

Quantitative Modeling of Norwegian Tonal Accents in Different Focus Conditions

Hansjörg Mixdorff¹, Bistra Andreeva², Jacques Koreman³

¹ Dept. of Computer Science and Media, Beuth University of Applied Sciences, Germany

² Institute of Phonetics, Saarland University, Saarbrücken, Germany

³ Dept. of Language and Communication Studies, NTNU, Trondheim, Norway

mixdorff@beuth-hochschule.de, andreeva@coli.uni-saarland.de, jacques.koreman@ntnu.no

Abstract

The present paper is a first attempt to use the Fujisaki model to parameterize the F₀ contours of utterances containing Accent 1 and Accent 2 tonal accents in Norwegian in different focus conditions. Differences in timing and amplitude of the accent commands are found, largely corresponding to descriptions in the literature. This shows that the model can be used as a basis for manipulating stimuli which can then be used in perception tests to determine phonetic differences between different accent as well as focus types.

Index Terms: Norwegian lexical tone, focus, Fujisaki model

1. Introduction

In the Trondheim model of intonation, which was developed for East Norwegian by Fretheim in the 1980's, an Intonation Unit (IU) consists of Intonational Phrases (IPs) headed by accented syllables. An IP consists of one or more Tonal Feet¹ (F) and Prosodic Words (ω), which are two interdependent categories. The Prosodic Word is the domain of the lexical tones (Accent 1 and Accent 2) and dominates all syllables² from the stressed syllable to the word boundary. The Foot is the domain of the phrase accent contrast between F[+focal] and F[-focal] and is headed by the ω . The right boundaries of F and ω do not necessarily coincide, since F can be longer. Thus the Foot extends from one lexically accented syllable to the next and contains a tonal accent on its left edge and a phrasal accent aligned at the end of the Foot [1].

East Norwegian Accent 1 consists of a L tonal accent realized on the lexically accented syllable, while Accent 2 is realized as HL either on the accented syllable or on the accented syllable and the syllable following it. Accent 2 could be said to have a later alignment than Accent 1. In both cases, the tonal accent is followed by a H phrasal accent, associated with the last syllable of the tonal Foot, and also with the Foot itself. The size of the rise corresponding to the H phrasal accent is related to the degree of prominence. Statements end

with an L% boundary tone associated at the end of the IU, while questions end with H%.

The realization of Accent 1 in different focus conditions was previously investigated in [3]. In this paper, the phonetic realization of Norwegian broad and narrow non-contrastive focus was compared to the realization in similar constructed utterances in German. The Norwegian data were restricted to Accent 1 (for comparability with the German data analysed in the same article), and the analyses were presented for reiterant "dada" utterances to enable easy comparison of segmentally similar material across the two languages. In Norwegian, the durations of the accented vowel, syllable and word were found to play an important role in signaling focus (cf. [4]), with much greater differences between focus conditions than was the case in German. With respect to F₀, German speakers were found to vary in their choice of the pitch accent, whereas Accent 1 in Norwegian is always realized as an L*+H accent. Nevertheless did we find considerable variation among the speakers in the way they realized this pitch accent in broad versus narrow focus conditions. A tendency towards later alignment of the H peak, closer to the accent phrase boundary, was observed in broad than in narrow focus conditions with the focus late in the utterance (which are otherwise similar to broad focus realisations).

The present study further investigates the phonetic realisation of Norwegian intonation in different focus conditions by comparing utterances containing two minimal pairs with Accent 1 versus Accent 2 in otherwise identical utterances, but with different foci. Additionally, statements are compared to structurally identical questions. The aim is to evaluate whether there are differences in the realization of the accents for these different conditions which listeners could use in their interpretation of the utterance.

