

Perception of the Lexical Accent Contrast in One Variety of East Norwegian

Language and Speech
2018, Vol. 61(3) 339–357

© The Author(s) 2017

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0023830917714014

journals.sagepub.com/home/las



Niamh E. Kelly

American University of Beirut, Lebanon

Rajka Smiljanić

University of Texas at Austin, USA

Abstract

The perception of the lexical pitch accents was examined in the Trøndersk dialect of Norwegian. Based on a production study, a categorization of stimuli with manipulated pitch contours was conducted. The experiment tested which acoustic cues (height and alignment of fundamental frequency (F0) minimum, and alignment of F0 maximum and turning point from maximum to minimum) are necessary for the perception of the tonal contrast. The results are consistent with the production findings in that changes in all of the examined acoustic cues contributed to the shift in accent categorization. The later alignment of the main F0 landmarks (F0 maximum, F0 minimum and turning point from maximum to minimum) induced accent 2 identification. Raising F0 minimum height also led to more accent 2 responses. The analysis of the perception patterns furthermore revealed that the effect of a later alignment of F0 minimum was weak unless combined with a later alignment of the other F0 landmarks, or a higher F0 minimum level, all of which contributed to more accent 2 responses. These results indicate that accent 1 is characterized by an early fall, and accent 2 by a salient initial high tone. Implications of these findings for the phonological analysis of the tonal contrasts in the Trøndersk dialect are also discussed.

Keywords

Scandinavian, Norwegian, perception, pitch accent, prosody

Introduction

In languages with a lexical tonal accent (or pitch accent) contrast, the pitch contour on a stressed syllable can change the meaning of a word. Varying instantiations of this phenomenon are found in Scandinavian languages (Bruce, 1977; Gårding, 1973; Fintoft, 1987), Lithuanian (Senn, 1966),

Corresponding author:

Niamh Kelly, Department of English, American University of Beirut, PO Box 11-0236, Riad El Solh 1107-2020, Beirut, Lebanon.

Email: nk114@aub.edu.lb / niamhkelly632@gmail.com

Latvian (Derksen, 1966; Karins, 1996), Japanese (Pierrehumbert & Beckman, 1988), and some varieties of Korean (Chang, 2013; Kim, 1988), Basque (Hualde, 1991), Dutch and German (Gussenhoven, 2004) and Serbian and Croatian (Smiljanić, 2006). This feature differs from lexical tone in a tone language because while lexical tone can surface on the syllable or mora (meaning that a polysyllabic word can bear more than one tone), tonal accent occurs at the level of the word, so even a polysyllabic word will bear just one accent. Tonal accent languages are described by Hualde as languages in which “words contrast in the tonal melody associated with the stressed syllable” (Hualde, 2012, p.1335).

Tonal accent is also known as “word accent”, particularly in reference to Swedish (Bruce, 1977; Gårding, 1973), while Fintoft (1987) refers to it as a “toneme” system in his description of Norwegian. Tonal accents of Norwegian and Swedish have been studied extensively (e.g., Bjerrum, 1948; Elstad, 1978; Gussenhoven, 2004; Lorentz, 1981; Riad, 1998, 2006; Segerup, 2003, 2004; Storm, 1884; Van Dommelen, 2002; Van Dommelen & Nilsen, 2003; Vanvik, 1957). Prosodically, Norwegian and Swedish are stress languages with a lexical contrast in the alignment of pitch contours. The two contrasting accents in Norwegian and Swedish are referred to as accent 1 and accent 2. The tonal accent found today in most varieties of Norwegian and Swedish is thought to have arisen historically from a tonal contrast between monosyllables and polysyllables in Old Norse (Kristoffersen, 2000; Oftedal, 1952). When monosyllables ending in an obstruent sonorant sequence became disyllabic due to vowel insertion, they retained the tonal contour of monosyllables, thus creating a contrast in tonal contour rather than in syllable number. Another analysis, whereby stress was replaced with a lexical accent (Kock, 1884/1885; Riad, 2003), has also been proposed. While most varieties of Norwegian and Swedish have retained the tonal accent, in most varieties of Danish it has been lost, although words with accent 1 in Norwegian and Swedish correspond to words with *stød* (laryngealization) in Danish (Grønnum & Basbøll, 2003; Wetterlin, 2010). In Stockholm Swedish, the difference between accent 1 and accent 2 is in the timing of the fundamental frequency (F₀) fall in relation to the stressed syllable (Bruce, 1977). This fall occurs later relative to the segmental string for accent 2 than for accent 1 (similar to Figure 1). Swedish dialect types have further been divided into one-peaked and two-peaked dialects. This distinction refers to the realization of sentence accent as a second peak in double-peaked dialects (in single-peaked dialects this is expressed by an expansion of pitch range), highlighting the importance of distinguishing between word and sentence-level tonal patterns (Gårding, 1973; Gårding & Lindblad, 1975). Similarly, most dialects of Norwegian have the lexical accent contrast, in which the main peak is always earlier in accent 1 and accent 1 never has more peaks than accent 2, regardless of the phonetic realization of the contours (Fintoft, 1987).

Norwegian dialects are divided into East, West, Central and North Norwegian. Some dialects in the north do not have a tonal accent contrast, while the central dialects tend to be grouped with East Norwegian in terms of tonal accent. East Norwegian is considered a low-tone dialect, since accent 1 is a low tone (L) in most varieties and accent 2 is a high–low (HL) contour. West Norwegian is a high-tone dialect, where accent 1 is a high tone (H) and accent 2 is a low–high (LH) contour (Almberg, 2004; Kristoffersen, 2000). Fintoft (1987) notes that changes in the accent contours (such as timing of the peak) occur not abruptly but gradually as one moves through the country. The current investigation centers on Trøndersk, a variety of East Norwegian spoken in the Trøndelag region in the center of the country. This dialect is understudied, and it will function as a test case for examining the acoustic cues that distinguish the accents in continuous speech. Specifically, this work highlights the usefulness of perception studies in determining the salient characteristics of the tonal accent contrast by examining what cues listeners use to identify the two accents in this dialect, and

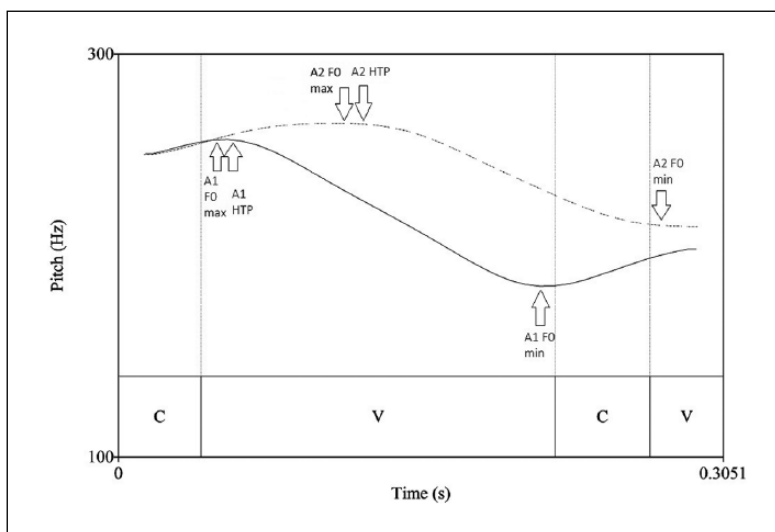


Figure 1. Trøndersk accent 1 (solid) and accent 2 (dashed), showing F0 maximum, F0 minimum and high turning point (HTP; turning point from the F0 maximum to the F0 minimum). These contours are based on average measurements of the landmarks for each accent.

whether all cues in production are necessary for a particular contrast to be perceived by listeners. The current investigation expands on previous perception studies by systematically manipulating more cues (F0 minimum height and alignment, F0 maximum alignment, and alignment of the turning point from maximum to minimum), in order to examine whether the perception of these cues will trigger categorical shifts in perception between the two accents. This will further provide insight into whether a combination of F0 cues will induce a change in identification responses. These results will speak more broadly to the perception of tonal accent contrasts in other varieties of Norwegian, and other languages with lexical tonal contrasts, by providing an acoustic analysis of tonal contours and examining the extent to which it coincides with phonological descriptions of the accents, and by determining which acoustic cues are salient in the perception of the contrasts.

