



FOURNIER, Rachel; GUSSENHOVEN, Carlos. **Measuring phonetic salience and perceptual distinctiveness: the lexical tone contrast of Venlo Dutch.** *Revista Diadorim / Revista de Estudos Linguísticos e Literários do Programa de Pós-Graduação em Letras Vernáculas da Universidade Federal do Rio de Janeiro*. Volume 12, Dezembro 2012. [<http://www.revistadiadorim.lettras.ufrj.br>]

MEASURING PHONETIC SALIENCE AND PERCEPTUAL DISTINCTIVENESS: THE LEXICAL TONE CONTRAST OF VENLO DUTCH¹

Rachel Fournier (Radboud University Nijmegen)

Carlos Gussenhoven (Radboud University Nijmegen)

ABSTRACT: Judgements of the phonetic difference between the two tones of Venlo Limburgish by both native speakers and speakers of standard Dutch were found to correlate weakly with the degree to which the tones were identified by native speakers, across a wide variety of intonational contexts. Of the acoustic measures we took, only f_0 appeared to explain the success with which the tones were identified by native speakers. Even though durational differences were significant, they made no contribution to the identification scores. It proved to be hard to explain the phonetic difference judgements on the basis of acoustic measures. The investigation provides some support for the claim that phonetic salience determines the degree to which phonological contrasts are perceived by native speakers. It also shows that perceived phonetic differences are consistent across different listener groups, but that the acoustic basis for these difference judgements may be hard to define.

RESUMO: O artigo mostra que as diferenças fonéticas entre os dois tons do dialeto Limburguês falado em Venlo, Holanda, percebidas por falantes nativos e por falantes do holandês padrão se correlacionam discretamente com o grau de identificação dos tons pelos falantes nativos, numa grande variedade de contextos entonacionais. Entre as medidas acústicas feitas, apenas a F_0 explica o êxito na identificação dos tons pelos falantes nativos. Ainda que diferenças de duração fossem significativas, elas não contribuiriam para os índices de identificação. Revelou-se difícil explicar julgamentos sobre diferenças fonéticas com base em medidas acústicas. A investigação fornece argumento para a tese de que a sali-

1. A slightly longer version of this text appeared as chapter 4 in the first author's 2008 dissertation *Perception of the tone contrast in East Limburgian dialects*. Utrecht: LOT Publications Nr. 196.

ência fonética determina o grau em que os contrastes fonológicos são percebidos pelos falantes nativos. Mostra, ainda, que as diferenças fonéticas percebidas são consistentes entre os diferentes grupos de ouvintes, mas que suas bases acústicas podem ser difíceis de determinar.

1. Introduction

The shape of the pitch contours in Venlo Dutch (Limburgish) depends not only on which of the two lexical tones occur (Accent 1 or Accent 2), but also on the intonation melody, the presence of an intonational pitch accent and whether the syllable is final or nonfinal in the Intonational Phrase (Gussenhoven & van der Vliet 1999). The dialect is 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 Westfalia, and the larger parts of the Belgian and Dutch provinces of Limburg.

There are some recent indications that the tone contrast is recessive. Our assumption in the investigation reported here is that the perceptual robustness of a contrast determines the path along which the contrasts is on its way out, meaning that a collapse of the system may begin in specific contexts. One indication for this is that the contrast between Accent 1 and Accent 2 is lost in nonfocal, nonfinal contexts (Gussenhoven 2000, Fournier, Verhoeven, Swerts and Gussenhoven 2006). Our aim here is to establish how well the contrasts Venlo Limburgish 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 would appear that to some extent, therefore, phonetic salience must provide the explanation of the structure of phonological

system. Speech perception in adverse conditions, like a noisy environment, will affect acoustically less salient features more than more salient features.

Going against this conclusion is the 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 unlikely to hear the distinction between the early and late falls of Venlo Limburgish declaratives, both of which signal a declarative intonation in the standard language. 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.

1.2. Perceivability

We define the perceivability of a contrast, its robustness, as the average recognition scores of any two forms. 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 isn't, 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 contrast between /h/ and its absence was better perceived by listeners who had it in their language. Peperkamp & Depoux (2002) demonstrate with short term memory tasks 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 2008). 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; Huang 2001) or could better discriminate (Lee 1996, Burnham et al. 1996) 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 decided to investigate phonetic salience from two different angles. First, we explore objective measures of acoustic distance, taking our cue from Hermes (1998). Second, we measure the perceived phonetic difference both with native and non-native listeners. By comparing the objective and subjective distance measures, we will 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 may gain some insight into the extent to which the phonological status of a phonetic difference influences perceived salience. After answering these questions, we will proceed 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, focusing 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 preceivability.

2. Stimuli

All tests described below are based on the same recordings, made by 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.

2.1. The corpus

The carrier sentences used in the perception experiments contained four tonal minimal pairs (*knien*^{III} = ‘rabbit(s)’, *bein*^{III} = ‘leg(s)’, *derm*^{III} = ‘intestine(s)’ and *stein*^{III} = ‘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 reveal their grammatical number in other words, such as in the articles or through verbal concord. The use of metalinguistic sentences allowed us 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 and Gussenhoven 2006).

The carrier sentences are listed in Table I. 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. 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 and van der Vliet (1999). These utterances were not included in any perception tests, but their acoustic properties will be described 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 *bein*, *derm*, *knien* and *stein* sentences, a further set of sentences was read aloud by the speakers, in which the word *beer^l* ('beer') contrasted with *baer^{ll}* ('bear, (ursus)') 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 I, subjects would find it easier to pronounce the sentences. In these practice sentences, which are given in Appendix 1, the word *beer^l* was mostly used in combination with the verb *drinke* ('to drink') and *baer^{ll}* with the verb *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 I 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 Veldeke spelling, a standard orthographic system developed for Limburgish dialects in general (see Bakkes, Crompvoets, Notten and 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 I 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. Table I shows the carrier sentences with *knien* as the experimental item. The utterances were analysed acoustically.

