................ 53
2.3.1 \( f_0 \) contours ........................................................................................... 54
2.3.1.1 Roermond............................................................................................ 54
2.3.1.2 Weert .................................................................................................. 57
2.3.2 Duration ...................................................................................................... 58
2.4. Perception experiment: Results ......................................................................... 59
2.4.1. Statistical analysis ................................................................................... 59
2.4.2 Discussion ................................................................................................... 64
2.5 Summary and conclusion................................................................................... 67
4.7.1 Sentences used in the training session ......................................................125
4.7.2 Words used the subjective distance experiment ........................................126
  4.7.2.1 Preparatory items (in order of presentation) .....................................126
  4.7.2.2 Answer sheet (extract) .................................................................126

5  LATERALIZATION OF TONAL AND INTONATIONAL PITCH PROCESSING: AN
MEG STUDY 127
5.1 Introduction .............................................................................................127
5.2 Method .......................................................................................................131
  5.2.1 Subjects .............................................................................................131
  5.2.2 Word stimuli .....................................................................................131
  5.2.3 Procedures .........................................................................................134
  5.2.4 Data acquisition and analysis .............................................................135
5.3 Results .......................................................................................................136
5.4 Discussion .................................................................................................141

6  SUMMARY AND CONCLUSIONS 145

7  REFERENCES 153

SAMENVATTING (SUMMARY IN DUTCH) 165

CURRICULUM VITAE 173
Preface

Monotonous speech is not only as unsettling as a face without expression, it is also a major impairment to language understanding. Deprive speech of its melody, and you will find yourself with a disembodied heart which will not function properly unless you bandage it with many more words and gestures. Speech melody is a highly efficient tool that allows for the expression of feelings, social or regional identity, discourse and word meaning. Of course, not all of these elements will necessarily be conveyed or detectable in all conditions. Their presence and perceivability depend on many factors, such as the speaker’s intentions and knowledge, and the listener’s experience, in addition to language-internal (i.e. grammatical) aspects. The present thesis will focus on the listener’s perspective, and the linguistic factors that are likely to influence the perception of pitch when it is used to distinguish words. Such a linguistic use of pitch at the word level is called lexical tone.

Lexical tones are found in more than half of the world’s languages, many of them spoken in Africa (e.g. Khoisan or Bantu varieties) and East Asia (e.g. most Chinese dialects, Vietnamese). Europe also has a few tonal islands. In one of them, we find the two language varieties examined in this thesis, the Dutch dialects of Roermond and Venlo. These dialects have a binary tone contrast referred to as Accent 1 vs. Accent 2, and they are of great typological interest due to the number of different contour shapes that are used to encode the two tones. An Accent 1 word pronounced in the middle of a sentence does not have the same melody as one pronounced at the end of the sentence; its shape also depends on whether it is accented (focused) or not, and as if this were not enough, the number of different contours is multiplied by the number of intonation contours that can be used in the dialect in question. For instance, the Roermond dialect has different sentence melodies for expressing statements and yes/no questions, which will result in two distinct sets of contour shapes for Accent 1 and Accent 2. It is quite exceptional to find a language or language family in which tonal contrasts vary drastically along so many dimensions. Such a special status of East Limburgian (and some of their neighbouring) dialects has given rise to a series of studies in the past decade. However, these studies have concerned themselves with the production of the tone contrast and have paid little attention to its perception. Our study intended to fill this gap. This was done by carrying out a series of experiments, all sharing the underlying research question on how lexical tones are perceived by native speakers of the dialects under consideration.

The first step towards answering our research question was to review the relevant literature on structural aspects of the Roermond and the Venlo dialect, as well as on earlier studies dedicated to tone perception. This review is given in chapter 1. Chapters 2, 3 and 4 report on three experiments with similar designs,
which involved the recording of minimal pairs representing Accent 1 and Accent 2 in different contexts (as determined by the factors of variation found in production studies), randomizing the resulting stimuli, and asking native listeners to identify the tone used in each stimulus. This general pattern was enriched with additional questions and tests in chapters 2 and 4.

In chapter 2, we describe an experiment designed to investigate the perception of the Roermond tonal contrasts as compared to a durational contrast found in the neighbouring dialect of Weert. The comparison between the Roermond and the Weert dialect allowed us to assess the relative importance of two of the factors of variation mentioned above, focus and position in the sentence, for the perception of pitch- vs. duration-based lexical contrasts. Chapter 3 is exclusively devoted to the dialect of Roermond and adds the dimension of intonation to the set of factors responsible for changes in contour shapes. Some unexpected results raised the issue on how perceptual salience is related to the identification of lexical tones. This issue was addressed more directly in our next experiment, reported in chapter 4. Based on the Venlo contrasts, this experiment compared the recognition performances of native listeners with their judgement on the phonetic distances perceived between members of the minimal pairs. These two dimensions were also compared with dissimilarity judgements made by non-native speakers as well as with various automatic measures of acoustic distance based on the same stimulus pairs.

Chapter 5 presents a new perspective on our research question, by investigating the neurocognitive processes underlying tone perception. This experiment represents a shift in the methods (by using an MEG scanner instead of a headphone and an answer sheet) as well as in the way we look at tone and intonation. The contextual variation studied in the previous experiments is still at the center of our interest, but this time, we treat tone and intonation as two distinct functional entities instead of considering intonation as a factor of variation for the production and perception of tone. More specifically, rather than asking ourselves whether the tone contrast is perceived (or processed) differently when it is pronounced as a statement or as a question, we want to determine whether the tone contrast is processed differently than the contrast between a statement and a question. The thesis concludes with chapter 6, which summarizes our research and discusses its results.
1 Introduction

1.1 The tone contrast in East Limburgian

1.1.1 General description

When someone is asked to provide an example of lexical tones, the audience is very likely to hear, once again, about the difference between a mother and a horse, or swearing and hemp. Mandarin Chinese is indeed a typical instance of a language that makes extensive use of lexical tones. Most syllables have their own tonal specification, and little explanation is necessary to convince us how crucial pitch contours are for understanding this language. Asia and Africa boast many other examples of such languages. In Europe, lexical tone is distributed less generously, both with respect to the proportion of languages and language varieties that have it and to its importance in these languages. The tonal contrast in the Swedish-Norwegian dialect continuum, Lithuanian, Serbo-Croatian, Basque and Limburgian, is confined to one syllable per word, and many words do not have the option of bearing tone. What makes these languages particularly interesting is that since tone does not occupy most of the phonetic space used for pitch contours, a substantial part of this space is left for intonational processes. This fruitful cohabitation of tone and intonation in Europe has been studied increasingly over the past decades (e.g. Bruce 1977 for Swedish, Lehiste and Ivić 1986 and Smiljanić and Hualde 2000 for Serbo-Croatian, Gussenhoven & van der Vliet 1999 for Limburgian, and Elordieta 1997 for Basque).

In the present study, the tone-intonation interaction will be examined from the perspective of two East Limburgian dialects, the ones spoken in the Dutch cities of Roermond and Venlo (see map in Figure 1.1). They belong to a dialect continuum within the larger Dutch-German dialect continuum that encompasses the provinces of Limburg in the Netherlands and Belgium as well as the North of Rhineland-Palatinate and the Southwest of North Rhine-Westphalia in Germany. Figure 1.1 shows the area in which these tonal dialects are spoken.

The Limburgian tone contrast opposes Accent 1 (traditionally called stoottoon, or ‘push tone’) to Accent 2 (sleptoon, or ‘drag tone’). There is no unique way of briefly describing the shape of Accent 1 and Accent 2 pitch contours. Let us look at three rather different ways of looking at them, which we may characterize as (a) impressionistic, (b) selective, and (c) (partially) enumerative:
(a) By “push tone” we mean the strong, aggressive pronunciation of a sound (which can be long or short); “drag tone” is the peculiar singsong tone that is typical of many Limburgian dialects. (Staelens 1989:9)
(b) The bug bie (‘bee’) is pronounced with a so-called “push tone”: You start high and then lower the tone. When someone in Limburg says kom er bie (‘come here’), he pronounces the last word differently. The tone starts high, and goes down, but eventually goes up again. This tone is usually called “drag tone”. (Oostendorp 1996:222)
(c) a. push tone: falling contour, which means that the tone of the syllable falls from begin to end (Dutch also has this contour).
b. drag tone: This one is typically Limburgian. It occurs in two ways: first, shortly before a long break; the pitch falls, rises higher than it started, and slightly falls again. This movement is realized on one syllable. It also causes the syllable to become longer. It sounds very un-Dutch. Second: just in the sentence, not before a break: the pitch hardly falls. This, too, sounds un-Dutch, but it does not stand out as much as the other case. (Bakkes 2002:27)

The first definition gives a (somewhat subjective) general idea of Accent 1 and 2, without making them identifiable with confidence in the speech signal. The reason for such a high level of abstraction is that it is virtually impossible to describe the Limburgian tones with general acoustic characteristics. Not only is there a great deal of variation between the dialects, so that the pitch contour for Accent 1 in one dialect may sound like an Accent 2 in the other dialect, but the shape of the contours can also vary within a dialect according to structural factors. As we will see in the course of this thesis, the pitch contour corresponding to Accent 1 and 2 words can only be predicted correctly if we know the position of this word in the sentence (within or at the end of an utterance, and within or outside the focus constituent), as well as the discourse meaning used in this sentence (e.g. ‘statement’ vs. ‘question’). Definition (b) depicts a selected set of Accent 1 and 2 instances, which we may see as their most common form: inside the focus constituent, and at the end of a declarative sentence. Perhaps for this reason, this form is also subject to less variation among dialects, so that the $f_0$ (fundamental frequency) contours given in Figure 1.2, which were recorded by a speaker of the Venlo dialect, are similar to those observed in a number of other Limburgian dialects.

---

1 We give here the cited passages in their non-translated (Dutch) version: (a) “Onder stoottoon verstaan we de krachtige, agessieve uitspraak van een klink (die kort of lang kan zijn); sleeptoon is de eigenaardige zangtoon die aan veel Limburgse dialecten eigen is.”
(c) a. stoottoon: dalende intonatie, wat inhoudt dat de toon van de lettergreep daalt van begin naar einde (het Nederlands heeft deze intonatie ook);
2 For instance, in interrogative sentences, a focused word ending with a rise will be interpreted as an instance of Accent 1 by Venlo speakers, but as an instance of Accent 2 by Roermond speakers (see sections 1.1.3.1 and 1.1.4.1).
3 The same opposition between a fall and a fall-rise (with a more or less pronounced rise) was observed in the dialects of Roermond (see section 1.1.3.1), Sittard (Hanssen 2005) and Helden (van den Beuken 2007). In Hasselt (Peters 2006b), Accent 2 in this context is characterized by a mid-high plateau.
Definition (c) adds one dimension to the description, by distinguishing contours in sentence-internal or sentence-final position. Figure 1.3 shows the nonfinal contours for Accent 1 and 2 in the Venlo and Roermond dialect.

An additional dimension, absent from the definitions above, is the focus condition of the sentence. Words may or may not belong to the focus constituent of an utterance, depending on the relative importance – usually in terms of newness – of the information that they convey. For instance, if the sentence Pieter helped paint the Easter eggs is produced as an answer to Who helped paint the Easter eggs? (as opposed to, say, What happened?, or What did Pieter do?) , only Pieter is the new element in the sentence, and it will be focused. The two examples below show two different interpretations of the Venlo clause Ièrs zei ik "knien" (‘First I said
“rabbits”). Both are pronounced with ‘continuation’ intonation, signalling that the speaker is to add something after these clauses. In the left panel, the clause may answer the question ‘What did you say first?’ (the answer being ‘rabbits’, which is thus focused), and in the right panel, the question may be something like ‘What did you do with the word ‘rabbits’?’ (‘First I SAID ‘rabbits’, then I WROTE it.’). As we can see, the word knien, which in both cases bears Accent 1, is not realized with the same pitch contour when it occurs in a focused position as when it does not.

Figure 1.4: Final contours for the word knien with ‘continuation’ intonation in the Venlo clauses Ièrs zei ik "KNIEN" vs. Ièrs ZEI ik "knien" (‘First I said ‘rabbit’’). In both figures, the voiced part of knien is the last non-interrupted line.

Finally, the shape of Accent 1 and Accent 2 contours can depend on the discourse meaning of the sentence in which they occur. For instance, if a speaker of the Roermond dialect asks Zaes-te “BEIN”? (‘Do you say “legs”?’), the melody on bein (realized with Accent 1) will be rising-falling, whereas it will be only falling in the statement De zaes “BEIN”. (‘You say “LEG”.’). This difference is shown in Figure 1.5.

The influence of all these factors on the realization of Accent 1 and 2 will be explained in more detail in the next sections. First, we give a short overview of the system used for describing tonal grammars. Then, we show how this system is applied on the grammars of the two dialects of Roermond and Venlo.

**1.1.2 Autosegmental Metrical theory**

The theoretical framework used for the description of the Venlo and Roermond dialects is the Autosegmental Metrical (AM) theory, as introduced by Pierrehumbert (1980). In this section, we will only give a brief introduction to the basic concepts of the AM theory, in the way they are used for the dialects investigated in this thesis. Some terms may be defined differently for other languages. For a more complete
The main idea of the AM theory is to represent all tonal and intonational phenomena phonologically as sequences of two tones, H (high) and L (low), on a structural tier that runs parallel to the segmental tier. In being a linear string of lexical and intonational tones, it differs both from models that are based on the placement of one contour on top of another, as in earlier models of Swedish intonation associated with Eva Gårding (cf Ladd 1996), and from models that are based on a linear string of movements, like the description in the ‘Dutch School’ of ‘t Hart, Collier & Cohen (1990). The autosegmental nature of the representation, which it has in common with other models, allows us to describe utterances on two levels simultaneously, just like a musical score which displays notes and words at the same time. In the AM framework, several tones may coincide with a single segment (or syllable), as in songs, or, conversely, one tone may spread over several segments. For instance, a falling melody used on the English words milk (with the phoneme sequence /mIk/) and chocolate (‘/tʃɒklət/) may be represented with the same sequence of tones, viz. HL:

(1)

\[
\begin{array}{c|c|c}
milk  & \text{chocolate} \\ \\
\text{H} & \text{L} & \text{H} \\ \\
\end{array}
\]

However, unlike their musical counterpart, speech tunes are not fully specified. First, the H and L tones are phonological entities and should not be identified with absolute pitch levels, and second, only the language-relevant tonal events, and not the pitch movements in between, are phonologically defined in terms of melodic structure and links with the segmental tier. This abstract level of observation allows us to define distinctive sequences, or prosodic constituents, on which the grammar can operate.
The most common type of prosodic constituents used in intonational phonology is called the intonational phrase (or IP). Within an IP, we can observe different types of tonal events. Boundary tones, usually marked with a percent sign (T%) or a subscript T(e.g. T), occur at the beginning and end of the IP, indicating phrasing and/or discourse meaning. For instance, the Dutch sentence *Zullen we morgen afspreken, om negen uur?* (‘Shall we meet tomorrow, at nine o’clock?’) is usually pronounced as a sequence of two IPs (*Zullen we morgen afspreken* and *om negen uur*), which both end with the boundary tone H, in the first case for showing continuation and in the second for marking a question and the end of the utterance. As we can see, the same tone or tone sequence may have different functions. It may also have different meanings (e.g. the discourse meaning ‘continuation’ or ‘yes/no question’). This also holds for another special part of prosodic constituents, the pitch accent.

Pitch accents are recognizable in a tone sequence by the star attached to their leading element (H* or L*, or, for instance, H*L). In languages such as Dutch, or the Limburgian dialects investigated in this thesis, they are used to mark focus4. For instance, in the first IP of the sentence above, the word *morgen* (‘tomorrow’) is focused, since it appears to convey the most important information of the utterance. Therefore, the pitch accent will occur on this word - more specifically, on its most prominent syllable (the ‘word stress’ or ‘primary stress’, see for instance Gussenhoven 2004:13f). (2) gives the tone sequence corresponding to our example:

(2) *Zullen we morgen afspreken, om negen uur?*  
L, H* L H L, H* L*H H.

In addition to boundary tones and pitch accents, tone languages can have lexically specified tones, that is, tonal information stored in the mental representation of words. In the case of the Venlo and Roermond dialects, the difference between Accent 1 and Accent 2 is signalled by the presence (Accent 2) or absence (Accent 1) of a lexical H tone.

Given the pitch accent (which we also call *focal tone*, since it signals focus in the dialects observed), boundary tones and lexical tone(s), we can construct sequences of tones that describe any single-IP Roermond or Venlo utterance. However, the concatenation of tonal events is only the first step towards a correct pronunciation of East Limburgian IPs. Not only are tone sequences phonetically underspecified, but they may also undergo some modifications before they are passed on to the pronunciation stage. An original sequence, stored in the underlying representation, may be submitted to a set of phonological rules. If these rules apply to the given sequence, a different representation will be constructed, called the surface representation. Based on the surface representation, the acoustic signal can finally be created by means of tonal targets, which combine elements of the segmental tier to

---

4 In other languages (eg Japanese), the function of pitch accents is lexical, not intonational. See Pierrehumbert & Beckman (1988) for more details.
(relative) pitch heights. With the introduction of tonal targets, we leave the realm of phonology to address phonetic implementation issues. At this level, the targets are also linked up for the underspecified stretches, in a process called interpolation.

Sometimes, phonetic implementation follows general rules, such as the ones that govern transitions between tonal events, but these rules can be dialect-specific. This will be shown in the next two sections, dedicated to the description of the Venlo and Roermond grammars. As we will see, both dialects have rules that apply to underlying representations (e.g. infixation of boundary tones, or assignment of Accent 2 lexical tones in a limited number of contexts), as well as phonetic implementation rules (e.g. pronunciation of a H tone as a ‘mid-high’ in a given context).

1.1.3 The Roermond grammar

1.1.3.1 Building blocks and phonological adjustments

Roermond is a small town of 50,000 inhabitants in the middle of the Dutch province of Limburg. As in most of its neighbouring dialects (but not Weert, see chapter 2), Roermond Dutch uses pitch to encode a two-way lexical contrast between Accent 1 and Accent 2. The contrast mainly expresses differences between lexical meanings (e.g. *haas*\(^1\) = ‘hare’, *haas*\(^2\) = ‘glove’)\(^5\) or grammatical categories, such as the distinction between singular and plural (e.g. *knie*\(^1\) = ‘rabbits’, *knie*\(^2\) = ‘rabbit’) or between word classes (e.g., *fage*\(^1\) = ‘to fold’, *fage*\(^2\) = ‘folds’ (n.pl)). It should be noted, however, that each one of these distinctions can also be expressed by segmental means, as in the standard languages spoken in the Limburgian area (mainly Dutch and German), and that the morphological function of tone is confined to a handful of lexical entries. For instance, there are far less cases of plural marking by means of pitch contour differences than by means of a suffix –e (equivalent to the plural suffix –en in Dutch, or –en/-e in German).

Still, despite its relatively limited number of tonal minimal pairs (we counted 82 of them in the Kats (1985) dictionary), the Roermond lexical tonal system remains a crucial element of the grammar, since a substantial class of syllables cannot be realized correctly if speakers do not know their tonal specification. This class, which all East Limburgian dialects have in common (Gussenhoven & Peters 2008), comprises all main-stressed syllables with at least two sonorant moras, that is, a long vowel, a diphthong or the combination of a short vowel and a sonorant consonant ([m n ɲ R l w ʃ]). In such syllables, the lexical contrast is expressed by the absence (Accent 1) or the presence (Accent 2) of a H tone on the second mora\(^6\). This rule has

---

\(^5\) The superscripts \(^1\) and \(^2\) refer to Accent 1 and Accent 2, respectively.

\(^6\) Strictly speaking, Accent 1 and Accent 2 are not lexical tones in the sense of, say, the Chinese lexical tone HL (Tone 4). The real tone is the H tone that characterizes Accent 2
two implications. First, the TBU is the mora, which is characteristic for East Limburgian dialects; in other Limburgian dialects, such as Hasselt or Borgloon (Peters 2006b and 2007), the tonal contrast is syllable-based. Second, if there is no second mora in a word, there is no point in asking ourselves whether this word has Accent 1 or Accent 2: This lexical entry simply cannot be specified for tone. A different situation is encountered in bimoraic syllables that do not occur in a focused or final position, that is, in syllables that are not associated with a focal tone (T*) or a boundary tone (T). In this case, the lexical entry in question is indeed specified for tone, but its surface representation will be the same for Accent 1 and Accent 2. In other words, the lexical tonal contrast is suppressed when a word specified for Accent 1 or Accent 2 is realized after the focus constituent and before the last syllable of the IP. (R1) and (R2), adapted from Gussenhoven (2000a), summarize the basic conditions in which the Roermond tonal contrast can take place:

**Underlying representation of**

<table>
<thead>
<tr>
<th>monomoraic syllables:</th>
<th>( \mu(m) ) (( m = ) sonorant ( \mu ), ( F = ) foot)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accent 1 syllables:</td>
<td>( \mu(m m) )</td>
</tr>
<tr>
<td>Accent 2 syllables:</td>
<td>( \mu(m m) )</td>
</tr>
</tbody>
</table>

**Neutralization of the lexical tonal contrast**

The lexical H is deleted from Accent 2 syllables unless they also contain a focal (T*) and/or a boundary (T(T)) tone.

Rule (R2) mentions focal and boundary tones. In the Roermond dialect, focal tones can be represented by \( H^* \) or \( L^* \) and (final) boundary tones by \( L \), or \( H L_L \), depending on the discoursal meaning to be encoded. While declarative intonation is expressed by the combination of a \( H^* \) focal tone and a \( L \), boundary tone, the Roermond underlying representation for interrogative (yes/no) intonation is \( L^* \) plus \( H L_L \). These melodies are depicted in (R3):

**Intonational contours**

| **Declarative intonation:** | \( H^* L_L \) |
| **Interrogative intonation:** | \( L^* H L_L \) |

only. In this sense, the term accent seems more accurate than the traditional push tone and drag tone.

7 The TBU (Tone Bearing Unit) is the element on the segmental tier to which a tone can be linked, for instance the syllable (\( \sigma \)) or mora (\( \mu \)).

8 All subsequent (Rx) rules also originate from Gussenhoven 2000a.
The representations given in (R3) can be interpreted quite literally when both the focal and the final syllables are monomoraic. We will then have, for declarative sentences, a gradual fall that results from the interpolation between a high and a low target, and a rise-fall for interrogative sentences. However, these representations may surface differently in the cases that we are mainly interested in, viz. sentences with Accent 1 or Accent 2 in bimoraic focal or final syllables. Two examples of such sentences are given below, using the word beinⁿ (“leg” – with Accent 1, it would mean ‘legs’). First, let us see how lexical tones are associated with elements of the segmental tier. According to (R1), beinⁿ has a lexical H on its second mora. In example (3), rule (R2) gives us two good reasons for preserving this lexical H: the syllable in which it appears also contains a boundary and a focal tone. In example (4), since the focus is assigned to voot and the boundary tone aligned with vas, beinⁿ looses its lexical H, which makes it look just like an Accent 1 syllable.

(3)  
\[ [\text{Miene BEIN}^\text{II}] \]  
\[ L_1 \quad [H^\text{H}_{\text{lex}} L_t] \]  
\[ \text{‘My LEG’} \]  

(4)  
\[ [\text{Miene VOOT}^\text{I} \text{ zit aan miene bein}^\text{II} \text{ vas}] \]  
\[ L_1 \quad [H^* \text{ L}_{\text{lex}} L_t] \]  
\[ \text{‘My FOOT is attached to my leg’} \]

There is, however, more to the word BEINⁿ in example (3) (and also to example (4), see (6)). Based on the representation obtained so far, we would not be able to pronounce its lexical tones correctly. BEINⁿ would be realized as a fall, as is BEINⁿ. In reality, the underlying tone sequence in (3) is modified by a rule that allows Accent 2 to be radically different from Accent 1. Instead of a fall, the IP ends with a rise, due to the insertion of the boundary tone L, directly after the focal tone H*. The rule may be formulated as follows:

(R4) Sequencing of tones (constraint AlignT_{lex}^\text{Right})
If the final mora of the intonation phrase has a lexical H, the boundary tones (L, and H, L) are realized before it.

The tone sequence in Miene BEINⁿ is thus transformed from H^*H_{lex} L, to H^* L, H_{lex}, which makes it a typical instance of the “peculiar singsong tone” that was referred to by Staelens (1989, see above). (5) illustrates how the contour will finally be pronounced:
Such a use of final boundary tones is quite surprising if we consider the traditional conception of boundary tones (Pierrehumbert and Beckman 1988), in which they were assumed to occur at the edges of the IP. This departure from the general view is due to the constraints defined for the Roermond dialect in an OT perspective (see for instance Kager 1999 for an introduction, and Gussenhoven 2004 or Yip 2002 for tone-related examples). One of these constraints is the alignment of the lexical tone to the right edge of the syllable. This alignment constraint is obviously in competition with the constraint of aligning the boundary tones to the edge of the IP, which in this case is the same position. In the Roermond case, the former constraint is ranked higher than the latter, resulting in tone sequences as in (5).

Another case of friction between lexical and boundary tones arises when the focus is realized IP-internally, leaving some space between the focal and the final boundary tone. Let us first consider what happens when there is no lexical tone on the second mora of the focused word, that is, when this word bears Accent 1. An example of this situation was given in (4); we repeat it in (6), together with its pitch contour.

The contour above could be predicted by the rules (R1) to (R3), if it wasn’t for the steep fall that immediately follows the focal H* on VOOT. Such a steep fall, instead of the gradual fall that would be generated according to basic phonetic implementation rules (as a steady line from the H* to the low (L) target), is due to another constraint involving boundary tones. This constraint stipulates that boundary tones should not only align their right edge to the right edge of the IP, but also their left edge as far left as possible in the IP. By this, the boundary tone is pulled well into the IP and associates with the second, so far empty, mora of the focused syllable, so that we obtain the sharp fall ending on the second mora of VOOT. The two alignment constraints are given in (R5); (R5') gives a more concrete representation of their effect in Roermond IPs:

(R5) a. ALIGNT,LEFT - (T,L):
The left edge of the Phrasal boundary tones is aligned leftmost.
b. ALIGNRT - (T,R,Phrase,R):
The right edge of the boundary tones coincides with the right edge of the Phrase.

(R5') Leftward tone spreading
Lₐ spreads to the free mora in a focused syllable.

\[(m m)_ₐ \ldots \]
\[\text{H} \quad \text{Lₐ}\]

Such an influence of the boundary tone is, however, challenged in the presence of a lexical tone (marking Accent 2) in the focused syllable. In this case, the pitch remains high until the end of the syllable, after which it will obey the general phonetic interpolation rule that simply draws a (roughly) straight line from a H to a subsequent Lₐ, as shown in (7)⁹.

(7)  
\[\text{Miene ERM'' zit aan mienie beim'' vas }\]
\[\text{Lₐ} \quad \text{H*H}_{\text{lex}} \quad \text{Lₐ}\]

When we try to apply the constraints (R5a) and (R5b) to IPs with interrogative intonation, we see another discrepancy between predicted and observed contours that calls for a phonological adjustment. Let us consider the contour for the question "Höbse BEINI'' gezag?  ("Did you say LEG(S)?"), in which the focused word is in a nonfinal position. Based on (R5), we should have a very similar contour for Accent 1 and Accent 2: In the presence of Accent 1, the contour L*HₐLₐ should be realized with a steep rise, since the left edge of HₐLₐ, a high tone, is supposed to spread leftwards to all available moras until the focal tone L*. Accent 2 words should also be assigned the tone sequence L*H, this time due to the lexical H on the second mora. Instead, we have a low plateau for Accent 2, and for Accent 1, a rise that reaches its end only after the focused syllable. Both cases are dictated by a constraint called NoRise, which bans full rises within a single syllable. NoRise causes the assimilation of the lexical Hₐlex to Lₐlex, so that the resulting Accent 2 contour is flat (L*Lₐlex), and since this constraint is ranked higher than the left-alignment constraint stated in (R5a), the rise stretches past the end of the focused syllable. We give in (R6) the formal specification of NoRise:

---

⁹ As explained in Gussenhoven (2000a), the constraint also fails to apply to monomoraic syllables, which equally lack an empty sonorant mora, so that the post-focal contour will also be gradually falling.
The NoRise constraint is motivated by the observed difficulty of quickly pronouncing \( f_0 \) rises, as compared to falls (Ohala 1978), resulting in the tendency to avoid them. Such a tendency appears to be a common feature in languages (such as Hausa, and other examples given in Gussenhoven 2000a:150).

NoRise, however, does not preclude rises within any Roermond Dutch syllable. Indeed, as can be seen on the summary table (Table 1.1, next section), there are quite a few cases of rising contours within syllables. Ruling out all cases whose tone sequence does not strictly correspond to the one indicated in (R6) (see next section for explanations of individual cases), we still need to account for the exception constituted by nonfocused, IP–final Accent 2 syllables with declarative intonation. Gussenhoven’s explanation for this LH contour is based on the position of the Accent 2 syllable in the IP. As was said, the use of NoRise was motivated by the difficulty of pronouncing a rise in a short time. This difficulty becomes less of an issue when the syllable occurs at the end of an IP, since final syllables are usually lengthened, leaving more time for articulatory gestures\(^{10}\).

In a further step towards a formal grammar, Gussenhoven (2000a:151) translates this exception to NoRise as a separate constraint. Instead of stating, in one single constraint, that NoRise can only occur IP-internally (“NoNonFinalRise”), we can combine NoRise with a constraint stating that the underlying last tone of an IP should not be modified (IdentFin(T)). This combination of rules, which we may phrase as “Repair any syllable-internal rise, as long as the last tone in the IP remains unchanged”, has exactly the same effect as “Repair any syllable-internal rise in nonfinal position” in the case discussed above (Accent 2, \([-focus, +final]\), declarative). However, it is important to notice that only the second formulation of

\[ \text{(R6) NoRise} \]

\[
\begin{array}{c|c}
\text{L} & \text{H} \\
\hline
\text{ \( f_0 \) rises, as compared to falls (Ohala 1978), resulting in the tendency to avoid them. Such a tendency appears to be a common feature in languages (such as Hausa, and other examples given in Gussenhoven 2000a:150).}

\]

\[\text{NoRise, however, does not preclude rises within any Roermond Dutch syllable. Indeed, as can be seen on the summary table (Table 1.1, next section), there are quite a few cases of rising contours within syllables. Ruling out all cases whose tone sequence does not strictly correspond to the one indicated in (R6) (see next section for explanations of individual cases), we still need to account for the exception constituted by nonfocused, IP–final Accent 2 syllables with declarative intonation. Gussenhoven’s explanation for this LH contour is based on the position of the Accent 2 syllable in the IP. As was said, the use of NoRise was motivated by the difficulty of pronouncing a rise in a short time. This difficulty becomes less of an issue when the syllable occurs at the end of an IP, since final syllables are usually lengthened, leaving more time for articulatory gestures.}

\]

\[\text{In a further step towards a formal grammar, Gussenhoven (2000a:151) translates this exception to NoRise as a separate constraint. Instead of stating, in one single constraint, that NoRise can only occur IP-internally (“NoNonFinalRise”), we can combine NoRise with a constraint stating that the underlying last tone of an IP should not be modified (IdentFin(T)). This combination of rules, which we may phrase as “Repair any syllable-internal rise, as long as the last tone in the IP remains unchanged”, has exactly the same effect as “Repair any syllable-internal rise in nonfinal position” in the case discussed above (Accent 2, \([-focus, +final]\), declarative). However, it is important to notice that only the second formulation of}

\]

\[\text{10 According to Beckman (Beckman, unpublished; see Gussenhoven & van der Vliet 199:108), a syllable is in a privileged location if it is stressed or occurs at the edge of an IP. In privileged locations, pitch contours (as well as segments) are pronounced with more care. Speakers tend to use more time to pronounce distinctive elements more clearly, or to make distinctions at all. This principle is determinant in the Roermond tonal grammar, not only because it causes the neutralization of the tonal contrast in non-privileged locations, but also because it seems to allow the existence of an otherwise illicit syllable-internal rise. We may then wonder why the other privileged position, viz. the stressed syllable (which is usually longer than non-stressed syllable, see Cambier-Langenveld & Turk 1999 for Dutch), is precisely the one in which a rise is forbidden. A plausible reason for this apparent contradiction is that in phrase-internal position, there is still enough time left after the focused syllable to delay a rise. At the end of the IP, this time is given due to the position of the target word, since final syllables are usually longer than their IP-internal counterparts.}

\]
NORISE is explicitly based on empirical data. Speakers may prefer to avoid rises since these were shown to be less efficient than falls, but in final position (which in fact corresponds to the last syllable, not the last tone or mora), rises may be maintained due to the extra time available, as observed in other languages. On the other hand, the first formulation (even if it were exactly equivalent to the second one) should simply be taken literally, as an instruction within the phonological grammar, irrespective of phonetic considerations.

Such a divorce between the formal and the phonetically-based description may not be desirable in all situations, but in the Roermond grammar it helps to solve a small mystery: If we accept the special status of final syllables as the explanation for a full rise to occur, it would seem reasonable not to use NORISE at all in this position. Yet one of the two principles postulated for transforming the underlying representation $L^* H_{lex} H_{lex}$ (via $L^* H L_{lex}$) into $L^* L L_{lex}$ is precisely NORISE! We must then, in order to reconcile these observations, abstract away from phonetic considerations and simply confront the underlying structure $L^* H L_{lex}$ as depicted in (8)) with the NORISE and IdentFin(T) constraints\textsuperscript{11}.

\begin{equation}
(m \ m)\overline{|} \overline{|} \overline{L^* H L_{lex}}
\end{equation}

According to the grammar, the combination of an L tone associated with the first mora and an H tone associated with the second mora is illicit. The H is thus a violation of NORISE and should be repaired. This adjustment is, strictly speaking, not in contradiction with the IdentFin(T) constraint, since the H is not the last tone in the IP.

We have now discussed all rules (or constraints) defined for the Roermond grammar. It may be helpful to try and apply them to all Accent 1 and Accent 2 contours. This is done in the next section.

1.1.3.2 Contour inventory

The Roermond contrast, illustrated in Table 1.1, can be summarized as follows: In most of the cases, Accent 1 contours end with a fall and Accent 2 contours with a rise. The only exception to this fall/rise opposition is found in questions, when the target words are IP-internal and focused. In that case, Accent 1 contours are rising

\textsuperscript{11} Gussenhoven’s reaction to this comment was that “grammatical constraints go back to functional considerations, but once in the brain are no longer necessarily functional in that sense. Quite generally, constraints on representations don’t have anything to say about pronunciations except indirectly, to the extent that they prevent illegal representations from arising in surface representations” (personal communication).
while Accent 2 contours are flat. The phonological structure and rules underlying these observations are given below.

Table 1.1: Phonological representations and the corresponding stylized contours of Roermond Dutch. Solid contours represent Accent 1, interrupted contours represent Accent 2. Shaded portions indicate the stressed syllable. Adapted from Hanssen 2005.

<table>
<thead>
<tr>
<th></th>
<th>Focus final</th>
<th>Focus nonfinal</th>
<th>Nonfocus final</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Declarative</strong></td>
<td><img src="image" alt="Declarative Focus Final" /></td>
<td><img src="image" alt="Declarative Focus Nonfinal" /></td>
<td><img src="image" alt="Declarative Nonfocus Final" /></td>
</tr>
<tr>
<td>Ac. 1:</td>
<td>H* L</td>
<td>H* L</td>
<td>L</td>
</tr>
<tr>
<td>Ac. 2:</td>
<td>H* L&lt;sub&gt;lex&lt;/sub&gt;</td>
<td>H* L&lt;sub&gt;lex&lt;/sub&gt;</td>
<td>L&lt;sub&gt;lex&lt;/sub&gt;</td>
</tr>
<tr>
<td><strong>Interrogative</strong></td>
<td><img src="image" alt="Interrogative Focus Final" /></td>
<td><img src="image" alt="Interrogative Focus Nonfinal" /></td>
<td><img src="image" alt="Interrogative Nonfocus Final" /></td>
</tr>
<tr>
<td>Ac. 1:</td>
<td>L* H&lt;sub&gt;lex&lt;/sub&gt;</td>
<td>L* H&lt;sub&gt;lex&lt;/sub&gt;</td>
<td>L&lt;sub&gt;lex&lt;/sub&gt;</td>
</tr>
<tr>
<td>Ac. 2:</td>
<td>L* L&lt;sub&gt;lex&lt;/sub&gt;</td>
<td>L* L&lt;sub&gt;lex&lt;/sub&gt;</td>
<td>L&lt;sub&gt;lex&lt;/sub&gt;</td>
</tr>
</tbody>
</table>

The declarative contours for Accent 1 are all falling, due to the exceptionless concatenation of a focal H* and a (non-infixed) final boundary tone L. Focused, final Accent 1 words bear the focal H and the boundary L on the same syllable, with the corresponding high and low targets that cause a steep fall. Such a steep fall is also observed in focused, nonfinal position, due to the spreading rule (R5). In the nonfocus, final Accent 1 case, the fall from H* to the IP-internal target of L, started earlier in the utterance (a gradual fall if the focused word is monomoraic or bears Accent 2, or a steep one in the presence of Accent 1), so that the pitch is already low at the beginning of the target word, and continues to fall until its end.

Accent 2 words in declarative sentences show more diverse contours than their Accent 1 counterparts. In a focused, final target word, the rule (R4) applies, so that instead of a falling H*HL, sequence, we obtain the fall-rise H*LH. The nonfinal
case surfaces just like the underlying contour, that is, as a succession of two high tones (H*H); its realization as a rise from mid-high to high instead of a high plateau, is a result of phonetic implementation of the first H, realized mid. NoRISE has no grip on this rise, since it is not phonologically defined. Another slight rise is found in nonfocus, final Accent 2 words. Rule (R5) allows the final L, to move to the left, causing the lexical H to be realized at the very end of the IP. As explained in the previous section, NoRISE does not apply to this contour either, even though the rise is phonologically defined. The L, associated to the first mora and H associated to the second mora may surface as such, due to the position of the target word in the IP.

Interrogative sentences reveal a straightforward use of the focal and boundary tones for Accent 1 (underlying and surface representations are identical), and for Accent 2, interesting combinations of the phonological adjustment rules described above.

In focused, final Accent 1 words, the contour simply corresponds to the concatenation of focal L* and the boundary tone H, within the target word. Here we observe, as for Accent 2 contours in the same context (see previous section), a quite complex use of the NoRISE constraint. Remember that there are two effects of NoRISE: first, in Accent 2 words, the H-to-L change where H is associated to the second mora, and second, in Accent 1 words, the non-association of a H, to the second mora. In this case, NoRISE thus prevents the association of H, with the second mora, which would result in a forbidden structure. From a grammatical point of view, NoRISE is applied flawlessly, although on the surface, a rise-fall is observed. This is thus another case of the use of NoRISE which does not prevent a phonetic rise.

IP-internally, focused Accent 1 words start with low pitch, as dictated by L* on the first mora, and end with a rise that reaches its highest point after the end of the syllable (due to NoRISE). At the end of the IP, the contour for unfocused Accent 1 words is falling, following the movement indicated by the boundary tone H,L.

The tone sequence for Accent 2 in a focused, final context with interrogative intonation is determined by the assimilation of the lexical H (due to NoRISE, see also previous section), which follows the infixation of the boundary tone H,L. NoRISE is also responsible for the shape of the Accent 2 contour in focused, nonfinal target words: the lexical H in L*H is assimilated, creating the low plateau observed in this context. As we will see in chapter 3, this contour is sometimes realized as a slight rise, reducing the salience of the tonal contrast. The last Accent 2 contour, a fall-rise that characterises nonfocused, final words, is readily explained by rule (R4), which causes the boundary tone sequence H,L, to be moved before the lexical H.

We can now turn to the description of the tonal contrast found in the Venlo dialect, the other variant of East-Limburgian examined in this thesis.
1.1.4 The Venlo grammar

1.1.4.1 Building blocks and phonological adjustments

With its 90,000 inhabitants, twice as many as in Roermond, Venlo is one of the largest municipalities in the Dutch province of Limburg. It is situated in the north of the province, close to the German border, but more importantly, to the boundary between tonal (Limburg) and non-tonal dialects spoken further to the north of the province of Limburg as well as in Gelderland and Noord-Brabant. Perhaps for this reason, the intonational system used in Venlo is more complex than the one in Roermond: it may have integrated intonational contours from neighbouring dialects into its tonal system. Instead of the two different intonational contours found in Roermond, viz. statement and question intonation, the Venlo dialect has four contours: statement, low question, high question and continuation.

Statement intonation is, as in any other language, the default intonation (i.e. the one used in citation form), and is used to assert some information. Question intonation, or more precisely, yes/no question intonation, is mostly used to enquire on the truth or validity of a statement. There are two different types of yes/no questions in the Venlo dialect, the neutral one, in which one simply requests information, and the less frequent ‘surprised’ one, in which the speaker also seems to emphasize some doubt about the truth of a statement. Originally, these two questions were tagged as ‘interrogative’ and ‘surprised question’ (Gussenhoven & van der Vliet 1999), but since the interpretation or the use of ‘surprised questions’ by native speakers was not always as clear-cut as the name would suggest (see section 4.2.2: a speaker consistently exchanged the two question types), we will rather use the more acoustically-based names of ‘low question’ and ‘high question’ for the ‘interrogative’ and ‘surprised question’ intonation, respectively.

