# Interaction of fundamental frequency contour and segment duration in the production and perception of Norwegian

*Wim A. van Dommelen*

Norwegian University of Science and Technology

wim.van.dommelen@ntnu.no

ABSTRACT

The aim of this study is to explore the interplay of fundamental frequency contour and the production and perception of segment duration in Norwegian. Production patterns were investigated in a case study involving one speaker producing isolated words and a database study analyzing speech material from 220 speakers. Measurements revealed inconsistent effects of word accent on segment duration, accent 2 vs. 1 tending to give rise to longer durations in isolated words but shorter durations in words extracted from sentence context. Regression analyses investigating the relation between $f_0$ rise or fall in a vowel and the duration of such $f_0$ movements showed that word accent was not a contributing factor. In a perception test, the vowel of a disyllabic word was manipulated to create a short-long vowel continuum. Manipulation of $f_0$ contours showed that a falling vs. a flat or rising contour in the first syllable caused perceptual lengthening of the vowel. Changing the second syllable's original rising $f_0$ contour into a falling one shortened the first syllable perceptually when its contour was falling but had the opposite effect for a flat or rising contour.

**Key words**: fundamental frequency contour, segment duration, production, perception, Norwegian, accent 1, accent 2

## 1. Introduction

Norwegian may exploit contrastive tonal melodies to distinguish between segmentally identical lexical items. Adopting Kristoffersen's [1] terminology, they will be called accent 1 and accent 2 here. While most Norwegian dialects include tonal contrast in their phonological system, its phonetic realization varies between dialects. Usually, two main types are distinguished, namely low tone and high tone dialects. This nomenclature refers to the realization of accent 1. Whereas in a low tone dialect accent 1 has a LH melody with L realized on the prominent syllable, this accent is realized as HL in a high tone dialect. Typical patterns for accent 2 in the two types of dialects are HLH and LHL, respectively (see Figure 1).

The present study investigates the interaction of fundamental frequency contour and temporal organization in the production and perception of Norwegian. With regard to speech production, it could be speculated that the more dynamic accent 2 contour would require more time than accent 1. All other things being equal, accent 2 words would be longer than their accent 1 counterparts. This will be investigated in the production part of this paper (section 2). To the best of my knowledge there are no previous publications on this issue to refer to. Swedish has corresponding tonal contrasts. Investigating Stockholm and Scania Swedish, Ambrazaitis & Tronnier [2] observed longer segment durations in disyllabic accent 2 vs. accent 1 words in both dialects. Articulatory data on Scania Swedish collected by Svensson Lundmark et al. [3] confirmed this observation. In contrast, Svensson Lundmark [4] reported inconsistent effects of word accent for Southern Swedish vowel durations. Phonemically long accent 2 vowels were longer, but short accent 2 vowels did not differ from or were even shorter than their accent 1 counterparts. At present, it is not clear how the robust effect of longer accent 2 durations in [2] can be explained. The authors could rule out tonal complexity as an explanation.

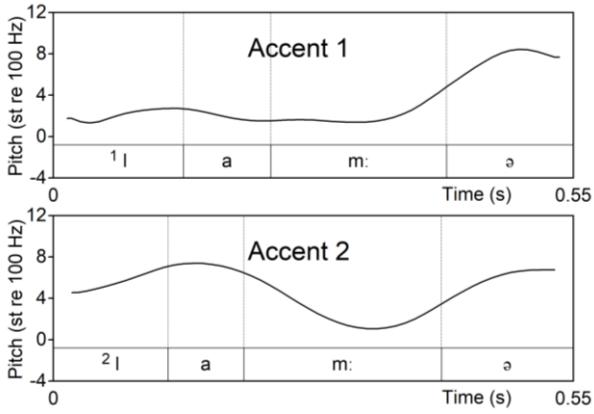


Figure 1: *Accent 1 (LH) pattern in <lammet> (the lamb); accent 2 (HLH) in <lamme> ([to] lame) spoken by a Norwegian low tone dialect speaker. Time-normalized word durations*

Mandarin Chinese has a fully-fledged tone system that may shed more light on the relation between tonal contour and segment duration. The picture for syllables bearing Tone (T) 1, 2, 3, or 4 presented by Wu & Kenstowicz [5] showed that the more complex T3 (314) had a longer duration than T2 (35). The level T1 (55) was in turn shorter than T2 but longer than T4 (51), which may reflect less physiological effort in the production of falling $f_0$. Matching results were obtained by Brotzman [6], Ho [7], Nordenhake & Svantesson [8], Whalen &Xu [9], and Xu [10]. Perception data obtained by Gussenhoven & Zhou [11] mirrored production in that syllables that had a longer duration in production (LL; corresponding to T3 as in [9, Figure 6]) sounded shorter than syllables that had a shorter duration in production (HH or HL; corresponding to T1 and T4, respectively), everything else being equal. They explain this result by assuming perceptual compensation of articulation strategies and constraints.

During the last 40-50 years or so, various studies have been performed examining the influence of a dynamic $f_0$ contour on perceived segment duration. In her classic investigation, Lehiste [12] observed a perceptual lengthening effect of a dynamic vs. level $f_0$ contour. More specifically, a rising contour had a stronger lengthening effect than a falling one. This hierarchy was confirmed by Wang et al. [13] and also, be it less robust, by Rosen [14]. A different hierarchy of falling > rising > level (> meaning 'perceived as longer') was found by Cumming [15]. Both rising (T15) and falling (T51) were perceived as longer than level (T33) in Yu [16]. In Rosen [17], a rising contour in a monosyllable caused perceptual shortening compared to a level one, a falling contour perceptual lengthening. Different heights of level contours had virtually no effect. On the whole, effects were small, and no statistics were performed. Different from these results for Swedish [17], van Dommelen [18] found only a lengthening effect of a falling, but no such effect of a rising vs. a flat contour in Norwegian mono- and disyllables. As appears from this brief overview, the effects of f0 contour dynamics and perceived duration are inconsistent across languages and experimental designs.

The perception part of this study (section 3) explores the effect of dynamic f0 contours in Norwegian disyllables via manipulation of contours in both syllables. Given the generally inconsistent picture it does not seem feasible to formulate any firm predictions, in particular regarding the question whether the interplay of f0 and perceived duration will be different in Norwegian than in other (especially non-tonal) languages.

Working hypotheses for the production (section 2) and the perception (section 3) parts will be formulated in the respective subsections.