Context	Sentence (using <i>knien</i> 'rabbit(s)')
[+focus,+final] declarative	In 't Venloos zaeste gewoeën "KNIEN" . In the Venlo dialect, you just say "RABBIT(S)".
[+focus,+final] low question	– Zaese gewoeën "KNIEN" ? [– Nae, ik zegk "BEER".] – Do you just say "RABBIT(S)"? – No, I say "BEER".
[+focus,+final] continuation	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)".

[+focus,+final] high question	[– In ‘t Venloos zaeste gewoeën “KNIEN”. – Ech waar?] Zaeste gewoeën “KNIEN”? [...] – Really? Do you just say “rabbit(s)”?
[+focus,– final] declarative	[– Waat hebste gehuurd?] Ik heb “KNIEN” gehuurd. – What did you hear? – I heard “RABBIT(S)”.
[+focus,– final] low question	– Hebse “KNIEN” gezag? [– Nae, ik heb “BEER” gezag.] – Did you say “RABBIT(S)”? – No, I said “BEER”. Ik heb ièrs “KNIEN” gezag, [en toen nog “DERM” gezag, en toen nog
[+focus,–final] continuation	“BEIN” gezag.] First I said “RABBIT(S)”, and then I said “INTESTINE(S)”, and then “LEG(S)”.
[+focus,–final] high question	[– Waat hebse gehuurd? – Ik heb “KNIEN” gehuurd. – Ech waar?] De hebs “KNIEN” gehuurd? What did you hear? – I heard “RABBIT(S)”. – Really? Did you hear “RABBIT(S)”?
[–focus,+final] declarative	Speaker YK: [– Nae,] ik SCHRIEËF “knien”. (see high question) Speaker KB: [– Nae,] ik FLUUSTER “knien”. (see low question) – No, I SHOUT/WHISPER “rabbit(s)”.
[–focus,+final] low question	Speaker YK: [– Ik SCHRIEËF “knien”] – De SCHRIEËFS “knien”? [Ik dach, de ZINGS “knien”!] Speaker KB: – De SCHRIEËFS “knien”? [– Nae, ik FLUUSTER “knien”.] – I SHOUT “rabbit(s)”. – You SHOUT “rabbit(s)”? I thought, you SING “rabbit(s)”! – You SHOUT “rabbit(s)”? – No, I WHISPER “rabbit(s)”.
[–focus,+final] continuation	Ièrs ZEI ik “knien”, [toen ZONG ik “knien”, en toen SCHREEF ik “knien”.] First I SAID “rabbit(s)”, then I SANG “rabbit(s)”, and then I WROTE “rabbit(s)”.
[–focus,+final] high question	Speaker YK: – ZAESE “knien”? [– Nae, ik SCHRIEËF “knien”.] Speaker KB: ZAESE “knien”? [Ik dach, de ZINGS “knien”!] YK: – Do you SAY “rabbit(s)”? – No, I SHOUT “rabbit(s)”! KB: – Do you SAY “rabbit(s)”? I thought, you SING “rabbit(s)”!

<i>[-focus,-final], prenuclear, low question & declarative</i>	– Hebse “knien” GEZÓNGE? – Nae, ik heb “knien” GEFIESPELD. – Did you SING “rabbit(s)”? – No, I WHISPERED “rabbit(s)”.
<i>[-focus,-final], postnuclear, low question & declarative</i>	– Huurt allein ANNIE “knien” good? – Nae, ouk MIEKE huurt “knien” good. – Does only ANNIE hear “rabbit(s)” well? – No, also MIEKE hears “rabbit(s)” well.

Table 1. Sentences representing tonal minimal pairs. Sentences in italics (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). The carrier sentences for the [-focus,+final] questions were different for the two speakers, but they triggered the same discourse meaning.

2.2. Recordings

The sentences were presented to the speakers in the form they are given in Table I 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 *knien*^{II} (‘rabbit’) in a [+focus, +final] context with high question intonation, was presented as follows:

-  – In ‘t Venloos zaeste gewoeën “**KNIEN**”.
– Ech waar? 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, neither speaker consistently used one contour in one condition. After an additional session, we ended up with complete data sets for both speakers. For the elicitation of the ‘continuative’ intonation, we used a ‘listing’ context, in which both speakers consistently used the intended contours. An earlier attempt with a context in which the target word occurred in a non-final clause led to a variety of intonation contours for both speakers.

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.

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 and van Gestel 1991).

Figure 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 contours in Figure 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 2. The differences between the apparent durations of Figure 1 and the actual durations in Figure 2 are nowhere more than 10 ms, except in YK’s realization of *derm*² 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.