As in the Roermond dialect, the Venlo TBU is the mora, and the same conditions apply for the tonal contrast to occur: We need a bimoraic main-stressed syllable, and Accent 2 syllables bear a lexical tone on their second mora if they are focused and/or IP–final. Not only does this lexical tone determine the shape of Accent 2 pitch contours, it also has a direct consequence on the duration of Accent 2 syllables, which will always be longer than their Accent 1 counterparts in final position. We will come back to this observation at the end of the section.

Besides the lexical tone, building blocks of IPs in the Venlo dialect are focal and boundary tones. Focal tones can have one of two forms, depending on the discourse meaning used in the IP. The most frequent form is H*, which is used in combination with statement, continuation and low question intonation. In high questions, the focal tone is underlyingly specified as L*, but as we will see below (rule V3), it also surfaces as an H* in Accent 2 syllables by an assimilation rule. The H* and L* focal

---

12 In other words, monomoraic, Accent 1 syllables and nonfocused, nonfinal Accent 2 syllables are toneless.
tones are combined with four different boundary tones (one for each discourse meaning) to create the following intonations:

(V1)  **Building blocks of the Venlo tonal grammar**
Lexical tone: H on the second mora of focused and/or final Accent 2 syllables

<table>
<thead>
<tr>
<th>Intonations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Declarative:</td>
</tr>
<tr>
<td>Low question:</td>
</tr>
<tr>
<td>Continuation:</td>
</tr>
<tr>
<td>High question:</td>
</tr>
</tbody>
</table>

(Hₙ denotes an optional utterance boundary tone, in addition to the IP boundary tone L, or Hₙ)

These melodies, enriched with an initial boundary L, and, possibly, with H lexical tones (in the presence of Accent 2 words in focused and/or final position), can accurately describe the underlying form of any Venlo Dutch intonational phrase. However, as we saw in section 1.1.3.1, the IPs may not always surface as the faithful concatenation of their underlying tones. Two elements of the grammar form the basis of possible modifications. First, the lexical H that comes with focused and/or final Accent 2 syllables changes the initial sequence and may motivate further adjustments (see below, from (V3) on). Second, the interpolation between the last tone of the focal syllable and the first tone of the final syllable (if these syllables do not coincide) is not always a straight line. The latter is explained by a rule related to the Roermond ‘leftward tone spreading’ rule (R5), which provides a low target to the first toneless mora after the focus in Accent 1 words. There is, however, an important difference between the Roermond and the Venlo version of the leftward tone spreading rule. Whereas in Roermond, only the Lₚ final boundary tone is concerned, the Venlo grammar extends this rule to any tone that follows the focal tone in the IP tone sequence (i.e. also Hₙ, or any lexical tone), provided this tone is different from the focal tone. The rule for Venlo is given in (V2) (adapted from Gussenhoven & van der Vliet 1999):

(V2)  **Leftward tone spreading**
A free mora in the focused syllable is filled with the next tone in the post-focal tone sequence, if this tone is different from the focal tone.

$$\begin{array}{ll}
\text{(m m)}_{\textbf{αT}} \ldots \text{m} & \ldots \text{m} \\
\text{\underline{αT}} & \underline{\text{αT}}
\end{array}$$

whereas $\alpha T = H$ or $L$; $-\alpha T = L$ if $\alpha T = H$, and vice-versa; $-\alpha T$ may be a lexical or a boundary tone

This rule can be interpreted in two ways, depending on the last syllable in the IP. First, if this last syllable is bimoraic with Accent 1 or monomoraic, the tone
following the focal tone will be a boundary tone (or the first part of a bitonal boundary tone). According to the contours given in (V1), there are three intonations in which the boundary tone is different from the focal tone: the declarative, low question, and high question intonation. In the case of declarative or low question intonation, the rule will result in a steep fall within the focused syllable. High questions, on the other hand, will be characterized by a steep rise (which is allowed to occur within the focused syllable, since NoRise is not a relevant constraint in the Venlo grammar), as shown in (9):

\[
\text{(9) } [\text{Zitte dien VEUT}^{1} \text{ aan dien bein}^{1\beta} ]
\]

\[
L_{i} \quad L^{*} \quad H H_{u}
\]

‘Are your feet attached to your legs?’

A second situation in which (V2) is used, is in sentences ending with an Accent 2 word. In this case, the tone following the focal tone will be the lexical H tone, which is different from the focal tone in sentences with high question intonation (the only melody with an L*). An example of this sequence of tonal events is shown in (10):

\[
\text{(10) } [\text{Zit diene VOOT}^{1} \text{ aan diene bein}^{1\beta} ]
\]

\[
L_{i} \quad L^{*} \quad H H_{u}
\]

‘Is your foot attached to your leg?’

As is evident from (V2), the tone following the focal tone can only fill the second mora of a focused word if this mora is available, that is, if the focused word bears Accent 1. In the case of Accent 2 in the focused syllable, another rule modifies the pitch contour of the IP. This rule is an assimilation rule, illustrated in (11) and spelled out in (V3), which replaces a focal L* by a H* when the second mora of the syllable is associated to a lexical H:

\[
\text{(11) } [\text{Zit diene BEIN}^{1} \text{ aan diene voort}^{1\beta} ]
\]

\[
L_{i} \quad H^{*} H \quad H H_{u}
\]

‘Is your leg attached to your foot?’

(V3) \text{L* assimilation}

L* is replaced by H* in Accent 2 syllables.
This assimilation rule represents a modification of the grammar presented in Gussenhoven & van der Vliet (1999). In the original grammar, two different underlying tones were assumed for Accent 1 and Accent 2 in high questions, respectively L* and H*. This tone-dependent allomorphy is here replaced by a more uniform underlying representation in high questions. We assume a unique underlying tone, L*, for high questions, which can be modified by the assimilation rule above. Not only does this view simplify the representation of the intonation contours, as in (V1), but the assimilation rule stated in (V3) is also more likely to occur in a Franconian dialect than the multiple underlying representations of a focal tone within a single intonation. As stated by Gussenhoven & van der Vliet (1999:122), “[t]he allomorphy statement […] differs from the situation in many other varieties of Dutch, where the occurrence of L* versus H* is uniquely determined by the semantics of the intonation (for instance, ‘Declarative’ vs. ‘Interrogative’”).

Assimilation rules raising L before H as in (V3) are legion in the world’s languages (see for instance Ohala 1990, Hyman 2007) and tone assimilations also occur in the Roermond and Cologne dialects. It is, then, not an abnormal coincidence that the Venlo dialect has two. The second assimilation rule, which also involves Accent 2 words, changes a lexical H into L if it occurs IP–finally, before a boundary L.

\[(V4)\] \(H_{\text{lex}}\text{ assimilation}\)

\(H_{\text{lex}}\) becomes \(L_{\text{lex}}\) before a L, boundary tone in the same syllable.

(12) illustrates the second assimilation rule. The sentence ends with a focused Accent 2 word, initially associated with the tone sequence H*H. Since the boundary tone L, is present in the same syllable, the lexical H will surface as an L, which results in the sequence H* L L. As shown in (12b), this rule has no effect on Accent 1 words, which, by definition, lack a lexical H (or L).

(12) a. Ich zegk KNIEN \[\]

\(H^* L_{\text{lex}} \ L_i\)

b. Ich zegk KNIEN \[\]

\(H^* L_{\text{lex}} \ L_i\)

‘I say “RABBIT”’

Based on the tone sequences given in (12), we might assume that the lexical contrast in this context depends on an additional L tone between the focal and the boundary tone for Accent 2. In reality, there is an extra cue to the lexical contrast, which we

\[13\] The fact that a unique underlying focal tone per intonation conforms with the intonational grammar of Standard Dutch (see also Gussenhoven 2004:247) does not necessarily guarantee the validity of the update – after all, many aspects of the grammar may be different since Standard Dutch is non-tonal. However, it seems to be a valid point in the case of Venlo Dutch, which probably borrowed more from the standard intonational grammar than many of its related dialects.
can observe in the contour of (12a). The contour for Accent 2 ends with a slight rise (although often, a level extension is observed instead). We will not expand on the reasons for this rise (see Gussenhoven & van der Vliet 1999:114f), but it should be noted that although the Accent 2 contour looks similar to the Roermond contour in the same context, the tone sequences in both dialects are fundamentally different \( (H^* L, H_{lex} \text{ vs. } H^* L_{lex} L) \), which shows the great variety of tonal processes even in closely related dialects.

A further difference between the Roermond and the Venlo dialect, is the influence of the lexical tone on another prosodic dimension, viz. duration. In the examples above, the Accent 2 syllable (12a) will be pronounced with a clearly longer duration than the Accent 1 syllable (12b). Gussenhoven & van der Vliet (1999:131) assume for the Venlo (but not for the Roermond) dialect a phonetic implementation rule that interprets the lexical tone in a final syllable as an instruction for extra lengthening of this syllable, as given in (V5):

\[
\text{(V5) Phonetic implementation: Final Accent-2 length}
\]

A lexical tone on an IP-final mora causes extra 'final lengthening'.

The Venlo dialect shares this property with a number of other Limburgian dialects (see section 1.2.2). It is viewed as an enhancement feature of the lexical contrast. In the case of the Venlo contrasts, duration is of crucial importance for the encoding (and perception) of the Accent 1/Accent 2 distinction. In some contexts, it seems to be the only cue at play. Examples (9) and (10) illustrate the duration-only based contrast between nonfocused final words in declarative sentences \( \text{beinI/II} \).

An important consequence of the lengthening rule (V5) is that since duration is now recruited as a cue for the lexical contrast, speakers avoid using it for other purposes, in particular for facilitating the pronunciation of a complex tone sequence in a single syllable. In one context, however, one would really need some more time to cope with tone crowding: the contour for Accent 1 in low questions, when the target word is focused and final, is a fall-rise \( (H^* L, H) \), which might well benefit from some extra time for it to be pronounced clearly. However, we cannot take this extra time, since this option is reserved for Accent 2 words. In order to save the contrast, speakers prefer to avoid the predicted contour altogether, and to use the one defined for Accent 1 in high questions instead.

Two more implementation rules are defined for contours in particular contexts. The first, called Phonetic \( H^* \)-lowering, causes the realization of the focal \( H^* \) at mid pitch when it occurs in sentences with continuation or high question intonation; the second, \( (H) \) Upstep, causes the \( H, H \) boundary tone in high questions to be realized as a rise instead of a high plateau. Since these rules are not crucial for a better understanding of the Venlo grammar, we will not comment on them further in this section. We will however see them at work in the next section.
1.1.4.2 Contour inventory and an alternative account

Table 1.2 gives a summary of the Venlo contours from their underlying representation to their expected pronunciation; Table 1.3 shows the actual (but stylized) contour shapes.

Recently, a voice arose (Kehrein 2007a, 2007b) to give a very different account of the intonational melodies of Franconian dialects, using as examples the dialects of Cologne and Venlo. Kehrein’s claim is that there is no solid reason to analyze the observed pitch contour differences as resulting from a tonal contrast. Instead, he proposes a moraic accent approach. The contrast does not oppose syllables that do not have an H tone on their second mora (Accent 1) to those which do have one (Accent 2), but rather opposes syllables in which the first mora is stressed (replacing Accent 1) to syllables in which the second mora is stressed (replacing Accent 2). Differences in intonational contours are then merely due to the fact that pitch accents are linked to the first or the second mora of the syllable (depending on where the prominent element is). In other words, the distinction between Accent 1 and Accent 2 is reduced to a simple timing difference.

Applied to the dialect of Venlo, this account explains adequately the apparent delay in ‘Accent 2’ contours, especially in a [+focus, –final] context. It also gives a simple explanation to the fact that some contrasts (in [–focus, +final] position, except with declarative intonation) are so subtle in terms of $f_0$ and may be largely based on duration. The interpretation given to the slight rises at the end of final Accent 2 words with declarative intonation may seem less convincing; these are seen as epitones that should enhance the contrast. In our view, duration seems to be robust enough in the other contexts and in the systems of many languages (see chapter 2 for the Weert dialect). Nevertheless, Kehrein’s account seems, on the whole, well worth a deeper analysis, including a confrontation with the contour inventories of other dialects in the area.

We will not include such an analysis in our thesis. The most important message in Gussenhoven’s account, namely the fact that the contours vary in function of the prosodic context, is not directly threatened by this new interpretation of the contrasts. The shapes of (what we will still call) Accent 1 and Accent 2 contours may be explained differently, but Kehrein does not deny that they differ per context; the issue of neutralization is not explicitly mentioned, but it does not seem to be incompatible with the alternative account.

This discussion concludes our introduction to the Roermond and Venlo contours. The rest of this chapter gives an overview of the literature on tone perception and its relevance to our research.
Table 1.2: Phonological representations of the Accent 1 and Accent 2 contours in the Venlo dialect.

<table>
<thead>
<tr>
<th>+FOCUS, +FINAL</th>
<th>+FOCUS, -FINAL</th>
<th>-FOCUS, +FINAL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Accent 1</strong></td>
<td><strong>Accent 2</strong></td>
<td><strong>Accent 1</strong></td>
</tr>
<tr>
<td>H* Li</td>
<td>H* LiHu</td>
<td>H* LiHu</td>
</tr>
<tr>
<td>flat tone</td>
<td>flat tone</td>
<td>flat tone</td>
</tr>
<tr>
<td>syllable</td>
<td>syllable</td>
<td>syllable</td>
</tr>
<tr>
<td>H* LiHu</td>
<td>H* LiHu</td>
<td>H* LiHu</td>
</tr>
<tr>
<td>flat tone</td>
<td>flat tone</td>
<td>flat tone</td>
</tr>
<tr>
<td>syllable</td>
<td>syllable</td>
<td>syllable</td>
</tr>
<tr>
<td>H* LiHu</td>
<td>H* LiHu</td>
<td>H* LiHu</td>
</tr>
<tr>
<td>flat tone</td>
<td>flat tone</td>
<td>flat tone</td>
</tr>
<tr>
<td>syllable</td>
<td>syllable</td>
<td>syllable</td>
</tr>
<tr>
<td>H* LiHu</td>
<td>H* LiHu</td>
<td>H* LiHu</td>
</tr>
<tr>
<td>flat tone</td>
<td>flat tone</td>
<td>flat tone</td>
</tr>
<tr>
<td>syllable</td>
<td>syllable</td>
<td>syllable</td>
</tr>
<tr>
<td>H* LiHu</td>
<td>H* LiHu</td>
<td>H* LiHu</td>
</tr>
<tr>
<td>flat tone</td>
<td>flat tone</td>
<td>flat tone</td>
</tr>
<tr>
<td>syllable</td>
<td>syllable</td>
<td>syllable</td>
</tr>
<tr>
<td>H* LiHu</td>
<td>H* LiHu</td>
<td>H* LiHu</td>
</tr>
<tr>
<td>flat tone</td>
<td>flat tone</td>
<td>flat tone</td>
</tr>
<tr>
<td>syllable</td>
<td>syllable</td>
<td>syllable</td>
</tr>
<tr>
<td>H* LiHu</td>
<td>H* LiHu</td>
<td>H* LiHu</td>
</tr>
<tr>
<td>flat tone</td>
<td>flat tone</td>
<td>flat tone</td>
</tr>
<tr>
<td>syllable</td>
<td>syllable</td>
<td>syllable</td>
</tr>
<tr>
<td>H* LiHu</td>
<td>H* LiHu</td>
<td>H* LiHu</td>
</tr>
<tr>
<td>flat tone</td>
<td>flat tone</td>
<td>flat tone</td>
</tr>
<tr>
<td>syllable</td>
<td>syllable</td>
<td>syllable</td>
</tr>
<tr>
<td>H* LiHu</td>
<td>H* LiHu</td>
<td>H* LiHu</td>
</tr>
<tr>
<td>flat tone</td>
<td>flat tone</td>
<td>flat tone</td>
</tr>
<tr>
<td>syllable</td>
<td>syllable</td>
<td>syllable</td>
</tr>
<tr>
<td>H* LiHu</td>
<td>H* LiHu</td>
<td>H* LiHu</td>
</tr>
<tr>
<td>flat tone</td>
<td>flat tone</td>
<td>flat tone</td>
</tr>
<tr>
<td>syllable</td>
<td>syllable</td>
<td>syllable</td>
</tr>
<tr>
<td>H* LiHu</td>
<td>H* LiHu</td>
<td>H* LiHu</td>
</tr>
<tr>
<td>flat tone</td>
<td>flat tone</td>
<td>flat tone</td>
</tr>
<tr>
<td>syllable</td>
<td>syllable</td>
<td>syllable</td>
</tr>
<tr>
<td>H* LiHu</td>
<td>H* LiHu</td>
<td>H* LiHu</td>
</tr>
<tr>
<td>flat tone</td>
<td>flat tone</td>
<td>flat tone</td>
</tr>
<tr>
<td>syllable</td>
<td>syllable</td>
<td>syllable</td>
</tr>
<tr>
<td>H* LiHu</td>
<td>H* LiHu</td>
<td>H* LiHu</td>
</tr>
<tr>
<td>flat tone</td>
<td>flat tone</td>
<td>flat tone</td>
</tr>
<tr>
<td>syllable</td>
<td>syllable</td>
<td>syllable</td>
</tr>
<tr>
<td>H* LiHu</td>
<td>H* LiHu</td>
<td>H* LiHu</td>
</tr>
<tr>
<td>flat tone</td>
<td>flat tone</td>
<td>flat tone</td>
</tr>
<tr>
<td>syllable</td>
<td>syllable</td>
<td>syllable</td>
</tr>
<tr>
<td>H* LiHu</td>
<td>H* LiHu</td>
<td>H* LiHu</td>
</tr>
<tr>
<td>flat tone</td>
<td>flat tone</td>
<td>flat tone</td>
</tr>
<tr>
<td>syllable</td>
<td>syllable</td>
<td>syllable</td>
</tr>
<tr>
<td>H* LiHu</td>
<td>H* LiHu</td>
<td>H* LiHu</td>
</tr>
<tr>
<td>flat tone</td>
<td>flat tone</td>
<td>flat tone</td>
</tr>
<tr>
<td>syllable</td>
<td>syllable</td>
<td>syllable</td>
</tr>
<tr>
<td>H* LiHu</td>
<td>H* LiHu</td>
<td>H* LiHu</td>
</tr>
<tr>
<td>flat tone</td>
<td>flat tone</td>
<td>flat tone</td>
</tr>
<tr>
<td>syllable</td>
<td>syllable</td>
<td>syllable</td>
</tr>
<tr>
<td>H* LiHu</td>
<td>H* LiHu</td>
<td>H* LiHu</td>
</tr>
<tr>
<td>flat tone</td>
<td>flat tone</td>
<td>flat tone</td>
</tr>
<tr>
<td>syllable</td>
<td>syllable</td>
<td>syllable</td>
</tr>
<tr>
<td>H* LiHu</td>
<td>H* LiHu</td>
<td>H* LiHu</td>
</tr>
<tr>
<td>flat tone</td>
<td>flat tone</td>
<td>flat tone</td>
</tr>
<tr>
<td>syllable</td>
<td>syllable</td>
<td>syllable</td>
</tr>
</tbody>
</table>
Table 1.3: Phonological representations and the corresponding stylized contours of Venlo Dutch (adapted from Hanssen 2005). Solid contours represent Accent 1, interrupted contours represent Accent 2. Shaded portions indicate the stressed syllable; the vertical dotted lines in the second and fourth columns indicate the longer syllable duration for Accent 2.

<table>
<thead>
<tr>
<th></th>
<th>Focus final</th>
<th>Focus nonfinal</th>
<th>Nonfocus final</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Declarative</strong></td>
<td><img src="image1" alt="Declarative Focus final" /></td>
<td><img src="image2" alt="Declarative Focus nonfinal" /></td>
<td><img src="image3" alt="Declarative Nonfocus final" /></td>
</tr>
<tr>
<td>Ac. 1: H* L</td>
<td>Ac. 1: H* L</td>
<td>Ac. 1: L</td>
<td></td>
</tr>
<tr>
<td>Ac. 2: H* L₁, L₂</td>
<td>Ac. 2: H* L₁, L₂</td>
<td>Ac. 2: L₁, L₂</td>
<td></td>
</tr>
<tr>
<td><strong>Low question</strong></td>
<td><img src="image4" alt="Low question Focus final" /></td>
<td><img src="image5" alt="Low question Focus nonfinal" /></td>
<td><img src="image6" alt="Low question Nonfocus final" /></td>
</tr>
<tr>
<td>Ac. 1: –</td>
<td>Ac. 1: H* L₁, L₂</td>
<td>Ac. 1: L₁, L₂</td>
<td></td>
</tr>
<tr>
<td>Ac. 2: H* L₁, L₂</td>
<td>Ac. 2: H* L₁, L₂</td>
<td>Ac. 2: L₁, L₂</td>
<td></td>
</tr>
<tr>
<td><strong>Continuation</strong></td>
<td><img src="image7" alt="Continuation Focus final" /></td>
<td><img src="image8" alt="Continuation Focus nonfinal" /></td>
<td><img src="image9" alt="Continuation Nonfocus final" /></td>
</tr>
<tr>
<td>Ac. 1: H* H</td>
<td>Ac. 1: H* H</td>
<td>Ac. 1: H</td>
<td></td>
</tr>
<tr>
<td>Ac. 2: H* L₁, H₂</td>
<td>Ac. 2: H* L₁, H₂</td>
<td>Ac. 2: L₁, H₂</td>
<td></td>
</tr>
<tr>
<td><strong>High question</strong></td>
<td><img src="image10" alt="High question Focus final" /></td>
<td><img src="image11" alt="High question Focus nonfinal" /></td>
<td><img src="image12" alt="High question Nonfocus final" /></td>
</tr>
<tr>
<td>Ac. 1: L* H₁H₂</td>
<td>Ac. 1: L* H₁H₂</td>
<td>Ac. 1: H₁H₂</td>
<td></td>
</tr>
<tr>
<td>Ac. 2: H* L₁, H₂</td>
<td>Ac. 2: H* L₁, H₂</td>
<td>Ac. 2: L₁, H₂</td>
<td></td>
</tr>
</tbody>
</table>
1.2 Research on tone perception

Studies on the perception of Dutch Limburgian tones are scarce: we found no more than two such experiments in the literature, both related to the dialect of Maastricht. De Bot, Cox and Weltens (1989) presented 14 pairs of Accent 1 and 2 words, which were recognized in only 60 to 65% of the cases. The authors gave two reasons for not drawing definitive conclusions from these rather poor results. First, in the lack of earlier production studies, they were not absolutely sure that the stimuli (recorded by one speaker only) were representative of the dialect under investigation, and second, the listeners had been living in another part of the country for at least a few years. The second study on the Maastricht tones (Gooskens & Rietveld 1995) also led to equivocal results (explained further in section 1.2.2). Our information on tone perception will thus have to be gathered more indirectly, via studies on languages that are typologically quite different from the dialects investigated in this thesis. Still, although the results may often be language specific, the issues raised as well as the methods used can provide useful insights for our own research. The first issue of interest concerns the perception of tones in context. In the first part of this chapter, we saw the crucial influence of context on the shape of East Limburgian tones. Section 1.2.1 inquires on the consequences of contextual influence on the perception of Chinese tones.

1.2.1 Tone in context

1.2.1.1 Tonal coarticulation and contrast neutralization

In languages like Chinese, each syllable is specified on the segmental as well as on the tonal level. When syllables are concatenated to form utterances, coarticulation effects may appear on both levels too. For instance, if the canonical form of a tone prescribes a high ending and the next tone is supposed to start with low pitch, speakers may be tempted to smooth out the transition between these tones by adjusting their original shape. The extent to which a tone is affected by the preceding (carry-over coarticulation) or the following tone (anticipatory coarticulation) is not universal. In Mambila, for instance, Connell (1999) found that adjacent tones have little influence on each other. Each Mambila tone has its own phonetic space and whenever this space is trespassed by another tone, correction occurs very quickly. By contrast, the shapes of Chinese (as well as Thai or Vietnamese) tones can be affected in their height and slope by coarticulation effects (Mandarin: Wu 1984, Xu 1994, Chen & Xu 2006; Thai: Abramson 1979, Gandour, Potisuk, Dechongkit and Ponglorpisit 1992; Vietnamese: Han and Kim 1974).

Xu (1994) investigated the consequences of coarticulation on the perception of Mandarin tones. In his study, he distinguished two sorts of contexts. On the one
hand, conflicting contexts were expected to favour coarticulation effects, for instance a high and a low tone surrounding a rising tone. On the other hand, the compatible contexts, such as two falling tones surrounding a rising tone, should have little impact on the shape of tones. The acoustic analysis of a corpus of three-syllabic words or phrases representing these contexts allowed to confirm their relative influence. Subsequently, the stimuli were manipulated segmentally but not tonally, so as to transform the words or phrases into non-sense sequences with their original tonal structure. Listeners were then asked to identify the second tone in the sequence. In general, recognition was fairly accurate, although better in compatible than in conflicting contexts. In a second experiment, the target syllable was presented without its original context. Whereas the recognition of syllables originating from a compatible context was hardly impaired by the lack of context, scores for the syllables originating from a conflicting context were very low. These results showed that listeners are able to recover the original form of coarticulated tones as long as they dispose of the tonal context responsible for their modification. Presented in isolation, strongly coarticulated tones are no longer identifiable.

Another context-induced transformation of a Mandarin tone is the so-called tone sandhi, in which the low (falling-rising) Tone 3 is replaced by a rising tone, Tone 2, when it precedes another Tone 3. Although articulatory constraints are very likely to play a role in this process, due to the difficulty of pronouncing two fall-rises in a row, Tone 3 sandhi is phonological (cf. chapter 3). A perception study by Wang and Li (1967) showed that listeners could not reliably disambiguate between sequences of Tone 2 – Tone 2 in which the first tone was a lexically specified Tone 2 and those in which the first tone was the result of a tone sandhi, suggesting that the contrast between Tone 2 and Tone 3 was neutralized in this position. This perceptual ambiguity was confirmed by Peng (1996, 1997), who, however, did find a small difference in the production of Tone 2 and the sandhi tone. This difference was not a difference in shape, but in overall $f_0$, with a slightly lower value for the sandhi tone than for Tone 2. Such a discrepancy between production and perception, in which small but systematic production differences are not acknowledged perceptually, is referred to as incomplete neutralization (Port & Crawford 1989). This phenomenon highlights the need of perception studies in order to evaluate the status of linguistic contrasts.

In the case of the tonal dialects studied in this thesis (Roermond and Venlo), no such differences were found in [–focus,–final] contexts. We aim at verifying this finding by recording new stimulus pairs in two different [–focus, –final] contexts, one in which the target word occurs before the nucleus (prenuclear) and one in which it occurs after it (postnuclear). We will analyze acoustically the new recordings and present them to native listeners in order to ensure that neutralization is complete. We will not check for coarticulation effects. Although these effects may occur (albeit to a lesser extent than in Chinese, due to the restrictions that apply to the presence of the lexical contrast), we choose to focus on the structurally motivated modifications of the Accent 1 and Accent 2 contours. Another modification of this type is the tone-intonation interaction, which is the object of the next section.
1.2.1.2 The tone-intonation interaction

The distinction between discourse meanings, such as the question-statement opposition, is often encoded by means of $f_0$. In languages which also use pitch for encoding lexical contrasts, we may expect some competition between the tonal and the intonational domain. This competition can be resolved by integrating both types of information into a pitch contour that will unambiguously represent the intended lexical tone and discourse meaning. As we saw in section 1.1, this option is the one used in East Limburgian dialects. In other languages, such as Mandarin or Cantonese, the situation is different. Both languages can use a rising pitch at the end of yes/no questions (as many other world’s languages do, see Gussenhoven & Chen 2000). If a question ends with a word that should bear a falling tone, the final contour will be, somehow, a (more or less fair) compromise between the intonational rise and the tonal fall. If the compromise prioritizes the intonational information, it may cause a neutralization of the tonal contrast.

Connell, Hogan and Roszypal (1983) simulated the possible influence of intonation on the perception of tones by lowering or raising the final syllables of Mandarin sentences in steps of 20 Hz. The resulting stimuli, which ended with $f_0$ values ranging from 80 Hz (the lowest declarative target) to 240 Hz (the highest interrogative target), were then presented to native speakers in an identification task. Results showed that the identification of tones was mostly stable in a 40 Hz range on both sides of the original signal for Tones 1 and 2, and 20 Hz for tones 3 and 4. These ranges were confronted with results of acoustic studies which assessed the influence of intonation on tones in natural speech, and they were considered wide enough to ensure accurate tone identification. In conclusion, intonation does not appear to threaten the recognition of Mandarin sentence–final lexical tones.

A recent study by Ma, Ciocca and Whitehill (2006) on Cantonese did not yield the same results. When pronounced sentence–finally in yes/no questions, tones 22, 23 and 21 (respectively Low Level, Low Rising and Low Falling) were shown to be acoustically very close to tone 25 (High Rising). This acoustic resemblance was verified perceptually, in that a large majority of the native listeners recruited for testing the perception of tones frequently misidentified the Low tones (22, 23 and 21) as High Rising (25) at the end of interrogative sentences. As pointed out by Ma et al. (2006), the discrepancy between their results an those found for Mandarin may be due to the different sizes of phonetic space used by the tones in these varieties. In Cantonese, more tones have to share a narrow phonetic space, since there are three level tones and two rising tones, so that a slight change is sufficient to make a tone resemble another. The same change applied to a Mandarin tone is less likely to make it ambiguous.

East Limburgian tones are, as we said, very different from (Mandarin or Cantonese) Chinese tones in that their canonical form is already co-determined by intonation. Rather than having one canonical form that may or may not be masked by intonational effects, East Limburgian tones have distinct canonical forms, depending
on the intonation. As we saw in sections 1.1.3.1 and 1.1.4.1, there is no case of neutralization due to the use of a different intonation. Even in the case of the Venlo low questions, where an Accent 1 contour was avoided due to a difficulty of pronunciation (cf section 1.1.4.1), speakers still maintained the binary contrast by importing an Accent 1 contour from another intonation category (‘high question’). Given these facts on the production side, it is unlikely that listeners will systematically fail to identify a lexical tone in one particular intonation. However, this does not mean that the contrast between Accent 1 and Accent 2 will be perceived in the same way in all intonations. Contours have different shapes depending on the intonation used, and the contrasts may be more or less salient across intonations. Such differences in salience may affect recognition; it could also be the case that speakers are in control of the contrasts in their language and that they are perfectly able to detect cues to identification even when these cues are more subtle. Chapter 4 aims at answering the question on the influence of phonetic salience on tone identification. We will see that this question is far from being trivial, especially when it comes to quantifying phonetic salience. On the one hand, salience depends on acoustic properties of the signal. On the other hand, it is determined by language experience, as any new learner of a tone language can testify. At first, it seems almost impossible to identify tones in speech when they are not overemphasized by an empathic teacher. Long-term exposure to the target language usually helps overcome this difficulty. These two aspects of phonetic salience, acoustic cues and language experience, are dealt with in the next sections.

1.2.2 Acoustic cues

Production studies have pointed out the contribution of several acoustic parameters for encoding lexical tones. The first and most important one (without which one wouldn’t even speak of tonal contrasts\textsuperscript{14}) is pitch.

In a great number of studies of tone and intonation, pitch descriptions are directly derived from $f_0$ (fundamental frequency) values, which reflect the frequency with which the vocal folds come together to produce (voiced) speech. $f_0$ is usually rendered in hertz (Hz), corresponding to the number of vocal fold oscillations per second (i.e. the number of times they come together). Although, strictly speaking, $f_0$ and pitch represent different quantities, the former belonging to acoustics and the latter to human perception, merging the two concepts is usually harmless as far as speech is concerned. That is, we may treat $f_0$ as a quite reliable approximation of pitch, so that it is acceptable to manipulate $f_0$ in perception experiments and interpret the results in terms of pitch. This has been done, for instance, to assess the relative importance of pitch in the context of tone perception. Other candidates were duration and intensity, which both have been found to vary systematically within

\textsuperscript{14} If pitch does not play any role in the contrast between words or morphemes, there is no reason to call this contrast “tonal”, except perhaps in cases such as the Venlo dialect where pitch is replaced by duration in a few contexts only.
tonal minimal pairs in various tone languages (e.g. Xu and Whalen 1990, Blicher et al. 1990 for Mandarin Chinese, Garding 1973:4 for Swedish).

In order to assess the relative importance of prosodic parameters in tone perception, one usually removes, in speech signals, all potential suprasegmental cues but one, and tests whether the tone can still be recognized. Such studies were conducted, for instance, for Thai (Abramson 1962:131-134), Mandarin Chinese (Howie 1972, Abramson 1975, Lin 1988), Yoruba (Hombert 1976), and Swedish (Malmberg 1967, Segerup 2004). They all showed that recognition was primarily based on $f_0$, in that the reduction of speech signals to monotonous stimuli (i.e. stimuli with constant $f_0$) has more dramatic consequences on tone recognition than the removal of any other prosodic cues. In many cases (but not in Mandarin Chinese, cf. Whalen & Xu 1992), recognition is simply impossible without $f_0$ information. This does not mean that amplitude and duration do not play any role in tone identification.

It could be shown, for instance, that the recognition of Thai tones was significantly improved in the presence of amplitude information (Abramson 1975, cited in Gandour 1978). This enhancing effect was also suggested for duration. Blicher et al. (1990), for instance, showed that duration could be used for telling apart Mandarin Tone 2 and Tone 3 when their $f_0$ contour was ambiguous: listeners tended to (correctly) identify longer instances of such ambiguous stimuli as Tone 3. Similar results, also related to Mandarin Chinese, were found by Liu and Samuel (2004).

Duration might play an even more significant role in the perception of East Limburgian tones, at least in a number of dialects. As we saw in section 1.1.4.1, the tonal contrast in the dialect of Venlo goes hand in hand with durational differences, Accent 2 being always longer than Accent 1 in sentence-final position. Such a systematic difference was also observed in the more southern dialect of Maastricht (Gooskens & Rietveld 1995) as well as Borgloon (Peters 2007) and Cologne (Gussenhoven & Peters 2004). In addition to a pitch contrast between a fall or rise-fall (Accent 1) and a rise or fall-rise (Accent 2), Accent 2 words were shown to be significantly longer than their Accent 1 counterparts. A perception experiment was designed (Gooskens & Rietveld 1995) in order to find out which prosodic cue, duration or pitch contour, was predominantly used in the process of tonal identification. Eighteen native listeners of the Maastricht dialect were presented minimal pairs of synthesized (standardized) contours for two words, *deur* and *numme*\(^1\). The stimuli, originally short and (rising-)falling for Accent 1 and long and (falling-)rising for Accent 2, were manipulated in order to obtain all combinations of three different durations (short, average and long) and two different contours (the typical Accent 1 and Accent 2 contours). If the judges identified a majority of long stimuli as Accent 2, the conclusion would be that duration is more important than $f_0$; if, on the other hand, the judges based their answers on the $f_0$ contours of the stimuli, duration could not be considered a primary cue for tonal identification. Unfortunately, results were inconclusive. Whereas all *numme* stimuli tended to be

---

\(^1\) *deur* = ‘door’, *deur*\(^2\) = ‘duration’; *numme*\(^1\) = ‘to take’, *numme*\(^2\) = ‘to name’
classified according to their duration, the judgment of the deur stimuli was based on $f_0$.

Although this experiment did not lead to definitive results, it showed in any case that duration should not be neglected in perceptual studies on Limburgian dialects. It is important to verify what role this dimension plays in the encoding of the dialect under investigation, in order to assess its potential influence on perception. As we will see in chapters 2 and 4, the importance of duration can vary from dialect to dialect, so that we will always measure this dimension in our stimuli\textsuperscript{16}. However, we are not going to manipulate the stimuli in order to isolate potential perceptual cues. Throughout this thesis, we will use natural stimuli. Our goal is to find out how contrasts, as they are pronounced in natural speech, are perceived by native listeners. The dimensions that we are interested in and that we want to vary systematically are not on an acoustic level, but on a structural level.

Let us now turn back to the most apparent cue to tonal identification, pitch. The perception of this acoustic variable has been examined further by trying to determine whether pitch height or pitch movement was more important for the identification of tones. Massaro, Tseng and Cohen (1985) found that both cues were necessary for a successful identification of Mandarin tones. Ma et al. (2006) showed the importance of pitch height in their study on the tone/intonation interaction in Cantonese. As we saw in section 1.2.1.2, level tones occurring at the end of yes/no questions were pronounced as rises, which caused the misinterpretation of the Low Level (22) tone as a High Rise (25). Interestingly, the two other level tones present in the Cantonese inventory, both higher than tone 22 (High Level, 55, and Mid Level, 33), attracted high recognition scores even though they were also pronounced as rises. Since Cantonese has no rising tone starting as high as the two higher level tones, the only possible confusion would have been to interpret 33 or 55 as a 25 tone, which was prevented due to the higher onsets of the modified tones. These results show the importance of tone height as a perceptual cue in Cantonese. At the same time, they underline the importance of the tonal inventory for the relative importance of the perceptual cues. In the next section, we report on more experiments that led to this conclusion.

\subsection*{1.2.3 Language experience}

In a perception experiment involving speakers of Cantonese, Mandarin, Taiwanese, Thai and English, Gandour (1983) examined the importance of language background in tone dissimilarity judgements. He presented synthesized stimulus pairs to all subjects, who were then asked to assess their dissimilarity on a scale from 0 to 10. By comparing the scores for different level and contour tones, it could

\footnote{\textsuperscript{16} It may well be the case that intensity displays regular patterns in the dialects under investigation. We will not pursue this issue, since we do not assume that it can have a primary role in the perception of Roermond or Venlo tones.}
be shown that the judgements indeed differed per language group. Speakers of English (being a non-tone language) and Cantonese (see also Ma et al. 2006, in the previous section) appeared to assign more importance to pitch height than the other three groups, while speakers of Thai were more sensitive to pitch movements. A more recent study by Lee, Vakoch and Wurm (1996) also highlighted the influence of language experience by showing that the discrimination of tones was better when subjects had to judge stimulus pairs from their own language. When presented with Mandarin and Cantonese tone pairs, speakers of Mandarin made less mistakes in assessing whether the members of the pairs were the same or different when the pairs represented Mandarin tones, and the same was verified for the Cantonese group. A third group of subjects was included in the experiment, this time involving speakers of a non-tone language (English). This group made significantly more mistakes than the Mandarin and the Cantonese group in judging pairs of the two dialects.

Based on these results, we might speculate that speakers of non-tone languages are somehow less sensitive to pitch-based differences than speakers of tone languages. This does not appear to be the case. Since Klatt’s (1973) study on minimal detectable pitch differences, it has become clear that the human ear is well able to perceive very subtle tonal variations (lower than or equal to 2 Hz). As Yip (2002:290) points out, languages tend to base their contrasts on much larger differences, starting from about 10 Hz. The difference between language groups is, then, not so much a matter of ability as of learned behaviour, on perceptual routines which are dictated by the contrasts defined in the native language. Connell (2000) found, for instance, differences between native and non-native perception of the four Mambila level tones (Tone 4, the highest, is about 10 Hz higher than Tone 3, which has the same distance with Tone 2, which has the same distance with Tone 1). While non-native speakers (who did not appear to have a lower sensitivity to pitch than the native speakers) perceived the four tones equally well, the native group had more difficulty in identifying the Mid tones than the High and Low tones. Similarly, Huang (2001, 2004) found a smaller perceptual distance between Mandarin Tone 2 and Tone 3 in native speakers only. This higher confusability could not be explained solely in terms of acoustic distances, since other tone pairs were comparable in this respect and still were further apart in the perceptual space of native listeners. Furthermore, different results were observed in a control group of non-native (English) speakers, who found the tone pairs 1 (High) - 2 (Rising), 2 - 4 (Falling) and 4 - 3 as confusable as the Tone 2 - 3 pair. Huan concluded that the smaller perceptual distance between Tone 2 and Tone 3 was most likely due to the tone sandhi rule involving these two tones (cf. section 1.2.1.1).

The issue of language-specific tone perception will be of great importance in our third experiment (reported in chapter 4), dedicated to the relationship between perceptual salience and tone identification. If we want to assess perceptual salience in a sensible way, we should be aware of possible language-dependent differences. We will address this question by having native and non-native speakers listen to Venlo stimulus pairs and judging the perceived differences. The next section shows
that language experience also determines the cognitive processes underlying the perception of contrasts.

1.2.4 Categorical perception and (neuro)cognition

Categorical perception, as opposed to continuous perception, refers to the perception of sounds (or other sensory stimuli, such as colours) as discrete classes of events. It is tested for by modifying sounds in a series of equal steps and asking subjects to identify the resulting stimuli in terms of phonological categories, or to discriminate stimulus pairs. If one end of the created continuum represents one phonological category (say, the sound /p/) and the other, a distinct category (e.g. /b/), listeners will not react to the different steps in a gradual way. Instead, they will identify the first few stimuli as a /p/ and after a given point, all other stimuli will be perceived as a /b/. More specifically, the perceived distance between members of the first group, which are all tokens of the same phonological category, is clearly smaller than that between any two members of the distinct groups. Listeners will then show an enhanced discrimination of instances at or near the boundary between categories. By contrast, if a continuum between [p] and an aspirated [ph] is created, no turning point between the two sounds will be observed in speakers of English, since these sounds are not phonologically distinct.

