Are tones aligned with articulatory events? Evidence from Italian and French

Mariapaola d'Imperio, Robert Espesser, Hélène Loevenbruck, Caroline Menezes, Noël Nguyen, Pauline Welby

► To cite this version:

Mariapaola d'Imperio, Robert Espesser, Hélène Loevenbruck, Caroline Menezes, Noël Nguyen, et al.. Are tones aligned with articulatory events? Evidence from Italian and French. Cole, Jennifer. Papers in Laboratory Phonology 9, Mouton de Gruyter, pp.577-608, 2007. hal-00244489

HAL Id: hal-00244489
https://hal.science/hal-00244489

Submitted on 7 Feb 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Chapter X
Are tones aligned to articulatory events? Evidence from Italian and French

1. Introduction

Tonal alignment work has suggested that the temporal location of tonal targets relative to segmental "anchors" might be governed by principles of synchrony and stability (Arvaniti, Ladd, and Mennen 1998; Ladd *et al.* 1999; *inter alia*). However, a number of discrepancies have emerged in the cross-linguistic study of alignment. For instance, despite some regularities in the alignment of L(ow) targets with segmental structure (Caspers and van Heuven 1993; Prieto, van Santen, and Hirschberg 1995; Arvaniti, Ladd, and Mennen 1998), the alignment of H(igh) targets appears to be quite controversial. While L tones of rising LH prenuclear pitch accents appear to be anchored to the left edge of the stressed syllable in a variety of stress-accent languages, the H of the same rising prenuclear accents can be variably aligned with either the middle of the associated stressed syllable, the end (i.e., the right edge) or even with the postaccentual syllable (for instance in prenuclear rises of Greek, cf. Arvaniti, Ladd, and Mennen 1998).

A specific contrast we will concentrate upon in this paper concerns the alignment contrast found between yes/no question and statement pitch accent LH rises in Neapolitan Italian (D'Imperio 2000, 2002). In this variety, as Figure 1 shows, both accents are characterized by a salient rise to a peak, with the H target realized earlier within the stressed syllable (around the middle of the vowel nucleus) for the statement L+H* than for the question L*+H. This alignment difference is also employed in identification (D'Imperio and House 1997; D'Imperio 2000). The relationship between the alignment properties of these accents and the notion of association is detailed in Prieto, D'Imperio and Gili Fivela (to appear).

[insert Figure 1 here]

The variable alignment of tonal targets has been partly accounted for by factors leading to the surface variability of underlying tones (Silverman and Pierrehumbert 1990; Xu 2002). For instance, gestural overlap can account for the earlier alignment of H peaks when immediately followed by another tonal event ("tonal crowding"), such as a phrase accent (for H targets of nuclear accents). However, there is currently no production model of the alignment of tones with segmental structure, though there are some proposals based on articulatory constraints. Xu and colleagues (Xu 1998, 2002; Xu and Wang 2001) claim that the phasing of tones and syllables is constrained by the speed at which speech is produced, as well as by general motor skills and coordination patterns (following Kelso 1984). These constraints permit (underlying) tones and syllables to be phased only at a 0º angle with each other. What this basically means is that, in a rising tone, the implementation of the LH target movement would be synchronized with the first oral gesture of the "host syllable." The implementation of the L target, according to Xu, would then be phased at the onset of the stressed syllable (i.e., at a 0º angle) and the H target would likewise be phased with the offset of the same syllable.

Note that this claim is partly based on Kelso's work on bimanual coordination, as well as on Xu and colleagues' findings on the speed of pitch change in lab speech for Mandarin Chinese (Xu 1998; Xu and Sun 2002). However, Kelso's original conclusions were twofold: first, in cycling finger movements, only two stable phase states (attractors) exist between the hands, namely symmetrical in-phase (0º angle) and antisymmetrical out-of-phase (180º angle). Secondly, at critical cycling frequency, an abrupt transition from the antisymmetrical attractor state to the in-phase state is observed. However the fact that only two stable states exist does not exclude other (non-stable) states, even at high cycling rates. In addition, the notion that the coordination of laryngeal and supralaryngeal movements would strictly obey principles observed with respect to the hand or wrist is controversial. Simply concerning the coordination of supralaryngeal articulators, Kelso and colleagues themselves have shown that the relative phasing of jaw vowel gestures and upper lip consonant gestures is not stable across manipulations in linguistic and nonlinguistic factors, including speaking rate, stress patterns, syllable structures and consonant identity (e.g., Nittrouer *et al.* 1988). Therefore it does not seem obvious to us that the synchronization between the implementation of pitch targets and their host syllables is as definite as claimed by Xu (2002). Furthermore, not all tones are associated to syllables. For example, there is a great deal of variability in the realization of the H of the French early rise, and few researchers maintain that it is associated to a syllable (see Welby 2003, 2004).

Speech rate has been also taken as a possible testing ground for the segmental anchoring hypothesis of tonal events. For instance, Ladd *et al.* (1999) show that prenuclear LH rises of English consistently align the L target with the onset of the stressed syllable (hence, a segmental/syllabic boundary) and that this alignment is independent of speech rate. However, in order to show stable alignment of H targets, the authors need to employ an alignment measure that is not relative to a segmental boundary but relative to a time interval (i.e., the interval stretching from the offset of the stressed vowel to the onset of the postaccentual vowel).

In sum, it is often difficult to find definite segmental landmarks to which tonal targets might be aligned. Moreover, most of the alignment proposals so far inherently assume that if some anchors for tonal alignment do exist they must be acoustic in nature. A plausible alternative would be to assume that some anchors are primarily articulatory in nature. This view would explain why in some cases no regularities are apparent in the acoustic signal. Interestingly, Byrd and Saltzman (2003) propose a theory of prosodic gestures that attempts to phase together segmental and suprasegmental structures in terms of their temporal and coordination properties within the Task Dynamics framework. While their prosodic gesture deals with rhythm *per se*, is it possible for tonal gestures to be phased to oral articulatory gestures? If yes, how?

To explore this possibility, we adopt a new experimental paradigm for alignment research in which articulatory measures are performed simultaneously with acoustic measures. Here we show the preliminary results of a larger study in which alignment to acoustic as well as articulatory events is compared for two Romance languages, Italian and French. The strong hypothesis tested here is that tonal targets are synchronized with articulatory events linked to consonantal production.[1] In other words, a stable, albeit large, temporal alignment of a tonal and a segmental event would not constitute sufficient evidence for synchronization of the laryngeal and the supralaryngeal system. In other words, we believe that in order for the synchronization hypothesis to hold, the main criterion is that the temporal lag between the tone and the alignment point (be it acoustic or articulatory) should be small. As a second criterion, we will also hypothesize that articulatory measures will be more stable than acoustic ones, though this will be further discussed below.

Most of the recent work on tonal alignment has mainly employed stress accent languages as testing ground, while the alignment characteristics of languages without lexically contrastive stress, such as French, are much less studied. In the present study, we will therefore include some preliminary data for French. We will focus on the L tone of the LH early rise that often occurs near the beginning of the accentual phrase. Although there is a fairly stable alignment of the L with respect to a segmental landmark – the tone is realized in the region of the function word-content word boundary – there is still a fair amount of variability. In order to further explore this alignment from a global production standpoint, we conducted a study to examine the potential relationship of this L tone to articulatory parameters.

For Neapolitan Italian, we specifically contrasted two hypotheses: (a) H pitch targets of questions and statements align with two acoustic landmarks usually found in the alignment literature, i.e., either the onset or the offset of the stressed vowel; (b) H targets are synchronous with some specific articulatory dimension related to the primary articulator for the offset of the stressed syllable. Moreover, to better evaluate the hypothesis of synchronization, two rates of speech were employed. If alignment with either acoustic or articulatory landmarks is the relevant anchor point, we expect H latency values to be stable regardless of rate of speech, as suggested by Ladd *et al.* (1999) and given the predictions of Xu (2002). Hence, our goal is to take a closer look at the timing between articulatory and F_0 events, which is a necessary step before any conclusion about relative phasing is drawn. The accompanying prediction is that the alignment anchors would vary for the question and the statement rise.