# 2. Production

## 2.1. Case study

### 2.1.1. Experimental hypothesis

This section deals with the relation between word accent and word duration using recordings of accent 1 – accent 2 minimal pairs. In view of the picture for closely-related Swedish it was hypothesized that in a Norwegian low tone dialect, realization of words carrying a HLH (accent 2) pattern would require a longer duration compared with a LH (accent 1) pattern.

### 2.1.2. Materials, recordings, and analysis

Speech material used for the case study consisted of three lists of accent 1 – accent 2 minimal pairs developed as training material for the Computer-Assisted Listening and Speaking Tutor [19] (CALST). The first list comprised 16 disyllabic word pairs of varying segmental composition but always with a voiced medial consonant (e.g., *skallet*, /ˈskalːə/, 'bald' – *skalle*, /ˀskalːə/, '[to] knock'). List 2 contained 58 disyllabic word pairs, most of them having voiceless medial consonants (e.g., *søket*, /ˈsøːkə/, 'the search' - *søke*, /ˀsøːkə/, '[to] search'). There were 50 trisyllabic word pairs in the third list also with varying types of initial and medial consonants (e.g., *lysene*, /ˈlyːsənə/, 'the lights' - *lysende*, /ˀlyːsənə/, 'shining').

Using a Shure KSM44 microphone, audio recordings were made in the sound-treated studio of the Department of Language and Literature at the Norwegian University of Science and Technology. Recordings were stored on hard-disk with a sampling frequency of 44.1 kHz and 16-bit quantization. A trained male speaker of South-East Norwegian in his late twenties read each of the three lists twice, the first time all minimal pairs in the order accent 1 – accent 2, the second time reversed.

Acoustic analysis of the recordings was performed using the Praat program [20] and involved visual inspection of waveform and spectrogram. Realization of Norwegian accents 1 and 2 extends over minimally two syllables, where L of LH and HL of HLH is realized on the prominent syllable. For the disyllabic as well as the trisyllabic words, total word duration was measured. Since all accent 1 vs. accent 2 words were minimal pairs, segmental content did not introduce any confounding factor.

### 2.1.3. Results case study

Results from the case study are presented in Table 1. Pooled across all conditions, accent 2 minimal pair members were 6 ms longer than their accent 1 counterparts. According to a repeated measures ANOVA with Word Accent (1 vs. 2), Order of production (Accent 1 – 2 vs. Accent 2 – 1), and List (1 – 3), this durational difference did not reach statistical significance ($F(1, 121) = 1.11$, $p = 0.294$). At the same time, the analysis revealed a significant Word Accent x Order interaction, accent 1 words being longer than accent 2 words under condition 1-2 (659 ms vs. 616 ms) but shorter under condition 2-1 (630 ms vs. 685 ms; $F(1, 121) = 171.9$, $p < 0.001$).

Separate analyses of each of the three lists revealed a significant effect of word accent for merely one of them. For List 1, accent 2 words were shorter than accent 1 words (3 ms; the difference not being significant: $F(1, 15) = 0.17$, $p = 0.684$) with a significant Accent x Order interaction ($F(1, 15) = 36.6$, $p < 0.001$). For Lists 2 and 3, accent 2 words were longer than their accent 1 counterparts (11 ms and 2 ms, only the former difference reaching significance: $F(1, 57) = 8.19$, $p = 0.006$ and $F(1, 49 = 0.21$, $p = 0.650$, respectively). For both lists, Accent x Order interactions turned out to be significant ($F(1, 57) = 112.9$, $p < 0.001$, and $F(1, 49) = 75.3$, $p < 0.001$, respectively). The three-way interaction Accent x Order x List did not reach significance ($F(1, 121) = 0.324$, $p = 0.724$).

Table 1: *Mean word duration and standard deviation (in ms) for accent 1 and accent 2 words read in that order and reversed (see text)*

| Materials | Order | Acc. 1 | sd | Acc. 2 | sd | n |
|---|---|---|---|---|---|---|
| List 1 | 1-2 | 615 | 58 | 556 | 72 | 16 |
| | 2-1 | 572 | 55 | 625 | 50 | 16 |
| List 2 | 1-2 | 624 | 66 | 587 | 68 | 58 |
| | 2-1 | 591 | 60 | 651 | 64 | 58 |
| List 3 | 1-2 | 714 | 68 | 668 | 59 | 50 |
| | 2-1 | 693 | 68 | 743 | 71 | 50 |
| Lists 1-3 | 1-2 | 659 | 80 | 616 | 78 | 124 |
| | 2-1 | 630 | 81 | 685 | 81 | 124 |
| Grand mean | | 644 | 82 | 650 | 87 | 248 |

## 2.2. Database study - vowel and word duration

### 2.2.1. Experimental hypothesis

While the speech material used in the case study presented above was systematic but of restricted size, larger amounts of speech material for both low tone and high tone dialects were available in the database exploited in the present section. Two different analyses were performed. Because the speech material does not contain minimal pairs but consists of read sentences with varying content, the first analysis examined only vowel durations. Consequently, only part of the relevant accent 1/accent 2 $f_0$ contours could be involved. Therefore, the second analysis focused on word duration. To reduce undesirable variation due to varying word length, only selected tokens were evaluated (see 2.2.2). Again, it was hypothesized that accent 2 tokens would be longer than those produced with accent 1. Due to inherently longer word duration, this effect was expected to be stronger for the words than for the vowels.

### 2.2.2. Speakers, speech materials and analysis

Materials used for this part of the investigation were chosen from the speech database NB Tale [21] provided by the National Library of Norway. The corpus was collected in 2012, and subsequent annotations in XSAMPA were completed 2013. Almost all recordings were made in a sound-treated booth using 48 kHz sampling frequency and 16-bit quantization. The database contains recordings of first and second language speakers of Norwegian. The former group comprises 240 speakers and is divided according to speakers' dialectal background into 12 subgroups containing 20 speakers each. The subgroups are balanced for age (< 40 years, > 40 years) and gender, and cover main four dialect groups: North, Central, West, and East Norwegian. While North and West Norwegian are classified as high tone dialects, Central and East Norwegian are low tone dialects. For the present purposes, the following numbers of speakers were selected: 60 native speakers of North Norwegian and Central/East Norwegian each, and 100 native speakers of West Norwegian dialects. Central and East Norwegian are prosodically close, which is why they were pooled to represent low tone dialects. One dialectically heterogeneous group of 20 speakers was not included.

Each speaker in NB Tale read a total of 20 sentences. While the first three sentences are identical for all speakers, the remaining 17 have different content across speakers. Because realization of the tonal contrast is dependent on stress, stressed vowels were selected from this material. Inspection revealed that the vowels originated from words varying in length between one and 24 phonemes. For shorter as well as longer words, distributions of phonemically short and long vowels carrying accent 1 and 2 appeared to be too dissimilar to be included for evaluation. Therefore, for subsequent evaluation of vowel duration, vowels from words having four to ten phonemes were selected.

