EXAMINING LEXICAL TONAL CONTRAST IN NORWEGIAN USING INTONATION MODELLING

Niamh Kelly[1] & Katrin Schweitzer[2]

[1]The University of Texas; [2]University of Stuttgart
niamh.kelly@utexas.edu, katrin.schweitzer@ims.uni-stuttgart.de

ABSTRACT

The lexical accents of one dialect of East Norwegian were examined using a parametric intonation model. The purpose of this investigation was to examine the tonal contrast using a different method, whereby analyses can be done automatically and additional acoustic cues examined. The disyllabic contrast of the Trøndersk dialect had been previously described using F_0-measurements in Praat, but the current investigation uses a parametric intonation model, PaIntE [21]. The results support the description of accent 1 and accent 2 having a HL contour, with accent 2 having a later alignment with the segmental string. Moreover, accent 2 was found to have a higher maximum and a smaller amplitude of the fall following the peak. The accents did not differ with respect to their rise or the steepness of the rise or fall. This research demonstrates how different approaches to linguistic analyses can inform one another.

Keywords: Norwegian, lexical accent, parametric intonation model

1. INTRODUCTION

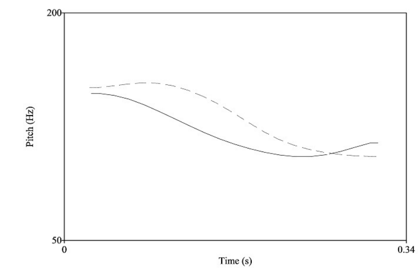
Tonal accent is a prosodic pattern found on main stressed syllables. Languages with tonal accent or pitch accent are described as "a class of stress languages where words contrast in the tonal melody that is associated with the stressed syllable" [16] and as having "invariant tonal contours on accented syllables" [14]. This means that the pitch contour on a word can change the word's meaning.

Some languages that exhibit this phenomenon - in various ways - are Lithuanian [28, 9], Latvian [17, 8], Serbian and Croatian [29], Japanese [23], and Basque [15], and some varieties of Dutch and German [12].

1.1. Norwegian Tonal Accent

Norwegian and Swedish also have a tonal accent contrast [10, 11, 5, 25, 19]. The tonal accent contrast is the same feature in both languages, and is thought to have arisen from a contrast between monosyllables and polysyllables in Old Norse [22, 19]. Stress placement and morphology are the main determiners for which accent a word takes [26, 13, 6]. The phonetic realisation of the contrast varies depending on the dialect. The Trøndersk dialect of Norwegian is an East Norwegian dialect spoken in the centre of the country, around the city of Trondheim. The accent contrast of this dialect has been described as a contrast in the alignment of the tones with the segmental string. Accent 1 and accent 2 have been described as both having a HL contour (i.e. a high target followed by a low target) but accent 2 has a later timing [20]. A recent acoustic analysis corroborated this, as shown in the following figure [18] (based on one representative token of each accent):

Figure 1: *Trøndersk accents in broad focus. (Solid line is accent 1; dashed line is accent 2.)*



Specifically, accent 2 was found to have a higher F_0 minimum, later timing of the F_0 maximum and minimum, later timing of where the F_0 maximum first starts to drop, and longer duration for the unstressed (final) vowel, than accent 1.

1.2. Investigating tonal accent

The studies mentioned above are based on F_0 measurements that were chosen specifically to analyse the tonal contrast in Norwegian. In the current paper, we exploit a parametric intonation model, PaIntE [21]. The model was created as a data-based method of intonation generation and has been used to analyse sentence intonation in various studies (e.g. [27]). To our knowledge, it has not been used to analyse lexical tone before. The model's parameters are linguistically meaningful so a large amount of data can be analysed automatically and the results can be directly interpreted in linguistic terms.

2. METHODOLOGY

The current investigation examines the tonal contrast with a computational intonation model using data from [18] together with additional recordings. This approach has several interesting aspects: firstly, it allows for a comparison to previous studies on the timing of the two accent types, thereby allowing a comparison of the two methods. Secondly, the model offers insights into characteristics of accent shape which have not been analysed before and might contribute to the tonal contrast. Thirdly, since we include data from three new speakers which had not been examined in [18], our analyses shed light on the robustness of the acoustic cues of the tonal contrast across speakers.