In order to determine the precise tonal alignments and magnitudes of excursions, F₀ contours are parameterized using the Fujisaki model [5]. This model decomposes a given log F₀ contour into a base frequency *Fb*, a phrase component, capturing slower changes in the F₀ contour as associated with intonation phrases, and an accent component that reflects faster changes of F₀ associated with accents and boundary tones. The phrase and accent components can be interpreted as smooth responses of the model to impulse-wise phrase commands and box-shaped accent commands. In previous studies, F₀ contours of other Nordic languages, such as Swedish and Finnish have been successfully decomposed [6, 7]. Whereas the modeling of Swedish required the use of negative accent commands, Finnish did not.

2. Speech Material and Analysis Method

Sentences were constructed around two minimal pairs which only differ in their tonal accent. These words only contain

¹ The Tonal Foot is not to be confused with the stress foot (Σ) of nonlinear phonology. In his analysis of Norwegian intonation Kristoffersen [2] referred to this constituent as an Accent Phrase (AP).

² Kristoffersen's analysis [2] is different in that the syllables are not dominated by the Prosodic Word but by the Accent Phrase (= tonal Foot, Fretheim). The reason for omitting the constituent ω from the intonational hierarchy is its hybrid (not purely phonological) status (its left edge is defined by phonology and its right edge by morphology).

voiced segments, yielding continuous F0 contours. The words *bøndene*, (pronounced /b2n@n@/, represented by SAMPA, E. ‘the farmers’) and *loven*, (/lO:v@n/, E. ‘the law’), both carry Accent 1, while *bønnene* (also pronounced /b2n@n@/, E. ‘the beans’) and *låven* (/lO:v@n/, E. the ‘barn’) both have Accent 2. These ‘critical words’ were embedded in the sentence *Det er i Bergen* (E. ‘This is/these are the in Bergen.’) and preceded by questions to elicit broad focus or narrow contrastive focus (note that the narrow focus in our previous work was non-contrastive). In the case of narrow focus, two versions were produced, one with focus on the so-called critical word (*bøndene*, etc., as the reply to the question *Er det bønnene i Bergen?*) and one with the contrast realized on the word *Bergen* (as the reply to *Er det bøndene i Lillehammer?*). In addition, the sentence was realized as a broad focus question. All sentences and their contexts were read from paper. An example of a test utterance with the tree structure representation according to the Trondheim model is shown in Figure 1. The figure shows the primary and secondary association of tones and terminal elements of the prosodic tree as well as the peripheral association of tones to higher nodes (based on [1, 8]).

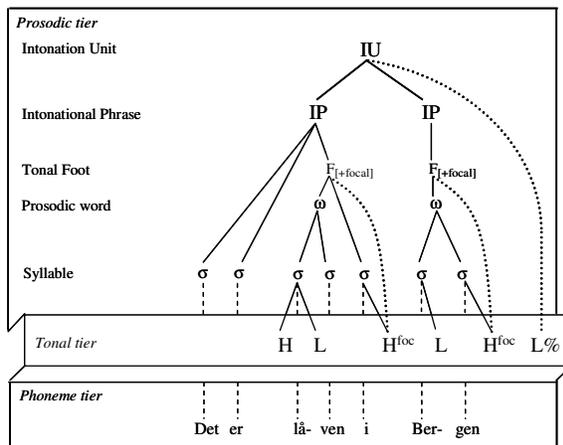


Figure 1: Tree representation for the intonation unit structure of one of the test sentences.

Five realizations of each sentence were recorded for 5 female and 3 male speakers of Urban East Norwegian. The recordings were made in a recording studio in the Department of Language and Communication Studies at NTNU. Since not all realizations were found to be satisfactory, for example due to hesitations or pauses within the utterances, a number of them were discarded from further analysis.

F0 values were extracted using the standard method in Praat [9] at a step size of 10 ms, and the F0 tracks were decomposed using an automatic method originally developed for German [10]. Resulting parameters were viewed in the *FujiParaEditor* [11] and corrected if necessary. Since Swedish appears to require negative accent commands to satisfactorily capture the difference between Accent 1 and Accent 2 [6], we examined the results for fitting errors on low tone syllables. We found that most instances could be well represented using positive accent commands only, only female speaker 3 seemed to have an occasional preference for an additional active lowering on the low accented syllable. For the majority of speakers, however, the low accent was simply achieved by the offset of a preceding accent command.

