

Constant Tonal Alignment in Swedish Word Accent II

Malin Svensson
Lund University

SPVR02, Language and Linguistics, Phonetics: Master's (One Year) Thesis
Fall semester 2012
Supervisor: Gilbert Ambrazaitis

Abstract

Studies on accentual tonal alignment of intonational languages suggest a segmental anchoring of tonal targets, more specifically that L in rising (LH) pre-nuclear accents anchors with a specific point in the segment, while the timing of H varies. This study tests if lexical accents, too, exhibit a constant alignment by testing the South Swedish word Accent II using speech rate as an experimental tool. When under the strain of tempo variability the L-target was shown not to be anchored with syllable onset, while the H-target anchored close to syllable offset. This could mean that H is an important phonological event in Accent II, while L is not. Further studies are suggested on the following L-target. Results suggest further that additional factors such as individual speaker strategies and effects of speech rate should be taken into account in future studies.

Keywords: tonal alignment, segmental anchoring, word accent, pre-nuclear accent, speech rate

Table of contents

| | |
|--|----|
| Abstract | 2 |
| Constant Tonal Alignment in Swedish Word Accent II | 4 |
| Introduction..... | 4 |
| Method..... | 9 |
| Speech materials | 9 |
| Speakers | 10 |
| Recording and analysis procedure | 10 |
| Results..... | 12 |
| Discussion..... | 17 |
| References | 22 |
| Acknowledgements | 24 |
| Appendix | 25 |
| Target sentences..... | 25 |
| The original material..... | 25 |
| Praat scripts..... | 27 |

Constant Tonal Alignment in Swedish Word Accent II

Introduction

The duality of phonetics and phonology can be explained as the relationship between the world of physically produced sounds and the systems of symbols that compose the sounds of speech. Over a period of 13 years there has been an on-and-off, sometimes lively, debate in intonational phonology on a certain aspect of this relationship. Principally, the debate deals with the matching of the suprasegmental tonal curve with the segments with which it is assumed to be linked. The above-mentioned debate has mainly focused on tonal alignment where tonal targets, rather than curve shapes, are aligned with the segmental string. Tonal alignment might be seen upon as a wider notion for other concepts such as timing, tonal association or segmental anchoring. In the present study *alignment* is used as the overall concept for alignment between tones and segments and *anchoring* is used when referring to constant alignment.

The idea of constant alignment of the tonal targets is in opposition to that of a fixed interval distance between the tonal targets. A fixed interval distance would, in theory, account for different shapes of the tonal curve, for example the higher the excursion the longer the rise time. However, this hypothesis has been more or less rejected by the growing body of evidence in favor of constant alignment (Arvaniti, Ladd & Mennen, 1998; Atterer & Ladd, 2004; Ladd, Faulkner, Faulkner & Schepman, 1999; Xu, 1998; Dilley, Ladd & Schepman, 2005). Ladd et al. (1999) instead proposed in their study that the anchoring and F0 excursion of the tonal targets together determine the shape of the accent. They found evidence on the basis of which they discarded the interval distance hypothesis and instead advocated the “segmental anchoring principle”, a principle originally suggested by Arvaniti et al. (1998).

Constant alignment has been tested in several languages and with various experimental tools/factors, such as different segmental construction, tonal contexts or speech rates. Most studies have focused on pre-nuclear rising accents in intonation languages, with the syllable as the main host for anchoring. The present study is a new approach since it concentrates on the production of word accents in Swedish with speech rate as an experimental tool. To the best of my knowledge, a language with lexical accents has not been taken into account in the recent research on constant alignment.

The use of speech rate as an experimental tool is based on the idea that speakers will try to retain primary features of phonological properties, while they will let other features be modified under time pressure (Caspers & Van Heuven, 1993). Speech rate has been used successfully by a number of researchers in studies concerning tonal alignment. Caspers and

Van Heuven studied Dutch pre-nuclear rising accents and tested various phonetic features with different speech rates. They found that some of the phonetic features were affected by speech rate, for example the rise was affected but not the fall. Although tonal alignment was not the main focus of Caspers and Van Heuven, they interpreted the results as the rise showing alignment with the segment, but for the fall other features such as shape were believed to be of greater importance.

Speech rate is normally defined by excluding pause time and is measured in average syllable duration (ASD) or syllables per second (see for example Crystal & House, 1990; Quené, 2008). An effect of speech rate has been recognized on F0 excursion; however studies are contradictory (see Ladd et al., 1999, for discussion). Because of the inconclusiveness of the effect of speech rate, Ladd et al. tested the experimental factor of speech rate in their study on tonal alignment in English pre-nuclear accents. They found a robust effect of speech rate on segment duration in all three speech rates for all speakers included in the study. There was also an effect of speech rate on rise duration in pitch accents for all speakers, but no effect on excursion size. Since speech rate had both an effect on segment length and rise time for pitch accents, they interpreted this as support for a model for tonal targets linked to specific points in the segmental string, the “segmental anchoring principle” (Ladd et al., 1999).

Speech rate has also been used as an experimental tool in a study on tonal alignment in Mandarin (Xu, 1998). In this study Xu combined the variables of speech rate and segmental composition. Xu’s main concern was to find the base segment for alignment of different lexical tones, thus the comparison with different segmental compositions was central. The results confirmed that across the three speech rates the same anchoring points with the syllable was found. By using speech rate as an experimental tool, Xu was able to identify the syllable as the main domain of lexical tones in mandarin.

Even though the growing use of speech rate as an experimental tool might signal that the experimental setting is valid, it presupposes the use of the auxiliary hypothesis that primary phonetic features of phonological entities will not be affected by speech rate. Perhaps because speech rate is an unverified variable, tonal alignment has also been tested with factors such as tonal context or segmental construction, the latter already mentioned above in the study from 1998 by Xu. Arvaniti et al. (1998) made use of segmental constructions for testing constant alignment with segments in Greek pre-nuclear accents. Their results showed that rise time varied between segmental compositions while F0 excursion stood intact, indicating that tones anchor with segments regardless of their duration. They also found that both the beginning and the end of the rise, L and H, were anchored. The L-target occurred before

syllable onset, while the H-target occurred after syllable offset, thus the tonal gesture rose all through the syllable. Arvaniti et al. saw this as evidence in favor of the segmental anchoring hypothesis and suggested their results to be interpreted as tonal targets being important distinct phonological events.

In their study on English pre-nuclear accents Ladd et al. (1999) were able to second the results of Arvaniti et al. on Greece pre-nuclear accents. Ladd et al. also found evidence in favor of constant alignment of the L-target with syllable onset, and the rise was again followed through the syllable. The H-target was found somewhere late in the consonant that followed the stressed vowel. In a second study on English from 2005, Dilley et al. found that the distance between L and H, the rise time, was in proportion with segment duration. Both Dilley et al. and Ladd et al. interpreted their results as evidence in favor of the segmental anchoring principle and considered it to be a valid model.

