ON THE ARTICULATORY BASES OF PROMINENCE IN ITALIAN

Cinzia Avesani, Mario Vayra° and Claudio Zmarich**

*Institute for Cognitive Sciences and Technologies - CNR, Padova, Italy
°Dipartimento di Studi Linguistici e Orientali - University of Bologna, Italy

{avesani, zmarich}@pd.istc.cnr.it; mario.vayra@unibo.it

ABSTRACT

This study reports the first results of a research aimed to investigate how segmental variation is conditioned by prosody in Italian, by examining the acoustic and articulatory properties of syllables that are prominent at different levels of the prosodic hierarchy. We examined lip movement kinematics of unstressed, stressed and nuclearly accented syllables in order to understand the kinematic characteristics of accent-induced articulatory strenghthening. The kinematic results are then interpreted within a Task Dynamics model to evaluate how prosodically-driven variation can be accounted for by a particular dynamical parameter setting in a mass-spring gestural model.

1. INTRODUCTION

Previous work on Italian has provided both acoustic and articulatory (jaw displacement) evidence that at word level unstressed vowels are shorter in duration as compared to stressed ones and exhibit reduction at least in the height dimension (F1), involving the global gesture for the vowel [14]. These results have been extensively confirmed: as shown in [11], all stressed vowels are longer than unstressed ones and show less displacement for the lip opening gesture independently of their height (with the exception of /u/). Moreover, electropalatographic results on high and low vowels tongue movements indicate both an increase in the “Posteriority Index” [6,7] for stressed /a/, /i/ and /u/, pointing to a more extreme jaw lowering movement, and an increase in the “Centrality Index” for high vowels, compatible with a more extended tongue raising movement towards the target, with a consequent tighter constriction. Finally, as shown in [15], the slope of the “locus of equation” that relates F2 at the midpoint of a vowel to F2 at the transition onset of a CV syllable is steeper in unstressed than in stressed and in accented syllables. The results indicate that the coarticulatory “resistence” of a syllable varies systematically with its prominence level.

These data are compatible with results reported in the vast literature on the acoustic correlates of stress in English: stressed vowels have longer duration, less centralization, higher F0 and, *ceteris paribus*, more intensity and resistence to coarticulation than unstressed vowels [e.g. 2,5,8]. Quite a number of studies on the kinematic properties of stressed syllables agree that actions of the articulators that implement gestures of stressed segments (jaw, tongue, lips) are longer in duration, show greater displacement (jaw and lower lip) and higher peak velocities than gestures for unstressed syllables [e.g., 1,2,4,5,8,10].

Studies of the jaw and tongue kinematics have yielded to different accounts of the prominence-enhancing strategies speakers employ in the production of accentually prominent syllables: they can enhance the “intrinsic sonority” of the segments [1], or, through a “localized hyperaticulation”, their distinctive features in a way that maximizes lexical distinctions [5]. The two strategies that make a low vowel more clear (sonority expansion) and more periferal (localized hyperarticulation) can be simultaneously at work in accented syllables, and their co-presence can explain the apparently conflicting results found in high vowels of an increase of jaw lowering and a simultaneous narrowing of the palatal or velar constriction.

The available evidence on Italian seems to suggest that, as in English, accent-induced prosodic strenghtening is achieved by both strategies, and that in low vowels sonority expansion and hyperarticulation of the lingual gesture sum up their prominence enhancing effects. However, most of the previosus studies on Italian prominence did not clearly control for the levels of the stress hierarchy involved, and even when there is an attempt to distinguish between prominences at the highest levels of the hierarchy [6] it is left undefined whether we are dealing with a stress vs accent or with an unstress vs accent contrast.

New findings on stress are due to investigators adopting a “dynamical system” perspective on speech production [e.g, 3,8,12,13]. In this framework, a speech gesture can be modeled as a mass-spring dynamical system with a fixed set of parameter values that characterizes a local constriction in the vocal tract.

In a dynamical system, stressing can be achieved by varying the values of few abstract dynamic parameters: intergestural phasing, target (underlying amplitude), stiffness and also “linear rescaling” [10], a proportional scaling of target and stiffness that can be conceived as a unique parameter [e.g., 4].

In the present work we intend to address the question of how prosodically conditioned prominence is articulated in Italian and to understand how the prosodically-induced kinematic variations can be accounted for by parameter settings in a mass-spring gestural model.