A summary of the production findings of research on disyllabic forms in dialects spoken in the Trøndelag region is provided in Table 1. From these descriptions, there is large variation in both what constitutes tonal categories and how they are implemented, with some indicating alignment as the difference between the accents and others describing the tonal makeup as differentiating accent 1 from accent 2. In Fintoft (1970), some tonal minimal pairs differed in F0 at vowel onset, while other pairs did not. There are also inconsistent findings regarding whether there is an initial H, particularly in accent 1. The somewhat conflicting findings of the studies listed in Table 1 may also arise from their use of words that were recorded in isolation (not in carrier sentences), thus potentially including sentential intonation in the descriptions, and of small numbers of speakers and words, such that for these experiments, the results may not be representative of the population of speakers.

Most recently, Kelly and Smiljanić (2017) examined in detail the acoustic correlates of the two accents produced phrase-medially and phrase-finally under broad and narrow focus by 10 speakers of the Trøndersk dialect (Kelly, 2015; Kelly & Smiljanić, 2014, 2017) (all studies used the same data). Disyllabic target words with either accent 1 or accent 2 were examined, all with initial stress

Table 1. Summary of previous production findings on Trøndelag dialects.

| Author | Dialect | Accent 1 | Accent 2 | Difference |
|--------------------------------|----------------|-----------------------|----------|----------------------|
| Fintoft (1970) | Trondheim | High–low contour (HL) | HL | Timing Onset? |
| Nilsen (1992): | Trøndersk | Low tone (L) | HL | H tone |
| Van Dommelen & Nilsen (2003) | West Trøndersk | HL | HL | Timing |
| Wetterlin (2010): | Trondheim | L | L | Timing F0 range? |
| Kristoffersen (2006) | Oppdal | HL | HL | Timing |
| Kelly & Smiljanic (2014, 2017) | Trøndersk | HL | HL | Timing F0 minimum |

and controlled for vowel and surrounding consonants. There were five different accent 1 words, each produced three times per speaker, and 2 different accent 2 words, each produced seven or eight times per speaker. These stimuli included one minimal pair, differing only in accent. Detailed acoustic analyses revealed that both accents consist of an HL contour, that is, an initial F0 maximum followed by an F0 minimum (see Figure 1). This suggests that the accent contrast in this differs from other varieties of East Norwegian, such as the Oslo dialect (Fintoft, 1970; Fintoft & Mártony, 1964; Kristoffersen, 2006), which do not have an initial H in accent 1. The important distinction between the accents was in the height and the alignment of the contour. The alignment of the whole contour (both F0 maximum and minimum) with the stressed syllable was found to be later for accent 2 compared to accent 1. The height of the L tone (F0 minimum height) was also higher for accent 2 than for accent 1.

The current study builds on the findings of the acoustic analyses by examining which cues are used by listeners to perceive the accent contrast. While a number of studies examined how the accent contrasts are realized acoustically in various Scandinavian dialects, fewer studies looked at the perception of the tonal accent contrast. One study that did however examine the perception of the accent contrast was Segerup (2004), which found that listeners could correctly identify naturally produced tonal minimal pairs 96% of the time for one variety of West Swedish. With regard to which cues listeners use to make lexical accent identification, Efremova, Fintoft, and Ormestad (1963), using gating experiments found that the shape of the F0 contour in the initial, stressed syllable contains important cues in terms of tonal contour shape for distinguishing the accents in disyllabic words in Swedish. Similarly, Norwegian listeners were able to identify the two Norwegian accents accurately even when presented with truncated words up to the end of the initial, stressed vowel (Fintoft, 1970). Bruce (1977) examined perception of the accent contrast in Stockholm Swedish and confirmed that listeners used the timing of the F0 contour (in particular, the beginning of the fall) in relation to the stressed syllable as the main cue in differentiating between the two accents.

Specifically, listeners identified accent 2 as long as the fall began 25% of the way into the vowel, or later. With regards to the perception of Norwegian tonal accents, two studies tested which aspects of the F0 contours were salient indicators of tonal categories. In a small-scale study, Fintoft and Mártony (1964) manipulated F0 peak height, alignment and the slope of the rise and fall to examine accent identification in the Oslo dialect. With just the first consonant–vowel syllable of the manipulated disyllabic words played to the listeners, a level or rising F0 in the stressed vowel was identified as accent 1, and a falling F0 at the end of the stressed vowel was identified as accent 2. In another perception study, Fintoft (1970) used synthesized stimuli

composed of sine wave signals with manipulated frequency contours, superimposed on segments. The results also showed that Norwegian listeners identified a level or rising frequency contour at the end of the stressed vowel as accent 1, and a falling contour at that point as accent 2. These results indicate that the alignment of the F0 contour is a salient cue for the perception of the lexical pitch contrasts in a variety of Scandinavian dialects. The current investigation expands on this by manipulating more cues, specifically based on the production results for the Trøndersk variety (Kelly & Smiljanić, 2017).

The specific cues manipulated here are: height of F0 minimum; alignment of F0 maximum and minimum; and alignment of the turning point from F0 maximum to F0 minimum. The alignment cues were found to be later for accent 2 than accent 1, while the height of the F0 minimum was found to be higher for accent 2 than accent 1. In order to investigate whether listeners used these cues in perception of the accentual distinctions, we systematically manipulate them and examine their effect on accent judgments. Listeners were presented with tokens in which these acoustic cues were systematically shifted from values representative of accent 1 to values representative of accent 2 and were asked to identify them. Based on the description of the accents in this dialect, it was hypothesized that the manipulated stimuli with a lower and earlier F0 minimum and earlier turning point would be identified as accent 1, and those with a higher and later F0 minimum and a later turning point would be identified more often as accent 2. When both F0 minimum height and alignment were changed, it was hypothesized that the alignment of the F0 minimum would be more important than its height, because previous descriptions of the accent contrast focus on the alignment differences rather than height differences. If found to be correct, this will indicate the importance of the F0 alignment dimension over the F0 range/height dimension in the accentual distinction, which suggests a salience of timing cues over pitch height cues in tonal accent processing. This will further allow us to draw a link between production and perception, to see how many of the acoustic cues that speakers reliably produce constitute salient information that listeners are sensitive to and use to identify the contrast.

2 Methods

2.1 Materials

One token of an accent 1 word (linet /'li:nə/ “the flax/linen”) produced by a female native speaker of Trøndersk, aged 20, excised from a broad focus position¹ was chosen as the template for the cue manipulations. This token was chosen as it was a natural production typical of accent 1 in broad focus. A single word was used rather than multiple words in order to limit the processing by the listeners, so they would tune into differences in the contour and not segments. Based on the measurements from the production study, words with stylized pitch contours were created. These manipulations were done using the PSOLA function in Praat (Boersma & Weenink, 2011). Four cues (as seen in Figure 1) were manipulated: (1) F0 maximum alignment; (2) high turning point (HTP) alignment; (3) F0 minimum height; and (4) F0 minimum alignment. Two Norwegian speakers who did not participate in the experiment provided written comments confirming that the stimuli sounded natural.

2.1.1 F0 maximum alignment. F0 maximum alignment is the alignment of the initial F0 maximum in relation to the stressed vowel onset. F0 maximum alignment was manipulated in six equal steps of 22 milliseconds (ms) (Figure 2). For these stimuli, F0 minimum alignment was manipulated in conjunction so as to keep the slope between them constant.

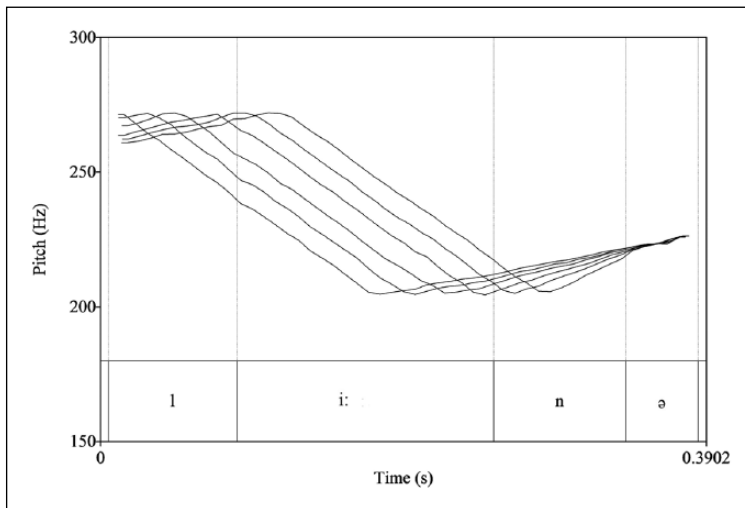


Figure 2. Six steps for F0 maximum alignment and F0 minimum alignment.