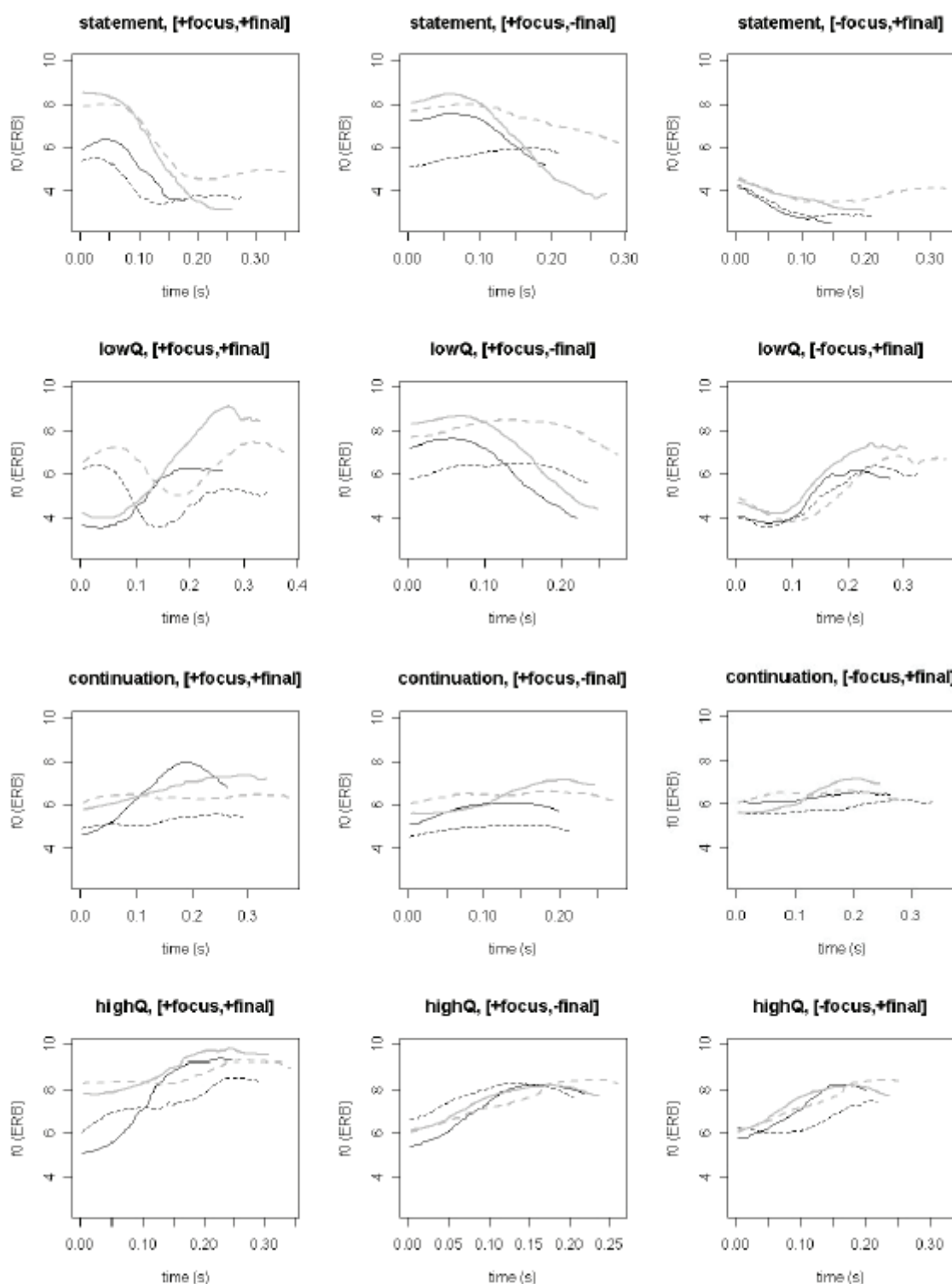


Figure 1: f_0 contours of Venlo Accent 1 and Accent 2 (ERB) in four intonation contours and three prosodic contexts, as a function of time (ms). Solid lines represent Accent 1 and dashed lines Accent 2. Gray lines give speaker YK's utterances and black lines give speaker KB's utterances.

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 non-final 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 non-final 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.

Gussenhoven & van der Vliet (1999) 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. As expected, 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 phonological, we included both contours, thus restoring symmetry to the Venlo tonal system.

In this 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 non-final context, the ‘low interrogatives’ resemble the corresponding ‘declarative’ contours. In the [-focus, +final] position, the contour falls slightly and then rises again, to end with a brief plateau. The difference between Accent 1 and Accent 2 would appear only durational, Accent 2 being longer.

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 approximately 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 and van der Vliet (1999). 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 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 non-focused 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. 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 4.2.5.

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, $p < .001$; see Figure 2². In fact, Accent 2 is longer than Accent 1

2. The analysis of absolute durations was done with an 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.

in all intonations and contexts, except in six individual cases (out of 96)³. Figure 2 shows the durational differences between Accent 2 and Accent 1 per context, intonation and speaker. Figure 3 shows the actual rhyme durations for Accent 1 and Accent 2 separately.

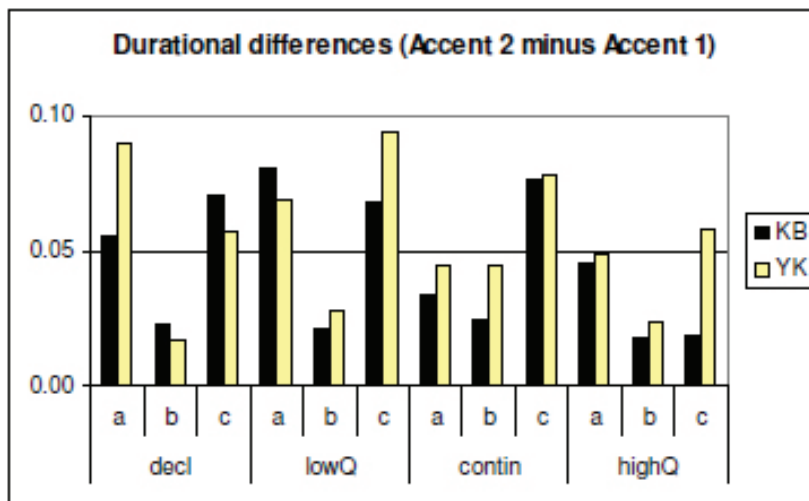


Figure 2: Durational differences (seconds) between Accent 2 and Accent 1 rhymes per intonation contour (‘declarative’, ‘low interrogative’, ‘continuation’ and ‘high interrogative’) and context (a= [+focus,+final], b=[+focus, -final], c=[-focus,+final]), as pronounced by two speakers.