2.1. Recording

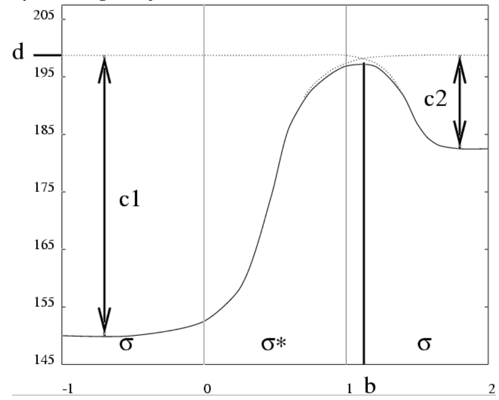
In the following, we will describe the recordings for the complete data set comprised of Kelly & Smiljanic's [18] data, as well as the three additional speakers who were recorded following the same procedure.

The tonal accent contrast was examined using high quality recordings of words with either accent 1 or accent 2 that had been produced in sentence-medial position. Stimuli consisted of disyllabic words with initial stress and either accent 1 or accent 2, with the vowel /i/ in the stressed syllable. For all words, the sounds on either side of the /i/ were voiced sonorants, for example *limet* 'the glue' (accent 1), *minne* 'the memory' (accent 2). The target words were sentence-medial, preceded by two unstressed syllables. In order to achieve broad focus on the target word, two words later in the sentence were contrastively focused with one another, and not with the target word, for example:
*Det var <u>limet</u> i en **film**, men ikke i et **stykke**.*
"There was the glue in a film, but not in a play."

There were five target words for accent 1, each repeated 3 times, and two target words for accent 2, each repeated seven or eight times. This gave 15 tokens per accent per speaker, a total of 390 tokens for the current experiment (15 tokens x 2 accents x 13 speakers). Three tokens were discarded due to the speaker pausing during it or saying a different word.

Taken together, participants were 13 native speakers (nine females, four males), aged 18-45, of the Trøndersk dialect of Norwegian. They were recruited by means of posters and fliers and were paid 110 NOK (approx. US$20) for 40 minutes, the duration of the experiment.

Figure 2: *The PaIntE model function and its parameters. Parameters a1 and a2, reflecting the steepness of the rise and the fall, are not displayed. Figure from [21].*



Recordings took place in the sound studio at NTNU, Trondheim. Sentences were presented in a randomized order one at a time in slideshow format, with the participant in control of when to move to the next sentence. Participants were seated in front of a desktop computer and read the sentences aloud into a microphone while being recorded through the program Adobe Audition. The sentences were segmented into separate sound files for analysis and automatically annotated for phones, syllables and words using forced alignment [24].

2.2. PaIntE: Parametric Intonation Modelling

To analyse the acoustic realisation of the tonal accents on the target words, the pitch contour was analysed with a parametric intonation model, PaIntE [21]. The model employs a function term to approximate a peak in the F_0 contour, comprising 6 parameters which are set by the model so that the resulting curve fits the actual F_0 shape best. The six parameters are linguistically meaningful: they specify the steepness of the rise before, and the fall after the peak (parameter *a1* and *a2*, respectively), the temporal alignment of the peak (*b*), the amplitude of the rise / fall (*a1/a2*) and the absolute peak height *(d)*. Figure 2 illustrates the parameters. They are calculated over the span of the accented syllable ($\sigma*$) and its immediate neighbours. The x-axis indicates time (normalised for syllable duration, i.e. the syllable bearing the accent spans from 0 to 1) and the y-axis displays the fundamental frequency in Hertz.

2.3. Statistical Analysis

Before carrying out the statistical analysis, all accent tokens where the PaIntE approximation failed to fit a curve were excluded. This was the case for only 7 tokens, so 380 observations went into the analysis, 188 accent 1 tokens, 192 accent 2 tokens.