HTP alignment is the alignment of the drop from the F0 maximum in relation to stressed vowel onset. This cue was manipulated in six equal steps of 19 ms (Figure 3). F0 minimum alignment was kept constant. Based on the production study, F0 maximum alignment and HTP alignment were manipulated separately because they were independent landmarks, often distantly separated from one another (Kelly & Smiljanić, 2017).

2.1.2 F0 minimum height and F0 minimum alignment. F0 minimum height is the height of the lowest point in the contour after the initial F0 maximum. This cue was manipulated in five equal steps of 8 Hertz (Hz).

F0 minimum alignment is the alignment of the F0 minimum in relation to stressed vowel onset. This cue was manipulated in five equal steps of 20 ms.

Each of the five F0 minimum height steps were combined with each of the five F0 minimum alignment steps, resulting in 25 stimuli. Two schematic representations of two sets of these manipulations—those with the lowest and second highest (for the purposes of clarity in the figure) levels of F0 minimum height—are shown in Figure 4.

The end points for each cue were determined based on the actual productions of the accent 1 and accent 2 versions of the minimal pair (in ms or Hz). The range was then divided evenly into five or six equal steps, continuing for one further step beyond each endpoint. This was done to examine if listeners were able to identify these more exaggerated values more accurately. The cue conditions thus have different numbers of steps to avoid having too large jumps between steps. For all stimuli, the final vowel duration was set to 50 ms and the stressed vowel to 170 ms, F0 at vowel onset was set to 97 semitones (st re 1 Hz; 264 Hz) (this final cue was only set at this value when the manipulation of the contour did not affect it (Figures 3 and 4)).

Table 2 shows the incremental changes for each acoustic cue.

It is important to note that some of these cues could not be manipulated independently, since the manipulation of one necessarily affected others. Therefore, for the manipulation of F0 maximum alignment, F0 minimum alignment was also changed so as to keep the slope constant, while for HTP, F0 minimum alignment was kept constant (thus changing the slope) because making it successively later would have resulted in it being at the end of the word. Furthermore, manipulating F0 minimum

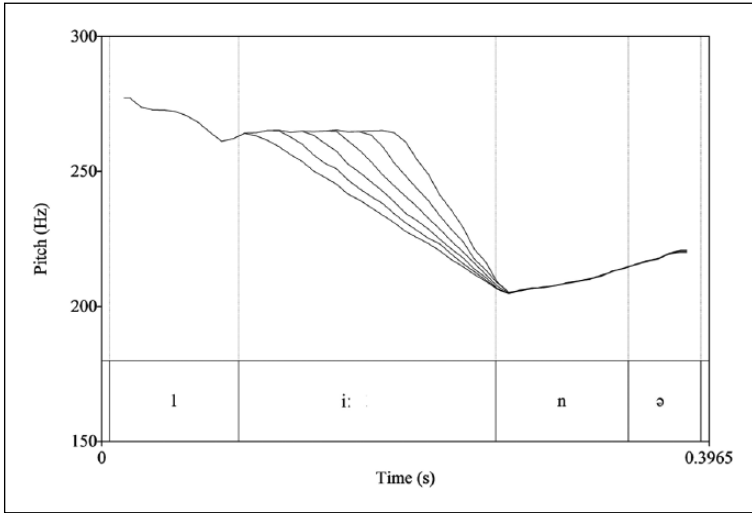


Figure 3. Six steps for manipulations of the high turning point.

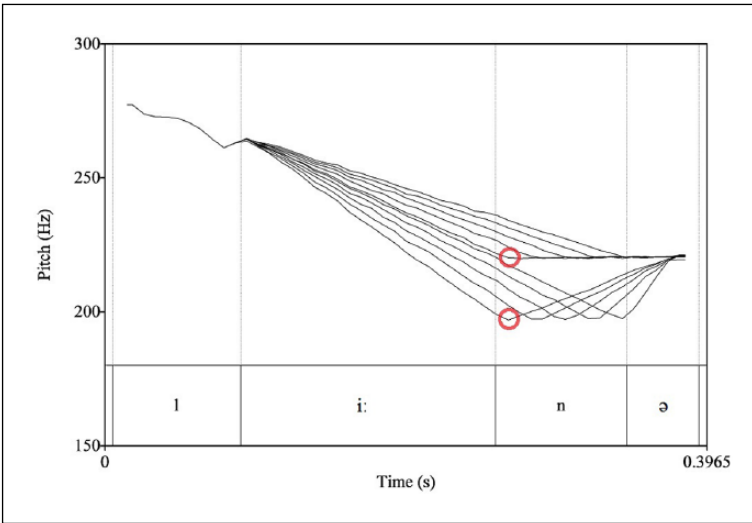


Figure 4. Five steps for F0 minimum alignment, shown at two levels of F0 minimum height (the first level of alignment is marked at both levels of height, for clarity).

alignment and height necessarily meant that the slope also changed. While this means that the results for each manipulation cannot be interpreted in isolation, we believe that the number of cues manipulated and examined still provides clear information about the salient characteristics of the accents.

2.2 Listeners

Listeners were 24 native speakers (16 females, 8 males) of the Trøndersk dialect, aged 18–45. As indicated by a language background questionnaire, they all grew up speaking Trøndersk at

Table 2. The F0 (Hz) and duration (milliseconds (ms) from vowel onset) at each step of the manipulated stimuli. The bold numbers represent the original accent 1 (earlier steps) and accent 2 (later steps) values.

| Measure | Steps | | | | | |
|-------------------------|-------|------------|-----|------------|-----------|-----|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| F0 Maximum timing (ms) | -76 | -54 | -32 | -10 | 12 | 34 |
| High turning point (ms) | 11 | 30 | 48 | 67 | 86 | 105 |
| F0 Minimum (Hz) | 196 | 204 | 212 | 220 | 228 | |
| F0 Minimum timing (ms) | 170 | 190 | 210 | 230 | 250 | |

home, and had not spent a significant amount of time living anywhere else. They were recruited using posters and fliers around the Norwegian University of Science and Technology (NTNU) campus in Trondheim.

2.3 Procedure

Participants were seated in the sound studio at NTNU, Trondheim in front of a desktop computer, and wore headphones. The stimuli were presented in the PsychoPy program (Peirce, 2007). Participants heard one word at a time and were asked to choose which of two words (*linet* or *Line*) on-screen they had heard, in a forced-choice task. There were 37 manipulations total: (5 F0 minimum × 5 F0 minimum alignment) + 6 F0 maximum alignment + 6 HTP. Each stimulus was presented 10 times, for a total of 370 stimuli. The 37 unique stimuli were divided into two blocks of 18 and 19 stimuli respectively, so as not to have any block too long. All the stimuli were heard, without repetition, every two blocks. The stimuli were split up differently for all sets of two blocks, so the combination of stimuli in a particular block was always different. Within each block, the order of stimuli was randomized. There were 20 blocks in total. A pause screen appeared after each block, and participants pressed a key when they were ready to continue to the next block. There were three practice trials to familiarize listeners with the task. Listeners were not given feedback on any trials. The participant's response for one trial initiated the following trial. There was a blink of the screen to indicate that a new trial had begun. If they did not respond within 1.5 seconds after the target word, the experiment moved to the next trial. For half of the participants, the accent 1 word (*linet*—'the linen') always appeared on the left of the screen and the accent 2 word (*Line*—a girl's name) always appeared on the right, and for the other half this was switched.

