INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

Strength and structure: Coupling tones with oral constriction gestures

Doris Mücke[1], Anne Hermes[1,2], Sam Tilsen[3]

[1]IfL-Phonetics, University of Cologne, Germany
[2]Laboratoire de Phonétique et Phonologie, UMR 7018, CNRS/Sorbonne-Nouvelle, France
[3]Cornell University, Ithaca, United States

doris.muecke@uni-koeln.de, anne.hermes@uni-koeln.de, tilsen@cornell.edu

Abstract

According to the segmental anchor hypothesis within the Autosegmental-Metrical approach, tones are aligned with segmental boundaries of consonant and vowels in the acoustic domain. In prenuclear rising pitch accents (LH*), the rise is assumed to occur in the vicinity of the accented syllable it is phonologically associated with. However, there are differences in the alignment patterns within and across languages that cannot be captured within the AM approach. In the present study, we investigate the coordination of tonal and oral constriction gestures within Articulatory Phonology. Therefore, we model the coordination of prenuclear LH* pitch accents in Catalan, Northern and Southern German with respect to syllable production on the basis of recordings with a 2D electromagnetic articulography. We provide an extended coupled oscillators model that allows for balanced and imbalanced coupling strengths. Based on examples, we show that the observed differences in alignment patterns for prenuclear rising pitch accents can be modelled with the same underlying coordinative structures/coupling modes for vocalic and tonal gestures and that surface differences arise from gradient variation in coupling strengths.

Index Terms: dynamical systems, tonal alignment, tonal gestures, oral constriction gestures, computational model of variability, imbalanced coupling

1. Introduction

1.1. Tonal alignment in the AM approach

Within the Autosegmental-Metrical (AM) approach, tones are associated with tone-bearing units of the textual string, such as stressed syllables in German pitch accents. The tonal alignment research extended this concept by developing the segmental anchoring hypothesis, measuring patterns of co-occurrences of pitch movements with boundaries of the segmental string in the acoustic dimension [1-6]. Segmental anchoring and the related measures of co-occurring events in the tonal and segmental string are not meant to be a direct reflex of the phonological categorization: Tonal *association* is categorical and low-dimensional (phonological representation), while tonal *alignment* is continuous and high-dimensional (phonetic representation).

Figure 1 provides an example for the alignment continuum of prenuclear rising pitch accents (LH*) measured on the acoustic surface for Northern German, Southern German (Viennese) and Catalan, adapted from [6,7]. The figure provides the respective acoustic alignment patterns for L and H* in terms of F0 turning points relative to segmental boundaries in the textual string. The patterns show that the nuclear rises are aligned earlier in Catalan than in the German varieties. Within the German varieties, the alignment is earlier in Viennese than in Northern German. Differences are subtle but consistent, and all of them are likely reflecting the same phonological categorisation [8].

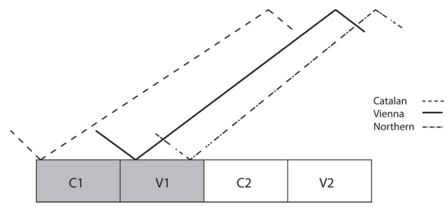

Figure 1: *Tonal alignment patterns with acoustic segments for nuclear rising accents in Catalan, Vienna German and Northern German.*

In the AM approach, L and H are associated with a tone-bearing unit. This is usually the accented syllable, carrying the tone. However, in the examples shown above, the H peaks always fall outside of the accented syllable on the unstressed syllable when looking at the acoustic surface. In Catalan, the H peak occurs at the end of the consonant of the unstressed syllable, while it occurs at the end of the following vowel of the unstressed syllable in Northern German. In terms of a phonological association in the AM framework, we might state that L is phonologically associated with the left periphery of the unstressed syllable in Catalan and with the right periphery in Northern German, but this analysis is not very convincing. The picture is even more complicated when comparing Northern German and Viennese German. In both varieties, the rise is aligned with the vowel in the unstressed syllable, but it occurs at the beginning of the V segment in Viennese German and at the end of the same segment in Northern German. It is somewhat problematic to express all of the phonetic variation through differences in phonological associations, because such an approach seems to over-complicate the theory and raises the question of whether we can arbitrarily add categories to a theory whenever a new surface pattern is observed [8]. It our aim to exemplify how to treat variability within a dynamical account.