To investigate possible relationships between characteristics in the shape of the two accents and the accent type, for each PaIntE parameter, we fit a linear mixed model predicting the parameter by the fixed factor *accent.type* (encoding whether the word carries accent 1 or 2) while controlling for the random factors *speaker* (the speaker identity), *word* (the target word), *sex* (the speaker's sex), *hometown* (the speaker's hometown), *other.lang* (other languages that the speaker speaks well), *vowel.length* (whether the target syllable had a long vowel ("limet" , "line", "linet", "slimet", and "smilet") or a short vowel ("glimtet", "minne").

For each analysed parameter, we determined the best fitting linear mixed model by carrying out model comparisons using likelihood ratio tests (in a similar fashion to [2, 3, 30]). After determining the best model with only random factors, we added *accent.type* as a fixed factor and tested whether it improved the model. A model was considered to be better than the predecessor if a) the improvement was significant ($p(\chi^2) < 0.05$) and if b) the AIC value (Akaike's information criterion, see [1]) was at least 2 points smaller [7].[1] The p-values reported were obtained by comparing the winning model to the model without accent-type.

3. RESULTS

3.1. Late peak in accent 2

To investigate the alignment of the peak with the segmental material, we predicted PaIntE parameter *b* by accent-type. The best null model comprising only random factors contained *speaker* and *word*. The fixed factor *accent.type* improved the model significantly ($p(\chi^2) < 0.005$).

The coefficient for *accent.type:accent2* was a positive one ($\beta = 0.38851 \pm 0.09035$ SE, standard error), i.e. the timing of the peak for accent 2 is about 0.39 units in syllable-normalised time (in other words, about 39% of the syllable duration) later than for accent 1.

3.2. High peak in accent 2

To detect any differences in the absolute peak height of the two accents, we predicted PaIntE parameter *d* by accent type. The best null model for this task comprised the random factors *speaker*, *word* and *sex*, as well as by-speaker random slopes for *accent.type*. The model incorporating *accent.type* as a fixed effect significantly improved the null model ($p(\chi^2) < 0.05$).

The coefficient for *accent.type:accent2* was a positive one ($\beta = 10.092 \pm 4.51$), i.e. the peak in accent 2 is about 10.1Hz higher than in accent 1.

3.3. Reduced fall in accent 2

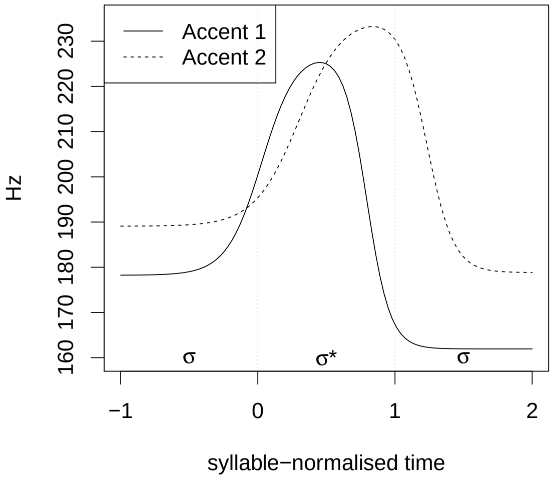
Accent 2 was previously found to have a higher F_0-minimum than accent 1 [18]. None of the parameters in the PaIntE model function offer direct access to that measure, since the function models the peak, i.e. the H-target, not the low tone. However, together with the result on peak height described above, PaIntE parameter *c2*, denoting the height of the fall, can be exploited to gain insight into the nature of the higher F_0-minimum: if the contours of accent 1 and 2 have a similar amplitude of the fall, then, together with the knowledge that accent 2 has a higher peak (i.e. a higher F_0-maximum than accent 1), we can infer that the F_0 minimum of accent 2 is higher than the F_0-minimum of accent 1, as well. A higher F_0 minimum can also be inferred if the fall in accent 2 is significantly smaller than the one in accent 1. However, if the fall in accent 2 is significantly bigger than in accent 1, we can infer either no difference in F_0-minima or a lower minimum in accent 2.

To compare the height of the fall in the two accent types, PaIntE parameter *c2* was predicted by accent type. The best null model comprising only random factors contained again the two random factors *speaker* and *word*, as well as by-speaker random slopes for *accent.type*. The fixed factor *accent.type* improved the model significantly ($p(\chi^2) <$ 0.05). The coefficient for *accent.type:accent2* was a negative one ($\beta = -8.867 \pm 4.161$), i.e. the amplitude of accent 2's fall is about 8.9Hz smaller than the one of accent 1.