An additional test was conducted with naturally produced tokens in broad and narrow focus, to confirm that listeners could accurately identify two naturally produced tokens of accent 1 and accent 2 words in isolation. This test was also used to determine whether all listeners were eligible to participate (i.e., had the required language background) and could do the task. This test actually took place after the experiment with manipulated contours, so as to avoid any effect of hearing a specific minimal pair on the identification task of stimuli with the manipulated cues. The natural stimuli were one accent 1 and one accent 2 word with either broad or narrow focus, produced by a male native speaker of the Trøndersk dialect, aged 36 (i.e., not the same speaker as in the rest of the experiment). The words were extracted from carrier sentences eliciting the accent type and broad versus narrow focus reading, and equalized for loudness. The stimuli were chosen because they were representative of the accents in broad and narrow focus conditions, in that they contained the cues found to differ between the accents and focus conditions in the production experiment (Kelly & Smiljanić, 2017). Each stimulus was presented 10 times for a total of 40 stimuli (2 focus

conditions \times 2 accents \times 10 repetitions). These were presented in randomized order in 2 blocks of 20 trials each. Listeners' accuracy in identification of the accents was measured as the proportion of correct responses per accent and condition (broad focus or narrow focus).

2.4 Analyses

Responses were analyzed by comparing how each manipulation (separately) affected identification of the accents, following Chang (2013). For each stimulus, the overall number of accent 1 and accent 2 responses for each manipulation was obtained. In order to determine whether any of the four manipulated cues significantly affected identification of an accent, the responses were subjected to a logistic regression analysis with the dependent variable being the response (accent 1 or 2) and the independent variables being each of the manipulations (F0 minimum height and alignment, F0 maximum alignment, HTP alignment), using the *glmer* function in R (R Development Core Team, 2008). Accent 1 was coded as 0 and accent 2 as 1. As in the figures above, each manipulation was coded 1 (lowest F0 height or earliest alignment) to 5 or 6 (highest F0 height or latest alignment). Listener was included as a random effect, as a likelihood ratio test using the *anova* function showed a model including this to be better than a model without this, for each measure.

3 Results

In the blocks containing naturally produced stimuli, three subjects scored almost 0% accuracy (0%, 0% and 10%, respectively),² so their responses to the manipulated stimuli were disregarded. This left 21 subjects (14 females, 7 males) for analysis. The results indicated that most listeners were highly accurate at identifying accent 1 (72%) and accent 2 (81%) words in broad focus even without any context. The results also revealed that listeners were more accurate at identifying the accents in narrow focus compared to broad focus (accent 1: 94%; accent 2: 90%). Only one participant had a higher accuracy in the broad focus condition. Four participants did equally well in both conditions, with two of these achieving 100% accuracy in both.

In the experiment using manipulated contours, an average of 14 trials (range: 0 to 62) per participant had no response (i.e., the next trial automatically started after the participant failed to respond in the allotted time), which amounted to 3.8% of the data. These trials were excluded from analysis.

Figures 5–7 show the identification results for each cue manipulation averaged across all listeners. In all graphs, the results show that tokens with a later alignment (to the right of the graphs) and higher F0 minimum level are perceived more often as accent 2. It can also be seen that the responses from accent 1 to accent 2 change gradually rather than in a more categorical manner. (This was the case for individual responses as well as pooled responses.) Figures 6 and 7 show more accent 2 responses from left to right (early to later alignment). Figure 5 shows more accent 2 responses from left to right within each graph and across all panels from left to right, corresponding to more accent 2 responses as F0 minimum gets higher and later.

A stepwise logistic regression test³ was used to determine how well the response (accent 1 or 2) could be predicted from each of the manipulations. All parameters significantly affected the responses, as shown in Table 3. Each of the manipulations significantly impacted accent identification, in the expected direction: those with a higher F0 minimum height or later F0 minimum alignment, later HTP alignment or later F0 maximum alignment was identified more often as accent 2. There was also a significant interaction between F0 minimum height and its alignment.

The odds ratio of the coefficient (shown in the Exp column) reveals that F0 maximum alignment had the greatest effect on responses, followed by HTP, then F0 minimum height and finally F0 minimum alignment. These will be discussed in turn.

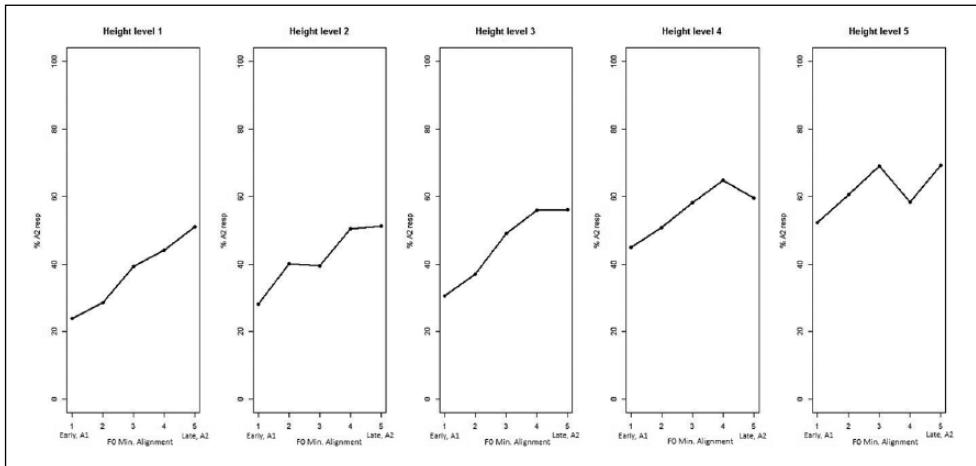


Figure 5. Percentage of accent 2 responses for all listeners when F0 minimum alignment was manipulated. The x-axis shows the manipulations of F0 minimum alignment relative to vowel onset, 1 being the earliest and 5 being the latest. Each panel shows a different F0 minimum height level, the left-most being the lowest and the right-most being the highest.

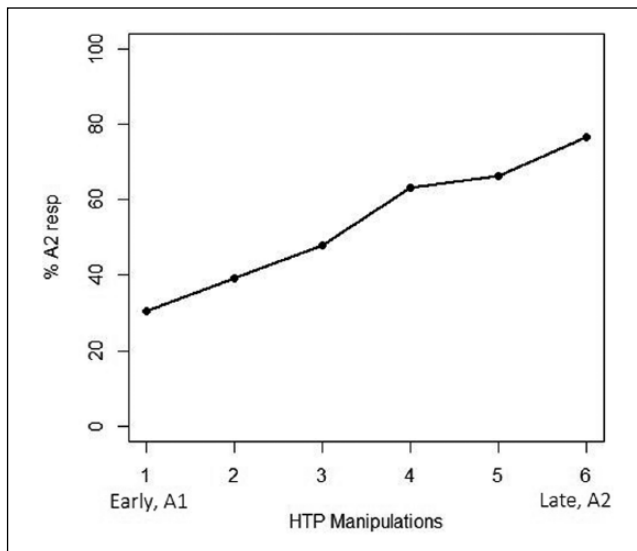


Figure 6. Percentage of accent 2 responses for all listeners when high turning point was manipulated. The x-axis shows alignment steps, with 1 being the earliest (accent 1-like) and 6 the latest (accent 2-like).

To examine the two-way interaction, the results for F0 minimum height and its alignment were examined separately, using a logistic regression. The effect of F0 minimum height was examined separately for each alignment step, and the effect of F0 minimum alignment was examined separately for each height level (as in Figure 5). The results showed that the effect of height was significant at all alignment steps ($p < 0.001$ at each alignment), and the effect of alignment was significant at all height levels ($p < 0.01$ at each height level). These results

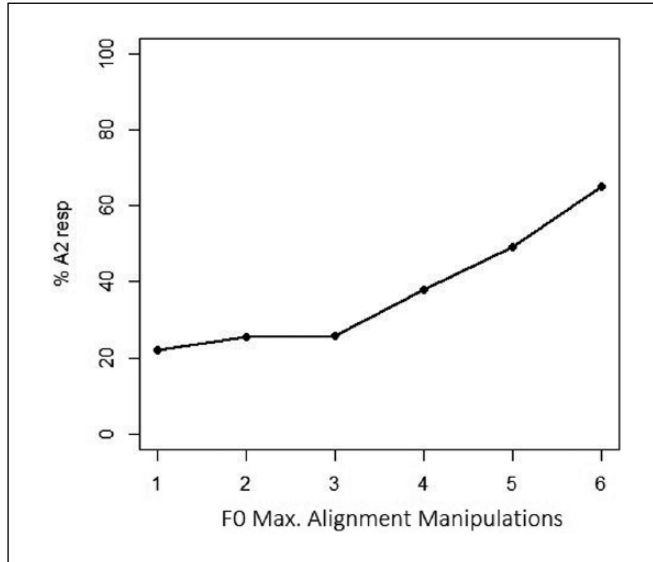


Figure 7. Percentage of accent 2 responses for all listeners when F0 maximum alignment was manipulated. The x-axis shows the alignment step, with 1 being the earliest (accent 1-like) and 6 the latest (accent 2-like).

