A pilot study: acoustic and articulatory data on tonal alignment in Swedish word accents

Malin Svensson Lundmark, Johan Frid and Susanne Schötz
Centre for Languages and Literature, Lund University

Abstract

This pilot study compares the timing of articulatory gestures to the timing of the tonal contour in South Swedish Accent 1 and Accent 2. Acoustic and articulatory data were collected with an EMA (Carstens AG501). Variables included the tonal alignment of the high tone H and the following low tone L to the vowel onset, the syllable offset, as well as to the lip aperture and the tongue body. Acoustic results point towards different units as host for the accents: Accent 1 aligns with the vowel while Accent 2 aligns with the syllable. The articulatory data shows alignment to different gestures: a stable tonal alignment with the lip aperture in Accent 1, and a less stable alignment with the movement of the tongue body in Accent 2.

Introduction

In the prosodic typology of Swedish intonation provided by the Lund Model (Bruce & Gårding, 1978; Bruce, 2007) the two word accents are assumed to be represented by a tonal fall associated with the stressed syllable. However, the tonal peak of Accent 1 always precedes the peak of Accent 2 in all tonal dialect types of Swedish. Moreover, there are extensive timing differences between dialects, e.g. Accent 1 by a speaker of Stockholm Swedish begins with a low tone, while in South Swedish it starts with a high tone.

There are morphological rules attached to the Swedish word accent distinction (see e.g. Bruce, 1998; Riad, 2014; Riad 2012). For example, a nominal monosyllabic word stem is assigned Accent 1 in the singular form, but receives Accent 2 when plural suffixes are added. Perception studies have found evidence that word accents indeed provide cues of the upcoming suffix (Roll et al., 2013).

However, views on the phonological typology differ: the Lund Model assesses an equipollent distinction where both accents have lexical tones. Other accounts have stressed a privative distinction where only Accent 2 is lexically marked (Riad, 2006; Engstrand, 1997). It has also been suggested that for some dialects the lexical tone in Accent 2 consists of a high tone, while in some dialects (including South Swedish) it is a low tone (Riad, 2006). In the revised Lund Model, Bruce (2007) made the specific assumption for South Swedish of a tonal fall, an H+L pattern, for Accent 1 and a rise, an L+H pattern, for Accent 2. The rise in South Swedish Accent 2 has indeed been shown to be relevant from a perceptual point of view (Ambrazaitis & Bruce, 2006). However, in this pilot study we will only look at the stability of H and L of the fall in both accents, and not L of the rise.

Studies in intonational languages have displayed an unambiguous case of the start of the rise L aligning with the syllable onset in pre-nuclear accents, e.g. Greek (Arvaniti et al., 1998), Italian (Niemann et al., 2011), Dutch (Caspers & Van Heuven, 1993), English (Ladd et al., 1999), and German (Atterer & Ladd, 2004), or just after the syllable onset in the tone language Mandarin (Xu, 1998). A similar consistent result has not been found for the high target H in either of the studies. However, in a study on tonal alignment in the South Swedish Accent 2, the L marking the beginning of the rise appeared to be less stable than H (Svensson Lundmark, 2014).

Recent tonal alignment studies have incorporated articulatory data (Hermes et al., 2008; Mücke et al., 2012; Niemann et al., 2014), following the articulatory phonology framework and the notion of articulatory gestures (Browman & Goldstein, 1992). By including intonation in the gestures, i.e. a tonal gesture, it is possible to couple it with the consonantal and vocalic gestures (Mücke et al., 2012; Niemann et al., 2011). Niemann et al. (2014) found a stable anchoring of H at the vocalic gesture in rising nuclear accents in German, hence presented evidence for stable L as well as H targets.

In this pilot study we adopt this promising account for the study of the Swedish word accents. This enables us to address a number of important questions: If the timing of the tonal curve equals a tonal gesture, are the tonal gestures of Accent 1 and Accent 2 coupled with separate articulatory gestures? Or do the articulatory gestures differ between the accents? Any such difference would open up for the possibility that the two word accents have different roles in speech motor control. Maybe this would shed a light on the issue of whether the word accents should be considered a privative or an equipollent distinction. To the best of our knowledge articulatory studies in the past on the Swedish word accents have been restricted to laryngeal control (e.g. Gårding et al., 1975).