We also compared durations between speakers, intonations and contexts, and found significant differences in all three categories. 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

3. In these cases, the differences in duration range from 3 to 19 ms. Half of the cases concern the nonfinal context.

Cologne (Peters 2006), while pre-final lengthening was also found for dialect of Roermond (Fournier, Verhoeven, Swerts & Gussenhoven 2006).

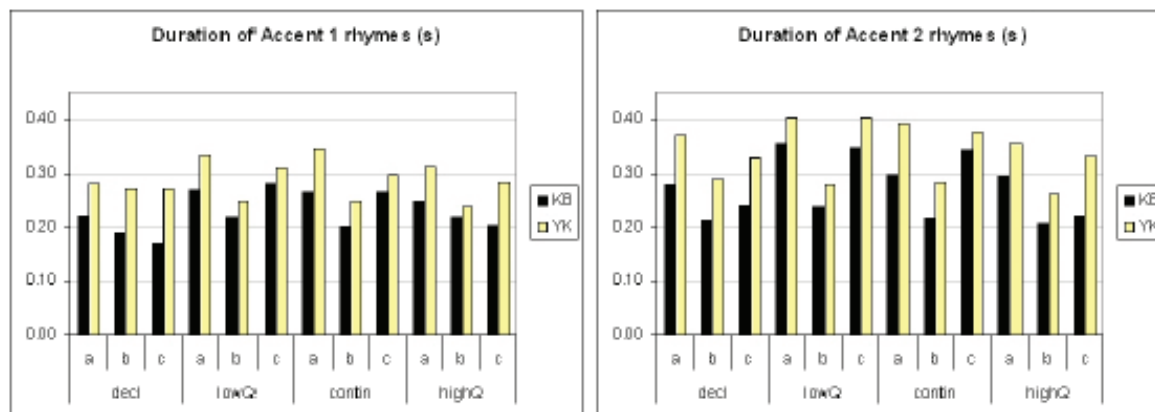


Figure 3. Mean rhyme durations of Accent 1 (panel a) and Accent 2 (panel b) in s., pooled over four lexical items in ‘declarative’, ‘low interrogative’, ‘continuative’ and ‘high interrogative’ intonations in focused final, focused non-final, and non-focused final contexts for speakers YK and KB.

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 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]). 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 effect⁴.

4. We also analyzed our data with the additional factor WORD. Although the choice of the lexical item does have an 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 *derm1-derm2* pair than, say, within a *stein1-stein2* pair. In the remainder of this study, we will not evaluate possible effects of the lexical item on the results.

3. Recognition: Tonal contrast in different prosodic contexts

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.

3.2. Results

Figure 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.

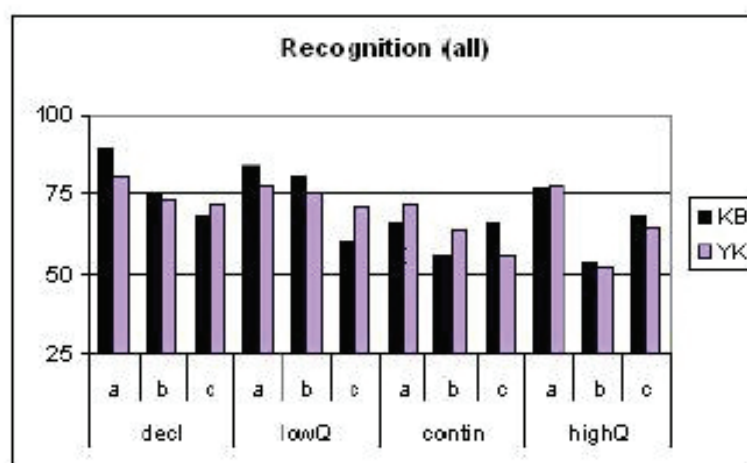


Figure 4: Recognition percentages averaged over Accent 1 and Accent 2 per intonation contour ('declarative', 'low interrogative', 'continuation' and 'high interrogative'), context (a= [+focus,+final], b=[+focus, -final], c=[-focus,+final]), and speaker (KB, YK).

The recognition scores displayed in Figure 4 show recognition scores per contrast. The mean recognition score 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, there was a significant interaction between tone and intonation. This is due to the fact that Accent 2 was better recognized than Accent 1 in sentences with 'continuative' intonation (Accent 2 scores are better than those for Accent 1 by 22%). The non-final context shows the greatest bias towards Accent 2, meaning that here Accent 1 is often misinterpreted as Accent 2.

The main effect for 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 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 INTONATION and 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 non-final contexts.

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 AGE_GROUP ($p < .001$). The difference was found to be fairly constant across contexts, intonations and speakers. Nevertheless, we found a 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.

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 somewhat 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 trend than we found for the Venlo group.

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 4.1 investigates the acoustic distance measures we calculated

between Accent 1 and Accent 2, Section 4.2 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 4.3 explains why we can safely limit our data set to the [+focus] and [+final] cases.

4.1. 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 takes 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 r 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. Whenever 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 Pearson's r is less appropriate. In order to see why this is the case, consider the hypothetical contours in Figure 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.

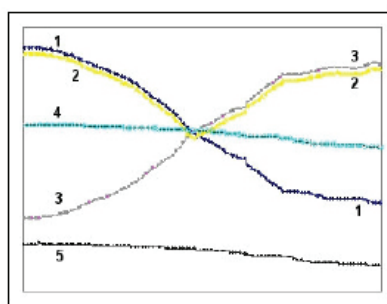


Figure 5. Five hypothetical f_0 contours, whose distance measures are to be calculated using (a) Root mean square error (RMSE) scores, (b) Pearson's correlation coefficients, and (c) cosine distance scores (converted from the original correlation coefficients).

