Exploring multidimensionality

Acoustic and articulatory correlates of Swedish word accents

Svensson Lundmark, Malin; Ambrazaitis, Gilbert; Ewald, Otto

Published in:
Interspeech 2017

DOI:
10.21437/Interspeech.2017-1502

2017

Document Version:
Publisher's PDF, also known as Version of record

INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

Exploring multidimensionality: Acoustic and articulatory correlates of Swedish word accents

Malin Svensson Lundmark, Gilbert Ambrazaitis, Otto Ewald

Centre for Languages and Literature, Lund University, Sweden
malin.svensson_lundmark@ling.lu.se, gilbert.ambrazaitis@ling.lu.se

Abstract

This study investigates acoustic and articulatory correlates of South Swedish word accents (Accent 1 vs. 2) – a tonal distinction traditionally associated with F0 timing. The study is motivated by previous findings on (i) the acoustic complexity of tonal prosody and (ii) tonal-articulatory interplay in other languages.

Acoustic and articulatory (EMA) data from two controlled experiments are reported (14 speakers in total; pilot EMA recordings with 2 speakers). Apart from the well-established F0 timing pattern, results of Experiment 1 reveal a longer duration of a post-stress consonant in Accent 2 than in Accent 1, a higher degree of creaky voice in Accent 1, as well as a deviant (two-peak) pitch pattern in Accent 2 for one of eight discourse conditions used in the experiment. Experiment 2 reveals an effect of word accent on vowel articulation, as the tongue body gesture target is reached earlier in Accent 2. It also suggests slight but (marginally) significant word-accent effects on word-initial gestural coordination, taking slightly different forms in the two speakers, as well as corresponding differences in word-initial formant patterns. Results are discussed concerning their potential perceptual relevance, as well as with reference to the c-center effect discussed within Articulatory Phonology.

Index Terms: speech production, pitch, lexical tone, voice quality, articulatory gestures, articulatory phonology, EMA

1. Introduction

This paper studies acoustic and articulatory manifestations of South Swedish word accents: a binary, phonological tonal distinction (Accent 1, Accent 2, henceforth A1 and A2). It thereby adds to a growing body of evidence arguing for a multidimensional nature of tonal prosody (cf. 1.1). To this end, we explore further possible phonetic cues beyond the timing of a (rise-)fall in fundamental frequency (F0), which is traditionally regarded as the distinction's primary phonetic correlate (cf. 1.1). In particular, we investigate patterns of F0, voice quality, durations, articulatory gestures, and formants.

1.1. Acoustic complexity of Swedish word accents

The Swedish word accent distinction has been recognized as a tonal, i.e. F0-related phenomenon since at least Meyer's early comprehensive dialect study [1,2], and F0 has been established as the primary cue for perception [3,4,5]. The F0-dinstinction has been analyzed in terms of timing of a (rising-) falling F0 gesture [6], a proposal which has been controversial when it comes to the Stockholm variety [7,8], but rather undisputed for the South Swedish variety studied in this paper [9,10] (cf. Fig. 2 for a display of typical F0-patterns).

Further (potential) phonetic correlates of the word accents beyond F0 have hardly attracted any attention, although duration and intensity have been revealed as secondary phonetic correlates already in [3]. In particular, for Stockholm Swedish, a longer duration has been observed for the stressed vowel in A1 than in A2, while the reversed pattern was attested for a post-stress consonant, i.e. a longer duration for A2 [11]. Assuming that tonal complexity would trigger a longer duration, we would expect a longer vowel in A2 in Elert's [11] data, because the Stockholm A2 surfaces as a two-peak F0 pattern. So what can, alternatively, explain Elert's findings? We suggest that it is not the tonal complexity *per se*, but the function of the tones involved that matters: it is the *focal accent* tone that causes lengthening. Following [6], this tone is realized in the stressed syllable in A1, but in the post-stress in A2, explaining the differential lengthening of stressed vowel (A1) and following consonant (A2) in Elert's [11] data.

A multidimensional view of tonal word accent encoding is well in line with an increasing body of evidence attesting acoustic complexity of tonal prosodic events in several other languages, often in connection with sentence-level intonation [12, 13, 14, 15].