1.2. Coordination of tonal and oral gestures in AP

Articulatory Phonology [9, 10] decomposes speech into a set of potentially overlapping units, articulatory gestures. The temporal organisation of gestures can be modelled by a

http://dx.doi.org/10.21437/Interspeech.2019-2650

network of coupled oscillators [11-17]. Within this self-organized network, each gesture is associated with an oscillator (or clock) and the oscillators are coupled to one another in a pairwise fashion. In the planning process of an utterance, coupling between the oscillators forces them to settle into a stable timing pattern. In execution, the oscillators then function to trigger the initiation of a specific gestures that they are coupled to. There are two coupling modes: gestures start either at the same time (in-phase, relative phase 0° of a gesture's oscillator) or one after the other (anti-phase) relative phase 180° of a gesture's oscillator). Figure 2 schematises different coupling structures for different syllable structures.

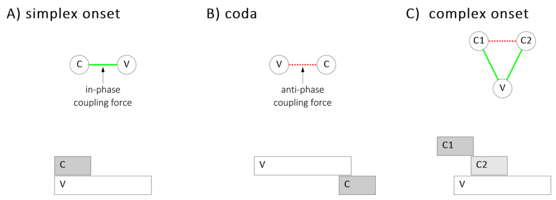

Figure 2: *Coupling graphs (top) and related gestural scores CV, VC and CCCV syllable. Solid lines (green) mark in-phase coupling and dashed lines (red) mark anti-phase coupling.*

In a CV syllable, C and V gestures are associated with in-phase coupling. They start at the same time, but the V movement is slower and on the acoustic surface, we get the impression of a CV sequence. In a VC syllable, the V and C gesture are associated with anti-phase coupling, i.e. they are initiated sequentially. In a C1C2V syllable, with a branching onset, the coupling forces compete with each other. Both Cs are in-phase with the V, but in anti-phase with one another. This coupling structure leads to a rightward shift C2 towards the following V and a leftward shift of C1 away from it. In a prototypical form, it is assumed that coupling forces are equal in strength, leading to symmetrical shift pattern on the surface representation.

Within the AP framework, F0 control is accomplished through tonal gestures, which are analogous to oral articulatory gestures [18, 7, 19, 20]. Therefore, tonal gestures can be directly coupled to syllable-internal gestures such as consonants and vowels in terms of in-phase and anti-phase relations. For Catalan and Viennese German, different coupling structures/topologies have been suggested for prenuclear rising pitch accents, LH* [7]: (i) In Catalan the high tonal gesture is coupled in-phase with the V gesture and anti-phase with the low tonal gesture. On the acoustic surface, the start of the rise co-occurs with the beginning of the syllable. (ii) In Viennese, a competing coupling structure is assumed, i.e. both tonal gestures are coupled in-phase with the V and anti-phase with one another. This leads to a later rise on the acoustic surface. However, the different coupling structures cannot account for the differences between Northern and Southern German, when equal coupling strength (i.e. balanced coupling strength) is assumed.

1.3. Aim of the present study

In the present study, we model prenuclear rising pitch accents LH* in Catalan and German varieties (Northern and Southern) by assigning the *same coordinative coupling structure* for the tonal and oral constriction gestures. To account for the observed variation in the output pattern, we adjust the coupling strength parameter that comes for free with a coupled oscillators model, which allows for imbalanced coupling strengths between gestures. We do this by using examples for each language and variability.

2. Method

Articulatory recordings were collected with a 2D Electromagnetic Articulograph (Carstens AG100). The kinematic data were sampled at 400 Hz, downsampled to 200 Hz and smoothed with a low-pass filter at 40Hz. Sensors were placed on the upper and lower lip, tongue tip, tongue blade, and tongue dorsum. The articulatory data was annotated within the EMU Speech Database System [21]. Landmarks in the articulatory domain for consonantal and vocalic gesture were identified in the vertical plane. The onsets and targets (local minima and maxima) of the respective tonal, consonantal and vocalic gestures were labelled using zero-velocity crossing in the velocity curve. The recordings of the datasets are related to [6, 7]. For the modelling of the tonal alignment data we used the recordings of 3 speakers (Northern German, Vienna and Catalan).

The speech material consists of disyllabic target words (CV.CV) that were embedded in carrier sentences (mini-dialogues) for German (1) and Catalan (2). All target words were elicited in contrastive focus condition and carried a rising (LH) pitch accent.