With only positive accent commands, the L tone immediately follows the end of the command. It is expected that one positive command occurs before (Accent 1) or around

the start of (Accent 2) the accented syllable, and another one around the accent phrase boundary.

3. Results

Figure 2 shows a typical example of the word *bøndene* (Accent 1), top, in the broad focal condition and *bønnene* (Accent 2), bottom, uttered by speaker SP1.

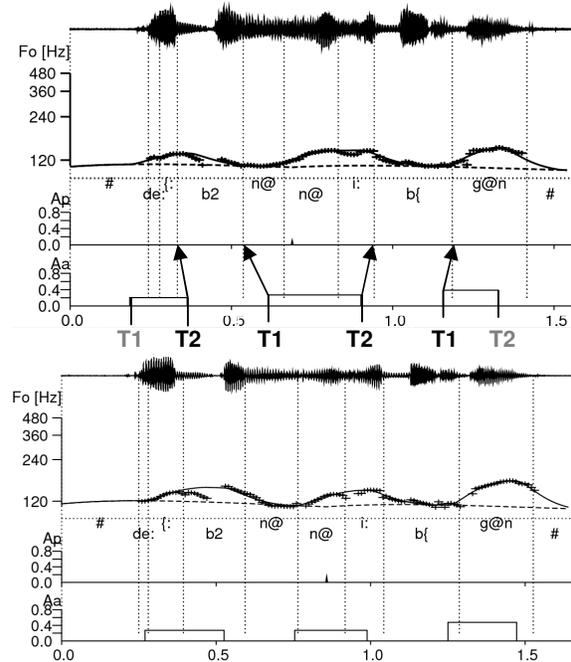


Figure 2: Two examples of analysis, Accent 1 (top) and Accent 2 (bottom), broad focus. Accent command onset times T1 and offset times T2 and their alignment with respect to syllable onsets illustrated by arrows in the top panel.

Each figure displays from the top to the bottom: the speech waveform, the F0 contour (extracted and modeled), as well as the underlying phrase and accent commands. The syllable boundaries are indicated by the dotted vertical lines. We shall now explain how the accent commands are aligned with respect to the syllables of the utterance. The first accent command starts close to the segmental onset of the utterance and ends at the beginning (Accent 1) or after the middle (Accent 2) of the accented syllable [b2]. In the subsequent analysis we therefore associate the offset time T2 of the first command with the syllable [b2] (see top panel for illustration of alignment indicated by arrows). It occurs early for Accent 1 (at the beginning of the fall into the low accented syllable) and late for Accent 2 (yielding a high accented syllable). In many cases of Accent 1 words, however, this first accent command is missing altogether (29.7% of all cases in the database) whereas in the case of Accent 2 it occurs consistently. The following accent command coincides with the rising tone delimiting the accent phrase *bønnene i*; its start is aligned with the syllable following the accented syllable (in this case [n@]), and the offset time T2 of this command is determined by the end of the accent phrase and the beginning of the following low accented syllable [b{] of the Accent 1 word *Bergen*. In our analysis we therefore associate T2 with the onset of the accented syllable [b{]. Furthermore, the third and last accent command in our example marks the end of the accent phrase ‘Bergen’ and we therefore relate its onset time T1 with the onset of the syllable [g@n] and its offset time T2

with the end of the same syllable. Just as the accent command preceding [b2], this last accent command is not always realized, especially when *bønnene* is narrowly focused (49.3% of all cases).

Table 1 lists the mean relative timing ($T1rel$ and $T2rel$) and the mean command amplitude Aa for Accents 1 and 2 in the broad focus condition. The timing is measured in ms relative to the onset of the relevant syllable (see illustration in Figure 2, top). A negative value therefore means that the onset $T1$ or offset $T2$ precedes the syllable onset. The alignment most relevant for the current syllable is indicated in bold face.

Table 1: Mean alignment in ms and amplitudes of accent commands for the broad focus condition, statements.