Whereas the existence of categorical perception is widely acknowledged in the segmental – more precisely, the consonantal - domain (Liberman, Harris, Hoffman & Griffith 1957, Eimas, Siqueland, Jusczyk & Vigorito 1971, Harnad 1987), the question whether tone perception is categorical or continuous still has not been answered with confidence (Francis, Ciocca & Ng 2003; see also Gandour 1978 and Hallé, Chang & Best 2004). A recent study by Hallé, Chang & Best (2004) aimed at giving a new impulse to the discussion by comparing tone perception in Taiwan Mandarin vs. French listeners. Three tone-to-tone continua of seven equal steps (e.g. 8 contours representing the high Tone 1 at one end and the rising Tone 2 at the other end) were constructed and presented for identification and discrimination to the native and non-native subjects. Results showed a qualitative difference between the two language groups. Taiwanese listeners appeared to be better than the non-native subjects in both identification accuracy and between-category discrimination performance. In the non-native group, identification was not biased by tone categories. However, the between-group difference was not reflected quantitatively by the typical patterns of categorical perception, so that the authors could only characterize the native performance as “quasi-categorical”. Obviously, the dispute on categorical tone perception is not settled.

A solution to this problem may be brought up by the use of alternative experimental methods in the area of cognitive brain research. A series of studies by Phillips, Pellathy & Marantz (2000) revealed that the categorical perception of phonemes could effectively be demonstrated by recording the brain activity of native listeners. The technique used for the recordings was magnetoencephalography (MEG), which
detects magnetic fields generated in the brain in reaction to experimental stimuli. As in traditional categorical perception experiments, subjects heard a number of acoustically diverse instances of two categories. The identification and discrimination tasks were not necessary. All the subjects had to do was to sit back while their brain reactions were measured by an MEG scanner. In order to highlight the specific brain reactions to phonological contrasts (as opposed to acoustic differences within categories), the stimuli were presented in an auditory mismatch paradigm. In this paradigm, sequences of frequent, (near-)identical auditory stimuli (the standards) are interrupted by infrequent contrasting stimuli (the deviants). The mismatch paradigm is based on the idea that when a sound is heard several times in a row, it leaves a specific pattern in the sensory memory, against which every incoming sound is matched. A deviant sound causes an increased activity in a brain area specialized for the processing of sounds, the auditory cortex. This brain response, called mismatch negativity (MMN), can be detected with an MEG scanner. By measuring the degree (or amplitude) of the MMN, one can assess how subjects apprehend the difference they heard. For instance, if a sequence of identical musical tones (say, a C) is interrupted by a G, the MMN will be larger than if it is interrupted by a D (see Näätänen 2001). In the realm of phonemes, Phillips, Pellathy & Marantz (2000) showed that the mismatch response is stronger when the deviant represents a category change than when it is a token of the same category.

As we can see, the mismatch paradigm seems adequate to enquire further on the categorical perception of tone, although to the present day, it has not been used for this purpose. Our MEG experiment did not intend to fill this gap. Its main goal was rather to answer a question that has been asked repeatedly in the last decades, without reaching any clear conclusions. We wanted to find out in which brain hemisphere tonal and intonational contrasts were processed. Earlier studies could not agree on the localization of these pitch-based linguistic differences. In our view, one of the reasons for this disagreement was that too often, the studies only focused on intonational processing without directly comparing it with tonal processing, or the other way around. Of the few studies that did test tone and intonation in parallel, no one used a language in which these functions were encoded on the same linguistic level. As was shown in the previous paragraphs, discrete and gradient differences are processed differently in the brain. In languages like Chinese (the one that was used for directly comparing tone and intonation), tone contrasts are discrete while intonational contrasts are arguably gradient. Such a difference does not exist in the dialects investigated in this thesis, which encode both lexical and intonational contrasts phonologically. This special status of East Limburgian dialects motivated a new try, which could be done using the mismatch paradigm: We could look for MMN reactions in both hemispheres (in particular in the left and right auditory cortices) and if it was found, we could determine in which hemisphere it was

---

17 The mismatch response was first detected using electroencephalography (EEG). A term more adapted to MEG than mismatch negativity, is magnetic mismatch negativity (MMNm) or mismatch field (MMF). However, the denotation MMN is more widely known than MMF or MMNm, and it is often used in the context of MEG experiments.
stronger, hence answering the question on the lateralization of intonational and tonal processing. This experiment is reported in chapter 5.

Taken together, the investigations reported in this thesis have increased our knowledge of the tone contrast of a typologically very interesting group of tone languages considerably, relative to the sparse and inconclusive data that were available when our research was begun.
2 Perceiving word prosodic contrasts as a function of sentence prosody in two Dutch Limburgian dialects\textsuperscript{18}

2.1 Introduction

Linguistic communication can generally be regarded as a function of the way meanings are encoded in the phonological structure of languages. The perceptual distinctiveness of segments partly depends on a variety of phonological and phonetic factors. For instance, while in many languages the distinction between voiced and voiceless consonants carries a high functional load in nonfinal positions, it is frequently neutralized at the end of a word or syllable. Similarly, the place of articulation of nasal stops may be contrastive in the syllable onset, but not in the coda. Such distributional patterns are generally due to the less favourable conditions obtaining in the neutralizing contexts for the contrasts concerned, and to the inclusion of these limiting conditions as categorical restrictions in the phonology of languages (e.g. Maddieson 1984; Flemming 1996; Boersma 1998). A prosodic factor acting on the vulnerability of contrasts is the variable precision in the articulation as a function of information status. When words represent new or contrastive information, they tend to be produced with greater articulatory care, yielding more prototypical realisations of the segments in them, while the pronunciation of contextually redundant words is often reduced, resulting in shorter words and the undershoot of articulatory targets (Eefting & Nooteboom 1993; van Bergem 1995).

There is growing evidence that such observations at the segmental level also apply to prosodic contrasts. Studies of various tone languages report that the realisation of a lexical tone may be affected by its prosodic context, such as the tones of neighbouring words (Xu 1997), even to the extent that a given configuration loses its

canonical shape as realized in isolated words (Kochanski, Shih & Jing 2003). Sentence intonation can have a strong impact on the melodic shape of tones as well. In Hausa, for instance, polar question intonation suspends the contrast between lexical H and HL (Inkelas & Leben 1991), and many languages change the lexically specified tone at the end of a phrase to distinguish between different utterance types (Ladd 1996:149; Yip 2002:273; Luksaneeyanawin 1993). In addition, analogous to what happens at the segmental level, melodic configurations of tones may be modified by the expression of focus or for emphasis effects (Gussenhoven 2004:15). For instance, when a word is focused in Shanghai Chinese, the phonetic realisation of its lexical tone uses an expanded pitch range, such that an L-tone becomes lower and H-tone higher (Selkirk & Shen 1990); the same is true for Mandarin (Xu 1999). Similarly, Longacre (1952) reported that under stress, the two highest tones of Trique (Oto-Manguean) are pronounced higher and the three lower tones lower than in non-stressed positions. As a result, tonal contrasts become more salient in such privileged positions (cf. also Beckman 1997).

The Dutch Limburgian dialects of Roermond and Weert are of particular interest in this regard. They belong to a group of Limburgian dialects spoken in the south-east of the Netherlands which display a great deal of overlap in their respective vocabularies, as illustrated by words like bein ‘leg’, derm ‘intestine’ and erm ‘arm’. Additionally, they share a word prosodic contrast between what has recently been referred to as Accent 1 and Accent 2 (Gussenhoven & van der Vliet 1999, Gussenhoven 2000a,b). However, while Roermond uses a pitch distinction, Weert relies largely on a duration distinction (Heijmans 2003). Roermond is tonal in the real sense of the word in that it marks lexical distinctions by melodic means. The exact realisation of these tones is quite complex, and varies with the intonation contour (declarative vs. interrogative intonation), the information status of the word (within the focus or outside it) and the position of the word in the intonational phrase (final or nonfinal). Interestingly, nonfocused, nonfinal contexts have been claimed to neutralize the distinction (Gussenhoven 2000a,b). The dialect of Weert, by contrast, may well have been tonal in the past, in which case it has replaced the melodic cues with durational ones (Heijmans 2003, Verhoeven 2003). Although it has been suggested that the vowel duration difference is accompanied by an alignment difference of a rising-falling pitch configuration (Verhoeven & Connell 1992), it would appear to be the vowel duration difference which is phonologically relevant. Focus will have a general effect on articulatory precision (e.g. de Jong 1995), and thus also on the realization of vowel quantity contrasts. Focus is expressed by means of pitch accents in Dutch, and pitch accented syllables are longer than equivalent unaccented syllables (‘accentual lengthening’, Cambier-Langeveld 1999). However, as pointed out by Sieb Nooteboom (personal communication), while quantity contrasts are commonly neutralized in unstressed syllables, as in Dutch (Rietveld, Kerkhoff & Gussenhoven 2004), there are no examples of languages that neutralize a quantity contrast outside the focus constituent.

A second possible difference between the two phonological contrasts concerns the extent to which acoustic cues may be non-local. The prosodic contrast employed for
the lexical distinctions in Roermond is integrated with the intonational structure of the sentence, and it is therefore possible that the cues to the perception of the lexical tone contrast are less reliably confined to the speech signal for the word in question, and may in part depend on the correct interpretation of the intonation contour for the sentence as a whole. On the other hand, a quantity contrast is locally detectable, and is unlikely to have acoustic cues across the boundaries of the word in which it occurs.

Given these observations, it is not unreasonable to assume that the difference between the tonal encoding in Roermond and the quantity encoding in Weert will have significant consequences for the perception of contrasts across intonational conditions. The discriminability of the tonal distinctions in Roermond is likely to vary from one intonational context to another, while the quantity opposition in Weert is likely to be less dependent on intonational context and position in the intonational phrase: in this respect, the quantity contrast is similar to segmental contrasts, which will be similarly robust against such variation in the context. Finally, the discriminability of the prosodic contrast in Weert should not be expected to depend on the interpretation of the intonation contour for the sentence, as it may be in Roermond. As a result, the excision of words from their sentence contexts is arguably more harmful to the perception of the contrast in the dialect Roermond than in the dialect of Weert.

In order to provide a better insight into the salience of the prosodic contrasts in these dialects and achieve an understanding of their communicative function, a cross-dialectal perception study was conducted. Following a methodological procedure for functional, crosslinguistic analyses of prosodic distinctions outlined in Swerts, Krahmer & Avesani (2002), we aimed to find out to what extent the perception of prosodic contrasts is dependent on whether it is signalled by $f_0$ (Roermond) or duration (Weert). It was expected that this distinction is perceivable in all prosodic contexts in Weert, whereas in Roermond the distinction may be neutralized in the nonfocused, nonfinal context. In addition, it was expected that the prosodic distinction in Weert remains robust when words are excised from their context, whereas for Roermond the contrast will be more vulnerable when listeners do not have access to the wider intonational contour.

Because we were interested in the perception of the word prosodic contrast in natural utterances, we conducted the perception experiment with the help of spoken data rather than with stimuli with artificially manipulated $f_0$ and duration specifications. In order to be able to relate the perception results to the stimuli we used to obtain them, we report $f_0$ and duration measurements for the experimental words for the two dialects separately, for each of the prosodic contexts used. This will be done in section 2.3. First, however, section 2.2 will describe the selection of lexical minimal pairs, the prosodic contexts and methodological procedures. Section 2.4 will present the results of the perception experiment, while section 2.5 will offer a conclusion.
2.2 Method

2.2.1 Target words and carrier sentences

Instead of directly eliciting judgements on the occurrence of the phonological categories by referring to ‘stoottoon’ and ‘slepeettoon’, terms that naive speakers in both communities are quite generally familiar with, it was decided to exploit the morphological use of the word accent distinction to express grammatical number. Not only did this enable us to use a task in which subjects identified the singular or plural form of nouns, it also provided us with a convenient base line, in that singular or plural forms of nouns could be included that differ in other aspects than the prosodic feature under investigation. As observed in section 2.1, the dialects of Roermond and Weert are sufficiently close together for there to be a usable number of nouns with singular and plural forms that are etymologically equivalent across the dialects. Nine nouns were selected, six of which distinguish their singular and plural forms purely by prosodic means. That is, barring tonal features and durational differences, the singular and plural forms are segmentally the same (but see section 2.5), and are spelt the same, as shown in Table 2.1. The last three nouns in the table were used to provide a base line for the recognition of grammatical number as expressed by segmental means. Two of these, teen and kleur, use a plural suffix /-e/, while man uses an unlauted vowel in the plural. Moreover, in kleur and man, the plural form differs from the singular both in having a different vowel and in having Accent 2 rather than Accent 1.

Table 2.1: Target words recorded for the perception experiment; I and II (in superscript) represent Accent 1 and 2, respectively.

<table>
<thead>
<tr>
<th>Singular form</th>
<th>Plural form</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>bein² (/bein¹/)</td>
<td>bein¹</td>
<td>‘leg(s)’</td>
</tr>
<tr>
<td>derrm² (/derrm¹/)</td>
<td>derrm¹</td>
<td>‘intestine(s)’</td>
</tr>
<tr>
<td>erm² (/erm¹/)</td>
<td>erm¹</td>
<td>‘arm(s)’</td>
</tr>
<tr>
<td>knien² (/knien¹/)</td>
<td>knien¹</td>
<td>‘rabbit(s)’</td>
</tr>
<tr>
<td>pin² (/pin¹/)</td>
<td>pin¹</td>
<td>‘wooden nail(s)’</td>
</tr>
<tr>
<td>sjtein² (/sjtein¹/)</td>
<td>sjtein¹</td>
<td>‘stone(s)’</td>
</tr>
<tr>
<td>teen² (/teen¹/)</td>
<td>tene²</td>
<td>‘toe(s)’</td>
</tr>
<tr>
<td>kleur¹ (/kleur¹/)</td>
<td>kleure²</td>
<td>‘colour(s)’</td>
</tr>
<tr>
<td>man² (/man¹/)</td>
<td>men¹</td>
<td>‘man (men)’</td>
</tr>
</tbody>
</table>
The singular and plural form of each of these nine nouns were embedded in five carrier sentences which each represent a different prosodic condition. These carrier sentences were short, and typically consisted of a single intonational phrase. An important requirement on their composition was that they should not contain morphological or semantic indications that reveal the grammatical number of the nouns in question. Since both dialects have premodifying elements that reveal this information (e.g. gene knien for ‘no rabbit’ vs gen knien for ‘no rabbits’ in Roermond Dutch), metalinguistic carrier sentences were used such as Ich höb KNIEN geheurd ‘I heard “RABBIT(S)”’.

The five prosodic conditions and the carrier sentences and glosses for both dialects are given in (1), with the noun for ‘rabbit(s)’. Two of the five prosodic conditions featured the noun in focus position, i.e. in final and in nonfinal position. The three other conditions keep the noun outside the information focus of the sentence. In one of these the noun is phrase–final, while in the two remaining ones the word occurs in nonfinal position, i.e., in prenuclear or in postnuclear position. In (1c), the noun is phrase–final, while in the remaining two conditions the word occurs in nonfinal position. In (1d1), it occurs before the nuclear syllable, while in (1d2) it occurs in the postnuclear stretch. These positions were selected on the basis of claims in Gussenhoven & van der Vliet (1999) and Gussenhoven (2000a,b) about the preservation of the contrast in the dialects of Venlo and Roermond. Specifically, in both dialects, the [+focus, +final], [+focus, –final] and [–focus, +final] conditions are described as having the contrast and the postnuclear [–focus, –final] condition as not allowing the contrast. The prenuclear [–focus, –final] position is not discussed in either of the above descriptions.

(1)  a. [+focus, +final]
    Roermond Dutch (RD): In ’t Remunjs zaes-se geweun ‘KNIEN’.
    Weert Dutch (WD): In ’t Wieërs zegkje gewoeën ‘KNIEN’.
    ‘In Roermond/Weert Dutch, you just say “RABBIT(S)”’.

b. [+focus, –final]
    RD: [Wat höbs-se geheurd?] Ich höb ‘KNIEN’ geheurd.
    WD: [Wat hejae gehuuërdj?] Ich heb ‘KNIEN’ gehuuërdj.
    ‘What did you hear? I heard “RABBIT(S)”’.

c. [–focus, +final]
    RD: [Eers ZAG hae ‘knien’,] toen SJREEFDE hae ‘knien’.
    WD: [ZEGJE ‘knien’?] Nein, ich HUUER ‘knien’.
    RD: ‘First he SAID “rabbit(s)”, then he WROTE “rabbit(s)”’.
    WD: ‘Do you SAY “rabbit(s)”? No, I HEAR “rabbit(s)”’.

d1. [–focus, –final], prenuclear
    RD: [Hōbs-se ’knien’ GEZAG?] Nae, ich höb ’knien’ GEZÖNGE.
    WD: [Hejae ’knien’ GEZAGDJ?] Nein, ich heb ’knien’ GEZÖNGE.
    ‘Did you SAY “rabbits”? No, I SANG “rabbits”’.
The two grammatical numbers (‘singular’ and ‘plural’), five prosodic conditions and nine nouns yielded 90 sentences, which were arranged in four different random orders. They were recorded by 3 female and 3 male native speakers in Roermond and 3 female and 2 male native speakers in Weert. The grammatical number of the orthographically ambiguous nouns was in each case clearly indicated on the reading list. For each dialect, an expert native speaker assisted in monitoring the correctness of informants’ deliveries. Incorrect deliveries were read again at the end of the session. Recordings were made by means of a DAT-recorder with a sampling rate of 44.1 kHz. Subsequently, two male and two female speakers were selected for each dialect who appeared to the expert native speakers as the most competent readers. Among the various pronunciations of the same sentences by the same speaker, the most fluent and representative version was selected in a similar way. This procedure yielded 360 sentences for each dialect.

Subsequently, the experimental nouns were excised from these 360 speech files for each dialect. This was done to test the hypothesis that excision of the nouns from the sentences is more detrimental to the perception of prosodically encoded grammatical number in a ‘tonal’ dialect like Roermond than in a ‘durational’ dialect like Weert, considering that tonal information is more likely to be linked to the prosody of the sentence as a whole than durational information is.

2.2.2 Experimental tapes

The stimuli were presented to native listeners in a perception experiment, for which two experimental tapes were prepared. The only difference between the two tapes was the order of the stimuli: where a singular form appeared on the first tape, its corresponding plural form was found on the second, and vice versa. This was done to compensate for order effects. Each tape consisted of two presentation blocks. The first block contained all 360 sentences and the second block contained 360 target words that had been extracted from the carrier sentences in the first block. The presentation order of stimuli in the second block was complementary to that of the first block in that a singular form appeared where a plural appeared in the first block.

19 All recordings and perception experiments were performed in classrooms, in Roermond and in Weert.
and vice versa. The interstimulus interval was 5 s in both blocks. The stimuli were not numbered on the tape, but after 10 stimuli informants heard a short orientation signal.

2.2.3 Judges and procedure

Eighteen judges participated in the Roermond experiment, and twenty-three in the Weert experiment. All judges had confirmed that they were native speakers of the respective dialects and reported using the dialect in most of their daily social encounters.

Thirteen of the 23 Weert judges listened to the first experimental tape and the other ten to the second. In Roermond, eight of the 18 judges listened to the first tape and the remaining ten to the second. They were recruited from the student population in the senior classes of local secondary schools. Their average age was 17.6 years in Weert and 17.2 in Roermond. They were not told about the real objectives of the experiment and were paid a small fee for their participation.

The judges were seated in a quiet room and were given a scoring sheet to record their judgements. This sheet contained the target words of all the stimuli in their singular form and for each stimulus two boxes, i.e. one marked ‘singular’ and the other ‘plural’. After a short introduction, informants listened to all sentence stimuli via headphones and recorded their judgements. The same procedure was used for the word stimuli. In total, the two sessions took about an hour, excluding the break.

2.3 Acoustic analysis

In this section, we present acoustic data for the prosodically marked minimal pairs of words that were excised from the sentence stimuli. They formed a set of 2 (‘singular’ and ‘plural’) × 5 (prosodic conditions) × 5 (nouns) × 4 (speakers), or 200 excised words in each dialect. The scores for one of the six nouns we recorded were discarded. After the experiment had been run it appeared that our (young) listeners were not familiar with either the meaning or the prosodic plural form of the word *pin* (‘wooden nail’), to which they would add a schwa-suffix to form a plural if they were to use the word.

Our objective is not just to provide the information necessary to understand the results of the perception experiments, but also to provide information on the phonetics of the dialects to supplement the information given in Verhoeven & Connell (1992), Verhoeven (2002, 2003) for Weert and Gussenhoven (2000a,b) for Roermond. We measured the $f_0$ and duration within the rhymes of each target word. All segmentations were done manually, taking auditory as well as spectral properties into account. Within the words, segment boundaries were determined between onset
and rhyme and inside the rhyme between nucleus and coda\textsuperscript{20}. We report the \( f_0 \) data in section 2.3.1 and the duration data in section 2.3.2.

### 2.3.1 \( f_0 \) contours

All \( f_0 \) values were computed automatically using the Praat software (Boersma & Weenink 1992-2004, version 4.2). The same script was run for both dialects, which computed the following measures:

- the \( f_0 \) at the beginning and end of the rhyme
- the maximum and minimum \( f_0 \) in the rhyme
- the maximum and minimum within the nucleus and the coda

The \( f_0 \) measurements were recorded together with the time of their occurrence. Below, we report the average data pooled across words and speakers. In general, we observed the same order and approximate timing of minima and maxima within the rhyme. The maxima preceded the minima and occurred near the beginning of the rhyme. There were 22 exceptions to this tendency in the Roermond data, of which 13 occurred in the [+focus,–final] context for Accent 2, and 20 in the Weert data, of which 14 occurred in the prenuclear context (1d1). Below, we have kept these exceptional cases out of the calculations of the means; we will discuss the exceptional cases wherever they present a consistent pattern in a given context.

#### 2.3.1.1 Roermond

In the [+focus,+final] context, the maximum \( f_0 \) in Accent 1 occurs early in the rhyme, either right at the beginning or preceded by a small rise or plateau. From this maximum, \( f_0 \) falls for 85 Hz on average, until a minimum is reached at 19 ms from the end of the rhyme. After that, the contour is virtually flat. The contour of Accent 2, represented by the interrupted graphs in Figure 2.1, reaches a maximum close to the beginning of the rhyme, which is likewise followed by a fall of 44 Hz. Other than in the case of Accent 1, a rise occurs after the fall’s end point at 125 ms from the end of the rhyme. This contour is shown in panel (a) of Figure 2.1, that displays average beginning and end points as well as minima and maxima of the \( f_0 \) contours.

\textsuperscript{20} A special situation was found for the words erm and derm, which were often pronounced with a schwa between /r/ and /m/. This schwa insertion in word-final consonant clusters is quite common in Dutch dialects (see for instance Gussenhoven & Aarts 1999, for the dialect of Maastricht) as well as in the standard language (Swerts et al. 2001). As a study by Donselaar et al. (1996) suggests, it does not imply an increased number of syllables in the transformed word. The same can be postulated for Roermond Dutch: In final bisyllabic trochaic words, like erre ‘R’s’, the \( f_0 \) features of nonfinal syllables are observed; by contrast, \textit{erm} has the \( f_0 \) characteristics of a phrase-final syllable.
Note that in this figure, duration has been standardized and can only be read off as a percentage of the rhyme duration.

<table>
<thead>
<tr>
<th>Roermond</th>
<th>Weert</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. +focus/+final</td>
<td>a. +focus/+final</td>
</tr>
<tr>
<td><img src="image1" alt="Graph" /></td>
<td><img src="image2" alt="Graph" /></td>
</tr>
<tr>
<td>b. +focus/-final</td>
<td>b. +focus/-final</td>
</tr>
<tr>
<td><img src="image3" alt="Graph" /></td>
<td><img src="image4" alt="Graph" /></td>
</tr>
<tr>
<td>c. -focus/+final</td>
<td>c. -focus/+final</td>
</tr>
<tr>
<td><img src="image5" alt="Graph" /></td>
<td><img src="image6" alt="Graph" /></td>
</tr>
</tbody>
</table>
Figure 2.1: Average contours for Accent 1 (solid lines) and Accent 2 (dashed lines) in five different contexts (panels (a) to (d2)), for Roermond (left) and Weert Dutch (right). Time is expressed as the percentage of the average duration in each combination of accent and context.

Like its counterpart in the final context, the Accent 1 contour in the [+focus,–final] context, shown in panel (b), is characterized by a fall, which may again be preceded by a rise, in more cases in fact than in the final context. The fall is steeper than that in (1a), covering 98 Hz in less time. This is in part explained by the fact that the fall needs to be completed within the syllable, which in phrase-internal position is shorter than in phrase-final position. In the contour for Accent 2 in this context, we observe a minimum at 38 ms from the beginning of the rhyme, after which the pitch goes up for 30 Hz until a maximum at 71 ms from the end of the rhyme (that is, some 120 ms later than for Accent 1 in this condition) and then falls again. However, these averages only represent thirteen out of the 20 instances of Accent 2 in the [+focus,–final] condition. According to Gussenhoven (2000a), we might have expected more cases in which the minimum \( f_0 \) precedes the maximum, suggesting a rising pattern. There are two reasons for an invered order of the minimum and maximum \( f_0 \). First, in the word *knen*, the rise already starts before the rhyme, during the onset /n/, continuing until roughly 50 ms from the start of the rhyme. The rise thus begins in the voiced part of the onset, leaving more space for the fall. Second, various minor fluctuations in the contours have led to somewhat misleading
minimum and maximum values. In such cases, an initial close-copy stylisation ('t Hart, Collier & Cohen 1990) would have brought out the overall rising pattern more adequately.

In the [-focus,+final] context, shown in panel (c), a slow, steady fall of 34 Hz is observed in Accent 1, with a flat continuation after the minimum is reached. For Accent 2, a fall of 25 Hz typically starts immediately at the beginning of the rhyme, and after reaching a minimum at 165 ms into the rhyme, the contour rises again to the phrase-end by some 10 Hz.

In the prenuclear context, there are three cases for Accent 1 in which the minimum occurs before the maximum. In the remaining 17 cases, \( f_0 \) falls over 25 Hz, but unlike what happens in final position outside the focus, there is a slight rise after the minimum, most probably in anticipation of the nuclear H* tone, as can be seen in panel (d1). In the three cases in which the maximum occurs after minimum, the same pattern is observed, except that the rise starts earlier in the rhyme and \( f_0 \) then exceeds the prenuclear maximum. As will be clear from panel (d1), the contour for the Accent 2 counterpart is much the same.

In the postnuclear nonfinal context, shown in panel (d2), a similar set of contours is found for both Accent 1 and Accent 2 as in the prenuclear context, except that the post-minimum stretch is virtually flat, and thus does not show the anticipatory rise towards the following H* and the fall is longer. As in panel (d1), the maximum occurs almost at the beginning and the \( f_0 \) fall is quite slow.

These data broadly conform to those in Gussenhoven (2000a,b). The neutralization that was claimed for the postnuclear nonfinal context would appear to be confirmed in our data, which also suggest that neutralization may occur in prenuclear contexts, which was not dealt with in our earlier investigations. However, the differences found in this context are not sufficiently clear to allow a conclusion at this point.

2.3.1.2 Weert

We briefly discuss the contour shapes we found in the same contexts in the Weert data, which were similarly pooled across words and speakers.

In the [+focus,+final] context, both Accent 1 and Accent 2 have a falling \( f_0 \), with no observable difference between them, as shown in panel (b). There are a fair number of cases where a slight rise can be observed towards the rhyme end. In the [+focus,—final] context, both Accent 1 and Accent 2 have relatively high \( f_0 \) peaks, from which the contour falls towards the end of the rhyme. While the peak in the nonfinal context is later than in the final context, the \( f_0 \) excursions of the falls in both contexts are comparable, amounting to some 60 Hz. The difference in peak alignment between final and nonfinal occurrences is reminiscent of a similar finding for English referred to by Pierrehumbert and Steele (1989).
In the three nonfocused contexts shown in panels (c), (d1) and (d2), similar contours are observed, again with no appreciable difference between the two word accents, with the possible exception of the final context, where the Accent 2 contour reaches a somewhat lower minimum (panel c). In all cases, there is a shallow fall throughout the rhyme. Since these contexts place the target word in unaccented position, it is reasonable to assume that the falling $f_0$ is due to declination.\footnote{In the final nonfocused context, the fall is wider, about 46 Hz, and steeper than in the nonfinal nonfocused contexts, where falls cover about 23 and 32 Hz, respectively. The difference will in part be due to the fact that fourteen words, seven for each accent, were excluded from the calculations, as their minima preceded the maxima. As it happened, all fourteen cases were spoken by female speakers.}

### 2.3.2 Duration

Vowel durations were determined for all 400 words, and two Analyses of Variance were performed on the data for Roermond and Weert separately, with ACCENT (two levels) and CONTEXT (five levels) as independent variables. In Roermond Dutch, Accent 2 is on average longer than Accent 1 (172 ms against 159 ms), which difference is significant ($F(1,2)=8.265$, $p=0.005$). However, as can be seen in panel (a) of Figure 2.2, the durational overlap between the two accents is considerable. Looking at individual results, we found that there were 30 cases out of 100 noun pairs in which Accent 1 was longer than Accent 2, distributed over all contexts. The effect of prosodic context was found to be significant ($F(1,4)=11.531$, $p<0.001$), with an average duration of 178 ms in the [+focus,+final] context, 162 ms in [+focus,–final], 186 ms in [+focus,+final], 143 ms in prenuclear [+focus,–final] and 159 ms in postnuclear [–focus,+final]. The [–focus,+final] context differed significantly from [+focus,–final] as well as from pre- and postnuclear [+focus,–final]; a significant difference was also found between [+focus,+final] and prenuclear [–focus,–final]. As we can see, these differences always involve one final and one nonfinal context, which reflects a general tendency for syllables to have different durations in final and nonfinal contexts.

In Weert, there was likewise a significant difference in vowel duration between Accent 1 and Accent 2 ($F(1,2)=63.637$, $p<0.001$). Average vowel duration in Accent 1 was 148 ms, while that for Accent 2 amounted to 213 ms, a larger difference than that found for the Roermond speakers. There was also less overlap, as will be clear from a comparison between panels (a) and (b). The difference is consistent with findings reported by Verhoeven & Connell (1992) and Verhoeven (2003). Unlike what we found for Roermond, vowel duration between the different contexts is not significant in the Weert data. In both dialects, the interaction between vowel duration and context is not significant.
Clearly, these data confirm that the Roermond dialect is largely ‘tonal’ and the Weert dialect largely or wholly ‘durational’. The absence of a significant interaction between context and accent in the Weert data suggests that the durational difference is equally present in all contexts. This finding should be contrasted with the obvious effect of context on the difference between the $f_0$ contours of Accent 1 and Accent 2 in the dialect of Roermond, as shown in the left half of Figure 2.1. We now turn to the results of the perception experiment.

2.4. Perception experiment: Results

2.4.1. Statistical analysis

Because we wanted to be sure that all judges in fact had sufficient knowledge of the dialect they claimed to speak, we carried out a post-hoc check on their performance by correlating their scores with the average of the scores for the other judges in the same group of listeners. Taking a Pearson correlation coefficient ($r$) of 0.25 as the minimum required for inclusion, two of the judges in the Weert group had to be excluded. In the Roermond group, no judge had a lower $r$ than 0.34. The mean $r$ for Roermond was 0.52, and for Weert 0.41.

The cross-dialectal perception experiment was designed to answer the two following important questions: whether the identification of singular and plural forms of
nouns, our operational definition of the contrast between Accent 1 and Accent 2, is influenced by the prosodic context in one or both dialects; and secondly, whether the recognition of isolated words is less successful than in stimuli representing the full sentences. Moreover, it would be interesting to compare the answers to these questions with those obtained for singular-plural differences that are segmentally encoded. It was decided to apply the CART method (Classification And Regression Trees, see Breiman et al. 1984), whose results more directly address the answers to these questions than an Analysis of Variance. Not only does CART enumerate the significant factors at play in a given situation, but it also gives a concise sketch of the way in which these factors are organized. Instead of a list of interactions, a tree is constructed that shows the relative importance of each factor on the dependent variable. In addition to this hierarchy, information is given about the exact value of the dependent variable in function of the significant effects.

Four CART analyses were performed on the responses from the perception experiment, one for the data from the sentence stimuli and one for those obtained from the word stimuli, for each dialect separately. The candidate predictors were NUMBER (singular or plural), STIMULUS_TYPE (sentence or word), CONTEXT (five levels), SPEAKER (S1 to S4) and WORD (w1 to w5 for the prosodic minimal pairs; w1 to w3 for the segmental ones).

Figure 2.3 illustrates the partitionings made for the five words with prosodic number marking bein, erm, derm, knien and sjein in Roermond Dutch. It indicates that no split was found for the predictors STIMULUS_TYPE, SPEAKER and WORD. The two predictors that were retained by the partitioning program were CONTEXT and NUMBER.

![Figure 2.3: CART found for the Roermond Dutch prosodic minimal pairs (7200 observations in total), presented as words or as sentences. The percentages specified in each node represent the mean recognition rates in this particular condition.](image)
Importantly, the [+focus] and [–focus,+final] contexts appear in one partition, leaving the nonfinal, nonfocused contexts in the complementary partition. The first three represent focus and/or phrase-final positions, and yield high scores for all stimulus types, for all speakers and words, in singular as well as in plural forms. The remaining prenuclear and postnuclear contexts represent the phrase-internal positions outside the focus. Recognition scores in this partition are considerably lower than in the complementary set of contexts. This is shown in panels (a) and (c) of Figure 2.4, which plot the recognition scores for the singular and plural forms in the sentence and word conditions separately for the dialect of Roermond. The prenuclear and postnuclear contexts attract lower scores in both modes of presentation.

**Figure 2.4: Recognition scores of the prosodically marked minimal pairs as a function of context, number and dialect, for the sentence (above) and word stimuli (below). Contexts a, b and c represent the nuclear and/or final contexts, and contexts d1 and d2 represent the prenuclear and postnuclear nonfinal positions.**

Within the scores for the prenuclear and postnuclear contexts, a significant difference was observed between singular and plural forms, as shown by the lower partition in Figure 2.3. From Figure 2.4, it is clear that singular forms (Accent 2) were more likely to be identified correctly than their plural counterparts (Accent 1). More specifically, both singular and plural forms were better identified when presented within their carrier sentence in a focused, final prosodic context, where
more than 90% of the items were correctly identified. Although somewhat lower, recognition scores remain high in the [+focus,–final] and [–focus,+final] positions. We begin to see that singulars are better recognized than plurals (84.7% against 72.8%) in the [–focus,+final] context. This difference is larger still in the nonfocused, nonfinal contexts, in which the proportion of correctly identified plural forms drops below 50% (40.6% in the prenuclear context and 45.3% in the postnuclear context, while the scores for singular forms are still well above chance (69.2% in both contexts).

The results obtained with the word stimuli show these tendencies to an even stronger degree. While the ‘best’ context, [+focus,+final], attracts scores above 90%, with recognition rates in the [+focus,–final] context only slightly below those in the sentence condition, the gap between plurals and singulars widens even more in the [–focus, +final] context than in the sentence stimuli. Singulars are recognized correctly in 91.7% of the cases, while those for plurals drop to 60.6%. In the nonfocused, nonfinal contexts, the difference between singular and plural increases from about 25% to more than 50%. While the differentiation in the scores as a function of context was expected, the difference in recognition rates between singular and plural forms was unexpected. Both issues are discussed in section 2.4.2.

The CART analysis for Weert Dutch led to the most basic tree, without any split from the root. The partitioning algorithm was also run on separate data sets, first on the sentence stimuli and then on the word stimuli, in all cases without effect on the results. Panels (b) and (d1) in Figure 2.4 show the recognition scores for the Weert data in the same format as for the Roermond data. Clearly, the Weert data do not show the degradation in the recognition in the phrase-internal nonfocused contexts, in either mode of presentation. Neither does there appear to be any differentiation between the scores for the singular and the plural forms. Scores for both sentence stimuli and the word stimuli are high in all five contexts, i.e. around 80%.

Turning now to the results for the segmentally encoded minimal pairs, these recognition scores are just over 95% in Roermond and 89% in Weert. As expected, the CART analysis produced no partitions in the Roermond data, but no fewer than seven splits in the Weert scores, involving all predictors except NUMBER. A closer look at the data revealed that this high number of significant effects characterized only half of the results, viz. the recognition scores for the excised words. Separate analyses for the two types of stimuli with the four remaining predictors showed no splits in the sentence data, whereas a complex tree appeared again with the word stimuli. This result can be attributed to the extremely low number of incorrect scores, which because they are so rare may easily be unevenly distributed over the various categories. In the absence of other sources of variation, the CART program will tend to exaggerate the importance of this uneven distribution and as a result, to produce splits based on a few incidental mistakes. These results will not be discussed further. The main finding is that, apart from a few recognition errors, recognition scores of singular and plural forms are very high when segmental cues are at play.
In order to evaluate the more detailed results displayed in Figure 2.4, two Analyses of Variance were carried out on the Roermond and Weert prosodic data separately, with scores pooled over the five words, and with STIMULUS_TYPE (sentence or word), NUMBER (singular or plural), CONTEXT (5), and SPEAKER (4) as factors. Broadly, this analysis confirmed the results in Figure 2.3. Not surprisingly, the number of significant interactions and main effects (as displayed in Table 2.2) is higher than could be revealed in the CART analyses. First, there were significant interactions and main effects for SPEAKER in both data sets, reflecting the fact that not all four speakers equally effectively conveyed the grammatical number of the nouns. Judging by the effect sizes, expressed by $\eta^2$ (which ranges from 0 to 1), differences were larger among the Weert speakers than the Roermond speakers (Table 2.2). Second, NUMBER and STIMULUS_TYPE were found to be significant in Roermond, but not in Weert, which confirms the results of the CART analyses. CONTEXT was significant in both data sets, but the effect was small in Weert, as is clear from panels (b) and (d) in Figure 2.4. The effect must be due to the very small but regular fall-off in recognition rates towards the less salient nonfocused contexts, while the interaction with NUMBER is reflected in the irregular paths of the graphs across the contexts and presentation conditions. Turning to the results for Roermond, significant effects for STIMULUS_TYPE, NUMBER and CONTEXT were found, with particularly high $\eta^2$'s for NUMBER and CONTEXT (Table 2.2). The $\eta^2$ for CONTEXT in the Roermond data is 0.90, compared to just 0.29 for Weert. The striking increase in the difference in the recognition rates between plurals and singulars across contexts is expressed by the interaction between CONTEXT and NUMBER, and the fact that this trend is more extreme in the single word than in the sentence stimuli is expressed by the three-way interaction between NUMBER, CONTEXT and STIMULUS_TYPE. Finally, when averaged over the five prosodic conditions (as shown in Figure 2.4), the singulars were recognized better, but plurals worse, in the single-word condition. This interaction between STIMULUS_TYPE and NUMBER is significant (Table 2.2), and the $\eta^2$ of 0.67 is substantial. These findings are discussed in section 2.4.2.

---

22 We similarly performed ANOVAs on the scores for the segmental minimal pairs. Unhelpfully, all effects and interactions appeared to be significant in both dialects, except for the STIMULUS_TYPE in Weert and the interaction NUMBER*SPEAKER in Roermond. In view of the very high recognition scores in all conditions, these significant results most probably reflect small variations within a narrow range just below a ceiling.
Table 2.2: Significant main effects and interactions for the factors NUMBER, CONTEXT, STIMULUS_TYPE and SPEAKER for the scores as pooled across five nouns with prosodically encoded grammatical number in the dialects of Roermond and Weert, with Huynh-Feldt corrected p-values (p ≤ 0.01) and partial \( \eta^2 \).

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>F</th>
<th>Sig.</th>
<th>Partial ( \eta^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Roermond</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stimulus type</td>
<td>1.00</td>
<td>9.581</td>
<td>0.007</td>
<td>0.360</td>
</tr>
<tr>
<td>number</td>
<td>1.00</td>
<td>44.157</td>
<td>0.000</td>
<td>0.722</td>
</tr>
<tr>
<td>context</td>
<td>3.16</td>
<td>146.459</td>
<td>0.000</td>
<td>0.896</td>
</tr>
<tr>
<td>speaker</td>
<td>3.00</td>
<td>12.683</td>
<td>0.000</td>
<td>0.427</td>
</tr>
<tr>
<td>stimulus_type * number</td>
<td>1.00</td>
<td>33.631</td>
<td>0.000</td>
<td>0.664</td>
</tr>
<tr>
<td>number * context</td>
<td>4.00</td>
<td>63.435</td>
<td>0.000</td>
<td>0.789</td>
</tr>
<tr>
<td>stimulus_type * number * context</td>
<td>4.00</td>
<td>10.785</td>
<td>0.000</td>
<td>0.388</td>
</tr>
<tr>
<td>stimulus_type * speaker</td>
<td>2.35</td>
<td>5.632</td>
<td>0.005</td>
<td>0.249</td>
</tr>
<tr>
<td>stimulus_type * number * speaker</td>
<td>3.00</td>
<td>5.945</td>
<td>0.001</td>
<td>0.259</td>
</tr>
<tr>
<td>number * context * speaker</td>
<td>11.17</td>
<td>6.923</td>
<td>0.000</td>
<td>0.289</td>
</tr>
<tr>
<td>stimulus_type * number * context * speaker</td>
<td>9.62</td>
<td>5.498</td>
<td>0.000</td>
<td>0.244</td>
</tr>
<tr>
<td><strong>Weert</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>context</td>
<td>3.12</td>
<td>8.165</td>
<td>0.000</td>
<td>0.290</td>
</tr>
<tr>
<td>speaker</td>
<td>2.85</td>
<td>21.737</td>
<td>0.000</td>
<td>0.521</td>
</tr>
<tr>
<td>number * context</td>
<td>4.00</td>
<td>13.866</td>
<td>0.000</td>
<td>0.409</td>
</tr>
<tr>
<td>stimulus_type * number * context</td>
<td>3.51</td>
<td>10.934</td>
<td>0.000</td>
<td>0.353</td>
</tr>
<tr>
<td>number * speaker</td>
<td>3.00</td>
<td>23.010</td>
<td>0.000</td>
<td>0.535</td>
</tr>
<tr>
<td>stimulus_type * number * speaker</td>
<td>2.97</td>
<td>5.322</td>
<td>0.003</td>
<td>0.210</td>
</tr>
<tr>
<td>context * speaker</td>
<td>9.84</td>
<td>5.163</td>
<td>0.000</td>
<td>0.205</td>
</tr>
<tr>
<td>number * context * speaker</td>
<td>9.53</td>
<td>6.179</td>
<td>0.000</td>
<td>0.236</td>
</tr>
<tr>
<td>stimulus_type * number * context * speaker</td>
<td>10.60</td>
<td>2.858</td>
<td>0.002</td>
<td>0.125</td>
</tr>
</tbody>
</table>

2.4.2 Discussion

The results of this cross-dialectal perception experiment supports the following conclusions.