To examine the effect of $f_0$ contour on temporal organization beyond the level of stressed vowels, words containing five phonemes were selected. For these tokens only word duration was evaluated. The selection of words having the same length in terms of phonemes reduced the confounding influence of unsystematically varying word length on the effect of accent.

Further, due to occasional errors in the database some vowels had unrealistically long durations. Hence, for the analysis of vowel as well as word duration, words with vowels longer than 300 ms were excluded. Analysis was performed using a Praat script extracting vowel and word durations from NB Tale's annotated TextGrids.

### 2.2.3. Statistical evaluation

Production data were statistically analyzed using the R program's package lme4 [22] to calculate Linear Mixed Effects Models (LMEM) [23]; see also [24]. Evaluation of results for word accent and vowel duration (section 2.2.4) involved the fixed factors Accent (1, 2), vowel Quantity (short, long), Dialect group (North, West, Central/East Norwegian), and Age (< 40 years, > 40 years) with by-subject random slopes and intercepts for the factors Accent and Quantity. Each analysis was performed by comparing this model with a model reduced by one of the fixed factors. Interactions were tested using the same method. Subsequently, to assess the factor's significance, likelihood ratio tests were performed comparing the two models.

### 2.2.4. Results – vowel and word duration

Mean vowel durations are presented in Table 2. Across all conditions, vowels in accent 2 words were on average only 1 ms longer than their accent 1 counterparts (99 ms vs 98 ms). Nevertheless, statistical analysis showed the effect of the factor Accent to be significant ($\chi^2(1) = 4.91, p = 0.027$). Breaking down data by Accent and Quantity revealed that the effect of slightly longer accent 2 vowels was not consistent. Across all dialects, phonologically short accent 2 vowels were shorter than accent 1 vowels (75 ms vs. 78 ms). Long vowels showed the opposite tendency (125 ms vs. 123 ms). This two-way interaction turned out to reach statistical significance ($\chi^2(1) = 34.8, p < 0.001$). As can be seen from the results for each of the three dialect groups, similar interaction patterns could be observed for West Norwegian and East/Central Norwegian. In contrast, North Norwegian accent 2 vowels were generally found to be somewhat shorter than their accent 1 counterparts (74 ms vs. 76 ms, and 119 ms vs. 121 ms, respectively). This Accent x Quantity x Dialect interaction was statistically confirmed ($\chi^2(7) = 57.6, p < 0.001$).

The division of speaker groups in NB Tale according to age (younger than 40 vs. older than 40 years) made it possible to evaluate age as a factor. Overall, older speakers produced longer vowels than younger ones (pooled across Accent and Quantity conditions 102 ms vs. 95 ms; a statistically significant effect: $\chi^2(1) = 21.0$, $p < 0.001$). The same age effect of 7 ms was found for both accent 1 (101 ms vs. 94 ms) and accent 2 vowels (103 ms vs. 96 ms). Therefore, the Accent x Age interaction failed to reach significance ($\chi^2(1) = 0.48$, $p = 0.488$).

Table 2: *Mean vowel duration and standard deviation (in ms) broken down across quantity (QT; Short/Long) and accent 1 and accent 2 words of 4-10 phonemes length*

| Dialect | QT | Acc. 1 | sd | n | Acc. 2 | sd | n |
|---|---|---|---|---|---|---|---|
| North | S | 76 | 26 | 1977 | 74 | 25 | 1947 |
| | L | 121 | 38 | 1508 | 119 | 35 | 1761 |
| West | S | 80 | 27 | 3483 | 77 | 27 | 3077 |
| | L | 124 | 40 | 2447 | 128 | 40 | 2884 |
| East/ Centr. | S | 77 | 26 | 1853 | 73 | 26 | 1869 |
| | L | 124 | 40 | 1566 | 126 | 38 | 1749 |
| All | S | 78 | 27 | 7313 | 75 | 26 | 6893 |
| | L | 123 | 39 | 5521 | 125 | 38 | 6394 |
| Grand mean | | 98 | 40 | 12834 | 99 | 41 | 13287 |

Inspection of the data for word duration revealed a 10 ms longer duration for accent 1 compared to accent 2 words (see Table 3). According to statistical analysis, this effect was significant ($\chi^2(1) = 18.8$, $p < 0.001$). As can be seen from the table, words with phonemically short vowels contribute most to the effect (durations of 427 ms for accent 1 vs. 409 ms for accent 2; correspondingly for long vowel words 458 ms vs. 453 ms), but the Accent x Quantity interaction is only marginally significant ($\chi^2(1) = 3.03$, $p = 0.082$). For each of the three dialect groups, the picture is comparable. This is reflected in the nonsignificant three-way interaction Accent x Quantity x Dialect ($\chi^2(7) = 10.5$, $p = 0.161$).

Table 3: *Mean word duration and standard deviation (in ms) broken down across quantity (QT; Short/Long) and accent 1 and accent 2 words of 5 phonemes length*

| Dialect | QT | Acc. 1 | sd | n | Acc. 2 | sd | n |
|---|---|---|---|---|---|---|---|
| North | S | 426 | 110 | 404 | 412 | 94 | 568 |
| | L | 454 | 109 | 263 | 452 | 104 | 457 |
| West | S | 426 | 120 | 745 | 414 | 102 | 916 |
| | L | 458 | 113 | 401 | 453 | 112 | 706 |
| East/ Centr. | S | 429 | 114 | 406 | 400 | 100 | 545 |
| | L | 462 | 124 | 302 | 455 | 128 | 439 |
| All | S | 427 | 116 | 1555 | 409 | 99 | 2029 |
| | L | 458 | 115 | 966 | 453 | 114 | 1602 |
| Grand mean | | 439 | 117 | 2521 | 429 | 108 | 3631 |

As for the factor age, speakers older than 40 were found to have longer word durations than those under 40 (447 ms vs. 419 ms; $\chi^2(1) = 21.9$, $p < 0.001$). The age effect was stronger for accent 1 words (456 ms vs. 422 ms) than for accent 2 words (440 ms vs. 418 ms), as confirmed by the significant Accent x Age interaction ($\chi^2(1) = 4.53$, $p = 0.033$).