Syllable	Accent 1			Accent 2		
	$T1rel$	$T2rel$	Aa	$T1rel$	$T2rel$	Aa
[b2]	-164	8	.171	-65	117	.343
[n@]	84	301	.370	172	390	.369
[b{]	-253	-30	.330	-186	24	.262
[g@n]	-33	129	.482	-49	161	.471
[IO:]	-155	28	.255	-23	80	.188
[v@n]	-11	225	.370	-69	156	.327
[b{]	-244	15	.239	-206	29	.275
[g@n]	-57	111	.599	-18	168	.320

The table shows that the accent command offset ($T2rel$) on the Accent 2 words occurs about 60-110 ms later than in Accent 1 words. This delay also results in a delay in the rise of the phrasal tones on [n@] and [v@n]. Furthermore, accent command amplitudes associated with Accent 2 words are generally higher than for Accent 1 words (where also they are not always present), indicating that the former is the marked case.

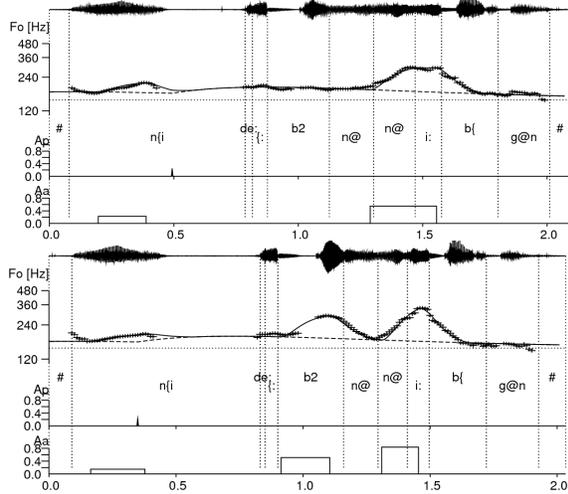


Figure 3: Two examples of analysis, Accent 1 (top) and Accent 2 (bottom), contrastive early focus: “*Nei, det er bønnene/bøndene i Bergen*”.

Figure 3 displays two examples uttered by SP4 of the condition in which *bøndene* (Accent1)/*bønnene* (Accent2) is contrastively focused, with the Accent 1 word at the top, Accent 2 at the bottom. As explained before, like in many other cases, there is no accent command preceding the low syllable [b2] in the Accent 1 case, whereas there is one starting shortly after the beginning of the syllable [b2] in the Accent 2 case. Subsequently, in contrast to the broad focus condition, the narrow focus is actually marked by the increased amplitude of the accent command associated with the phrasal tone, which

starts near the end of the syllable [n@] following the accented syllable. This accent command ends shortly before the beginning of the next low accented syllable [b{]. In contrast to the broad focus condition, the rising phrasal tone on *Bergen* is deleted.

Table 2 lists the mean relative timing and mean accent command amplitudes Aa for Accents 1 and 2 in the early narrow focus condition. The timing is measured in ms relative to the onset of the relevant syllable.

Table 2: Mean alignment in ms and amplitudes of accent commands for early narrow focus condition, statements.

syllable	Accent 1			Accent 2		
	$T1rel$	$T2rel$	Aa	$T1rel$	$T2rel$	Aa
[b2]	-158	24	.171	-61	155	.382
[n@]	90	361	.535	168	367	.659
[b{]	-299	-20	.425	-228	-28	.592
[g@n]	-115	114	.131	-159	24	.146
[IO:]	-170	15	.165	-67	114	.303
[v@n]	-16	247	.537	50	250	.651
[b{]	-286	-5	.412	-220	-24	.594
[g@n]	-110	83	.231	-88	49	.119

As mentioned before, the early narrow focus is reflected by the increase in accent command amplitude Aa associated with the phrasal tone connected with the syllables [n@] and [v@n], respectively, as well as the reduction of Aa associated with the phrasal tone on the syllables [g@n] of *Bergen*. This effect was not found in [3], where a non-contrastive narrow focus was elicited, so that the question can be raised whether it is caused by the contrastive nature of the narrow focus condition in the present data.