Table 3. Logistic regression results.

| Condition | Coefficient | Exp() | p-value |
|-----------------------------------------------|-------------|-------|---------------|
| F0 Maximum alignment | 0.5 | 1.65 | $p < 0.001^*$ |
| High turning point | 0.47 | 1.6 | $p < 0.001^*$ |
| F0 Minimum height | 0.3 | 1.35 | $p < 0.001^*$ |
| F0 min. alignment | 0.26 | 1.3 | $p < 0.001^*$ |
| Interaction (F0 minimum height and alignment) | -0.05 | 0.95 | $p < 0.01^*$ |

suggest that these two factors—how high the F0 minimum was and how early/late it was aligned with the segmental string—in conjunction contributed to accent identification. That is, the later and higher the F0 minimum was, the more likely listeners were to identify accent 2. Table 4 lists the steps in the acoustic cue continuum at which the shift in accent identification response from accent 1 to accent 2 occurs (more than 50% responses).

With regard to the interaction between F0 minimum alignment and F0 minimum height, for the earliest alignment step (step 1), the responses change from accent 1 to accent 2 between F0 minimum height steps 4 and 5. For the latest F0 minimum alignment (step 5), majority responses are accent 2. Likewise, for the lowest F0 minimum height (step 1), the majority of the responses change from accent 1 to accent 2 between F0 minimum alignment steps 3 and 4.

The results for F0 minimum height and its alignment provide an insight into which features listeners may use to make their accent judgments. The odds ratio results indicate that F0 minimum height has a similar effect on responses (1.35) to F0 minimum alignment (1.3). The lowest F0 minimum height values elicit a majority of accent 1 responses regardless of how late the F0 minimum alignment is. This can be seen in Figure 5, left-most panel. Only at height level 3 (middle panel)

Table 4. Majority response crossover points for each condition.

| Condition | Crossover point is between steps: |
|------------------------------|------------------------------------|
| High turning point alignment | 3 and 4 |
| F0 maximum alignment | 4 and 5 |
| F0 minimum alignment step 1 | F0 minimum height steps 4 and 5 |
| F0 minimum alignment step 2 | F0 minimum height steps 4 and 5 |
| F0 minimum alignment step 3 | F0 minimum height steps 3 and 4 |
| F0 minimum alignment step 4 | F0 minimum height steps 2 and 3 |
| F0 minimum alignment step 5 | all accent 2 |
| F0 minimum height step 1 | F0 minimum alignment steps 4 and 5 |
| F0 minimum height step 2 | F0 minimum alignment steps 4 and 5 |
| F0 minimum height step 3 | F0 minimum alignment steps 3 and 4 |
| F0 minimum height step 4 | F0 minimum alignment steps 2 and 3 |
| F0 minimum height step 5 | all accent 2 |

when the alignment is late (stimuli 4 and 5, 80% of the way into the following consonant, or 95% of the way into the VC unit (stressed vowel and following consonant), or later), do listeners give 50% accent 2 responses. By height levels 4 and 5 (the two right-most panels), we get consistently more accent 2 responses (at height level 4, alignment step 1 has almost 50% accent 2 responses, and this increases as alignment gets later for both height levels 4 and 5). Likewise, when alignment is early (steps 1 and 2 the end of the vowel or 70–78% into the VC unit), there is a tendency towards accent 1 responses. Only when alignment gets later and height increases are there more accent 2 responses. Figure 4 further shows that the interaction of the F0 minimum height and alignment affects the slope of the contour. When the height of the F0 minimum is high (step 4), the slope of the fall from H to L is very shallow. When it is at its lowest level (step 1), the fall is much steeper.

These results are in accord with the production study where the F0 minimum height of accent 2 was significantly higher than that of accent 1. At the highest level of F0 minimum, listeners could be perceiving a high tone, especially as alignment gets later, whereas when it is at its lowest level, no high tone is perceived, regardless of how late the alignment becomes (at the latest alignment, accent 2 responses reach about 50%).

The hypotheses were thus supported in that a higher and later F0 minimum was perceived as accent 2, but it was in fact a combination of a higher and later F0 minimum that induced more accent 2 responses, rather than either of these cues alone. And in contrast to the hypothesis, F0 minimum alignment had a slightly smaller effect on responses than F0 minimum height. The importance of alignment will be discussed in detail below.

Examining the alignment of the HTP (Figure 3) shows that accent 2 responses increase as the HTP is delayed, that is, aligned later (40% through the vowel, step 4), (Figure 6). This is the case even though the F0 minimum alignment is not later. Thus, if the initial drop from H to L gets late enough in the vowel the perception of accent 2 is induced. This shift occurs regardless of the F0 minimum alignment. Figure 3 shows that making the HTP later creates a high plateau across most of the stressed vowel which induces accent 2 identification as well. It seems that listeners use the H tone for accent 2 identification only when it is aligned late enough into the stressed vowel. In these responses, this occurred around stimulus 3 or 4, where the HTP is 48 ms (28%) into the vowel. This is in line with findings on the perception of F0 changes at various points in the segmental string (House, 1990, 1996, 2004). During the transition from a consonant to a following vowel there is a high level of spectral change, which decreases listeners' sensitivity to F0 changes at these points.

House showed that for the F0 to be perceived as a high tone, the high F0 has to occur past the point of spectral change, and into the middle of the vowel. If a tone is to be perceived as falling, the fall needs to start 30–50 ms after vowel onset. It is the F0 during the middle of the vowel that determines what the listener perceives. If this is the case, the initial H is only salient for the perception of accent 2, and is not far enough into the vowel to be a salient cue for accent 1 perception.

Finally, we turn to the F0 maximum alignment series. Once again, the later the F0 maximum is aligned, the more accent 2 responses there are. According to the odds ratio of the coefficient, this is the manipulation that had the greatest effect on responses. The alignment of the HTP in the HTP manipulation steps 1 and 2 is comparable to the alignment of the HTP in the F0 maximum alignment steps 5 and 6 (Figure 2). However, comparing responses, F0 maximum alignment steps 5 and 6 have many more accent 2 responses (50–70%) than HTP stimuli 1 and 2 (30–40%). The main difference between these stimuli is that in these F0 maximum alignment stimuli, the F0 minimum alignment is also later, past the end of the vowel (and thus the slope of the fall is also shallower). It appears that this combination of later landmarks is key to changing the perception of the accents, such that when the whole contour is later, accent 2 is most likely to be identified.

This finding—that a later alignment of more than one F0 landmark increases accent 2 responses more than a later alignment of just one F0 landmark—provides an insight into the difference in responses between F0 minimum alignment steps 1 and 2 (Figure 5) and F0 maximum alignment steps 5 and 6 (Figure 7, with F0 maximum 7% and 20%, respectively, into the vowel). Even though these stimuli have comparable F0 minimum alignment, the later F0 maximum alignment shifts majority responses to accent 2. The different responses suggest again that F0 minimum alignment (and the subsequent change in slope) is not sufficient to induce a strong change in responses (as in Figure 5, left panel). F0 minimum alignment can be used by listeners to aid in accent identification but it seems to be salient only in combination with another cue, the alignment of F0 maximum. Combined, these responses suggest that the initial H (as in HTP stimuli 4–6) is a more salient cue compared to the F0 minimum alignment cue, further supporting the notion that the initial H is only perceptible in accent 2. This is discussed further in the following section.

Listeners can perceive this initial high tone when the slope is shallow as in F0 minimum height level 5 or timing level 5, or when there is a high plateau (as in HTP stimuli 4–6). Accent 2, then, is perceived when there is an initial high tone, or a later contour (as in F0 maximum alignment stimuli 5 and 6). Overall, the cue interactions suggest that accent 1 is characterized by an early F0 fall (a fall that begins at or before vowel onset), and furthermore, that the slope of the F0 fall has to be steep enough and/or the F0 minimum height has to be low enough for it to be perceived as a fall.