2. METHOD

We have analyzed the kinematics of the lip opening and closing movements occurring in CV<u>CV(C)</u>CV nonce-words (where C = [b, m]; V= [a, i]), produced by two female speakers of the Tuscan and Northern variety of Italian, in medial position of a declarative sentense at a normal and fast tempo. Of the 8 repetitions recorded, we analysed 6-7 tokens of each syllable type for speaker SG and 5-6 for speaker CA. The penultimate syllable of the nonce-words can be unstressed, stressed or nuclearly accented in a contrastively focused constituent. Sentences have been embedded in short dialogues in order to elicit the intended focus interpretation.

Articulatory data have been collected by a fully automatic optotracking movement analyzer for 3D kinematics data acquisition (ELITE), which also allows a synchronous recording of the acoustic signal. The data we report on here are relative to Lip Aperture (the distance between the markers attached to the central point of the vermillion border of upper and lower lip). The pitch contour of each sentence has been ToBI transcribed in order to check for a uniform realization of the accented syllable: both speakers always realized the accented syllable with a L+H* pitch accent [9]. The results concern CV syllables with V=[a], uttered at a normal tempo. Statistical evaluation of the systematic influence of prominence levels is based on one-way ANOVAs, separately computed for each speaker. Several dependent articulatory variables have been calculated at or between the moments of onset, target and peak velocity of the opening gesture (from the maximum closure for C to the maximum opening for V) and of the closing gesture (from the maximum opening for V to the maximum closing for the postvocalic C). The variables include: movement duration, displacement, peak velocity, Time-to-Peak velocity (duration of the movement accereleration phase). Stiffness of each gesture is calculated as the ratio between peak velocity and displacement.

3. RESULTS

3.1. Acoustic durations

For both speakers, the acoustic duration of both the target syllable and of the vowel is significantly different across the different levels of stress (syllable: SG: $F(2,15)=31,1$ $p<0,0001$; CA: $F(2,10)=26,1$ $p=0,0001$. Vowel: SG: $F(2,15)=21,5$ $p<0,0001$; CA: $F(2,10)=29,3$ $p<0,0001$). Fisher’s post hoc test confirms longer durations of syllables and vowels for progressively higher levels of stress: U(stressed) < S(tressed) < A(ccented).

3.2. Kinematic results

The opening gesture into the vowel and the closing gesture out of it generally show progressively longer duration, bigger displacement and faster velocity as the prominence level progresses from no stress to lexical stress to nuclear accent. Results for kinematic and dynamic variables are summarized in Table 1.

Duration. Both opening and closing gestures of the two speakers show significant differences in duration (opening gesture: SG: $F(2,15)=22.1$ $p<0.0001$; CA: $F(2,10)=23.3$ $p<0.05$. Closing gesture: SG: $F(2,15)=19.8$ $p<0.0001$; CA: $F(2,10)=15.8$ $p<0.05$). All paired comparisons confirm a significant steady increase in duration (with the only exception of the duration of unstressed vs stressed closing gesture of CA).

Displacement. Gestures of both speakers significantly differ in displacement at all prominence levels (opening gesture: (SG: $F(2,15)=22.9$ $p<0.0001$; CA: $F(2,10)=31.5$ $p<0.0001$). Closing gesture: SG: $F(2,15)=41.7$ $p<0.0001$; CA: $F(2,10)=31.5$ $p<0.0001$). Paired comparisons confirm a significant progressive trend, simmetrically for both gestures: U< S< A. Separate analyses on the targeted endpoint of the movement reveal that the larger displacement reflects a significantly greater vocal tract openess due to a lower position in the vowel (and not a higher position in the consonant).

Peak velocity. Data on the maximum velocity reached by the opening and closing gestures are significantly different (opening gesture: SG ($F(2,15)=10.1$ $p<0.05$); CA ($F(2,10)=15.3$ $p<0.05$). Closing gesture: SG: ($F(2,15)=17.4$ $p=0.0001$; CA $F(2,10)=13.3$ $p<0.05$).

The numerical values show a progressive trend in the expected direction (U<S<A), but the post hoc test reveals that the difference is significant, for both gestures, only at the extreme degrees of prominence: unstressed vs accented. Stressed gestures are not significantly longer than unstressed ones for SG, and are not significantly shorter than accented ones for CA.

Time-to-Peak velocity and stiffness. The time from the movement onset to its peak velocity (TTP) is crucial in a mass-spring gestural model because gestures with lower stiffness will have later occurring peak velocities. Results on TTP can then illuminate the dynamics underlying the different duration patterns shown by syllables with varying degrees of prominence.