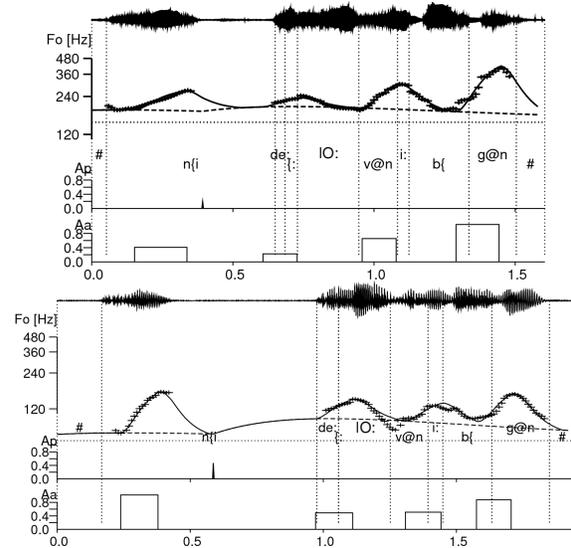


Figure 4: Two examples of analysis, Accent 1 (top) and Accent 2 (bottom), contrastive late focus: “*Nei, det er loven/låven i Bergen*”.

Figure 4 displays two examples of the condition in which *Bergen* is contrastively focused, with the Accent 1 word at the top (uttered by speaker SP4), Accent 2 at the bottom (uttered by speaker SP3). As can be seen the narrow focus causes the accent command on *Bergen* to be boosted. However, we found different strategies in different speakers for achieving this. Speaker SP3, for instance, connects the phrasal tone on [v@n] with an even higher accent command on [g@n], see Figure 5, top. In these late contrastive focus conditions, the accent

command amplitude assigned to *Bergen* reaches a mean of .789. In utterances of questions (see example in Figure 4, bottom) this value rises to .902. The higher target at the end of questions is modelled by a higher command amplitude A_a than at the end of statements. The accent command thus combines the H tone of the accent phrase with an H% boundary for questions, and with an L% boundary tone for statements.

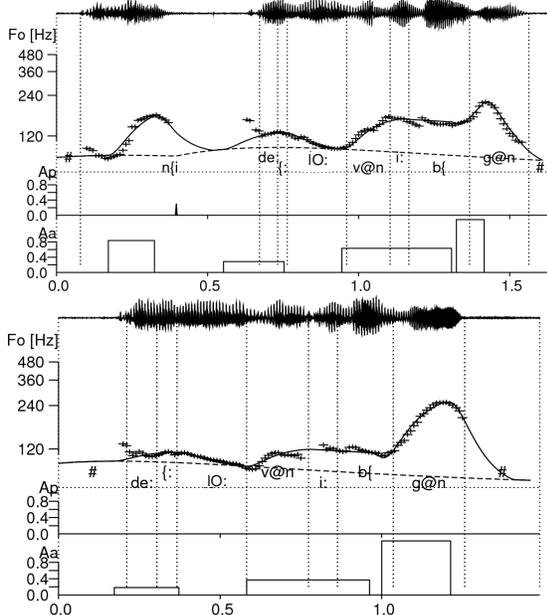


Figure 5: Top: contrastive late focus, “*Nei, det er loven i Bergen*”, statement; bottom: Example of a broad focus question “*Det er loven i Bergen?*” (both uttered by speaker SP3)

4. Discussion

A privative contrast between Accent 1 and Accent 2 is assumed by [12], among others, in which Accent 1 is toneless and only has intonational tones, while Accent 2 has an H tone in its underlying representation. This description is supported by our observation that Accent 1 is not necessarily connected with an accent command, whereas Accent 2 always is, and usually with one of higher amplitude. The timing of the following tonal transitions is delayed in cases of Accent 2. Another important finding is that the rising tonal accent is initiated already in the vicinity of the syllable following the accent syllable, but the tonal gesture, that is, the associated accent command, extends until the end of the accent group.