1.2. The interplay of tonal and articulatory gestures

The framework of Articulatory Phonology [16, 17] has in recent years started to include prosodic information [18, 19, 20, 21, 22, 23]. More specifically tones have been proposed to represent articulatory gestures, i.e. tone gestures, comparable to consonantal and vocalic gestures [19, 21], and tonal alignment has been shown to be more stable relative to articulatory landmarks [21, 24, 25, 26] than to acoustic ones [27, 28, 29, 30, 31]. Lexical tone gestures in Mandarin have even been proposed to compete with consonantal gestures in onset [19]. This type of competitive coordination between articulatory gestures is the source for a phenomenon known as the *c-center effect,* where the start of consonantal gestures shifts depending on an inter-competitive relationship with the vowel [32]. Thus, if vocalic, consonantal and tone gestures were coordinated in onset we would find cues of this coordination pattern in the consonantal and vocalic gestures.

1.3. Hypotheses

Based on the background presented in the previous sections we hypothesize, in general, that South Swedish word accents exhibit further acoustic or articulatory correlated beyond F0 timing. For instance, given the different tonal timing patterns of A1 and A2 we assume that the coupling of tone gestures with consonantal and vocalic gestures would differ between the word accents. We expect to find cues of this coupling in the coordination pattern of the consonantal and vocalic gestures, e.g. a difference in c-center effect in onset.

http://dx.doi.org/10.21437/Interspeech.2017-1502

Another prediction based on our interpretation of [11] (cf. 1.1) is that, for South Swedish, *no* durational effect should occur for the vowel, as the focal accent, in this dialect, is produced through the word accent gesture, and within the stressed vowel for both A1 and A2; we might, however, still predict a longer post-stress consonant in A2, as the A2 pitch rise-fall crosses this consonant.

2. Method

Two experiments are presented, designed independently of each other, both investigating South Swedish word accent production. Apart from the same target word pair being used in both data collections – *bilen* ('the car') for A1, and *bilar* ('cars') for A2 – they differ – and complement each other – in several respects. Experiment 1 was originally designed to investigate word accent production as a function of discourse context (cf. 2.1). It involves 12 speakers (6 female, 6 male) and provides a relatively large amount of data (576 recorded tokens), however, limited to acoustic recordings. Experiment 2 is a pilot study involving 2 female speakers (39 recorded tokens), combining acoustic with articulatory recordings using Electromagnetic Articulography (EMA). The recordings from both experiments were acoustically segmented into consonants and vowels. Additionally, the occlusion phase of /b/ was segmented.

For experiment 1, the target words (A1, A2) were embedded in short carrier phrases such as 'yes, by car', and eight different conditions (discourse contexts) were created in order to elicit different readings of the test phrase, e.g. as an assertion, a confirmation, a correction, an exclamation etc. The context was presented in written form, and for some conditions, there was an additional auditory context question. Participant's read the contexts (quiet) and the target phrase (aloud) from a computer screen in an experimental studio.

In order to avoid unnecessary F0 analysis errors, F0 calculation was performed in the time-domain based on 'pulses' automatically determined by Praat [33] which we manually corrected using ProsodyPro [34].

In experiment 2 kinematic data was recorded in an Electromagnetic Articulograph (EMA, Carstens AG501, sampling rate 250 Hz) at the Lund University Humanities Lab. Sound was recorded simultaneously using an external condenser microphone (t.bone EM 9600). The target words were produced in the carrier phrase *Det var TARGET jag sa* ('It was TARGET I said'). Two female speakers of South Swedish (age 38 and 49) read the material ten times each. The sentences were shown on a prompter in a random order.

The consonantal gesture of the stressed syllable /bi:/ is a bilabial closure, and the vocalic gesture of /bi:/ is a palatal narrow. Hence, EMA data from sensors on the upper and the lower lips (=lip aperture), and on the tongue body, were collected and further processed in R [35] and normalized for head movements.

3. Results and discussion

3.1. Fundamental frequency (F0) (Experiment 1)

As for F0, we restrict this paper to a visual analysis of the general shape or patterning of F0, since correlates beyond F0 shall be in focus in the first place. Figure 1 displays mean F0 curves for A2, averaged across our 6 female speakers, separately for all eight discourse contexts involved, as an example. Intonational expressions due to sentence- or discourse-level functions are outside the scope of this paper, but we include this display for two reasons: First, it demonstrates the relative stability of tonal timing, despite discourse-induced variation in parameters such as peak shape, height, or range. It thus replicates and confirms what is known about the South Swedish word accents' F0 patterning (cf. 1.1).