In a cross-linguistic study on English and two German dialects Atterer and Ladd (2004) also interpreted their results as being in support of tonal anchoring. The L target constantly aligned with syllable onset and H with the end of the syllable. Nonetheless, they did find small differences between the languages. Revising the view by Arvaniti et al. (1998) that tonal targets are distinct phonological events, Atterer and Ladd explained the cross-linguistic differences as being due to different phonetic realization rules. Instead of belonging to discrete patterns of alignment the segmental anchoring of the rise belongs to a continuum of phonetic realizations (Atterer & Ladd, 2004).

Regardless of whether studies on tonal alignment support the constant segmental anchoring hypothesis or not, many of them display the same pattern, namely that L aligns with syllable onset. Results include the L-target occurring just before the onset of the accented syllable (Arvaniti et al., 1998, for Greek; Niemann, Mücke, Nam, Goldstein, & Grice, 2011, for Italian), at syllable onset (Caspers & Van Heuven, 1993, for Dutch; Ladd et al., 1999, for English; Atterer & Ladd, 2004, for German), or after syllable onset (Xu, 1998, for Mandarin). The precise timing seems to vary, but the studies do show anchoring of L with the beginning of the syllable, while the same consistent result does not exist for the H target. Even though several studies have found H anchoring, not all of them have. Results include H anchored after syllable offset (Atterer & Ladd, 2004; Xu, 1998; Arvaniti et al., 1998), or somewhere late in the syllable (Ladd et al., 1999). Nonetheless, Ladd et al. discovered inconclusive results on the alignment of H, where the slow speech rate and two out of six speakers varied considerably from the others. When excluding the two speakers and the slow speech rate, they found a much more constant alignment of H somewhere late in the syllable. Ladd et al.

concluded that the slow speech rate allured individual speaker strategies and inserted unexpected pauses in the material which affected the measurement of F0 maximum. Caspers and Van Heuven (1993) found that the end of the rise, the H-target, varied considerably under time pressure and thus discarded that it did anchor with the segment. However, they did also consider whether or not the variation had to do with segmental construction.

It can be concluded that the L-target appears to be more stable than the H-target in rising pre-nuclear accents in intonation languages. Niemann et al. (2011) tried to explain the phenomenon by confronting the cross-linguistic results that displayed the differences and the discussion on whether that is because of phonological properties or phonetic details. They interpret the strong anchoring of L target with syllable onset in rising accents as the start of a tonal high gesture, which aim to reach a certain F0 level, the H-target. The alignment of the tonal gesture competes with the oral gestures; the vocalic and consonantal gestures, and the competition forces the precise alignment of the H-target to vary. According to Niemann et al. the cross-linguistic variation is systematic and can be explained by different phonological structures between the intonation languages.

Another turn in the debate on tonal alignment has been taken by Niebuhr, D'Imperio, Fivela and Cangemi (2011). They addressed the effect of individual speaker strategies when defining pitch accent, and why the timing of tonal targets insufficiently only accounts for some of the speakers. Niebuhr et al. propose that segmental anchoring might be a statistical approach and that it is instead other related features that are responsible for the consistent results. Their proposition contains the analysis of individual speaker strategies whenever investigating tonal alignment.

Features other than segmental anchoring that define pitch accents have been found elsewhere. Caspers and Van Heuven (1993) looked into falling accents and found that neither the beginning nor the end, neither L nor H, of the falling accent anchored. Their conclusion was that alignment, as a feature, was of more importance for the rising accent. To the best of my knowledge, the only study on alignment in a tonal language other than Swedish found different features as well for falling and rising tones (Xu, 1998). Xu found that in the rising tone (tone 2) in Mandarin, the L-target tended to occur after syllable onset, towards the center of the syllable and the H-target towards syllable offset. But for the falling accent no anchoring was corroborated. In other words, there might be both inter and intra-individual speaker strategies, as well as different strategies for rising and falling accents.

One of the first accounts for the timing of tonal targets and an attempt to categorize the prosodic variation in a tonal language can be found in the prosodic typology of Swedish

intonation in the Lund Model (Bruce & Gårding, 1978; Bruce, 2007). One of the key features of the earlier version of the Lund Model is that the two Swedish word accents are assumed to be represented by a fall in the stressed syllable in a prosodic word, where the two accents differ in timing of the fall. In the Central Swedish dialect (Svea) the high level in Accent I is associated with the pre-stressed syllable and in Accent II with the stressed syllable. Although the pitch gesture of Accent I precedes Accent II in all Swedish dialect types (Svea, Göta, North, Gotland, Dala and South) there is a variation between the Swedish dialects which is also included in the Lund Model. For example in the South Swedish dialect (South) both accents are timed considerably later than in the Central Swedish dialect (Svea): the high level in Accent I is associated with the stressed vowel and in Accent II with offset of the stressed syllable. The first categorized variation between dialects has later been revised by Bruce (2007), who instead made a rough categorization into two groups - early and late dialect types. Bruce also identified, for all dialects, an LHL tonal gesture from which bitonal gestures are extracted; either a fall, H+L, or a rise, L+H. Bruce generalized an association of a fall in the dialects with an early timing of the accents (Svea and Göta) and a rise in the dialects with a late timing (South, Gotland, Dala, North). For the South Swedish dialect, a late timed dialect type, Bruce made the specific assumption of a fall, an H+L pattern, for Accent I and a rise, an L+H pattern, for Accent II. A reason for the distinction between the early and the late timing of word accents can be explained by the existence or non-existence of the focal word accent in the dialects, as Bruce proposes. The revised prosodic typology includes the old distinction between single-peaked and double-peaked dialect types, meaning that word accents in focused position attract an additional tonal peak in the double-peaked dialects while in single-peak dialects the focus is believed to be signaled by an enhanced tonal gesture of the word accent. The revised approach instead distinguishes the early dialect types, the same as the double-peaked types, as attracting an additional focal gesture, as opposed to the late dialect types, the single-peaked types, that seem to lack the prominence level of focus accentuation altogether. With the time pressure of the extra focal gesture the word accent gesture in the early dialect types necessitates an earlier timing, while as Bruce explains:

“...compared with the corresponding situation in South Swedish, where the accentual gesture is less complex consisting of a simple rise-fall with no extra focal pitch peak. In this case, therefore, there is no extra time pressure on the pitch realization of the corresponding accent II gesture within the time window. This gives room for a relatively later timing of the accentual gesture.” (Bruce, p 143, 2007)

Although prominence level is not under investigation in this study, the explanation by Bruce on the categorization of early versus late dialect types in the Swedish language is an interesting aspect of time pressure on the realization of pitch gestures. Apart from focal accentuation Bruce states phrasing as another feature that might affect timing of the tonal gestures in the Swedish word accents.

To conclude: The rising gesture of the lexical word Accent II in South Swedish is assumed to have a late timing with the stressed syllable. As if by chance, there is a rough phonetic match with the pre-nuclear accents in the already mentioned studies on tonal alignment in intonation languages. The rising accents in the intonational languages seem to phonetically match the timing of the lexical Accent II in South Swedish. It would be interesting to see whether or not additional phonetic features are similar or whether both the low and the high target, L and H, are anchored with a segment, as is assumed by the constant segmental anchoring principle. Previous studies have displayed an unambiguous case of the tonal target L aligning with syllable onset in pre-nuclear accents, though the precise timing seems to vary across languages. Comparing lexical accent with pre-nuclear accent might tell us something on the phonological status of an accent and would make an important contribution to the field.