Method

Speech material

The material consisted of simplex disyllabic Accent 1 and Accent 2 target words with stress on the penult produced in the carrier phrase [Det var TARGET jag sa.] (It was TARGET I said.). In the sentence context the target words elicited a focal accent, which in the South Swedish dialect does not differ distinctively from a non-focal accent [8]. The words were matched so that each Accent 1 – Accent 2 pair consisted of the same nominal word stem, but with different suffixes: definite singular for Accent 1 and indefinite plural for Accent 2. The target words also conditioned either an open or a closed stressed syllable (CV:.CVC or CVC.CVC), which in turn contained either a closed or an open vowel (Table 1). Thus, the material consisted of eight target words; each accent pair conditioned syllable type and vowel type.

Data collection

Two female speakers of South Swedish (age 38 and 49) read the material ten times each, i.e. 80 target words per speaker. The sentences were shown on a prompter in a random order.

Sound recordings and kinematic data were recorded simultaneously using a 3D Electromagnetic Articulograph (Carstens AG501) with an external condenser microphone (t.bone EM 9600). Articulatory movements were tracked by sensors on the upper and lower lips (at the vermilion border in the sagittal plane), on the tongue body and also on the bridge of the nose and behind one ear; the latter two sensors were used to correct for head movements during the recordings.

Table 1. The conditions of the eight target words.

		Closed vowel	Open vowel
Accent 1	CVC	/bilden/ (the picture)	/valen/ (the mound)
	CV:	/biːlen/ (the car)	/vɑːlen/ (the whale)
Accent 2	CVC	/bilder/ (pictures)	/valar/ (mounds)
	CV:	/biːlar/ (cars)	/vɑːlar/ (whales)

Measurements

The recorded samples amounted to 160 tokens (2 speakers x 8 target words x 10 repetitions). Acoustic segmentation and annotation of F0 turning points was made manually in Praat (Boersma & Weenink, 2014). For the high tone (H) maximum pitch was labelled in the tonal peak or high plateau. For the following low point (L2) the minimum F0 was used unless an apparent turning point appeared later or earlier.

Speaker MS occasionally used atypical South Swedish accent patterns. These samples were omitted. Some target words were produced with creaky voice, which resulted in some missing data for the L2 point. Out of 160 observations 144 with H-targets and 126 with L2 were used in our further analyses. The data of both speakers were collapsed in the analysis. From the acoustic data the following variables were obtained (Figure 1):

1) the distance from H to the vowel onset
2) the distance from L2 to the vowel onset
3) the distance from H to the syllable offset
4) the distance from L2 to the syllable offset

The two articulatory targets were automatically annotated in R (R Core Team, 2015) (marked as ‘x’ in Figure 1) and the following variables were collected:

5) the distance from H to the max velocity of the lip aperture
6) the distance from L2 to the max velocity of the lip aperture
7) the distance from H to the articulatory target of the vocalic gesture (max tongue body for open vowel, min for closed vowel)
8) the distance from L2 to the articulatory target of the vocalic gesture (max tongue body for open vowel, min for closed vowel)

9) the sync difference (distance between the two articulatory targets)

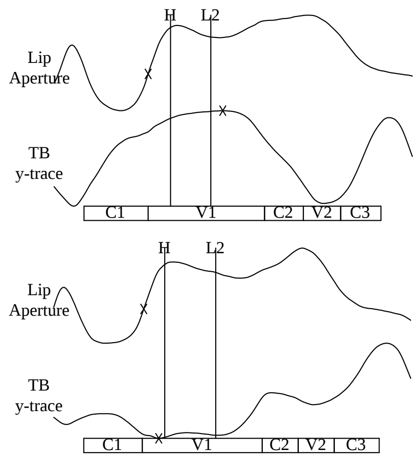

Figure 1. Acoustic and articulatory landmarks for the tonal targets H and L2. Accent 1 words by speaker SS: open syllables (CV:.CVC) with a closed vowel (top) and an open vowel (bottom).

Results

Acoustic results

Duration

The duration was similar for Accent 1 and Accent 2 in the different target words, stressed syllables or vowels (Table 2). We also found similar durations in the syllable types (CVC and CV:), except for the obvious difference in vowel duration. Also, a significant difference was found between the two stressed vowel types closed and open (SE=11.39, t=-2.421, *p*<.05).

Table 2. Duration (ms) of the variables used in the study.

Predictor variables	Target words	Stressed syllable	Stressed vowel
Accent 1	724	433	209
Accent 2	722	426	212
CVC	721	426	143
CV:	725	433	276
Closed vowel	732	442	197
Open vowel	714	416	224

Acoustic tonal alignment

In Accent 1 both tonal targets seem stable in relation to the vowel onset: H is about 50 ms after the vowel onset and L2 about 200 ms (see Figure 2). In Accent 2 H is about 240 ms and L2 about 400 ms after the vowel onset, and also more variable than in Accent 1. A regression analysis shows that the syllable type only affects the alignment of H in Accent 2 (SE=0.007, t=-4.209, *p*<.001). However, the target word /bilder/, which has a stressed CVC syllable with a closed vowel, appears to stand out. When we removed the target word /bilder/ from the data, the syllable type also influenced L2 in Accent 2 (SE=0.031, t=-2.79, *p*<.01). The effect of the syllable type indicates that Accent 2 is not aligned to the vowel.