For French, we hypothesize that the L tone of the initial rise would be better aligned relative to articulatory events specific to the onset of the first syllable of the first content word in the accentual phrase than to articulatory events for the preceding syllable. In *le mélomane* 'the music lover', for example, we expect the L to be aligned with respect to the first syllable of the content word *mélomane* ([me]) rather than that of the function word *le* (see Figure 8). This would strengthen the hypothesis (Welby 2002, 2003, 2004) that this tone is actually associated with the left edge of the first content word of the accentual phrase, and not with the immediately preceding function word.

2. Tonal alignment in Neapolitan Italian

A preliminary study (D'Imperio, Nguyen, and Munhall 2003) was conducted in which various latency measures, both acoustic and articulatory, were analyzed for 80 utterances produced by one speaker (the first author). Specifically, the kinematics of OPTOTRAK[2] markers attached to the speaker's upper and lower lip was tracked over time during the production of the corpus sentences. The melodic target considered was the H tone of LH nuclear rises in Neapolitan Italian, both for questions and statements (see Figure 1).

In order to test the hypothesis of stable anchoring of H targets, the materials were produced with two self-paced rates of speech, normal and fast. Summarizing the results, H targets of nuclear rises in Neapolitan statements and questions appeared to be more closely phased with the articulatory dimension of the Euclidean distance of the lips (henceforth LA, “Lip Aperture”), for target syllables starting with labial consonants, than with two of the most commonly employed acoustic segmental landmarks for tonal alignment (i.e., onset and offset of stressed vowel). Specifically, the H peaks of statements appeared to be phased with maximum LA within the stressed syllable. Note that this location does not correspond to any identifiable segmental boundary, acoustic event or phonological unit, and does not overlap with RMS peak amplitude. In fact, RMS peaks were generally much earlier than articulatory peaks, hence further away from H peaks. On the other hand, the question H peaks appeared to be phased with a local minimum of LA (see Figure 2). While these peaks were often located around the right edge of the stressed syllable, their location could not be more precisely described with respect to segmental landmarks. These preliminary results suggested that articulatory alignment could be a fruitful avenue of research. We therefore collected and analyzed additional articulatory data (including jaw and tongue movements, two parameters as yet unexplored) to shed light on tonal alignment issues.

[insert Figure 2 here]

2.1. Corpus

In a follow-up study, we collected a Neapolitan Italian corpus consisting of real words inserted in the carrier sentences given in (1). The two carriers differ only in that (1a) is a statement, while (1b) is a question.

(1) (a) *Vedrai il/la X domani.*
‘You will see the X tomorrow.’

(b) *Vedrai il/la X domani?*
‘Will you see the X tomorrow?’

All utterances contained bisyllabic target words with stress on the first syllable, which were produced with narrow focus. The target words varied in the structure of the stressed syllable: they were either of the form 'NV.CV or 'NVN.CV. The stressed (first) syllable always had a bilabial or apical nasal onset (e.g., *mamma* ['mam.ma] ‘mom’, *nonno* ['non.no] ‘grandfather’). This syllable was either open or closed, and coda consonants were bilabial or apical nasals. All open syllable words contained the labial or apical nasal in onset position (e.g., *Mapa* ['ma.pa] (proper name), *nono* ['no.no] ‘ninth’). Coda consonants were always followed by homorganic onset consonants (e.g., *mamma* ['mam.ma] ‘mom’, *Nando* ['nan.do] (proper name)). There were 10 repetitions of the target utterances for both modalities (question and statement), randomized together with a set of sentences for an unrelated study (e.g., *Dico salsa lentamente* ‘Say salsa slowly’). The stimuli were read by two speakers of Neapolitan Italian (one of whom was the first author) at two self-paced rates, normal and fast, for a total of 320 sentences each. Here we will show the results of only one of the speakers.

2.2. Method

Articulatory and acoustic data were recorded simultaneously and were time-synchronized. Articulatory movements were recorded using the electromagnetic articulograph system (Carstens AG-100, 2-D EMA system) at a sampling rate of 500 Hz. Receivers were placed on the upper and lower lips (at the vermilion border), the lower jaw (at the root of the incisors), and on four positions on the tongue surface, from approximately 9 mm to 50 mm posterior, corresponding to the apex, blade, body and back. Two receivers placed on the nose and on the upper incisor served as references. All data were subsequently corrected for head movements and rotated such that the measured occlusal plane coincided with the x-axis. The acoustic signal was simultaneously recorded[3] at a sampling rate of 44.1 kHz using a DAT recorder and a microphone mounted on the headpiece worn by participants. The acoustic data were transferred to computer using a digital-to-digital cable and then downsampled to 20 kHz. When necessary, to compensate for a drift due to the microphone, the acoustic signal was filtered using a linear-phase high-pass filter (obtained with the minimax method, Remez algorithm). The recordings were conducted at the Institut de la Communication Parlée, Grenoble, France.

All articulatory analyses were conducted using the Unix software program *Motif Édition Signal* (MES, Espesser 1996)[4], which displays the acoustic waveform, the fundamental frequency contour, the articulatory trajectories, and the linear velocities of the smoothed trajectories, and also allows the use of multilayered labels. Before plotting the articulatory trajectories, the articulatory data were further smoothed by a low pass filter (25 Hz). Acoustic measurements were made using the Praat speech analysis software (Boersma and Weenink 2005). Praat scripts were written to semi-automate and therefore speed the labelling process. For each utterance, word and syllable boundaries were tagged, using waveforms and spectrograms to guide the segmentation. The following segmental landmarks were also marked: onset of the stressed syllable (s1), onset of the stressed vowel (v1), offset of the stressed vowel (v2), onset of the postaccentual vowel (v3). The peak of the nuclear pitch accent (the H of the LH sequences) was automatically detected and labelled.

Displacement measures of the maxima and minima of crucial articulatory trajectories involved with the stressed syllable, as well as their peak vertical velocities were estimated. We also calculated the Lip Aperture (LA) value (in order to compare the current data with those of D'Imperio, Nguyen, and Munhall 2003). The labels were automatically obtained through a peak-picking procedure. Temporal measurements included (a) duration of stressed syllable; (b) duration of the target word; (c) latency of the F_0 peak (H) from v1, as well as from v2 and (d) latency of the F_0 peak from the selected maxima, minima and peak velocities of the articulatory trajectories. Three independent variables were included in the study: rate (RATE: fast or normal), open/closed syllable (OC: open or closed) and place of articulation (ART: apical or labial).

2.3. Results

2.3.1. Acoustic alignment

In order to test whether the self-imposed rate difference in the production of the recording material was reflected by the data, a one-way ANOVA was conducted on the duration of the target word with RATE as a main factor.

The results showed, as expected, an effect of rate [$F(1,317) = 95.278$; $p<0.05$].[5]

To test the hypothesis that H peaks are aligned with the onset of the stressed vowel (v1), ANOVAs were performed separately for questions and statements (given that their relative pitch accents might be aligned to different points within the syllable). Figure 3 displays the latency of H peaks from v1 (upper panels) and v2 (lower panels) in boxplots[6], separately for statements (left panels) and questions (right panels). In all the following boxplots, the 0 on the y-axis (marked by a dashed line) corresponds to the temporal location of a relevant acoustic or articulatory event. In the left panel (statements) we can observe the absence of an effect of rate on the alignment of H peaks relative to v1 [$F(1,151) = 1.8$; $p = 0.18$]. In the same panel we also see that the H tone occurs around 81 ms. after v1, within the stressed syllable. Hence, despite their stability with respect to tonal anchoring, we cannot consider H targets of statements to be "synchronized" with the stressed vowel onset.