First, the fact that the recognition scores for the word accent contrast in both Limburgian dialects are around 90% in the focused contexts is in itself an important finding, in particular because this contrast has been informally characterized as very subtle and difficult to learn (Notten 1974).

Second, while the recognition scores for Weert were virtually invariant under different prosodic conditions, those for the Roermond dialect varied drastically depending on the position in the sentence and information status of the experimental
words. Specifically, recognition rates were high if the word appeared at the end of
the sentence and/or was the focus of the sentence. The positional factor was
phonologically interpreted as phrase–final versus phrase-internal, while the
information status most probably resolves as accented versus unaccented. The two
contexts in which the contrast was poorly recognized are the prenuclear and the
postnuclear nonfinal positions. Given the discourse context set up by the
contextualizing sentences, it is to be expected that a focus-marking intonational
pitch accent occurred just on the focused words, causing the experimental words in
the prenuclear and (nonfinal) postnuclear positions to be unaccented. The production
data, too, showed that no contrast was made in either of these contexts, and so the
generalization is that poor recognition was found in unaccented phrase-internal
contexts. Although Gussenhoven (2000a,b) only dealt with postnuclear cases, this
generalization corresponds with that for the neutralization of the tone contrast in
Roermond.

A third finding was that recognition rates in Weert, but not in Roermond, were
unaffected by the way in which the stimuli were presented. The grammatical number
of words excised from sentences and presented in isolation was as easily recognized
by the Weert-speaking listeners as that of words presented in their sentences. This
implies that the acoustic cues to the distinction are located inside the signal for the
words in question, rather than being spread across the sentence as a whole. This
finding is compatible with the idea that the grammatical number distinction in Weert
is expressed through a quantity contrast. On average, recognition in the Roermond
data deteriorated in the word condition, which indicates that the cues to grammatical
number in this tonal dialect are partly located in the sentence as a whole. It is at first
sight somewhat puzzling that the presentation of excised words led to higher scores
for singular forms and to lower scores for plural forms. The fact that the
deterioration of recognition rates across contexts works out differently for singulars
and plurals when the surrounding sentence fragments are removed from the stimuli
is reflected in the three-way interaction NUMBER*CONTEXT*STIMULUS_TYPE.
Apparently, subjects shifted significantly towards ‘singular’ judgements in the
nonfocused conditions, thus boosting the recognition scores for singulars and
depressing those for plurals.

In principle, there are two possible explanations for this shift. First, the bias towards
singulars may be due to the fact that singular forms are somehow more
representative of the words in question, which was for instance reflected in the fact
that on the answer sheets the words were listed in the singular form. As listeners
were less certain of the correct answer, they may have increasingly opted for the
morphologically unmarked term of the number opposition. The fact that the bias
towards singulars is already evident in the scores for the sentence stimuli supports
this explanation, as it suggests that any increase in difficulty leads to more singular
judgements. The alternative to this morphological explanation is a phonological one.
Inspection of the pitch contours in the nonfocused conditions in panels (d1) and (d2)
of Figure 2.1 reveals an essentially flat contour, while the same is true for the
Accent 1 contour in panel (c). The presentation of these excised words as isolated
utterances necessarily triggers an interpretation of a final focused utterance, and
since in that condition Accent 1 is signalled by a sharp $f_0$ fall, the relatively flat contours are more reminiscent of an Accent 2 realisation, leading listeners to Accent 2 judgements, i.e. singular forms. It is to be noted that there is no bias in the Weert data, which may however be due to the generally high recognition scores. In the absence of further data, it is difficult to choose between these explanations. A prediction of the morphological account is that if more clearly discernible number contrasts, like the segmentally encoded ones we used in our experiment, were presented in noise conditions, there would be a bias towards singular judgements. This bias would, moreover, increase as the noise more effectively masks the speech signal.

Finally, we consider the extent to which the listeners derived any information from the phonetic parameter which signals the number distinction in the other dialect. It was already clear that Weert listeners cannot have relied on $f_0$ information, because of the general similarity of the $f_0$ contours for the two word accents in all contexts, and also because excision of the words from their sentences did not lead to a deterioration in recognition. In the case of Roermond, a durational difference between the two word accents was observed in that Accent 2 was 13 ms longer than Accent 1. Duration is a well-attested enhancing feature of the tone contrast in many Limburgian areas (Gussenhoven & Aarts 1999; Gussenhoven & Peters 2004; Peters 2008) and it is conceivable that this is also true for the Roermond dialect. While the average difference is small, the variation is sufficiently large for there to be a considerable number of clearly ‘long’ and ‘short’ tokens. As it happens, recognition scores were not better for those tokens than for the ones in which the difference was less clear, or, for that matter, for those tokens in which the durations were the reverse. Thus, while Weert is unambiguously durational, Roermond is unambiguously tonal.

These results are in agreement with phonological analyses of these dialects. Gussenhoven (2000a) describes the Roermond word prosodic contrast as one between the absence of a lexical tone (Accent 1) and a lexical tone that interacts with the tones of the intonation contour (Accent 2). Heijmans & Gussenhoven (1999) describe the word prosodic contrast of Weert as a quantity contrast in the case of monophthongs (e.g. /kniːtn/ vs. /kniːn/ for plural and singular forms of the word for ‘rabbit’) and a contrast between vowel-glide combinations and diphthongs in the case of words like stein ([ʃtein] vs /ʃtein/) for singular and plural forms of the word for ‘stone’.) In this latter case, they claim there may be a vowel quality difference between the plural and singular forms, such that the tongue glide for the diphthong is less extensive than that for the vowel-glide combination, but that more noticeably the duration of the vowel-glide combination is shorter than that of the diphthong.
2.5 Summary and conclusion

Phonological contrasts are not equal. Phonetic salience varies across contrasts and across contexts, rendering some contrasts more robust than others. In a cross-dialectal investigation, we considered the vulnerability of a word prosodic contrast, generally referred to as stoottoon or ‘Accent 1’ vs. sleeptoon or ‘Accent 2’, as a function of the sentence prosodic context. This was motivated by recent findings suggesting that this word prosodic contrast may interact with sentence prosody, such that it would appear to be impaired or neutralized as a function of information status and position in the sentence. Specifically, lack of focus and phrase-internal position had been reported to be detrimental to the contrast in dialects like Venlo and Roermond (Gussenhoven & van der Vliet 1999; Gussenhoven 2000a). Equally, however, there is considerable variation in the phonological and phonetic accounts of dialects which share this word prosodic contrast. While the opposition is a unitary phenomenon in going back to a single source, phonological and phonetic manifestations vary substantially. The opposition in the dialect of Roermond is tonal, while that in Weert appears to be encoded largely as a durational difference (Verhoeven & Connell 1992; Verhoeven 2002; Heijmans 2003). Since the neutralization had been reported for dialects with lexical tone contrasts, the suggestion was that it was specifically its tonal nature which caused it to interact with the intonational tone structure. By contrast, if the opposition is encoded durationally, the dependence on the intonational structure might be less strong.

The perception experiment used naturally spoken utterances and exploited the circumstance that the word prosodic distinction is used for grammatical number contrasts in some nouns. It showed that native speakers of Roermond recognize the members of the opposition in focused contexts and phrase-final contexts, but fail to do so in phrase-internal, nonfocused contexts. This suggests a categorical neutralization of the contrast in these latter positions. Moreover, the recognition of grammatical number on the basis of the presentation of excisions of the words from their sentences reduced the recognizability significantly, in part due to a bias in the scores towards singular forms, i.e. forms with Accent 2. Significantly, these effects were virtually absent in the parallel Weert data, where grammatical number is encoded durationally. While there is a significant context effect, the size of this effect is very small, and readily explained as due to increased attention levels in the case of focused or phrase-final words. There is no effect of the removal of the sentence context.

An acoustic investigation of the stimuli, which was confined to declarative intonation patterns, revealed large $f_0$ differences and minor duration differences in the Roermond stimuli, and large duration differences with negligible $f_0$ differences in the Weert stimuli. Duration did not have any effect on the perception of the number opposition in Roermond and neither could a case be made for a role of $f_0$ differences in Weert. In Roermond, duration failed to compensate for ambiguities
that appeared in a nonfocused, nonfinal context. In Weert, the contrast was insensitive to the prosodic context, just as segmental contrasts are.
3 The perception of lexical tones in declarative and interrogative intonation contours in the dialect of Roermond

54,000 inhabitants in the Dutch province of Limburg, is one of the better studied tonal varieties of German and Dutch, among which are those of Cologne in Germany (Gussenhoven & Peters 2004, Peters 2006a) and Borgloon in Belgium (Peters 2007). Generally, there is a binary tone contrast in the word stress syllable, referred to as Accent 1 vs Accent 2. The lexical tones of these dialects, known as ‘Limburgian’ in the Netherlands and Belgium, occur in a number of different shapes owing to their interaction with intonational tones. The question we aimed to answer was to what extent the contrast between Accent 1 and Accent 2 is perceivable by native speakers in the various contexts the tone contrast occurs in.

In this introduction, we briefly consider the Limburgian tone contrast in a typological perspective by comparing its contextual variation with the contextual variation found in Chinese languages. Next, we give the phonology of the tone contrast and its intonational contexts.

A version of this chapter will be submitted to Linguistics as: Fournier, R.: Perception of lexical tone in the Dutch dialect of Roermond.
3.1.1 Context-dependence of lexical tone realisation

Variation in the pitch shapes of lexical tones has been reported to result from the presence of boundary tones (Hyman 1996, Ladd 1996:151), to adjustments due to immediately following or preceding lexical tones (Xu 1997, Chen & Xu 2006), or to variation in pitch range (cf. Ladd 1996, Gussenhoven 2004). Such sources of variation may affect the shape of lexical tones considerably, as in the case of rising Mandarin Tone 2, which instead of its canonical rising shape may be realized with high level pitch after a tone ending in high pitch, Tone 1 or Tone 2 (Kuo, Xu & Yip 2007). Occasionally, it may cause lexical tone contrasts to be virtually neutralized, or at least detectable only for some listeners, as in the case of Cantonese Low, Low Rising and Low Falling tones in final syllables of sentences with rising intonation (Ma, Ciocca & Whitehall 2006). However, these various allophonic realizations usually allow the general shapes of the tones to be more or less recognizable, comparable to the way a contextually nasalized vowel may to a large extent retain the tongue position which it has when appearing in an equivalent oral context. The basic shapes of the lexical pitch contours are identifiable as the pronunciation in isolation, and the various adjustments that affect those basic shapes to greater or lesser extents are attributable to phonetic effects of surrounding context. In this situation, the phonological representation of the tone remains intact, and their phonetic realisation is influenced by neighbouring lexical or intonational tones or by variations in pitch register.

Although both types are often referred to by the term ‘tone sandhi’, the above allophonic adjustments in the realisation of lexical tones are to be distinguished from what Chen (2000:20) refers to as ‘morphotonemic’ tone sandhi in Chinese languages, where pitch shapes may change in specific structural positions, like initially or finally in some tone domain. Typically, the lexical tone is no longer recognizable out of context, and is replaced with another from the set of lexical tones of the language in question (Vance 1976). For instance, Beijing Mandarin third-tone sandhi replaces low-pitched Tone 3 with rising Tone 2 before another Tone 3 in the same tone domain (Yip 2000:180). Morphotonemic tone sandhi is thus discrete, phonological and typically neutralizing with some other tone.

One of the striking features of the two lexical tones in Limburgian dialects is that their realisation varies drastically, as well as discretely, from one context to the next. The fact that there are different phonological forms for the same underlying tone category makes the variation they are subject to comparable to morphotonemic tone sandhi in Chinese. However, in the case of Limburgian, each tone category is manifested by a number of distinct phonological forms, which do not occur as realisations of the other category. The tonal forms occur in mutually distinct sets, one for Accent 1 and the other for Accent 2. The tone categories thus do not share phonological forms with the other tone category, as is typically the case in Chinese tone sandhi. The variation is governed by three variables. First, the intonation contour co-determines the melodic shapes of the lexical tones. Second, in the final syllable of the Intonational Phrase (IP), the tonal contrast is realized differently from
what is observed in phrase-internal syllables. Third, in syllables with an intonational pitch accent, henceforth focused syllables, the tonal contrast is realized differently from what is seen in unfocused syllables. Thus, the tonal representations of Limburgian are assembled from intonational and lexical sources. Indeed, instead of saying that the realisation of the lexical tones depends on the intonation, it would be equally correct to say that the realisation of the intonational contrasts depends on the choice of words, to the extent that these come with different lexical tone categories.

The binary tone contrast of Limburgian dialects occurs in the syllable with word stress, and it is thus comparable to Norwegian and Swedish. Unlike the Scandinavian languages, the Limburgian dialects can have the contrast on word-final syllables. The number of different instantiations of the tone contrast may become particularly large in the dialects spoken in the central part of the Dutch province of Limburg, which are known in the Dutch dialectological literature as East Limburgian. The dialect of Roermond has two intonation contours, referred to a declarative and an interrogative intonation, and together with the variation in position in the IP and in the presence of a pitch accent, this leads to $2 \times 2 \times 2$ or eight contexts in which the tone contrast is potentially instantiated in some specific way. Other dialects, like that of Venlo (Gussenhoven & van der Vliet 1999), have more intonation patterns, and in them the number of different contexts increases accordingly. While there are several descriptions of the tone contrast in these East Limburgian dialects, there has never been an attempt to establish the extent to which the tonal contrast is perceivable for native speakers in any complete set of contexts in some dialect.

### 3.1.2 Goal

Our goal in this investigation, then, was to establish whether, and if so to what extent, native speakers of a dialect can identify the lexical tones in spoken utterances in which the factors that determine the shapes of the lexical tones are varied systematically. The dialect under investigation is the East Limburgian dialect of Roermond, for which we already have information about the perceivability of the contrast in a selection of contexts. Specifically, the contrasts as occurring with declarative intonation have been investigated in a cross-dialectal experiment (Fournier, Verhoeven, Swerts & Gussenhoven 2006). Broadly, the result of this investigation for the Roermond dialect was that the lexical tones are identified virtually as reliably as vocalic contrasts are, but that this was true only in three contexts: in IP-final and IP-internal focused syllables, and in IP-final unfocused syllables. 'Focused' here refers to the condition in which the word concerned was

---

24 De Bot, Cox and Weltens (1990) carried out a perception experiment on the dialect of Maastricht, but only using words in isolation. Gooskens and Rietveld (1995), who also tested the Maastricht contrasts, did take sentence position (final/nonfinal) and intonation contour (statement/question) into consideration, but all their stimuli were accented. To our knowledge, no other studies have been conducted to assess the perception of related dialects.
pitch accented due to corrective focus. Henceforth, we refer to this condition as the ‘focused’ context. In IP-internal nonfocused syllables, which context was investigated both in pre-nuclear and in post-nuclear positions, the contrast was effectively neutralized. We were interested in establishing whether this pattern also applies to the interrogative forms.

Table 3.1 gives schematic representations of the contrasts in the four contexts identified above on the basis of the production data in Gussenhoven (2000) and Fournier et al. (2006). The contrast at issue is the second data cell in the second row, in which the contrast occurs in an IP-internal nuclear syllable.

Table 3.1: Schematic pitch contours of two intonation contours in accented and unaccented final and nonfinal positions in the intonational phrase for Accent 1 (solid lines) and Accent 2 (dotted lines). The shaded boxes represent the syllable in the position concerned. The parts of the contours following the boxes are the contours that occur when the final syllable in the IP has Accent 1 or no accent.

<table>
<thead>
<tr>
<th></th>
<th>[+focus, +final]</th>
<th>[+focus, −final]</th>
<th>[−focus, +final]</th>
<th>[−focus, −final]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Declarative</strong></td>
<td>![Declarative][1]</td>
<td>![Declarative][2]</td>
<td>![Declarative][3]</td>
<td>![Declarative][4]</td>
</tr>
<tr>
<td><strong>Interrogative</strong></td>
<td>![Interrogative][1]</td>
<td>![Interrogative][2]</td>
<td>![Interrogative][3]</td>
<td>![Interrogative][4]</td>
</tr>
</tbody>
</table>

### 3.1.3 The contextual variation in detail

In this section we discuss the phonetics and phonology of the contours given in Table 3.1, following Gussenhoven (2000)\(^\text{25}\). The lexical tone contrast is privative: Accent 1 has no lexical tone, either underlyingly or on the surface, and Accent 2 is a H-tone on the second sonorant mora of the syllable with main stress. This distribution implies that words whose main stressed syllables have one sonorant mora, like [kat] ‘cat’ or [opal] ‘thorn’, cannot have the tone contrast. Examples of minimal pairs are [haas\(\text{I}\)] ‘hare’ vs. [haas\(\text{II}\)] ‘glove’, [iteit\(\text{I}\)] ‘stones’ vs. [iteit\(\text{II}\)] ‘skewer’, where the superscript I or II at the end of the syllable represents the lexical tone. Roermond has some ten minimal pairs of monosyllabic words whose singular form has Accent 2 and whose plural forms have Accent 1, like the forms for ‘stone’.

---

\(^{25}\) See also chapter 1 for more detail on the Roermond grammar.
Assuming utterances with one sentence accent, the intonational tones are provided either by the interrogative intonation, which is L* H* L, or by the declarative intonation, H* L. For Accent 1, these are the only tones in the representation; for Accent 2, they combine with the H on the second mora. Unexceptionally, the intonational tones are ordered before the lexical tone in any syllable in which they come together. In the first and third data columns of Table 3.1, the final position of the H is shown as a weak final rise in the final portion of the dotted contours, the declarative contours for Accent 2. The focused forms are shown in (2), where the singular form for ‘stone’ in (2a) has the lexical H after the intonational pitch accent H* and the boundary tone L, while the plural in (2b) has Accent 1. The nonfocused forms are given in (3a,b). In (3a), the intonational pitch accent is on zag ‘says’, leaving the postfocal final syllable for the intonational boundary tone and the lexical tone, in that order.

\[
\begin{align*}
\text{(2) a.} & \quad \text{Hae zag ‘stein’} \\
& \quad | \quad | \quad | \\
& \quad L_i \quad H^*L_iH \\
& \quad ‘He says “stone”’ \\
\text{b.} & \quad \text{Hae zag ‘stein’} \\
& \quad | \quad | \\
& \quad L_i \quad H^*L_i \\
& \quad ‘He says “stones”’
\end{align*}
\]

\[
\begin{align*}
\text{(3) a.} & \quad \text{Hae ZAG ‘sjtein’} \\
& \quad | \quad | \quad | \\
& \quad L_i \quad H^* \quad L_iH \\
& \quad ‘He SAYS “stone”’ \\
\text{b.} & \quad \text{Hae ZAG ‘sjtein’} \\
& \quad | \quad | \\
& \quad L_i \quad H^* \quad L_i \\
& \quad ‘He SAYS “stones”’
\end{align*}
\]

In the focused nonfinal contours in the second data column, the H appears after H*, as shown in (4a). As a result, the pitch fall occurs after the accented syllable stjein, while in (6b) the pitch falls inside the accented syllable. The explanation for the sharp fall in (4b) is that the L boundary tone associates with the second mora of the stressed syllable.

\[
\begin{align*}
\text{(4) a.} & \quad \text{Ich höb ‘sjtein’ gezag} \\
& \quad | \quad | \quad | \quad | \\
& \quad L_i \quad H^*H \quad L_i \quad L_i \\
& \quad ‘I said “stone”’ \\
\text{b.} & \quad \text{Ich höb ‘sjtein’ gezag} \\
& \quad | \quad | | \quad | \\
& \quad L_i \quad H^* \quad L_i \quad L_i \\
& \quad ‘I said “stones”’
\end{align*}
\]

The fourth data column of Fig. 3.1 is assumed to represent a neutralization in unaccented IP-internal contexts, and is accounted for by deletion of the lexical H when it is unaccompanied by intonational tones.

For the interrogative contours in the first three data columns, a constraint banning nonfinal phonological LH-contours is relevant, given in (2). NoRise plays a role in the tonal phonology of many languages (Yip 2002: 80, 190, 280) and is believed to
reflect the harder an slower production of pitch rises than pitch falls (Ohala & Ewan 1973, Sundberg 1979).

(5) **NORISE:**

\[
\begin{array}{c}
* \sigma \\
\text{m} \quad \text{m} \\
\text{L} \quad \text{H}
\end{array}
\]

**NORISE** has two effects. First, when \(L^*\) and lexical \(H\) come together on the same syllable, as in *stein* ‘stone’ in (6), the lexical \(H\) is assimilated to \(L\). This is how the low level pitch arises in accented nonfinal syllables with Accent 2, as shown in Table 3.1 (second row, second data column, dotted contour). Second, when \(L^*\) is alone in a syllable, i.e. combines with Accent 1, the following \(H_i\) fails to associate with the free mora, causing the rise from \(L^*\) to \(H_i\) to extend beyond the accented syllable, within which it rises only to mid. Its pitch here is between mid and high, depending on the number of syllables till the IP-end. This is shown in (6) (cf. the solid line in the same cell of Table 3.1). For comparison, we give the phonological representations of the equivalent contours for the declarative in (4a) for Accent 2 and (4b) for Accent 1. In (4b), \(L_i\) does associate with the second mora, and a steep fall is produced, as indicated in Table 3.1.

(6) a. b.

Höbse “stjein” gezag?  Höbse “sjtein” gezag?

\[
\begin{array}{c}
L_i \\
L^*H \\
H_iL_i \\
\rightarrow L_i \\
L^*L \\
H_iL_i
\end{array}
\]

‘Did you say “stone”?’  ‘Did you say “stones”?’

The interrogative for Accent 1 is a rise-fall, the realisation of contour \(L^*H_iL_i\) on a single syllable, as shown in (7b). In the form of the interrogative Accent 2, the underlying representation \(L^*H_iL_iH\) is changed to \(L^*L_iL_iH\) (effectively \(L^*L_iH\)) on account of **NORISE** (5), realized as rising pitch, as shown in (7a).

(7) a. b.

\[
\begin{array}{c}
L_i \\
L^*L_iH \\
\rightarrow L_i \\
L^*H_iL_i
\end{array}
\]

‘Did he say “stone”?’  ‘Did he say “stones”?’

Lastly, the unfocused interrogative final positions (cf. the third column in Table 3.1) are distinguished by the final addition of the \(H\)-tone for Accent 2, realized as a late,
weak rise, as shown in (8a). The fall-rise on the final syllable contrasts with a plain fall due to H_{i}L_{i} in Accent 1 (8b).

\[
\begin{array}{ll}
(8) a. & b. \\
\text{Zag hae ‘sjtein’?} & \text{Zag hae ‘sjtein’?} \\
L_{i} & L_{i} \\
L^{*}H_{i}L_{i}H & L^{*}H_{i}L_{i} \\
\text{‘Did he SAY “stone”?} & \text{‘Did he SAY “stones”?}
\end{array}
\]

Intuitively, the contrasts represented in Table 3.1 and analysed in (2)-(4) and (6)-(8) would seem to represent different degrees of perceptual salience. For instance, the focused declarative final contours seem more distinct in the interrogative intonation than in the declarative intonation. There are no indications, however, that speakers have greater difficulty recognizing a tone in one context better than in another, if we disregard the IP-internal nonfocused position illustrated in the fourth data column of Table 3.1.

### 3.1.4 Hypothesis

On the basis of the results of Fournier et al. (2006), which showed that listeners could not reliably discriminate between the lexical tones in unaccented IP-internal positions, whether before or after the nuclear syllable, we hypothesized that the same pattern of discrimination holds for the interrogative contours. As in the present investigation, discriminability of contrast was established by investigating the identification of each lexical tone separately. The stimuli consisted of monosyllabic words with Accent 2 whose plurals are formed by changing the tone into Accent 1 (as shown in section 3.1.3), and listeners were asked whether a word in a given utterance was a singular or a plural form. Particularly in the contexts for which discrimination was poor, low recognition scores were found for Accent 1, but high recognition scores for Accent 2. It was unclear what the bias towards Accent 2 was to be attributed to, but we certainly did not assume that it was related to the intonational contour. We thus hypothesized that this bias would also be found in the interrogative condition.

As will be explained in section 3.3, the results were not entirely in agreement with this expectations. For that reason, in addition to the main perception experiment, we elicited production and perception data in a follow-up experiment, in order to help us understand the unexpected results of the main experiment.
3.2. Method

3.2.1 Stimuli

We drew a corpus of sentences in which the lexical tone (Accent 1 and Accent 2), the intonation contour (declarative and interrogative), and the prosodic context (5) were varied orthogonally. The five contexts are the four contexts given in the columns of Table 3.1, plus one which is due to splitting up the fourth context into a pre-nuclear and a post-nuclear condition. These contexts are given as the rows in Table 3.2. The ten conditions were instantiated with the help of four nouns: bein (‘leg’), derm (‘intestine’), knien (‘rabbit’) and sjtein (‘stone’), each of them occurring both in its singular form (Accent 2) and its segmentally identical plural form (Accent 1), giving 120 sentences in all. Since the contrast between Accent 1 and Accent 2 is used to encode grammatical number in a small number of nouns, we could instruct listeners to decide whether the stimulus they heard contained a singular or a plural form rather than having them identify the phonological categories ‘Accent 1’ and ‘Accent 2’. Even though there exist popular names for these tone categories, it appears to be more difficult for listeners to identify a phonological category than a morphological one. The singular and plural forms of the target words were embedded in ten carrier sentences, corresponding to the combination of two intonation contours with the five prosodic contexts, as shown in Table 3.2.

Table 3.2: Carrier sentences used in the perception experiment, with the target word knien. Sentences or sentence parts in brackets were used during the recordings in order to favour a correct focus assignment, but were not presented to the listeners.

<table>
<thead>
<tr>
<th>CONTEXT</th>
<th>DECLARATIVE INTONATION</th>
<th>INTERROGATIVE INTONATION</th>
</tr>
</thead>
</table>
| [+focus, +final] | **In ’t Remunjs zaes-se geweun “KNIEN”**. In Roermond Dutch, you just say “RABBIT(S)”.
Zaes-se geweun "KNIEN"? [Nae, ich zae "HAAS".]
Do you just say "RABBIT(S)?"
No, I say "HARE". |
| [+focus, –final] | **[Wat höbs-se geheurd?] Ich höb “KNIEN” geheurd.**
[What did you hear?] I heard “RABBIT(S)”.
Höbs-se "KNIEN" geheurd? [Nae, "KIENDJ;
Did you hear "RABBIT(S)?"
No, "CHILD". |
| [–focus, +final] | **[Eers ZAG hae “knien”,] toen SJREEFDE hae “knien”.**
[First he SAID “rabbit”] then he WROTE “rabbit(s)”.
SJREEFDE hae "knien"? [Nae, hae ZAG "knien".]
Did he WRITE "rabbit(s)?"
[No, he SAID "rabbit(s)." ] |

26 Accent 1 is commonly referred to as stoottoon (‘push tone’) and Accent 2 as sleeptoon (‘drag tone’).
3.2.2 Recordings

The recordings were made in two sessions. The first session involved the declarative sentences only, which were also used for the experiment reported in Fournier et al. (2006). For that experiment, we had used six nouns, the four mentioned above plus *pin* ‘peg’ and *erm* ‘arm’, and recorded six native speakers, three male and three female. The second session involved interrogative sentences only, and was conducted with one male and one female speaker from the original set of six. Since the recordings of all speakers had been judged adequate for the first experiment, the selection of the two speakers, NG and PI, was simply based on their availability at the time of the recording. In the second session, we excluded the word *pin*, with which some speakers were unfamiliar (pegs being hardly needed these days), as well as *erm*, because of its resemblance with *derm*. By reducing the materials in this way, we managed to keep the duration of the test down to about 20 minutes, which is similar to the duration of the test of Fournier et al. (2006).

In both recording sessions, sentences were presented to speakers in four blocks, which displayed different randomized orders. Accented words appeared in capital letters; before each sentence, the grammatical number of the target word was specified in brackets (e.g. *eine knien* ‘one rabbit’ or *twee knien* ‘two rabbits’). A native speaker of Roermond Dutch was present during the recordings and asked the speakers to repeat any sentences which she judged to be mispronounced in any way at the end of each block. These mistakes were usually incidental, with one exception. One of our speakers (PI) systematically pronounced target words in the prenuclear context as separate IPs, which made the difference between Accent 1 and Accent 2 rather clearly perceivable. Since this problem was only noticed at the end of the recording session, we recorded new versions of all prenuclear cases (intermixed with a reduced set of other sentences). Also, since speaker PI found it difficult to switch from one focus condition to another, he was instructed to replace the unfocused target word with the word *get* (‘something’), which is typically unaccented, and then to pronounce the sentence again, this time with the original target word. This strategy resulted in realisations that satisfied the native speaker who was monitoring the recordings. Nevertheless, because the ‘nonfocus’ condition is a vital part of our design and that of Fournier et al. (2006), we backed up our data with recordings from a representative group of 17 speakers, who were asked to
pronounce sentences with prenuclear target words embedded in contexts that naturally allowed target words to be deaccented. We will report on these additional data in section 3.2.4.

3.2.3 The perception experiment

Twenty-two native speakers of Roermond participated in the experiment, 9 male and 13 female, aged between 42 and 85 years, with a mean age of 61. All subjects had spent the greater part of their lives in Roermond, and spoke the dialect at home and in social encounters outside the home, except at work, where Standard Dutch was more often used. The listening experiment was carried out in a quiet room where subjects listened to the stimuli through headphones in groups of five to eight persons. On their answer sheet, the target word for each stimulus was printed in its singular form together with two boxes labelled *enkelvoud* ‘singular’ and *meervoud* ‘plural’. The randomized set of stimuli was presented in a single session. During the session, the subjects were instructed to listen to the target word in each sentence and to check either the singular or the plural box, depending on which form they heard.

The next section presents the results of this perception test.

3.3. Analysis

3.3.1 Statistical analysis of the perception scores

The collected answers were pooled over the four nouns and analysed with an ANOVA (repeated measures), using the within-subject factors INTONATION (2 levels); NUMBER (2 levels); CONTEXT (5 levels) and SPEAKER (2 levels). The index for effect size is partial $\eta^2$, which ranges from 0 to 1. Table 3.3 lists effects and interactions of effects which have been found to be significant at a 1% level.

<table>
<thead>
<tr>
<th>Factor</th>
<th>F</th>
<th>sig.</th>
<th>part. $\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>intonation</td>
<td>18.247</td>
<td>&lt;.001</td>
<td>0.465</td>
</tr>
<tr>
<td>number</td>
<td>25.478</td>
<td>&lt;.001</td>
<td>0.548</td>
</tr>
<tr>
<td>context</td>
<td>277.210</td>
<td>&lt;.001</td>
<td><strong>0.930</strong></td>
</tr>
<tr>
<td>intonation*context</td>
<td>21.594</td>
<td>&lt;.001</td>
<td>0.507</td>
</tr>
<tr>
<td>number*context</td>
<td>50.551</td>
<td>&lt;.001</td>
<td><strong>0.739</strong></td>
</tr>
</tbody>
</table>

Table 3.3: Significant main effects and interactions of effects on recognition scores. Bold-faced rows represent the three strongest (interactions of) effects.
As shown in the table above, the strongest factor affecting the recognition scores is CONTEXT, with an effect size of .93. Pairwise comparisons between contexts reveal that [+focus,+final] cases are recognized significantly better than all others (p<.001), with more than 98% of correct answers. Also, scores are lower in the [+focus, –final] contexts than in the [+focus, –final] and [–focus,+final] contexts. CONTEXT significantly interacts with all other effects, especially with NUMBER (η² = .739). This interaction reflects the fact that singular forms attract better scores than plural ones in the [–focus, –final] contexts, both pre-nuclear and post-nuclear. Figure 3.1 gives a summary of the scores obtained for declarative and interrogative sentences, in which the difference between singular and plural forms is apparent in both intonations. This difference in fact appears to be larger in interrogative than in declarative sentences, which is shown in the interaction INTONATION*NUMBER*CONTEXT.

Figure 3.1. Recognition rates for singular and plural forms of nouns in declarative (left panel) and interrogative sentences (right panel), as a function of prosodic context (a= [+focus,+final]; b = [+focus,–final]; c = [–focus,+final]; d1 = [–focus, –final], prenuclear; d2=[–focus,–final], postnuclear). Pooled over the two speakers.

With a comparable effect size (η² = .732), the interaction INTONATION*SPEAKER highlights another important reason for score variation. Compared with NG’s sentences which attract similar scores in questions and statements, PI’s recordings lead to better scores in statements than in questions. Interestingly, however, intonation is not globally responsible for the results for speaker PI. Scores are rather boosted in one particular context in declarative sentences and depressed in another
context in interrogative sentences: while NG’s [–focus,–final], prenuclear sentences with declarative intonation attract 56% of correct answers, the score for PI’s sentences is no less than 77%, but in [+focus,–final] interrogative sentences, only 66% of the answers for PI’s utterances are correct, against a high 91% for NG’s. These speaker-related score differences, which we will discuss in more detail in Sections 3.3.2.3 and 3.3.2.4, are illustrated in Figure 3.2.

Figure 3.2: Recognition rates per speaker, intonation and context. The two most striking differences in recognition scores are found in PI’s realization of [+focus, –final] target words with interrogative intonation (see Section 3.3.2.3), and in the same speaker’s prenuclear target words with declarative intonation (see Section 3.3.2.4).

We will not comment on the interactions CONTEXT*SPEAKER and CONTEXT*NUMBER*SPEAKER since they do not highlight any new facts. The interaction INTONATION*CONTEXT is partly related to a cause mentioned before, namely, the score differences associated with PI’s recordings. More importantly, it shows that, for both speakers, in a [–focus, +final] context, number identification is better in questions than in statements (p = .001). These score differences are observed for both PI’s and NG’s realizations, and for Accent 1 as well as Accent 2 forms. In search of possible explanations of this finding as well as of the speaker-related differences reported above, we will first have a closer look at the stimuli used in the experiment.

---

27 The fact that different contexts are involved in score differences per intonation explains that the interaction INTONATION*SPEAKER*CONTEXT is weaker than INTONATION*SPEAKER.  
28 This score is based on a repeated-measures ANOVA using, as dependent variable, the scores obtained for [-focus, +final] words only. The independent variables were INTONATION, NUMBER and SPEAKER. The ANOVA revealed no other effect or interaction of effects than INTONATION.
3.3.2 Acoustic analysis of the stimuli

3.3.2.1 General description

All sentences were segmented manually, after which a Praat script was run to extract maximum, minimum, begin and end f0 values in the nucleus and coda of the target words. Figure 3.3 shows the average f0 contours for statements and questions in final and/or focused contexts. The contours for declarative sentences are very similar to the ones described in our previous experiment (Fournier et al. 2006), which in turn corresponded to Gussenhoven's descriptions in Gussenhoven (2000).

In the [+focus, +final] context, f0 falls quite steeply throughout the rhyme in the case of Accent 1, but only until the last third of the rhyme in Accent 2, after which a rise of 10 Hz is observed. Contours in the [-focus, +final] cases follow the same scheme, only with smaller f0 excursions, which makes the tonal contrast less salient. As we saw in Section 3.2, this reduced f0 excursion may have led to more mistakes in the perception test, but the recognition scores are still high enough to show that the contrast is well rooted in the listeners’ tonal grammar. In the [+focus, –final] cases, pitch is mostly falling in Accent 1, but slightly rises in Accent 2. No difference was found between Accent 1 and Accent 2 in the postnuclear contours, which exhibit an overall fall of about 20 Hz. In the prenuclear ones, Accent 1 differs from Accent 2 in the final part of the contour, in which a slight rise occurs. This rise will be discussed in Section 3.2.3.

When used with question intonation, Accent 1 in a [+focus, +final] context is pronounced with a 120 Hz rise in the first half of the rhyme followed by a fall of 160 Hz in the second half. Accent 2 starts with a slow rise that becomes steeper in the second half, with an overall f0 excursion of 120 Hz. When the target word appears in a [-focus, +final] context, Accent 1 is again characterized by a high fall (without the initial rise). In Accent 2, the fall is shorter (80 Hz) and followed by a 30 Hz rise. Compared with the [-focus, +final] cases pronounced with declarative intonation, this contrast seems more salient, which might explain the better recognition scores reported in section 3.1. Turning to the [+focus, –final] cases, we observe that Accent 1 has a steep rise throughout the rhyme, whereas pitch is maintained at a low level until the end of the rhyme in Accent 2. The slight rise in the average contour shown in Figure 3.3 reflects a difference amongst (and within) speakers that will be discussed in section 3.3.2.2.

As for the pre- and postnuclear cases, our data match Gussenhoven’s description, in that the contours do not depend on the accent used: a steady slow fall is observed in all instances of [-focus, –final] words. Due to a methodological issue that arose during the recording of the questions (see section 3.2.2), the analysis of the prenuclear cases was based on new measurements, which are described in section 3.3.2.4.
In addition to the analysis of $f_0$, vowel durations were compared across accents, contexts and intonation contours. An univariate ANOVA only revealed a significant effect of the context ($F = 19.95$, $p < .001$), with significant differences between all contexts except between [+focus, +final] (178 ms) and [–focus, +final] (171 ms), as well as between [+focus, –final] (149 ms) and the postnuclear [–focus, –final] (145 ms). We can infer from these results that the position of the target word in the IP has more effect on duration than its focus status, and that, contrary to observations made...
on other Limburgian dialects (see for instance Peters 2008), duration does not necessarily go hand in hand with the choice of the accent.\footnote{This does not appear to hold for words spoken in isolation. A set of 36 words recorded by one of our informants (NG) exhibits systematic differences between accents as well as between intonations, Accent 2 being longer than Accent 1 and questions longer than statements. See chapter 5.}

### 3.3.2.2 Additional tests

In order to check the validity of one of our speakers' realizations of prenuclear words with interrogative intonation, we asked 17 native speakers of the Roermond dialect (about 20 years old on average) to record the following sentences (which were intermixed with other sentences):

- *Waas hae vorige maondj in MILAAN? Nae, hae was vorige maondj in BERLIEN.* ('Was he last month in Milan? No, he was last month in Berlin."
- *Waas hae vorige waek in MILAAN? Nae, hae was vorige waek in BERLIEN.* ('Was he last week in Milan? No, he was last week in Berlin."

The recordings were made with a Sharp MiniDisc, at different places where background noise might occur. If a recorded sentence was masked by noise, the speakers were requested to repeat it; otherwise, they were allowed to proceed to the next sentence even when the pitch contour used did not seem to correspond to the expected one. Apart from one case in which a speaker split a sentence into several IPs, all recordings were considered valid for the acoustic analysis of the target words *maondj* and *waek*. This analysis confirmed the neutralization of the tonal contrast in prenuclear position. A 20 Hz pitch fall was observed in both accents, only reflecting the overall fall that precedes the focused word *MILAAN*. Deviations from this average fall were tested in an ANOVA, with the overall \( f_0 \) excursion on the target word as the dependent variable and the accent as a factor. This factor was not significant (\( p=.374 \)). The original data used for the perception experiment can hence be viewed as typical instances of prenuclear words pronounced with question intonation. The answer part of these sentences will be dealt with in section 3.3.2.4.

In addition to the *maondj/waek* sentences, our 17 native speakers recorded sentences reproducing the \([+focus, –final]\) context with question intonation, in order to facilitate the interpretation of the puzzling result reported in section 3.3.1. For the same reason, the subjects were also asked to perceptually judge prenuclear words with declarative intonation. We will turn to these new data in the next two sections, in search of a link between unexpected identification rates and possibly deviant acoustic realisations.
3.3.2.3 Focused, nonfinal target words with interrogative intonation

In the perception experiment, it appeared that recognition scores for Accent 2 in a [+focus,–final] context with interrogative intonation are speaker-dependent. PI’s pronunciations have a lower likelihood of being correctly identified than NG’s, with 40% of correct answers against 84%. The analysis of the data shows that, instead of the plateau described in Gussenhoven's study, a 42 Hz rise occurred on PI's Accent 2 contours. Although the $f_0$ excursion in the rise is clearly larger in Accent 1 (109 Hz), this weak rise must have led listeners to perceive Accent 1 where Accent 2 was intended. The data for NG are less homogeneous, just as the listeners' scores are: Whereas there were virtually no recognition mistakes for her Accent 2 versions of *knien*, *sjein* and *bein*, which are falling or only very slightly rising (14 Hz), the 27 Hz rise observed on the word *derm* led to the poor recognition rate of 41%. Figure 3.4 illustrates the differences between Accent 1 and 2, and between the two speakers.