## 2.3. Database study - $f_0$ slope and segment duration

### 2.3.1. Experimental hypothesis

The goal of this section is to explore whether realization of the Norwegian word accents has any implications with regard to the slope of rising or falling tonal movements during a vowel. Two multiple regression models were run each with word accent as one of the independent variables. (More predictors will be introduced in section 2.3.3.) The first model investigated the relation between slope height and slope duration of $f_0$ during a vowel, the second one the relation between slope steepness and slope duration. Two alternative outcomes were examined (see Figure 2). The first possible outcome (a) would be that the slope of rise (or fall) of an $f_0$ contour during a vowel tends to be constant. This would imply that slope height (h1, h2) increases with increasing duration of the slope.

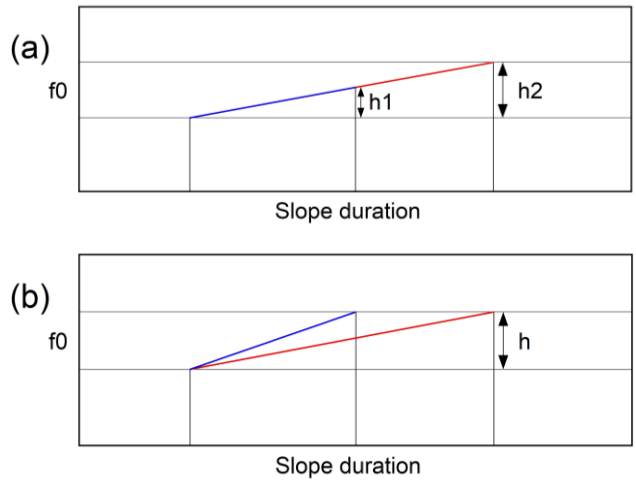


Figure 2: *Possible relations between slope and slope duration. (a) The slope of $f_0$ rise is constant, (b) Slope height (h) is constant*

Outcome (b) assumes that it is required to achieve a certain slope height (h) independent of slope duration. Following possibility (a), one would expect to find a relatively strong correlation between slope height and slope duration. In contrast, outcome (b) would imply a strong correlation between steepness of the slope and slope duration. Results supporting the latter could be interpreted to indicate a preference to reach an intonational target regardless of the movement's duration. As an experimental hypothesis it was assumed that in both regression models word accent would be a strong predictor of the dependent variable (slope height in model a, slope steepness in model b), thus indicating the influence of word accent on $f_0$ contours in Norwegian accent 1 and accent 2 words.

### 2.3.2. Speakers, speech materials and analysis

For this part of the study, the same speech materials produced by the same speakers as described in section 2.2.2 were used. Evaluation involved $f_0$ movements in stressed vowels from accent 1 and accent 2 words. Figure 3 depicts the relevant parameters for this evaluation. Using a Praat script, the values (in Hz) and the corresponding time points for $f_0$ minimum and maximum in the vowel were extracted. The difference in $f_0$ height (called *slope height*) divided by the time difference between the two points (called *slope duration*) gives the *$f_0$ slope* for the vowel. Because rising as well as falling $f_0$ contours might be expected to occur, subsequent calculations used absolute values of $f_0$ slope.

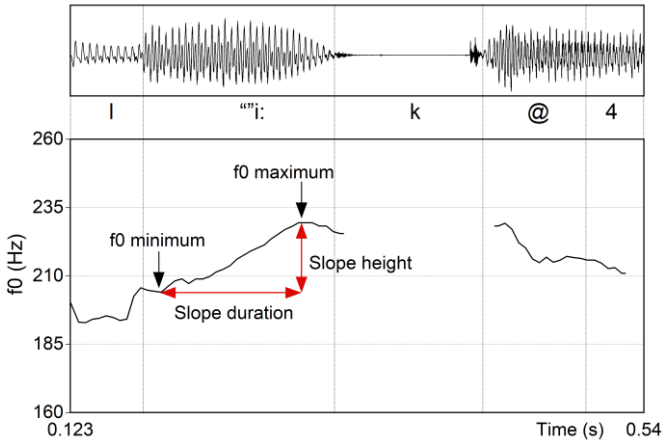


Figure 3: *Illustration of measurements in a vowel performed to calculate the slope of $f_0$*

### 2.3.3. Statistical evaluation

To test the experimental hypothesis formulated above, two types of multiple regression tests were performed, one with slope height, and the other type with steepness of the slope as dependent variable. In both cases, the following independent variables were included in the equations: slope duration, vowel quantity (short or long), accent (1 or 2), word duration, relative position of the beginning of the vowel in the word (expressed as percent of word duration), and relative position of the beginning of the vowel in the utterance (similarly in percent). Predictor variables were included using forced entry.

### 2.3.4. Results – $f_0$ slope and segment duration

Let us start by looking at the results of the multiple regression calculations pooled across all dialects. They revealed that the model correlating slope height with the six predictor variables explained 37.2 % of the total variance. Individual contributions of the predictor variables to this model are presented in Table 4. As shown by a standardized correlation coefficient of beta = 0.633, slope duration is by far the most influential one. The contribution of word accent did not reach statistical significance. Although predictor variables quantity, word duration, and position of the vowel in the word were significant, small beta values indicate that their contributions to the regression model are almost negligibly small.

Table 4: *Multiple regression results for dependent variable slope height and six predictor variables. V = Vowel. Beta = standardized coefficient*

| Predictor | Beta | t value | p | |
|---|---|---|---|---|
| Slope duration | 0.633 | 103.93 | < 0.001 | *** |
| Quantity (short, long) | -0.059 | -9.54 | < 0.001 | *** |
| Accent (1, 2) | 0.003 | 0.59 | 0.553 | n.s. |
| Word duration | 0.034 | 6.54 | < 0.001 | *** |
| V position in word | -0.030 | -5.54 | < 0.001 | *** |
| V position in utterance | -0.006 | -1.09 | 0.276 | n.s. |

Table 5 shows the results of the alternative regression model which included steepness of slope as dependent variable. This model explained just 2.4 % of the total variance. For the present purposes it is important to note that also here the contribution of the variable accent is small (beta = -0.015). The relevance of the corresponding value of p = 0.025 is therefore limited.

