

JAW MOVEMENTS IN TWO TONAL CONTEXTS

Malin Svensson Lundmark¹, Johan Frid²

¹Centre for Languages and Literature, Lund University, Sweden; ²Humanities Lab, Lund University, Sweden
malin.svensson_lundmark@ling.lu.se; johan.frid@humlab.lu.se

ABSTRACT

Intra-syllabic articulatory movements have previously been shown to vary depending on tonal context. In this study we concentrate on the jaw and explore its movements in two tonal contexts, i.e. the falling and the rising tone of the Swedish word accents. EMA was used to track the mandible of 19 speakers while they read sentences with target words carrying similar word onsets. Evaluation of the acceleration (second derivative) of the jaw movements revealed three well-defined intervals: jaw opening, jaw open posture and jaw closing. Mixed effects models showed that the jaw opening and the jaw open posture were longer in the rising tone context. Concurrently, spatial data revealed lower jaw trajectories in the falling tone context. Our results resemble that of previous findings on the tongue body. In conclusion, the two tonal contexts induce different jaw movements, presumably because of the involved physical mechanism of two distinct tonal targets.

Keywords: jaw, acceleration, EMA, tone

1. INTRODUCTION

This study explores the movements of the jaw in the two tonal categories of the Swedish word accents (henceforth SWA). SWA are morpho-phonological and sometimes referred to as lexical pitch accents. They are used to differentiate monosyllabic and polysyllabic words, while in disyllabic words they are associated with different suffixes. Phonologically a falling or a rising tone, respectively, is associated with the stressed syllable [1]. This results in an early f₀ peak in Accent 1 (A1), and a late f₀ peak in Accent 2 (A2) (Fig. 1, from [2]).

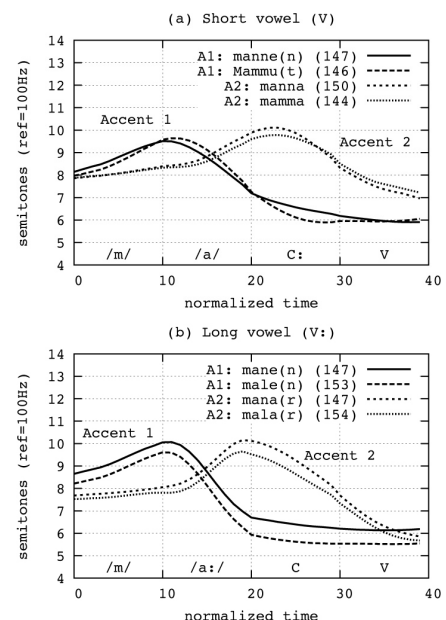
What we present here is complementary to previous articulatory findings on the lips and the tongue body (henceforth TB) during the word-initial CV sequence /ma/ [2] in A1 and A2. In [2] a slightly longer lip closure was found in A1 than in A2, while at the same time a longer lowering yet with a higher positioned trajectory of TB in A2 than in A1. In this study we are using the same material as in [2] to perform measurements on the jaw movements.

Previous studies on the coordination of articulatory movements with f₀ have mainly focused

on Mandarin lexical tones, finding evidence of lip and TB coordination differences between the tones [3], as well as TB height differences [4-6]. In addition, [5-6] found an effect of tone height on the jaw movement: a lower jaw position in Mandarin tone 3 (a low tone). Their results correspond well with the findings by [7] that low tones encompass lower larynx movements and downward movements of the jaw, while higher tones display the jaw moving backward as the tone falls.

These previous findings suggest that the jaw, the tongue, and the lips are highly coordinated with f₀ during speech, although the details of the effect of the tonal context are not fully understood. Therefore, in order to get an overall picture an examination of the jaw movements in different tonal contexts is required.

Figure 1: Mean f₀ contours of A1 and A2 target words by South Swedish speakers (from [2]): short vowels (a); long vowels (b). Normalized time scale. The vertical lines are segment boundaries.



1.1. Research questions

The goal was to investigate the spatiotemporal role of the jaw during the word-initial CV sequence /ma/ comparing the two tonal categories of the SWA. The research is exploratory and guided by the following questions:

1. Does jaw movement differ between A1 and A2 word-initial CV sequences?

2. Will the jaw follow the same pattern as TB, i.e. have longer intervals in A2?

Because the speech material includes /ma/ sequences with both long and short vowels, and different word-endings, we furthermore suspect the jaw to be affected by these circumstances, e.g. that the closing of the jaw is affected by the post-vocalic consonant, as has previously been found in [8].