![Figure 3.4: Realization of Accent 1 (thick line) and Accent 2 (thin line) for both speakers.](image)

Whereas the contrast between Accent 1 and Accent 2 is clearly visible in NG's realisations, PI tends to reduce the difference in pitch excursions. A closer look at all recordings, including those which were not selected for the perception experiment, revealed that in a few cases, NG also produced instances of Accent 2 in which $f_0$ was in fact rising throughout the rhyme, although the overall pitch excursion was never larger than 30 Hz. This observation led us to suspect that PI's rise on Accent 2 was not just an individual variant but may reflect a more general tendency of reducing the tonal contrast in this context.

The behaviour of PI, and to a lesser extent NG, in nonfinal interrogative focused contexts suggests that the dialect may be on its way to a neutralization of the lexical
tone distinction in this particular context. In order to see to what extent their behaviour is more general, recordings from 17 additional speakers were collected and analysed. This analysis confirmed our conjecture. The following sentences were recorded, in the same conditions as explained in the previous section:

Zit diene BEIN aan diene vooot? ('Is your LEG attached to your foot?')
Zitte dien BEIN aan dien veut? ('Are your LEGS attached to your feet?')
(1 knien) Höb-se KNIEN gezag? ('Did you say RABBIT?')
(2 knien) Höb-se KNIEN gezag? ('Did you say RABBITS?')

Of the 34 sentence pairs recorded, only 11 could be treated as acceptable instances of the tonal contrast to be analysed. A sentence was included in the statistics if no segmental errors were found in or around the word of interest, and if the pitch contour used on that word roughly corresponded to the instances recorded for the perception experiment. More specifically, Accent 1 had to be pronounced with a rising contour, and Accent 2, with a flat or rising contour. Contours in which the f0 fell for more than 40 Hz were excluded. In addition, since the target words were to be analyzed in pairs, both members of the contrast needed to be accepted in order for a pair to be included in the analysis. We found 7 pairs in which a difference smaller than 20 Hz in the f0 excursions was measured in the contours, and 11 pairs in which this difference was larger. Of these eleven pairs, only three were instances of the minimal pair represented by BEIN. We will thus compute the average f0 excursions based on the eight instances of the pair represented by KNIEN, which minimal pair also figured in the perception experiment.

The contours for the plural form of KNIEN, i.e. the form with Accent 1, are very similar to NG and PI's instances. The average f0 excursion between minimum and maximum is 113 Hz. In the singular form (Accent 2), only two speakers produced an f0 excursion of less than 10 Hz; the average difference between minimum and maximum is 44 Hz, about the same as what we found for PI's contours.

Since we only have production data from young speakers, we cannot make strong statements about the change of the contrast over time, but it is quite likely that it has been moving towards neutralization. Of the small proportion of speakers who produced a clear contrast, only two actually used the flat Accent 2 contour that was best recognized by the older listeners. A possible reason for the disappearance of this contrast is its lack of acoustic salience. As Hume & Johnson (2001) put it: "If a contrast is perceptually weak in a certain position, synchronic phonology works to enhance or sacrifice it by way of epenthesis, metathesis, dissimilation, assimilation or deletion". In the present case, within-word acoustic differences - measured in terms of f0 excursions across the relevant time span - cannot be the only factor at

30 The first two sentences were borrowed from Gussenhoven (2000a), as are the sentences discussed in the next section.
31 This fall suggests that some speakers realize polar questions with a declarative intonation contour. This is not uncommon in the Venlo dialect. Note that this contour is also, and systematically this time, used for encoding WH-questions.
play, since the contrast in IP-final position in declarative utterances led to significantly better recognition scores, even though the acoustic difference is smaller. This relationship between salience and contrast will be discussed further in section 3.4.

3.2.4 Prenuclear words with declarative intonation

Another unexpected result is the good performance of listeners in a prenuclear context with declarative intonation, when the stimuli are pronounced by PI: 70% of the plural forms and 84% of the singular forms were correctly identified, against 45% (pl.) and 65% (sg.) of NG’s utterances. In our previous experiment (Fournier et al. 2006), which also used NG’s and PI’s utterances as stimuli, the scores for NG’s stimuli were broadly comparable, but only 56% of the plural forms and 60% of the singular forms in PI’s stimuli were recognized correctly. Since the stimuli were the same, this difference must be attributed to the judges, who were older and might have detected a hint to Accent 1 more easily than the younger listeners used in Fournier at al. (2006). Possibly, there might have been a contrast in unaccented prenuclear position a generation ago, as is found today in the related dialects of Cologne (Gussenhoven & Peters 2004) and Sittard (Hanssen 2005). It is also possible that the metalinguistic use of the target word made it more difficult to treat as an unimportant, and hence unaccented, word. The speaker, then, may have realized this word with more prominence, adding a pitch accent where he was not supposed to. In any event, the resulting Accent 1 shape does not have the same contour as a regular focused case. Rather than a falling Accent 1 and a rising Accent 2 contour, we observe a slight rise throughout the rhyme in Accent 1 words, which contrasts with a flat or slightly falling contour in Accent 2 words. These contours are shown in Figure 3.5.

Figure 3.5: $f_0$ computed for PI’s instances of Accent 1 (left panel) and Accent 2 (right panel) in a prenuclear context, with declarative intonation. Each line type corresponds to a lexical item.
According to these plots, either a putative old contrast or an additional prenuclear pitch accent has been realized, in three out of four cases, with a weak rise for Accent 1. In order to verify whether the rise was indeed a cue for the recognition of Accent 1 in a prenuclear context, we performed a follow-up listening experiment. In this short test, 17 speakers were asked to judge various prenuclear instances of Accent 1 and Accent 2 as recorded by PI and NG, which either had rising pitch throughout the rhyme or a late rise or low pitch peak in the second half of the rhyme as well as more or less level pitch. The results showed a tendency to identify weakly falling and flat contours with Accent 2 (64%) and rising ones with Accent 1 (78%). Less clear-cut cases led to scores close to the chance level (55%).

Additional recordings performed with the same subjects failed to show a rise for Accent 1 in a different prenuclear context. The target word maondf in the sentence Nae, hae was vorige maondf in BERLIEN (as opposed to Nae, hae was vorige waek in BERLIEN, see section 3.2.2) was realized with a flat contour, just like its Accent 2 counterpart in waek. Figure 3.6 shows all contours measured for Accent 1 in maondf.

Figure 3.6: f0 contours for the Accent 1 word maondf in prenuclear position.

In sum, there is apparently a rule of thumb underlying the listeners’ association of rising contours with Accent 1 in prenuclear contexts, but this rule was not used during the production of nonfocused prenuclear target words, except in PI’s case. It might be that listeners still retain knowledge of an older contrast, perhaps still present in older speakers, even though they do not use it themselves. On the other hand, it is improbable that PI is conservative in one area (see previous section) and progressive in another. We are thus more inclined to assume that our speaker added a pitch accent in prenuclear position, even though our sentence was not supposed to elicit a second pitch accent in the utterance. In the absence of other examples in our data or in earlier production studies (which focused on the stretch from the last focused syllable to the end of the IP, see Gussenhoven 2000a:135), we can only hypothesize that a prenuclear pitch accent is characterized by a rise. What remains
unclear, however, is why older judges in the main experiment reported here should have recognized the Accent 1 instances more accurately than younger judges of Fournier at al. (2006).

3.4 Discussion and conclusion

In this study, we examined the perception of all pitch contours representing the tonal contrast between Accent 1 and Accent 2 in the dialect of Roermond. Results of a forced-choice identification task showed that the different shapes of Accent 1 and Accent 2 forms are recognized well above chance by native speakers, except in three cases.

The first case of poor recognition concerns target words which appear in [–focus, –final] contexts. The average recognition rate for these words (55%) is substantially lower than for words in a focused and/or final context (more than 90% on average). This confirms the results of previous production (Gussenhoven 2000a, Gussenhoven 2000b) and perception (Fournier et al. 2006) studies, which all claim a neutralization of the tonal contrast in [–focus, –final] contexts. However, this confirmation did not materialize as a chance-level recognition of Accent 1 and Accent 2 individually. Instead, we observed fairly high scores for Accent 2 and low ones for Accent 1. The same was found in Fournier et al. (2006). For such a bias towards Accent 2, two possible explanations had been envisaged, which we can also apply to the present study. The first is related to the fact that in the test, Accent 2 words systematically represented singular forms. The singular form may be seen as a default, the option to which listeners tend to escape when they have doubts on what they heard. The second interpretation takes us back to the speech signal, which displays a rather flat contour for words in [–focus, –final] position. This flat contour may be more readily interpreted as an instance of Accent 2, which often exhibits, across contexts, smaller pitch excursions than Accent 1 (see Figure 3.3). Both explanations seem equally plausible (cf. Fournier et al). We can now confirm that the bias towards the singular (or Accent 2) is irrespective of the intonational contour, since it applies to interrogative as well as in declarative sentences. Unfortunately, because (relatively) level pitch is also a feature of nonfinal Accent 2 in interrogative contours, this finding does not bring us closer to an answer to the question whether the bias has a morphological or a phonological explanation.

There was one particular set of utterances in which the bias towards Accent 2 was considerably reduced. One speaker’s realizations of prenuclear words with declarative intonation triggered high scores for Accent 2 as well as for Accent 1. The reason for this deviant result could be traced in the acoustic shape of the Accent 1 contour, a slight rise that contrasted with the flat or falling contour observed for Accent 2. Further tests with more native speakers suggested that the rise was used as a cue for the identification of prenuclear Accent 1 words. However, the slight rise did not appear in the new production data for prenuclear, nonfocused Accent 1. The
most likely explanation for these data is that this one speaker produced an additional pitch accent in the prenuclear position, characterized by a slight rise throughout the rhyme. If this rise indeed corresponds to the regular case of an Accent 1 prenuclear pitch accent, it is not surprising that listeners would identify the target word as an instance of Accent 1. They did not use it themselves during the production of our experimental sentences since these sentences were not supposed to elicit them.

A second exception to the high recognition scores reported for the Roermond contrasts seems to offer us an opportunity to observe an earlier stage of a neutralization process. This case concerns focused words that appear in nonfinal interrogative contexts. According to Gussenhoven’s production studies and the pronunciation of one of our speakers, the contrast should oppose a rise (for Accent 1) to a low plateau (for Accent 2). However, our second speaker consistently replaced the low plateau by a slight rise, which, due to its resemblance with Accent 1, caused a drop in recognition scores for Accent 2. A follow-up test confirmed that this speaker’s tendency to reduce the salience of the contrast was in fact quite common among younger speakers. These results suggest that the tonal contrast in this position is on its way out. Phonological contrasts are more vulnerable if they are not easily perceivable. In our case, we may assume that at some point, speakers started to reduce the acoustic salience of the contrast by turning the plateau, which is immediately followed by a rise in the post-accentual syllable, into an anticipatory rise, so as to compromise its identity and to make its recognition more difficult. This, in turn, is likely to encourage the listener-turned-speaker to realize the contrast even less clearly, thus precipitating its disappearance.

The example above underlines the importance of acoustic salience in the study of a single phonological contrast that surfaces in various shapes. The third example of lower scores found in our data, which affected [-focus, +final] declarative cases, shows that this issue is more complex than it may seem at first sight. Based on the example reported in the previous paragraph, we may be tempted to quantify the acoustic salience of pitch contours based on Euclidean distances (i.e. the difference between the $f_0$ of Accent 1 and the $f_0$ of Accent 2 at each time point), and to use this measure to account for lower recognition scores. By this means, we could readily explain why the scores for [-focus, +final] declarative cases were lower than those for [+focus] contexts, which indeed exhibit larger surfaces between the $f_0$ curves. Yet if we compare the [-focus, +final] declarative cases with the [+focus, -final] interrogative cases mentioned above, it becomes evident that excursion sizes are only part of the story. Although $f_0$ excursions in the [-focus, +final] declarative cases are clearly smaller than those between a slight rise and a more pronounced rise (see Figure 3.3 and the left panel of Figure 3.4), recognition is better in the first than in the second case. To account for this, we could assume that acoustic salience also depends on the direction of the contour, and consider the sheer presence or absence of a rise as the main cue at play during tone discrimination in this context. An attempt to deal with the issue of salience in a more systematic way is reported in the next chapter. Using all contours of a neighbouring dialect, that of Venlo, we will try to assess the importance of salience, as measured automatically and auditorily, for the perceivability of tonal contrasts.
4 Phonetic salience and the perception of the lexical tone contrast of Venlo Dutch

4.1 Introduction

Studies on the tonal systems of East Limburgian dialects, which have a lexical tone contrast broadly comparable to the Scandinavian distinction between Accent 1 and Accent 2, have shown that the tone contrast varies as a function of the prosodic context. Specifically, the shape of the pitch contours on the relevant syllables depends not only on the tone, but also on the accentuation (as in the expression of information focus), the position of the syllable in the Intonational Phrase, and the intonation contour, for instance declarative as opposed to interrogative intonation (for Venlo, Gussenhoven & van der Vliet 1999; for Roermond: Gussenhoven 2000a,b; for Sittard: Hanssen 2005; for Helden: van den Beuken 2007). The dialects are spoken in the extreme north-west corner of a larger tonal area, which consists of the northern half of Rhineland Palatinate, the southern half of North Rhine-Westphalia, and the larger parts of the Belgian and Dutch provinces of Limburg. There are indications that the tonal area is shrinking. Gilles (1999) found no contrast in Luxembourgish, which is traditionally included as the south-western corner of the area on the basis of the older literature, while Heijmans (2003) showed that the dialect of Weert, spoken on the north-western edge of the area in the Netherlands, has largely reinterpreted the contrast as a quantity contrast.

Our assumption in the investigation reported here is that the perceptual robustness of a contrast determines the path along which the contrast is on its way out. A collapse of the system may begin in specific contexts, while remaining in others. One indication that this applies to the east Limburgian tone contrast between Accent 1 and Accent 2 is the loss of the opposition in nonfocal, nonfinal contexts as opposed to its firm retention in focused declarative contours (Fournier, Verhoeven, Swerts & Gussenhoven 2006, or chapter 2 of this thesis). Interestingly, in a follow-up experiment involving the recognition of the lexical tones in interrogative contours, the contrast between a rising contour (Accent 1) and a low flat contour (Accent 2), was compromised in the speech of one speaker, whose

---

32 A version of this chapter will be submitted to *Phonetica* as: Fournier, R. & C. Gussenhoven: Phonetic salience and the perception of the lexical tone contrast of Venlo Dutch.
Accent 2 was mistaken for Accent 1 in 61% of the cases on nonfinal syllables (against 17% in the case of the other speaker). This poor recognition of Accent 2 could be explained by a slight pitch rise where other speakers produced low level pitch. Since the form for Accent 1 in the same context is rising pitch, the failure to continue the fully low pitch apparently impaired the recognizability of Accent 2, and hence the perceivability of the contrast. In this case, the behaviour of this speaker may herald the disappearance of the contrast in nonfinal interrogative contexts.

The aim of our investigation was to establish how well the contrasts of one of these peripheral dialects are perceived by native speakers, and to answer the question whether the extent to which the lexical tones are recognized is related to the phonetic salience of the contrast in the location concerned. The relation between phonetic salience and contrast maintenance is a recurrent theme in recent phonological theories. Beckman (1998) proposes that some structural positions favour the presence of phonological contrasts while others are prone to neutralize them (‘positional faithfulness’). Among these positions, which are taken to have a privileged role to play in word processing, are root-initial syllables, stressed syllables, and syllable onsets (as opposed to unstressed syllables, root-internal syllables and syllable codas). Domain-final syllables may equally have a ‘strong’ position. Although Beckman’s treatment stresses the grammatical role of positional faithfulness constraints in Optimality Theory rather than the phonetic properties of contrasts, the functional connections to perception and word processing are evident. Another way in which the connection between contrast and phonetic salience has been incorporated into Optimality Theory is through Steriade’s (2008) P-map, which allows faithfulness violations to be penalized in proportion to the phonetic difference of the contrast involved. Contrasts are thus predicted to be lost in places where they are least perceivable.

It is tempting to assume that phonetic salience can explain the structure of phonological systems. Speech perception in adverse conditions, like a noisy environment, will affect acoustically less salient features more dramatically than more salient features. However, it is also a common experience that native listeners appear to deal with phonetically smaller contrasts as easily as with phonetically larger contrasts. In the words of Labov, “there is no such thing as a small difference in sound” (1991:38). A belief that subtle differences are hard to hear may stem from the experience of L2 listeners, whose L1 phonological categories may include phonetic forms that are to be assigned to different categories in the L2. For instance, unless the difference is pointed out to them, non-native listeners with a Dutch or German language background are likely not to hear the distinction between the early and late falls of East Limburgian declaratives, which in the standard languages signal a declarative falling intonation in both cases. This view would imply that in a stable situation there is no correlation between salience and recognition: even though the difference is small by objective acoustic measures, as long as listeners know what to pay attention to, the acoustically smaller contrasts are functionally equivalent to acoustically larger ones. This suggests that a relation between salience and the contrast perception is more likely to be found in less
stable situations, as in the dialects on the periphery of the tonal area, in which there may be an ongoing process of tone loss. The dialect of Venlo is such a peripheral dialect, which moreover has a large number of contextually defined tone contrasts, due to the presence of four distinct intonation contours, which together with the variation in position in the sentence and the optionality of the a focus-marking accent, define twelve contexts in which the contrast is realized.

We define the perceivability of a contrast, its robustness, as the average recognition scores of the two forms involved in the contrast. A definition of phonetic salience is less straightforward. At first sight, an acoustic measure should be derivable from a comparison of two speech signals, in our case based on the fundamental frequency. However, there are reasons to believe that such a procedure fails to reflect the way human listeners perceive phonetic forms. For instance, listeners may well assign more importance to some parts of the signal, say, the end of a contour, or to higher pitch more than to lower pitch. In addition, there is the issue of whether perceived phonetic salience is the same for native and non-native listeners, and if it is not, which measure should figure in our investigation. It has generally been found that discriminability varies as a function of the phonological status of the contrast in the language of the listener. For instance, Mielke (2003) showed that the same contrast, viz. that between /h/ and its absence, was better perceived by listeners who had it in their language. Peperkamp & Depoux (2002) demonstrate that short term memory tasks show that French listeners do not effectively distinguish stress contrasts, not even after extensive L2 exposure to a language with contrastive stress (Depoux, Sebastián-Gallés, Navarrete & Peperkamp ms.). In the realm of tone, Gandour (1983), Burnham et al. (1996), Lee (1996) and Huang (2001) show that native listeners perceived larger differences (Gandour 1983: on a scale from 0 to 10; Huang 2001: reaction times recorded during same/different judgements) or could better discriminate (Lee 1996, Burnham et al. 1996: same/different judgements) between tones than non-native listeners. This may in part be related to the language-dependent choice of cues used for tonal identification or discrimination. For instance, Gandour (1983) observed that English speakers seemed to prioritize tone height, while Thai speakers rather focused on the direction of the tone contours (rising vs falling).

We therefore investigated phonetic salience from two different angles. First, we explored objective measures of acoustic distance, taking our cue from Hermes (1998). Second, we measured the perceived phonetic difference both with native and non-native listeners. By comparing the objective and subjective distance measures, we would be able to estimate the success with which acoustic measures reflect perceived phonetic salience, and by obtaining difference scores on the same contrasts from both non-native speakers and native speakers of the Venlo dialect, we could gain some insight into the extent to which the phonological status of a phonetic difference influences perceived salience. After answering these questions, we proceeded to the question whether the robustness of contrasts, as established on the basis of recognition scores, is related to phonetic salience. Thus, the research questions we address in this article concern two possible factors in the perception of the lexical tones in the dialect of Venlo:
1. To what extent does the recognition of the lexical tones vary with context, as defined by accentuation, position in the Intonation Phrase, and the intonation contour?

2. Does the variation in recognition rate across the conditions given in Question 1 correlate with the variation in phonetic salience?

In order to answer the first question, a male and a female speaker of the dialect recorded a corpus of sentences in which a number of tonal minimal pairs were embedded in a number of positions in sentences spoken with different intonation contours with and without a focus marking accent. In section 4.2, we report on the acoustic properties of the experimental words, focussing on duration and $f_0$ differences. The sentences were used in a perception experiment in which native listeners were asked to indicate whether Accent 1 or Accent 2 occurred in the experimental words. This Recognition Experiment is reported in section 4.3. In order to answer the second question, we obtained two acoustic difference measures from the data reported in section 4.2, one based on duration and the other on $f_0$ properties in each of the relevant prosodic and intonational contexts. In addition, we collected difference judgements from both native and non-native speakers for these same stimulus pairs. These measures are discussed and reported in section 4.5. With the help of these measures, we determined whether there is a correlation between phonetic salience, defined acoustically or perceptually, and recognition rates, and in addition assessed the extent to which the perception of the phonetic difference between two utterances is accounted for by the acoustic difference measure. This is done in section 4.5.

Our hypotheses for the two research questions are thus (1) that recognition rates vary across prosodic contexts, and that (2) the phonetic salience of a contrast correlates with its perceivability.

4.2 Stimuli

All tests described below are based on data recorded from two native speakers of the Venlo dialect. For the recognition task, whole sentences were used, whereas objective and subjective distance scores between members of minimal pairs were established based on words excised from these sentences. In this section, we will first give an overview of the sentences recorded, and then proceed to the acoustic analysis of the target words within the sentences.
4.2.1 The corpus

The carrier sentences used in the perception experiments contained four tonal minimal pairs (\textit{knienn} = 'rabbit(s)', \textit{beinn} = 'leg(s)', \textit{derm} = 'intestine(s)' and \textit{stein} = 'stone(s)'). All four minimal pairs involve a nominal number contrast, with Accent 1 denoting the plural and Accent 2 denoting the singular form. The sentences elicited these words in a large number of contexts in which they were used metalinguistically. The reason for this was that sentences in which these words are embedded in conventional grammatical structures often inevitably reveal their grammatical number in other words, such as the articles or verbal concord. The use of metalinguistic sentences allowed us to place the words in otherwise identical sentences. Importantly, words that are used metalinguistically are incorporated in the intonational structures of the sentence as a whole, and do not necessarily introduce additional prosodic boundaries (cf. Fournier, Verhoeven, Swerts & Gussenhoven 2006).

The carrier sentences are listed in Table 4.1. They represent the combination of three different prosodic contexts ([+focus,+final], [+focus,–final] and [–focus,+final]) with four intonation contours (declarative, low interrogative, continuative, high interrogative). We also included a sentence for eliciting the Accent 1 contour in [+focus,+final] cases with low question intonation, although Gussenhoven & van der Vliet 1999 reported that in this condition, there was no specific contour and that the speakers filled this gap in the grammar by using the corresponding high question contour instead (see section 1.1.4.1). We will see in the next section that there was indeed a difference, which justified the inclusion of both contours in the perception experiment. In addition to all +focus and/or +final cases, a number of instances of target words in [–focus,–final] contexts were recorded with declarative and interrogative intonation, in order to verify the neutralization reported in Gussenhoven & van der Vliet (1999). These utterances were not included in any perception tests, but their acoustic properties will be described in the same way as for those we did include, and their acoustic difference scores were calculated as for the contrasts we did include in the perception tests.

Prior to the recording of the \textit{bein}, \textit{derm}, \textit{knienn} and \textit{stein} sentences, a further set of sentences was read aloud by the speakers, in which the word \textit{beer} ('beer') contrasted with \textit{baer} ('bear') in a number of prosodic contexts. This was done because it was felt that by using each target word in a natural context, rather than in a metalinguistic context as in the sentences in Table 4.1, subjects would find it easier to pronounce the sentences. In these practice sentences, which are given in Appendix 1, the word \textit{beer} was mostly used in combination with the verb \textit{drinke} ('to drink') and \textit{baer} with the verb \textit{jage} ('to hunt'). By reading these sentences first, speakers had an opportunity to acquaint themselves with the use of target words in different prosodic contexts.
Both the experimental sentences in Table 4.1 and the practice sentences in the Appendix typically appear as part of a mini dialogue, which was read aloud in full by the same speaker. The sentences were presented to the speakers in the Veldeke spelling, a standard orthographic system developed for Limburgian dialects in general (see Bakkes, Crompvoets, Notten & Walraven 2001). Turn-taking in the dialogues was indicated by means of a hyphen, while the focal accent in the sentences was indicated by bold-faced capital letters (cf. Table 4.1 and Appendix). The main session was divided into two parts. In the first part, we recorded three different orders of all sentences with 'declarative' and 'low interrogative' intonation, intermixed with half of the sentences that were intended to have ‘continuative’ intonation. The other half of the ‘continuative’ sentences were recorded in a second block, together with sentences intended to have ‘high interrogative’ intonation, again in three different orders. In total, each speaker recorded $28 + 320 + 135 = 483$ sentences in about two hours. The utterances we elicited with the help of all the sentences shown in Table 4.1 were analysed acoustically, but those in the [−focus,–final] condition were not used in the perception tests, as their sole purpose was to allow us to evaluate the claim in Gussenhoven & van der Vliet (1999) that the lexical tone contrast is neutralized in this context. Inclusion in the Recognition Experiment and the Subjective Distance Experiment would have made those tests too long, while the inclusion of non-contrasts among the contrasts would have blurred the data we need to answer the question if the salience of a contrasts is a measure of its perceivability.

Table 4.1. Sentences representing tonal minimal pairs, with knien as the experimental item (the other words were bein, derm and stein). The [−focus, –final] cases were recorded and analyzed but not used in the Recognition Experiment. Clauses or sentences in brackets, which helped eliciting the right discourse meaning, were read aloud but cut off prior to the Recognition Experiment (i.e. they were not used as stimuli for the context specified in the first column of the table). Although the carrier sentences for the [−focus, +final] questions were different for the two speakers, they triggered the same discourse meaning.

<table>
<thead>
<tr>
<th>Context</th>
<th>Sentence (using knien ‘rabbit(s)’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[+focus, +final]</td>
<td><strong>In ’t Venloos zaeste gewoeën &quot;KNIEN&quot;</strong>. In the Venlo dialect, you just say “RABBIT(S)”.</td>
</tr>
<tr>
<td>declarative</td>
<td></td>
</tr>
<tr>
<td>[+focus, +final]</td>
<td>Zaeste gewoeën &quot;KNIEN&quot;?</td>
</tr>
</tbody>
</table>
| low question           | − [− Nae, ik zegk “BEER”.] Do you just say “RABBIT(S)”? − No, I say “BEER”.
|                        |                                                                                                                       |
| [+focus, +final]       | Ièrs zei ik "KNIEN", [toen zei ik ”DERM”, en toen nog ”BEIN”]. First I said “RABBIT(S)”, then I said “INTESTINE(S)”, and then “LEG(S)”. |
| continuation           |                                                                                                                       |
| [+focus, +final]       | [− In ’t Venloos zaeste gewoeën “KNIEN”]. − Ech waor?] Zaeste gewoeën "KNIEN"?” […] – Really? Do you just say “rabbit(s)”?
| high question          |                                                                                                                       |
| [+focus, −final]       | [− Waat hebste gehuùrd?]                                                                                              |
| declarative            | Ik heb "KNIEN” gehuùrd. − What did you hear? − I heard “RABBIT(S)”.


### 4.2.2 Recordings

The sentences were presented to the speakers in the form given in Table 4.1 and Appendix 1, but in the case of the experimental sentences the grammatical number of each target word was symbolized by means of either one or two small icons.
representing the meaning of the word in question. For instance, the mini-dialog eliciting *kniend* (‘rabbit’) in a [+focus, +final] context with high question intonation, was presented as follows:

– In ’t Venloos zaeste gewoeën "KNIEN".
– Ech waor? Zaeste gewoeën "KNIEN"?

The speakers, one female and one male, were 50 and 62 years old, respectively. Both were or had been language teachers and were at ease with the reading task, so that the utterances were pronounced in a fluent way and without fluffs or segmental mistakes. Our aim was to obtain a homogeneous corpus which was representative of what we knew about Venlo Dutch. This meant that we coached our speakers in the sense that if they did not produce the intonation contour that a given sentence was intended to elicit, we usually led them to produce that contour, even though it was not their first choice. Two additional recording sessions were required in the case of speaker KB in order for us to obtain two complete sets of minimal pairs in all relevant contexts. One difficulty here was the choice between the ‘low interrogative’ and the ‘high interrogative’ intonations. Although the presentation of the sentences in distinct recording blocks facilitated the distinction between these two contours, both speakers tended to substitute one for the other, with, in KB’s case, an apparent preference for the ‘low interrogative’. In the [–focus,+final] context, YK initially pronounced *ZAESE ‘bein’?* with the ‘high interrogative’ intonation, where we intended to elicit the ‘low interrogative’, while the reverse happened in *De SCHREEFS ‘bein’?*, which was intended to elicit the high interrogative. Since the sets were complementary, we decided to keep them as they were. However, for speaker KB, whom we recorded after speaker YK, we decided to change the context sentences for the sentences which we intended to elicit the ‘high interrogative’ and the ‘low interrogative’. This had the expected effect for *De SCHREEFS ‘bein’?*, which elicited the ‘low interrogative’ as intended. After additional attempts, a set of utterances of *ZAESE ‘bein’?* with ‘high interrogative’ intonation was also obtained.

The second difficulty in the elicitation of the intonation contours concerned the ‘continuative’ intonation. In the first session, both a ‘listing’ context was used and a context in which the experimental item occurred in a non–final clause. This second context is illustrated for the [+focus,+final] context by *Ik zei wel “KNIEN”, maar ik meinde “KIEND”*. ‘I may have said “RABBIT”, but I meant “CHILD”’. However, whereas speaker YK consistently realized the lists with the expected ‘continuation’ intonation, she used a variety of intonation contours in the non–final clauses. We therefore decided to discard the utterances produced in the non–final clause context, and to elicit the ‘continuation’ intonation from speaker KB with the help of the listing context only. As it happens, speaker KB produced a similar kind of variation in his lists as YK had produced in the pre–final clauses. We decided to retain the ‘continuation’ intonations that YK did produce, and had a third session with him in which we completed the elicitation of this contour.
A third reason for the additional recording session was that after a preliminary acoustic analysis of the utterances by speaker KB, it turned out that he had produced Accent 1 pronunciations of all experimental words in the ‘declarative’ and ‘low question’ intonations in the [–focus, +final] context. This could either point to a systematic neutralization of the contrast in [–focus, +final] context in KB’s utterances, or simply to a mistake. We redid the recordings during the second session. Although we did not mention the problem at hand, the contrast reappeared in almost all new versions, probably due to the fact that many sentences in the new corpus contained target words in [–focus, +final] position, making it easier for the speaker to produce the correct discourse meaning without neglecting the contrast.

4.2.3 Acoustic analysis of the [+focus] and [+final] cases

We selected one version of each sentence per speaker. The target word within each sentence was segmented manually into onset and rhyme. The rhyme was then used for the acoustic analysis of the stimuli, reported in the following section.

4.2.3.1 Fundamental frequency

For the analysis of the $f_0$ contours, which was carried out with Praat (Boersma & Weenink 2007), each rhyme was inspected for gaps in the $f_0$ measurements. We found that 81 out of the 256 contours had such gaps, typically due to creakiness, and decided to interpolate $f_0$ values between the beginning and end of each gap. The duration of these interpolations was 34 ms on average (s.d. 21). Subsequently, we extracted an average $f_0$ measurement from each of 100 time windows equally divided over each rhyme. The scale we used was ERB (Equivalent Rectangular Bandwidth), which is closer to human perception than the Hz scale (Hermes & van Gestel 1991).

Figure 4.1 displays the contours for Accent 1 and 2 in all [+focus] and [+final] contexts for the four intonation contours ‘declarative’, ‘low interrogative’, ‘continuation’ and ‘high interrogative’. The phonological analysis of these contours, slightly adapted from Gussenhoven & van der Vliet (1999), was given in chapter 1. The contours in Figure 4.1 reflect the sections in the speech signal in which $f_0$ values were actually computed, not the durations as determined by manual segmentation, which are given in Figure 4.2. The differences between the apparent durations of Figure 4.1 and the actual durations in Figure 4.2 are nowhere more than 10 ms, except in YK's realization of *derm* in the [–focus, +final] context spoken with the ‘high interrogative’ intonation, where no $f_0$ measurements were obtained in the last 25 ms of the rhyme. As a result, the averaged contour for Accent 2 in this context looks shorter than it actually is.
Figure 4.1: $f_0$ contours of Venlo Accent 1 and Accent 2 (ERB) in four intonation contours and three prosodic contexts, as a function of time (ms). Solid lines represent Accent 1 and dashed lines Accent 2. Gray lines give speaker YK's utterances and black lines give speaker KB's utterances.

We give some general observations about these contours, noting any discrepancies between these data and those reported in Gussenhoven & van der Vliet (1999). Focused ‘declarative’ Accent 1 has a sharply falling pitch contour in final as well as nonfinal positions, while a weak low falling contour occurs in the final
unfocused context. In final position, Accent 2 resembles Accent 1, but has a rising part after the fall starting approximately at the time point where the contour ends in Accent 1. In nonfinal position, the contour for Accent 2 is slightly rising in KB’s utterances and slightly falling in YK’s utterances, the peak being aligned later in the KB’s contour. The durational difference between Accent 1 and Accent 2 is 20 ms in this context, against more than 90 ms in the final ones. We also observe that all other things being equal (context, intonation and tone), YK’s rhymes are consistently longer than KB’s. However, the difference in duration between Accent 1 and Accent 2 rhymes is not always larger in YK’s realizations.

As we said in section 4.2.1, we recorded all six instances of the ‘low interrogative’ intonation, instead of the five contours reported in Gussenhoven & van der Vliet (1999), who claimed that there was no specific contour for Accent 1 in the [+focus, +final] context with ‘low interrogative’ intonation. The contour that contrasts with Accent 2 in this context is borrowed from the set of high question contours. In other words, the Accent 1 contour in [+focus, +final] position is phonologically identical in low and in high questions. Indeed, in our data, the contours in the low and the high question instances have quite similar shapes. However, we can also observe a systematic register difference between them: while the low questions start around 4 ERB, the starting pitch in high questions is 5 ERB for KB, and almost 8 ERB for YK. Although this register difference may not be phonologically relevant, we decided to treat the pronunciation used for the ‘low interrogative’ as a contour in its own right, thus restoring symmetry to the Venlo tonal system.

In this same context, Accent 2 has a more complex shape than Accent 1 and, in fact, than any other contour in the Venlo system: a fall-rise. An initial short rise is followed by a fairly steep fall and a rise. The contour ends in a slight fall (YK) or a plateau (KB), but inspection and auditory evaluation of the wave form seem to indicate that these movements fall outside the effective speech signal (cf. Gussenhoven 2004:9), and in the tonal analysis only the steep fall and the subsequent rise are phonologically specified. In the nonfinal context, the ‘low interrogratives’ resemble the corresponding ‘declarative’ contours. In the [-focus, +final] position, the contour falls slightly, rises again, and ends with a brief plateau. The difference between Accent 1 and Accent 2 would appear to be only a durational one with Accent 2 being longer than Accent 1.

In utterances pronounced with ‘continuation’ intonation, Accent 1 and Accent 2 exhibit the same kind of difference in all cases, namely, a rise (Accent 1) vs. a near-plateau (Accent 2). We did not observe the brief initial steep rise for Accent 2 found in Gussenhoven & van der Vliet (1999). In [+focus,+final] cases, KB’s realizations of Accent 1 differ from YK’s in that the rise is steeper and followed by a fall, whereas YK’s contour is rather a steady rise throughout the rhyme. In the other two contexts, YK’s realizations exhibit a steeper rise than KB’s, but the difference in \( f_0 \) excursions between the two speakers is smaller than in the [+focus,+final] context.
In the ‘high interrogative’ intonation, contours usually start with high pitch (around 6–7 ERB), except for KB’s realisation of Accent 1 in [+focus,+final] position, which clearly contrasts with the (rise-)plateau-rise of Accent 2. In [+focus,–final] position, both tones rise to reach a target at ca. 8 ERB, with Accent 2 starting at a higher pitch in speaker KB (6.4 instead of 5.7 ERB). The relatively high beginnings of the [+focus] Accent 2 contours for Speaker YK are unexpected on the basis of the description in Gussenhoven & van der Vliet (1999) and the contours given there in Figure 4.9. In [–focus,+final] position, the rise in Accent 1 becomes somewhat steeper than the one for Accent 2 after 170 ms, and plateaus at a point where Accent 2 is still rising, but the $f_0$ excursions for both accents are small (max. 1 ERB) and there is no point in the contours where the difference between Accent 1 and Accent 2 amounts to more than 0.4 ERB.

Summarizing, in the ‘declarative’, ‘continuative’ and ‘high interrogative’ intonations, contours for Accent 1 look broadly similar across the three contexts, with smaller $f_0$ excursions in the nonfocused than in the focused cases. In ‘declarative’ contours, all target words have falling pitch, in high questions, a rise followed by a plateau, and in continuative clauses, a rise followed by a slight fall. For Accent 2, there are more substantial differences between final and nonfinal contexts, due to the interaction of the lexical H with the boundary tone (see chapter 1). Pitch contour differences appear to be subtler with the ‘continuation’ and ‘high interrogative’ intonations. In such cases, durational differences are likely to play an important role in the perception of the lexical tone distinction, Accent 2 being longer than Accent 1 in most cases. By and large, the forms we elicited agree with those given in Gussenhoven & van der Vliet (1999), except for the focused Accent 1 contours in the ‘high interrogative’ intonation for speaker YK, which begin with mid or high pitch instead of the low pitch reported in Gussenhoven & van der Vliet (1999).

The next section gives an overview of rhyme duration in the [+focus] and [+final] cases. The [–focus,–final] cases, which will not appear in the perception test, will be analysed in terms of their $f_0$ and duration in section 4.4.1.3.

### 4.2.3.2 Duration

A first analysis of duration per tone over all positions revealed that Accent 2 is significantly longer than Accent 1 (47 ms in average, p<.001). In fact, Accent 2 is longer than Accent 1 in all intonations and contexts, except in six individual cases.

---

33 The analysis of absolute durations was done with a univariate ANOVA with the factors TONE, INTONATION, CONTEXT and SPEAKER. The ANOVA found significant effects for all factors, plus the interactions CONTEXT*TONE (p=.003) and CONTEXT*INTONATION (p=.001). Since there are already quite a few statistical analyses in this chapter, we will restrict ourselves to the discussion of the main effects only.
Figure 4.2: Durational differences (seconds) between Accent 2 and Accent 1 rhymes per intonation contour (‘declarative’, ‘low interrogative’, ‘continuation’ and ‘high interrogative’) and context (a= [+focus, +final], b= [+focus, –final], c= [–focus, +final]), as pronounced by two speakers.

(out of 96)34. Figure 4.2 shows the durational differences between Accent 2 and Accent 1 per context, intonation and speaker.

We also compared Accent 1 and Accent 2 durations between speakers, intonations and contexts, and found significant differences in all three categories. Figure 4.3 gives mean durations of Accent 1 and Accent 2 rhymes per intonation contour and prosodic position for each speaker separately. As for speakers, it is clear that YK’s utterances are systematically longer than KB’s (63 ms in average, p<.001). We found the following rankings for the durations in the different intonations and contexts:

- **Intonation:** Rhymes pronounced with continuation or low question intonation are significantly longer than rhymes with declarative or high question intonation (p<.001). The mean duration in each context is: low question: 307 ms, continuation: 294 ms, high question: 264 ms, declarative: 260 ms.
- **Context:** [+focus, +final] rhymes are longer than [–focus, +final] cases (‘accentual lengthening’; 313 ms and 292 ms, respectively), which are longer than [+focus, –final] cases (‘pre–final lengthening’; 239 ms; p<.001 in all comparisons). Accentual lengthening and pre–final lengthening were also found for Cologne (Peters 2006a), while pre–final lengthening was also found for dialect of Roermond (Fournier, Verhoeven, Swerts & Gussenhoven 2006).

---

34 In these cases, the differences in duration range from 3 to 19 ms. Half of the cases concern the nonfinal context.
Figure 4.3. Mean rhyme durations of Accent 1 (panel a) and Accent 2 (panel b) in s., pooled over four lexical items in 'declarative', 'low interrogative', 'continuative' and 'high interrogative' intonations in focused final, focused non–final, and nonfocused final contexts for speakers YK and KB.