Table 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 r is 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. It is also insensitive to register, as is clear from a comparison of contours 4 and 5, which differ in register, but whose slopes are identical, and thus have a Pearson of $r=1$. By contrast, RMSE scores (also shown in Table 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.

f_0 contours	r_{Pearson}	RMSE	d_{Cos}
1 vs. 2	0.24	2.61	5.33
1 vs. 3	-1	4.28	22.51
1 vs. 4	0.95	1.87	4.25
1 vs. 5	0.95	4.98	1.44
3 vs. 4	-0.95	2.41	7.59
4 vs. 5	1	4.50	0.98
3 vs. 5	-0.95	5.02	13.60

Table 2. Difference measures between the hypothetical contours in Figure 5. 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.

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 f_0 contours as sequences of straight lines. This would appear to be the explanation of Hermes' results, whose stimuli were straight-line stylizations of natural f_0 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 6, which represents Accent 1 and Accent 2 contours in $[-\text{focus}, +\text{final}]$ declarative position. The upper pair is a time-normalized version of KB's $[-\text{focus}, +\text{final}]$ declarative contours, averaged over four words (see Figure 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 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 taken into account, which explains the dramatic consequences of the irregularities in the contours for the similarity measure concerned. In this case, RMSE scores reflect more accurately the resemblance between the two contour pairs, with rather low distance values in both cases.

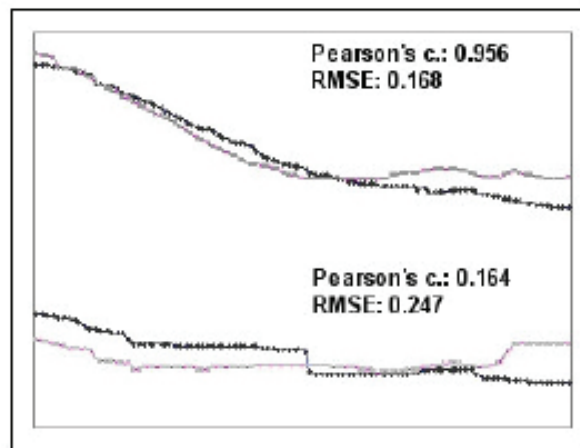


Figure 6. Pairs of Accent 1 – Accent 2 f_0 contours and two objective distance measures (Pearson's coefficients, RMSE scores) found for each pair. Upper panel: contours measured in four different words, time-normalized and averaged. Lower panel: time-normalized contour found in one word, *bein*. Irregularities 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. 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).

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

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.

$$r_{cos} = \frac{\sum_{i=1}^p g_i h_i}{\sqrt{\sum_{i=1}^p (g_i)^2} \sqrt{\sum_{i=1}^p (h_i)^2}} \quad (2)$$

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 $d_{cos} = 100*(1-r_{cos})$, which leaves the properties of the cosine method unaffected.⁵ The similarity scores are given in the third column of Table 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 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.

4.2. Objective distances between Accent 1 and Accent 2

Measures for f_0

RMSE and cosine distance scores between Accent 1 and Accent 2 were computed separately for each

5. The correlations between cosines and transformed cosines have opposite signs, because cosine coefficients, like Pearson’s coefficients, increase as the distance between contours decreases (negative correlation), while their transformed values increase as distance increases (positive correlation).

context and intonation pattern, and for each speaker, based on 100 f_0 values per contour. The results are given in Figure 7.

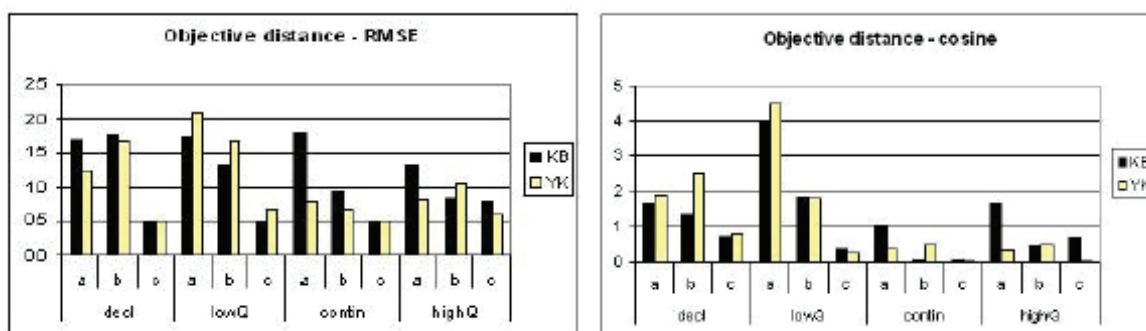


Figure 7: RMSE (left panel) and cosine (right panel) distance measures between Accent 1 and Accent 2 for each combination of context ($a=[+focus,+final]$, $b=[+focus,-final]$, $c=[-focus,+final]$), intonation ('declarative', 'low question', 'continuation' and 'high question') and speaker (KB, YK).

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 dominates 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, followed by those with 'declarative' 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 lo-

wer pitch than the corresponding contours in KB's realizations. This difference between speakers is not obviously reflected in the recognition scores (see Figure 4), neither is it reflected in the subjective distance judgements, as we will see in section 4.2.

A measure for duration

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. As we saw in section 2.3.2 (Figures 2 and 3), Accent 2 is longer than Accent 1, in particular in the final contexts. When we included the durational contrasts (Figure 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 identifying information. These generalizations seem robust, but it should be noted that they are only partly supported by the statistical analysis. As was said in section 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, while for the cosine distances only intonation ($p=0.036$) and intonation*context ($p=0.004$) were 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 2 is neutralized.