The present study is a production study where the hypothesis of constant tonal alignment is tested on the Swedish word Accent II. The hypothesis is that if tonal targets anchor with the syllable, then this anchoring would not alter when under the strain of tempo variability. More specifically, if speech rate does have an effect on timing of the tonal targets then this would be evidence against the segmental anchoring principle as a valid model for word accents. Because the (new) Lund Model assumes that both L and H are phonologically relevant, anchoring of both L and H is expected.

Method

Speech materials

The material used was initially recorded by the researcher for a different study and originally consisted of 40 test sentences. From the original material, three semantically identical words were extracted that fit the criterion of the current study; an unbroken tonal curve, a word Accent II, identical segmental surroundings and that neither syllable and vowel onset, nor syllable and vowel offset coincided. Additional conditions of the target words such as location in the original material can be seen in Table 1. The target word “många”, [‘mɔŋa], is an adjective, meaning *many* in English.

| Sentence | Word | Segmental surroundings | Placement | Syllable structure | Accent |
|----------|-------|---|-------------|--------------------|--------|
| 2 | många | [9 syllables] bisyllabic target word [2 syllables] | Pre-nuclear | CVCCV ONCd.ON | II |
| 22 | många | [9 syllables] bisyllabic target word [2 syllables] | Pre-nuclear | CVCCV ONCd.ON | II |
| 36 | många | [9 syllables] bisyllabic target word [2 syllables] | Pre-nuclear | CVCCV ONCd.ON | II |

Table 1. Segmental surroundings, location in material and syllable structure of the target words.

Speakers

The material was initially recorded for a different study in which two age groups were recorded. For this study only the older speakers were tested due to technical issues. Since they made up the largest age group and thus enabled more test material, this was not seen as a problem. There are no indications that tonal alignment is affected by speaker age. All speakers were voluntary and spoke the same variety of the South Swedish dialect. Criterion for the speakers was that they had all lived most of their lives in the same area in the northeastern part of the South Swedish region. Moreover, their parents also had to have lived most of their lives in the area. There were seven speakers, four males and three females, and the average age of the speakers was 72 years. Below is a list of the speakers.

- M87 - Man, 87 years old.
- M83 - Man, 83 years old.
- F83 - Woman, 83 years old.
- F67 - Woman, 67 years old.
- M64 - Man, 64 years old.
- M63 - Man, 63 years old.
- F58 - Woman, 58 years old.

Recording and analysis procedure

The original material was read twice by each speaker at each speech rate; normal, slow and fast. The recording leader set the pace of the speech rate with the leading question and the speaker was asked to answer the question and to follow the speech rate of the recording leader. Normal speech rate was controlled through the normal speaking rate of the speaker. For the slow speaking rate a visual metronome was used, only for the recording leader to see. Again the recording leader set the pace. For the faster speech rate, the speaker was asked to speed up as fast as possible. An IMG Stage boundary microphone (table-microphone) with phantom power was used (ECM-302B). This particular setting was chosen

because of the different recording environments; all of the recordings were recorded in people's homes and the recording facilities were easy to bring along and to set up. A table microphone was chosen since it is non-invasive and the speakers were expected to be naïve with no prior recording experiences. A headset would have been preferred but the quality of the available headsets was not good enough. Because of the different recording environments, nuisance variables could not be under control. However, the recording devices were the same. Accommodation effect was held as constant as possible; the same person (the researcher) read all the leading sentences and made all the recordings. Each speaker was recorded in 48 kHz on a Marantz digital recorder (PMD660). The uncompressed 16-bit wav-files were then transferred onto a laptop where analyses took place.

The researcher performed segmentation and annotation in Praat (www.Praat.org). Since each speaker was recorded twice, the material consisted of 126 items. For each speaker this means six target words at each speech rate. Each target word was segmented into syllables and the accented vowel (Figure 1). Since the border between the two syllables was difficult to distinguish, offset of accented syllable was calculated to occur in between offset of accented vowel and onset of the following vowel. Because of testability reasons the tonal curve has been semi-automatically annotated for the tonal targets L and H. If there was a visible tonal peak, the target word has been marked by the researcher and the highest target was automatically extracted using the "Move cursor to maximum pitch" in the Praat menu. This procedure was used even when there was a tonal plateau. When the tonal curve was broken and a clear, visible tonal peak was missing, the "Move cursor to maximum pitch" was not used and the item was labeled by the researcher as missing data. For the preceding low target, i.e. the start of the rise, it was also extracted semi-automatically, but using instead the "Move cursor to minimum pitch". For this, only the first syllable and the preceding vowel were marked by the researcher. If the automatically extracted low target was followed by a disrupted rise uncovering yet another low target, the automatically extracted low target was ignored and the second one was instead manually extracted by the researcher. When the tonal curve was broken and a clear visible tonal valley was missing, the "Move cursor to minimum pitch" was not used and the item was labeled as missing data.

The data was collected with Praat scripts (see Appendix). Each speaker suffered from missing data and any target words that missed either L or H were excluded. Across all seven speakers the normal speech rate contained 39 usable items, the fast speech rate 19 items and the slow speech rate 26 items. Each speaker had at least one item in each speech rate, most

had several. The data was organized in Microsoft Excel and then transferred to IBM SPSS program for statistics.

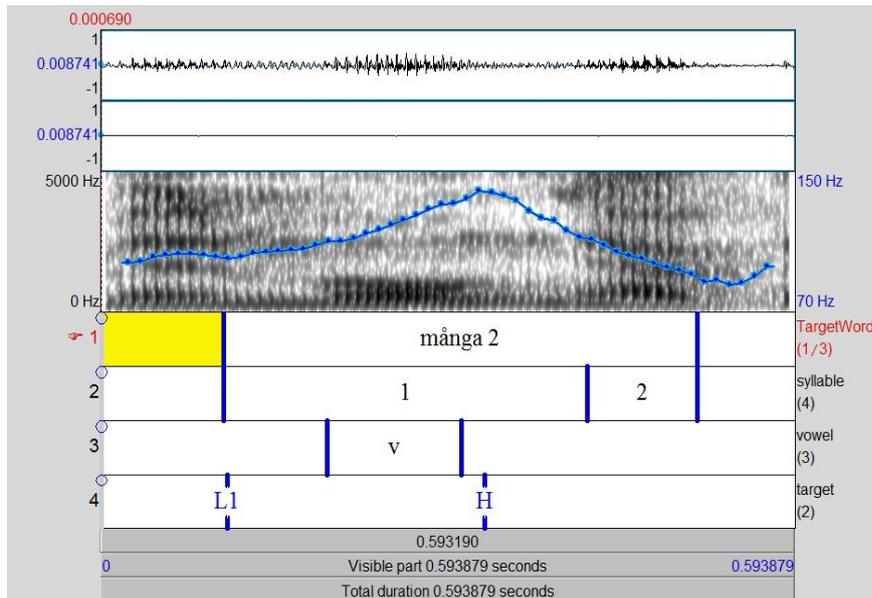


Figure 1. Image of praat window with three interval tiers and one point tier. Target word 2 as pronounced by speaker M63 in normal speech rate. Tier 2, *syllables*, did not mark syllables, but instead the borders signified onset of stressed syllable as well as onset and offset of following vowel.