The durational differences between Accent 1 and Accent 2 are likely to play a role in speech processing, despite the striking difference between the two speakers, which causes Accent 1 rhymes pronounced by YK to be about as long as Accent 2 rhymes in KB’s target words. In real life situations, normalization will typically prevent speaker differences from being interpreted as linguistic differences. In the context of our experiment, it will be interesting to know if the durational difference between Accent 1 and Accent 2 varies across intonation contours and contexts. Figure 4.2 suggests it does, since differences appear smaller in the [–final] context than in the [+final] ones. Accordingly, we submitted the difference between Accent 2 and Accent 1 to a univariate ANOVA with the fixed factors CONTEXT and INTONATION, and the random factor SPEAKER. The only effect that this analysis found significant is CONTEXT (p=.007). A post–hoc test confirmed that this effect is due to significantly smaller differences in the nonfinal context in comparison to the final ones (p<.001 in [+focus,–final] vs. [+focus, +/–final] while p=.461 in [+focus,+final] vs. [–focus, +final])35. Thus, while individual durations significantly depend on all three factors mentioned above, when it comes to a comparison between the members of tonal minimal pairs only CONTEXT appears to have a significant effect36. The durational differences are discussed further (in comparison with other distance measures) in section 4.4.1.2.

35 This is in line with Gussenhoven’s ‘extra lengthening’ rule for the final Accent 2 words, explained in section 1.1.4.1.
36 We also analyzed our data with the additional factor WORD. Although the choice of the lexical item does have an overall effect on rhyme duration (with the following significant relationships: derm > bein > stein > knien), it does not affect durational differences: the difference between Accent 1 and Accent 2 will not be significantly larger within a dermI-dermII pair than, say, within a steinI-steinII pair. In the remainder of this study, we will not evaluate possible effects of the lexical item on the results.
4.3 Recognition: Tonal contrast in different prosodic contexts

In this section we describe the design of the Recognition Experiment and report its results.

4.3.1 Procedure

The 192 sentences (3 contexts * 4 intonation contours * 4 word pairs * 2 accents * 2 speakers) were randomized automatically (all categories mixed). The stimuli were arranged in blocks of 10 stimuli, and a short orientation signal was inserted at the end of each block. In order to neutralize possible order effects, we created a second set of sentences based on the first randomized set, in which all singular forms were replaced by their corresponding plural forms, and vice-versa. Half of the subjects (group A) were then presented the sentences in the first order, and the other half (group B) listened to the sentences in the second order.

We recruited nineteen native speakers of the dialect from the students and the teaching and administrative staff at a secondary school in Venlo. According to a form that all subjects filled in prior to the test, the dialect of Venlo was their primary language at home and in most social encounters (besides Standard Dutch, which was used at work or at school). No subject reported hearing problems. The average age was 29, but subjects were either between 15 and 17 years old (12 students), or between 40 and 61 (7 members of the teaching or administrative staff). The two age classes were distributed more or less evenly over groups A and B, so that age was not confounded with presentation order. The same was true for gender (11 female, 8 male).

The Recognition Experiment was carried out in a quiet room where subjects listened to the stimuli through headphones, in two groups of about ten subjects each. Each subject received an answer sheet, on which the target word for each stimulus was printed in its singular form together with two boxes labelled 'enkelvoud' ‘singular’ and ‘meervoud’ ‘plural’. Even though they heard sentences, only the experimental words were printed. Their task was to listen to the sentences and focus on the word printed on their sheet, and check the ‘singular’ or ‘plural’ box according to what they heard. All sentences were presented in a single session, which took about 25 minutes. The subjects were remunerated for their participation.
4.3.2 Results

Figure 4.4 summarizes the results obtained for each intonation, context and speaker. The bars represent the mean recognition rates for Accent 1 and Accent 2 as percentages of the number of trials.

![Recognition percentages averaged over Accent 1 and Accent 2 per intonation contour (‘declarative’, ‘low interrogative’, ‘continuation’ and ‘high interrogative’), context (a = [+focus,+final], b = [+focus,–final], c = [–focus, +final]), and speaker (KB, YK).](image)

The recognition scores displayed in Figure 4.4 represent the perceivability of the tone contrast in the different contexts, and can as such be compared with the phonetic difference measures to be computed on the basis of the acoustic data presented in section 4.2 as well as with the subjective difference measures to be collected in the perception tests we ran as part of the Phonetic Salience Experiment. This will be done in section 4.5. Overall, the mean recognition rate was 70.1% (s.d. 13.7%).

While a recognition score expresses the salience of a contrast rather than of a single member of this contrast, in order to evaluate the functioning of the accentual contrast in the dialect we are interested in knowing whether Accent 1 and Accent 2 are equally recognizable. In the repeated measures ANOVA we ran on the results, we therefore included a factor TONE (Accent 1, Accent 2) to the set of within-subjects factors defined for the previous analyses (INTONATION, CONTEXT, SPEAKER). The analysis yielded main effects for INTONATION and CONTEXT (both p < .001), as well as the interactions INTONATION*TONE, INTONATION*CONTEXT, and INTONATION*TONE*CONTEXT. In addition, there were the interactions INTONATION*CONTEXT*SPEAKER and TONE*CONTEXT*SPEAKER (all p< .001).

The absence of a main effect for TONE means that, in general, Accent 1 is not easier or more difficult to recognize than Accent 2. However, as showed by a significant interaction between tone and intonation, there are cases in which Accent 2 is better recognized than Accent 1. These cases are mostly found in sentences...
with ‘continuative’ intonation, in which scores for Accent 2 are better than those for Accent 1 by 22%, as opposed to 2% in ‘declaratives’, 5% in ‘low interrogatives’, and 10% in ‘high interrogatives’, the latter three all non–significant. This is true for all three contexts, but the nonfinal context shows the greatest bias towards Accent 2, meaning that here Accent 1 is often misinterpreted as Accent 2.

The main effect for \textit{intonation} is due to the higher recognition scores in ‘declarative’ and ‘low interrogative’ intonations, which are significantly different from those in the ‘continuative’ and ‘high interrogative’ intonations. The main effect for \textit{context} is due to the significantly higher scores in the focused final context than in the focused nonfinal and nonfocused final contexts. The interaction between \textit{intonation} and \textit{context} must be due to the fact that in the focused final context recognition scores are better in the ‘declarative’ and ‘low interrogative’ intonations than in the other contexts, and that the scores in the ‘high interrogative’ intonation are particularly low in nonfinal context.

We observed a systematic difference between generations. Subjects in the older age group (7 subjects between 40 and 61 years old) performed significantly better than those in the younger (12 subjects between 15 and 17 years old), with a difference of 17.3% in the average scores over all contexts and intonation contours (81.03% s.d. 3.96 for the older subjects; 63.8% s.d. 9.3 for the younger subjects). The effect of age was verified by means of an ANOVA, using the same within-subjects factors as above and the between-subjects factor \textit{age group} (p <.001). The difference was found to be fairly constant across contexts, intonations and speakers. Nevertheless, we found a \textit{tone/context/age group} interaction (p=.008), which reflects the fact that in the case of the younger listeners the better scores that are found in the [+focus, +final] context concern Accent 2 only, while for the older listeners both Accent 1 and Accent 2 are recognized well in this context.

### 4.3.3 Recognition Experiment: Conclusion

A recognition rate of the lexical tone contrast in the Venlo dialect of 70% is fairly good, although less than that found in the same three contexts in the related dialect of Roermond, where a recognition score of 91% was obtained. Even if we restrict ourselves to the two contours that are also found in the Roermond dialect, the ‘declarative’ and the ‘interrogative’ (‘low question’), the difference between the dialects is still present, with Venlo reaching 76%.

A second indication that the contrast is vulnerable in the Venlo dialect is provided by the difference between the age groups (17.3%). Interestingly, no difference between generations could be established in the experiments on the Roermond dialect, where an older group of 22 subjects obtained the recognition scores of 91% (s.d. 3.7%) mentioned above, and a younger group of 18 subjects still obtained
87% (s.d. 11.1%). Although the exact percentages in the Roermond group should be compared with some caution, due to the different sets of data presented to the subjects (the younger group had to judge declarative sentences only, intermixed with segmental minimal pairs, while the older group was presented declarative as well as interrogative sentences, all displaying tonal minimal pairs), they certainly show a different tendency than in the Venlo group.

### 4.4 Distance measures between Accent 1 and Accent 2

This section reports on the investigation of the relation between the perceivability of the tone contrast and its phonetic salience. Section 4.4.1.1 investigates the acoustic distance measures we calculated between Accent 1 and Accent 2, Section 4.4.1.2 gives the values of two of these measures (RMSE and cosine correlation) for our data set, and before we go on to the subjective distance measures (in Section 4.4), Section 4.4.1.3 explains why we can safely limit our data set to the [+focus] and [+final] cases.

#### 4.4.1 Objective distances

##### 4.4.1.1 Method: comparison between measures

A number of methods have been used to calculate acoustic differences between pitch contours. Largely, these have been based on the root mean square error (RMSE) and correlation coefficients (Pearson’s $r$). The two measures are to a large extent complementary. Whereas RMSE directly expresses the difference between pitch values at each time point, and thereby take account of pitch range differences, correlation coefficients rather express the differences in the general trajectories of pitch contours, abstracting away from pitch range differences.

In an experiment which involved both subjective and objective similarity measures between synthesized pitch contours, Hermes (1998) found that Pearson’s correlation coefficients showed the strongest correspondence with human perceptual ratings. While this would appear to make the correlation method a sensible candidate for assessing the distance of our Venlo contour contrasts, it is improbable that this measure is the best approximation of an objective human ear in all experimental situations. In particular, as Hermes points out, the pitch range normalization implied by this method may be undesirable. Wherever pitch range is likely to be an important criterion in the assessment of pitch contour differences, automatic measures such as the mean distance or the root-mean-square distance should be preferred to correlation coefficients. In the case of our Venlo data, it
would appear that the Pearson’s coefficients are less appropriate. In order to see why this is the case, consider the hypothetical contours in Figure 4.5. Contour 1 is a fall, resembling declarative Accent 1 contours in [+focus,+final] position; contour 3 is its mirror image, and contour 2 is a combination of the first half of contour 1 with the second half of contour 3, resulting in a fall-rise. Finally, contour 4 is a slow fall, while the shape of contour 5 is identical to that of contour 4, but is realized in a lower register.

![Figure 4.5](image)

*Figure 4.5. Five hypothetical $f_0$ contours, whose distance measures are to be calculated using (a) Root mean square error (RMSE) scores, (b) Pearson’s correlation coefficients, and (c) cosine distance scores (converted from the original correlation coefficients). See Table 4.2.*

Table 4.2 shows the Pearson’s coefficients computed for several pairs of contours. First, observe that Pearson’s coefficients are sensitive to direction. A comparison between contour 1 and its mirror image yields a correlation of −1, meaning that the contours follow opposite directions (falling vs. rising), and if two contours partly follow the same direction (contour 1 vs. contour 2), the correlation between them will obviously be closer to zero. Importantly, Pearson’s coefficients are insensitive to slope, as is clear from the barely lower distance measure between contours 1 and 4, both of which fall, but differ dramatically in slope. They are also insensitive to register, as is clear from the distance measure between contours 4 and 5, which differ in register, but whose slopes are identical, and are therefore characterized as being identical by Pearson’s $r$ of 1. By contrast, RMSE scores (also shown in Table 4.2, along with a third measure that will be discussed below) take differences in range and register into consideration. For instance, whereas the Pearson’s coefficient of contour 4 vs. contour 5 is 1, the RMSE score is a low 4.5, reflecting the register shift. Also, the low RMSE of 1.87 between contours 1 and 4 is due to the difference in slope, and which compares with the near-identity as expressed by the Pearson’s $r$ of 0.95.
Table 4.2. Difference measures between the hypothetical contours in Figure 4.6. Pearson’s scores represent similarities which range from –1 (mirror image) to 1 (identical), with a 0 point meaning that the contours have nothing in common, whereas RMSE and cosine scores represent dissimilarities which range from 0 (identical) to an unknown maximum.

<table>
<thead>
<tr>
<th>f0 contours</th>
<th>rPearson</th>
<th>RMSE</th>
<th>dCos</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 vs. 2</td>
<td>0.24</td>
<td>2.61</td>
<td>5.33</td>
</tr>
<tr>
<td>1 vs. 3</td>
<td>-1</td>
<td>4.28</td>
<td>22.51</td>
</tr>
<tr>
<td>1 vs. 4</td>
<td>0.95</td>
<td>1.87</td>
<td>4.25</td>
</tr>
<tr>
<td>1 vs. 5</td>
<td>0.95</td>
<td>4.98</td>
<td>1.44</td>
</tr>
<tr>
<td>3 vs. 4</td>
<td>-0.95</td>
<td>2.41</td>
<td>7.59</td>
</tr>
<tr>
<td>4 vs. 5</td>
<td>1</td>
<td>4.50</td>
<td>0.98</td>
</tr>
<tr>
<td>3 vs. 5</td>
<td>-0.95</td>
<td>5.02</td>
<td>13.60</td>
</tr>
</tbody>
</table>

As is clear from these examples, Pearson’s coefficients reflect a specific type of information, the direction of the contours. In other words, they interpret the contours in terms of the basic movements, reminiscent of the approximation of natural f0 contours as sequences of straight lines. This would appear to be the explanation of Hermes’ results, whose stimuli were straight-line stylizations of natural f0 contours. When the criteria used for resynthesis prioritize the same elements as Pearson’s coefficients do, the RMSE scores may look like so many outliers. To illustrate this point, consider the two contour pairs in Figure 4.6, which represents Accent 1 and Accent 2 contours in [–focus,+final] declarative position. The upper pair is a time–normalized version of KB’s [–focus,+final] declarative contours, averaged over four words (see Figure 4.1, panel c), and the lower pair corresponds to the word *bein* pronounced in the same context. It should be noted that the pitch ranges observed in each contour are 1.05 ERB on average, which makes them comparable to contours 4 and 5 in Figure 4.5. We may assume that fluctuations within this narrow range are ignored by the human listener, but Pearson’s coefficients will not, as these are designed to take only the co-variation of the data points into account, not the size of the differences between them. Any difference in direction will be treated as important, which explains the dramatic consequences of the irregularities detected in the contours for the similarity measure concerned. In this case, RMSE scores seem to reflect more accurately the resemblance between the two contour pairs, with rather low distance values in both cases.
Figure 4.6. Pairs of Accent 1 – Accent 2 $f_0$ contours and two objective distance measures (Pearson’s coefficients, RMSE scores) found for each pair. Upper pair: contours measured in four different words, time–normalized and averaged. Lower pair: time–normalized contour found in one word, bein. Irregularities in the contours have more dramatic consequences for Pearson’s coefficients than for the RMSE values.

Since several of our contour pairs are defined within small $f_0$ ranges, we expect RMSE scores to represent more realistic distance measures than Pearson’s $r_s$, at least for these pairs. On the other hand, giving up Pearson’s coefficients implies that we do not take full account of differences in contour trajectories which are likely to be reflected in human perception. A compromise may be found in the cosine correlation function, which computes distances in a way that resembles the Pearson correlation function, while also taking pitch range and register into consideration, as is shown by a comparison between them. First, the Pearson correlation between two contours $g$ and $h$ (which, in our case, are vectors of 100 values each) is defined as in (1).

$$ r_{Pearson} = \frac{\sum_{i=1}^{p} (g_i - \bar{g})(h_i - \bar{h})}{\sqrt{\sum_{i=1}^{p} (g_i - \bar{g})^2} \sqrt{\sum_{i=1}^{p} (h_i - \bar{h})^2}}, $$

where $\bar{g}$ and $\bar{h}$ are the average $f_0$ values of the two contours. The formula for cosine coefficients is almost the same, as shown in (2), except that the average $f_0$ is not subtracted, which prevents normalization.
The similarity scores produced by (2), \(r_{\cos}\), can be more easily compared with the RMSE measures if they are transformed into distance measures \(d_{\cos}\) by means of \(d_{\cos} = 100 \times (1 - r_{\cos})\), which leaves the properties of the cosine method unaffected.\(^{37}\) The similarity scores are given in the third column of Table 4.2. Observe that these scores would appear to be a compromise between Pearson’s \(r\) and RMSE. It is reassuring to see, for instance, that the \(d_{\cos}\) between the steep and slow falls (contours 1 and 5) is smaller than that between slow fall and the steep rise (contours 3 and 5). By contrast, the RMSE scores are almost the same, while the Pearson’s scores would appear to exaggerate the difference by giving them scores that lie close to the theoretical extremes of 1 and –1. Similar results were found in the real-life examples shown in Figure 4.6. For the upper pair, the cosine distance was 0.13, against 4.20 for the lower pair, which nicely reflects the irregularities observed in the lower pair, and at the same time acknowledges the similarities observed within each of the pairs of contours (both falling at first, and evolving in about the same pitch register).

In view of these considerations, we decided not to report Pearson’s \(r\)s nor employ them in our further exploration of the connection between contrast salience and recognition. Since both RMSE and cosine scores appear to provide more realistic distance measures, we decided to retain both distance measures, to see how a purely range based distance measure compares with one that also includes information about direction.

### 4.4.1.2 Objective distances between Accent 1 and Accent 2

RMSE and cosine distance scores between Accent 1 and Accent 2 were computed separately for each context and intonation pattern, and for each speaker, based on 100 \(f_0\) values per contour. The results are given in Figure 4.7.

\(^{37}\)When computing correlations between cosine coefficients and other measures (see section 4.4.1.2), we found exactly the same results for the cosine and the transformed cosine measures, except for their sign: where coefficients involving cosine values were positive, those involving the transformed cosine values were negative, and vice-versa. This is directly related to the fact that the cosine coefficients, like Pearson’s coefficients, increase as the distance between contours decreases (negative correlation), while their transformed version would increase as distance increases (positive correlation).
A first observation to be made on these plots is that the cosine distance measures have a wider range, causing differences to be more pronounced, but that broadly the same pattern of results is obtained as that shown in the RMSE graph. In both scales, the largest distance is found in the [+focus,+final] context with ‘low interrogative’ intonation. In the corresponding [–final] case, the distances are very much smaller. In declarative sentences, the [+focus] contexts also clearly dominate the [–focus,+final] one. When the target words were pronounced with ‘continuation’ or ‘high interrogative’ intonation, distances are generally rather small, but the lowest scores are again found in the [–focus,+final] context. That is, on the basis of both measures we can conclude that tone contrasts with ‘continuation’ intonation obtain lower distance scores than contrasts with ‘high interrogative’ intonation, while the highest scores are obtained in the ‘low interrogative’ intonation. At a higher level of aggregation, the cosine distance measures suggest the main difference is between low–scoring ‘continuation’ and ‘high interrogative’ intonations on the one hand and ‘declarative’ and ‘low interrogative’ intonations on the other.

As might be expected, there is no general pattern in the way the differences between speakers are characterized. In five out of the twelve comparisons, both distance measures agree there is a larger contrast for one speaker than for the other, and in seven cases the distance measures disagree as to which speaker has the largest contrast. One case of agreement is worth pointing out. Both measures reveal a large difference in the case of speaker KB in the [+focus,+final] context with ‘continuation’ and ‘high interrogative’ intonation. This is explained by KB’s steep rise in Accent 1, which starts from a lower pitch than the corresponding contours in KB’s realizations. This difference between speakers is not obviously reflected in the recognition scores (see Figure 4.4), neither is it reflected in the subjective distance judgements, as we will see in section 4.4.2.
The RMSE and cosine distances measures are based on $f_0$ measures only, and thus ignore differences in duration between Accent 1 and Accent 2. However, as we saw in section 4.2.3.2 (Figures 4.2 and 4.3), Accent 2 is longer than Accent 1, in particular in the final contexts. Accordingly, we included the durational contrasts, as shown in Figure 4.2, as an additional vector in the analysis. All three distance measures reveal clear differences between Accent 1 and Accent 2 words in [+focus,+final] contexts, but generally the pattern for the durational distance measure is rather different from those found for the RMSE and cosine distances. The [+focus,—final] context, in particular, shows that durational differences between Accent 1 and Accent 2 are more sensitive to the position of the target word in the IP than to its focus situation, whereas the opposite is true for $f_0$ differences. In declaratives and low questions, there is thus always at least one acoustic cue to highlight the tonal contrast. By contrast, in the ‘continuation’ and ‘high interrogative’ intonations, the contrasts in the [+focus,—final] context show little salience, since neither $f_0$ nor duration seem to provide solid indications of tonal identity. These generalizations seem sufficiently robust, but it should be noted that they are only partly supported by the statistical analysis. As was said in section 4.2.3.2, the factor CONTEXT was found significant for durational differences. However, univariate ANOVAs with the factors intonation (fixed), context (fixed) and speaker (random) showed no significant effects or interactions of effects for RMSE at the 5% level, and for the cosine distances, only INTONATION (p=0.036), and INTONATION*CONTEXT (p=0.004) were found significant. Before moving on to a comparison between these objective scores and the subjective scores, and an analysis of the relation between phonetic salience and recognition, we provide an evaluation of the claim that in [–focus,—final] contexts, the contrast between Accent 1 and Accent is neutralized.

### 4.4.1.3 Ruling out the [–focus,—final] context

Figure 4.8 gives the $f_0$ contours for Accent 1 and 2 for the ‘declarative’ and ‘low interrogative’ intonations in [–focus,—final] position. As will be clear, they are highly similar within each speaker’s set of utterances.

We are now in a position to give a quantitative measure of the acoustic differences between Accent 1 and Accent 2 in these contexts. RMSE and cosine distance measures were computed as we did for the other contexts. We found cosine correlation coefficients between 0.9979 and 0.9999, or 0.21 and 0.01 in our converted scale, which means there is a great similarity between the contours in this context compared to the other contexts. The RMSE distance scores gave similar results. In addition to $f_0$, a few more measures were computed in order to

38 There was one exception to the low RMSE distance scores, a not-too-low RMS (.7) in the prenuclear case, when pronounced with interrogative intonation. This score is due to differences in pitch register, especially in one speaker. Such pitch register differences are likely to be related to the intonation used in the carrier sentence. Speakers generally use a
a. prenuclear

Statement

decl, [focus, final], prenuc., KB
decl, [focus, final], prenuc., YK

Low question

lewQ, [focus, final], prenuc., KB
dewQ, [focus, final], prenuc., YK

b. postnuclear

Statement

decl, [focus, final] postnuc., KB
decl, [focus, final] postnuc., YK

higher pitch register in questions than in statements (Ohala 1984, Gussenhoven 2002, Haan & van Heuven 2000 for Dutch). They can then decide to which extent pitch register is raised in comparison with the more standard statement register. Some speakers, as YK, are fairly regular in their choice of pitch register for questions, but others, as KB, may introduce a greater variation in a set of equivalent utterances. Such variation has an influence on automatic distance measurements, but looking at the contours in each condition, we could not find any systematic difference between Accent 1 and Accent 2.
exclude other possible ways of encoding a tonal contrast. First of all, we examined durations in Accent 1 and Accent 2 rhymes. We found very little difference between both tones. In prenuclear cases, the difference between Accent 2 and Accent 1 is smaller than 10 milliseconds (4 ms in statements, minus 7 ms in questions). In postnuclear cases, in which we saw that durational differences are the largest (cf. section 4.2.3.2), we found an average difference of 2 ms in statements and 12 ms in questions. Second, we computed F1, F2, F3 and intensity values for all [–focus,–final] cases. Again, for each variable we considered, we found a great deal of overlap between curves for Accent 1 and Accent 2. This observation was put to the test by averaging 100 values in each rhyme, in two halves (the first 50 values and the remaining 50), and by comparing these averages by means of paired t-tests (Accent 1 vs. Accent 2). Out of the 32 tests carried out on F1, F2, F3 and intensity averages for the pre- and postnuclear cases, one yielded a significant difference between means for Accent 1 and Accent 2, the comparison between Accent 1 and Accent 2 intensity values in the second part of postnuclear rhymes. This exception is likely to be accidental, and thus unlikely to reflect a true difference between Accent 1 and Accent 2. Without testing this conclusion in a perception experiment, we assume that there is no difference between the dialects of Venlo and Roermond (see Fournier et al. 2006) in the way that the tonal contrast is neutralized in [–focus,–final] contexts.

4.4.2 Subjective distances between Accent 1 and Accent 2

4.4.2.1 Method

The excised portions of the speech wave forms corresponding to the four experimental words bein, derm, knien and stein were arranged in pairs in the order Accent 1-Accent 2 as well as Accent 2-Accent 1, one for each of the 12 conditions.
defined by the four intonations and the three contexts, for each speaker separately. This yielded 12 x 4 (words) x 2 (speakers) x 2 (orders), or 192 pairs. In addition, we prepared 48 pairs of identical stimuli, which were to serve as a baseline for the minimal distance score (“no difference”). In order to neutralize a possible influence of presentation order within the pairs and of the order of presentation of the pairs, we prepared two tests, A and B, each with 96 minimal pairs and 48 identical items, which were each other’s mirror images both with respect to the stimulus order in each pair and with respect to the order of presentation. Since we could not provide a reference for a maximal (or medial) distance without compromising the objectivity of the test, the experimental stimuli were preceded by twelve stimulus pairs representing all the combinations of intonation and context, plus two pairs of identical stimuli, in order for the subjects to get an impression of the range of differences they were asked to assess. This orientation set is given in Appendix 2. Each pair was presented twice, followed by a short piano tone and a 3 s pause during which subjects recorded their judgements. Stimulus pairs were arranged in blocks of ten, with a longer piano tone and an extra 2 s pause occurring between blocks.

Twenty native speakers of the Venlo dialect and twenty native speakers of Standard Dutch with no knowledge of any Limburgian dialect were recruited as judges. Half the judges in each group were presented with test A, and the other half with test B. The mean age of the Venlo group was 17 years and that of the Standard Dutch group was 20 years. In both language groups, there were more women than men (12 women and 8 men in the Venlo group, and 11 women and 9 men in the Standard Dutch group). No subject reported a hearing problem. Nine subjects in the Venlo group also participated in the Recognition Experiment (section 4.3). For these judges, the subjective distance test took place one hour after the recognition test. The listening experiments took place in class rooms in Venlo (for the Venlo group) and Nijmegen (for the Standard Dutch group). Subjects were instructed to judge the size difference between the members of each experimental pair. They listened to the stimuli through headphones and were provided with an answer sheet which listed the word used in each stimulus pair, followed by a 10-point scale in which a distance score could be registered (an example is given in Appendix 2). The instructions briefly described the structure of the stimulus blocks and explained how the scale should be interpreted: 0 meant that there was no difference between words in a pair, and 9 that a very clear difference could be heard, of the sort that would even be audible in a very noisy room. No mention was made of the kind of difference that subjects should focus on.

### 4.4.2.2 Results for the Standard Dutch listeners

In a first step, we checked our data for outliers, in order to ensure a homogeneous set of scores. This was done by computing Pearson’s coefficients between each judge’s scores and the mean scores over all subjects. Coefficients (r) ranged from 0.56 to 0.88, which are high enough to allow us to keep all judges in the data set.
We then tested our data for effects of presentation order. To this end, we first ran an ANOVA with the within-subjects factors INTONATION, CONTEXT and SPEAKER and PRESENTATION_ORDER as a between-subjects factor. There was not no main effect of PRESENTATION_ORDER and no interactions with any of the other factors, and we could therefore consider our results a single data set. Accordingly, we ran the ANOVA again, this time without PRESENTATION_ORDER. We found main effects for INTONATION and CONTEXT, as well as the interactions INTONATION*SPEAKER, INTONATION*CONTEXT and INTONATION*CONTEXT*SPEAKER (all effects<.001).

The upper panel of Figure 4.9 shows these results. The effect of INTONATION can be seen in the higher distance scores for ‘declarative’ and ‘low interrogative’ intonation than for ‘high interrogative’ intonation, which differences are significant by post-hoc tests done for all pairs of intonations (p<.001). The main effect of CONTEXT is visible in that [+focus,+final] words were judged more salient than the other two contexts across all intonations, while [+focus,–final] is more salient than [–focus,–final], as confirmed by post-hoc tests (p<.001). As suggested by the INTONATION*CONTEXT interaction, different context rankings do emerge depending on the intonation used. In the ‘high interrogative’ intonation, word pairs in the [–focus, +final] context are judged to be somewhat more salient than in the [+focus,–final] context, while for the other intonations this is the least salient contrast. The interaction INTONATION*SPEAKER must be due to the fact that KB consistently makes larger contrasts than YK in the ‘declarative’ intonation, but smaller contrasts in the ‘high interrogative’ intonation. Finally, the INTONATION*CONTEXT*SPEAKER interaction reflects the mixed picture for the other two intonations: the [+focus,+final] context is better for YK in the ‘low interrogative’ intonation, but for KB in the ‘continuative’ intonation.

Figure 4.9: Subjective distance by non-native (upper panel) and native speakers (lower panel), averaged per intonation (declarative, low question, continuative and high question), context (a= [+focus,+final], b= [+focus,–final] and c= [–focus, +final]) and speaker (KB, YK).
4.4.2.3 Results for the Venlo listeners

We applied the same treatment to the results of the Venlo group of listeners, whose results are given in the lower panel of Figure 4.9. The correlations between each subject and the mean of all subjects yielded significant $r$’s between 0.43 and 0.88 for 18 listeners, while the scores for the remaining two did not correlate with the mean scores ($r=-0.05$ and 0.08). We excluded these two subjects from further processing. The first ANOVA again included PRESENTATION_ORDER as a between–subjects factor by the side of INTONATION, CONTEXT and SPEAKER and PRESENTATION_ORDER as a within–subjects factors, and again yielded no significant effects involving PRESENTATION_ORDER. The ANOVA without PRESENTATION_ORDER yielded all the effects we found for the Dutch listeners, at the same levels of significance, except INTONATION*SPEAKER which was significant only at $p<.05$. It is to be noted that in the results for the Venlo listeners, we do not find the consistently larger contrasts for YK in the ‘high interrogative’ intonation.

Post-hoc tests showed that the effect of INTONATION is due to significant differences for all comparisons except for that between the ‘low interrogative’ and ‘continuative’ intonations ($p<.001$). As in the case of the Dutch listeners, the effect of CONTEXT is due to the higher scores in [+focus,+final] words than in the other two contexts ($p<0.001$), but for the Venlo listeners [+focus,–final] is not more salient than [–focus,–final].

4.4.2.4 Native and non-native judgements compared

In view of the very similar results for the two groups of listeners, the question arises whether it is meaningful to keep the groups separate. An ANOVA with a between–subjects factor LANGUAGE_GROUP (2 levels) and with the same within-subjects factors as in the earlier analyses showed the same effects and interactions as did the analysis of the scores of the Standard Dutch group by itself, and in addition yielded a four-way interaction LANGUAGE_GROUP*INTONATION*CONTEXT*SPEAKER ($p<0.01$). This effect can be seen in the upper and lower panels of Figure 4.10, which shows that KB is judged by the Venlo listeners to

---

39 Even after this exclusion, there remained a great deal of variation amongst the subjects’ judging strategies. While most judges used a broad range of possible scores (the lowest score was always 0 or 1, and the highest one was 8 or 9 in 15 cases), there were three cases in which only four or less levels were used. However, we did not opt for range normalization when comparing the scores (i.e. stretching the four or less levels to an average score range and adjusting all scores accordingly). By normalizing score ranges, we would assume that the maximum level in each speaker’s scores always corresponds to a very large distance, while it may well be the case that some subjects simply did not perceive any large distances at all and hence kept their scores low.
have a larger contrast than YK in the [–focus] syllables in the ‘continuative’ and ‘high interrogative’ intonations, but not by the Standard Dutch listeners. Inspection of the realizations in Figures 4.1 and 4.2 suggests that the Venlo listeners may have been more sensitive to the $f_0$ differences in these contrasts, since the duration differences are larger for YK in this context.

Apart from this case, the results are very comparable. The correlation (Pearson’s $r$) between native vs. non-native scores is a high 0.86. We therefore pooled the results over the two groups of subjects (38 subjects in total), as has been done in Figure 4.10. In our evaluation of the salience of the recognition scores, we will use these pooled data.

![Subjective distance scores given by 38 native and non-native speakers, averaged per intonation (declarative, low question, continuative and high question), context (a= [+focus, +final], b= [+focus, –final] and c= [–focus, +final]) and speaker (KB, YK).]

4.5 Contrast salience and recognizability

In the previous section, we showed that the phonetic salience of the Venlo tone contrast varies with the intonation contour of the utterance, the accentuation of the words in question and their position in the sentence. In this section, we will investigate whether low salience of a contrast leads to poor recognizability. In section 4.4.1.3 we have seen that the dialect neutralizes the tone contrast in unaccented (nonfocused), nonfinal positions. Assuming the neutralization was a historical process, as suggested by the presence of the contrast in the dialects of Cologne (Gussenhoven & Peters 2004) and Sittard (Hanssen 2005), our conjecture is that the contrast disappeared from this position due to its lack of salience. A relation between salience and recognizability may therefore signal the imminent loss of non-salient contrasts.

Table 4.3 gives Spearman’s $\rho$’s and significance levels for the recognition scores (Recog), the subjective distance scores (SubjDist), and the three objective distances, the root mean square error (RMSE), $d_{cos}$ and duration (Dur). They are...
based on vectors of 96 elements, i.e. three contexts, four intonations, four words and two speakers. We used the non-parametric Spearman’s ρ, because the objective distance measures are not normally distributed.

There is no correlation between duration differences and the recognition scores, which suggests that recognition is based on \( f_0 \) differences only. This is an unexpected result in the view of our acoustical analysis of the stimuli, which suggests that duration may be more important than \( f_0 \) in quite a few cases. We tried to look at our data from different angles in order to exclude two possible blurring factors. First, we recomputed the \( ρ \) coefficients with the results in a [–focus, +final] context only. The new coefficients were hardly higher than the old ones. Second, we took one step back and considered the recognition scores and durations for Accent 1 and Accent 2 separately, in the [–focus, +final] context. Rather than using absolute duration values (determined by manual segmentation), we compared the durations with average Accent 1 and Accent 2 durations, as computed per speaker and within the [–focus, +final] context only. Departures from these averages were assigned a minus sign if they induced potential confusions between Accent 1 and Accent 2, and a plus sign otherwise. For instance, if an Accent 2 rhyme pronounced by YK had a duration of 300ms whereas the average for YK’s Accent 2 [–focus, +final] context was 360ms, its “relative duration” value was set to -60; if the rhyme was 390ms long, its new value was 30. With this method, we could sort our data in a way that on the one side of the ladder (the very small, negative numbers), we had the most non-typical instances of Accent 1 and Accent 2, and on the other side, the most typical ones. We then computed the correlation between these new values and the recognition scores. Again, the result was not significant, showing that the long instances of Accent 2 were not recognized better than the potentially confusing instances of Accent 1 and Accent 2 (in terms of duration). We must conclude that the Venlo listeners did not, after all, rely on duration during recognition.

Although significant, the correlation between recognition and the two objective \( f_0 \) distances is low. Of these, \( d_{cos} \) explains the recognition scores best, with \( r=0.47 \), which suggests that a combination of range differences and contour shape is superior to a distance measure based on contour shape differences alone. However, a breakdown over the three positional contexts revealed that this correlation is entirely due to the correlation in the [+focus, –final] context, meaning that in phrase–final syllables we observe no relation between acoustic differences and recognition. The tenuous nature of the relation is further underscored by the fact that a breakdown over intonation contours only allows the correlation of \( d_{cos} \) to

---

40 By computing correlations between absolute duration values (as determined by manual segmentation) and recognition scores, we would have answered the question whether long words were recognized better than short ones, which obviously was not our question. We rather wanted to know whether abnormally long instances of Accent 1 or short instances of Accent 2 attracted worse scores than short Accent 1 or long Accent 2 words.
survive for the ‘continuative’ intonation, while no correlation with RMSE was significant.

The correlation between the recognition scores and the subjective distances is 0.38, which means that a bare 14% of the variation in the recognition success of the Venlo lexical tones is explained by the phonetic salience of the difference between Accent 1 and 2 in the context concerned. A breakdown over the three contexts leaves no correlation intact, and a breakdown over intonation contours shows correlations only for the ‘declarative’ (0.54) and ‘low interrogative’ (0.41) intonations (both \( p < 0.5 \)).

It could be argued that the recognition success of a particular tone, say Accent 1, in a particular context is determined not just by its discriminability from the contrasting tone, Accent 2, in the same context, but in addition with its discriminability from both tones in all other contexts, and that we should therefore establish the mean phonetic difference of each of the 24 forms with the mean of the 23 other forms. These measures could then be correlated with the recognition scores to see if a tone’s distinctiveness in the phonetic space used by the dialect correlates with its recognizability. Quite apart from the practical problem of obtaining subjective distance measures for 23*24 or 553 stimulus pairs, this procedure would fail to reflect the fact that the loss of a form always means the loss of an opposition in a particular context. In conclusion, we find that the recognition success of the Venlo tones is explained by the differences in the \( f_0 \) contour shape and pitch range between them in a given context.

Table 4.3. Spearman’s \( \rho \) and significance levels between recognition scores (Recog), difference judgements (SubjDist), Root Mean Square Error (RMSE), inverted cosines and duration difference. \( N = 96 \). Values in grey are duplicates.

<table>
<thead>
<tr>
<th></th>
<th>recogV</th>
<th>SubjDist</th>
<th>RMSE</th>
<th>d_cos</th>
<th>dur abs</th>
</tr>
</thead>
<tbody>
<tr>
<td>RecogV</td>
<td>Corr. coeff.</td>
<td>1,000</td>
<td>0,379</td>
<td>0,322</td>
<td>0,466</td>
</tr>
<tr>
<td></td>
<td>Sig. (2–tailed)</td>
<td>.</td>
<td>0,000</td>
<td>0,001</td>
<td>0,000</td>
</tr>
<tr>
<td>SubjDist</td>
<td>Corr. coeff.</td>
<td>0,379</td>
<td>1,000</td>
<td>0,645</td>
<td>0,471</td>
</tr>
<tr>
<td></td>
<td>Sig. (2–tailed)</td>
<td>0,000</td>
<td>.</td>
<td>0,000</td>
<td>0,000</td>
</tr>
<tr>
<td>RMSE</td>
<td>Corr. coeff.</td>
<td>0,322</td>
<td>0,645</td>
<td>1,000</td>
<td>0,791</td>
</tr>
<tr>
<td></td>
<td>Sig. (2–tailed)</td>
<td>0,001</td>
<td>0,000</td>
<td>.</td>
<td>0,000</td>
</tr>
<tr>
<td>d_cos</td>
<td>Corr. coeff.</td>
<td>0,466</td>
<td>0,471</td>
<td>0,791</td>
<td>1,000</td>
</tr>
<tr>
<td></td>
<td>Sig. (2–tailed)</td>
<td>0,000</td>
<td>0,000</td>
<td>0,000</td>
<td>.</td>
</tr>
<tr>
<td>dur abs</td>
<td>Corr. coeff.</td>
<td>0,119</td>
<td>0,216</td>
<td>-0,115</td>
<td>-0,108</td>
</tr>
<tr>
<td></td>
<td>Sig. (2–tailed)</td>
<td>0,249</td>
<td>0,034</td>
<td>0,266</td>
<td>0,296</td>
</tr>
</tbody>
</table>
The extent to which the subjective distance measures are explained by the objective distance measures can be answered by inspecting their correlations.\(^{41}\) We find that both \(d_{\cos}\) and RMSE correlate moderately with the subjective distance scores, and that there is a weak correlation with the duration differences. A breakdown of these coefficients over contexts and listener groups revealed that the significance of the correlation with the duration difference was due only to the \([-\text{focus},+\text{final}]\) context (0.44, \(p=0.012\) for the Standard Dutch listeners and 0.49, \(p=0.005\) for the Venlo listeners). Conversely, the correlations with RMSE and \(d_{\cos}\) disappeared in this context, except for a weak correlation for the Standard Dutch listeners of 0.37 (\(p<0.04\)). For both groups, the phonetic salience scores are therefore solely explained by \(f_0\) differences in the focal contexts and largely by duration differences in the final unfocused context.

4.6 Discussion and conclusion

We have shown that the lexical tone contrast between Accent 1 and Accent 2 in the dialect of Venlo appears in a large variety of contour shapes, depending on the intonation used, the position in the Intonational Phrase, and if final in the IP, on the presence of a focus marking accent. The recognition of the tones varies considerably across these contexts, from mean rates of 77% and 75% in the ‘declarative’ and ‘low interrogative’ intonations to 63% and 66% in the ‘continuative’ and ‘high interrogative’ intonations. In addition, the contrast is better recognized in focused final syllables than in nonfinal or nonfocused syllables. In nonfinal nonfocused syllables, the contrast could be shown to be neutralized on the basis of the production data, confirming the description in Gussenhoven & van der Vliet (1999) on this point. The results for the two speakers in the experiment were very similar. Other than in the dialect of Roermond (Fournier et al. 2006), we found that older speakers were better at recognizing the tones than younger speakers. This is an indication that the tone contrast is subject to erosion; in fact, even in the older group recognition was not as good as in the Roermond group.

In a perception experiment in which listeners were asked to judge the perceived phonetic difference between the two lexical tones in each of the twelve prosodic contexts, we found that native and non-native listeners strongly agreed on the degree of phonetic salience of the phonetic contrasts. There was no indication that listeners were influenced in their phonetic judgements by the phonological status of the difference in their language. This finding is reminiscent of a study on Chinese by Hallé, Chang & Best (2004) which compared the judgement of native

---

\(^{41}\) We did not run regression analyses to predict either the recognition scores or subjective distance scores because of the multicollinearity present in the data, where some of the predictor variables have higher correlations with each other than with the dependent variable.
(Taiwanese) and non-native (French) speakers on Taiwanese tones. Results suggested that in spite of their non-tonal background, French subjects were as sensitive to pitch differences as Taiwanese were; what distinguished the two groups was found at a deeper cognitive level, in that Taiwanese showed a quasi-categorical mode of processing while French perception was psychoacoustically based.