For questions, we can notice from Figure 4 (upper, right panel) that the rate effect for the alignment of H to v1 is indeed significant [$F(1,152) = 20.1$; $p<0.05$]. Moreover, since the H peak is realized later in questions than in statements, the distance between the tone target and the onset of the stressed vowel is even larger (closed syllable, question, fast: 139 ms.; closed syllable, question, normal: 146 ms.; open syllable, question, fast: 147 ms.; open syllable, question, normal: 162 ms.).

Another possible acoustic alignment point, especially for questions, is v2, i.e., the offset of the stressed vowel. In fact, when v2 was considered as a possible alignment point, rate was highly significant only for statements [$F(1,155) = 29.23$; $p<0.05$], but not for questions [$F(1,156) = 1.74$; $p=0.14$]. Despite the stability of tonal alignment relative to v2 in questions, which would suggest an anchoring relative to the offset of the stressed vowel, there was no exact synchronization between the H tone and the acoustic event. This is clear in Figure 4 (lower, right), which shows a mean latency of 50 ms. between H and v2 (closed syllable, question, fast: 53 ms.; closed syllable, question, normal: 47 ms.; open syllable, question, fast: 51 ms.; open syllable, question, normal: 49 ms.).

[insert Figure 3 here]

2.3.2. Articulatory alignment

From the acoustic results for statements we might infer that the stressed vowel onset (v1) is a stable site for H anchoring, given the absence of a rate effect. However, we also found that the actual latency value between the H target and this acoustic landmark was quite large. This result replicates the finding of D'Imperio, Nguyen, and Munhall (2003). In this earlier work we had also found that location of the H target of statements could be more precisely characterized relative to the articulatory dimension of the maximum Euclidean distance (Lip Aperture, LA) for labial onset consonants. Also, question H peaks appeared to coincide with a local minimum for LA, somewhere around the right edge of the acoustic stressed syllable. While the earlier study provided evidence for articulatory alignment only for labial consonants, the present study examines whether this pattern also holds for syllables with apical consonants. In the current study, we first analyzed data for labial consonant targets in relation to the between-distance measure of the lips (MaxLA = maximum Lip Aperture; MinLA = minimum Lip Aperture) in order to make comparisons with the preliminary study (D'Imperio, Nguyen, and Munhall 2003).

[insert Figure 4 here]

As we can see from Figure 4, H targets of statements occurred on average around 32 ms. after maximum LA. However, contrary to the results reported in D'Imperio, Nguyen, and Munhall (2003), the results of an ANOVA showed a significant effect of rate [$F(1,75) = 10.46$; $p<0.05$].

[insert Figure 5 here]

Since maximum LA might be thought of co-occurring with an energy maximum due to the output of the lips when maximally open, we also calculated the latency of H peaks from RMS peak amplitude. As we can see in Figure 5, there was no near-synchronization of H targets with RMS peak amplitude location. In fact, statement H targets were realized on average 52 ms. after this reference point (closed syllable, statement, fast: 48.5 ms.; closed syllable, statement, normal: 53 ms.; open syllable, statement, fast: 57 ms.; open syllable, statement, normal: 48 ms.). Moreover, we also found a great deal of variability, as shown in the boxplots. On the other hand, an

ANOVA found no rate difference for this measure [$F(1,75) = 0.15$; $p = 0.7$].

Next we investigated the role of the vertical trajectories of the primary articulators for apicals and labials. Informal observations showed that the tone appeared to be synchronized around the region of the peak velocity for the target trajectory in statements. Peak velocity is a crucial articulatory parameter. Simple (unimodal) articulatory movements have been shown to feature two phases: an acceleration phase, in which velocity increases, followed by a deceleration phase, in which velocity decreases. These two phases are delimited by the time of peak velocity. A remarkable observation about human limb, jaw or tongue motion is that the relationship between movement peak velocity and movement amplitude is stable and linear (see e.g. Cooke, 1980; Nelson, 1983; Ostry, Keller & Parush, 1983; Kelso et al., 1985; Perrier, Abry & Keller, 1989). Interestingly, in speech movements, prosodic changes (such as manipulations of rate) have been shown to influence the slope of this relationship.

Some investigators have tried to relate peak velocity to physical entities and to model the linear peak velocity-amplitude relationship in terms of second-order mechanical systems with parameters such as mass, stiffness and viscosity. In undamped second-order (mass-spring) modeling of articulatory movement, the ratio of peak velocity over displacement (Vmax/disp hereafter) is proportional to the square root of the normalized stiffness (i.e. of the stiffness divided by the mass). An increased slope of the linear peak velocity-displacement relationship can be modeled by an increase in stiffness. According to Cooke (1980) "This observation accords with the common experience of tensing or co-contracting in the expectation of performing a very rapid movement."

Although the adequacy of the equivalence between stiffness and the Vmax/disp ratio in speech movements is controverted since it only holds for undamped systems and untruncated movements (see e.g. Lœvenbruck, 1996; Byrd and Saltzman, 1998), a relationship between peak velocity and effort has been shown to exist. Nelson (1983) describes a set of physical costs associated with accomplishing a skilled movement. The impulse cost is defined as proportional to the total impulse (i.e. the time integral of the force magnitude) occurring during the movement. Nelson showed that "it equals the peak velocity, V, for all movements where the frictional forces are negligibly small compared to the applied forces, and where the velocity patterns are unimodal (have a single velocity peak). For all other cases, the peak velocities are always somewhat greater than this effort-cost."

According to Nelson, peak velocity is therefore related to the amount of effort during a movement. For more recent results on the notion of effort in speech see also Perkell *et al.* (2002).

Some researchers have tried to relate the articulatory effects of prosodic changes (such as accent or boundary effects) to changes in the parameters of second order articulatory models. In that framework, accented speech movements have sometimes been associated with a decrease in stiffness, defined as the Vmax/disp ratio (e.g. Ostry *et al.*, 1983). It has been shown however that the relationship between accent and the Vmax/disp ratio is far from that simple and that several parameters, including timing, seem to interplay in accent production (e.g. Lœvenbruck, 1996; Cho, in press). Furthermore, some researchers have suggested that absolute time from movement onset to peak velocity is a better indicator of stiffness than the Vmax/disp ratio (Byrd and Saltzman, 1998; Byrd *et al.*, 2000). In this line, interesting data on the timing of the peak velocities of several speech articulators has been provided by Keller (1987) and Gracco (1988) *inter alia*.

The time at which peak velocity occurs can therefore be considered as a likely candidate among the possible centrally-controlled articulatory parameters in prosody. As such it could well be phased with the time at which a crucial laryngeal target (such as the peak H tone) occurs. The latency between peak velocity and peak F_0 value was therefore examined.

Figure 6 shows H latencies from the maximum velocity of the y trajectory of the lower lip for labial onsets (upper left) and maximum velocity of the y trajectory of the tongue apex for apicals (upper right) for statements. For questions, we considered the peak value of the specific displacement trajectory (either lower lip y or tongue apex y), as shown in the lower panels of Figure 6. This peak value in the displacement trajectory corresponds to a null value for velocity.

[insert Figure 6 here]

In Figure 6, the timing of the H peak on average tends to converge around zero in the faster rate, except for fast, open syllable questions and labial articulation. The figure shows that the latency of the H peak to peak velocity for statements converged around zero with closer alignment for fast rate of speech. In fact, the rate effect was significant for both statements containing labials [$F(1,75) = 18.25$; $p<0.05$] and those containing apicals [$F(1,75) = 20.98$; $p<0.05$], though the differences were quite small, except for normal/open items (Apicals: fast, closed = 4 ms.; fast, open -13 ms.; normal, closed = -19; normal, open = -44 ms. Labials: fast, closed = 5 ms.; fast, open = -8 ms.; normal, closed = -10 ms.; normal, open = -37 ms.).