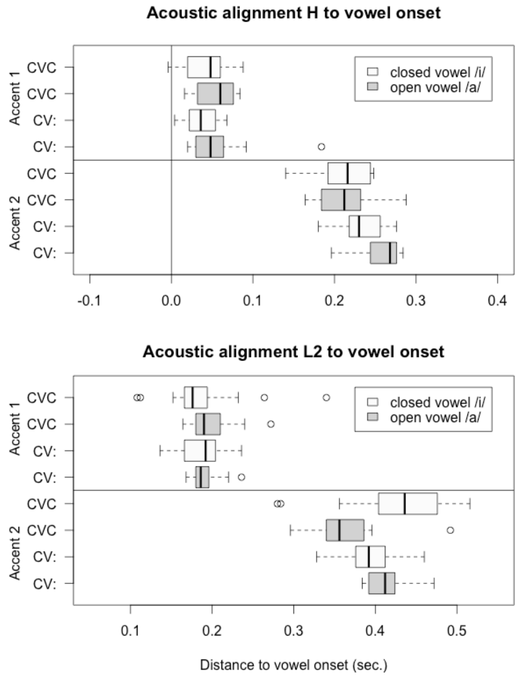

Figure 2. Alignment of H (top) and L2 (bottom) to the vowel onset (vertical line at 0 sec.).

Both targets seem to vary in their alignment to the syllable offset (see Figure 3) in Accent 1, while in Accent 2, H is quite stable about 140 ms before the syllable offset and L2 is aligned almost perfectly to the syllable offset. According to a regression analysis the syllable type affects the timing of H and L2 significantly in both Accent 1 (H: SE=0.009, t=5.576, *p*<.001; L2: SE=0.011, t=5.637, *p*<.001) and in Accent 2 (H: SE=0.006, t=3.936, *p*<.001; L2: SE=0.012, t=5.664, *p*<.001). However, removing the target word /bilder/, which in Figure 3 clearly deviates from the others, the significant effect disappears in Accent 2 for H (SE=0.004, t=-1.475, *p*=0.146) and for L2 (SE=0.029, t=-1.196, *p*=0.237). This indicates a stable alignment with the syllable in Accent 2.

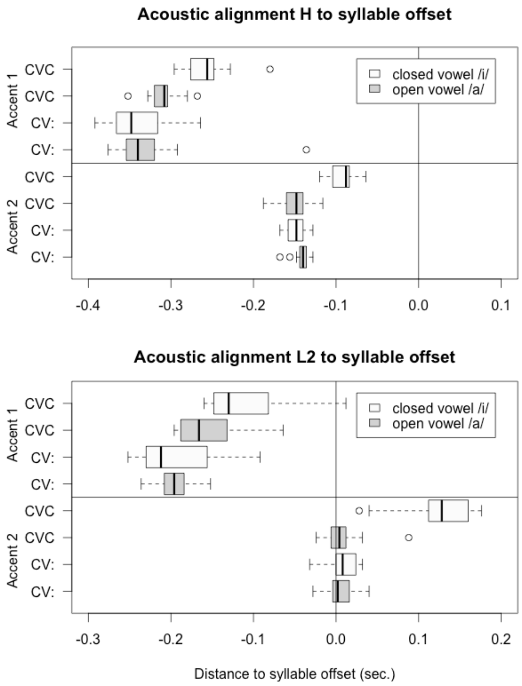

Figure 3. Alignment of H (top) and L2 (bottom) to the syllable offset (vertical line at 0 sec.).