4 Discussion

The aim of the perception experiment conducted here was to examine which acoustic cues listeners were sensitive to in identifying accent 1 and accent 2 words in the Trøndersk dialect of Norwegian. Specifically, the study sought to determine whether changes in F0 minimum alignment, F0 maximum alignment, F0 minimum height, and HTP alignment contribute to the shifts in accent identification. In the production study, Kelly and Smiljanić (2017) found that in broad focus, accent 2 has a higher and later F0 minimum, later F0 maximum alignment and later HTP alignment compared to accent 1. The perception experiment conducted here showed that listeners were most sensitive to an early F0 fall, leading to a majority of accent 1 identifications, and an initial salient high tone, which led to a majority of accent 2 identifications. A higher F0 minimum and later F0 minimum alignment, or a later alignment of HTP or F0 maximum, all led to more accent 2 responses.

Although each cue impacted accent responses, a later F0 minimum alignment alone did not consistently increase the likelihood of accent 2 responses. While at every level of F0 minimum

height there was a significant effect of F0 minimum alignment on responses, there was still a majority accent 1 response for the lowest two levels of height, regardless of how late alignment was. The apparent lack of importance of the F0 minimum alignment on its own (that is, when not combined with a higher F0) in perception of the contrast may be surprising in light of production studies that focus on this as the main correlate of the contrast in the Trondheim variety of Trøndersk. For example, Wetterlin (2010) describes the timing of the L as the main difference between the accents. Previous research has found that while a phonological contrast may have multiple acoustic cues, these are not all equally salient in the perception of the contrast (Abramson & Lisker, 1985; Francis, Ciocca, Ma, & Fenn, 2008; Haggard, Ambler, & Callow, 1970; Whalen, Abramson, Lisker, & Mody, 1993). While F0 minimum alignment is consistently different between the accents, this does not mean that listeners rely on it to differentiate the accents. In fact, examining the relative contributions of the manipulations here indicates that F0 minimum alignment is the cue that had the weakest effect on responses, while F0 maximum alignment had the greatest effect. The perception results here thus indicate that listeners rely on other cues to differentiate the accents. The later alignment of the F0 minimum in accent 2 could simply be a consequence of the presence of a high tone before it, or of a later alignment of the entire HL contour for accent 2. The results here highlight that finding consistent acoustic correlates of the tonal contrasts in production does not necessarily mean that listeners use them in perception. In fact, it is the combined late F0 minimum alignment with a higher F0 minimum that lead to accent 2 identification in Trøndersk.

The perception results further indicated that an early fall to the F0 minimum was the most salient cue for accent 1 identification, while a later alignment of the entire HL contour was important for accent 2 identification. The results for accent 2 are somewhat similar to what Bruce (1977) found for Stockholm Swedish, where the F0 fall had to start at least 25% into the vowel to induce accent 2 responses. In the current study, the perception of an initial high tone (due to a high plateau or a very shallow fall) was also necessary for accent 2 identification. While it is tempting to use these perception results to conclude that only accent 2 has an initial phonological high tone and that the contrast is privative and not one of alignment alone, this assumption may be premature. The perception results simply indicate that since accent 1 has an earlier fall, it will be perceived differently from accent 2, not necessarily proving that accent 1 does not have an initial high tone⁴. However, the findings for accent 1 here may be interpreted in the context of what House calls “recoding from movement to levels” (House, 1990, p.75). If the onset of a fall takes place before the area of maximum spectral change (the transition from a consonant to a vowel, up to about 25 ms into the vowel), the fall could be perceived as a low tone. Similarly, Remijsen (2013) notes that “[i]f an F0 change...sets in during the onset consonant or at the beginning of the vowel...it would be perceived in terms of level targets, with the end target likely to predominate” (p. 324). Following this, the current results suggest that the initial, early-aligned H is not perceived as indicative of accent 1, rather, it is the early F0 fall that listeners use to identify this accent. The question arises then of whether the initial fall in accent 1 is a fall from an early phonological high tone (as posited by Kristoffersen, 2007), making the contrast one of timing where both accents are HL, or simply a fall from a phonetic high that is used by speakers to make the low target salient (the explanation favored by Nilsen, 1992). This latter explanation would make the accent contrast in this dialect a privative one between L and HL. The current data do not allow us to distinguish between these possibilities, because both of them include an early F0 maximum, whether it is phonological or phonetic. However, these alternatives need not contradict one another. If the L is the most salient feature of accent 1, and the fall enhances this salience, this does not preclude the presence of a high tonal target occurring before it. The accent contrast in the Oslo dialect has been characterized as accent 1 being L and accent 2 HL, where accent 1 has no initial fall (Fintoft, 1970; Fintoft & Mártony, 1964; Kristoffersen, 2006). Thus, the question arises: if Trøndersk has no initial

phonological H in accent 1, why is an initial fall consistently found in this accent? It appears that the Oslo dialect does not need a phonetic high target to make the following low tone salient, thus perhaps indicating that the initial fall in accent 1 in Trøndersk is in fact part of the phonological makeup of the tonal contour rather than a phonetic effect. This would suggest that there is a H tone preceding the target (accent 1) word in Trøndersk, leading to an initial fall in the target word. Thus, the conclusion here is reminiscent of that in Ritter and Grice (2015) on German intonation, whereby tonal movement preceding the stressed syllable is important. The fact remains that the initial F0 of accents 1 and 2 is perceived as qualitatively different, so while accent 2 undoubtedly has an initial H, it is an initial fall that seems to be most relevant for listeners to perceive accent 1.

Accent 1 is identified when the F0 fall is early, its slope is steep enough and/or the F0 minimum height is low enough that it will be perceived as a fall. In contrast, when the F0 minimum height is lower or alignment is later, and thus the slope of the fall is shallow, accent 2 is identified. However, this does not mean that accent 2 has to have a shallow fall. A later alignment of F0 maximum induced accent 2 responses without any change in the slope, and a later HTP alignment also induced accent 2 responses, and these stimuli had a steep F0 fall. Therefore, it appears it is not directly the slope that affects responses, rather, it is whether the other cues combine to create the perception of a fall early in the vowel (accent 1) or an initial high and later fall (accent 2). Research on lexical tone languages has examined the roles of F0 height and slope in tone perception (e.g., Abramson, 1978; Connell, 2000; Gandour, 1983; Hombert, 1976). There is evidence for the importance of the F0 slope or contour, for example, in Taiwanese (Lin & Repp, 1989), and Thai (Mixdorff, Luksaneeyanawin, Fujisaki, & Charnavit, 2001; Zsiga & Nitsaraj, 2007). In particular, Zsiga and Nitsaraj (2007) note that it is not simply slope but the alignment of F0 landmarks with specific segmental landmarks (in this case, mora boundaries) that allow for tone identification in Thai. Gandour and Harshman (1978) found that direction and slope of F0 are used in synthesized pitch discrimination for speakers of Thai and Yoruba. Overall, both F0 height and movement appear to be important cues for pitch perception. In addition, some cues found in production are simply acoustic correlates of other, more salient cues, and not always relevant to the listener. In this study, a higher or later F0 minimum induces the perception of a late initial high tone or plateau, inducing accent 2 identification. This combination of cues resulted in a shallower slope of the fall. However, in the stimuli with a late HTP, also identified as accent 2, there is a steeper fall, yet this does not mean the accent is perceived as accent 1. Thus, it appears that in these stimuli, while slope corresponds to other landmarks, it in and of itself does not appear to be important in accent identification.

With regard to the identification curves shown in Figures 5–7, it is important to note that there are no 100% accent 1 responses or accent 2 responses. This is likely due to the fact that other cues that were not being manipulated for the experiment (including vowel duration) were set at specific, intermediate values, thus possibly making the stimuli ambiguous or unnatural in these respects.⁵ The stimuli used in this experiment were created using just one token base (accent 1) which could have introduced other, non-F0 factors may be biasing listeners toward one response type over another. If there were other correlates of the accent contrast, such as voice quality, these could result in listener bias toward accent 1. To ensure against this, a small number of tokens was created from an accent 2 base, and these were tested on five participants. The same controls and manipulations were made on this base, and for these participants, their responses for the accent 2 base did not differ from their responses for the accent 1 base for the comparable manipulations. This indicates that using accent 1 as a base token for all manipulations likely did not bias responses. While a future investigation will include a full set of tokens created from both accent 1 and accent 2 words, the results here, nonetheless, provide some suggestions about the nature of the lexical tonal contrast in perception as well as some insights into the phonological characterization of the contrast.