Crucially, in order to support our hypothesis that articulatory alignment values are more stable and smaller than the acoustic ones presented in the preceding section, we performed two types of analysis. First, we carried out variance tests between pooled articulatory and acoustic latency data, separately for questions and statements, place of articulation and syllable structure. The tests did not reveal a significant difference in variance between the acoustic and the articulatory data. Hence, a set of four ANOVAs was performed with pooled acoustic and articulatory alignment data, and with SIGNAL (acoustic or articulatory) as the main factor. Specifically, we first contrasted the latency values of H relative to peak velocity for statements (articulatory alignment) vs. the latency relative to stressed vowel onset (acoustic latency), for both apical and labial items and for both syllable structures. The differences were statistically significant [labials: $F(1,154)=513$, $p<0.001$; apicals: $F(1,156)=869$, $p<0.001$]. Similar ANOVAs were performed for the labial and apical question items, taking as dependent articulatory measure the latency relative to peak trajectory (of the main constrictor) and the stressed vowel offset as the acoustic alignment point. Again, the differences were significant [labials: $F(1,156)=207$, $p<0.001$; apicals: $F(1,117)=218$, $p<0.001$]. All the results confirmed the observation that articulatory latencies are smaller than acoustic ones.

Summarizing the results, based on the data from one speaker, H targets of nuclear rises in Neapolitan statements appear to be more closely, though not more stably, synchronized with the articulatory dimension of peak velocity within the trajectory of the primary constrictor than with two of the most commonly employed acoustic segmental landmarks for tonal alignment (i.e., onset and offset of stressed vowel). Also, for questions, it appears that the H target is not synchronous with stressed vowel acoustic offset, but it appears to be synchronized with the location at which a local maximum for the relevant constrictor trajectory is reached, around the end of the stressed syllable.

3. Tonal alignment in French

French intonation is typically characterized by an F_0 rise on the last full syllable of a non-utterance-final phrase and an optional early rise occurring somewhere before the late rise (see Di Cristo 1998, and references therein). The late rise is a marker of the "primary accent" and is accompanied by syllabic lengthening (Pasdeloup 1990; Jun and Fougeron 2002, who claim that the early L "can spread over all of the clitic syllables preceding the AP initial content word"; *inter alia*). The early rise is a marker of the "secondary accent" and is accompanied by strengthening of the onset consonants (Pasdeloup 1990, Mertens *et al*. 2001). There is general agreement that French prosody includes these two rises, although there is considerable disagreement about their structure.

A minimal pair, with and without the optional early rise, is given in Figure 7. There is no difference in meaning between the two pronunciations. The first tier shows words and word boundaries and the second tier shows tones for the target accentual phrase (*le mélomane* 'the music lover'). Note that in both patterns, the early L is realized at the determiner-noun boundary, resulting in an inflection or "elbow" in the F_0 curve.

[insert figure 7 here]

Welby (2003, 2004) conducted a study of the tonal alignment of the French early and late rises with a group of seven speakers. She found that the two peripheral tones of the common two-rise (<u>L</u>HL<u>H</u>) pattern were the most stably aligned to segmental landmarks. The findings of interest for the current study involve the distribution of the L of the early rise. Welby (2003, 2004) reports that 73% of early L tones were realized within 20 ms. of the function word-content word boundary, with some tones realized shortly before the boundary (i.e., in the function word) and some shortly after (i.e., in the content word). Welby proposed a double association of the early L (2003, 2004). She argued that this L is associated to the left edge of the first content word of the phrase and to the left edge of an earlier syllable, often that of the first syllable of the accentual phrase. This proposed association is at odds with earlier accounts, which claim that the L of the early rise is associated with the end of the function word (e.g., Vaissière 1997, Jun and Fougeron 2002, *inter alia*).

The relationship between articulatory events and their acoustic consequences is not a one-to-one relationship. Taking the bilabial nasal [m] as an example, from an acoustic point of view, we typically use a drastic drop in intensity in the spectrogram (and the waveform) to mark the consonant beginning, assuming that the airflow through the oral cavity has been blocked by complete closure of the lips. However, articulatory studies show that the velum is lowered even before oral closure is reached (Vaissière 1988, Krakow 1993) and this affects the drop in intensity even before the oral closure. Since the acoustic signal is the result of many simultaneous articulatory adjustments it is worthwile to examine articulatory alignment of tonal events, in addition to acoustic alignment, to have a better understanding of the coordination between laryngeal and supralaryngeal movements.

3.1. Corpus

The French corpus consisted of sentences containing nonwords, real words, and reiterant speech (in which the speakers replaced all syllables with the syllable [ma]) in target position. The nonwords consisted of three minimal pairs of identical segment sequences, differing only in segmentation, as in (2). In (2a), [la] is the first syllable of the (pseudo) proper name *Lanamileau*, while in (2b), [la] is the definite article *la* in the phrase *la namileau* (*namileau* is nonword).

(2) a. *Lanamileau*
[la.na.mi.lo]
'Lanamileau' (proper name)

b. *la namileau*
[la.na.mi.lo]
'the namileau'

Each critical sequence was four-syllables long, with the first syllable interpretable as a determiner or a possessive adjective. The other two pairs were *Manumulat* (pseudo proper name)/ *ma numulat* 'my numulat' [ma.ny.my.la], *Nomalomue* (pseudo-proper name)/ *nos nomalomues* 'our nomalomues' [no.ma.lo.my]. The target sequences were embedded in two types of carrier sentences, either sentence-initially, in (3a), or sentence-medially, as in (3b). Each nonword item appeared 10 times in each position in the sentence list read by participants. The sentences containing nonword targets were randomized together with those containing real word targets and reiterant speech.

(3) a. *Mais ______ marmonna en râlant.* [sentence initial]
'But ______ muttered while grumbling.'

b. *Elle marmonna à ______ en râlant.* [sentence medial]
'She muttered to ______ while grumbling.'

We expected that most of the target sequences would be produced with an early rise. While the early rise is optional, it is quite common in accentual phrases being at least three syllables long and in read speech (Pasdeloup 1990, Jun and Fougeron 2000, Welby 2003, *inter alia*). Specifically, we expected that the early rise would begin at the left edge of the content word (e.g., at [la] for *Lanamileau* and at [na] for *la namileau*), in line with earlier observations. The beginning of this rise is visible as an "elbow" in the F_0 curve when not in absolute utterance-initial position.

Three native speakers of standard French spoken in metropolitan France read the list of utterances at a self-determined normal rate. Unlike the Italian study the items were recorded at one rate and rate effects were not analyzed. Only items with nonwords produced by one speaker are reported here.

3.1. Method

Within this list, a total of 22 utterances were discarded from the analyses. These items were either truncated by error in the original recording, produced without the optional early rise, produced with hesitations or disfluencies, or produced with an incorrect segmentation of the target sequence. In total 155 utterances were analyzed for French. The procedures for the articulatory and acoustic recordings were identical to those for the Italian experiment (see section 2.2).

[insert Table 1 here]

As for Italian, articulatory analyses were conducted using MES and acoustic measurements using Praat, as described in section 2.2. For the French data, Low (L) elbows of the early rise could not be tagged in a consistent manner by F_0 minimum detection or by hand, since a L target does not always correspond to a local F_0 minimum. Fundamental frequency and time information was hence first extracted for each file using a Praat script (Boersma, 2001). Then, the L position was then calculated using an R script[7] encoding a procedure described in D'Imperio (2000: 92) (see also Pierrehumbert & Beckman, 1988).

3.2. Results

3.2.1. Acoustic alignment

To determine the alignment of the L tone to the acoustic syllable onset, we calculated latency to the L tone (the early elbow) from the beginning of the acoustic syllable onset. Similarly, we calculated the latency to the L tone from the peak velocity of the primary articulator.