Results

Descriptive statistics show an apparent difference in speech rate between the recordings. The stressed syllable of the target word was measured. Syllable duration for all items and all speakers is for slow speech 291 ms (standard deviation 46), normal speech rate 218 ms (standard deviation 34) and fast speech rate 178 ms (standard deviation 28). Looking at each speaker, they all show the same pattern of decreasing syllable duration for each condition (see Figure 2).

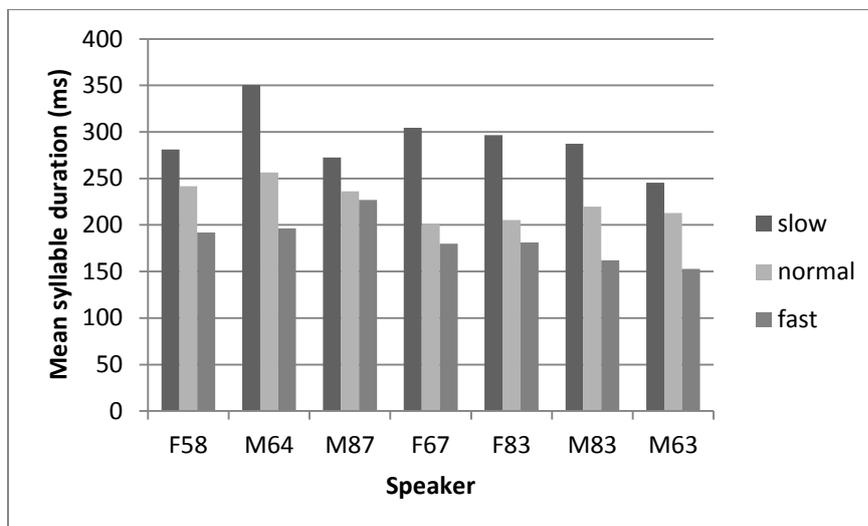


Figure 2. Graph of mean syllable duration for each speaker in each speech rate.

To test the effect of speech rate on the collected data two different tests have been used for the inferential statistics. One of them is a repeated measures design of test of within subject effects, with the random sample of speakers. In the first ANOVA, speech rate was labeled as the independent variable and syllable duration as dependent variable. To account for missing data, mean values across the obtainable items for each speaker were used as raw data. The ANOVA showed that speech rate had a significant effect on syllable duration ($F=66,490$, $df 1$, $p < .000$). A pairwise comparisons test (Bonferroni) found a significant difference between all three speech rates: slow and normal speech rate ($p = .003$), normal and fast speech rate ($p = .008$) and slow and fast speech rate ($p = .001$). It can be concluded that the rate manipulation was successful, given the large effect of speech rate on segment duration, and that consequently additional statistics can be followed by using speech rate as an experimental tool.

Since the segments are affected by speech rate, the distance between the tonal targets L and H should also be affected by speech rate, unless there is a fixed interval distance between them. The mean distance between timing of tonal targets for each speaker was tested in an ANOVA with distance between L and H as dependent variable and speech rate as independent variable. The ANOVA displays a significant effect of speech rates ($F=17.129$, $df 1$, $p < .006$). The Bonferroni pairwise comparison test showed significant difference only between slow and fast speech rate ($p = .018$) and not between slow and normal speech rate ($p = .927$), nor between normal and fast speech rate ($p = .348$). Since the overall ANOVA showed statistical significance it does, however, imply as expected that the distance between the tonal targets L and H is not constant.

The effect of speech rate on distance between tonal targets indicates that the distance between them is affected by segment duration due to alignment with the segment. If there is some sort of anchoring of the tonal curve, this would necessitate a correlation between segment duration and distance between tonal targets. In order to test the relationship further the second inferential statistic method used in this study, beside the ANOVA, is the Correlations Pearsons (2-tailed) test. The correlation is tested between syllable duration and distance between tonal targets L and H. Since there is an uneven amount of missing data for all speech rates and all speakers, the correlation is calculated independently of speaker or speech rate. There appears to be a weak to moderate positive linear relationship ($R = 0.433$, $N = 74$), which indicates a correlation. However, there does not seem to be a convincingly strong correlation. The scatter plot shows the weak to moderate positive relationship between the two factors. It also shows that the relationship is linear and not logarithmic, which could

have accounted for the weak to moderate relationship, and that there is a difference between the speech rates.

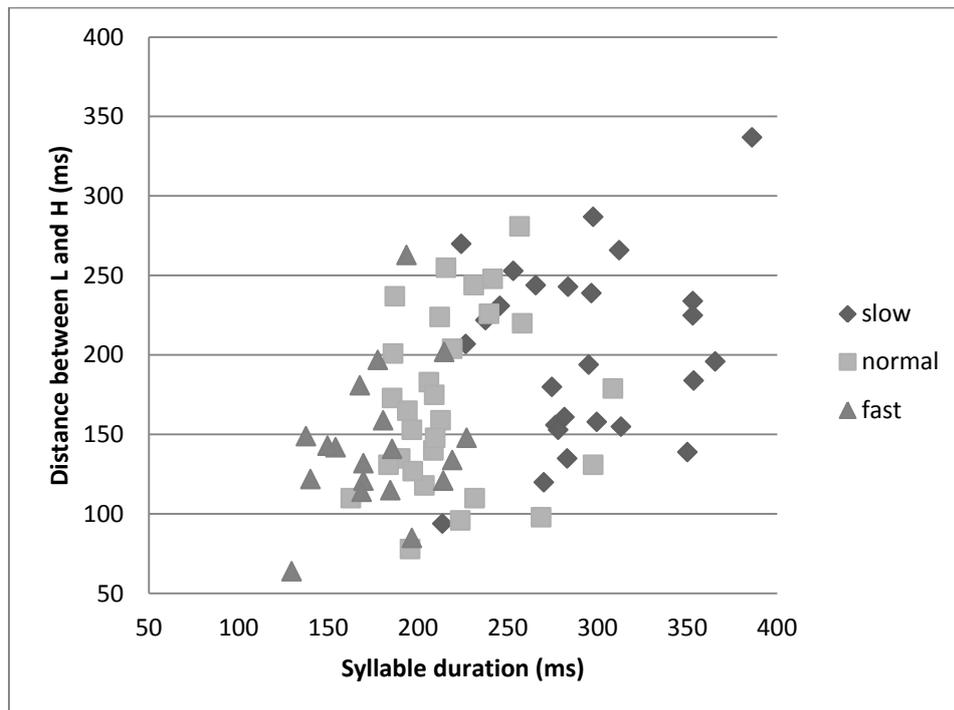


Figure 3. Scatter plot of the relationship between syllable duration and the distance between tonal targets L and H, with the three speech rates included.