2. METHOD

Articulatory data was collected from 19 South Swedish speakers (12 female, \bar{x} =40 yrs, sd =12.3 yrs) using a Carstens AG501. Each speaker read leading questions + target sentences from a prompter (presented eight times in random order), an arrangement employed to put a contrastive focus onto the last element in the target sentence. This left the target word in a low-prominence inducing context, hence controlling for possible effects of sentence intonation. In Swedish high prominence is associated with an additional f_0 peak or a higher f_0 [1, 9].

The original data set consisted of 18 target words, divided into nine word accent pairs with identical stressed syllables (e.g. *bilen-bilar* or *manen-manar*). For this study we only used the four word accent pairs that shared the similar word-initial CV sequence /ma/. However, the target words differed in stressed vowel length and post-vocalic segments (Table 1). The target words were embedded in individual but similarly structured target sentences. The VCV sequence preceding the target words was identical (/ade/). 1191 tokens were recorded.

Table 1: The target words according to SWA and vowel length; long vowels denote open syllables.

	Accent 1	Accent 2
Short vowel	man.nen <i>the man</i> mam.mut <i>mammoth</i>	man.na <i>semolina</i> mam.ma <i>mom</i>
Long vowel	ma:.nen <i>the mane</i> ma:.len <i>the catfish</i>	ma:.nar <i>urge</i> ma:.lar <i>catfish (pl.)</i>

2.1. Procedure

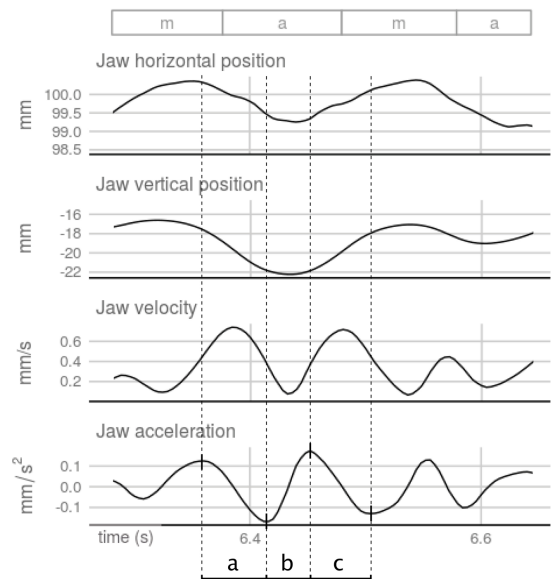
Articulatory data was recorded at 250 Hz, and audio was recorded simultaneously at 48 kHz. For this study we used data from a sensor on the mandible (just below the incisors). In order to correct for head movements, three additional sensors were employed; one behind each ear, and one on the nose ridge. The sagittal angle was not controlled for during the recordings but is not expected to have an effect since we apply the combined x and y dimension. After post-processing including head correction in Carstens software, the articulatory data was transferred to R [11] where it was smoothed using locally weighted regression by the R function *loess* (span=0.1). The

first author segmented the acoustic data manually in Praat [12] using ProsodyPro [13]. The TextGrids were further used in R to automatically extract articulatory data from selected time frames of each target word. The time frames were manually adjusted for each speaker using visual cues of the y-trace jaw movements and the tangential velocity profile.

2.2. Measurements

We used the second derivative of the xy trace movement (first derivative of tangential velocity) to locate the moment of maximal acceleration and deceleration (Fig. 2). Maximal de-/acceleration as onset and offset of an articulatory movement is preferable, since it reflects the damped mass-spring systems of articulatory dynamics ([14] for an account on a Dynamical Systems Theory for speech). In other words, with smoothed acceleration data it is feasible to track speed changes of the articulators, e.g. as they slow down before a complete closure.

Figure 2: Horizontal and vertical jaw trajectories, tangential velocity, and acceleration during an item (*mamma*, A2, segments in boxes). Intervals based on maximal acceleration: jaw opening (a), jaw open posture (b), and jaw closing (c).



After a visual inspection of the jaw acceleration during /ma/, it became evident that there were three well-defined intervals: the jaw opening, the jaw open posture, and the jaw closing (Fig. 2). For this reason, we collected the time points of the maximal de-/acceleration and used them to calculate the three intervals, and also combined them to obtain the total jaw duration. We predicted the temporal measurements to reflect the downward-backward movement, but nonetheless collected data on the vertical position and made a qualitative analysis of the normalized y-trace.