Table 5: *Multiple regression results for dependent variable slope steepness and six predictor variables. V = Vowel. Beta = standardized coefficient*

| Predictor | Beta | t value | p | |
|---|---|---|---|---|
| Slope duration | -0.090 | -11.86 | < 0.001 | *** |
| Quantity (short, long) | -0.054 | -7.04 | < 0.001 | *** |
| Accent (1, 2) | -0.015 | -2.24 | 0.025 | * |
| Word duration | 0.066 | 10.27 | < 0.001 | *** |
| V position in word | -0.065 | -9.56 | < 0.001 | *** |
| V position in utterance | 0.002 | 0.375 | 0.708 | n.s. |

Comparable results were obtained by running multiple regression analyses for each of the dialect groups North, West, and East/Central Norwegian. Predictor variables were the same as previously. Results for correlating slope height with the six predictor variables revealed similarly high amounts of explained variance as for the pooled data (30.7 %, 34.9 % and 42.8 % for North, West, and East/Central Norwegian, respectively). Models involving slope steepness achieved 3.7 %, 2.8 %, and 2.1 %, respectively. These results are reflected in the generally high beta values for slope height in comparison with slope steepness (see Table 6). Noticeably lower beta values than for steepness of slope were found for the variable accent (varying between 0.015 and 0.070), revealing its modest contribution to the models.

In addition to native speakers of Norwegian, the NB Tale database contains recordings from 11 different groups of L2 speakers originating from both European and non-European countries. Multiple regression calculations as described above were performed for the whole of these 220 subjects. For these speakers, too, the amount of explained variance in a model with slope height as a dependent variable was high (32.9 %) but low with slope steepness as a dependent variable (2.1 %). Finally, the factor accent had virtually no impact (beta = 0.022 and beta = 0.021, respectively). The significance of corresponding p values (p = 0.005 and p = 0.028) can be ascribed to large numbers of observations.

Table 6: *Part of multiple regression results for dependent variables slope height and steepness, and predictor variables slope duration and accent. Beta = standardized coefficient*

| Dialect | Predictor variables | | | |
|---|---|---|---|---|
| | Slope duration | | Accent (1, 2) | |
| | Beta | p | Beta | p |
| North | | | | |
| Slope height | 0.570 | < 0.001 | -0.040 | < 0.001 |
| Slope steepness | -0.147 | < 0.001 | -0.040 | 0.003 |
| West | | | | |
| Slope height | 0.614 | < 0.001 | 0.015 | 0.077 |
| Slope steepness | -0.108 | < 0.001 | 0.015 | 0.128 |
| East/Central | | | | |
| Slope height | 0.654 | < 0.001 | 0.070 | < 0.001 |
| Slope steepness | -0.094 | < 0.001 | 0.020 | 0.136 |

### 2.4. Summary production study

The amount of speech material investigated in the case study (section 2.1) was limited but could serve to indicate possible effects of word accent on temporal structure. Across the three examined word types, accent 2 realizations were shown to be unsubstantially longer than their accent 1 counterparts, the mean duration difference merely being 6 ms. Only for the disyllabic minimal word pairs from List 2, there was a significant effect of word accent, accent 2 words being 11 ms longer than accent 1 words. In addition, the effect of order of production (accent 1 – accent 2 vs. reversed) was stronger than that of accent.

The word material in the database study (section 2.2) was heterogeneous, thus causing a certain amount of noise in the data. It seems reasonable to assume, however, that the large number of observations would average out unsystematic variation. So, even small effects could reach statistical significance, e.g., the accent 1 – accent 2 difference in vowel duration of 1 ms across all conditions. Pooled over the three explored dialect groups, phonologically short accent 2 vowels were shown to be shorter, but phonologically long accent 2 vowels were longer than their accent 1 counterparts. North Norwegian patterns were different in this respect, but because accent 1 - accent 2 differences were on average just a few milliseconds small, differences between dialects can be said to be negligible. Summarizing, it seems justified to conclude that Norwegian word accent does not affect vowel duration.

Selection of words containing a fixed number of phonemes made it possible to estimate the effect of accent on word duration keeping random variation at a minimum. Contrary to the effect of word accent on vowel duration, accent 1 words were longer than accent 2 words. As was the case with vowel duration, the relative size of the effect was small (around 2 – 3 % of word duration).

Investigation of the relation between duration of an $f_0$ contour and its slope in section 2.3 revealed that slope height (as defined above) to a large degree covaried with its duration. There was no such relation between slope duration and its steepness. More importantly, the results of the regression analyses indicated that the factor word accent had virtually no impact. Similar results were obtained for L2 users of Norwegian. Altogether, the present production data thus suggest that Norwegian word accents are produced without affecting temporal word structure.

## 3. Perception

### 3.1. Experimental hypotheses

The goal of the perception part of this study was to shed more light on the effect of varying $f_0$ contours on perceived duration in Norwegian. As test paradigm, perception of the minimal pair <hakker> /²hɑkːər/ (*hoes*) - <haker> /²hɑːkər/ (*chins*) was chosen. In speech production, long and short /ɑ(ː)/ are spectrally very similar. Therefore, shortening the /ɑː/ vowel in <haker> will change perceived word identity into <hakker>.

Apart from creating a vowel duration continuum from short to long, the original course of $f_0$ in the first as well as the second syllable was manipulated (see Figure 4). The originally falling contour (FA) in the first syllable was replaced by a rising (RI) and a flat (FL) one. Additionally, the second syllable's original rising (RI) contour was replaced by a falling (FA) one. To keep the $f_0$ transition from the first to the second syllable comparable, the latter contour started at approximately the same height as the original one.

Based on previous evidence [18] it was speculated that a falling vs. a flat $f_0$ contour in the first syllable would cause perceptual lengthening of the vowel. For a rising vs. a flat contour, no such effect was expected.

The exchange of the original rising $f_0$ contour in the second syllable by a falling one corresponds to substituting a high boundary tone (H%) by a low (L%) one. Examining the perception of vowel duration, Steffman & Jun [25] found that listeners perceived the presence of utterance-medial L% replacing H% as signaling a slowdown in speech rate. Consequently, the vowel following the phrasal boundary was perceptually shortened. Therefore, it was speculated that the presence of a low vs. a high boundary tone in <hakker> - <haker> would strengthen the impression of finality, such that listeners through a compensation strategy [11] would experience shortening of the first syllable's vowel.

### 3.2. Generation of stimuli

Stimulus material for the listening test was prepared in the following way. Using a Shure KSM44 microphone, audio recordings were made in the sound-treated studio of the Department of Language and Literature at the Norwegian University of Science and Technology. Recordings were stored on hard-disk with a sampling frequency of 44.1 kHz and 16-bit quantization. A 20-year-old male speaker of South East Norwegian produced the minimal pair <hakker> - <haker> in isolation. The latter was selected to generate sound stimuli. All manipulations were performed using Praat.

As a result of pilot tests run to find workable segment duration values, the duration of the /ɑː/ was shortened from originally 274 ms to 200 ms. The latter duration was found to allow perception of a long vowel quantity, while at the same time reducing the number of stimuli. Subsequently, vowel durations varying in 10 ms steps from 70 ms to 200 ms were generated. Further, closure duration was lengthened to 192 ms, i.e., the

value mid-way between those in <haker> (152 ms) and <hakker> (232 ms). Similarly, the duration of the /ər/ part was set to the value midway between the two original tokens ([319 + 250]/2 = 285 ms).