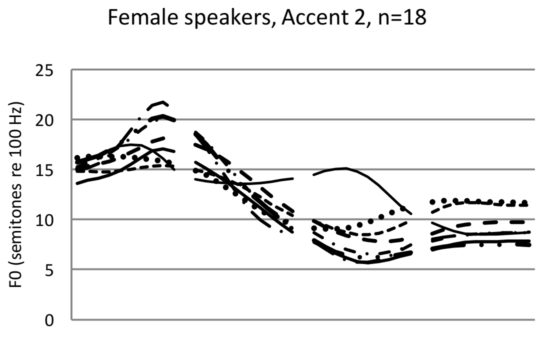

Figure 1: *Mean F0 curves across 6 speakers and 3 repetitions (n=18) for the target word (b)ilar 'cars', initial /b/ not included; 8 conditions (discourse contexts, represented by the separate lines); time is normalized, breaks in the curves indicate acoustic segment boundaries: /i/, /l/, /a/, /r/.*

Second, however, the data also provide an unexpected result, as the F0 pattern for one of the conditions (an 'exclamation') is crucially deviating; it indeed reminds of a Stockholm Swedish pattern, where A2 surfaces as a two-peak F0 curve. What is not evident from this mean curve is that the two-peak pattern was produced exclusively in this condition, in 13 of the 18 tokens, and at least once by each of the 6 speakers. We conclude that it represents a regular pattern of this dialect, occurring on certain discourse conditions, but will leave further discussion of this pattern for future research.

Figure 2 offers an average display comparing A1 and A2, across all female speakers and seven of the conditions (condition 'exclamation' excluded). We obtain equivalent results for male speakers.

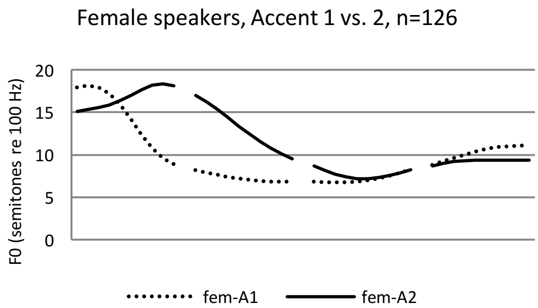

Figure 2: *Mean F0 curves across 6 speakers, 7 conditions (discourse contexts), and 3 repetitions (n=126) for the target words (b)ilen 'the car' (A1) and (b)ilar 'cars' (A2); for further explanations cf. Fig. 1.*

3.2. Voice quality (Experiment 1)

Informal observations during the annotation process revealed a high degree of creaky voice, mostly during the second (final) syllable of the target word (/lɛn/ or /lar/, respectively), but also, sometimes during the preceding (stressed) vowel /iː/. To study possible effects of the word accent on the occurrence of creaky voice, we annotated creaky voice (a) during the vowel /iː/ and (b) during the following consonant /l/, following a simple scheme deciding between absence/presence of creaky voice. Results are displayed in Table 1.

Table 1: *Annotations of creaky voice in %, broken down by word accent (A1, A2) and speaker gender.*

Condition	/iː/	/l/
A1 (all speakers)	62	79
A2 (all speakers)	1	54
A1 (female/ male)	74/50	78/79
A2 (female/ male)	1/1	63/44
Female/Male (both A)	38/25	71/62

Table 1 suggests a strong effect or word accent on the occurrence of creaky voice, despite a certain effect of gender, which is seen both in the stressed vowel and in the following voiced consonant. A linear mixed model for /iː/ with word accent and gender as fixed factors (speaker and context has random effects) reveals significant effects for word accent (t= -18.629, df=555, p<.001***), gender (t=-3.874, df=15.7, p=.0014**), and their interaction (t=4.180, df=555, p<.001***). For /l/, there is an effect of word accent (t=-3.290, df=555, p=.0011**), of the interaction of word accent and gender (t=-2.961, df=555, p=.0032**), but no main effect of gender.

A comparison of these findings with the F0 patterns obtained in 3.1 suggests that the occurrence of creaky voice relates to low or sharply falling F0, thus explaining its higher frequency of occurrence in A1, and in particular its occurrence already during the vowel, where it hardly ever occurs for A2.