Figure 8 plots the latency to the L tone from the acoustic syllable onset. Positive values correspond to a L tone realized within the onset consonant of the first syllable of the content word (e.g., the [m] of *Manumulat* or the [n] of *la namileau*). Negative values correspond to a L tone realized earlier than the acoustic syllable onset. Generally, we see that the L tone is realized within 10 ms. of the syllable onset. Mean latency of the elbow from the syllable onset is 17 ms. for the one-word condition, and 3 ms. for the two-word condition. This difference was significant ($t = 4.17$, $p<0.05$).

[insert Figure 8 here]

In the one-word segmentation condition, the L tone of the early rise is rarely realized before the acoustic onset of the first syllable of the content word. This observation is understandable if we consider that in the two-word segmentation condition, there is an extra syllable preceding the first syllable of the content word. For example, in the carrier sentence in (3b), there are two syllables (*à la*) preceding the content word in the target sequence *la namileau*, but only one syllable (*à*) in the target sequence *Lanamileau*. This probably allowed the speaker more time to achieve the L target at the beginning of the early rise.

[insert Figure 9 here]

Figure 9 plots the latency of the L tone (early elbow) relative to the peak velocity of the lower lip trajectory on the y-axis (up-down movement). This is the crucial articulator for the target in the one-word *Manumulat* and the two-word *nos malomues*, where the syllable contains the bilabial [m] in onset position. Focusing on these two cases, we observe a compact distribution, with the L tone generally within 30 ms. of the articulatory event, and with a mean distribution at -6 ms. An exception is the case of *nos malomues* in phrase-medial position, where there is a larger spread of the distribution and the tone can even occur about 15 ms. before the peak velocity. This larger spread could be due to the uncertainty in the location of the L tone in these sequences. Hence, these data seem to suggest the existence of near-synchronization between L and the peak velocity of the primary constrictor trajectory. However, these two cases only make up a third of the data presented here. Our additional data contain syllable onsets that have the tongue tip as the crucial articulator. It is interesting to note that the mean latency values for the two-word segmentation conditions are almost identical for the different onset consonant articulator (i.e. lower lip: -6 ms, tongue tip: -6.3 ms) suggesting that the alignment of the tone to the peak velocity of the syllable onset articulatory gesture is rather stable regardless of the consonant type. While this lends credence for articulatory alignment with tones, more research has to be still conducted.

[insert Figure 10 here]

Figure 10 plots the latency of the L tone (early elbow) from peak velocity of the tongue tip trajectory on the y-axis (up-down trajectory). Data for the tongue apex show more variation than the data for the lower lip. This variation is expected given that there are two different apical consonants included in our corpus, the nasal [n], and the lateral [l], in different vowel contexts. The relevant sequences here are *Lanamileau/la namileau*, *ma numulat,* and *Nomalomue*. Despite these differences, the L tone is realized on average 17 ms. after peak velocity of the tongue apex trajectory for one-word segmentation cases (all word types and utterance positions included), and -6 ms. for two-word segmentation cases, starting with apical consonants.

As for the Italian data, we tested whether the articulatory latencies were less variable than the acoustic ones, before comparing which of these latencies were smaller. The results of the variance tests performed on pooled latencies for word type and segmentation type were not significant, except for the comparison between latency data for apical items for two-word segmentation, where the articulatory latencies were more variable. Hence we performed a set of four ANOVAs with pooled acoustic and articulatory alignment data, and with SIGNAL (acoustic or articulatory) as the main factor. Specifically, we first contrasted the latency values of the L relative to peak velocity (articulatory alignment) vs. the latencies relative to stressed syllable onset (acoustic latency) for one-word items. Here the results were significant only for the apical items [apicals: $F(1,94)=17.1$, $p<0.0001$; labials: $F(1,48)=1.65$, $p=0.2$], with mean values of 18 ms. for the acoustic latencies and of 36 ms. for the articulatory latencies. Similar ANOVAs were performed for the labial and apical two-word items. However, none of the results were statistically significant [apicals: $F(1,57)=0.38$, $p=0.54$; labials: $F(1,52)=3.06$, $p=0.06$].

4. Discussion

We started this enterprise with the specific goal of finding some articulatory events to which the tonal targets (as measured within the F_0 curve) might be synchronized. Remember that our strong hypothesis is that tonal targets are timed to co-occur with specific articulatory events. Therefore, a mere stable anchoring of the tonal target at a fairly large distance from an acoustic or articulatory event would not support this hypothesis.

Summarizing the results for Neapolitan, we saw that it is impossible to find an acoustic event to which the H target of the LH rise of questions and statements might be anchored within this view. Two possible candidates for anchoring were v1 (onset of the stressed vowel) for the H of statements and v2 (offset of the stressed vowel) for the H of questions. Though no rate effects were found on the alignment of the H with v1 for statements and with v2 for questions, both H latencies were quite large. Specifically, the statement H peak was found at a mean distance of 81 ms. after v1, while the question H peak was found at least 47 ms. after v2. Therefore, though one might suppose that question H targets might be aligned with the offset of the stressed syllable (as it has been suggested for H targets of prenuclear rises in various languages), this was clearly not the case given the results for our open syllable items.

On the other hand, when articulatory targets were considered, we obtained a closer synchronization of tone and anchor point. When looking at LA for labial consonant targets (in order to compare the present data to those of D'Imperio, Nguyen, and Munhall 2003) we already find a closer locking of tone and anchor, in that the mean latencies found are rarely more than around 30 ms. A near-synchronization of the events is, however, found for the latency to H from the peak velocity of the trajectory of the primary articulator for the consonantal target. That is, for labial targets in statement utterances, the H target latency from peak velocity of lower lip y movement trajectory is only between 5 and -8 ms. for fast renditions (both open and closed syllables), while for apical targets, the latency values for fast renditions are between 4 and -13 ms. Moreover, the articulatory event to which the tone might be phased does not seem to correspond to any identifiable segmental boundary or phonological unit. A potential acoustic candidate for the phasing of a tonal event is the peak RMS amplitude within the target syllable. However, when measuring the latency to H targets of statements from RMS peaks, the absolute values were quite large (H targets were located more than 50 ms. before the RMS peak).

Hence, the claim (Xu 2002) that tones and syllables can only be phased at a 0º angle with each other is contradicted by our results for Italian, in that H peaks of statements align with peak velocities of primary constrictor trajectories, while H peaks of questions seem to align with peak values of the same trajectories, offering evidence for at least two stable attractor states. Thus it is too simplistic to assume only one kind of phasing for the speech mechanism given the rapid changes in speech production moment to moment. However, much more work on speech motor control must be done to truly understand to what extent articulation constrains tonal alignment.

Our current formulation of the alignment of tones with articulatory events has two shortcomings. First, we are examining two different dimensions: actual articulatory events for the segmentals and an acoustic parameter (F_0) for the tonal specification. This concern can be addressed using techniques such as EGG to study tonal gestures at the glottal source. Second, no exact synchronization was found in relation to the articulatory dimensions under study. There was often a time lag (albeit a short one) between the tonal target and the measured articulatory dimension. It might be possible to 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 the realization of any two targets.

Related to this, we found an effect of rate on the articulatory alignment measures for the Italian EMA study. Remeber that for fast rates the latency values were generally closer to the zero (i.e., the reference point). In previous literature, Ladd et al. (1999) used the maintenance of temporal cohesion across different speaking rates as an indicator of stable alignment. However, the authors examined alignment only with respect to F_0 values, while we measured latencies between F_0 and articulatory events. Laryngeal and supralaryngeal articulators have different biomechanical properties and feature different inherent inertia, for instance. It could be that these articulators have different inherent responses under rate pressure, and that their relative timing could therefore be altered. The fact that the preliminary study by D'Imperio et al. (2003) did not show a rate effect, might be due to rate differences across the two studies. Very few studies have examined the effect of rate in both the articulatory and acoustic dimensions, so we need to be cautious in interpreting this result. Still, we believe that looking at the appropriate parameter (like peak velocity rather than displacement) may reveal more stability, though we still do not know how to account for all the other sources of variability.