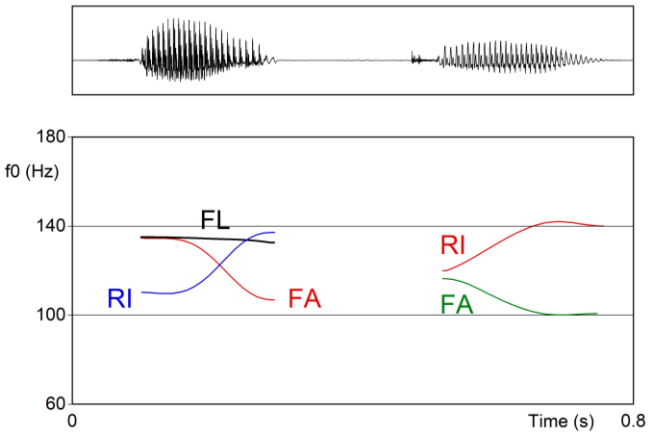


Figure 4: *Waveform of <haker> with vowel duration 200 ms and examples of $f_0$ contour manipulations (see text). FA-RI is the original accent 2 contour*

Additionally, $f_0$ contours in both the first and the second syllable were manipulated (see Figure 4). The original contours were falling (FA) from 135 Hz to 107 Hz in /a:/, and rising (RI) from 119 Hz to 139 Hz in /ər/. From these contours mirrored versions were created. The rising contour in /a:/ (RI) had values approximately ranging from the corresponding ones in the original. The values of the manipulated contour in /ər/ fell from approximately 119 Hz to 102 Hz. Finally, a nearly flat (FL) version of the /a:/'s contour was generated (to avoid a metallic sound quality slightly falling from 135 Hz to 131 Hz). Thus, for each vowel duration six intonational variants were created (FA-RI, RI-RI, FL-RI, and FA-FA, RI-FA, FL-FA).

### 3.3. Listening test procedure

In an identification test, the 84 (14 vowel durations x 6 intonational variants) different stimuli were presented to listeners four times each in randomized order. Randomization order was different for individual listeners. Before performing the test, listeners received written instructions.

A group of 21 subjects (9 f, 12 m) aged between 15 and 49 years (median: 22 years) from various dialect backgrounds and with no reported hearing problems participated in the listening test. Their task was to respond to the stimuli by clicking on one of two alternatives (*hakker* or *haker*) displayed on a computer screen. Since listeners' response prompted the program to present the next stimulus, presentation pace was set by the subjects. The majority of them needed less than 15 minutes to judge the 336 stimuli. Whereas five individual participants performed the test seated in the department's studio, the remaining 16 subjects accomplished the task online using their own equipment. Subjects were paid for their participation.

### 3.4. Listening test evaluation

Listening test data were evaluated by estimating the 50% point in perception of *hakker-haker* responses for each listener and each of the six intonational variants. Using the *predict* function for logistic regression in R, in each case the probability value at or closest to 0.50 was selected to find the corresponding value at the 70-200 ms vowel duration continuum. Estimated individual vowel durations representing perceptual phoneme boundaries were used to run a repeated measures ANOVA testing statistical significance of the effects of $f_0$ contour manipulations.

### 3.5. Results listening test

Figure 5 depicts the number of short vowel (*hakker*) responses as a function of vowel duration and $f_0$ contour pooled across the second syllable's two $f_0$ conditions. Whereas a falling vs. a flat $f_0$ contour in the /a(:)/ vowel decreased the number of short vowel responses, the response distributions for the rising vs. the flat $f_0$ contour overlap to a large degree. Further, Figure 6 shows that pooled across the first syllable's three $f_0$ conditions there is no effect of a rising vs. a falling contour in the second syllable.

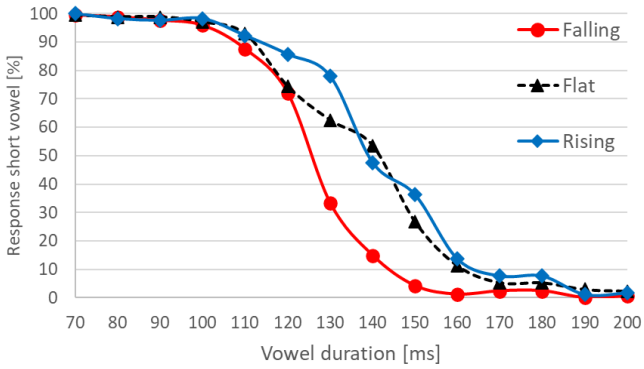


Figure 5: *Percentage of short vowel responses for <hakker – haker> as a function of vowel duration and $f_0$ contour pooled across two $f_0$ contours in the second syllable*

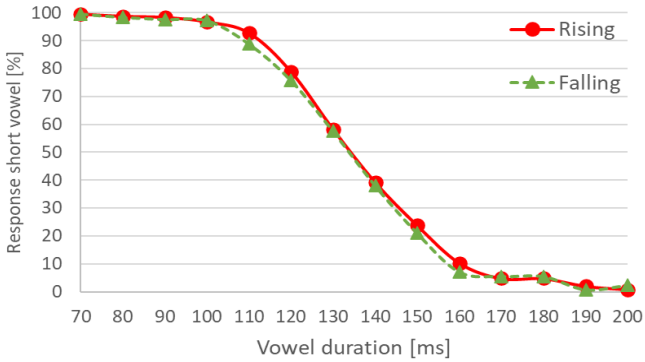


Figure 6: *Percentage of short vowel responses for <hakker – haker> as a function of vowel duration and $f_0$ contour in the second syllable pooled across three $f_0$ contours in the first syllable*

Inspection of the data specified for each of the six intonational conditions revealed a more complex picture (cf. Table 7). According to a repeated measures ANOVA with $f_0$ contour of the first syllable and $f_0$ contour of the second syllable as factors, the effect of the former factor was statistically significant ($F(2, 40) = 28.9$, $p < 0.001$). In contrast, $f_0$ contour variation in the second syllable had no significant effect, underpinning the picture presented in Figure 6 ($F(1, 20) = 2.04$, $p = 0.169$). The interaction of the two main factors, however, reached statistical significance ($F(1.55, 30.97) = 5.66$, $p = 0.013$; after Greenhouse-

Geisser correction). This interaction can be seen from Table 7. Following a falling $f_0$ contour in the vowel /a/, a falling vs. rising $f_0$ in the second syllable caused a shift of the perceptual boundary towards longer durations (127.4 ms vs. 124.2 ms). According to a t-test for correlated samples this shift is significant ($t(20) = 2.18$; $p = 0.041$). For stimuli with a flat or rising contour in the first syllable, the effect was the opposite (flat: 137.7 ms vs. 139.3 ms; rising: 138.5 vs. 143.5 ms). A repeated measures ANOVA only involving flat and rising contours in the first syllable showed that this reversed shift reached significance ($F(1, 20) = 11.8$, $p = 0.003$).