4.3. Neutralization in the [-focus,-final] context

Figure 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.

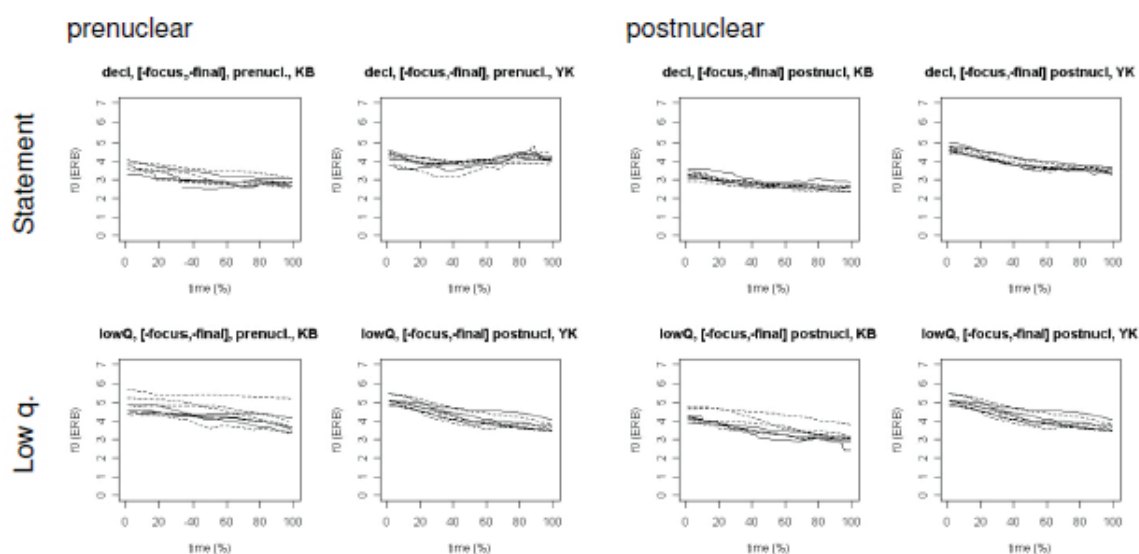


Figure 8. f_0 contours for Accent 1 (solid lines) and Accent 2 (interrupted lines) in ‘declarative’ and ‘interrogative’ intonations in the non-focused, non-final positions for speaker YK and speaker KB separately.

In this context, the contrast between Accent 1 and Accent 2 has been reported to be neutralized.

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 great similarity between the contours in this context compared to the other contexts. The RMSE distance scores gave similar results⁶. In addition to f_0 , a few more measures were computed in order to 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 the 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

6. 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 higher pitch register in questions than in statements (Ohala 1984, Gussenhoven 2002, Haan and 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.

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 unlikely to reflect a categorical difference between Accent 1 and Accent 2. We conclude 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. Subjective distances between Accent 1 and Accent 2

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 Limburgish 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 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 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.

4.4.1. 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 r between each judge's scores and the mean scores over all subjects. Coefficients 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 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.

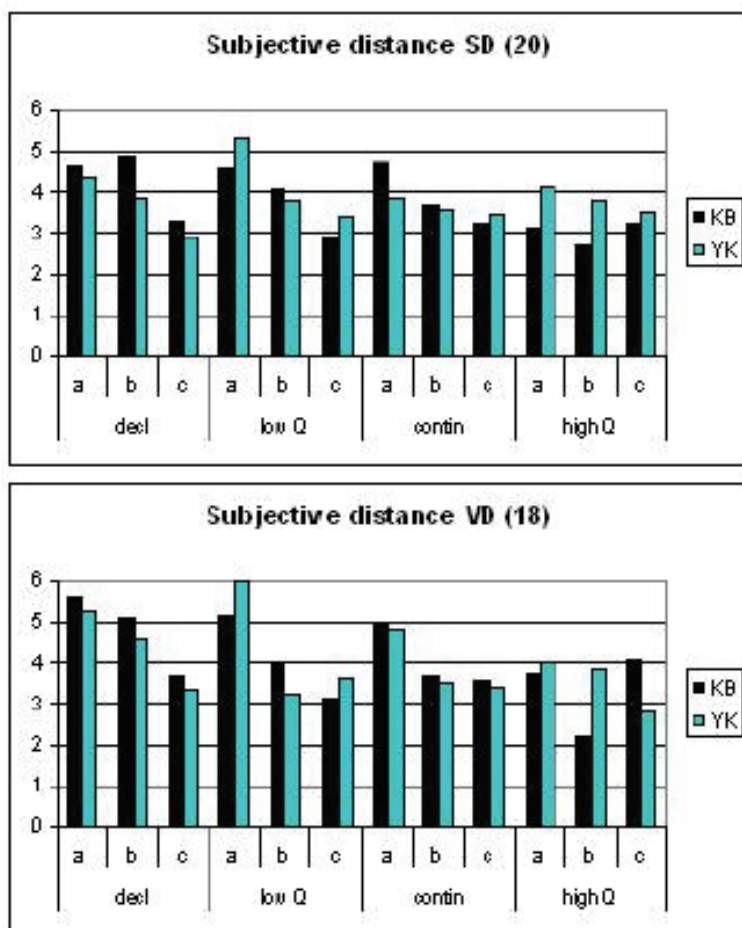


Figure 9: Subjective distance by non-native (upper panel) and native speakers (lower panel), averaged per intonation (declarative, low question, continuative and high question), context (a= [+focus,+final], b= [+focus,-final] and c= [-focus,+final]) and speaker (KB, YK).

