An articulatory hypothesis for the alignment of tonal targets in Italian

Mariapaola D'Imperio*, **Noël Nguyen*** and **Kevin G. Munhall**†

* Laboratoire Parole et Langage, UMR 6057 CNRS
Université de Provence, FRANCE

† Departments of Psychology and Otolaryngology
Queen's University, Ontario, CANADA

dimperio@lpl.univ-aix.fr, nguyen@lpl.univ-aix.fr, munhallk@psyc.queensu.ca

ABSTRACT

Recent work on tonal alignment has suggested that the temporal location of tonal targets is determined relative to segmental "anchors" which are defined in acoustic terms. However, whether tonal targets are phased with respect to these acoustic anchors in a stable and systematic manner is controversial. According to an alternative hypothesis, the anchor points used for the alignment of tonal targets are articulatory in nature. In this perspective, the complex character of the mapping between the tonal targets and the acoustic signal would be partly attributable to articulatory-to-acoustic non-linearities. Hence, a new experimental paradigm for alignment research is adopted here in which optoelectronic data for orofacial movements are collected simultaneously with acoustic data. A corpus of Italian questions and statements was produced according to two different rates of speech, i.e. normal and fast. Preliminary results suggest that there is a tendency to synchronize specific pitch events with articulatory gestures.

1 INTRODUCTION

Recent work on the temporal alignment of tonal targets with the speech chain has suggested that specific peaks and valleys in the intonation contour are consistently aligned with well defined acoustic segmental boundaries, such as the onset or the offset of the stressed vowel, the onset of the stressed syllable, etc. Such regularities have been suggested for a number of languages. For instance [1] shows that the L target of rising prenuclear accents in Greek cooccurs with the onset of the stressed syllable. Moreover, [2] shows that prenuclear F_0 rises in English consistently align their L target with the onset of the stressed syllable and that this alignment is independent of speech rate.

Despite such regularities in the alignment of L targets, the alignment of H targets appears to be more controversial. In fact, it appears to be somehow more difficult to find specific acoustic events to which such targets might be aligned. For instance, in order to find support for a constant segmental anchoring of H targets in prenuclear English rises, [2] needs to do away with traditional segmental units or boundaries and assumes that H targets are aligned with an interval stretching from the offset of the stressed vowel to the onset of the postaccentual vowel. Moreover, in an earlier work, [3] employs the notion of syllabic "rhyme" in order to account for the alignment of H targets of American English H* accents. Hence, in these studies H targets are assumed to be aligned with acoustic events and/or intervals that do not necessarily correspond to traditional segmental boundaries and/or units.

A different view is represented by work such as [4]. In order to account for surface variability in tonal target realization, the author proposes that underlying pitch targets (which would not be the actual F_0 peaks and valleys measured in the signal) are synchronized to the entire syllable as opposed to a subsyllabic unit.

The present work stems from two observations regarding the existing alignment literature. First, both the segmental and the syllabic anchor hypotheses mentioned above inherently assume that if some anchors for tonal alignment do exist they can be extracted either directly or indirectly from the acoustic record, i.e. by measuring alignment of some F_0 event relative to an acoustic anchor. A plausible alternative would be to assume that if any anchors for tonal alignment do exist, they are primarily articulatory in nature and hence are not readily identifiable with any of the observable acoustic landmarks. This would explain why in some cases the underlying regularities would be masked.

On the other hand, all of the views presented above appear to maintain that all LH rises align in the same way to the segments/syllable. However, it has been previously observed that in the Neapolitan variety of Italian there exists an alignment contrast between question

ISBN 1-876346-48-5 © 2003 UAB

and statement pitch accent rises [5, 6]. In this variety, narrow focus statements are in fact characterized by a L+H* nuclear rise, while questions present a L*+H nuclear rise. One of the differences between the two rises is that the alignment of the H peak is later in questions than in statements.

Here we show the results of a preliminary study in which we contrasted two hypotheses. The first hypothesis was that H pitch targets of Neapolitan Italian questions would align with two acoustic landmarks usually found in the alignment literature, i.e. either the onset or the offset of the stressed vowel. The alternative hypothesis was that pitch targets would instead be synchronous with some specific articulatory dimension, i.e. either maximum or minimum between-lip distance within the stressed syllable.