Table 7: *Mean phoneme boundaries (in ms) in <hakker – haker> perception for three $f_0$ contours in the first, and two in the second syllable (FL= flat, FA= falling, RI= rising). n = 21*

| | FL-RI | FA-RI | RI-RI |
|---|---|---|---|
| Phoneme boundary | 139.3 | 124.2 | 143.5 |
| Standard deviation | 13.5 | 8.3 | 12.3 |
| | FL-FA | FA-FA | RI-FA |
| Phoneme boundary | 137.7 | 127.4 | 138.5 |
| Standard deviation | 13.6 | 8.9 | 13.3 |

### 3.6. Summary perception study

As far as the influence of $f_0$ dynamics in a stressed vowel is concerned, the results from the perception test were in line with expectations. A falling vs. a flat $f_0$ contour caused a shift of the short-long phoneme boundary towards shorter durations. This can be interpreted as a perceptual lengthening of the vowel. For a rising vs. a flat contour no such effect was observed.

Findings for the effect of $f_0$ in the second syllable are harder to interpret. Pooled across all three $f_0$ conditions in the first syllable, there was virtually no effect. Following a flat or a rising contour in the first syllable, however, a falling $f_0$ caused a boundary shift towards shorter durations. In contrast, after a falling contour in the first syllable, the boundary shifted in the opposite direction. The discussion will attempt to explain this outcome.

## 4. Discussion

The aim of this study was to explore the interplay of fundamental frequency contour and the production and perception of segment duration in Norwegian. While the production part looked into the effect of tonal word accent on vowel and word duration, the perception part investigated the impact of a dynamic $f_0$ contour in a disyllabic word on perceived duration.

Concerning production, it was speculated that words carrying accent 2 would have longer durations than accent 1 words. The results from the production part did not support this expectation. To begin with, durational effects of accent 1 vs. accent 2 were generally small. For the isolated words that showed a significant effect (List 2), the relative difference amounted to approximately 1.9 %, for the context-embedded words from the database < 2.5 %. Investigating the perception of speaking rate, Quené [26] estimated just noticeable difference for tempo in speech to be about 5 %. Therefore, it seems doubtful that the present durational effects would have the potential to be perceptually relevant. Moreover, the direction of the word accent factor varied depending on conditions. While isolated accent 2 words from List 2 were longer than their accent 1 counterparts, the opposite direction was found for words in context. Considering the evidence of consistently longer segment durations in Swedish accent 2 compared to accent 1 words in [2], it can only be speculated why the picture for Norwegian should be different. The tonal systems of these two languages are basically comparable [27, 28]. An explanation of the incompatible results might be sought in the speech material used. The material in [2] was more controlled and tested on a larger group of speakers (24) than the word lists spoken by one speaker in the current investigation. Moreover, the word material used in the database study varied unsystematically. Although the selection criterion of a word length of five phonemes certainly reduced the influence of confounding factors, remaining random variation of segmental content was inevitable. Future research using more controlled material may find out if the picture for Norwegian emerged from the current data is representative.

As was pointed out in the introduction, investigations on Mandarin Chinese have shown that syllables bearing more complex tones tend to be longer than syllables with level tones [5, 6, 7, 8, 9, 10]. This is not necessarily in conflict with the absence of a complexity factor in [2] because in Mandarin the domain of the whole $f_0$ contour is just the syllable. A complex contour is articulatorily more demanding and may, therefore, lead to a longer syllable duration. Because the realization of tonal contrast in Swedish or Norwegian extends over two or more syllables, there is no such temporal constraint.

The results from the present perception study are in line with previous findings for Norwegian in that a falling - but not a rising - $f_0$ contour in a stressed vowel perceptually lengthened its duration [18]. At the same time, this outcome is in conflict with research reporting similar effects for both falling and rising contours [12, 13, 14, 15]. Possibly the discrepancy is due to differences in experimental design. While the present listeners' task was real word identification, both Lehiste [12], Wang et al. [13] and Rosen [14] asked listeners to judge which of two isolated vowels sounded longer. Differences in experimental procedures may also lie behind conflicting findings by Lehnert-LeHouillier [29] and van Dommelen [30]. The former explored the effect of a falling vs. a level $f_0$ in CV non-sense words through a categorial AXB forced-choice task. Only Japanese, but not German, Spanish, and Thai listeners exploited $f_0$ and perceived a vowel with a falling contour as longer. In contrast, in a series of identification experiments German listeners in [30] were shown to use the $f_0$ cue. The fact that these listeners did not perform a metalinguistic task but identified members of existing minimal pairs might have been crucial in this context. Further, the effect of a rising $f_0$ contour on the identification of Swedish /ɛt/ (*one*) - /ɛ:t/ (*eat;* imp.) in [17] is at odds with the absence of such an effect in /²hak:ər/ - /²ha:kər/. The decisive difference may be the absence of tonal contrast in Swedish monosyllabic words as against the relevance of tonal contrast in the Norwegian disyllables. It could be speculated that the Norwegian accent system imposes certain perceptual constraints that override the effect of a rising contour.

The influence of a falling vs. rising contour (L% vs. H%) in the second syllable of <hakker> - <haker> depending on the direction of the contour in the first syllable is hard to explain. The results suggest that the boundary shifts in perception are not random. Following a flat as well as a rising contour in the first syllable, substituting H% by L% caused a perceptual lengthening of the vowel. These two contours have a high $f_0$ value at the end of the vowel in common. For a falling contour in the first syllable, the effect was the opposite. This interaction may be explained with reference to the production of tones. Faytak & Yu [31] have presented evidence showing that in tone languages, there is a negative correlation between tone height and duration, i.e., the lower the tone, the longer its duration. This production pattern is in line with the shorter *perceived* duration for lower tones in Yu [16] (cf. [11]). Taking this into account, following a high-ending flat or rising contour, the low height of L% would be especially prominent, and the second syllable would sound shorter, in turn causing perceptual lengthening of the first one. The opposite effect of H% - perceptual shortening of the second syllable - would be weaker because of its height matching the end of the flat or rising contour in the first syllable. Following the same reasoning, the influence of L% following a falling contour would be weaker compared to H%. Therefore, the latter would make the second syllable sound longer and the first one shorter.