4.4.2. 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 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 fur-

ther processing⁷. 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].

4.4.3. 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 9, which shows that KB is judged by the Venlo listeners to 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 1 and 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.

Other than 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 10. In our evaluation of the salience of the recognition scores, we will use these pooled data.

7. 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.

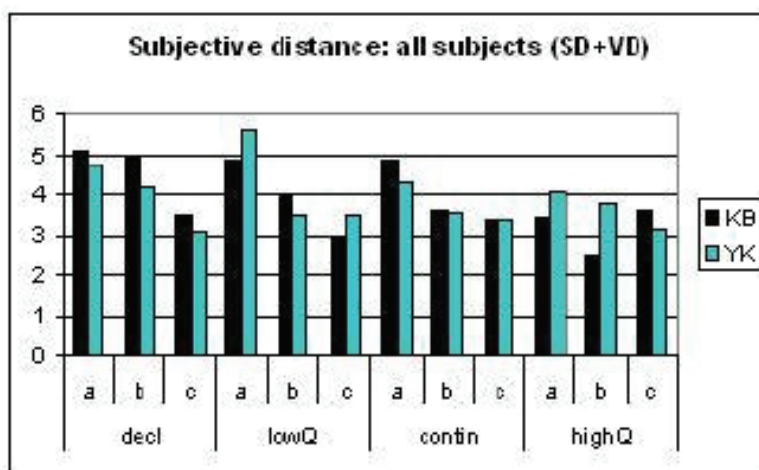


Figure 10. Subjective distance scores given by 38 native and non-native speakers, averaged per intonation (declarative, low question, continuative and high question), context (a= [+focus,+final], b= [+focus,-final] and c= [-focus,+final]) and speaker (KB, YK).

5. Contrast salience and recognizability

Section 4 showed that the phonetic salience of the Venlo tone contrast varies with the intonation contour of the utterance, the accentuation of the target words and their position in the sentence. In this section, we will investigate whether low salience of a contrast leads to poor recognizability. In section 4.1.3 we saw 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 2006), a reasonable assumption 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 3 gives Spearman’s ρ ’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 inspected our data from different angles in order to exclude two possible blurring fac-

tors. 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)⁸, 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] rhymes is 360ms, its "relative duration" value was set to -60; if the rhyme was 390ms long, its new value was 30. Accent 1 received the opposite treatment: if an Accent 1 rhyme pronounced by YK was 30ms *longer* than the average duration, its new value was -30, and if it was 30ms *shorter*, the 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 and short instances of Accent 1 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 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

8. 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.

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.

		recogV	SubjDist	RMSE	d_cos	dur_abs
RecogV	Correlation Coeff.	1,000	0,379	0,322	0,466	0,119
	Sig. (2-tailed)	.	0,000	0,001	0,000	0,249
SubjDist	Correlation Coeff.	0,379	1,000	0,645	0,471	0,216
	Sig. (2-tailed)	0,000	.	0,000	0,000	0,034
RMSE	Correlation Coeff.	0,322	0,645	1,000	0,791	-0,115
	Sig. (2-tailed)	0,001	0,000	.	0,000	0,266
d_cos	Correlation Coeff.	0,466	0,471	0,791	1,000	-0,108
	Sig. (2-tailed)	0,000	0,000	0,000	.	0,296
dur_abs	Correlation Coeff.	0,119	0,216	-0,115	-0,108	1,000
	Sig. (2-tailed)	0,249	0,034	0,266	0,296	.

Table 3. Spearman’s ρ and significance levels between recognition scores (Recog), difference judgements (SubjDist), Root Mean Square Error (RMSE), inverted cosines and duration difference. $N=96$.

The extent to which the subjective distance measures are explained by the objective distance measures can be answered by inspecting their correlations.⁹ 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

9. We did not run regression analyses to predict either the recognition scores or subjective distance scores because of the multicollinearity in the data, where some of the predictor variables have higher correlations with each other than with the dependent variable.

significance of the correlation with the duration difference was due only to the [-focus,+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.

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 across 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. 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), our acoustic difference measures showed that the contrast between Accent 1 and Accent 2 was encoded with two kinds of acoustic information, f_0 and 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 largely 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 significant. Although native listeners do take durational differences into account in their phonetic distance judgements, they do not use them reliably during recognition.

The acoustic difference measures had a moderate correlation with the subjective distance measure, suggesting that phonetic salience judgements are more complex than can be captured by the measures we used, RMSE, *d-cos* and duration. The closest measure was RMSE, although *d-cos* showed a higher correlation with recognition scores.

7. Appendix: Stimuli

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:].

Context	Accent 1: <i>beer</i> ([be:r], ‘beer’)	Accent 2: <i>baer</i> ([be:r], ‘bear’)
[+focus,+final] declarative	– Waat hebse gedrónken? – 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 gedrónke? – Nae, allein KÓFFIE . – 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 DÓNKER 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 DÓNKER ? – Nae, ‘t beer is BLOND . – Is the ale DARK? – No, it is BLOND.	– Is d’n baer BROÉN ? – Nae, d’n baer is ZWART . – Is the bear BROWN? – No, the bear is BLACK.
[-focus,-final], postnuclear, interrogative & declarative	– Hebse DÓNKER beer gedrónke? – Nae, ik heb BLOND beer gedrónke. – Did you drink DARK ale? – No, I drank BLOND ale.	– Hebse ‘nen BROÈNE baer ge- jaag? – Nae, ik heb ‘n ZWARTE baer ge- jaag. – Did you hunt a BROWN bear? – No, I hunted a BLACK bear.