In order to better evaluate the hypothesis of synchronization, two rates of speech were employed. Hence, on one side, if alignment with either acoustic or articulatory landmark is the relevant anchor point, we expect H latency to be constant irrelevant of rate of speech. The accompanying prediction is that the alignment anchors would be different for the question and the statement rise.

2 METHOD

2.1 CORPUS

The corpus consisted of a group of sentences in which modality (question or statement; QS henceforth) as well as structure of the stressed syllable were varied within the target sentence. The stressed syllable within the target words was always initial and could either be closed or open (Open/Closed, henceforth). The closed syllable word contained a labial geminate nasal (*MAMma* "mother"), while the open syllable word contained a labial singleton stop (*MApa*).

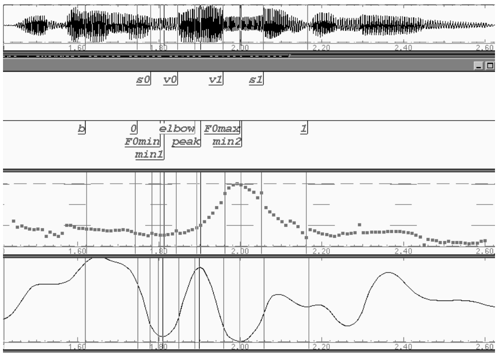

Figure 1: Waveform, labels, F_0 track and lip distance curve for the utterance "Vedrai [la MAMmma] domani?".

Target stressed syllables were always penultimate and the word was embedded as the direct object in a fixed carrier sentence *Vedrai la [...] domani* "You will see the [...] tomorrow", uttered either as a statement or as a question, such that focus scope was always narrow over the object. One Neapolitan speaker, the first author, produced each target sentence 10 times, in randomized order. The speaker read the materials at 2 self-imposed rates, normal and fast, for a total of 80 sentences.

The kinematics of markers attached to the speaker's face were tracked over time during the production of the corpus sentences by means of an OPTOTRAK.[1] Specifically, we collected data for upper and lower lip markers, together with head markers (used to correct for head motion from absolute lip movement data). The acoustic data was recorded on DAT tape and analyzed using ESPS Waves$^+$. F_0 was extracted using a 10-ms frame step using the get_f0 ESPS pitch tracker. The data were gathered at the Speech Production and Perception Laboratory of Queen's University, Canada.

2.2 MEASUREMENTS

The F_0 tracks were inspected in combination with waveform and spectrogram for each utterance. Within the rising accentual contour, the L target and the H target were marked (though here we present data only relative to the H target). In most cases, a peak value corresponding to the H target was easy to locate. Hence, this target was mostly measured at the single highest F_0 point within the accented syllable and was labeled "F0max" (see Figure 1). As shown in Figure 1, we also marked locations for the onset of the stressed syllable (s0), the onset of the stressed vowel (v0), the onset of the postaccentual vowel (s1) and the offset of the stressed vowel (v1).

As for the articulatory measures, the Euclidean distance between the three-dimensional positions of the lip markers was employed in order to calculate a global measure of "between-lip" distance. Within the lip distance curve (see Figure 1), we marked the minimum distance relative to the onset (*min1*) and the offset (*min2*) of the stressed syllable as well as the maximum distance within the stressed vowel (*peak*). These labels were automatically obtained through a peak-peaking procedure.

Duration and latency measurements included (i) duration of the portion going from s0 to s1; (ii) the temporal distance of F0max (H) from v0, as well as from v1 and (iii) the temporal distance of F0max from the articulatory measures of *peak* and *min2*. Moreover, in order to test if *peak* location would correlate with v0 location, we also measured the latency from v0 to *peak*. Three independent variables were included in the study, that is rate (RATE: fast or normal), open/closed syllable (OC: open or closed) and modality (QS: question or statement).

[1] The OPTOTRAK is an electronic three-dimensional movement tracking device.