3.4. Equal rise in both accents

To investigate potential differences in the amplitude of the rise, PaIntE parameter *c1* was predicted by accent type by means of a linear mixed model. The best null model comprised random intercepts for *speaker* and *word*, as well as by-speaker random slopes for *accent.type*. However, the model with the fixed factor *accent.type* did not improve the prediction significantly. That is, the two accents do not differ in the height of their rising parts.

Figure 3: *Schematic description of the two accent types: accent shapes computed with the mean values for each PaIntE parameter.*



3.5. Equal steepness in both accents

The two last PaIntE-parameters, *a1* and *a2*, encoding the steepness of the rise and fall, respectively, did not yield a significant difference between the two accent types, either. That is, both accents are similar in how steep the F_0 contour rises before, and falls after, the peak.

3.6. Conclusion

The findings in terms of timing were in line with the descriptions from previous work on this dialect [10, 19, 18] in that accent 2 has a later F_0 peak. Some additional results were also found: accent 2 has a higher F_0 maximum than accent 1. Also, accent 2 has a smaller amplitude of the fall (from the F_0 peak) than accent 1, which, together with the result about a greater peak height confirms studies reporting a higher F_0minimum [18]. The amplitude of the rise and the steepness of both, rise and fall, did not differ significantly between the accent types.

Figure 3 displays two accents composed of the mean PaIntE values for accent 1 and 2 separately. That is, for each accent type, we averaged over each PaIntE parameter and assigned these values to the free parameters in the PaIntE function. The two resulting curves are displayed. Of course, other than in the linear mixed models, these two curves do not take speaker specific differences into account. So for instance the Hertz-values given for differences between the two accents in the descriptions of the results above are more accurate than the difference in the averaged values displayed here, because the linear mixed models control for speaker variation.

4. DISCUSSION

The current findings support previous work conducted using Praat [4] scripts, indicating that PaIntE is a reliable tool for analysing the nature of lexical accent differences automatically. That is, its linguistically meaningful parameters can shed further light on the nature of the tonal contrast by looking at features of accent shape that have not been looked at before, e.g. the amplitude of rise and fall and their steepness. The fact that the results from [18] were corroborated, even though three new speakers had been added, also indicates that the disyllabic accent contrast has robust cues across speakers. Furthermore, these speakers were from a variety of towns south and west of Trondheim, but including *hometown* as a random factor did not improve the model, indicating that these speakers are from a homogeneous group, at least in terms of the lexical accents.

Overall, we can confirm that timing is an important aspect in the accent contrast in disyllabic words. Both accents have an initial peak and subsequent fall - that is, a HL contour - with accent 2's contour being aligned later with the segmental string, as in Figure 3. Specifically our results showed that the peak of accent 2 tokens is about 39% of the syllable duration later than in accent 1 tokens. In addition, our PaIntE analysis indicates that both accent types are similar with respect to the steepness of of both the F_0 rise and fall, supporting the description of both accents having the same contour.

Our new findings, the higher peak and smaller fall for accent 2 (resulting in a higher F_0 minimum which has been reported before [18]) also indicate that accent 2 might generally be higher pitched than accent 1. A PaIntE analysis on the original data set from [18] yielded no significant effect between the accents for this measure, indicating that the larger data set used in the current analysis allowed for this new result. It remains to be determined whether this effect is robust across even more speakers.

They also offer interesting opportunities for future work: for example, it would be interesting to examine whether the tonal contrast is not solely related to timing, but also to higher landmarks by looking at other data sets and by investigating whether there is perceptual evidence for the slight but significant differences in accent height.

5. ACKNOWLEDGEMENTS

This research was supported by the National Science Foundation (Doctoral Dissertation Research Improvement Grant No. 1322700) and the German Research Association DFG (SFB 732, project INF).

6. REFERENCES

[1] Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. Petrov, B. N., Csaki, F., (eds), *2nd International Symposium on Information Theory* 267–281.

[2] Baayen, H. 2008. *Analyzing Linguistic Data*. Cambridge University Press.