Many acoustic studies have already shown a stability of tonal alignment to acoustic events, especially to the left edge of the stressed syllable (Caspers and van Heuven 1993, Ladd *et al.* 1999). Moreover, for French, Welby (2003, 2004), similarly to our present data here, found that 73% of L tones were realized within 20 ms. of the content word beginning. Hence, the motivation for the articulatory study for French did not have as an explicit goal to find a better alignment candidate for L than the acoustic one. On the other hand, the cross-linguistic comparison gives us enough reason to suppose that some of the regularities in the laryngeal/supralaryngeal coordination might be due at least partly to general motor control constraints. In fact, French L tones, much like the Italian H tones of statements, appear to be closely aligned to peak velocity of the primary articulator for the consonantal constriction, with average values for L latency from lower lip velocity as low as -2 ms. for *ma numulat* in initial position, or around 9 ms. for *Manumulat* and *Nomalomue* (i.e., one-word segmentation cases) in initial position. However, cases with very large alignment values were also found (such as a high mean value for *Lanamileau* in initial position), for which we cannot presently account. As expected, close alignment with the acoustic syllable onset was also found in the present study, spanning from 36 ms. for *Nomalomue* (one-word segmentation) to 6 ms. for *Manumulat* (both in initial position). However, the mean values for the acoustic alignment were smaller for two-word segmentation cases, with values as small as 3 ms. in initial position for *la namileau* and *ma numulat*.

Moreover, the articulatory alignment data were as stable as the acoustic ones and were not smaller for most of the comparisons carried out in the French data. While the first fact mirrors the Italian data, the synchronization hypothesis for articulatory data is not supported for French, though it is not disconfirmed either. Further studies with more recordings and more speakers are therefore needed before we can draw any definite conclusion. In French, another candidate for articulatory alignment may be the H of the late rise (the peak of this rise). Although this H is realized late in the last syllable of the accentual phrase, its exact position is a good deal more variable than that of the L of the early rise (see Welby and Lœvenbruck, to appear, for discussion). Finding articulatory alignment for the French H tone would strengthen the assumption of tonal events co-occurring with articulatory events.

5. Conclusion

This work attempts a new characterization of tonal target alignment based on articulatory rather than strictly acoustic anchor points and, with D'Imperio *et al.* (2003), is an initial empirical attempt to show how tonal gestures might be phased to oral articulatory gestures. Our results suggest that H targets of LH nuclear rises in Neapolitan might be synchronous with peak velocity for the primary consonantal constriction trajectories as far as statement H peaks are concerned. In questions, H peaks appear to be timed to occur at the peak trajectory of the relevant articulator (that is, at a zero velocity point). Though for French the acoustic alignment data for the L tone of the early rise are as variable and show similar latency values than the articulatory latencies, we noticed also for this language a trend for the production of L tones around the peak velocity of the primary constrictor, at least for some of our target words and utterance positions. However, more evidence for an articulatory alignment might come from studying more data for the early L as well as examining the H of the late rise in French, as this is known to be less stable for acoustic analyses. Finally, 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.

6. Acknowledgments

This research was supported by ACI grant 0220244 from the French Ministry of Research to Mariapaola D'Imperio. The authors would like to give special thanks to Pascal Perrier and Christophe Savariaux at ICP, Grenoble, France, for their help with the EMA system as well as data collection and pre-processing. Further thanks are addressed to Pascal Perrier for helpful discussions on motor control issues. Thanks also to Kevin Munhall for the collection of OPTOTRAK data at Queen's University, Kingston, Ontario, Canada. Finally, we would like to thank Dani Byrd and the editors of this volume for their very helpful suggestions and comments on an earlier version of this paper.

References

Arvaniti, Amalia, D., Robert Ladd, and Ineke Mennen 1998 Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics* 36: 3–25.

Boersma, Paul, and David Weenink 2005 Praat: doing phonetics by computer (Version 4.2) [Computer program]. Retrieved from http://www.praat.org/

Byrd, Dani, and Elliot L. Saltzman 1998 Intragestural dynamics of multiple phrasal boundaries. *Journal of Phonetics* 26:173–199.

Byrd, Dani, Abigail Kaun, Shrikanth Narayanan, and Elliot Saltzman 2000 Phrasal signatures in articulation. In *Papers in Laboratory Phonology V, Acquisition and the Lexicon*, Michael B. Broe and Janet B. Pierrehumbert (eds.), 70–87. Cambridge: Cambridge University Press.

Caspers, Johanneke, and Vincent J. van Heuven 1993 Effects of time pressure on the phonetic realization of the Dutch accent-lending pitch rise and fall. *Phonetica* 50: 161–171.

Cho, Taehong in press Manifestation of prosodic structure in articulation: Evidence from lip movement kinematics in English. *Laboratory Phonology 8*. Berlin/New York: Mouton de Gruyter.

Cooke, J. D. 1980 The organization of simple, skilled movements. *Tutorials in Motor Behavior*. G. E. Stelmach and S. O. Requin (eds.), 199–211.

Di Cristo, Albert 1998 Intonation in French. *In Intonation Systems: A Survey of Twenty Languages*, Daniel Hirst and Albert Di Cristo (eds.), 197–220. Cambridge: Cambridge University Press.

D'Imperio, Mariapaola 1999 Tonal structure and pitch targets in Italian focus constituents. In *Proceedings of the 14th International Congress of Phonetic Sciences*, John Ohala (ed.), vol. 3, 1757–1760, San Francisco, U.S.A.

2000 The Role of Perception in Defining Tonal Targets and their Alignment, Ph.D. diss., Department of Linguistics, The Ohio State University.

2001 Focus and tonal structure in Neapolitan Italian, *Speech Communication* 33(4): 339–356.

2002 Italian intonation: An overview and some questions. *Probus* 14: 37–69.

D'Imperio, Mariapaola, and David House 1997 Perception of questions and statements in Neapolitan Italian. In *Proceedings of Eurospeech '97,* vol. 1, G. Kokkinakis, N. Fakotakis, and E. Dermatas (eds.), 251–254, Rhodes, Greece.

D'Imperio, Mariapaola, Noël Nguyen, and Kevin Munhall 2003 An articulatory hypothesis for the alignment of tonal targets in Italian. In *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, Spain, 3–9 august 2003, 253–256.

D'Imperio, Mariapaola, Caterina Petrone, and Noël Nguyen to appear How autonomous is fundamental frequency information? Effects of tonal alignment on lexical identification in Italian. *Proceedings of Tone and Intonation in Europe*, Santorini, Greece.

Espesser, Robert 1996 MES : Un environnement de traitement du signal. *Proceedings of XXIe Journées d'Étude sur la Parole*, 447. Avignon, France.

Gracco, Vincent L. 1988 Timing factors in the coordination of speech movements. *Journal of Neuroscience* 8 (12): 4629–4639.

Jun, Sun-Ah, and Cécile Fougeron 2000 A phonological model of French intonation. In *Intonation: Analysis, Modelling and Technology*, Antonis Botinis (ed.), 209–242. Boston: Kluwer.

Kelso, J.A.S. 1984 Phase transitions and critical behavior in human bimanual coordination. *American Journal of Physiology: Regulatory, Integrative and Comparative* 246: R1000–R1004.

1987 The variation of absolute and relative measures of speech activity. Journal of Phonetics: 15, 335–347.

and Eric Vatikiotis-Bateson, Elliot L. Saltzman and Bruce Kay 1985 A qualitative dynamic analysis of reiterant speech production: phase portraits, kinematics, and dynamic modeling. *Journal of the Acoustical Society of America*, 77 (1): 266–280.