Finally, it should be pointed out that the results of perception tests involving manipulation of $f_0$ contours as in the current one are inherently difficult to interpret. The token used in the test was originally pronounced with low tone dialect accent 2. Manipulation of its $f_0$ contour resulted, necessarily, in loss of its accent 2 tonal pattern. According to participants' remarks some contours made the impression of representing a high tone dialect or even having no word accent at all. It is thinkable that in this way the Norwegian word accent system interacted with general auditory mechanisms in the perception of duration. This question remains to be investigated.

## 5. Conclusion

The results from the present investigation suggest that Norwegian word accents do not impose any constraints on temporal word structure. Future research using tailored material may shed more light on this issue. It remains an open question why similar accents in closely-related Swedish do have an impact on segment durations. From the perception part it could be concluded that a falling vs. a flat contour in a stressed vowel lengthens its perceived duration. In contrast, a rising vs. a flat contour did not have an impact on perceived duration. The data suggest that the $f_0$ contour of a following unstressed syllable affects the perceptual duration of that syllable, in turn also changing the stressed syllable's perceived duration.

## 6. References


1. Kristoffersen, G. (2000). *The Phonology of Norwegian*. Oxford University Press.
2. Ambrazaitis, G., & Tronnier, M. (2021). Segmental durations as a correlate of Swedish word accents: Evidence from Stockholm and Scania Swedish. *Proc. Fonetik 2021, Centre for Languages and Literature, Lund University, Sweden*, 13-16.
3. Svensson Lundmark, M., Frid, J., Ambrazaitis, G., & Schötz, S. (2021). Word-initial consonant–vowel coordination in a lexical pitch-accent language. *Phonetica*, 78, 515-569.
4. Svensson Lundmark, M. (2022). Evidence of segmental articulations: Acceleration determines vowel segment duration in Swedish Word Accents. *Proc. 1st International Conference of Tone and Intonation (TAI 2021), Sønderborg, Denmark*, 156-160. doi: 10.21437/TAI.2021-32
5. Wu, F., & Kenstowicz, M. (2015). Duration reflexes of syllable structure in Mandarin. *Lingua*, 164, 87–99.
6. Brotzman, R. (1964). Progress report on Mandarin tone study. *Project on linguistic analysis report*, 8, 1–35.
7. Ho, A. T. (1976). The acoustic variation of Mandarin tones. *Phonetica*, 33, 353--367.
8. Nordenhake, M., & Svantesson, J.-O. (1983). Duration of standard Chinese word tones in different sentence environments. *Lund University Working Papers in Linguistics*, 25, 105–111.
9. Whalen, D., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica*, 49, 25-47.
10. Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, 25, 61-83.
11. Gussenhoven, C., & Zhou, W. (2013). Revisiting pitch slope and height effects on perceived duration. *Proc. 14th Annual Conference of the International Speech Communication Association* (*Interspeech 2013*)*, Lyon, France*, 1365-1369.
12. Lehiste, I. (1976). Influence of fundamental frequency pattern on the perception of duration. *Journal of Phonetics,* 4, 113-117.
13. Wang, W. S-Y., Lehiste, I., Chuang, C.-K., & Darnovsky, N. (1976). Perception of vowel duration. *The Journal of the Acoustical Society of America*, 60, S92. doi: 10.1121/1.200360
14. Rosen, S. M. (1977a). The effect of fundamental frequency patterns on perceived duration. *Speech Transmission Laboratory Quarterly Progress and Status Report,* 18(1), 17–30.
15. Cumming, R. (2011). The effect of dynamic fundamental frequency on the perception of duration. *Journal of Phonetics,* 39, 375-387.
16. Yu, A. C. L. (2010). Tonal effects on perceived vowel duration. *Laboratory Phonology,* 10, 4(4), 151–168.
17. Rosen, S. M. (1977b). Fundamental frequency patterns and the long–short vowel distinction in Swedish. *Speech Transmission Laboratory Quarterly Progress and Status Report,* 18(1), 31–37.
18. Van Dommelen, W. A. (1995). Interactions of fundamental frequency contour and perceived duration in Norwegian. *Phonetica,* 52, 180-187.
19. CALST - Computer-Assisted Listening and Speaking Tutor, https://www.ntnu.edu/isl/calst
20. Boersma, P., & Weenink, D. (2021). *Praat: Doing phonetics by computer* [Computer program]. Version 6.1.48. http://www.praat.org/
21. *NB Tale - en grunnleggende akustisk fonetisk taledatabase for norsk [NB Speech – a fundamental acoustic-phonetic speech database for Norwegian].* http://www.nb.no/sprakbanken/show?serial=sbr-31&lang=nb
22. R Core Team (2021). *R: A language and environment for statistical computing.* Retrieved from https://ww.R-project.org/

23. Baayen, R.H. (2008). *Analyzing linguistic data: A practical introduction to statistics using* R. Cambridge University Press.
24. Barr, D.J., Levy, R., Scheepers, C., & Tily, H.J. (2013). Random-effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255-278. doi: 10.1016/j.jml.2012.11.001
25. Steffman, J., & Jun, S.-A. (2021). Tonal cues to prosodic structure in rate-dependent speech perception. *The Journal of the Acoustical Society of America*, 150, 3825-3837.
26. Quené, H. (2007). On the just noticeable difference for tempo in speech. *Journal of Phonetics,* 35, 353–362.
27. Myrberg, S., & Riad, T. (2015). The prosodic hierarchy of Swedish. *Nordic Journal of Linguistics*, 38(2), 115-147. doi: 10.1017/S0332586515000177
28. Myrberg, S. (2021). Using Prosogram to study final rises in South Swedish: Implications for the Scandinavian tone accent typology. *1st International Conference on Tone and Intonation (TAI), Sonderborg, Denmark*, 264-268.
29. Lehnert-LeHouillier, H. (2010). A cross-linguistic investigation of cues to vowel length perception. *Journal of Phonetics,* 38, 472–482.
30. Van Dommelen, W. A. (1993). Does dynamic F0 increase perceived duration? New light on an old issue. *Journal of Phonetics*, 21, 367–386.
31. Faytak, M., & Yu, A. C. L. (2011). A typological study of the interaction between level tones and duration. *Proc. 17th International Congress of Phonetics Sciences*, *Hong Kong, China*, 659-662.