3.3. Durations (Experiment 1 and 2)

Segmental durations of the stressed syllable and the following consonant (/b/, /iː/, /l/) were measured in the acoustic domain for data from both experiments. The results from Experiment 1 show no effect of word accent on the stressed syllable (neither on /b/ nor on /iː/), but a significant effect in the duration of the post-stress consonant /l/, which was on average 10 *ms* longer in A2 (89 *ms*) than in A1 (79 *ms*) (linear mixed model with word accent as fixed factor; speaker and context has random effects: t=8.90, df=556, p<.001***). This result is perfectly in line with our prediction formulated in 1.3. The results of Experiment 2, however, are less conclusive. They are inconsistent between the two speakers, and partly contradictory to our predictions. We explain these inconsistencies by the relatively small amount of data.

3.4. Articulatory gestures (Experiment 2)

In this section, we explore two different articulatory dimensions: (i) time lags of consonantal (bilabial closure, i.e. lip aperture, LA) and vocalic (palatal narrow, i.e. tongue body, TB) gesture onset at stressed-syllable onset (cf. 3.4.1), and (ii) timing of the target of the vocalic gesture (cf. 3.4.2). While (i) was motivated by the prediction of a c-center effect (cf. 1.3), (ii) was motivated by an initial qualitative assessment of the data, which suggest a later timing of TB gesture target in A1 than in A2.

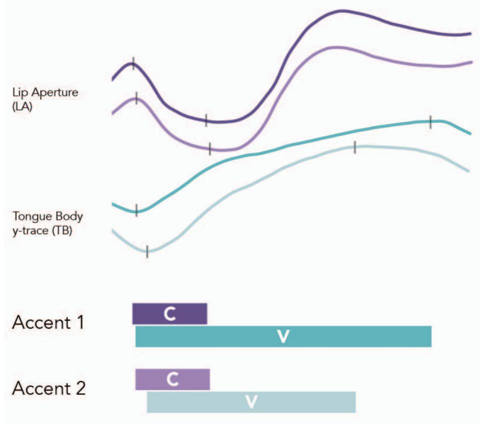

Figure 3: *Trajectories with marked onset and target of the consonantal gesture (bilabial closure, LA), and the vocalic gesture (palatal narrow, vertical position of TB), of an A1 and an A2 word.*

3.4.1. *Gestural co-ordination in onset*

Figure 3 above displays an example of articulatory gestural trajectories for an A1 token and an A2 token, as pronounced by speaker M. In this example, we see that the consonantal gesture starts slightly earlier in A2 than in A1. To corroborate this observation, we measured the time lags between onsets of the consonantal (LA) and the vocalic (TB y-trace) gestures. Results are displayed in Figure 4, suggesting for speaker M that the consonantal (LA) gesture starts somewhat earlier for A2 than for A1. This difference is significant for speaker M (t=-2.75, df=14.77, p=.015*), however, not for speaker S (t=-.44, df=18.91, p=.66; cf. Fig. 4).

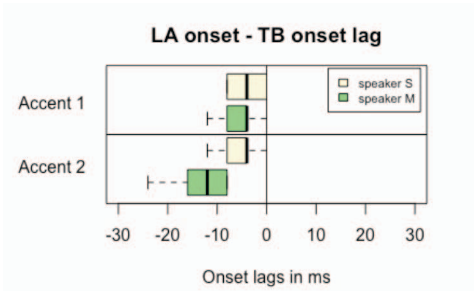

Figure 4: *Time lags of the consonantal gesture onset (LA) and the vocalic gesture onset (TB) in word onsets. Speaker M displays a small but significant c-center effect in Accent 2.*

An earlier onset of a consonantal gesture can be explained as the result of a c-center effect, when articulatory (consonantal or tonal, cf. 1.2 above) gestures co-occur and compete at syllable onset. In our case, it could be a result of (i) coordination with different tone gestures: presumably a H-gesture and a L-gesture, respectively. Studies on Mandarin have indeed reported different coordination patterns between the rising Tone 2 and the falling Tone 4 [19, 23]. Different coordination patterns could also be due to (ii) different coupling relations for A1 and A2. However, this is not within the scope of this article. An alternative explanation could assume (iii) a lexical tonal target in A2, but not in A1 [8], and relate the competitiveness to the phonological status of the tone. Indeed, post-lexical tone gestures have been shown to not result in a c-center effect, hence they seem to not compete with the consonantal gestures in onset [21].