Krakow, Rena A. 1993 Nonsegmental influences on velum movement patterns. In *Nasals, Nasalization, and the Velum*, M. K. Huffman and Rena A. Krakow (eds.). New York: Academic Press. [= Phonetics and Phonology, vol. 5].

Ladd, D. R., Dan Faulkner, Hanneke Faulkner, and Astrid Schepman 1999 Constant “segmental anchoring” of F0 movements under changes in speech rate. *Journal of the Acoustical Society of America* 106: 1543–1554.

Lœvenbruck, Hélène 1996 Pistes pour le contrôle d’un robot parlant capable de réduction vocalique. Doctoral diss., Institut National Polytechnique de Grenoble, Grenoble.

Mertens, Piet, Jean-Philippe Goldman, Éric Wehrli, and Arnaud Gaudinat 2001 La synthèse de l'intonation à partir de structures syntaxiques riches. *Traitement Automatique des Langues* 42(1): 145–192.

Nelson, W. L. 1983 Physical principles for economies of skilled movements. *Biol. Cybern.*, 46: 135–147.

Nittrouer, Susan, Kevin Munhall, J. A. Scott Kelso, Betty Tuller, and Katherine S. Harris 1988 Patterns of interarticulator phasing and their relation to linguistic structure. *Journal of the Acoustical Society of America* 84 (5) : 1653–1661.

Pasdeloup, Valérie 1990 Modèle de règles rythmiques du français appliqué à la synthèse de la parole. Doctoral diss., Institut de Phonétique d’Aix-en-Provence, Université de Provence.

Pierrehumbert, Janet B., and Mary E. Beckman 1988 *Japanese tone structure*. Cambridge, MA: MIT Press.

Ostry David J., Eric Keller, and A. Parush 1983 Similarities in the control of the speech articulators and the limbs: Kinematics of the tongue dorsum movement in speech. *Journal of Experimental Psychology (Human Perception and Performance)* 9: 622–636.

Perkell, Joseph S., Majid Zandipour, Melanie L. Matthies, and Harlan Lane 2002 Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues. *Journal of the Acoustical Society of America* 112(4): 1627–41.

Perkell, Joseph S.., and Majid Zandipour. 2002 Economy of effort in different speaking conditions. II. Kinematic performance spaces for cyclical and speech movements. *Journal of the Acoustical Society of America*. 112(4): 1642–51.

Pascal Perrier, Christian Abry, and Eric Keller 1989 Vers une modélisation des mouvements du dos de la langue. *Journal d'Acoustique* 2: 69–77.

Prieto, Pilar, Mariapaola D'Imperio, and Barbara Gili Fivela to appear Pitch accent alignment in Romance: primary and secondary associations with metrical structure. *Language and Speech*, special issue on Intonation in Language Varieties, edited by Paul Warren.

Prieto, Pilar, Jan. P. van Santen, and Julia Hirschberg 1995 Tonal alignment patterns in Spanish. *Journal of Phonetics* 23: 429–451.

Silverman, Kim E. A. and Janet B. Pierrehumbert 1990 The timing of prenuclear high accents in English. In *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, J. Kingston and M. E. Beckman (eds.), 72–106. Cambridge: Cambridge University Press.

Vaissière, Jacqueline 1988 Prediction of velum movement from phonological specifications. *Phonetica* 45: 122–139.

Vaissière, Jacqueline 1997 Langues, prosodies et syntaxe. *Traitement Automatique des Langues* 38: 53–82.

Vatikiotis-Bateson, Eric, Kevin G. Munhall, and David J. Ostry 1993 *Optoelectronic measurement of orofacial motions during speech production, In Measuring Speech Production*, Maureen Stone (ed.), Acoustical Society of America, 1993.

Welby, Pauline 2003 The slaying of Lady Mondegreen, being a study of the association and alignment of French intonational rises and their role in speech segmentation. PhD diss., The Ohio State University.

The structure of French intonational rises: A study of text-to-tune alignment. In *Proceedings of the Speech Prosody 2002 Conference*, Bernard Bel and Isabelle Marlien (eds.), 127–130. Nara, Japan.

and Lœvenbruck, Hélène (to appear). Anchored down in anchorage: Syllable structure and segmental anchoring in French. In *Italian Journal of Linguistics*, special issue on Current Issues on Tonal Alignment in European Languages, edited by Mariapaola D'Imperio.

Xu, Yi 1998 Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica* 55: 179–203.

2002 Articulatory constraints and tonal alignment. In Proceedings of the Speech Prosody 2002 Conference, Bernard Bel and Isabelle Marlien (eds.), 91–100. Aix-en-Provence, France.

and Xuejing Sun 2002 Maximum speed of pitch change and how it may relate to speech. *Journal of Acoustical Society of America* 111: 1399–1413.

and Q. Emily Wang 2001 Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication* 33: 319–337.

Notes

[1] Data for several points on the tongue as well (which provide information on vowel articulation), but we did not find any interesting regularities concerning these measurements.

[2] The OPTOTRAK is a three-dimensional electronic movement tracking device (Vatikiotis-Bateson, Munhall, and Ostry 1993). The experiment was conducted at the Speech Perception and Production Laboratory, Queen's University, Kingston, Ontario, Canada.

[3] The EMA systems sends short synchronization beeps at the beginning and end of each articulatory recording. The two synchronization signals, which feature a sharply rising ramp followed by a sharply falling ramp, are recorded on the right channel of the DAT. After the recording session, a Matlab script uses the two synchronization beeps to extract the audio cuts the audio signals corresponding to the articulatory signals.

[4] MES is a signal editor based on a plug-in system, able to display and/or compute various types of information: acoustic as well as articulatory data, including EPG data, and multilayered labelling tiers.

[5] At fast rate, the speaker's production was 15% faster, with a mean rate of about 6.17 syl/s., as opposed to 5.43 syl/s. for the normal rate.

[6] The black point inside the box is located at the median of the data; the height of the box is equal to the interquartile distance (IQD); the whiskers extend to the extreme values of the data or a distance 1.5 IQD from the point, whichever is less.

[7] R Development Core Team, 2004. R is a computer language for statistics, sometimes called GNU S. It is a freeware similar to S.

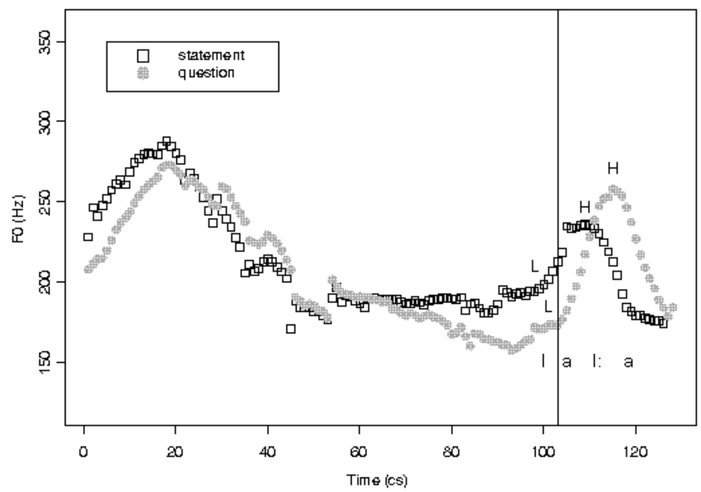


Figure 1. Fundamental frequency (F_0) values of a narrow focus yes/no question utterance (filled circles) and a narrow focus statement utterance (empty squares) produced by a Neapolitan speaker (from D'Imperio 2002). Note that the LH tone sequence is later in the question utterance, relative to the onset of the stressed vowel, which served as the synchronization reference for the two utterances.