If both targets anchored at a specific point then the relationship would be expected to be stronger. The weak to moderate relationship might indicate that either only one of the targets is anchored or that neither of them is anchored. To test whether there in fact is no anchoring of any of the targets with the segment, a repeated measures design of test of within subject effects is used. In the ANOVA with speakers as random sample speech rate was labeled as the independent variable and anchoring as dependent variable. To account for missing data, mean is first calculated for each speaker across the available items for each condition. Anchoring is measured as distance in milliseconds between targets L and syllable onset, and H and syllable offset, as proposed by earlier studies.

If speech rate has an effect on the distance between H and syllable offset, this is evidence against the segmental anchoring principle. There was no significant effect of speech rate on anchoring of H, as expected ($F = .702$, $df = 1$, $p = .434$). The graph displays the possible anchoring that seems to occur late in the syllable, close to syllable offset (Figure 4).

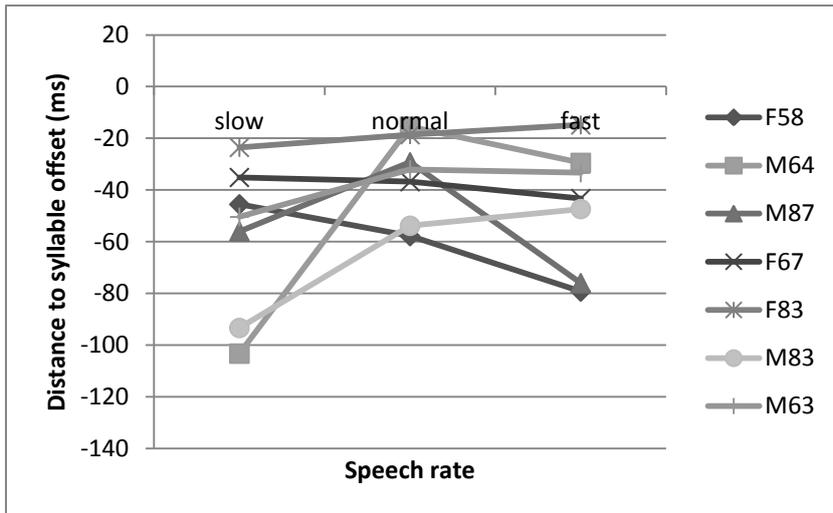


Figure 4. Graph displaying the mean distance between H and syllable offset for each speaker. A negative number indicates that timing is before offset.

Next, the anchoring of L with syllable onset is tested, with the same conditions as above for the H target. An ANOVA confirms that there is no significant effect of speech rate on the distance between the tonal target L and syllable onset ($F=0.460$, $df 1$, $p= .523$). However, even though inferential statistics state there is no effect, the graph below sends a different message. Figure 5 displays that for three of the speakers (F58, F67, M83) there is a pattern of an effect of speech rate, while for the other speakers there is an apparent larger variance.

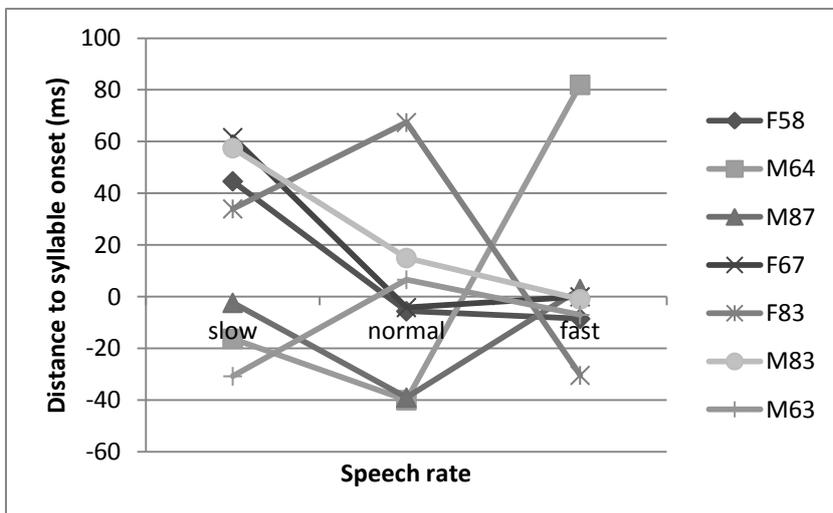


Figure 5. Graph displaying the mean distance between L and syllable onset for each speaker. A negative number indicates that timing is before onset.

In Table 2 below the alignment data is collected, displaying quite a large standard deviation as well as the anomaly of only one available item in some of the conditions for three of the speakers (M64, M87 and F83), which might explain the variance displayed in Figure 5.

To avoid a type II error, additional ANOVAs were made, this time with target words as random sample.

Mean distance between target and segment border (ms)

| Speaker | L - syllable onset | | | H - syllable offset | | |
|--------------|--------------------|----------|----------|---------------------|----------|----------|
| | slow | normal | fast | slow | normal | fast |
| F58 | 45 (47) | -6 (97) | -9 (1) | -45 (18) | -58 (14) | -79 (32) |
| M64 | -16 (22) | -40 (*) | 82 (*) | -103 (49) | -16 (*) | -29 (*) |
| M87 | -2 (43) | -39 (44) | 3 (*) | -56 (32) | -29 (48) | -76 (*) |
| F67 | 62 (56) | -4 (9) | 0 (4) | -35 (21) | -37 (10) | -43 (17) |
| F83 | 34 (*) | 67 (39) | -30 (28) | -24 (*) | -19 (13) | -15 (18) |
| M83 | 57 (26) | 15 (38) | -1 (45) | -93 (9) | -54 (20) | -47 (7) |
| M63 | -31 (36) | 7 (29) | -7 (8) | -50 (26) | -32 (25) | -33 (31) |
| All speakers | 24 (51) | 9 (54) | -6 (36) | -65 (37) | -38 (23) | -40 (27) |

Table 2. Table displaying mean distance between tonal target and segment border (ms). Standard deviation in parentheses.

* Only one available item in this condition.

In the ANOVA, with target word as random sample, speech rate was labeled as the independent variable and the anchoring as dependent variable. To account for missing data, the mean is first calculated for each target word across each condition. The ANOVA for the effect of speech rate on the distance between L and syllable onset showed that speech rate did in fact have a significant effect ($F= 14.095$, $df 1$, $p= .013$). A pairwise comparisons test (Bonferroni) found a significant difference between slow and normal speech rate ($p= .034$) and slow and fast speech rate ($p= .013$), but no significant difference between normal and fast speech rate ($p= .062$). In other words, alignment of L with syllable onset is affected by speech rate. This is evidence against the segmental anchoring principle. Below is a graph displaying the mean of the six target words in each speech rate (Figure 6).