The identification curves are also more continuous than categorical. The results are not sigmoid curves with an abrupt change from one accent category to the other, as found for consonant contrasts, for instance, rather they are more gradual slopes (both for individual participants and across averages). Previous work on the perception of pitch in language has shown mixed results. Some research has found that lexical tone in Mandarin is perceived categorically (Hallé, Chang, & Best, 2004; Xu, Gandour, & Francis, 2006). Francis, Ciocca, and Ng (2003) showed that in Cantonese, level tones were perceived more continuously, similar to vowels, while contour tones were perceived more like consonants, that is, more categorically. Generally, at the crossover point from one category to another (determined by an identification task), there is a peak in discrimination accuracy, indicating that listeners distinguish stimuli across the category boundary better than within the category (Repp, 1984). However, work on the perception of lexical tone does not always find such correspondences between identification and discrimination functions (DiCanio, 2012; Francis et al., 2003). The results from a brief discrimination task conducted with a subset of the listeners from this experiment indicated that discrimination accuracy was not higher across the category boundaries than within categories. In a larger discrimination study related to the current investigation, the results of a discrimination experiment by Kelly and Dogil (under review) also support an analysis where the accents are perceived continuously rather than categorically. Therefore, both the identification and discrimination results on this dialect indicate a continuous perception of the lexical accents.

In summary, this investigation provided novel insights into the lexical tonal accent contrast in Trøndersk approaching it from the perspective of perception. The results demonstrated that accent 1 is perceived when there is an initial fall, and accent 2 when there is a salient initial H with later alignment. This provides evidence that the contrast in this dialect is predominantly one of timing rather than of tonal makeup. The results also add to the findings from tonal languages in that the F0 accent contrasts are perceived continuously rather than categorically. The production results on the accent contrast guided the manipulation of stimuli to examine which cues listeners use to distinguish the accents. Previous work described F0 minimum alignment as the main cue to the contrast, but the perception experiment here revealed that this alone is not the most relevant cue, rather the timing of the F0 fall and the initial F0 height are what listeners use to distinguish the accents. The results thus revealed an interesting mismatch between the way speakers produce the lexical contrasts and the use of these cues in perception, highlighting the need to examine production and perception patterns in parallel. Finally, this work demonstrated that, in this dialect, a cluster of F0 cues in combination, rather than any one individually, contributed to the perception of the contrast.

Acknowledgements

Thanks to Wim van Dommelen for the use of the recording studio at NTNU, Trondheim. We thank attendees of TAL 2014 for comments, particularly Bert Remijsen. Special thanks to Gjert Kristoffersen for help in setting up the production experiments and for comments on this paper. Thanks also to Scott Myers, Megan Crowhurst and Harvey Sussman, as well as Grzegorz Dogil, Katrin Schneider and all at the IMS Stuttgart. This research was conducted with the support of the National Science Foundation Doctoral Dissertation Research Improvement Grant No. 1322700. The authors would also like to thank reviewers at various stages of the manuscript for their insights and guidance.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research used for this investigation was conducted with the support of the National Science Foundation Doctoral Dissertation Research Improvement Grant No. 1322700.

Notes

1. The target sentence was: *Det var linet i en film, men ikke i et stykke*. “There was the linen in a film, but not in a play.”
2. The reasons for this are unclear: they may have a hearing problem they were unaware of, or misunderstood the task. However, if this were the case, it would be expected that they would perform at chance level. Scoring so low means they are misidentifying the accents (they still had a high response rate; it was not the case that they failed to respond) suggesting that they switched response keys. It is also possible although less likely that their perception was affected by another dialect, although this did not show up in their responses to the background questionnaire.
3. R code for the independent variable F0 minimum height is as follows: `glmer (response ~ MinHeight + (1|Participant), family = binomial, data = minimum)`. R code for F0 minimum alignment: `glmer (response ~ MinAlign + (1|Participant), family = binomial, data = minimum)`. R code for F0 maximum alignment: `glmer (response ~ MaxAlign + (1|Participant), family = binomial, data = maximum)`. R code for HTP: `glmer (response ~ HTP + (1|Participant), family = binomial, data = htp)`.
4. A perception experiment involving a stimulus with just a flat contour (L tone) could determine whether listeners classify this as a natural token of accent 1 or as no natural accent, and thus shed further light on whether accent 1 requires an initial high tone or not.
5. It is also possible that the manipulated stimuli simply did not sound completely natural to the listeners, resulting in the lack of 100% responses in either direction.

References

- Abramson, A. (1978). Static and dynamic acoustic cues in distinctive tones. *Language and Speech*, 21(4), 319–325.
- Abramson, A. S., & Lisker, L. (1985). Relative power of cues: F0 shift versus voice timing. In V. Fromkin (Ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged* (pp. 25–33). New York, NY: Academic Press.
- Almberg, J. (2004). Tonal differences between four Norwegian dialect regions – some acoustic findings. In G. Bruce, & M. Horne (Eds.), *Nordic prosody IX* (pp. 19–28). Lund, Sweden: Peter Lang.
- Bjerrum, M. (1948). *Felstedmaalets tonale accenter*. Aarhus, Denmark: Universitetsforlaget i Aarhus.
- Boersma, P., & Weenink, D. (2011). Praat: doing phonetics by computer [Computer program]. Version 5.3.03. Retrieved from <http://www.praat.org/>
- Bruce, G. (1977). *Swedish word accents in sentence perspective*. Lund, Sweden: Travaux de L’Institut de Linguistique de Lund 12.
- Chang, S.-E. (2013). Effects of fundamental frequency and duration variation on the perception of South Kyungsang Korean tones. *Language and Speech*, 56(2), 211–228.
- Connell, B. (2000). The perception of lexical tone in Mambila. *Language and Speech*, 43(2), 163–182, 2000.
- Derksen, R. (1966). *Metatony in Baltic*. Amsterdam, The Netherlands: Rodopi.
- DiCanio, C. T. (2012). Cross-linguistic perception of Itunyoso Trique tone. *Journal of Phonetics*, 40(5), 672–688.
- Efremova, I. B., Fintoft, K., & Ormestad, H. (1963). Intelligibility of Tonic Accents. *Phonetica*, 10(3–4), 203–212.
- Elstad, K. (1978). Det nordnorske cirkumsflekstonemet [The Northern Norway Circle Spotstone]. In E. Gårding, G. Bruce, & R. Bannert (Eds.), *Nordic Prosody* (pp. 165–174). Lund, Sweden: Peter Lang. [In Swedish.]
- Fintoft, K. (1970). *Acoustical analysis and perception of tonemes in some Norwegian dialects*. Oslo, Norway: Universitetsforlaget.
- Fintoft, K. (1987). Toneme patterns in Norwegian and in Swedish dialects. In R. Channon, & L. Shockey (Eds.) *In Honor of Ilse Lehiste* (pp. 33–50). The Netherlands: Foris Publications.
- Fintoft, K., & Mártony, J. (1964). Word Accent in East Norwegian. *STL-QPSR*, 5(3), 8–15.
- Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics*, 36(2), 268–294.