Although the results do not suggest a c-center effect for speaker S, we observe traces of another, albeit related effect at syllable onset: For speaker S, the duration of the bilabial closure gesture (defined as the distance of LA target from LA onset) was on average slightly shorter in A2 than in A1 (marginally significant at t=1.97, df=11.79, p=.072), while no such effect did appear for speaker M (t=-.57, df=13.40, p=.58).

We suggest that these two effects observed for S and M might be two different outcomes of the same underlying mechanism, as they might have the equivalent effect of an earlier bilabial closure release (which need not, but might possibly following from a shorter LA gesture) with respect to the vocalic gesture. This would provide an earlier *acoustic* vowel onset in A2 than A1, which in turn might be perceptually motivated, supporting listeners anticipating the nature of the upcoming lexical accentual tone (a low tone in A2). An early recognition of lexical tones is advantageous as it has been shown for Swedish how word stem tones are used by listeners to predict upcoming word endings [36]. We will leave the verification of this perception-based interpretation to future research.

3.4.2. Timing of the vocalic gestural target

A salient articulatory difference between A1 and A2 consistently observed for both speakers was a later-timed gesture target (defined as the duration target–onset) of the vocalic gesture (i.e. the maximum height of the tongue body in /i/) in A1. This effect is seen in Figure 5; it is significant for both S (t=-5.84, df=14.28, p<.001***) and M (t=-2.53, df=11.33, p=.027*).

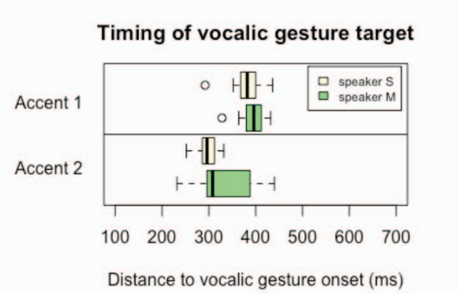

Figure 5: *Timing of the vocalic gesture target. For both speakers the vocalic gestural target (maximum height of the tongue body) is later in A1 than in A2.*

For a tentative interpretation of this effect we consider the possibility of a coupling to the durational effect attested in Experiment 1, as these two effects – a longer acoustic duration of the post-vocalic consonant (Exp. 1) and an earlier timing of the vocalic gesture target (Exp. 2) in A2 – are among the most consistent results of the respective experiments (after the expected results for F0). A longer acoustic duration of the consonant in A2 (Exp. 1) is possibly achieved by an earlier consonantal gesture; this in turn might trigger an earlier timing of the (pre-consonantal) vocalic gesture target. A partial support for this interpretation comes from additional measurements of the timing of the tongue tip gesture target for /l/, which indeed proved significantly earlier in A2 than in A1, however only for speaker S (t=-3.94, df=16.61, p=.0011**), but not for M (t=-.18, df=16.00, p=.86). However, an alternative interpretation is that the timing of the vocalic gesture target is very much related to the coupling of the tone gesture in A2, which could force the vocalic target timing to be earlier than in A1. A stable coordination pattern between the tone peak and the target of the vocalic gesture has been found in German pitch accents [26].

3.5. Formant patterns (Experiment 2)

If the articulatory findings reported in 3.4 are to be considered potential perceptual cues of Swedish word accents, then they should exhibit some acoustic correlate. We therefore measured the following formant frequencies: Following [37], who found F2 and F3 during stop closure to be affected by gestural overlap, we measured mean F2 and F3 of the occlusion phase of /b/, as well as F2 and F3 at vowel onset. The formants of the vowel onset were measured as a specific point immediately following the release of /b/. Two criteria were used to establish the point: 1) it was the second pulse with three distinct formants that constitute the vowel; 2) there was a steep rise in the amplitude at the vowel onset.