2.3. Analysis

The calculations based on the temporal measurements (the three intervals and the total duration) were statistically tested to see whether they differed between the two tonal categories (A1 and A2). Generalized linear mixed effects models (GLMM) were used to account for sensor placement variation and speaker variability (*speaker* as random effect: random intercept and slope). We added *word* as random effect (random intercept) to avoid certain factors (e.g. word frequency, word endings, or target sentence information structure) enhancing the possibility of a Type I error (=a lower p-value). *Word accent* and *vowel length* were set as fixed effects. We performed a likelihood ratios test for each interval to test whether the additional complexity of vowel length was warranted. The models were run in R using the lme4-package [15]. P-values were obtained by using the lmerTest-package in R [16].

For the qualitative analysis of spatial movements, the words were time-normalized by z-transforming the time points. The jaw sensor vertical positions were then normalized for each word or each speaker by z-transforming positions per word/speaker on the basis of all positions in the utterance it occurred in.

3. RESULTS

3.1 Temporal results

After a model comparison, word accent was fit as the only fixed effect on the jaw opening interval data. The model showed a significant difference between the word accents ($t=2.41$, $p<.05$) (Table 2). Fig. 3A visualizes the small but significant effect of tonal context on the jaw opening. Although there was some variation, A1 had shorter jaw opening intervals than A2 in all four word accent pairs.

The measurements on the jaw open posture interval warranted adding vowel length as a fixed effect to the model. The more complex model revealed a significant effect of word accent ($t=3.17$, $p<.05$), vowel length ($t=-8.66$, $p<.001$) as well as the interaction between them ($t=-2.76$, $p<.05$) on the jaw open posture (Table 2), as illustrated in Fig. 3B. The target words with long vowels (the four bars to the right) have longer and more varied intervals, and in addition seem affected by the word accent as opposed to the short vowels (left bars) (Fig. 3B).

The results on the jaw closing intervals revealed no effect by the fixed effects. Only word accent was warranted, which showed no effect on the jaw closing interval ($t=0.26$, $p=.8$). Error bars in Fig. 3C display some difference between vowel lengths.

Total jaw duration (all three intervals) showed an effect by word accent ($t=3.17$, $p<.05$): the overall jaw

movements in A2 are longer (Fig. 3D). The model warranted adding vowel length as well. Table 2 shows the full effects of the model including the interaction.

Fig. 3. Mean intervals in milliseconds (y-axis). A-C: with error bars, all target words divided into SWA pairs with short vowels (left) and long vowels (right). D: all words divided into SWA

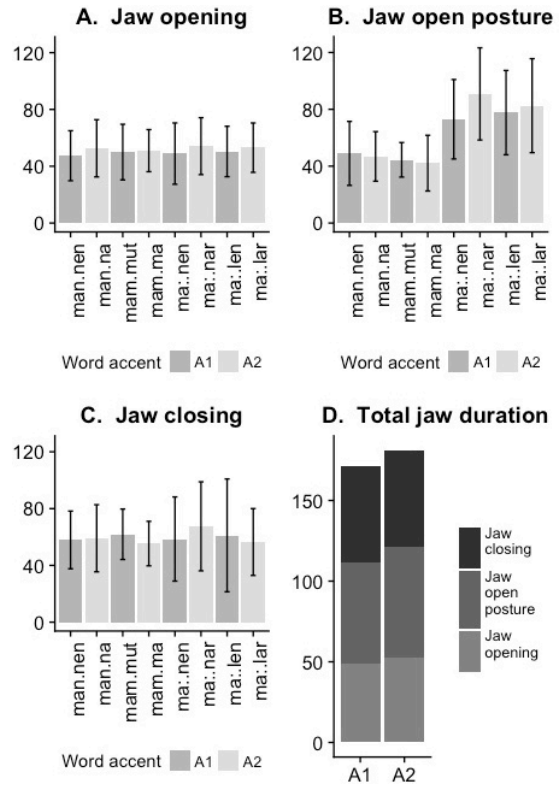


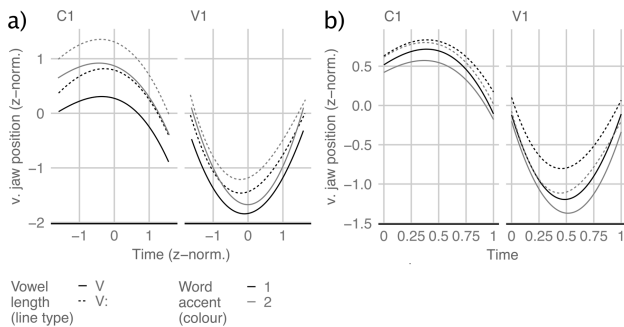
Table 2: GLMM results. Significant effects in bold. Abbr.: intercept (I/C); fixed effects word accent (A1), vowel length (V:), and interaction (A1*V:).