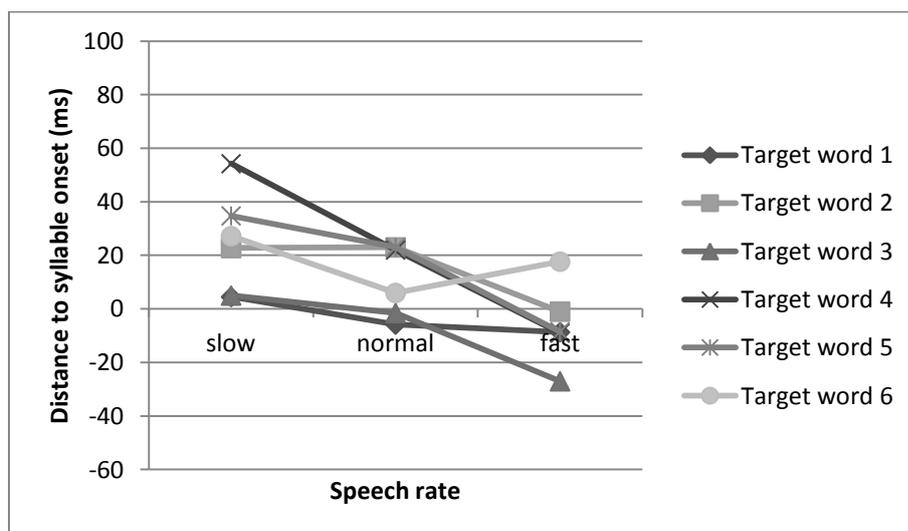


Figure 6. Graph displaying the mean distance between L and syllable onset for each target word. A negative number indicates that timing is before onset.

An ANOVA on the distance between H and syllable offset is also calculated. The results show a low p-value; however not statistically significant ($F= 5.159$, $df 1$, $p= .072$). Speech rate appears to not affect the possible anchoring of H. Means and standard deviations are shown in Figure 7 and Table 3.

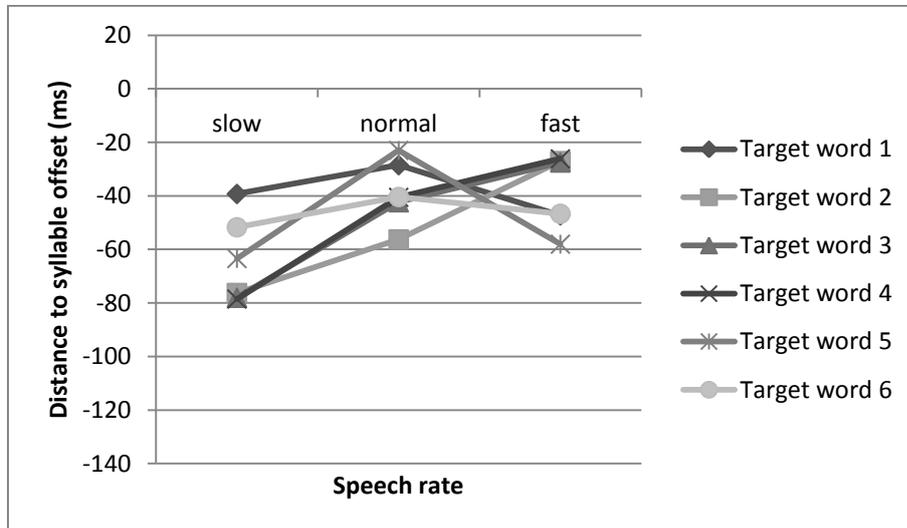


Figure 7. Graph displaying the mean distance between H and syllable offset for each target word. A negative number indicates that timing is before offset.

Mean distance between target and segment border (ms)

| Target word | L - syllable onset | | | H - syllable offset | | |
|------------------|--------------------|---------|----------|---------------------|----------|----------|
| | slow | normal | fast | slow | normal | fast |
| 1 | 4 (22) | -6 (37) | -9 (37) | -39 (32) | -28 (25) | -47 (37) |
| 3 | 23 (65) | 23 (51) | -1 (18) | -76 (25) | -56 (19) | -27 (37) |
| 3 | 5 (50) | -2 (59) | -27 (34) | -78 (54) | -42 (19) | -28 (17) |
| 4 | 54 (56) | 22 (78) | -9 (11) | -79 (39) | -40 (21) | -26 (22) |
| 5 | 35 (50) | 23 (59) | -8 (9) | -63 (33) | -23 (17) | -58 (16) |
| 6 | 27 (64) | 6 (60) | 18 (59) | -52 (30) | -40 (29) | -47 (18) |
| All target words | 24 (51) | 9 (54) | -6 (36) | -65 (37) | -38 (23) | -40 (27) |

Table 3. Table displaying mean distance between tonal target and segment border (ms). Standard deviation in parentheses.

Discussion

It was expected to find that the tones aligned with the segments, which was tested through the correlation between segment length and rise time, and the effect of speech rate on each of them. However, the place for alignment seemed to vary both between and within speakers, and between speech rates. As for constant alignment, the results actually showed evidence against segmental anchoring of the L-target, but no conclusive evidence against H anchoring. This can be interpreted as meaning that the end, but not the start, of the rise of a word accent has a tendency of anchoring with a segment.

The rise has been examined in this study since earlier studies indicated that across languages the start of a rising accent was anchored around syllable onset. This study does not support the anchoring of L in a lexical accent; instead evidence has been found against anchoring. It might be that L is not relevant for Accent II in the South Swedish dialect. The rise is surely an important feature of the word accent, but if the start of the rise is not anchored it is possible that L is not a phonological event and where it starts is not relevant. The end of the rise, H, might in fact be an important phonological feature, and the timing of the target essential. Independent of syllable duration the timing of H was found approximately 40 ms before syllable offset. The measurements are not based on relative duration of segments and the location of alignment has not been tested in this study, yet the indication that the timing of H is about 40 ms before syllable offset supports The Lund Model and the accent typology which incorporates the South Swedish dialect. In the earlier version of the Lund Model the association of the start of the fall, H, was found to vary across the different dialects and was seen as an important feature of the accent. The revised version included a falling pattern in, HL, in Accent I and a rising pattern, LH, in Accent II in the South Swedish dialect. This study does not support an LH tonal gesture in Accent II in the given dialect, since L was not anchoring with syllable onset. Further studies are suggested on the alignment of L given the uncovered variability between speakers.

The results raise some questions regarding the variability of tonal alignment. The data displayed a great variability both between and within speakers. The variability between speakers might be because of small samples of some of the speakers. The anomalies that were expected to underlie this phenomenon were at first interpreted to be the reason for why no effect of speech rate was found in the ANOVA on speakers. Even though the ANOVA on target words enabled larger samples, they still included those speakers that in one way or another diverted from the others. Other studies have had problems with certain speakers that seem to have a different strategy to define pitch accent and when excluding them they were able to find support for segmental anchoring (Ladd et al., 1999). It might be that constant alignment is a strategy for most of the speakers but for some of the speakers other features are used to define pitch accents. It might also be that the same speaker uses different strategies for aligning the tones to the segment. This is a possible interpretation of the displayed anomalies of larger standard deviation for some of the speakers. The proposition by Niebuhr et al. (2011) to include speaker strategy in investigation on tonal alignment is thus a valid suggestion.