The results for the occlusion phase revealed an effect of word accent on the F3-F2 difference for speaker S (t=2.14, df=17.25, p=.047*), but not for F2 or F3 separately, and not at all for speaker M. For vowel onset, we found a complementary pattern: no effects of word accent for speaker S, but an effect on F3-F2 for M (t=2.83, df=12.62, p=.015*). That is, the c-center effect observed for M (cf. 3.4.1) relates to an acoustic effect at vowel onset, while the (marginal) difference in LA-duration in speaker S seems to relate to formant differences during the stop occlusion.

These results suggest that articulatory correlates of the word accents have acoustic effects and thus qualify as candidates for perceptual word accents cues.

4. Conclusion

Probably due to the strong attested power of F0 as a perceptual cue to the (South) Swedish word accent contrast (cf. 1), additional secondary cues, have so far hardly attracted any attention. This study suggests that South Swedish word accents are distinguished in speech production by means of several phonetic (articulatory and acoustic) dimensions: F0 timing, creaky voice predominantly in A1 , a longer duration of a post-stress consonant in A2, different gestural coordination at word onset (also mirroring in the acoustic domain in terms of the F3-F2 difference), and an earlier reached target of the vocalic gesture in A2. Future research will need to evaluate the perceptual relevance of these potential correlates of Swedish word accents and their usefulness, e.g., in the on-line prediction of upcoming suffixes [36].

5. Acknowledgements

We gratefully acknowledge the Humanities Laboratory, Lund University for offering experimental studios and EMA recording facilities. We also wish to thank Dr. Johan Frid (SWE-CLARIN) for assistance with data analysis (Exp. 2), as well as Anneliese Kelterer (student assistant) for annotating data from Exp. 1. Exp. 1 was conducted by the second author within a grant from the Swedish Research Council (PI Johan Frid, 2009-1566). Last but not least, we thank our participants!

6. References

[1] E. A. Meyer, "Die Intonation im Schwedischen," *Erster Teil: Die Sveamundarten*, Stockholm: Fritzes, 1937.

[2] B. Malmberg, "Bemerkungen zum schwedischen Wortakzent", *Zeitschrift für Phonetik*, vol. 12, pp. 193–207, 1959.

[3] B. Malmberg, "Nyare fonetiska rön," Lund: Gleerup, 1966.

[4] D. House and G. Bruce, "Word and focal accents in Swedish from a recognition perspective," in *Nordic Prosody V – Papers from a Symposium, Turku,* pp. 156–173, 1990.

[5] G. Ambrazaitis and G. Bruce, "Perception of South Swedish Word Accents," *Working Papers 52, Proceedings from Fonetik 2006*, G. Ambrazaitis, and S. Schötz, Eds. Lund, pp. 5–8, 2006.

[6] G. Bruce, "Swedish Word Accents in Sentence Perspective," *Travaux de l'institut de linguistique de Lund*, no. 12, Lund: Gleerup, 1977.

[7] O. Engstrand, "Phonetic interpretation of the word accent contrast in Swedish," *Phonetica*, vol. 52, pp. 171–179, 1995.

[8] T. Riad, "Scandinavian accent typology," *Sprachtypologie und Universalienforschung (STUF),* vol. 59, no. 1, pp. 36–55, 2006.

[9] G. Bruce and E. Gårding, "A prosodic typology for Swedish dialects," In *Nordic Prosody – Papers from a Symposium, Lund*, pp. 219–228, 1978.

[10] B. Malmberg, "Sydsvensk ordaccent – en experimentalfonetisk undersökning," *Lunds Universitets Årsskrift. N. F. Avd. 1*, vol. 49, Lund: Gleerup, 1953.

[11] C.-C. Elert, "Phonologic Studies of Quantity in Swedish. Based on Material from Stockholm Speakers," Stockholm: Almqvist & Wiksell, 1964.

[12] M. Heldner, "Focal accent – F0 movements and beyond", *Ph. D. thesis*, Umeå University, 2001.

[13] V. Van Heuven and E. van Zanten, "Speech rate as a secondary prosodic characteristic of polarity questions in three languages," *Speech Communication* 47, pp. 87–99, 2005.

[14] O. Niebuhr, "The acoustic complexity of intonation," *Nordic Prosody XI,* E-L. Asu and P. Lippus, Eds. Peter Lang, pp. 25–38, 2013.

[15] G. Ambrazaitis and J. Frid, "F0 peak timing, height, and shape as independent features," *in Proc. 4th International Symposium on Tonal Aspects of Languages, Nijmegen,* pp. 138–142, 2014.