Dep. variable		Est.	SE	df	t	p
Jaw opening	I/C	49.37	1.8	19.28	27.38	.000
	A1	3.03	1.26	19.80	2.41	.026
Jaw open posture	I/C	75.36	3.47	19.69	21.70	.000
	A1	11.25	3.55	9.17	3.17	.011
	V:	-29.19	3.37	8.04	-8.66	.000
	A1*V:	-13.07	4.73	7.80	-2.76	.025
Jaw closing	I/C	59.21	4.37	23.18	13.56	.000
	A1	0.90	3.46	12.93	0.26	.800
Total jaw duration	I/C	184.39	7.00	23.12	26.47	.000
	A1	20.60	6.49	8.40	3.17	.012
	V:	-33.99	6.28	7.43	-5.41	.001
	A1*V:	-24.32	8.89	7.45	-2.74	.027

3.2 Spatial results

Regression lines on the vertical jaw data were used to visualize the differences between the word accents (Fig. 4). When normalized by word we find a lower jaw trajectory in A1 than in A2. Comparing this to a normalized vertical jaw data by speaker instead reveal a slightly higher jaw in A1. Noticeably, long vowels showed higher jaw trajectories in both word accents.

Figure 4: Regression lines of vertical jaw positions during /ma/ (C1, V1). Normalized by word (a) and speaker (b). Effects: word accent and vowel length.



4. DISCUSSION

The temporal results revealed that the jaw opening and the jaw open posture were shorter in the falling tone (A1) than in the rising tone context (A2). The jaw movements in A1 seemed truncated, as if adjusting for the presence of an early tonal target. On this note, SWA have similar tonal contours, but distinct tonal timing (see Fig. 1). The tonal patterns result in a jaw opening either in a high tonal context (A1) or a low tonal context (A2). Thus, as a shorter jaw opening interval suggests an earlier jaw open posture onset, this might be prompted by the early high tone in A1. Similarly, the open jaw posture was prolonged in A2, as if to cover the rise to the high tone before the jaw closes completely. In other words, the tonal contexts of SWA seem to cooccur with systematic and distinctively timed jaw movements. We suggest that the jaw, and the coordinated articulatory movements (based on findings in [2]), mechanically adjust for the distinct tonal targets of A1 and A2.

Our temporal analysis based on GLMM takes into account the variation between speakers and target words. It also includes dynamical movements, since we use acceleration time points of the downward-backward movements. The results correspond well with earlier findings of shorter TB movement in A1 [2]. In addition, the prolonged jaw movements (total jaw duration) in A2 is consistent with previous acoustic studies on SWA segments [9, 17].

Inspection of the vertical data revealed a lower jaw trajectory for the falling tone context (A1) during /ma/. This could suggest that the shorter intervals in A1 is due to an already low jaw starting position, hence a more open mouth as a possible effect of the high tone context. However, when our data is normalized by speaker the patterns are shifted instead towards a slightly lower jaw for the rising tone (A2). Hence, our two means of normalization either leave out substantial information on the individual target words or speaker variability. Although normalization

is fruitful for prominence or lexical frequency effects on segment duration, it leaves out information about the dynamical systems of position, velocity and acceleration, all of which contribute to the shortening or the prolongation of articulatory movements.

The closing of the jaw seems affected by manner and place of articulation for the second consonant, as has partly been reported elsewhere [8], and possibly also by the second vowel, which differs between the target words in our material. However, our data on this is hard to read because of the presence of multiple factors and we leave this for a subsequent study.