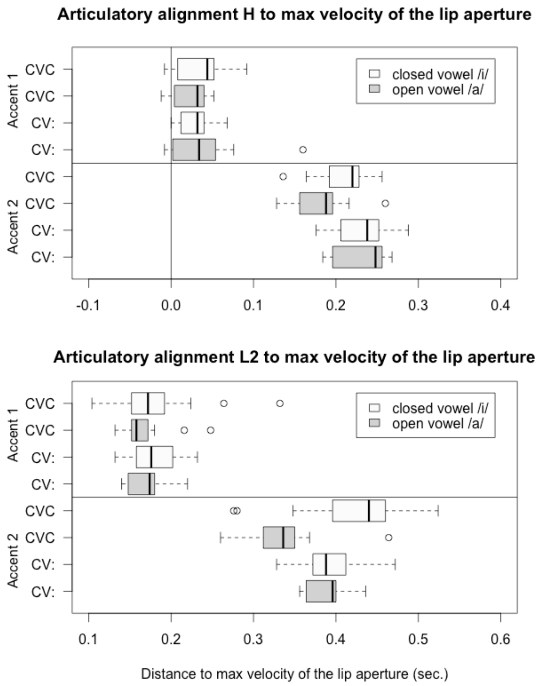

Figure 4. Alignment of H (top) and L2 (bottom) to max velocity of the lip aperture (vertical line at 0 sec.).

Articulatory tonal alignment

The tonal alignment differs significantly between the accents in relation to the max velocity of the lip aperture in H (SE=0.006, t=-33.63, *p*<.001) as well as in L2 (SE=0.008, t=-26.03, *p*<.001). Accent 1 displays a more stable relationship to the lip aperture than Accent 2 (see Figure 4). A significant effect of syllable type is only found in Accent 2 for H (SE=0.008, t=-4.581, *p*<.001). By excluding /bilder/ once again a significant effect in Accent 2 is also found for L2 (SE=0.031, t=-3.188, *p*<.01), indicating an alignment to the lip aperture in Accent 1.

In Figure 5 the target of the vocalic gesture displays less stability in the tonal alignment than in the lip aperture, but a significant difference is still found between the accents in both H (SE=0.011, t=-17.64, *p*<.001) and L2 (SE=0.013, t=-17.80, *p*<.001). The syllable type affects Accent 1 in both H (SE=0.016, t=4.211, *p*<.001) and L2 (SE=0.016, t=4.368, *p*<.001). Excluding /bilder/ results in no significant effect of syllable type on neither H nor L2 in Accent 2, suggesting a tonal alignment to the vocalic gesture (if /bilder/ remains in the data the significant effect is only present in L2, SE=0.018, t=2.293, *p*<.05).

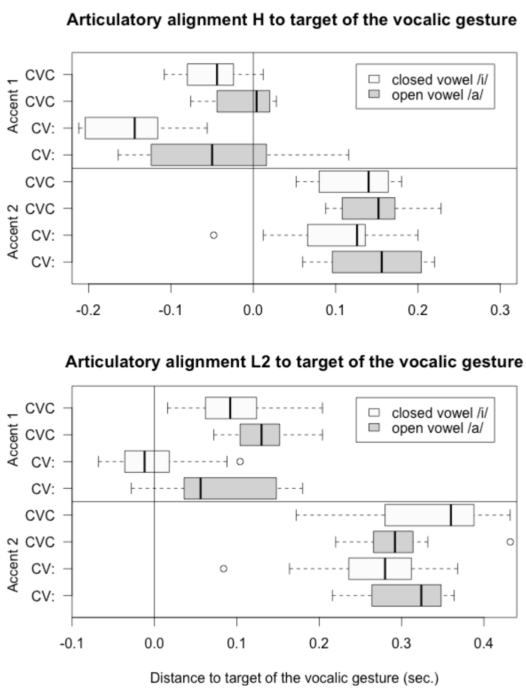

Figure 5. Alignment of H (top) and L2 (bottom) to target of vocalic gesture (vertical line at 0 sec.).

There appears to be no synchronization difference between the two articulatory variables: both articulators seem to differ between the syllable types, but not between accents (see Figure 6). No significant difference is found between Accent 1 and Accent 2 (SE=0.01, t=-1.193 , *p*=0.235).

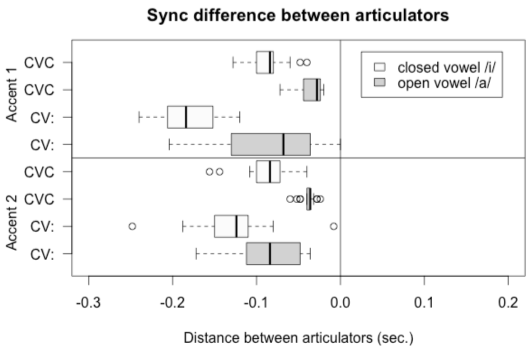

Figure 6. Distance between the articulators. The target of the vocalic gesture is at 0 sec.

Discussion and conclusion

The syllable structure affects the acoustic tonal alignment in both accents but to different units. We interpret this as an alignment of the fall in Accent 1 to the vowel and of Accent 2 to the syllable. This might call for a revision of the Lund Model, which associates both accents with the stressed syllable.