(1) Q.: Hat sie die Nanni oder die Mahmi bestohlen) *(lit: Has he the Manni or the nanni robbed?)*
A: "Er hat die MAHmi bestohlen" (lit: *He robbed the Mahmi*).

(2) Q: Va venir la Mimamila? *(lit: Came the Mimamila?)*
A: "No, la MiMAmi" *(lit: No, the Mimami)*

To model differences in the alignment of the tonal target relative to articulatory targets, we used the following variables: the start (L) and the target of the tonal rise gesture (H) relative to the onset of the vocalic gesture (LV and HV).

3. The coupled oscillators model

The extended model we used is based on a standard implementation of the coupled oscillators model of Articulatory Phonology [22, 23], schematized in Fig. 3 below. In this model, consonantal gestural planning oscillators are anti-phase coupled to one another and are in-phase coupled to a vocalic planning oscillator. The potential functions and forces for in-phase (V+/F+) and anti-phase (V-/F-) coupling are shown in Fig. 3A, 3B. The relative strengths of these forces determine a stable equilibrium of relative phase for the oscillators. Relative phases can be viewed as angles on the phase circle (Fig. 3C), or as distances in time between any given phase of two oscillations (e.g. peaks, as in Fig. 3D). Prior to utterance initiation, the relative phases of the oscillators evolve towards a stable equilibrium. Once all relative phases have stabilized (arrow in Fig. 3E), the gesture corresponding to each planning oscillator is initiated when the oscillator reaches a phase of 0 (i.e. arrows in Fig. 3D). Note that the frequency (*f*) of the oscillators and their relative phases determine the temporal pattern of movement initiation, where $\Delta_{C1V} = \phi_{C1V}/(2\pi f)$ and $\Delta_{C2V} = \phi_{C2V}/(2\pi f)$.

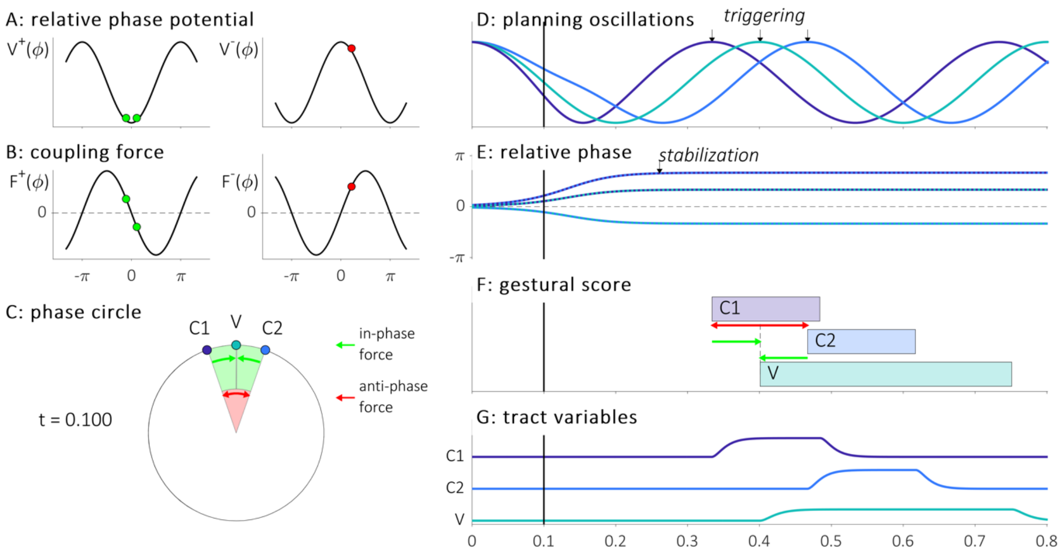

Figure 3: *Overview of the coupled oscillators model of the c-center effect. (A) In-phase (V+) and anti-phase (V-) relative phase potentials. (B) Forces on oscillator phases are the opposite of the derivative of the relative phase potential. (C) Planning oscillator phases at time t=0.100. (D) Planning oscillations over time; arrows indicate when each oscillator triggers activation of the corresponding gesture. (E) Relative phases over time; after stabilization the oscillators trigger the initiation of gestural activation. (F) Gestural scores. (G) Tract variables.*