[3] Baayen, R. H., Davidson, D. J., Bates, D. M. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4), 390–412.

[4] Boersma, P., Weenink, D. 2011. Praat: doing phonetics by computer [Computer program]. Version 5.3.03. http://www.praat.org/.

[5] Bruce, G. 1977. Swedish word accents in sentence perspective. *Travaux de L'Institut de Linguistique de Lund* 12.

[6] Bruce, G., Hermans, B. 1999. Word tone in Germanic languages. In: van der Hulst, H., (ed), *Word Prosodic Systems in the Languages of Europe*. Berlin: Mouton de Gruyter 605–658.

[7] Burnham, K. P., Anderson, D. R. 2002. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer 2 edition.

[8] Derksen, R. 1966. *Metatony in Baltic*. Amsterdam: Rodopi.

[9] Dogil, G., Möhler, G. 1998. Phonetic invariance and phonological stability: Lithuanian pitch accents. In: *Proceedings of the 5th International Conference on Spoken Language Processing (ICSLP)*. Sydney: 75–86.

[10] Fintoft, K. 1970. *Acoustical Analysis and Perception of Tonemes in Some Norwegian Dialects*. Oslo: Universitetsforlaget.

[11] Gårding, E. 1973. The Scandinavian word accents. In: *Working Papers 8*. Phonetics Laboratory, Lund University.

[12] Gussenhoven, C. 2004. *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

[13] Haugen, E. 1967. On the Rules of Norwegian Tonality. *Language* 43(1), 185–202.

[14] Hayes, B. 1995. *Metrical stress theory: principles and case studies*. Chicago: University of Chicago Press.

[15] Hualde, J. I. 1991. *Basque phonology*. London: Routledge.

[16] Hualde, J. I. 2012. Two Basque accentual systems and the notion of pitch-accent language. *Lingua* 122, 1335–1351.

[17] Karins, A. K. 1996. *The prosodic structure of Latvian*. PhD thesis University of Pennsylvania.

[18] 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: 95–99.

[19] Kristoffersen, G. 2000. *The Phonology of Norwegian*. Oxford: Oxford University Press.

[20] Kristoffersen, G. 2006. Tonal melodies and tonal alignment in East Norwegian. In: Bruce, G., Horne, M., (eds), *Nordic Prosody IX*. Frankfurt am Main: Peter Lang.

[21] Möhler, G. 2001. Improvements of the PaIntE model for F_0 parametrization. Technical report Institute of Natural Language Processing, University of Stuttgart. Draft version.

[22] Oftedal, M. 1952. On the origin of the Scandinavian tone distinction. In: *NTS XVI*. 201–25.

[23] Pierrehumbert, J., Beckman, M. 1988. *Japanese tone structure*. Cambridge: MIT Press.

[24] Rapp, S. 1995. Automatic phonemic transcription and linguistic annotation from known text with Hidden Markov models—An aligner for German. *Proc. of ELSNET Goes East and IMACS Workshop "Integration of Language and Speech in Academia and Industry" (Russia)*.

[25] Riad, T. 1998. Towards a Scandinavian accent typology. In: *Phonology and morphology of the Germanic languages (Linguistische Arbeiten 386)*. Tübingen: Niemeyer 77–109.

[26] Rischel, J. 1963. Morphemic tone and word tone in Eastern Norwegian. *Phonetica* 10, 154–164.

[27] Schweitzer, K., Walsh, M., Calhoun, S., Schütze, H., Möbius, B., Schweitzer, A., Dogil, G. 2015. Exploring the relationship between intonation and the lexicon: Evidence for lexicalised storage of intonation. *Speech Communication* 66(0), 65–81.

[28] Senn, A. 1966. *Handbuch der litauischen Sprache*. Heidelberg: Carl Winter: Universitätsverlag.

[29] Smiljanić, R. 2006. Early vs. late focus: Pitch-peak alignment in two dialects of Serbian and Croatian. In: Goldstein, L., Whalen, D., Best, C., (eds), *Papers in laboratory phonology (8)*. Berlin: Mouton 495–518.

[30] Winter, B. 2013. Linear models and linear mixed effects models in r with linguistic applications. arXiv: 1308.5499.