The tonal alignment to the articulatory targets is also affected by syllable type. Moreover, no synchronization difference was found between the accents. These results indicate that the tonal gestures of the accents couple with different articulators: Accent 1 with the consonantal gesture of the lip aperture and Accent 2 to the vocalic gesture of the tongue body. The articulatory, and the acoustic, results indicate that the word accents are different in their phonological nature. It seems plausible that they are separate, which leans more towards the privative than the equipollent distinction.

If the consonantal gesture was to be equivalent to the syllable and the vocalic gesture to the vowel, the articulatory results would contradict the acoustic results, but clearly a more complex relationship is expected, seeing that the relationship between articulatory movements and acoustics is nonlinear. Furthermore, the deviating data on the target word /bilder/ may be explained by coarticulation effects of /i/ and /l/ by speaker SS. It can also be due to the acoustic segmentation, as the deviation is not equally clear in the articulatory alignment.

To be able to further add to the phonological typology and the lexical distinction of the two accents, a follow-up study would benefit from additional segmental structures, as well as measures on the start of the rise. Another aim in future research would be to find a more stable target for the vocalic gesture, but also to include other articulators.

References

Ambrazaitis G & Bruce G (2006). Perception of south Swedish word accents. In: Ambrazaitis G & Schötz S, eds, *Working Papers 52, Proceedings from Fonetik 2006.* Lund, 5-8.

Arvaniti A, Ladd D R & Mennen I (1998). Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics*, 26: 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: 177-197.

Boersma P & Weenink D (2014). Praat: doing phonetics by computer [Computer program], version 5.4.01. http://www.praat.org/

Browman C P & Goldstein L (1992). Articulatory phonology: an overview. *Phonetica*, 49: 155-180.

Bruce G & Gårding E (1978). A prosodic typology of Swedish dialects. In: Gårding E, Bruce G & Bannert R, eds, *Nordic prosody, Papers from a symposium*. Lund, 219-228.

Bruce G (1998). *Allmän och svensk prosodi*. *Praktisk Lingvistik 16.* Lund: Dept. of linguistics and phonetics.

Bruce G (2007). Components of a prosodic typology of Swedish intonation. In: Riad T & Gussenhoven C, eds, *Tones and Tunes, Volume 1: Typological Studies in Word and Sentence Prosody*. Berlin, 113-146.

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: 161-171.

Engstrand O (1997). Phonetic interpretation of the word accent contrast in Swedish: evidence from spontaneous speech. *Phonetica*, 54: 61-75.

Gårding E, Fujimura O, Hirose H & Simada Z (1975). Laryngeal control of Swedish word accents. *Working papers 10*. Lund, 53-82.

Hermes A, Becker J, Mücke D, Baumann S & Grice M (2008). Articulatory gestures and focus marking in German. *Proc. Speech Prosody 2008*, Campinas, 457-460.

Ladd D R, Faulkner D, Faulkner H & Schepman A (1999). Constant ‘segmental anchoring’ of F0 movements under changes in speech rate. *J. Acoust. Soc. Am.*, 106(3): 1543-1554.

Mücke D, Nam H, Hermes A & Goldstein L (2012). Coupling of tone and constriction gestures in pitch accents. In: Hoole P, Bombien L, Pouplier M, Mooshammer C & Kühnert B, eds, *Consonant Clusters and Structural Complexity*. Munich: Mouton de Gruyter, 157-176.

Niemann H, Grice M & Mücke D (2014). Segmental and positional effects in tonal alignment: An articulatory approach. *Proc. 10th ISSP*, Cologne, 285-288.

Niemann H, Mücke D, Nam H, Goldstein L & Grice M (2011). Tones as gestures: the case of Italian and German. *Proc. ICPhS XVII*, Hong Kong, 1486-1489.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Riad T (2014). *The phonology of Swedish.* Oxford: Oxford University Press.

Riad T (2012). Culminativity, stress and tone accent in Central Swedish. *Lingua*, 122: 1352-1379.

Riad T (2006). Scandinavian accent typology. In: Viberg Å, ed, *Special issue on Swedish. Sprachtypologie und Universalienforschung (STUF) 59(1).* Berlin, 36-55.

Roll M, Söderström P & Horne M (2013). Word-stem tones cue suffixes in the brain. *Brain Research*, 1520: 116-120.

Svensson Lundmark M (2014). Constant tonal alignment in Swedish word accent II. *Proc. Speech Prosody 7*, Dublin, 987-991.

Xu Y (1998). Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica,* 55: 179-203.