7.2. Words used the subjective distance experiment

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 exact 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 exact 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

7.2.2 .Answer sheet (extract)

Blokje 1

	geen verschil	heel veel verschil
	0 1 2 3 4 5 6 7 8 9	
bein	o o o o o o o o o o	
stein	o o o o o o o o o o	
[...]		
derm	o o o o o o o o o o	
	0 1 2 3 4 5 6 7 8 9	

References

Alsters, A.C.T, J.H.O.P Alsters–van der Hor, H.H.E. van Daelen, A.J. Reijnders, G.A. Rongen, and J.L.L. Vissers: Venloos Woordenboek. Venlo: Stichting Henric van Veldeke (1993).

Bakkes, P; H. Crompvoets, J. Notten, F. Walraven: Veldekespelling 2000, Veldeke, Jaargang 76 – 2001 nummer 1, p. 13–14 (2001).

Beckman, J.N.: Positional faithfulness. Ph.D. dissertation, University of Massachusetts Amherst, Massachusetts (1998). [Retrieved June 23, 2008, from Dissertations & Theses: A&I database. Publication No. AAT 9823717].

Boersma, P. & D. Weenink (2005). Praat: Doing phonetics by computer (Version 4.3.11) [Computer program]. Retrieved 8 May 2005 from <<http://www.praat.org>>.

Burnham, D; E. Francis, D. Webster, Di, S. Luksaneeyanawin, C. Attapaiboon, F. Lacerda and P. Keller (1996): Perception of lexical tone across languages: evidence for a linguistic mode of processing, In ICSLP–1996, 2514–2517.

Depoux, Sebastián-Gallés, Navarrete & Peperkamp (2008). Persistent stress ‘deafness’: the case of French learners of Spanish. *Cognition*, 106, 682-706.

Fournier, R., Verhoeven, J., Swerts, M. & Gussenhoven, C. (2006). Perceiving word prosodic contrasts as a function of sentence prosody in two Dutch Limburgian dialects. *Journal of Phonetics* 34, 29-48.

Gandour, J (1983). Tone perception in Far Eastern languages. In *Journal of Phonetics* 11, p. 149–176.

Gussenhoven, C. & van der Vliet, P. (1999). The phonology of tone and intonation in the Dutch dialect of Venlo. *Journal of Linguistics*, 35, 99-135.

- Gussenhoven, C. (2000a). The lexical tone contrast of Roermond Dutch in Optimality Theory. In M. Horne (Ed.), *Intonation: Theory and Experiment* (pp. 129-167). Amsterdam: Kluwer. [Also ROA-382]
- Gussenhoven, C. (2000b). The boundary tones are coming: On the non-peripheral realization of intonational boundary tones. In M. Broe & J. Pierrehumbert (Eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon* (pp. 132-151). Cambridge: Cambridge University Press.
- Gussenhoven, C. (2002). Phonology of intonation. State-of-the-Article. *GLOT International* 6 (Nos 9/10). 271-284.
- Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.
- Gussenhoven, C. & Peters, J. (2004). A tonal analysis of Cologne Schärffung. *Phonology*, 21, 251-285.
- Hanssen, J (2005): *Tone and intonation in the dialect of Sittard*. MA Radboud University Nijmegen, the Netherlands.
- Hermes, D.J. and J.C. van Gestel: The frequency scale of speech intonation. *Journal of the Acoustical Society of America*, 90, pp. 97–102 (1991).
- Hermes, D.J.: Measuring the perceptual similarity of pitch contours. *Journal of Speech, Language, and Hearing Research*, 41:73–82 (1998).
- Heuven, V. van and J. Haan (2000). Phonetic Correlates of Statement versus Question Intonation in Dutch. In: Botinis, A (ed). *Intonation: Analysis, Modelling and Technology*. Kluwer.
- Huang, Tsan (2001). The interplay of perception and phonology in Tone 3 sandhi in Chinese Putonghua. In Hume, E. and K. Johnson (eds.), *Studies on the interplay of speech perception and phonology* (OSU working papers in linguistics, No. 55), pp.23-42.
- Hyman, L. M. (1990). Boundary tonology and the prosodic hierarchy. In Sharon Inkelas & Draga Zec (eds), *The phonology-syntax connection*, 109-125. Chicago and London: University of Chicago Press.
- Labov, W.(1991): *The three dialects of English. Analysis of Sound Change in Progress*. Ed. Penelope Eckert. New York, NY: Academic Press .
- Lee, Y-S, D.A. Vakoch, & L. H. Wurm (1996). Tone Perception in Cantonese and Mandarin: A Cross-Linguistic Comparison. *Journal of Psycholinguistic Research*, 25:527–542.
- Mielke, J. (2003): The Interplay of Speech Perception and Phonology: Experimental Evidence from Turkish. In: *Phonetica* 60:208–229.
- Ohala, J. J. (1984). An Ethological Perspective on Common Cross-Language Utilization of Fo Of Voice. *Phonetica* 41 (1984), p. 1–16.
- Peperkamp, S. & E. Dupoux, E.(2002): A typological study of stress ‘deafness’. In C. Gussenhoven & N.

Warner (Eds.) *Laboratory Phonology 7*. (Mouton de Gruyter, Berlin/New York), p. 203–240.

Peters, J. (2006). “The Cologne word accent revisited”. *Zeitschrift für Dialektologie und Linguistik* 131, 107-133. [= Beihefte; Michiel de Vaan (ed.), *Germanic tone accents. Proceedings of the First International Workshop on Franconian Tone Accents, Leiden, 13-14 June 2003*. Stuttgart: Steiner]

Steriade, D. (2008). The phonology of perceptibility effects: the P-map and its consequences for constraint organization. In K. Hanson and S. Inkelas (Eds). *The Nature of the Word: Essays in Honor of Paul Kiparsky*. MIT Press, Cambridge MA.