Although our finding that the narrow and broad focus accents differ phonetically seems to contradict Fretheim’s description of them as indistinguishable [13], we must point out that the material used in our analyses may have been too selective: A relatively large number of utterances in the broad focus condition were discarded because a native listener felt the accent on the critical word was too strong for the utterance to be a typical broad focus realization. The assumption was that speakers might have produced the stronger accent because it is difficult, even for Norwegians, to produce the utterances with the critical words differing only in their tonal accent. This criterion may have been too restrictive, and possibly the utterances should have been classified as perfectly acceptable realizations of broad focus utterances. This might have led to less clear differences between broad and narrow focus realizations than we observed in our data, and brought to light

that the same realization can be appropriate in both a narrow and a broad focus utterance.

The next step in our work is to manipulate the Fujisaki model parameters to investigate whether our observations from the production study presented in this paper are also perceptually relevant. In particular, we shall investigate the timing of the last H tone in the LH (Accent 1) and HLH accents (Accent 2) for the different focus conditions. In addition, we shall investigate the effect of peak scaling, also in relation to non-contrastive versus contrastive focus.

5. Conclusions

It was shown that Fujisaki modeling of the Accent 1–Accent 2 opposition can reflect the expected differences in their realization. In addition, the modeling shows that there are differences in the alignment and size of the accent command in different focus conditions. Listeners may make use of these differences in their perception. The Fujisaki model can therefore be used to manipulate utterances for perception tests to verify our conclusions.

6. Acknowledgements

The work presented in this article was funded by the German Research Council, grant no. BA 737/10-1, 2006-2008, as well as an Erasmus staff exchange grant to the first author.

7. References

- [1] Fretheim, T. (1990). The Form and Function of Foot-External Syllables in Norwegian. Intonation. In K. Wiik and I. Raimo (eds.), Proc. Vth Nordic Prosody Conference, University of Turku. 87–110.
- [2] Kristoffersen, Gjert (2000). The phonology of Norwegian. Oxford: Oxford University Press.
- [3] Koreman, J., Andreeva, B., Barry, W.J., Van Dommelen, W. and Sikveland, R. (2009). Cross-language differences in the production of phrasal prominence in Norwegian and German, Proc. Xth Nordic Prosody conference, August 4 - 6 2008, Helsinki, Finland.
- [4] Fant, G. and Kruckenberg, A. (1993). Towards an integrated view of stress correlates, Dept. of Linguistics, Lund University, Working Papers 41, 42-46.
- [5] Fujisaki, H. and Hirose, K. (1984). Analysis of voice fundamental frequency contours for declarative sentences of Japanese”, J. of the Acoust. Society of Japan (E) 5(4), 233-241.
- [6] Fujisaki, H., Ljungquist, M. and Murata, H. (1993). Analysis and modelling of word accent and sentence intonation in Swedish, Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, 211-214.
- [7] Mixdorff, H., Vainio, M., Werner, S. and Järviö, J. (2002). The Manifestation of Linguistic Information in Prosodic Features of Finnish. Proc. Speech Prosody Conference, 511-514, Aix, France.
- [8] Pierrehumbert, J. and Beckman, M.E. (1988). Japanese Tone Structure. Linguistic Inquiry Monographs 15, Cambridge/MA: MIT.
- [9] Boersma, Paul (2001). Praat, a system for doing phonetics by computer. *Glott International* 5:9/10, 341-345.
- [10] Mixdorff, H. (2000). A novel approach to the fully automatic extraction of Fujisaki model parameters, Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP2000), vol. 3. Istanbul, 1281.1284.
- [11] Mixdorff, H. (1/10/2009). FujiParaEditor. <http://public.beuth-hochschule.de/~mixdorff/thesis/fujisaki.html>
- [12] Gussenhoven, C. (2004). The Phonology and Tone of Intonation. Cambridge: Cambridge University Press.
- [13] Gundel, J.K. and Fretheim, T. (2004). Topic and Focus. In: L. Horn and G. Ward (eds.), Handbook of Pragmatic Theory. Oxford: Blackwell. 174-196.