In cross-linguistic studies on pre-nuclear rising accents it has been found that L anchors close to syllable onset. A possible explanation for this is that a tonal gesture starts at

syllable onset (Niemann et al., 2011). The explanation also includes that speech gestures compete with the aligning of the tonal gesture and thus create the different anchoring data among dialects and languages, which is found in the timing of the end of the tonal gesture that varies due to the competition with vocalic gestures. In this study on word accents contradictory evidence has been found; L is not constantly aligned, while H seems to be. For the hypothesis of a tonal gesture to still be valid it could be that the timing of the end of the tonal gesture in lexical accents is stable because it has no real competitors. That L seems not to be anchored would in that case mean that the start of the tonal gesture instead has to compete with speech gestures, which is why it would vary both within and between speakers. Such an interpretation would support the view by Niemann et al. The relationship between the distance of tonal targets (the rise time) and the segment length is linked to this. Is there a casual relationship and what happens if two tonal gestures compete? It would be interesting to further investigate the notion of tones as gestures in Swedish accents.

There is a risk that the statistical results are not displaying the variability of tonal alignment, because of the effect of the independent variable speech rate. Speech rate, especially slow rate, might be affecting speech in a way that contradicts the auxiliary hypothesis on the effect of time pressure on important features of speech. The ANOVA on speakers showed no significant difference between any of the speech rates on the anchoring of either L to syllable onset or H to syllable offset. The results of the first ANOVAs do not mean that there is no difference between the speech rates. It is therefore not advisable to interpret the results as a tendency towards segmental anchoring, since it can also indicate that there was a difference between the speech rates but that the samples were too small in order for this difference to be found. According to the descriptive statistics in this study, for the normal and fast speech rate, four out of seven speakers showed no difference between the rates and L seemed to be aligned constantly with syllable onset; however standard deviation indicates a somewhat wide spread on some of them. Moreover, for the normal and fast rate no significant effect of speech rate was found on the anchoring of L. The effect of speech rate was only significant in comparing with slow speech rate. The auxiliary hypothesis that primary features of phonetic properties will try to be retained by speakers, while other features will be allowed to be modified by time pressure might be the case for the normal and the fast rate. However, the slow rate seems to divert from the others and induces other compensatory prosodic features than the normal or the fast speech rate. This can also be seen in the scatter plot of the correlation test (Figure 3) with an apparent anomaly in the slow rate, which is much more scattered across the graph than the normal or the fast rate. The anomalies found in this study

on slow speech rate have also been reported in other studies, where difficulties with the slow speech rate seem to have brought forth additional prosodic features to enable the, perhaps, unnaturally slower speech. In conclusion, as Ladd et al. stated: “This seems to confirm that there is something special about slow rate.” (Ladd et al., p 1551, 1999)

Regarding speech rate it should also be noted that for many of the speakers there seems to be an unexpected effect of speech rate. When measuring distance between L and syllable onset, compared to normal rate, the slow speech rate tends to shift the start of the rise later, while the faster speech rate starts the rise earlier. The H target aligning with syllable offset has the opposite phenomenon: the slower speech rate shifts the end of the rise earlier in the syllable and in the normal and faster speech it is later. This might be an anomaly due to the measurements not having been made on relative duration of segments, but it might also indicate something that is essential – because you have more time in slower speech you are able to act out more of the prosodic features, whilst in faster rate efficiency is more crucial. It could very well be that the most essential feature is the effect of a rise that follows through the accented syllable and in faster speech rate that is the one feature you need to accomplish. In other words, faster speech rate would induce less important features to be excluded. This discussion on speech rate points out a problem with the use of the auxiliary hypothesis that primary features of phonetic properties will try to be retained by speakers, while other features will be allowed to be modified by time pressure. Even though the results of the study confirmed that the manipulation of speech rate was successful, a future use of speech rate as an experimental tool needs to be further investigated.

It is unclear whether the statistical results display the constant nature of tonal alignment that the segmental anchoring principle presupposes. The accuracy of measurements needed may not be ensured. This leads to the testability of the study. In recent years there have been claims that tones are aligned with articulatory events rather than with acoustic data. In phonetic research phonetic knowledge is essential for analyzing acoustical data, since articulatory features constantly need to be taken into account when segmenting and annotating data. It takes phonetic craftsmanship to interpret acoustic anomalies in the data, which is a problem in testability aspects. For example, concerning the measurement of tonal curve it has been suggested that turning-points alone cannot account for listener categorization (Barnes, Veilleux, Brugos & Shattuck-Hufnagel, 2010). Moreover, a tonal plateau confounds the notion of the H-target, the top of the tonal curve. Different measurement methods have been elicited but since there is no consensus the measurement of tonal targets does not have the

highest attainable degree of testability. Thus, in phonetic research it is a challenge to ensure that no arbitrary measurement occurs.

Tonal targets indicate where a tonal curve starts and where it ends. An LHL tonal gesture consists of the rise and the fall of a tonal curve. It has been reported in other studies that for falling and rising accents different features might be important. In the introduction the coincidence was pointed out that the rise of the pre-nuclear accent in intonation languages phonetically roughly matched the lexical Accent II in South Swedish. According to the results there is no matching phonologically, the L-target of the rising pre-nuclear accent in an intonation language needs to be anchored, which does not seem to be the case for a South Swedish Accent II rise. Since the (new) Lund Model does state an LH gesture for Accent II, further investigations are proposed on the hypothesis raised here as to why the L-target is allowed to vary while the H-target needs to be constant. Moreover, the Lund Model proposes a falling pattern for Accent I, HL, in South Swedish. Since falling accents have hardly been studied before it would be relevant to investigate further which features are important for Accent I.

For the Swedish dialects, Bruce (2007) proposed an LHL tonal gesture from which different bitonal gestures are extracted. Since the study brought evidence against segmental anchoring of L in Accent II, this would be in favor of the old version of the Lund Model. It could also indicate that Accent II in South Swedish dialect is categorized by an HL gesture instead of an LH gesture. However, the discussion on speaker strategies and experimental tools points to the need for further investigation into the anchoring of L with syllable onset in Accent II. The scope of this study does not include the fall of the Accent II. If an investigation on the following L-target in Accent II would find evidence towards constant alignment, this would strengthen the suggestion that the categorization of Accent II should shift from a rising pattern to a falling pattern. On the other hand, evidence that would actually include both rise and fall of an accent would open up for the possibility of discussing the notion of tritonal accents as opposed to bitonal accents. Hence, further studies including perception experiments on the anchoring of the LHL tonal gesture in Swedish Accent II are suggested.