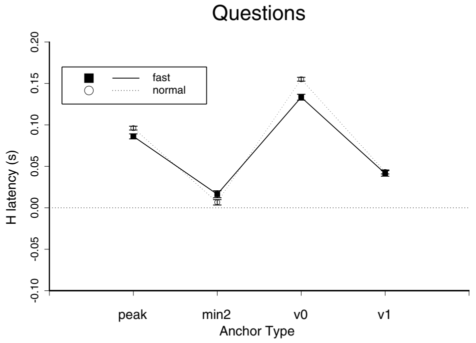

Figure 2: Alignment of H targets in questions relative to *peak*, *min2*, v0 and v1. The dotted line represents the relative location of each landmark.

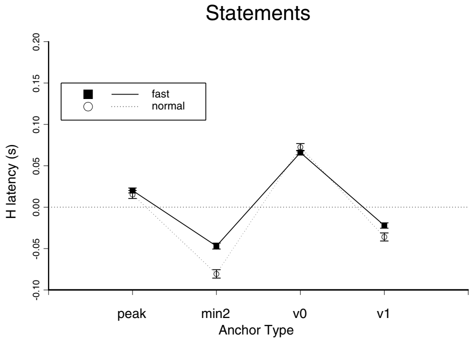

Figure 3: Alignment of H targets in questions relative to *peak*, *min2*, v0 and v1. The dotted line represents the relative location of each landmark.

3 RESULTS

3.1 ACOUSTIC LATENCIES

In order to test whether the self-imposed rate difference in the production of the recording material was reflected by the data, an ANOVA was conducted on the duration of the portion of the target word going from s0 to s1 (roughly corresponding to the stressed syllable). The results showed, as expected, a very strong effect of RATE $[F(1,72) = 574.89; p < 0.05]$.

We then went on to test the hypothesis that H peaks would be aligned with the onset of the stressed vowel (v0). The results of an ANOVA performed on the pooled results for questions and statements revealed a significant effect of RATE $[F(1,72) = 24.37; p < 0.05]$ and QS $[F(1,72) = 549.18; p < 0.05]$, while neither OC nor any of the interactions turned out to be significant. This rate difference appears though to be noticeable only for questions (see Figures 2 and 3). Hence, we performed an ANOVA on the statements data alone and found that RATE was indeed not significant $[F(1,37) = 1.66; p > 0.05]$.

RATE had no significant effect on the alignment of H targets relative to v1 $[F(1,72) = 1.73; p > 0.05]$, while QS was highly significant $[F(1,72) = 443.42; p < 0.05]$. Because of the QS effect, we tested the effect of RATE separately for questions and statements and found that the effect was indeed significant for statements alone $[F(1,37) = 6.26; p < 0.05]$.

3.2 ARTICULATORY LATENCIES

The acoustic results presented above seem to suggest that v1 might be a potential site for H alignment for questions. However, note from Figure 2 that despite the absence of a RATE effect, there is no synchronization between H target and this acoustic landmark. From qualitative observation of the articulatory data we had noticed instead a high correlation between the location of the H target in questions and the location of *min2* (see for instance Figure 1). As we can see from Figure 2, H targets occurred on average around 10.5 ms after *min2*. The results of an ANOVA showed in fact no significant effect of RATE $[F(1,37) = 1.47; p > 0.05]$.

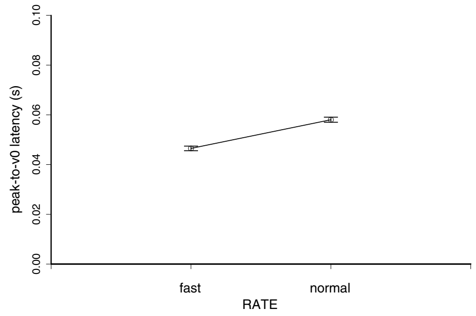

Figure 4: Mean and standard error for *peak* to v0 distance in the two rates.

As for statements, we saw that the possible alignment site for H targets is v0. This was indeed suggested in previous work [5]. Again, though, the data do not reveal a clear tendency for the tonal event to be phased with the acoustic event (see Figure 3). On the other hand, from inspection of the articulatory data, we noticed a high correlation between the location of the H target of statements and the location of *peak*, as observable in Figure 3. In fact, the ANOVA performed on the data revealed no effect of RATE $[F(1,37) = 0.85; p > 0.05]$.