There appeared to be a weak correlation between the subjective salience measures and the recognition rates, providing weak support that phonetic salience determines the perceivability and hence the robustness of a phonological contrast. However, the correlation was smaller than expected, with only 14% of the variation being explained by the subjective phonetic salience.

In accordance with the description in Gussenhoven & van der Vliet (1999), who report only durational differences in all intonations except ‘declarative’ in nonfocused final syllables, our acoustic difference measures showed that the contrast between Accent 1 and Accent 2 was encoded with different kinds of acoustic information, $f_0$ or duration. While the recognizability of focused (accented) syllables should mainly be explained by the $f_0$ differences between the members of each pair, we expected the recognition of the final nonfocused syllables to be only explained by the durational differences between them. However, the correlation coefficients computed between recognition and duration in the [–focus,+final] context, whether based on averages between Accent 1 and Accent 2 or on individual values, were not found significant. Although native listeners do acknowledge durational differences in this context (as shown by the subjective distance judgements), they do not use them reliably during recognition.

The acoustic difference measures, RMSE, $d$–cos and duration, had a moderate correlation with the subjective distance measure. The closest measure was RMSE, although $d$–cos showed a higher correlation with recognition scores. These results suggest that phonetic salience judgements are more complex than can be captured by the measures we used. This is not too surprising if we consider the general difference between automatic and human speech recognition; it often happens that a speech signal which is not correctly interpreted by a computer device will still be easily identified by human ear, thanks to the capacity of listeners to abstract away from irregularities of the signal (e.g. background noise or small $f_0$ fluctuations) and prioritize the relevant cues. Future research should focus more on the exact relevance of individual cues for similarity judgements and recognition. By manipulating $f_0$ and duration, we may be able to establish with more accuracy the weighting of cues for recognition on the one hand, and phonetic distance judgements on the other.
4.7 Appendix: Stimuli

4.7.1 Sentences used in the training session

The type of context (focus situation, position in the sentence, and intonation contour) is specified in the first column. Note that according to the official spelling (found in Alsters et al. 1993), the Venlo words for ‘beer’ and ‘bear’ are spelled differently. However, this had no effect on the pronunciation of the nucleus, which in both cases was \([e:]\).

<table>
<thead>
<tr>
<th>Context</th>
<th>Accent 1: beer ([be:r], ‘beer’)</th>
<th>Accent 2: baer ([be:r], ‘bear’)</th>
</tr>
</thead>
</table>
| [+focus, +final] declarative | – Waat hebse gedronken?  
– Ein glaas BEER.  
– What did you drink?  
– A glass of BEER. | – Waat hebse gejaag?  
– Einen BAER.  
– What did you hunt?  
– A BEAR. |
| [+focus, +final] interrogative | Drinkse ein glaas BEER?  
Are you having a glass of BEER? | Jaagse ein BAER?  
Are you hunting a BEAR? |
| [+focus, –final] interrogative | – Haet hae BEER gedronke?  
– Nae, allein KÖFFIE.  
– Did he drink BEER?  
– No, just coffee. | – Haet hae einen BAER gejaag?  
– Nae, allein einen HAAS.  
– Did he hunt a BEAR?  
– No, just a HARE. |
| [+focus, +final] interrogative & declarative | – Is det DÖNKER beer?  
– Nae, ’t is BLOND beer.  
– Is this DARK ale?  
– No, this is BLOND ale (= lager). | – Is det ’nen BROÈNE baer?  
– Nae, ein ZWARTE baer.  
– Is this a BROWN bear?  
– No, a BLACK bear. |
| [+focus, –final], prenuclear, interrogative & declarative | – Is ’t beer DÖNKER?  
– Nae, ’t beer is BLOND.  
– Is the ale DARK?  
– No, it is BLOND. | – Is d’n baer BROÈN?  
– Nae, d’n baer is ZWART.  
– Is the bear BROWN?  
– No, the bear is BLACK. |
| [+focus, –final], postnuclear, interrogative & declarative | – Hebse DÖNKER beer gedronke?  
– Nae, ik heb BLOND beer gedronke.  
– Did you drink DARK ale?  
– No, I drank BLOND ale. | – Hebse ’nen BROÈNE baer gejaag?  
– Nae, ik heb ’n ZWARTE baer gejaag.  
– Did you hunt a BROWN bear?  
– No, I hunted a BLACK bear. |
4.7.2 Words used the subjective distance experiment

4.7.2.1 Preparatory items (in order of presentation)

bein, declarative, [+focus,+final], speaker KB, Accent 1 then Accent 2
stein, high question,[+focus,+final],speaker YK, Accent 2 (twice the same stimulus)
knien, low question, [–focus,+final], speaker YK, Accent 2 then Accent 1
stein, high question, [–focus,+final], speaker YK, Accent 1 then Accent 2
derm, continuation, [+focus,–final], speaker KB, Accent 2 then Accent 1
bein, high question, [+focus,–final], speaker YK, Accent 1 then Accent 2
knien, continuation, [–focus,+final], speaker KB, Accent 2 then Accent 1
stein, declarative, [+focus,+final], speaker YK, Accent 1 then Accent 2
bein, continuation, [–focus,+final], speaker KB, Accent 1 (twice the same stimulus)
derm, low question, [+focus,+final], speaker KB, Accent 2 then Accent 1
bein, low question, [+focus,–final], speaker YK, Accent 1 then Accent 2
knien, high question, [–focus,+final], speaker YK, Accent 2 then Accent 1
stein, continuation, [+focus,–final], speaker KB, Accent 1 then Accent 2
derm, declarative, [+focus,+final], speaker KB, Accent 2 then Accent 1

4.7.2.2 Answer sheet (extract)

Blokje 1
geen verschil ........................................... heel veel verschil

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>bein</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>stein</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>[..]</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>derm</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
</tbody>
</table>

0 1 2 3 4 5 6 7 8 9
5 Lateralization of tonal and intonational pitch processing: An MEG study

5.1 Introduction

Although the number of studies on speech prosody has grown considerably in the last decades, there is no consensus on how its different components are apprehended by the human brain (see Baum & Pell 1999, for a review). First, it has not so far been established with certainty whether prosody should be considered a mosaic of parameters ($f_0$, duration, intensity) which may be processed in different areas of the brain (van Lancker & Sidtis 1992, Zatorre, Evans, Meyer and Gjedde, 1992), or whether the parameters form an entity that is treated differently depending on its cognitive function (van Lancker 1980). This so-called functional lateralization hypothesis holds that the processing of linguistically-relevant prosodic features is lateralized to the left hemisphere (LH), while emotional aspects of prosody are processed in the right hemisphere (RH), or at least, not predominantly in the LH (Starkstein, Federoff, Price, Leiguarda & Robinson 1994).

The LH specialization for the linguistic use of prosodic cues has been identified, for instance, in a PET (positron emission tomography) study by Gandour et al. (2000) based on Thai. As a tone language, Thai may distinguish lexical items only by means of a difference in pitch contour. The authors could show that native speakers used their left frontal operculum (in the vicinity of Broca's area, well-known for its role in language processing) when discriminating pitch patterns in Thai words. By contrast, native speakers of English (a non-tone language), but also of Chinese (a tone language), who were not familiar with Thai, did not activate this area when exposed to the same stimuli and rather exhibited a RH dominance during the pitch discrimination task. These results suggest that the LH is activated only if the input can be interpreted in a linguistic way, within a known system of tonal contrasts that convey lexical meaning.

42 A version of this chapter is to appear in Brain Research as: R. Fournier, C. Gussenhoven, O. Jensen and P. Hagoort (200X). Lateralization of tonal and intonational pitch processing: An MEG study.
The general picture whereby lexical tonal contrasts are more LH-lateralized than non-linguistic pitch processing, fails to address a number of issues. First, lexical meaning is not the only linguistic component in which pitch is involved. Using the same phonetic parameter on the intonational level, speakers are also able to express *discoursal meanings*, such as the distinction between questions and statements. Recent studies on German (Friederici & Alter 2004; Meyer, Steinhauser, Alter, Friederici & von Cramon 2004) as well as on Mandarin Chinese (Gandour et al. 2003) point to an increased RH activity during the processing of intonation. However, the stimuli used in these studies were not of the same kind as those used in most experiments on lexical tones. Instead of words, intonation studies usually have recourse to phrases or sentences, which are typical domains for intonational contrasts. As Baum and Pell (1999) point out, there might be a difference in lateralization depending on the domain size of stimuli, such that larger domains tend to be assigned to the RH. A recent cross-language (Chinese, English) fMRI experiment by Gandour et al. (2004) shows that when using equally long stimuli in tone and intonation discrimination tasks, the general RH dominance is in fact less obvious. Regardless of the task, both language groups do exhibit a RH asymmetry, but this asymmetry is confined to some regions, while other regions show a LH dominance in the Chinese speakers only. The interpretation proposed is that the RH mediates the prosodic analysis of complex sounds, while the LH is responsible for language-related processing.

A second issue which is not captured by the lateralization view identified above concerns the timing of pitch processing. Purely topographic hypotheses and research findings significantly underspecify the brain activation patterns that can be assumed to exist. Once a fuller conception is formulated about the tasks that subjects are faced with in pitch-related experiments, more detailed hypotheses readily suggest themselves. For instance, regardless of the status of a pitch contrast (lexical, structural intonational, or emotional), subjects will need to process the acoustic information in the stimuli before their linguistic or emotional status can be ascertained. Given the brain imaging results, this might imply that all stimuli will need to be processed in the RH first, and that subsequently only those that embody linguistic contrasts are further dealt with in the LH. Hypotheses that take a processing perspective as their point of departure cannot of course be tested without taking temporal information into account.

In sum, we believe there to be a need for information about brain activation patterns that are temporally fine-grained enough to be able to trace differently timed activations for similar stimuli. In addition, we need to be assured that we present stimuli that unambiguously represent both lexical and intonational contrasts in otherwise equivalent conditions. This latter requirement implies that we need to adopt a stimulus set comparable to that used in Gandour et al. (2004), in which the stimuli that represented lexical contrasts were equally long as and segmentally comparable to those representing intonational contrasts. Perhaps more so than was the case in Gandour et al. (2004), we need to ensure that the language from which we take our stimuli encodes intonational differences phonologically, i.e. by means of different tonal representations, rather than paralinguistically, i.e. by means of pitch
range variation. On the basis of the literature (Shen 1990, Wu 2000), it could be argued that the difference between questions and statements in Mandarin Chinese may not be phonological in the way that lexical differences are. For instance, the syllable yu pronounced with a falling tone (Tone 4) represents the word for ‘pen’, whereas the same syllable pronounced with a rising tone (Tone 2) represents the word for ‘fish’. By contrast, regardless of whether it is said as a statement or as a question, Tone 2 is always rising. What distinguishes both modes is the higher pitch register used in questions.

The present study aims at providing an account on the processing of tonal and intonational contrasts in Roermond Dutch, a dialect that unambiguously encodes both contrast types phonologically. Like the related tonal dialects spoken in the Netherlands, Belgium (together also referred to as Limburgian) and Germany, the dialect has two lexical tones, called Accent 1 and Accent 2. It is not difficult to find minimal pairs, like vaerI ‘feather’ and vaerII ‘ferry’, or knienI ‘rabbits’ and knienII ‘rabbit’. Since native listeners perceive a lexical and discoursal meaning in any (grammatically correct) pitch contour of a word spoken in isolation, the Roermond dialect allows us to use units of the same length to compare tonal and intonational processing. Unlike Standard Chinese, the Roermond lexical tones can have drastically different pitch contours depending on whether they are said with statement or with interrogative intonation. For instance, a syllable like [kni:n] said with a pitch fall signifies the plural form for ‘rabbits’ spoken as a statement. When the same syllable is spoken with a rise followed by a fall, it signifies the same plural form ‘rabbits’ said with question intonation, and when it is said with a fall followed by a rise, it represents the singular form for ‘rabbit’ as spoken with a statement intonation (Gussenhoven 2000a). In this way, every monosyllabic pitch contour represents a unique combination of discoursal and lexical meaning. There can be no doubt, therefore, that the intonational differences are phonological in the same way as the lexical differences are.

In our experiment, Roermond Dutch word stimuli were used in order to create lexical contrasts (e.g. vaerI ‘feather’ vs. vaerII ‘ferry’) as well as intonational contrasts (e.g. vaerI with statement intonation vs. vaerII with question intonation), or a combination of these two (e.g. vaerI ‘feather’ with statement intonation vs. vaerII ‘ferry’ with question intonation). The different experimental conditions were set up according to the auditory mismatch paradigm, in which sequences of frequent, (near-)identical sounds are interrupted by infrequent, deviant sounds. The idea behind this method is that the repeated presentation of a sound creates a pattern in the listener’s sensory memory, against which every incoming sound is matched. When a deviant sound is presented, the detection of change is reflected by an increased activity in the listener’s auditory cortex in a time window from 150 to 250 ms after stimulus onset. This brain response is referred to as mismatch negativity (MMN), an ERP component that has been used extensively in speech perception studies because, interestingly, its amplitude directly depends on the subject’s language experience. When a deviant phoneme or word pertains to the subject’s mother tongue, the MMN will be stronger than in the case of unfamiliar stimuli (Dehaene-Lambertz 1997, Näätänen 2001, Endrass et al. 2004). It could also be
shown (Phillips, Pellathy & Marantz 2000) that acoustic variation within a phoneme
category triggers weaker and later MMN than the difference between phonemes.

A closer look at the spatial distribution of MMN revealed that the increase in the
MMN amplitude for native (as compared to non-native) differences and for
phonological (as compared to acoustic, i.e. subphonemic) differences is left-
lateralized (Näätänen et al. 1997), reflecting the LH specialization for phonological
processing at the segmental level. However, when elicited in other experimental
conditions, MMN can also have similar or even higher amplitudes in the RH. For
instance, Shtyrov et al. (1998) showed that in the presence of environmental noise,
the MMN amplitude in reaction to phoneme contrasts decreases in the LH while it
increases in the RH. In the musical domain, it could also be shown (Fujikawa,
Trainor, Ross, Kakigi, & Pantev 2004) that LH is not the only host of MMN: no
laterality effect was found in subjects who were presented with five-note piano
melodies differing either in contour (rising vs. falling ending) or in interval (last note
raised by one tone without changing the pitch contour). This bilateral reaction to
pitch contour differences was verified for linguistic pitch as well, in an experiment
which compared intonational (statement vs. question) and segmental contrasts (/a/
vs. /e/) in Japanese (Imaizumi, Mori, Kiritani, Hosoi & Tonoike 2003). While both
contrasts triggered LH activity, only the intonational distinction between falling and
rising contours required an additional RH contribution.

In the light of these results, mismatch responses seem to provide a very convenient
handle for a comparative study of the processing of tone and intonation. On the one
hand, it can be used for highlighting possible task-related laterality effects, and on
the other hand, its sensitivity to language background can help to discriminate
between acoustic and phonological processing. More specifically, with respect to
our experimental goals, we intended to use MMN to verify a number of hypotheses.
First, the functional difference between tonal and intonational contrasts should be
reflected by the lateralization of MMN. We expected tonal processing to be more
left-lateralized than intonational processing, since it is more directly concerned with
lexical characteristics, which have been shown to be left-lateralized (Petersen, Fox,
Second, since all contrasts used in the experiment are well-formed utterances for
native speakers of the Roermond dialect, we assumed that we would find stronger
MMNs in those speakers than in non-native subjects. In order to test this hypothesis,
we presented the same stimuli to speakers of Standard Dutch, a non-tone language.
Since these subjects had no knowledge of Roermond Dutch or any other tone
language, any perceived differences in pitch contours could only be processed
within an intonational or emotional framework, thus shifting the MMN to the RH
more than they would have during a lexical task.

Since the comparison between contrast types as well as between language groups
intended to cover both temporal and spatial aspects of the data, it was decided to
record brain activity with the help of a whole-head MEG system.
Magnetoencephalography (MEG), like EEG (electroencephalogram), is a non-
invasive technique that can measure neuronal brain activity on a millisecond time
scale and hence makes it possible to detect the expected mismatch negativity, or rather magnetic mismatch fields, the magnetic equivalent to MMN. In addition to this outstanding temporal resolution, MEG provides more fine-grained spatial information than EEG. With this technique, we intended to collect new data on tonal and intonational processing which could be analysed against the background of earlier PET/IMRI-based as well as MMN/MMNm-based results.

5.2 Method

5.2.1 Subjects

Twelve right-handed native speakers of the Roermond dialect, between 21 and 34 years old (average age: 25.9 years), took part in the experiment. All subjects claimed to speak their dialect on a daily basis and obtained satisfactory scores in a short test that assessed their active and passive knowledge of Roermond Dutch. The test did not refer to any of the linguistic material that was to be investigated in the MEG experiment. Due to very high numbers of eye or muscle artifacts during the MEG recordings, four speakers were excluded from the analysis, which left us with 5 male and 3 female subjects. For the control group, 11 right-handed native speakers of Standard Dutch were recruited. They matched the Roermond subjects with respect to gender, age (average 26.2) and education (tertiary level), and they were not familiar with any Limburgian dialect or any other tone language. We selected 5 males and 3 females within the control group with the least MEG artefacts. Prior to the experiment, all subjects of the Roermond and control group gave written consent and filled a hand dominance questionnaire to attest their right-handedness. The subjects reported no history of neurological disorders or hearing impairments.

5.2.2 Word stimuli

The stimuli consisted of two sets of Roermond Dutch words recorded by a native speaker, alternately with statement and with question intonation. The first set contained three tonal minimal pairs, that is, word pairs with the same phoneme sequence but different tones (Accent 1 or 2), resulting in different lexical meanings. The second set contained seven words with distinct phoneme sequences, so that their meaning could be identified unambiguously at the segmental level. All recorded words were validated by 20 native speakers, who were asked to identify the lexical meaning as well as the discoursal meaning (‘statement’ or ‘question’). The validation of the stimuli ensured that the test words conveyed a clearly recognizable lexical and discoursal meaning, which was crucial in the case of the tonal minimal pairs. These minimal pairs were: haas\(^1\) (i.e. the syllable [ha:s] pronounced with Accent 1, meaning 'hare') and haas\(^2\) (i.e. the syllable [ha:s] pronounced with Accent 2, meaning 'glove'), graaf\(^1\) (‘canal’) and graaf\(^2\) (‘grave’), and finally, vaer\(^1\) (‘feather’).
and vaer\textsuperscript{\textastersign} (‘ferry’). Since each word was pronounced with two different intonations, it was also a member of an intonational minimal pair. For instance, haas\textsuperscript{\textastersign} (i.e. haas pronounced with Accent 1 and as a statement) formed a tonal minimal pair with haas\textsuperscript{\textdagger}, and an intonational minimal pair with haas\textsuperscript{\textdagger}q (Accent 1, question). The use of minimal pairs allowed us to isolate lexical or intonational differences from phoneme-level differences, in order to facilitate the interpretation of brain responses. However, in natural speech, truly minimal pairs do not exist. There is always some acoustic variation amongst realizations of a word, even in the sections that are phonologically identical. We took this variation into account by using three versions of each expression, i.e. each word as said with either statement or question intonation. In this way, we ensured that any effects of the deviant stimuli would be attributable to the phonological difference with the preceding stimulus, and not to uncontrollable acoustic differences between one pronunciation and the next. This resulted in 36 one-word utterances that were members of tonal minimal pairs (3 versions of 6 words with 2 intonations). For the words which were not members of tonal minimal pairs, namely diek\textsuperscript{\textasteriskcentered} (‘dike’), daak\textsuperscript{\textasteriskcentered} (‘roof’), lien\textsuperscript{\textstar} (‘line’), pien\textsuperscript{\textstar} (‘pain’), bank\textsuperscript{\textdagger} (‘bank’), bandj\textsuperscript{\textdagger} (‘roof’) and huur\textsuperscript{\textdagger} (‘rent’), only one token was used, one with a question intonation for diek\textsuperscript{\textasteriskcentered}, lien\textsuperscript{\textstar} and pien\textsuperscript{\textstar}, and one with a statement intonation for daak\textsuperscript{\textasteriskcentered}, bank\textsuperscript{\textdagger} and huur\textsuperscript{\textdagger}. These words were to be used in a detection task unrelated to the experimental question (see Procedures section) and brain responses to these words were not analyzed. Another set of words for which the brain responses were not analyzed is the minimal pair graaf\textsuperscript{\textdagger} / graaf\textsuperscript{\textdagger}. We decided to ignore these results because of an erroneous labelling of one of the sound files containing a version of graaf\textsuperscript{\textdagger}. Therefore, the experimental results will be based on the minimal pairs haas\textsuperscript{\textdagger}/haas\textsuperscript{\textdagger} and vaer\textsuperscript{\textdagger}/vaer\textsuperscript{\textdagger}.

Figure 5.1 shows the $f_0$ contours of all instances of haas (pronounced [ha:s]) and vaer (pronounced [fe:z\textsuperscript{\textdagger}]) used in the experiment. Although the average word length is 586 ms, the information used for tonal or intonational identification only covers about 300 ms, corresponding to the voiced part of the signal (mostly the vowel [a:] or [e:] plus the transitions between [h] and [a:] and between [e:a] and [x]). There is some variation with respect to the temporal alignment of these voiced parts within the words, but they generally start around 100 ms after word onset. More variation can be found in the timing of the prosodic contrasts. For instance, the statement and question contours for Accent 1 begin to diverge at least 100 ms earlier than the statement and question contours for Accent 2. These timing differences were taken into consideration for the time-locked averaging of our MEG data by shifting the ERPs accordingly (see Data analysis). Two additional systematic differences may be observed in Figure 5.1. First, the voiced part of the word is significantly longer for Accent 2 compared to Accent 1 (43 ms, $p<.001$), and for questions compared to statements (26 ms, $p=.014$; univariate ANOVA; factors: ACCENT and INTONATION). Second, as can be seen in Figure 5.1, the difference between maximum and minimum $f_0$ within a word is smaller in statements than in questions. The possible effects of these various systematic differences on brain reactions will be dealt with in the Discussion section. Other differences can be observed on Figure 5.1, within each cell of the table. These differences reflect the inevitable variation
between instances of the same token. In order to assess the size of the acoustic
differences between different pronunciations of the same linguistic category, we
computed Pearson’s correlation coefficients for all possible pairs of \( f_0 \) contours
within and between categories, based on 100 pitch values per contour, and found
that the correlations within categories were always higher than correlations between
categories. The highest correlation between categories was found for the lexical
contrast (Accent 1 vs. Accent 2) with declarative intonation (average Pearson’s
coefficient: .19, stdev: .2), which is still lower than the lowest coefficient found
within categories (.75). This means that in all cases the differences between tokens
of the same category are small compared to those among categories.

![Figure 5.1: The \( f_0 \) contours of haas and vaer with Accent 1 and 2 combined with
statement or question intonation, as a function of time. Each cell shows the contours
of 3 haas and 3 vaer words. Horizontal arrows represent lexical contrasts (e.g.
haasl_s vs. haasn_s or vaerl_q vs. vaern_l), vertical arrows intonational contrasts
(e.g. haasl_s vs. haasn_q), and diagonal arrows combined contrasts (e.g. haasl_s vs.
haasn_q).]
5.2.3 Procedures

The stimuli were presented in 36 blocks of 100 words implementing a mismatch negativity paradigm. Each 100-word block represented a lexical, an intonational or a combination of a lexical and intonational contrast between *standard* (78%) and *target* stimuli (18%). Both standard and target stimuli belonged to the set of *haas*/*haas* or *vaer*/*vaer* words. A second type of deviant stimuli, the *novel stimuli* (4%), consisted of one-word utterances of one of the seven words which were not members of tonal minimal pairs (*diek*, *daak* and the like). The words in each block were pseudo-randomized in such a way that target stimuli always occurred after at least three standard stimuli, and that these three standard stimuli represented at least two different versions of the same word/intonation combination. We give here as an example the first 12 items in a block representing a lexical contrast, using *haas*/*haas* pronounced as statements, with targets in bold print:

\[
\text{haas}_s(v3) - \text{haas}_s(v1) - \text{haas}_s(v2) - \text{haas}_s(v1) - \text{lien}_q - \\
\text{haas}_s(v1) - \text{haas}_s(v3) - \text{haas}_s(v3) - \text{haas}_s(v1) - \text{haas}_s(v3) - \text{lien}_q - ... 
\]

In this sequence, three different versions of *haas* (v1,v2 and v3) are used as standard stimuli, contrasting with different versions of the target stimuli *haas*.

The word *lien* is used repeatedly as a novel stimulus. We constructed similar blocks with the words *haas* (standard), *haas* (target) and *pien* (novel), *haas* (standard), *haas* (target) and *lien* (novel), and finally *haas* (standard), *haas* (target) and *pien* (novel). The same was done using the *vaer* and *graaf* words. This resulted in 12 blocks for lexical contrasts, of which 8 were retained for the analysis (after exclusion of the *graaf* blocks). Intonational and combined contrasts were modelled in the same way.

All blocks were randomized and presented binaurally to Roermond and control subjects. During presentation, the subjects were requested to focus on a fixation cross and to press a button when they heard a “completely different word”, corresponding to the novel stimuli. This condition was not designed to answer our experimental question, but rather to keep the subjects alert. The instructions concerning the task were deliberately kept vague in order to avoid an explicit reference to the prosodic properties of the stimuli, but the correct execution of the task was supervised during a training session consisting of 10 words (30% target, 10% novel). Corrections only had to be made for two subjects of the Roermond group, who at first pressed the button after both target and novel stimuli. During the main session, all subjects showed the same behaviour with respect to the task, pressing the button for novels only, with almost no mistakes. The main session was divided into three groups of 12 blocks. Within each group, which lasted 20 minutes, blocks were separated by 3-second breaks, during which the fixation cross was replaced by a text announcing the next block. Longer breaks (1 minute) were allowed between groups of 12 blocks. Words in the blocks were delivered every
second. In total, the training and main session took about 65 minutes. The software Presentation (version 0.70, www.neuro-bs.com) was used for delivering the stimuli and recording the behavioral responses.

5.2.4 Data acquisition and analysis

Ongoing brain activity was recorded with a whole-head magnetoencephalography system with 151 axial gradiometers (CTF/VSM Systems, Port Coquitlam, British Columbia, Canada). Data were acquired with a 600 Hz sampling frequency after applying a 150 Hz low-pass filter. Head localization was done at the beginning and end of each recording session, using marker coils placed at the cardinal points of the head (nasion, left and right ear canal). The magnetic fields produced by these coils allowed us to measure the position of the subject’s head with respect to the MEG sensor array. In addition to the MEG, the electrooculograms were recorded from electrodes placed above and below the left eye (vertical EOG) and at the outer canthus of each eye (horizontal EOG), for later eye artifact removal.

MEG data were processed using the FieldTrip toolbox developed at the F.C. Donders Centre for Cognitive Neuroimaging (http://www.ru.nl/fcdonders/fieldtrip/). Event-related fields (ERFs) were calculated from the data time-locked to the stimuli (1 s trials; 0.2 s baseline interval). Trials contaminated by eye movement or sensor jump artifacts were eliminated, leaving about 100 target stimulus trials and 340 standard stimulus trials for each subject and condition (lexical, intonational or combined contrast). The novel stimuli trials were not analyzed since they contained a motor response. Likewise, trials immediately following the novel stimulus trials were also excluded from the study.

In order to take into account possible differences in the timing of prosodic contrasts, we estimated, for each pair of contrasting pitch contours, the time point at which these contours started to differ acoustically. This was done by visually inspecting the plots of the relevant contour pairs. For instance, in order to determine the divergence point in a lexical contrast involving haas₁_s and haas₂_s, we compared all three versions of haas₁_s with the three versions of haas₂_s. The divergence points estimated in these comparisons (as measured from word onsets) were then averaged, giving a single value for the lexical contrasts involving statements. We repeated this operation for the lexical contrasts involving questions, the intonational contrasts involving Accent 1, and so on. Table 5.1 lists the resulting divergence points.

Subsequently, the ERFs were aligned in time according to the points of divergence, while retaining the baseline intervals time-locked to the onset of the words. The ERF were low-pass filtered at 35 Hz. The analysis focused on the difference waveform between standards and targets.
Table 5.1 Divergence points estimated for each experimental condition after word onset. A1/A2 refer to the two lexical tones (Accent 1 and Accent 2) and s/q to the two discourse meanings (s: statement; q: question). Divergence points are the same when stimuli are used as targets or as standards.

<table>
<thead>
<tr>
<th>contrast</th>
<th>standard ↔ target</th>
<th>divergence point</th>
</tr>
</thead>
<tbody>
<tr>
<td>lexical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1, s</td>
<td>A2, s</td>
<td>280 ms</td>
</tr>
<tr>
<td>A1, q</td>
<td>A2, q</td>
<td>115 ms</td>
</tr>
<tr>
<td>intonational</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1, s</td>
<td>A1, q</td>
<td>110 ms</td>
</tr>
<tr>
<td>A2, s</td>
<td>A2, q</td>
<td>215 ms</td>
</tr>
<tr>
<td>combined</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(lex.&amp;int.)</td>
<td>A1, s</td>
<td>150 ms</td>
</tr>
<tr>
<td></td>
<td>A2, q</td>
<td>130 ms</td>
</tr>
</tbody>
</table>

From the fields measured by axial gradiometers, we calculated the planar gradients of the MEG field distribution using a nearest-neighbour method yielding results compatible with the method described by Bastiaansen & Knoche (2000). The horizontal and vertical components of the estimated planar gradients approximate the signal measured by MEG systems with planar gradiometers. The combined planar gradient was then calculated from the root mean square of the horizontal and vertical planar field components (Helenius & Salmelin, 2002). The signals of the combined planar gradient fields are typically largest in magnitude directly above a given source (Hämäläinen et al. 1993). This is particularly advantageous when interpreting distributions of the magnitude of the ERFs at the sensor level.

5.3 Results

In order to characterize the ERFs corresponding to different linguistic uses of pitch differences (intonational or lexical), the brain responses to the standard stimuli were compared to the target stimuli. Figure 5.2a shows the ERFs for the axial gradiometers for the standard and target stimuli combined over all conditions and aligned with respect to the divergence point. The baseline was calculated with respect to the data in a 200 ms time window prior to the stimulus (not prior the divergence point). Systematic deflections in the ERFs are observed over the left and right hemisphere. Note that the difference between standards and targets form two dipolar patterns with polarities consistent with the negative and positive deflections in the fields. The dipolar distributions of the fields (indicated by the circles) suggest a source in the left and the right hemisphere each situated between the negative and positive fields. The strongest deflections are also associated with the strongest differences between the standards and targets. These distributions point to a left and right hemisphere source producing the ERFs and show that the magnitude of these sources is higher for targets than standards. Figure 5.2b displays the difference
between standards and targets, computed by combining equal-polarity fields of the 24 sensors circled in Figure 5.2a. A large difference in the grand average of the waveforms in these 24 sensors was observed around 200 ms. This latency, as well as the region over which the difference in activity occurs, corresponds to the magnetic equivalent of the mismatch negativity response (MMN), as reported for instance in Phillips (2001) and Näätänen (2001). Thus, in the rest of the study we used the interval 150-250ms after the divergence point as the time interval of interest when comparing the different conditions.

To further characterize the MMNm, we converted the ERFs to the combined planar gradient. The main advantage of this approach is that the largest planar gradient is observed directly above the source (Hämäläinen et al. 1993; Knoeche and Bastiaansen 2000). Figure 5.2c shows the difference in the planar gradient for ERFs when subtracting standards from targets, also aligned in time to the divergence point. The sensors with the largest differences clustered over LH and RH temporal areas, again pointing to a source in each hemisphere. The left source was stronger than the right source. We selected five channels from the region of the largest MMNm and the five corresponding channels over the RH (marked in Figure 5.2c).
Figure 5.2: The event related fields (ERFs) from the axial gradient when comparing standard to targets. (a) The grand average of the ERFs for the standard (black) and target (grey) stimuli aligned to the divergence point. Each plot represents the field
of a sensor arranged topographically according to the sensor position in the helmet. The two dipolar field patterns over the left and right hemisphere are indicated by the circles. Note the stronger deflection for targets compared to standards. (b) The grand average of the ERFs for the difference between standards and targets based on the sensors circled in (a), computed as the sum of the activity registered in the circles L- and R- minus the sum of the activity in the circles L+ and R+. A large difference is observed around 200 ms. (c) A topographic plot of the combined planar gradient of the ERFs for the difference between target and standard stimuli in the interval 0-500 ms after the divergence point. The sensors selected for the subsequent statistical analysis (MLT/MRT 13, 23, 24, 25, 33) are marked with *. Figure 5.3 shows the topographic plots of the difference between targets and standards in the three conditions (lexical, intonational and combined contrast) in the 150-250 ms window for the combined planar gradient. It appears that the left hemisphere is dominant in all conditions and in both groups, except for the intonational contrasts in the Roermond group. The difference between targets and standards is clearly larger in intonational and combined contrasts than in lexical contrasts.

<table>
<thead>
<tr>
<th></th>
<th>Lexical</th>
<th>Intonational</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roermond</td>
<td><img src="image1" alt="roermond_lexical" /></td>
<td><img src="image2" alt="roermond_intonational" /></td>
<td><img src="image3" alt="roermond_combined" /></td>
</tr>
<tr>
<td>Control</td>
<td><img src="image4" alt="control_lexical" /></td>
<td><img src="image5" alt="control_intonational" /></td>
<td><img src="image6" alt="control_combined" /></td>
</tr>
</tbody>
</table>

Figure 5.3: Topographic plots of the difference in magnetic fields between target and standard stimuli in the interval 150-250 ms after point of divergence, in the 3 conditions and 2 language groups. The maps represent the combined planar gradient.

In order to put these observations to the test, we averaged the magnitudes of the combined planar gradients from the selected sensors (one average per hemisphere
and per stimulus type: standard or target) and time window (150-250 ms). First of all, we checked whether the difference between targets and standards was significant in all conditions, including the lexical contrasts. This was indeed the case (p\textsubscript{lexical}=.001, p\textsubscript{intonational}<.001, p\textsubscript{combined}=.003). We then submitted the differences between standards and targets to a repeated-measure ANOVA using the within-subject factors CONTRAST (3 levels: lexical, intonational, combined) and HEMISPHERE (2 levels: left and right), and the between-subjects factor LANGUAGE\_GROUP (2 levels: Roermond and control). Results were adjusted using the Greenhouse-Geisser correction. This omnibus ANOVA showed a significant main effect of CONTRAST (F(1.5,20.9)=5.17, p=.022), as well as a significant interaction between LANGUAGE\_GROUP, CONTRAST and HEMISPHERE (F(1.7,23.8)=4.67, p=.024, for 3 contrasts). Further analyses per contrast revealed no significant effect or interaction of effects in the combined condition. We therefore excluded this condition from the analysis, focussing on the difference between lexical and intonational contrasts. A new ANOVA was performed, with the same factors as above but only two levels for CONTRAST, which again highlighted the effect of CONTRAST (F(1,14)=12.42, p=.003) and the LANGUAGE\_GROUP\_CONTRAST\_HEMISPHERE interaction (F(1,14)=10.93, p=.005).

A possible interpretation of this interaction, which would match the observations made on the topographic plots, is that language experience determines whether the hemispheric specialization of prosodic processing depends on the type of contrast. More precisely, the interaction CONTRAST\_HEMISPHERE should be significant for the Roermond group but not for the control group, reflecting the particular Roermond lateralization pattern for intonational contrasts. This was tested with the help of separate ANOVAs per language group (with the factors CONTRAST and HEMISPHERE), and within each language group, separate ANOVAs per contrast (with the factor HEMISPHERE). In the Roermond group, the ANOVA using CONTRAST (2 levels) and HEMISPHERE highlighted the expected effect of the contrast on lateralization, with a significant CONTRAST\_HEMISPHERE interaction (F(1,7)=9.93, p=.016). In the control group, there was only a significant main effect for CONTRAST (F(1,7)=11.98, p=.011), reflecting an overall stronger MMNm in intonational as compared to lexical contrasts (1.99 fT/m against 1.36 fT/m). The analyses per language group and per hemisphere revealed that this contrast effect was only significant in the LH for the control group (F(1,7)=12.64, p=.009), and in the RH for the Roermond group (F(1,7)=12.45, p=.010). Finally, the analyses per language group and per contrast, which allowed us to isolate the HEMISPHERE factor, showed lateralization effects in the Roermond group only (F(1,7)=6.32, p=.040 in lexical contrasts; F(1,7)=6.26, p=.041 in intonational contrasts).

In sum, the topographic plots as well as the statistical analysis of our data reveal clearer and more differentiated activation patterns in the Roermond group than in the control group. In particular, unlike the non-native speakers, Roermond subjects process intonational patterns predominantly in the right hemisphere. Figure 5.4 gives
5.4 Discussion

In our MEG study of lexical-tonal and intonational processing we found a clear MMNm in a time window from 150 to 250 ms after the divergence point of standard and deviant pitch contours. The most remarkable finding was a clear difference in lateralization between the native Roermond speakers and a non-native control group. While non-native speakers showed a stronger MMNm over the left temporal cortex in all conditions, the native Roermond group showed a stronger response over the left temporal cortex for lexical contrasts, but a predominantly RH response for the intonational contrasts. This finding is suggestive of the possibility that the native Roermond speakers processed lexical contrasts predominantly in the LH, but intonational contrasts predominantly in the RH.

The differential distribution of brain activation in the Roermond group compared to a more uniform pattern in the control group can be seen as an example of the functional plasticity of the brain (cf Hagoort, Brown, & Wassenaar 2003). Temporal cortices are recruitable bilaterally for pitch processing, and lateralization patterns emerge during language acquisition according to the options available in the grammar. Our results cannot be explained on the basis of the phonetic features in which standards and deviants differed, since the same features (namely pitch, enhanced with duration) were used for encoding lexical and intonational contrasts. Moreover, if activation patterns solely depended on acoustic cues, they would be...
identical in native and non-native subjects, which was shown not to be the case. Our results must therefore be due to the linguistic functions that the phonetic features encoded.

The finding that Roermond subjects process lexical contrasts predominantly in the LH temporal cortex is consistent with earlier results on phonological and lexical processing. The shift towards the RH in reaction to intonational contrasts is similar to that reported by Imaizumi et al. (1998), who found that Japanese lexical segmental contrasts triggered LH activity without any significant RH activity, while bilateral activity was found for intonational contrasts. The Roermond RH dominance during intonational processing also agrees with earlier work (Gandour et al. 2003; Friederici and Alter 2004; Meyer et al. 2004). However, these earlier findings do not seem compatible with the general LH dominance found in the control group. Since tonal contrasts are absent from Standard Dutch (and therefore from the speech of the non-native participants), pitch contour differences are assumed to be interpreted as intonational or affective differences, which are both traditionally assumed to be right-lateralized (Pell 2002; Buchanan et al. 2000). Rather than trying to reconcile our results with findings based on other languages or postulating different universal lateralization patterns in pitch processing, it seems more appropriate to acknowledge that there is in fact no universal brain region for pitch processing (Baum & Pell 1999), and we must therefore not expect the processing of a linguistic function like intonation to be allocated to the same part of the brain by speakers of different languages. Instead, we can assume that the greater complexity of the linguistic system that was acquired by the Roermond speakers, who learned a lexical tone distinction that combines in intricate ways with an intonational distinction, led to a different topography for the processing of pitch contrasts from that developed by the Standard Dutch subjects, who acquired a system with intonation contrasts only.

Our results confirm the conclusion reached by Gandour (2007) that studies of brain activation involving tone and intonation show ‘a mosaic of multiple regional asymmetries that allows for different regions being differentially weighted depending on language experience, stimulus properties, and cognitive processes evoked by task demands’. It is to be expected that combining spatially as well as temporally sensitive registrations of stimuli in subjects with a greater variety of language backgrounds will not just confirm the functional plasticity of the brain, but reveal in more detail in what ways different regions in the brain collaborate in the processing of prosodic contrasts.

The sensitivity of MMNm to language background had led us to formulate the hypothesis that our stimuli would trigger a stronger MMNm in the native than in the non-native group. Such a difference could not be established in the present data, presumably because the stimuli were fully compatible with Standard Dutch phonological forms. Haas [ha:s] is the Standard Dutch word for ‘hare’, while vaer is well-formed phonologically, even if not representing a word. More importantly, the four pitch patterns are readily interpretable as Dutch intonation contours. The contour for the Accent 1 statement represents a neutral, declarative intonation, that for the Accent 1 interrogative represents an emphatic declarative. The falling-rising
contour for the Accent 2 statement is interpretable either as a non-emphatic question or as a polite suggestion, while that for the Accent 2 question is an interrogative intonation. Because all four contours are phonologically (and semantically) distinct in Standard Dutch, the MMNm amplitudes were as strong as the ones found in the native group. We did find, however, significantly different MMNm amplitudes due to phonetic differences in the stimuli. As can be seen in Figure 5.3, the Roermond interrogative contours have a wider pitch span than the declarative contours, making the difference between statements and questions always larger than the one between Accent 1 and Accent 2, where pitch spans are more similar. According to Näätänen (2001), an increase in acoustic differences between simple tones can cause an increase in MMNm amplitude. It is reasonable to assume that this also holds for speech stimuli.