- Francis, A. L., Ciocca, V., & Ng B. K. C. (2003). On the (non)categorical perception of lexical tones. *Perception and Psychophysics*, 65(7), 1029–1044.
- Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, 11(2), 149–175.
- Gandour, J., & Harshman, R. (1978). Cross-language differences in tone perception. *Language and Speech*, 21(1), 1–33.
- Gårding, E. (1973). The Scandinavian word accents. In *Working Papers 8*. Lund, Sweden: Phonetics Laboratory, Lund University.
- Gårding, E., & Lindblad, P. (1975). Constancy and variation in Swedish word accent patterns. In *Working Papers 7* (pp. 36–100). Lund, Sweden: Phonetics Laboratory, Lund University.
- Grønnum, N., & Basbøll, H. (2003). Two-phased stød vowels – a cognitive reality? In M. Heldner (Ed.), *Proceedings from Fonetik 2003; Phonum 9* (pp. 33–36). Löfvånger, Sweden. Retrieved from <https://pdfs.semanticscholar.org/b5b1/070d0d74d9e00503cd967b13b048680e0f8f.pdf>
- Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge, UK: Cambridge University Press.
- Haggard, M., Ambler, S., & Callow, M. (1970). Pitch as a voicing cue. *Journal of the Acoustical Society of America*, 47(2), 613–617.
- Hallé, P. A., Chang, Y. C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, 32(3), 395–421.
- Hombert, J.-M. (1976). Perception of tones in bisyllabic nouns in Yoruba. *Studies in African Linguistics, Supplement 6*, 109–122.
- House, D. (1990). *Tonal perception in speech*. Lund, Sweden: Lund University Press.
- House, D. (1996). Differential perception of tonal contours through the syllables. In *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP)*, Philadelphia, USA, 3–6 October 1996 (pp. 2048–2051). Retrieved from <http://www.speech.kth.se/prod/publications/files/3166.pdf>
- House, D. (2004). Pitch and alignment in the perception of tone and intonation: Pragmatic signals and biological codes. In *Proceedings of Tonal Aspects of Languages (TAL) 2004*, Beijing, China, 28–30 March 2004 (pp. 93–96).
- Hualde, J. I. (1991). *Basque phonology*. London, UK: Routledge.
- Hualde, J. I. (2012). Two Basque accentual systems and the notion of pitch-accent language. *Lingua* 122(13): 1335–1351.
- Karins, A. K. (1996). *The prosodic structure of Latvian*. PhD Dissertation. University of Pennsylvania, USA.
- Kelly, N. E. (2015). *An experimental approach to the production and perception of Norwegian tonal accent*. PhD Dissertation. University of Texas at Austin, Austin, Texas, USA.
- Kelly, N. E., & Dogil, G. (under review). Accent specification in East Norwegian: The perceptual magnet effect.
- Kelly, N. E., & Smiljanić, R. (2014). The effect of focus on Norwegian tonal accent. In *Proceedings of Tonal Aspects of Languages (TAL) 2014*, Nijmegen, The Netherlands, 13–16 May 2014 (pp. 95–99).
- Kelly, N. E., & Smiljanić, R. (2017). The Effect of Focus and Phrase Position on East Norwegian Lexical Tonal Accents. *Phonetica*, 74(4), 193–230.
- Kim, G.-R. (1988). *The pitch accent system of the Taegu dialect of Korean with emphasis on tone sandhi at the phrasal level*. PhD Dissertation. University of Hawaii, USA.
- Kock, A. (1884/1885). *Språkshistoriska undersökningar om svensk accent*. [Historical investigations of Swedish tonal accent from a linguistic perspective] Volume II. Lund, Sweden: Gleerup. [In Swedish.]
- Kristoffersen, G. (2000). *The phonology of Norwegian*. Oxford, UK: Oxford University Press.
- Kristoffersen, G. (2006). Tonal melodies and tonal alignment in East Norwegian. In G. Bruce, & M. Horne (Eds.), *Nordic Prosody IX* (pp. 157–166). Frankfurt am Main, Germany: Peter Lang.
- Kristoffersen, G. (2007). Dialect variation in East Norwegian tone. In T. Riad, & C. Gussenhoven (Eds.), *Tones and tunes, Volume 1: Typological and comparative studies in word and sentence prosody* (pp. 91–111). Berlin, Germany: Mouton de Gruyter.
- Lin, H. B., & Repp, B. H. (1989). Cues to the perception of Taiwanese tones. *Language and Speech*, 32(1), 25–44.
- Lorentz, O. (1981). Adding tone to tone in Scandinavian dialects. In T. Fretheim (Ed.), *Nordic Prosody II* (pp. 166–180). Trondheim, Norway: Tapir.

- Mixdorff, H., Luksaneeyanawin, S., Fujisaki, H., & Charnavit, P. (2002). Perception of tone and vowel quality in Thai. Paper presented at the 7th International Conference on Spoken Language Processing, Denver Colorado, September, 2002.
- Nilsen, R. A. (1992). *Intonasjon i interaksjon- sentrale spørsmål i norsk intonologi* [Intonation in interactional questions in Norwegian intonology]. PhD Dissertation. University of Trondheim, Norway. [In Norwegian.]
- Oftedal, M. (1952). On the origin of the Scandinavian tone distinction. *Norsk Tidsskrift for Sprogvidenskap* 16(1): 201–225.
- Peirce, J.W. (2007). PsychoPy – Psychophysics software in Python. *Journal of Neuroscience Methods*, 162(1–2), 8–13.
- Pierrehumbert, J. B., & Beckman, M. (1988). *Japanese tone structure*. Cambridge, MA: MIT Press.
- R Development Core Team. (2008). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org>
- Remijsen, B. (2013). Tonal alignment is contrastive in falling contours in Dinka. *Language* 89(2), 297–327.
- Repp, B. (1984). Categorical perception: Issues, methods and findings. In N. Lass (Ed.), *Speech and Language: Advances in Basic Research and Practice (Volume 10)* (pp. 243–335). New York, NY: Academic Press.
- Riad, T. (1998). Towards a Scandinavian accent typology. In W. Kehrein, & R. Wiese (Eds.), *Phonology and morphology of the Germanic languages (Linguistische Arbeiten 386)* (pp. 77–109). Tübingen, Germany: Niemeyer.
- Riad, T. (2003). Diachrony of the Scandinavian accent typology. In P. Fikkert, & H. Jacobs (Eds.), *Development in prosodic systems (Studies in generative grammar, Volume 58)* (pp. 91–144). Berlin, Germany/New York, NY: Mouton de Gruyter.
- Riad, T. (2006). Scandinavian accent typology. *Sprachtypologie und Universalienforschung (STUF)*, 59(1), 36–55.
- Ritter, S., & Grice, M. (2015). The role of tonal onglides in German nuclear pitch accents. *Language and Speech* 58(1), 114–128.
- Segerup, M. (2003). Word accent gestures in West Swedish. In M. Heldner (Ed.), *Proceedings from Fonetik 2003; Phonum 9* (pp. 25–28). Umeå, Sweden: University of Umeå. Retrieved from <https://lup.lub.lu.se/search/ws/files/5960204/1624485.pdf>
- Segerup, M. (2004). Gothenburg Swedish word accents: A fine distinction. In P. Branderud, & H. Traunmüller (Eds.), *Proceedings from Fonetik 2004* (pp. 28–31). Stockholm, Sweden: Department of Linguistics, University of Stockholm. Retrieved from http://www2.ling.su.se/fon/fonetik_2004/segerup_fonetik2004.pdf
- Senn, A. (1966). *Handbuch der litauischen Sprache* [Handbook of the Lithuanian language]. Heidelberg, Germany: Carl Winter. [In German.]
- Smiljanić, R. (2006). Early vs. late focus: Pitch-peak alignment in two dialects of Serbian and Croatian. In L. Goldstein, D.H. Whalen, & C. Best (Eds.), *Papers in laboratory phonology (Volume 8)* (pp. 495–518). Berlin, Germany: Mouton de Gruyter.
- Storm, J. (1884). *Norvegia. Tidsskrift for det norske folks maal og minder* [Norvegia. Journal of the Norwegian people's goals and memories]. Kristiania, Norway: Grøndahl and Søn. [In Norwegian.]
- Van Dommelen, W. A. (2002). Toneme realization in two North Norwegian dialects. *Proceedings of Fonetik*, 44(1), 21–24.
- Van Dommelen, W. A., & Nilsen, R. A. (2003). Toneme realization in two East Norwegian dialects. *Proceedings of Fonetik, PHONUM 9*, 21–24.
- Vanvik, A. (1957). Norske tonelag [Norwegian tonemes]. *Maal og Minne* 92–102. [In Norwegian.]
- Wetterlin, A. (2010). *Tonal Accents in Norwegian: Phonology, Morphology and Lexical Specification*. Berlin, Germany: Walter de Gruyter.
- Whalen, D. H., Abramson, A. S., Lisker, L., & Mody, M. (1993). F0 gives voicing information even with unambiguous voice onset times. *Journal of the Acoustical Society of America*, 93(4), 2152–2159.
- Xu, Y., Gandour, J., & Francis, A. (2006). Effects of language experience and stimulus complexity on categorical perception of pitch direction. *Journal of the Acoustical Society of America*, 120(2), 1063–1074.
- Zsiga, E., & Nitisaroj, R. (2007). Tone features, tone perception, and peak alignment in Thai. *Language and Speech*, 50(3), 343–383.