There is a significant difference in TTP of the opening and closing gestures for both speakers (opening: SG: $F(2,15)=44.1$ $p<0.0001$; CA: $F(2,10)=20$ $p<0.05$. Closing: SG: $F(2,15)=11.7$ $p<0.05$; CA: $F(2,10)=25.1$ $p=0.0001$). As shown in the upper part of Fig. 1, TTP values are progressively longer as prominence level increases. All paired comparisons are significant only for the closing gesture of SG, though. Stressed vs accented gestures do not differ in the opening gesture of both speakers and unstressed vs stressed gesture do not differ in the closing gesture of CA.

For SG, stiffness data (Fig. 1, lower part) display a perfect inverse relationship with TTP data, with unstressed syllables showing the highest values, and a significant regression from the lowest to the highest prominence that is symmetrical in the opening and the closing gesture (SG, opening: $F(2,15)=44.2$ $p<0.0001$; closing: $F(2,15)=43$ $p<0.0001$). As for CA, while stiffness values are globally different and show the same downtrend (see Fig. 1, lower part) (CA, opening: $F(2,10)=28$ $p<0.0001$; closing: $F(2,10)=17.8$ $p<0.05$), a significant difference in post hoc pairwise comparisons is found only for the extreme degrees of prominence. Unstressed gestures are stiffer than accented gestures, but stressed opening gestures pattern with accented gestures and stressed closing gestures with unstressed ones.

Figure 1: TTP (upper part: time-to-peak velocity) and stiffness (lower part) of the opening and closing gesture for speaker SG and CA.

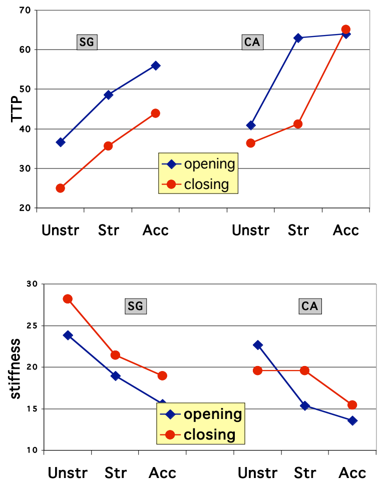

Table 1: Summary of prominence effect on the kinematic and dynamic variables for the two speakers. U= unstressed, S= stressed, A= accented syllables.

variable	opening		closing	
	SG	CA	SG	CA
Dur.	U<S<A	U <S<A	U <S<A	{U,S}<A
Displ.	U<S<A	U <S<A	U <S<A	U <S<A
P.Vel.	{U,S}<A	U<{S,A}	{U,S}<A	U<{S,A}
TTP	U<{S,A}	U<{S,A}	U <S<A	{U,S}<A
Stiffn.	U>S>A	U>{S,A}	U>S>A	{U,S}>A

4. DISCUSSION

Can we derive the set of prominence-driven acoustic and kinematic variations we observed in our data from the dynamical parameters of a mass-spring gestural model? And, if yes, can we single out a unique parameter setting as the mechanism that underlies the production of prominent syllables? In answering these questions we will interpret the data of each speaker separately.

For speaker SG, compared to stressed, unstressed syllables show shorter acoustic and articulatory duration, smaller displacement, equal peak velocity, shorter TTP and higher stiffness for both opening and closing gestures. All these properties are predicted in a dynamic model by a proportional change of the gesture's target and of its stiffnes. The same properties displayed by the opening and the closing gesture at the lower level of the prosodic hierarchy characterize the constrast of syllables prominent at the next higher level: stressed vs accented syllables (TTP is numerically, even if not significantly, lower in the opening gesture of stressed syllables). Setting progressively higher values for target and smaller values for stiffness gives rise to syllables with progressivelly higher prominence. A single dynamical mechanism appears to control the production of their opening and closing gestures: a linear rescaling that affects the target and the duration of the gesture. For speaker CA, the dynamics underlying opening and closing gestures seems to be different at the different levels of prominence. The opening gesture of unstressed syllables, compared to stressed one, shows the same set of kinematic properties that are at work for speaker SG and that can be derived positing a linear rescaling as the underlying dynamical mechanism. For the closing gesture, equal duration, less displacement, lower peak velocity and equal stiffness can all be derived from a pure change in target values, i.e. from the specification of a lower value for the underlying amplitude of the gesture. As for the stressed vs accented contrast, less duration and displacement but equal TTP and peak velocity indicate a change in the relative phasing for the gestures, with a specification of a later phase for the activation of the following closing gesture. Finally, the kinematic properties of the closing gesture are compatible with a linear rescaling.