In prototypical, canonical implementations of the coupled oscillators model, coupling forces are assumed to be balanced in two ways: (1) the strength of the anti-phase force and the average of the in-phase forces are equal: ($b/\hat{a} = 1$), where $\hat{a} = (a_1+a_2)/2$; and (2) both consonantal gestures are in-phase coupled to the vocalic gesture with equal strength ($a_1=a_2$). The relevant matrix of coupling strength parameters is shown below. We refer to the anti-phase to in-phase ratio ($b/\hat{a}$) as the *the strength of anti-phase coupling relative to in-phase coupling*, and the difference between a_1 and a_2 (i.e. a_1-a_2) as the *coupling difference*. The condition that $a_1=a_2$ is referred to as *balanced coupling*.

Under the assumption of balanced coupling, the stabilized relative phases of C1/C2 to V are $\phi = \pm\pi/3$. The balanced coupling model always generates LE/RE shifts which are equal and in opposite directions. In addition to the balanced coupling model, we optimized two expanded models, in order to assess whether apparent deviations of empirical data could be accommodated by gradient variation in coupling strength and perturbations which bias measurements.

$$\begin{array}{cccc} & \mathrm{C1} & \mathrm{C2} & \mathrm{V} \\ \mathrm{C1} & & b & a_1 \\ \mathrm{C2} & b & & a_2 \\ \mathrm{V} & a_1 & a_2 & \end{array}$$

Both of the expanded models allowed for a non-zero coupling difference, i.e. imbalanced coupling.

4. Results

In our model, we assign *the same coordinative patterns* for tones and oral constriction gestures in Catalan and the German varieties. To account for differences of the acoustic surface, we assigned different coupling strengths between the tonal and the vocalic gesture to model the different patterns. We assumed that the mechanism of *imbalanced coupling strength* can account for variation in tonal alignment across languages and dialects, making an analysis which posits distinct tonal categories in our view is unnecessary.

Figure 4 shows the coupling patterns in the dynamic articulatory model and the respective acoustic outcome for Catalan, Viennese and Northern German. Our model holds that the low tonal gesture and the high tonal gesture in an LH rising pitch accent are both coupled in-phase to the vowel, and for perceptual recoverability, the tonal gestures are anti-phase coupled with one another. This forms a complex coupling structure of competing coupling forces between L, H and V, analogous to the modelling of branching onsets on the syllable level. While the underlying coordinative structure is invariant in our model, the coupling strengths between tonal gestures and the vowel can vary. These were optimized to fit the surface timing patterns. Note that because we did not have an empirical measure of the onset of the L tone, we assumed that the low tonal gesture is initiated at a constant interval preceding the vocalic gesture, which we determined as the average V-to-H onset interval. The different strength parameters for LV and HV coupling for one speaker in Northern German, Viennese German, and Catalan in terms of coupling matrices are displayed in Figure 4A. For the Northern speaker, stronger LV coupling compared to HV coupling results in a long delay of the initiation of the H tonal gesture relative to the vocalic gesture. On the acoustic surface (see Fig. 4C), this corresponds to a late alignment pattern. The Viennese speaker had more balanced LV and HV coupling, and thus exhibits a more symmetric shift of L and H onset relative to V onset. In contrast, the Catalan speaker showed a stronger HV coupling leading to the relatively early initiation of the H gesture.