At a first glance, one could think that this regularity might simply stem from the occurrence of the H target at a fixed distance from the left edge of the stressed vowel. In such a case, if *peak* occurs always at the same distance from vowel onset, then we could not distinguish between an alignment of H with v0 or *peak*. Note though from Figure 4 that this was not the case: *peak* was in fact later at the normal rate of speech, which was confirmed by an ANOVA $[F(1,76) = 68.56; p < 0.05]$.

4 DISCUSSION

Summarizing the results, H targets of nuclear rises in Neapolitan statements and questions appear to be more closely phased with the articulatory dimension of between-lip distance than with two of the most commonly employed acoustic segmental landmarks for tonal alignment (i.e., onset and offset of stressed vowel). Specifically, for questions, it appears that the H target is not phased with vowel offset, though its latency relative to this landmark appears to be independent of speech rate. On the other hand, this target seems to be phased with the gesture associated with minimum between-lip distance at the end of the stressed syllable.

Likewise, though a certain stability of alignment was found when the alignment of the H target of statements was measured relative to stressed vowel onset, these two events cannot be claimed to be synchronous. This tonal target appeared instead to be phased with peak between-lip distance. Note that this location does not correspond to any identifiable segmental boundary or phonological unit. It is possible, though, that this location might correlate with the p-center dimension [7], whose role for tonal alignment has received much less attention in the literature.

One might argue that no exact synchronization was found even in relation to the articulatory dimensions under study. However, we might account for this discrepancy by invoking the mechanical response of the tonal production system. Due to factors such as inertia of the articulators, time lag to activate the muscles involved, etc., surface realization might never allow precise synchronization between any two targets.

One of the issues in the current investigation was also the different alignment behavior of question and statement pitch accents in Neapolitan. The preliminary results presented here seem to point to some kind of fine alignment specification for the H target. Specifically, at the level of phonological specification we can hypothesize that H target commands of Neapolitan rises are phased with commands of the supralaryngeal articulator involved to produce the segments to which the accent is associated. This phasing relation might be represented in a way similar to phasing between laryngeal and oral gestures in segmental production. Above all, we take these results to suggest that not all rises align in the same way with the associated syllable.

Hence, we believe that though the role of articulatory constraints is important, the exact phasing properties of different prosodic events is specific to the phonology of the language under investigation. Note that prosody has recently become the realm of investigation of the Task Dynamics program [8]. Our future research will focus on the dynamics of pitch accent alignment under an articulatory perspective.

5 CONCLUSION

This work has attempted a new chracterization of tonal target alignment based on articulatory rather than strictly acoustic anchor points. The preliminary results suggest that H targets of LH nuclear rises in Neapolitan might be phased with either peak or minimum between-lip distance for labial consonant production. More data and more segmental as well as prosodic contexts are needed in order to determine whether phasing of the laryngeal and the supralaryngeal system might be revealed through direct articulatory measures.

ACKNOWLEDGMENTS

Thanks to Michel Pitermann for help setting up the experiment and processing the articulatory data. This work was partly funded through an Ameritech Fellowship from the Ohio State University and ACI grant 0220244, *Ministère de la Recherche* (France).

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 f_0 movements under changes in speech rate," *The Journal of the Acoustical Society of America*, vol. 106, no. 3, pp. 1543–1554, 1999.

[3] J. P. H. van Santen and J. Hirschberg, "Segmental effects on timing and height of pitch contours," in *Proceedings of the International Conference on Spoken Language Processing*, Yokohama, Japan, 1994, vol. 2, pp. 719–722.

[4] Y. Xu, "Articulatory constraints on tonal alignment," in *Proceedings of SP2002*, Aix-en-Provence, France, 2002.

[5] M. D'Imperio, *The role of Perception in Defining Tonal Targets and their Alignment*, Ph.D. thesis, The Ohio State University, 2000.

[6] M. D'Imperio, "Focus and tonal structure in Neapolitan Italian," *Speech Communication*, vol. 33, no. 4, pp. 339–356, 2001.

[7] B. Pompino-Marshall, "On the psychoacoustic nature of the P-center phenomenon," *Journal of Phonetics*, vol. 17, pp. 175–192, 1989.

[8] D. Byrd and E. Saltzmann, "The elastic phrase: modeling the dynamics of boundary-adjacent lengthening," *Journal of Phonetics*, in press.