In conclusion, our MMNm study supports the view of function-driven, language-dependent pitch processing. It also highlights a clear discrepancy between linguistic and cerebral representations: What is considered universal amongst languages, such as the expression and recognition of discourse meanings by means of intonation, is not necessarily realized in an identical way in the human brain.
6 Summary and conclusions

The present study investigated, by means of four experiments, the perception of the tone contrast observed in the Dutch dialects of Roermond and Venlo. These dialects are characterized by an opposition between two tones, Accent 1 and Accent 2. What makes them different from typical tone languages like Chinese or Yoruba, besides the number and distribution of contrasting tones, is that the shape of Accent 1 and Accent 2 varies drastically as a function of the information status and the position of the Accent 1 or Accent 2 word in the sentence. An additional factor that conditions the contrast is the discourse meaning used in the sentence. For instance, in the dialect of Roermond, a focused Accent 1 word will be pronounced as a sharp fall at the end of a declarative sentence, but as a rise-fall when it concludes a yes/no question. In the same position, Accent 2 is realized with a fall-rise (statement) or a full rise (question). In order to characterize the distinction between the two East Limburgian tones in terms of identification and discrimination performance, it was crucial to take these structural criteria into consideration. This was done by constructing, for each dialect, a set of stimuli in which the variables of information status, position in the sentence and intonational contour (as determined by the discourse meaning) were varied systematically in combination with the two tones. In the first three experiments described below, the minimal pairs that were selected in this way represented the morphological opposition between singular (Accent 2) and plural (Accent 1) forms of a number of nouns. The fact that the tone contrast is used in this way made it possible to give subjects the simple task of saying whether a form was a singular or a plural throughout the experiments. In the fourth experiment, which was different in its aims and set up, we recorded stimuli that differed in their lexical meaning (e.g. *haas' = 'hare*, *haas'' = 'glove').

The first experiment was aimed at evaluating the identification of the Roermond tones as a function of two of the criteria mentioned above, information status (focused/nonfocused) and position in the sentence (final/nonfinal). In particular, we wanted to establish whether a specific combination of the structural variation factors, the case of nonfocused, nonfinal words (recorded in two versions: as prenuclear and postnuclear target words), was responsible for a neutralization of the tonal contrast. This had been demonstrated in Gussenhoven’s production-based account of East Limburgian dialects, but we needed to assess whether it was reflected on the perception side. A second objective of our experiment was to find out the extent to which the presence of the sentence context in the stimuli contributed to the recognition of Accent 1 and Accent 2. These questions were addressed using a cross-dialectal point of view. We conducted two tests in parallel involving the Roermond dialect on the one hand, and the neighbouring dialect of
Weert on the other. The motivation for comparing these two dialects was that in some aspects (e.g. vocabulary and syntax), they display a great deal of overlap, even to the extent of having very similar singular and plural forms of a large number of nouns. However, they differ in an important detail. While Roermond realizes the word prosodic contrast involved in the marking of grammatical number by means of tone, it had been claimed by Heijmans that the Weert dialect realizes it by means of a length opposition. We thus recorded etymologically related minimal pairs in both dialects, in which the prosodic contrast marked an opposition in grammatical number. In Roermond, this opposition was encoded with Accent 1 (plural) and Accent 2 (singular), and in Weert, with short (plural) and long syllables (singular). Acoustic analysis of the stimuli showed for Roermond that the contour shapes corresponded to Gussenhoven’s previous descriptions. In addition, we found that Accent 2 words tended to be slightly longer than their Accent 1 counterpart, but there was a great deal of overlap between the two tones, with about 30% of the Accent 1 instances being longer than Accent 2. The Weert data were, by contrast, characterized by a consistent duration difference between Accent 1 (short) and Accent 2 (long), but we found no systematic pitch-based difference in the minimal pairs. The Roermond and Weert stimuli were randomized and presented for identification to native listeners of the respective dialects. In order to assess the importance of the sentence context for recognition, we presented, in a separate session, the stimuli excised from their carrier sentences. Results showed high recognition scores (about 80%) in all contexts for the Weert group, irrespective of the presentation mode (with or without carrier sentence). This regularity in the scores suggests, first, that the prosodic context does not affect the perception of the Weert contrasts, and second, that these contrasts are located within the target word, rather than being spread across the sentence as a whole. By contrast, the perception of Roermond contrast varied considerably across prosodic contexts, and also depended on the presentation mode. When the stimuli were presented with their carrier sentence, scores were very high for focused contexts (89%) and in [–focus, +final] position (79%), but they dropped to 57% in [–focus, –final] position. The analysis per tone revealed that in [–focus, –final] contexts, scores for Accent 2 were significantly higher than those for Accent 1. Since this pattern was radically different from the one observed in focused and/or final contexts, and since our production data did not exhibit any clear difference between Accent 1 and Accent 2, we concluded that the opposition in the [–focus, –final] contexts was completely neutralized. Moreover, it was observed that this difference increased dramatically when the stimuli were presented excised from their context. This was also the case, albeit to a lesser extent, in the [–focus, +final] context.

We identified two possible explanations for the bias towards Accent 2 (or singular) forms. First, it could be due to the fact that singular forms are somehow more representative of the words in question. As listeners were less certain of the correct answer, they may have increasingly opted for the morphologically unmarked term of the number opposition. Alternatively, we may postulate a phonological explanation, stating that the flat contour observed in this context may be more readily interpreted as an instance of Accent 2, which often exhibits, across contexts, smaller pitch excursions than Accent 1. In the absence of further data, it is
difficult to choose between these explanations. We may in any case conclude that in Roermond, but not in Weert, the contrast highly depends on the intonational context. The fact that in the Roermond dialect lexical and intonational tones are integrated in the same phonological grammar thus turns out to have significant consequences for the functionality of the word prosodic contrast.

The second experiment represents an extension of the first, in that we added one dimension in the factors responsible for contour shape variation. In addition to information status and position in the sentence, the intonation contour (statement and yes/no question) was included as a variable in our perception test. This allowed us to present a complete set of contrasts to native listeners. Results were, on the whole, in line with the ones obtained in our first experiment, with high scores for focus and/or final contexts and a poorer recognition in [–focus, –final] contexts. However, there were departures from this general pattern. In particular, the bias found for Accent 2 (singular) was not verified in a prenuclear, declarative context when the stimuli were pronounced by one of our speakers. Instead, we observed high scores for both members of the minimal pairs. Acoustic analysis of this speaker’s utterances revealed that his Accent 1 (but not the Accent 2) contours in prenuclear position were realized with a slight rise. We performed a small-scale production and perception experiment which suggested that although young native speakers of the Roermond dialect do not realize nonfocused prenuclear Accent 1 words with a rising contour, they do use this rise as a cue for their identification. These results are likely due to the fact that our speaker mistakenly added a pitch accent in prenuclear position. We have no knowledge of how pitch accents preceding the last one are realized, since our data as well as the data collected for earlier production studies were not designed to elicit more than one pitch accent per stimulus sentence. We may, however, hypothesize that a prenuclear pitch accent is characterized by a rise, which would explain why listeners correctly identified the Accent 1 instances: they use prenuclear pitch accent themselves, only not in the context that was given in our experimental sentences. A second exception to the usual score patterns in focus and/or final vs. [–focus, –final] cases was found in one of the [+focus] contexts (viz. nonfinal words with interrogative intonation). In this context, we observed unusually low scores for Accent 2 words. This poor recognition was explained by the contours of one speaker, who systematically reduced the contrast between a rise (Accent 1) and a plateau (usually observed for Accent 2) to an opposition between a rise and a slight rise. A follow-up test confirmed that this speaker’s tendency to reduce the salience of the contrast was in fact quite common among younger speakers. These results support the claim that contrasts are more prone to be lost in contexts where they are not easily perceivable (Ohala 1981, Hume and Johnson 2001). This claim is also likely to explain why the tone contrast is neutralized in [–focus, –final] contexts, which do not favour a careful articulation of the Accent 1 and Accent 2 words.

The finding that the contrast in [+focus, –final] interrogative contexts is on its way out brings up the question of how perceptual salience is to be defined. If we assume that the poor results in the [+focus, –final] interrogative context were merely due to smaller $f_0$ excursions between the $f_0$ contours, we may wonder why
scores were better for [–focus, +final], declarative words, in which the difference between Accent 1 and Accent 2 is clearly smaller according to this criterion: In this context, the tone contrast opposes a slight fall and a slight rise, and at each point of the contour, the $f_0$ differences between Accent 1 and Accent 2 are smaller than in the interrogative case mentioned above. It may be the case that the listeners paid more attention to the direction of the contour than to actual $f_0$ excursions, considering the sheer presence or absence of a rise as the main cue at play for the identification of the stimuli. As we can see, the issue of perceptual salience is far from being straightforward and deserves to be addressed more directly.

This was done in the following experiment, which was designed to systematically compare recognition and perceptual salience (as expressed in measures of phonetic distance) in the dialect of Venlo. The city of Venlo is situated close to the boundary between tonal (Limburg) and non-tonal dialects. Perhaps for this reason, the intonational system used in Venlo is more complex than the one in Roermond. It may have integrated intonational contours from neighbouring dialects into its tonal system. Instead of the two different intonational contours found in Roermond, viz. statement and yes/no question intonation, the Venlo dialect has no less than four contours: statement, low question, high question and continuation. Combining this variable with information status and position in the sentence, we disposed of a broad assortment of tonal minimal pairs that allowed us to test for the relation between recognition and perceptual salience. In a first step, we carried out a recognition test very similar to the ones reported above, except that the minimal pairs representing words in the two [–focus, –final] contexts (pre- and postnuclear) were not included in our set of stimuli. Since the Accent 1 and Accent 2 words in these contexts showed no difference on the production side, we assumed that the contrast was neutralized in the same way as in the Roermond dialect. We therefore restricted ourselves to the study of contexts in which a contrast was postulated. Results of the recognition test in these (focused and/or final) cases showed that the scores varied considerably across contexts. Recognition was better in the ‘declarative’ and ‘low interrogative’ intonations than in the ‘continuative’ and ‘high interrogative’ intonations, and the contrast was better recognized in focused than in unfocused syllables. Interestingly, the recognition performance appeared to depend on the age of the subjects, the older subjects showing significantly better scores than the younger subjects. Such a difference was not found in the Roermond group. It was also observed that on the whole, scores were about 20% higher in the Roermond than in the Venlo group. These two observations suggest that the tone contrast is more threatened in the Venlo than in the Roermond dialect. We were interested to find out whether such an erosion of the contrast could be explained by its lack of perceptual salience.

We therefore collected, in a second step, different measures of phonetic distance, which were to be correlated with each other and with the recognition scores. On the one hand, we computed objective distance measures between members of tonal minimal pairs in each prosodic context. Two different measures assessed the pitch-based differences between the experimental words, RMSE (which directly expresses the difference between pitch values at each time point)
and cosine distance (which takes the direction of the contour into account), and one measure stated durational differences, since these appeared to be quite large in some contexts. On the other hand, we presented the minimal word pairs, intermixed with pairs of identical stimuli, to native listeners as well as to a group of (non-Limburgian) Dutch listeners. The subjects were asked to rate the perceived acoustic distance in each word pair on a scale from 0 (no difference) to 10 (very large difference). Contrary to our expectations, both language groups displayed very similar distance scores, indicating that at least in this experimental condition, listeners were not influenced in their phonetic judgements by the phonological status of the difference in their language. A second remark to be made on the distance scores, is that they shared two context-related characteristics with the recognition scores: They were higher in [+focus, +final] contexts, and in the ‘declarative’ and ‘low question’ intonations. The same can be said for the objective distance scores related to pitch, RMSE and cosine distance; in the case of duration, the difference between [+focus, +final] and [–focus, +final] was substantially reduced, as was the variation across intonations. These tendencies may have been partially responsible for the significant correlations found between recognition scores and all the distance measures except for duration. It should be noted, however, that the correlation coefficients were rather low (between 0.32 and 0.46). Also, the acoustic difference measures had a moderate correlation with the subjective distance measure. The closest match was found between subjective distance and RMSE. In conclusion, our data provide weak support that phonetic salience determines the perceivability and hence the robustness of a phonological contrast, and phonetic salience judgements are more complex than can be captured by the measures we used. This is not too surprising if we consider the general difference between automatic and human speech recognition; it often happens that a speech signal which is not correctly interpreted by a computer device will still be easily identified by human ear, thanks to the capacity of listeners to abstract away from irregularities of the signal (e.g. background noise or small $f_0$ fluctuations) and prioritize the relevant cues. What is more surprising, however, is that we found no clear difference between phonetic distance judgements of the two language groups. We may have expected that the phonological status of the stimuli would have more impact on the native vs. non-native judgements. For instance, duration could have been judged more important by Standard Dutch speakers since in their system, it represents a phonological contrast at the word level whereas $f_0$ does not. Moreover, we saw that duration was not as important as we predicted to the native recognition of the tone contrast. The fact that the acoustic cues were treated in about the same way by our judges is reminiscent of a study on Chinese by Hallé, Chang & Best (2004) which compared the judgement of native (Taiwanese) and non-native (French) speakers on Taiwanese tones. Results suggested that in spite of their non-tonal background, French subjects were as sensitive to pitch differences as Taiwanese were; what distinguished the two groups was found at a deeper cognitive level, in that Taiwanese showed a quasi-categorical mode of processing while French perception was psychoacoustically based.

Our last experiment was an attempt at understanding the mechanisms of tone perception from a neurocognitive perspective, again by comparing native and
non-native reactions to pitch-based contrasts. Earlier research has not been able to establish with certainty how such contrasts are processed in the brain. In particular, although it is well-known that the left cerebral hemisphere is associated with language processing, there is no consensus on which hemisphere is used for the processing of pitch-based linguistic distinctions such as tonal and intonational contrasts. The lack of agreement in the results is in part due to the fact that the findings are not always based on the same language, and often not even on the same type of language, when results with listeners with and without a tone language background are compared. In the few studies that did test tone and intonation in parallel, the language used was Chinese, which is not unproblematic with respect to the linguistic encoding of the two functions. Whereas lexical tone contrasts are encoded phonologically, it may be argued that intonational contrasts are based on phonetic differences such as pitch range variations.

We succeeded in controlling the phonological status of intonation contrasts and lexical tone contrasts by using the dialect of Roermond, which encodes both contrasts phonologically. An MEG experiment was carried out in order to compare the processing of the Roermond lexical-tonal and intonational contrasts. A set of words with identical phoneme sequences but distinct pitch contours, which represented different lexical meanings or discourse meanings (statement vs. question), were presented to native speakers as well as to a control group of speakers of Standard Dutch. The stimuli were arranged in a mismatch paradigm, which involves the presentation of sequences of frequent, (near-) identical sounds interrupted by infrequent, deviant sounds. This design triggers a brain reaction called mismatch negativity (MMN) which has interesting properties. In particular, if a contrast represents the opposition between two phonological categories, the MMN effect will be stronger than if it is purely acoustic. The mismatch paradigm was used under three experimental conditions: in the first condition (lexical), the pitch contour differences between standard and deviant stimuli reflected differences between lexical meanings; in the second condition (intonational), the stimuli differed in their discourse meaning; in the third condition (combined), they differed both in their lexical and discourse meaning. In all three conditions, native as well as non-native responses showed a clear MMNm (magnetic mismatch negativity) in a time window from 150 to 250 ms after the divergence point of standard and deviant pitch contours. In the lexical condition, a stronger response was found over the left temporal cortex of native as well as non-native speakers. In the intonational condition, the same activation pattern was observed in the control group, but not in the group of native speakers, who showed a right-hemisphere dominance instead. Finally, in the combined (lexical and intonational) condition, brain reactions appeared to represent the summation of the patterns found in the other two conditions. In sum, the lateralization of pitch processing is condition-dependent in the native group only, which suggests that language experience determines how processes should be distributed over both temporal cortices, according to the functions available in the grammar.

The main conclusions that can be drawn from these studies are, first, that unlike segmental and quantity contrasts, the recognition of the lexical tone contrast
in East Limburgian dialects depends on the prosodic context in which it appears. In favourable contexts, the tone contrast is perceived just as well as phonological contrasts on the segmental level. On the other hand, in both the Venlo and the Roermond dialect, it is neutralized in unaccented positions, except when it occurs in the final syllable of the phrase. The second important conclusion is that there are a number of indications that the tone contrast in the East Limburgian dialects is showing signs of disappearing, or at least of becoming frayed at the edges. The neutralization in the [+focus, –final] contexts may be followed in Roermond by the loss of the [+focus, –final] interrogative contrast. Younger speakers were shown to reduce the contrast between a rise and a flat contour to a rise and a slight rise, which affects the perceivability of the contrast and eventually, is very likely to cause its disappearance. Venlo has less robust recognition scores than Roermond. This may reflect the fact that it lies closer to the periphery of the tonal area. The third conclusion is that it is hard to establish a connection between salience and robustness experimentally. We did find some indications, but we feel more progress must first be made in defining phonetic salience. The fourth conclusion is that because of their experience with both lexical tone and intonation in their native language, speakers of tone languages like the dialect of Roermond have a more complex brain response pattern than speakers of standard Dutch, whose perceptual strategies may be more straightforward.

Future research should pursue the analysis of tonal dialects, many of which are still undescribed: Kerkrade, Vaals, Eupen, St Vith, to mention just some. More typological data on these grammars will provide us with a deeper understanding on how lexical tone works, and along which dimensions neighbouring dialects differ from each other. An important issue that would deserve more attention is the exact contribution of the sentence context for recognition, this time with a more phonetically-based approach. We saw in our first experiment that in some cases, the recognition of tone seems to benefit from the information stored in the sentence surrounding the target word. It would be useful to know why this is the case; in this context, production as well as perception studies are necessary.

More generally, research on tone would benefit considerably from more perception studies, not only in order to complete production studies but to refine the methods used so far and address issues that are specific to this side of communication. For instance, we may want to conduct experiments with manipulated $f_0$ and duration in order to establish the weighting of cues for recognition on the one hand, and phonetic distance judgements on the other. In the context of brain research, we may also want to control more for the acoustic cues at play in the tonal contrast. Also, it would be useful to better separate the lexical from the tonal effects, by conducting experiments which directly compare segmentally-based and pitch-based lexical contrasts. On another level of analysis, we may want to explore the possibilities of brain research for providing more insight on the possible categorical perception of tone.
References


Hanssen, J. (2005): Tone and intonation in the dialect of Sittard. MA Radboud University Nijmegen, the Netherlands.


Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). Psychophysiology, 38, 1-21.


International Conference on Spoken Language Processing (ICSLP 2000), 1, B1-B12.


In dit proefschrift wordt, aan de hand van vier experimenten, de waarneming van het tooncontrast in de Oost-Limburgse dialecten van Roermond en Venlo onderzocht. Deze dialecten worden gekenmerkt door een oppositie tussen twee tonen, Accent 1 en Accent 2. De toonoppositie kan betekenisverschillen met zich meebrengen, vergelijkbaar met het contrast tussen klinkers of medeklinkers. Als de /b/ in het Nederlandse woord *baas* door een /k/ vervangen wordt, verandert ook de betekenis van het woord (*kaas*). Het Roermonds en het Venloos kennen nog een andere manier om betekenisverschillen uit te drukken. Sommige woorden, zoals *haas*, kunnen verschillende betekenis hebben, afhankelijk van de melodie waarmee ze uitgesproken worden: met Accent 1 is *haas* een langoor, en met Accent 2, een handschoen; met Accent 1 betekent *kniën* ‘konijnen’ (meervoud) en met Accent 2, ‘konijn’ (enkelvoud). Er zijn veel andere talen, vooral in Azië (bv. het Chinees) en Afrika (bv. het Yoruba), die zulke toon-gebaseerde contrasten gebruiken. Een belangrijk verschil tussen deze talen en dialecten zoals het Roermonds en Venloos is het aantal en de verdeling van deze tonen: terwijl er in meer typische toontalen vaak minstens vier verschillende tonen zijn die vrijwel op elke lettergreep kunnen voorkomen, is het tooncontrast tussen de twee Limburgse tonen beperkt tot enkele lettergrepen in de zin.

Een ander, en voor onze studie nog belangrijker verschil, is dat de uitspraak van (Roermondse of Venlose) Accent 1- en Accent 2-woorden drastisch varieert als functie van focus (of een woord benadrukt wordt of niet) en positie in de zin (finaal of niet finaal). Bovendien is de melodie van Accent 1 of 2 afhankelijk van de zinsintonatie. In het Roermonds wordt, bijvoorbeeld, een beklemtoond Accent 2-woord met een lichte stijging uitgesproken als het binnen een mededeling voorkomt, maar als het een zin beëindigt, hoort men een dalende gevolgd door een stijging. Als het beklemtoonde Accent 2-woord binnen een vraag uitgesproken wordt, is zijn melodie weer anders, namelijk een vlakke lage toon. Voor Accent 1 zijn er in deze prosodische contexten natuurlijk evenveel verschillende melodiecontouren. Het contrast tussen Accent 1 en Accent 2 neemt dus veel vormen aan, en aangezien de twee contouren met elkaar ook verschillende melodiecontouren kunnen karakteriseren, hebben we perceptie-experimenten uitgevoerd waarin de bovengenoemde factoren (focus, positie in de zin en zinsintonatie) systematisch varieerden in combinatie met de twee tonen. Deelnemers luisteren naar zinnen waarin een woord met Accent 1 of Accent 2 beklemtoond en finaal was in een mededeling (b.v. *Ik zeg “KONLIN”*), verder naar mededelingen waarin hetzelfde woord niet beklemtoond maar wel finaal was (b.v. *Ik ZEI niet “konijn”, maar ik ZONG “konijn”*), maar bijvoorbeeld ook naar vragen waarin het woord beklemtoond en finaal was (b.v. *Zei je “KONLIN”?*). In onze eerste drie experimenten stelden de gebruikte Accent 1/Accent 2-woordparen het verschil
tussen enkel- en meervoud voor. De proefpersonen hoorden dus twee versies van elke zin: een keer droeg knien, het woord voor ‘konijn’, de melodie voor Accent 1 (‘konijnen’), en een keer droeg knien Accent 2 (‘konijn’). Na elke zin moesten de deelnemers zeggen of ze het woord knien in het enkelvoud of meervoud hadden gehoord. Aan de hand van het percentage correcte antwoorden die ze daarbij gaven, konden we vaststellen of ze de twee versies goed van elkaar konden onderscheiden. In het derde experiment hebben we deze vraag ook rechtstreeks aan de proefpersonen gesteld: ze hoorden woordparen en hun taak was om zelf de grootte van het verschil te schatten. In het vierde experiment, dat andere doelen en opzet had dan de eerste drie, hebben we woordparen gebruikt waarbij de tegenstelling tussen Accent 1 en 2 geen getalsverschil uitdrukt maar een verschil in semantische betekenis, zoals in het bovengenoemde woordpaar haasI (‘langoor’) en haasII (‘handschoen’). Hieronder worden de vier experimenten in meer detail beschreven.

Het doel van het eerste experiment was om de identificatie van de Roermondse tonen te beoordelen als functie van twee van de drie bovengenoemde criteria, te weten focus en positie in de zin. In het bijzonder wilden we weten of een specifieke combinatie van deze factoren, het geval van niet-beklemtoonde, niet-finale woorden (voor of na het zinsaccent), verantwoordelijk was voor een neutralisatie van het tooncontrast. Dit was door Gussenhoven al aangetoond aan de hand van akoestische studies van het Roermonds en het Venloos, maar we moesten zijn resultaten nog bevestigen met beoordelingen van moedertalige luisteraars. Een tweede doel van ons experiment was om te bepalen in hoeverre de aanwezigheid van de zinscontext in de stimuli de herkenning van Accent 1 en Accent 2 vergemakkelijkte. Deze vragen werden vanuit een vergelijkend perspectief behandeld. We voerden twee toetsen in parallel uit, de eerste in Roermond en de tweede in Weert. De reden voor deze vergelijking was dat het Weerts en het Roermonds in sommige aspecten nauw verwant zijn, maar toch genoeg van elkaar verschillen om relevante eigenschappen van het Roermonds duidelijk te kunnen maken. In het bijzonder konden we relatief gemakkelijk een aantal woorden vinden die bijna dezelfde enkelvouds- en meervoudsvormen hebben, op één kenmerk na: terwijl het Roermonds de oppositie tussen enkel- en meervoud door middel van toonhoogtebewegingen realiseert, gebruikt het Weerts duurverschillen. Zo betekent een lange uitspraak van knien ‘konijn’, en een korte uitspraak, ‘konijnen’. We veronderstelden dat de perceptie van het Roermonds tooncontrast gevoeliger zou zijn voor de context dan het Weertse duurcontrast, dat in dit opzicht meer gemeen heeft met segmentale contrasten zoals het verschil tussen /a/ en /e/. Met andere woorden, we verwachtten dat de herkenning van woordparen voor de Roermonds deelnemers eenvoudiger zou zijn wanneer het woord in een prominente positie in de zin stond, en/of wanneer de zinscontext aanwezig was, terwijl deze factoren voor de Weertse luisteraars van minder belang zouden zijn. Om deze veronderstelling na te gaan, namen we in beide dialecten soortgelijke zinnen met doelwoorden in het enkel- en meervoud op, waarin de prosodische contexten systematisch varieerden als functie van focus en positie in de zin. De doelwoorden in de zinnen werden vervolgens door moedertalige luisteraars beoordeeld. Om de invloed van de zinscontext op herkenning te bepalen, voerden we de toets in twee
versies uit: eerst luisterden de proefpersonen naar hele zinnen, en toen alleen naar de doelwoorden die we uit de zinnen geïsoleerd hadden.

De luistertoetsen leverden vrij hoge herkenningsscores in alle contexten voor de Weert-groep, ongeacht de prosodische context (beklemtoond of niet, finaal of niet) en de aanbiedingsvorm (zinnen of uitgeknipte woorden). Daarentegen was de herkenning door Roermondenaren wel afhankelijk van deze twee factoren. Ten eerste waren de scores voor beklemtoonde en/of finale doelwoorden veel beter dan voor onbeklemtoonde, niet-finale woorden, en ten tweede was de invloed van de prosodische context op herkenningsscores veel groter wanneer de proefpersonen alleen de doelwoorden hooriden. De lage scores voor onbeklemtoonde, niet-finale woorden weerspiegelen de neutralisatie van het tooncontrast in deze context. De resultaten vertoonden echter een onverwacht verschil tussen de scores voor Accent 1 en Accent 2 in deze context, met een duidelijke tendens om de doelwoorden als enkelvoudsvormen (Accent 2) op te vatten. Het kan zijn dat luisteraars in geval van twijfel op een standaardwaarde terugvielen, waarbij we aannemen dat het enkelvoud de standaard is. Het is ook mogelijk dat ze Accent 2 dachten te horen omdat de vlakke melodie in onbeklemtoonde, niet-finale woorden meer op andere Accent 2-contouren lijkt dan op de meestal steilere Accent 1-contouren. In ieder geval is het herkenningspatroon duidelijk anders in deze prosodische context dan in de beklemtoonde en/of finale gevallen, wat voor neutralisatie spreekt. Het feit dat de presentatie van de stimuli als uitgeknipte woorden een invloed had op de herkenningsscores suggereert bovendien dat de informatie over het tooncontrast, anders dan het duurcontrast in het Weerts, niet beperkt is tot het doelwoord. Het gebruik van toonhoogteverschillen op zowel woord- als zinsniveau heeft dus belangrijke consequenties voor de werkwijze van het onderzochte prosodische contrast.

Het tweede experiment is een aanvulling op het eerste, waarin we nog de dimensie van de zinsintonatie in de luistertoets integreerden. Naast de “neutrale”, mededelende intonatie werd de vraagintonatie in combinatie met de andere twee variatiefactoren (focus en positie in de zin) getoetst. Daarmee konden we voor het eerst alle verschillende tooncontrasten binnen een dialect tegelijkertijd perceptief evalueren. De resultaten waren, in het algemeen, vergelijkbaar met wat we in het eerste experiment gevonden hadden, met dezelfde voorkeur voor enkelvouds- of Accent 2-vormen zowel in mededelingen als ook in vragen. Een afwijking van dit vertrouwde patroon werd echter gevonden in beklemtoonde, niet-finale woorden met vraagintonatie. Veel Accent 2-doelwoorden werden als meervoudsvormen geïnterpreteerd. De akoestische analyse van de opgenomen stimuli toonde aan dat één van onze sprekers de neiging had om het contrast tussen een stijging (Accent 1) en een vlakke melodie (Accent 2) terug te brengen tot een contrast tussen een steile (Accent 1) en een minder steile (Accent 2) stijging. Aan de hand van extra opnames konden we vaststellen dat de reductie van het contrast niet ongewoon was onder jongere mensen. Het is derhalve zeer goed mogelijk dat het contrast in deze zinscontext aan het verdwijnen is. Dit resultaat ondersteunt de bewering dat contrasten eerder bedreigd worden als ze niet opvallend genoeg zijn. De neutralisatie van het contrast in ongeaccentueerde en niet-finale contexten kan
waarschijnlijk ook daardoor verklaard worden, want in deze contexten heeft men de neiging om woorden minder zorgvuldig uit te spreken. Wat echter nog niet vaststaat, is hoe de opvallendheid van een contrast te definieëren is. Op het eerste gezicht lijkt het dat de slechte herkenningsscores in beklemtoneerde, niet-finale woorden met vraagintonatie zonder meer toe te schrijven zijn aan de kleinere verschillen in toonhoogte tussen Accent 1 en Accent 2. We vonden echter een andere context in onze data waarin de verschillen nog kleiner waren (lichte stijging vs. lichte daling) zonder dat de scores daaronder te lijden hadden. Het kan dus zijn dat de luisteraars niet alleen op de absolute verschillen in toonhoogte letten, maar ook, en misschien vooral, op de richting van een contour. Als de stijging het belangrijkste kenmerk van Accent 1 in een bepaald context is, dan worden in deze context alle stijgingen, zelfs de minste stijgingen, als Accent 1 geïnterpreteerd, waardoor herkenningsfouten kunnen ontstaan.

De vraag naar de exacte bijdrage van toonhoogte en richting van de contour is blijkbaar nog lang niet opgehelderd en verdient het om directer getoetst te worden. Dit hebben we in het derde experiment gedaan, dat de Venlose tooncontrasten gebruikte om de herkenning en de fonetische opvallendheid systematisch te vergelijken. Het Venloos heeft hetzelfde contrast tussen Accent 1 en Accent 2 als het Roermonds, maar het heeft meer intonatiecontouren. Naast de neutrale intonatie (mededeling) worden er twee verschillende vraagintonaties (lage en hoge vraagintonatie) en een komma-intonatie (zoals in Eerst doe ik dit, dan dat, en dan...) gebruikt. De combinatie van de twee lexicale tonen met deze melodieën en de verschillende zinscontexten levert een groot aantal contouren op, en daarmee groeit ook de variatie in de fonetische opvallendheid van het tooncontrast. We maakten opnames van alle mogelijke combinaties van deze factoren en legden onze verzameling van zinnen voor aan moedertalige luisteraars. Net als in de vorige experimenten vroegen wij hen of het doelwoord in elke zin een enkel- of meervoudsvorm voorstelde. De resulterende scores toonden opnieuw een grote invloed van de zinscontext voor de herkenning van de doelwoorden. De scores varieerderen per intonatiecontour (bv. veel betere scores voor woorden in neutrale zinnen dan in zinnen met kommaintonatie), en waren over het algemeen beter voor beklemtoneende dan voor onbeklemtoneerde woorden. Bovendien waren de prestaties van oudere proefpersonen duidelijk beter dan die van jongere proefpersonen, een verschil dat we in Roermond niet hadden gevonden. In het algemeen was het percentage correcte antwoorden in Roermond ook hoger dan in Venlo. Deze twee verschillen tussen de dialecten interpreteerden we als teken van een snellere erosie van het tooncontrast in Venlo dan in Roermond. Dit zou kunnen samenhangen met het feit dat de stad Venlo aan de grens tussen het tonale en het niet-tonale gebied ligt. We wilden ook weten of er een verband was tussen de mogelijke verwijding van het contrast en zijn fonetische opvallendheid.

Om een maat van fonetische opvallendheid te verkrijgen, vroegen we twee groepen van proefpersonen, namelijk moedertaligen van het Venloos en Nederlanders die dit dialect niet kenden, om de afstanden binnen alle mogelijke Accent 1/Accent 2-paren op een schaal van 0 (geen verschil) tot 9 (heel groot verschil) te plaatsen. Aangezien we geen significant verschil tussen de twee groepen vonden, werden de
gemiddelde verschilscores vergeleken met de scores van het herkenningsexperiment. De analyse van de scores per context toonde overeenkomsten tussen de verschil- en de herkenningsscores, bijvoorbeeld hogere scores voor beklemttoonde, finale woorden dan voor andere woorden. We nemen aan dat deze overeenkomsten dels verantwoordelijk zijn voor de significante correlatie die gevonden werd tussen de herkennings- en de verschilscores. De correlatiecoëfficiënt was echter vrij laag (0.38). De herkennings- en verschilscores werden ook vergeleken met diverse automatisch berekende verschilmaten (RMSE, cosinus-afstanden en duurverschillen). In de analyse per context stelden we vooral overeenkomsten vast tussen de herkenningsscores, de subjectieve verschilscores en de RMSE- en cosinus-scores. Duurverschillen leken minder gerelateerd te zijn aan de andere maten. De berekende correlaties tussen herkenningsscores, subjectieve verschilscores en RMSE- en cosinus-scores waren allemaal significant maar wederom vrij laag: behalve de correlatie tussen RMSE de subjectieve verschilscores (0.65) waren alle coëfficiënten lager dan 0.5. In conclusie kunnen we stellen dat er wel een verband bestaat tussen de fonetische opvallendheid en de herkenbaarheid (en daarmee de duurzaamheid) van het tooncontrast, maar dat dit verband niet heel sterk is. Subjectieve beoordelingen van fonetische opvallendheid zijn bovendien complexer dan wat we met onze automatische maten aan konden tonen. Dit is vergelijkbaar met de moeilijkheden die aangetroffen werden bij de automatisering van spraakherkenningsprocessen. Het gebeurt niet zelden dat een uiting verkeerd geïnterpreteerd wordt door een computerprogramma en toch probleemloos ontcijferd kan worden door het menselijke oor, dankzij de cognitieve vaardigheid van luisteraars om onregelmatigheden van het spraaksignaal (bv. achtergrondruis of kleine \( f_0 \)-fluctuaties) te negeren en alleen de relevante informatie te gebruiken.

In ons laatste experiment hebben we geprobeerd om de mechanismen van tonperceptie vanuit een neurocognitief perspectief te bekijken en beter te begrijpen. Er is al veel onderzoek gedaan om te bepalen hoe \( f_0 \)-gebaseerde contrasten verwerkt worden in het menselijke brein, maar tot nu toe is er weinig overeenstemming gevonden over de geboekte resultaten. In het bijzonder, ofschoon al lang vaststaat dat de linkerhersenhalf geassocieerd is met spraakverwerking, weten we nog niet met zekerheid in welke hersenhelft spraakmelodieën verwerkt worden. Het gebrek aan definitieve resultaten hangt samen met het feit dat spraakmelodieën verschillende linguïstische functies kunnen hebben. Zoals we al zagen, kunnen toonhoogteverschillen contrasten uitdrukken op woordniveau (toon), maar ook op zinsniveau (intonatie, zoals het verschil tussen een vraag en een mededeling). Het wordt vaak aangenomen dat de lateralisatie van toongerelateerde processen afhankelijk is van de functie die de spraakmelodieën bekleden. Als er een bepaalde hersenhelft geïdentificeerd wordt in een studie die zich op de analyse van tooncontrasten beperkt, kunnen de resultaten niet zomaar worden toegespit op intonatiecontrasten. Het is belangrijk om de twee functies van toonhoogteverschillen direct te kunnen vergelijken om een correct beeld te verkrijgen van \( f_0 \)-verwerking. Oost-Limburgse dialecten zijn, wat dat betreft, ideale objecttaalen. Anders dan bijvoorbeeld in het Standaard Nederlands kunnen zowel lexicale als intonationele contrasten aan de hand van fonologisch-gebaseerde \( f_0 \)-
contrasten worden uitgedrukt, wat een eerlijke vergelijking tussen toon- en intonatieverwerking mogelijk maakt. Deze vergelijking werd aan de hand van een MEG- (i.e. magnetoencefalogram-)experiment gemaakt.

Een groep Roermondenaren en een groep Nederlanders zonder ervaring met Limburgse dialecten luisterden naar woorden met dezelfde reeks van klinkers en medeklinkers maar verschillende toonhoogtebewegingen terwijl hun breinreacties met de MEG-apparatuur opgenomen werden. We presenteerden de woorden volgens de mismatch negativity (MMN)-methode, waarbij een reeks van identieke klinkers (standard) af en toe onderbroken wordt door een andere klank (deviant). De ‘verrassing’ die daardoor ontstaat manifesteert zich in de vorm van een breinreactie die interessante eigenschappen heeft. In het bijzonder is de MMN-reactie groter wanneer het verschil tussen standard en deviant een verschil tussen fonologische categorieën represents (bv. /p/ vs. /t/ in het Nederlands) dan wanneer het verschil alleen akoestisch is (bv. /p/ vs. /pʰ/). We gebruikten het MMN-paradigma in drie condities. In de eerste conditie (lexicaal) was de standard een Accent 1-woord en de deviant een Accent 2-woord (of omgekeerd), in de tweede conditie (intonationeel) was de standard een vraag en de deviant een mededeling, en in de derde conditie (gemengd) was de standard bijvoorbeeld een Accent 1-woord als vraag en de deviant een Accent 2-woord als mededeling. We vonden, in de drie condities en de twee groepen van deelnemers, een duidelijke MMN-reactie op de deviants. In de lexicaal conditie was de reactie sterker in de linker- dan in de rechterhersenhelft van alle deelnemers, ongeacht hun moedertaal. In de intonationele conditie was de reactie dezelfde in de Standaard Nederlandse groep, maar niet in de Roermondse groep, die meer hersenactiviteit vertoonde in de rechter- dan in de linkerhelft. In de gemengde conditie vond de reactie bij moedertalige deelnemers zowel links als rechts plaats, en alleen links bij niet-moedertalige deelnemers. Samenvattend kunnen we zeggen dat de verwerking van toon- en intonatiecontrasten afhankelijk is van de taalervaring van de luisteraar. De verdeling van de verwerkingsprocessen over de twee hersenhelften wordt bepaald door de functies die in zijn taalsysteem beschikbaar zijn.

Uit de vier uitgevoerde experimenten trekken we de volgende conclusies. Ten eerste hangt de herkenning van het Oost-Limburgse lexicaal tooncontrast sterk af van de prosodische context waarin het voorkomt. In gunstige contexten wordt het tooncontrast net zo goed waargenomen als een duur- of segmenteel contrast. Het contrast wordt echter, zowel in het Roermonds als in het Venloos, geneutraliseerd in onbeklemtoonde contexten, tenzij het aan het eind van de zin voorkomt. Ten tweede zijn er aanwijzingen dat het tooncontrast in Oost-Limburgse dialecten niet overal even stabiel is. De neutralisatie in de onbeklemtoonde, niet-finale contexten zou kunnen worden gevolgd door de verdwijning van het contrast in vragen, als het woord met Accent 1 of 2 beklemtoond maar niet-finaal is. Het kon worden aangetoond dat jongere sprekers de neiging hebben om in deze context het verschil tussen een vlakke toon en een stijging te reduceren tot een verschil tussen een lichte en een steile stijging. Dit maakt het contrast minder duidelijk en kan als gevolg hebben dat het uiteindelijk helemaal verdwijnt. Bovendien hebben
we vast kunnen stellen dat de herkenning van tooncontrasten in het algemeen minder accuraat is in Venlo dan in Roermond. Een mogelijke verklaring daarvoor is dat de stad Venlo aan de grens tussen het tonale en het niet-tonale gebied ligt. De derde conclusie is dat het moeilijk is om experimenteel een verband vast te stellen tussen de fonetische opvallendheid en de duurzaamheid van een contrast. Onze poging om dat te doen heeft een aantal aanwijzingen opgeleverd, maar om duidelijker resultaten te verkrijgen moet fonetische opvallendheid eerst beter worden gedefinieerd. Ten slotte vertoont de verwerking van spraakmelodieën in het brein een complexer patroon bij sprekers van toontalen zoals het Roermonds dan bij sprekers van het Standaard Nederlands. Dit hangt samen met de ervaring van Roermondenaren met zowel toon- als intonatiecontrasten.
Curriculum vitae

Rachel Fournier was born in Sion, Switzerland, in 1971. In 1991, after a five-year teacher training and a few language trips to Germany, England and Australia, she started studying for a master’s degree at the University of Fribourg (CH). Her major was German Language and Literature, with a minor in Computer Science. The Erasmus program allowed her to spend a year in Germany, studying at the Freie Universität Berlin. She graduated in 1998, with a master’s thesis on the use of prosody in automatic speech recognition. She then worked as a research assistant within the Computational Linguistics group at the University of Zurich, and in parallel gave support and lessons for the use of computer programs at Lloyds Bank. In 2000, she left Switzerland for a trip around Asia, which ended with a Mandarin course in China. She moved to the Netherlands in 2002. Until 2008, she worked at the Radboud University Nijmegen on the PhD project that resulted in this dissertation, as a member of the Chair of General and Experimental Phonology. She currently holds a postdoc position in the same research group.