[16] C. Browman and L. Goldstein, "Towards an articulatory phonology," *Phonology Yearbook 3*, pp. 219–252, 1986.

[17] C. Browman and L. Goldstein, "Articulatory phonology: An overview," *Phonetica*, vol. 49, no. 3–4, pp. 155–180, 1992.

[18] D. Byrd and E. Saltzman, "The elastic phrase: modeling the dynamics of boundary-adjacent lengthening," *Journal of Phonetics*, vol. 31, pp. 149–180, 2003.

[19] M. Gao, "Tonal alignment in Mandarin Chinese: An articulatory phonology account," Doctoral dissertation, Yale University, 2008.

[20] H. Niemann, D. Mücke, H. Nam, L. Goldstein, and M. Grice, "Tones as Gestures: the Case of Italian and German," in *Proceedings of ICPhS XVII, Hong Kong,* pp. 1486–1489, 2011.

[21] D. Mücke, H. Nam, A. Hermes, and L. Goldstein, "Coupling of tone and constriction gestures in pitch accents," In *Consonant Clusters and Structural Complexity*, P. Hoole, L. Bombien, M. Pouplier, C. Mooshammer, and B. Kühnert, Eds. Munich: Mouton de Gruyter, pp. 157–176, 2012.

[22] A. Katsika, J. Krivopapic´, C. Moonshammer, M. Tiede, and L. Goldstein, "The coordination of boundary tones and its interaction with prominence," *Journal of Phonetics*, vol. 44, pp. 62–82, 2014.

[23] H. Yi and S. Tilsen, "Interaction between lexical tone and intonation: an EMA study," in INTERSPEECH 2016 – 17th Annual Conference of the International Speech Communication Association, Sep 8–10, San Francisco, USA, Proceedings, pp. 2448–2452, 2016.

[24] 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.

[25] H. Niemann, M. Grice, and D. Mücke, "Segmental and positional effects in tonal alignment: An articulatory approach," *in Proceedings of 10th ISSP, Cologne,* pp. 285–288, 2014.

[26] H. Niemann and D. Mücke, "Effects of phrasal position and metrical structure on alignment patterns of nuclear pitch accents in German: Acoustics and articulation," in *Proceedings of 18th International Congress of Phonetic Sciences, 10–14 August, Glasgow, UK*, 2015.

[27] J. Caspers and V. J. Van Heuven, "Effects of time pressure on the phonetic realization of the Dutch Accent-lending pitch rise and fall," *Phonetica,* vol. 50, pp. 161–171, 1993.

[28] A. Arvaniti, D. R. Ladd, and I. Mennen, "Stability of tonal alignment: the case of Greek prenuclear Accents," *Journal of Phonetics*, vol. 26, no. 3–25, 1998.

[29] Y. Xu, "Consistency of tone-syllable alignment across different syllable structures and speaking rates," *Phonetica,* vol. 55, pp. 179–203, 1998.

[30] D. R. Ladd, D. Faulkner, H. Faulkner, and A. Schepman, "Constant 'Segmental anchoring' of F0 movements under changes in speech rate," *J. Acoust. Soc. Am*., vol. 106, no. 3, pp. 1543–1554, 1999.

[31] M. Atterer and D. R. Ladd, "On the Phonetics and Phonology of 'Segmental anchoring' of F0: Evidence from German," *Journal of Phonetics,* vol. 32, pp. 177–197, 2004.

[32] H. Nam, L. Goldstein, and E. Saltzman, "Self-organization of syllable structure: a coupled oscillator model," in *Approaches to phonological complexity*, F. Pellegrino, E. Marisco, and I. Chitoran, Eds. Berlin/New York: Mouton de Gruyter, pp. 299–328, 2009.

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

[34] Y. Xu, "ProsodyPro — A Tool for Large-scale Systematic Prosody Analysis," in Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013), Aix-en-Provence, France, pp. 7–10, 2013.

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

[36] M. Roll, P. Söderström, and M. Horne, "Word-stem tones cue suffixes in the brain," *Brain Research*, no. 1520, pp. 116–120, 2013.

[37] E. C. Zsiga, "Acoustic evidence for gestural overlap in consonant sequences," *Journal of Phonetics*, vol. 22, pp. 121–140, 1994.