References

- Arvaniti, A., Ladd, D.R. & Mennen, I. (1998). Stability of Tonal Alignment: the Case of Greek Prenuclear Accents. *Journal of Phonetics* 26, pp 3-25
- Atterer, M. & Ladd, D.R. (2004). On the Phonetics and Phonology of “Segmental Anchoring” of F₀: Evidence from German. *Journal of Phonetics* 32, pp 177-197
- Barnes, J., Veilleux, N., Brugos, A. & Shattuck-Hufnagel, S. (2010). *The effect of global F₀ contour shape on the perception of tonal timing contrasts in American English intonation*. In *Speech Prosody*, Chicago
- Boersma, P. & Weenink, D. (2010). *Praat: doing phonetics by computer [Computer program]*. Version 5.2.03, retrieved 24 November 2010 from <http://www.praat.org/>
- Bruce, G. (2007). Components of a prosodic typology of Swedish intonation. In Tomas Riad & Carlos Gussenhoven (Eds.), *Tones and Tunes, Volume 1: Typological Studies in Word and Sentence Prosody*. Berlin
- Bruce, G. & Gårding, E. (1978). A prosodic typology of Swedish dialects. In Eva Gårding, Gösta Bruce & Robert Bannert (Eds.) *Nordic prosody. Papers from a symposium*. Lund
- Caspers, J. & Van Heuven, V.J. (1993). Effects of Time Pressure on the Phonetic Realization of the Dutch Accent-lending Pitch Rise and Fall. *Phonetica* 50, pp 161-171
- Crystal, T.H. & House, A.S. (1990). Articulation Rate and the Duration of Syllables and Stress Groups in Connected Speech. *The Journal of the Acoustical Society of America* 88, pp 101-112
- Dilley, L., Ladd, D.R. & Schepman, A. (2005). Alignment of L and H in Bitonal Pitch Accents: Testing Two Hypotheses. *Journal of Phonetics* 33, pp115-119
- Ladd, D.R., Faulkner, D., Faulkner, H. & Schepman, A. (1999). Constant “Segmental Anchoring” of F₀ Movements under Changes in Speech Rate. *The Journal of the Acoustical Society of America* 106 (3), pp 1543-1554
- Niebuhr, O., D’Imperio, M., Fivela, B.G. & Cangemi, F. (2011). *Are there “Shapers” and “Aligners”?* *Individual Differences in Signaling Pitch Accent Category*. In *ICPhS XVII*, Hong Kong
- Niemann, H., Mücke, D., Nam, H., Goldstein, L. & Grice, M. (2011). *Tones as Gestures: the Case of Italian and German*. In *ICPhS XVII*, Hong Kong
- Quené, H. (2008). Multilevel Modeling of Between-speaker and Within-speaker Variation in Spontaneous Speech Tempo. *The Journal of the Acoustical Society of America* 123 (2), pp 1104-1113

Xu, Y. (1998). Consistency of Tone-Syllable Alignment across Different Syllable Structures and Speaking Rates. *Phonetica* 55, pp179-203

Acknowledgements

While working with an earlier version of this thesis I was supervised by associate professor Hugo Quené and the late professor Gösta Bruce. I am very grateful for having had their excellent supervision in the initial steps of the study.

I also want to extend a big thank you to my supervisor Gilbert Ambrazaitis for his invaluable help, understanding and enthusiasm.

Last but not least thank you to all the speakers from Vittsjö for their time and effort. Their welcoming manner and interest in the project made the recordings even nicer.

Appendix

Target sentences

2. Nej, Anna bilar till Polen med många VÄNNER.
22. Nej, Erik har skrivit normen till många LAGAR.
36. Nej, Linn träffar endast norrmän med många BÅTAR.

The original material

1. Nej, hon ska resa runt världen med vännen LASSE.
2. Nej, Anna bilar till Polen med många VÄNNER.
3. Nej, Alf känner endast NORRMÄN med flera båtar.
4. Nej, Sam såg mystiska TECKEN vid sängens gavel.
5. Nej, hon kammar Jennys bena med nya KAMMEN.
6. Nej, gänget ska ses vid kullen om tjugo MINUTER.
7. Nej, han brukar vara BUREN av mamma Eva.
8. Nej, den katten satt på PÅLEN med sina vänner.
9. Nej, Gun hade spelat LJUDEN på högsta volym.
10. Nej, mamma la mina täcken vid sängens GAVEL.
11. Nej, Örjan vill räfsa tomten för hundra KRONOR.
12. Nej, hon hade skrivit NORMEN till många lagar.
13. Nej, Jan betalar för PÅLLEN var tredje månad.
14. Nej, vi har rest runt VÄRLDEN med vännen Kalle.
15. Nej, Bo är framme i KULLEN om tjugo minuter.
16. Nej, jag brukar städa BUREN med gammalt diskmedel.
17. Nej, jag hade frågat JUDEN hur sabbat firas.
18. Nej, hon kammar Gunnels BENA vid varje klippning.
19. Nej, Siv tar piller mot POLLEN var tredje månad.
20. Nej, du är framme i Kullen om tjugo MINUTER.
21. Nej, hon hoppar in för TOMTEN när han är sjuk.
22. Nej, Erik har skrivit normen till många LAGAR.
23. Nej, Elon har spelat ljuden med detta INSTRUMENT.
24. Nej, Britt hade samlat VÄRDEN till denna kalkyl.
25. Nej, den katten satt på pålen med andra KATTER.
26. Nej, det fanns mystiska tecken vid sängens GAVEL.
27. Nej, Örjan ska räfsa TOMTEN för hundra kronor.
28. Nej, han brukar vara buren av mamma EVA.

29. Nej, John betalar för pållen var tredje MÅNAD.
30. Nej, Eskil har frågat juden när sabbat BÖRJAR.
31. Nej, Jens hade samlat värden till denna KALKYL.
32. Nej, Lisa har skrivit B:na med samma PENNA.
33. Nej, gänget ska ses vid KULLEN om tjugo minuter.
34. Nej, pappa la mina TÄCKEN vid sängens gavel.
35. Nej, Jonny bilar till POLEN med många vänner.
36. Nej, Linn träffar endast norrmän med många BÅTAR.
37. Nej, far brukar städa buren med gammalt DISKMEDEL.
38. Nej, hon hoppar in för tomten när han är SJUK.
39. Nej, Siv tar piller mot pollen var tredje MÅNAD.
40. Nej, barnen har skrivit B:NA på tavlan idag.

Praat scripts

Tonal targets

```
# tonal targets

n = Get number of points... 4
for i to n
  tekst$ = Get label of point... 4 i
  if tekst$ <> ""
    t = Get time of point... 4 i
    printline 'tekst$' 't:3'
  endif
endfor
```

Syllables

```
# längd och stavelse 1

n = Get number of intervals... 2
for i to n
  tekst$ = Get label of interval... 2 i
  if tekst$ = "1"
    t1 = Get starting point... 2 i
    t2 = Get end point... 2 i
    printline 'tekst$' 't1:3' 't2:3'
  endif
endfor

# längd och stavelse 2

n = Get number of intervals... 2
for i to n
  tekst$ = Get label of interval... 2 i
  if tekst$ = "2"
    t1 = Get starting point... 2 i
    t2 = Get end point... 2 i
    printline 'tekst$' 't1:3' 't2:3'
  endif
endfor
```

Target words

```
# längd och vilka målord

n = Get number of intervals... 1
for i to n
  tekst$ = Get label of interval... 1 i
  if tekst$ <> ""
    t1 = Get starting point... 1 i
    t2 = Get end point... 1 i
    printline 'tekst$' 't1:3' 't2:3'
  endif
endfor
```