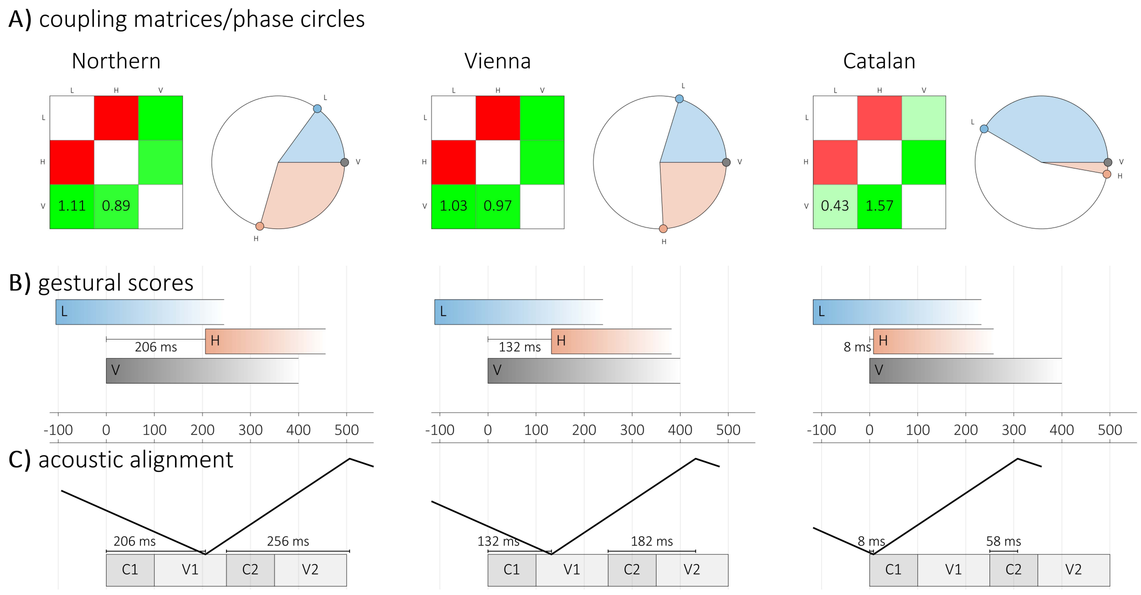

Figure 4: *Comparison of three tonal alignment patterns which result from gradient differences in coupling strengths for Northern German, Viennese German and Catalan. (A) Coupling matrices and phase circles: A greater imbalance of LV and HV coupling strengths results in a larger shift of the H gesture initiation. (B) Gestural scores. (C) Acoustic alignments.*

5. Discussion and conclusion

We have shown that in a dynamic model the same topology of network organisation can generate variation in the timing of pitch accents in Catalan, Viennese and Northern German (in line with [18,7,19]. More specifically, in our model we assume that the tonal gestures, L and H, are coupled in an in-phase mode to the V in the accented syllable (the tone-bearing unit) and in an anti-phase mode with respect to each other. This assumed coordinative structure can account for two effects: first, since we are dealing with coordinated movements instead of co-occurrences of tones and segments, we can generate a pattern in which the H peak falls outside the accented syllable (the tone-bearing unit), due to the competitive structure of the tones and the V (triggering H later than V with respect to the accented syllable). Secondly, by allowing for imbalanced coupling strength in the coordinative organization, differences in tonal alignment on the acoustic surface were modelled. While Viennese shows a balanced coupling strength pattern, the dominance of the L-V coupling (stronger L-V coupling) in Northern German leads to relatively late alignment of the H peak on the acoustic surface. Vice versa, the dominance of the H-V coupling (stronger H-V coupling) in Catalan leads to an early alignment of the H peak.

We assume a coupling of the tonal gestures to the V gesture of the accented syllable (in line with the Autosegmental-Metrical approach of the V gesture carrying the tone). The coupling of the two tonal gestures with the V gesture has the advantage that we do not need a further explanation why – on the phonetics surface – the peak falls outside of the accented syllable. Thus, the late peak is the result of an underlying phonological coordination pattern and not a result of the co-occurrences between F0 turning points and segmental boundaries. A larger database is necessary to prove the concept. But it is our primary aim to point out a way how to treat systematic variability in speech production within in a dynamic system and relate it to traditional phonology theory.

6. Acknowledgements

This work was supported by the German Research Foundation (DFG) as part of the SFB 1252 “Prominence in Language” in the project A04 “Dynamic Modelling of Prosodic Prominence” at the University of Cologne.

7. References

[1] A. Arvaniti, D.R. Ladd and I. Mennen. “Stability of tonal alignment: the case of Greek prenuclear accents.” *Journal of Phonetics* vol. 26, pp. 3-25, 1998.

[2] D.R. Ladd, D. Faulkner, H. Faulkner and A. Schepman. “Constant “segmental anchoring” of F0 movements under changes in speech rate.” *Journal of the Acoustical Society of America* vol. 106, pp. 1543-1554, 1999.

[3] D.R. Ladd and A. Schepman. “Phonological conditioning of peak alignment of rising pitch accents in Dutch.” *Journal of the Acoustical Society of America* vol. 107, pp. 2685-2696, 2000.