5. CONCLUSION

Results show that different kinematic properties systematically distinguish unstressed, stressed and accented syllables. Opening and closing gestures of syllables with higher prominence are progressively longer, faster and more displaced. These properties, together with the longer acoustic duration that distinguish less from more prominent syllables and vowels, are compatibile with a strategy that enhances the intrinsic sonority of a segment, and can be accounted for by specific dynamical parameter settings in a mass-spring gestural model. For one of our two speakers a single mechamism underlies the production of the opening and closing gestures of syllables with progressively higher degrees of prominence: linear rescaling, that can be viewed as a proportional change in stiffness and amplitude. Moreover, the systematicity of our results shows that the effects of prominence at the higher level of the prosodic hierarchy are not ancillary to the associated intonation pattern [2].

6. REFERENCES

[1] Beckman M., Edwards J., Fletcher M. 1992. Prosodic structure and tempo in a sonority model of articulatory dynamics. In: Docherty G. e Ladd R. (eds), *Papers in Laboratory Phonology II. Segment, gestrure and tone*, Cambridge, U.K., Cambridge University Press, 68-86.

[2] Beckman, M., Edwards, J., 1994. Articulatory evidence for differentiating stress categories. In: Keating P.A. (ed*),* *Papers in Laboratory Phonology III: Phonological Structure and Phonetic Form,* Cambridge, U.K., Cambridge University Press, 7-33.

[3] Browman, C. P., Goldstein, L. 1995. Dynamics and Articulatory Phonology. In Port R., Van Gelder T. (eds), *Mind in Motion: Explorations in the Dynamics of Cognition*, Cambridge, MA: The MIT Press, 175-193.

[4] Cho, T., 2006. Manifestation of Prosodic Structure in Articulation: Evidence from Lip Kinematics in English. In: Goldstein L., Whalen D., Best C. (eds), *Laboratory Phonology 8*, Berlin/New York, Mouton de Gruyter, 519-548.

[5] de Jong, K.. 1995. The supraglottal articulation of prominence in English: linguistic stress as localized hyperarticulation. *J. Acoust. Soc. Am.* 97, 491-504.

[6] Farnetani, E. 1997. The phonetic word: the articulation of stress and boundaries in Italian. *Quaderni del Centro di Studio per le Ricerche di Fonetica* XVI, 157-172.

[7] Farnetani, E., Vayra, M. 1996. The role of prosody in the shaping of articulation in Italian CV syllables. In: *From Control Strategies to Acoustics. Proc. of the 1st ESCA Tutorial and Research Workshop on Speech Production Modeling*, Autrans, May 21-24 1996, 9-12.

[8] Fowler, C. 1995. Acoustic and kinematic correlates of contrastive stress accent in spoken English. In: Bell-Berti F., Raphael L. (eds), *Producing speech: contemporary issues*, New York, AIP Press, 355-373.

[9] Grice M., D'Imperio M.P., Savino M., Avesani C. 2005. Towards a strategy for ToBI labelling varieties of Italian. In: Sun-Ah Jun (ed.) *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford, Oxford University Press, 362-389

[10] Harrington, J., Fletcher, J., Roberts, C. 1995. Coarticulation and the accented/unaccented distinction: evidence from jaw movement, *J. of Phonetics* 23, 305-322.

[11] Magno Caldognetto E., Vagges, K., Zmarich, C. 1995. Visible articulatory characteristics of the Italian stressed and unstressed vowels. *Proc. 13th ICPhS,* Stockholm, 1, 366-369.

[12] Saltzman, E. 1995. Dynamics and Coordinate Systems in skilled sensorimotor activity. In: T. van Gelder e R. Port (eds)., *Mind as Motion: Explorations in the Dynamics of Cognition*, MIT Press, Cambridge, MA: 150-173.

[13] Saltzman, E., Munhall, K. G. 1989. A dynamical approach to gestural patterning in speech production. *Ecological Psychology* 1, 333-382

[14] Vayra, M., Fowler, C. 1992. Declination of supralaryngeal gestures in spoken Italian. *Phonetica*, 49, 48-60.

[15] Zmarich C., Avesani C., Marchiori M. In press. Coarticolazione e accento, *Atti del III Convegno AISV.*