The results show that the biggest differences in duration are between phonemically short and long vowels; A1 vs. A2 effects are very small by comparison. However, the open posture intervals of the long vowels appear to differ between A1 and A2, and also display more variance than for the short vowels, which gives some insight into the variability of different vowel lengths. As the longest jaw open posture was found in *manar* (A2) with the highest mean f_0 (Fig. 1), prominence (as in high f_0) is probably the reason for the longer intervals of this particular target word. However, *manen* (A1), which also carries a high mean f_0 , has a shorter open posture interval. It seems that prominence has a positive effect on this interval only in late tonal peaks (rising tone), while an earlier peak (falling tone) leads to a negative effect of prominence on the jaw open posture interval. This observation supports our suggestion of a connection between the jaw and f_0 , though not necessarily reflected in absolute jaw height but rather, as we propose, in timed jaw movements mechanically adjusted to the tonal contexts of the SWA.

5. LIMITATIONS AND FUTURE STUDIES

Analyzing spatial data is intricate. Jaw height has previously been measured at lowest jaw position [5] and at midpoint of the vowel [6], but because of the multiple tonal targets such measurements are not informative for the SWA. Our results suggest that spatial measures need to represent several tonal targets and in addition control for speaker variation.

The data analysis was restricted to the mandible. However, since head posture has been shown to vary depending on tone [6] it should be accounted for in the future. A feasible way is to include data from sensors on the nose ridge or behind the ear.

6. ACKNOWLEDGMENTS

The authors gratefully acknowledge Lund University Humanities Lab. Many thanks to all the participants who graciously worked the EMA.

7. REFERENCES

- [1] Bruce, G. 2005. Intonational Prominence in Varieties of Swedish Revisited. In: Jun, S. (ed), *The Phonology of Intonation and Phrasing*, Oxford: Oxford University Press, 410–429.
- [2] Svensson Lundmark, M., Frid, J., Ambrazaitis, G., Schötz, S. Consonant-vowel coordination in a lexical pitch-accent language. Submitted.
- [3] Gao, M. 2008. *Tonal alignment in Mandarin Chinese: An articulatory phonology account*. (Unpublished doctoral dissertation). Yale University, New Haven.
- [4] Shaw, J., Chen, W., Proctor, M., Derrick, D. 2016. Influences of tone on vowel articulation in Mandarin Chinese. *J. Speech Lang. Hear. Res.* 59(6), 1566–1574.
- [5] Erickson, D., Iwata, R., Endo, M., Fujino, A. 2004. Effect of tone height on jaw and tongue articulation in Mandarin Chinese. *Proc. Tonal Aspects of Languages Beijing*.
- [6] Hoole, P., Hu, F. 2004. Tone-vowel interaction in standard Chinese. *Proc. Tonal Aspects of Languages Beijing*, 89–92.
- [7] Honda, K., Hirai, H., Masaki, S., Shimada, Y. 1999. Role of vertical larynx movement and cervical lordosis in f0 control. *Lang. Speech* 42(4), 401–411.
- [8] Mooshammer, C., Hoole, P., Geumann, A. 2007. Jaw and order. *Lang. Speech* 50(2), 145–176.
- [9] Svensson Lundmark, M., Ambrazaitis, G., Ewald, O. 2017. Exploring multidimensionality: Acoustic and articulatory correlates of Swedish word accents. *Proc. Interspeech 2017 Stockholm*, 3236–3240.
- [10] Löfqvist, A., Gracco, V. 1999. Interarticulator programming in VCV sequences: Lip and tongue movements. *J. Acoust. Soc. Am.* 105, 1864–1876.
- [11] R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org/>.
- [12] Boersma, P., Weenink, D. 2018. Praat: doing phonetics by computer. [Computer program] Version 6.0.37 <http://www.praat.org/>
- [13] Xu, Y. 2013. ProsodyPro — a tool for large-scale systematic prosody analysis. in *Proc. Tools and Resources for the Analysis of Speech Prosody 2013 Aix-en-Provence*, 7–10.
- [14] Iskarous, K. 2017. The relation between the continuous and the discrete: A note on the first principles of speech dynamics. *J. Phon.* 64, 8–20.
- [15] Bates, D., Maechler, M., Bolker, B., Walker, S. 2015. Fitting linear mixed-effects models using lme4. *J. Stat. Software* 67(1), 1–48.
- [16] Kuznetsova, A., Brockhoff, P., Christensen, R. 2017. lmerTest Package: tests in linear mixed effects models. *J. Stat. Software* 82(13), 1–26.
- [17] Elert, C. 1964. *Phonologic Studies of Quantity in Swedish*. Stockholm: Almqvist & Wicksell.