[4] M. D’Imperio, C. Petrone and N. Nguyen. “Effects of tonal alignment on lexical identification in Italian.” In C. Gussenhoven & T. Riad (eds.), *Tones and tunes, Vol. 2.* Berlin: Mouton de Gruyter, pp. 79-106, 2007.

[5] P. Prieto, and F. Torreira. “The segmental anchoring hypothesis revisited. Syllable structure and speech rate effects on peak timing in Spanish.” *Journal of Phonetics,* vol 35(4), pp. 473–500, 2007

[6] D. Mücke, M. Grice, J. Becker and A. Hermes. “Sources of variation in tonal alignment: evidence from acoustic and kinematic data.” Journal of Phonetics, vol. 37(3), pp. 321–338, 2009.

[7] D. Mücke, H. Nam, A. Hermes, and L. Goldstein, L. (2012). Coupling of tone and constriction gestures in pitch accents. *Consonant clusters and structural complexity*, pp. 205-230. Berlin, New York: Mouton De Gruyter, 2012.

[8] M. Atterer and D.R. Ladd. "On the phonetics and phonology of "segmental anchoring" of F0: evidence from German." *Journal of Phonetics* vol. 32(2), pp. 177-197, 2004.

[9] C.P. Browman, and L. Goldstein. "Towards an articulatory phonology". In C. Ewen & J. Anderson (eds.), *Phonology Yearbook 3*. Cambridge: Cambridge University Press, 219–252, 1986.

[10] C.P. Browman, and L. Goldstein (2000). "Competing constraints on intergestural coordination and self-organization of phonological structures." *Les Cahiers de l'Institut de la Communication parlée* 5, pp. 25-34, 2000.

[11] H. Nam, L. Goldstein and E. Saltzman (2010). „Self-organization of syllable structure: a coupled oscillator model." In F. Pellegrino, E. Marisco & I. Chitoran (eds.), *Approaches to phonological complexity*. Berlin: de Gruyter, pp. 299-328, 2010

[12] L. Goldstein, H. Nam, E. Saltzman and I. Chitoran. "Coupled oscillator planning model of speech timing and syllable structure." In G. Fant, H. Fujisaki & J. Shen (eds.), *Frontiers in Phonetics and Speech Science*. Beijing: The Commercial Press, 239-250, 2009.

[13] M. Pouplier, "The gestural approach to syllable structure: Universal, language-and cluster-specific aspects." *Speech planning and dynamics*, 63-96, 2012.

[14] J.A. Shaw, A. Gafos, P. Hoole, and C. Zeroual, C. "Dynamic invariance in the phonetic expression of syllable structure: a case study of Moroccan Arabic consonant clusters. *Phonology*, vol. *28*(3), pp. 455-490, 2011.

[15] A. Gafos, S. Charlow, J.A. Shaw, and P. Hoole. "Stochastic time analysis of syllable-referential intervals and simplex onsets." *Journal of Phonetics*, vol. *44*, pp. 152-166, 2014.

[16] A. Hermes, D. Mücke, and B. Auris, B. "The variability of syllable patterns in Tashlhiyt Berber and Polish." *Journal of Phonetics*, vol. *64*, pp. 127-144, 2017.

[17] S. Tilsen. "Selection and coordination: The articulatory basis for the emergence of phonological structure." *Journal of Phonetics*, vol. *55*, pp. 53-77, 2016

[18] M. Gao. "Gestural coordination among vowel, consonant and tone gestures in Mandarin Chinese." *Chinese Journal of Phonetics*, vol. *2*, pp. 43-50, 2009.

[19] H. Niemann, D. Mücke, H. Nam, L. Goldstein & M. Grice. „Tones as Gestures: The Case of Italian and German." *Proceedings of the 17th International Congress of Phonetic Sciences,* Hongkong, China. 1486-1489, 2011.

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

[21] S. Cassidy and J. Harrington. "Multi-level annotation in the EMU speech database management system." *Speech Communication* vol. 33, pp. 611-677, 2011.

[22] E. Saltzman, H. Nam, J. Krivokapic and L. Goldstein. „A task-dynamic toolkit for modeling the effects of prosodic structure on articulation". *Proceedings of the 4th International Conference on Speech Prosody,* Campinas, Brazil. Pp. 175-174, 2008

[23] S. Tilsen, S. "Exertive modulation of speech and articulatory phasing." *Journal of Phonetics*, *64*, 34-50, 2017.