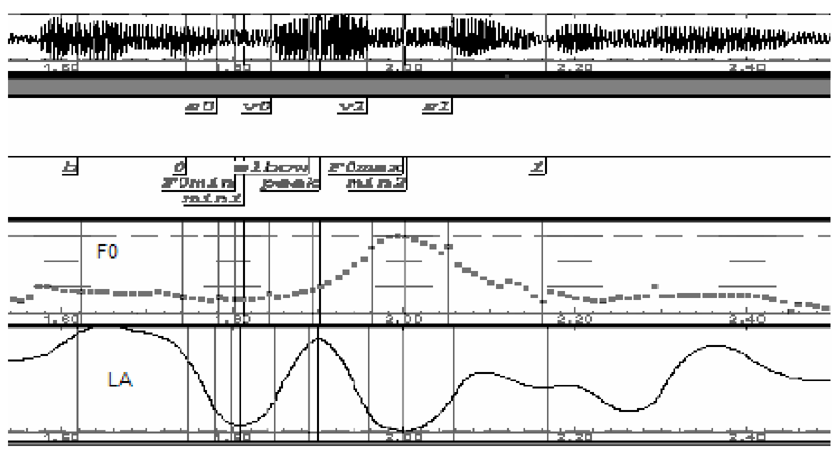


Figure 2. Waveform, labels, F_0 curve and between-lip distance values for the Italian question *Vedrai mamma domani?* 'Will you see Mom tomorrow?' produced in a normal speech rate. Note that the F_0 peak (F0max) is aligned with a local minimum of the between-lip distance curve (solid line), around the right edge of the stressed syllable [mam]. v0 = onset of the stressed vowel; v1 = offset of the stressed vowel; s0 = onset of the stressed syllable; s1 = offset of the stressed syllable. (from D'Imperio, Nguyen, and Munhall 2003).

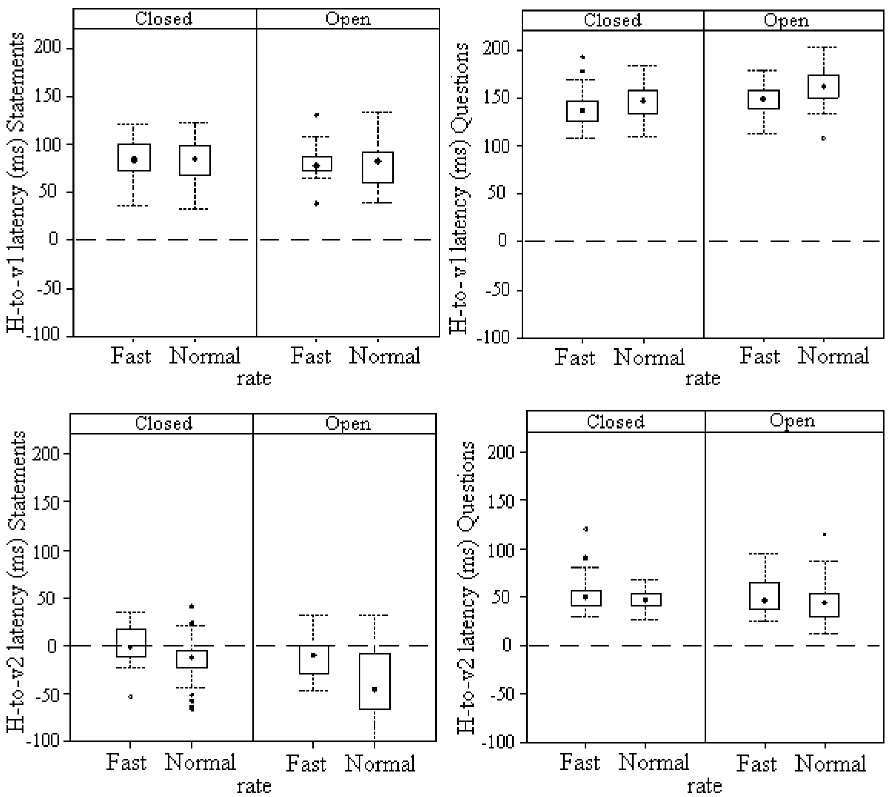


Figure 3. Boxplot of H latencies relative to the onset of the stressed vowel (v1) for Italian statements (upper left) and questions (upper right), H latencies relative to the offset of the stressed vowel (v2) (statements: lower left; questions: lower right) for both closed (C) and open (O) syllables and across fast (F) and normal (N) speech rates.

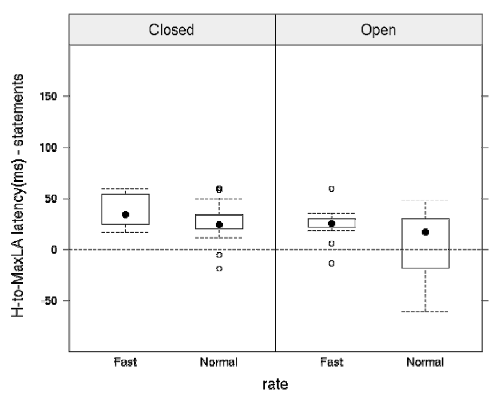


Figure 4. Boxplot of H latencies relative to MaxLA for Italian statement utterances for both closed and open syllables and across fast and normal speech rates.

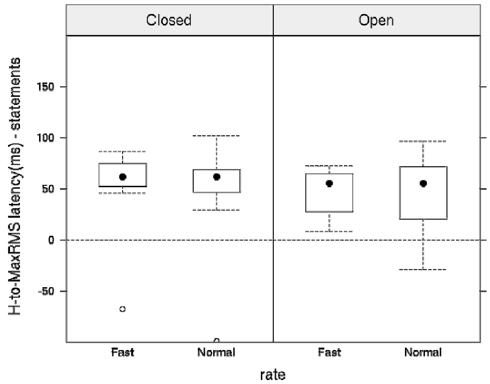


Figure 5. Boxplot of H latencies relative to Max RMS amplitude for Italian statement utterances (labial consonants), for both closed and open syllables and across fast and normal speech rates.

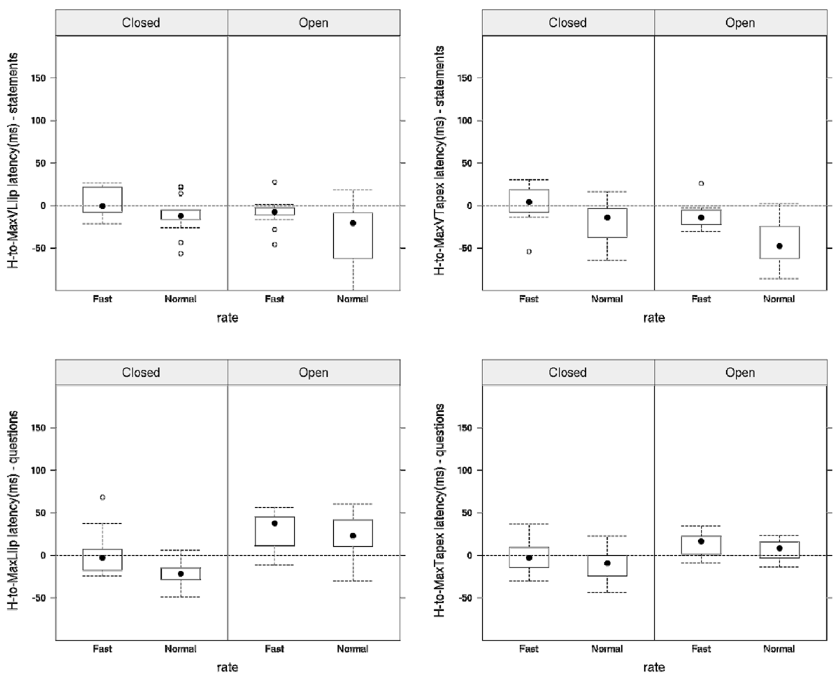
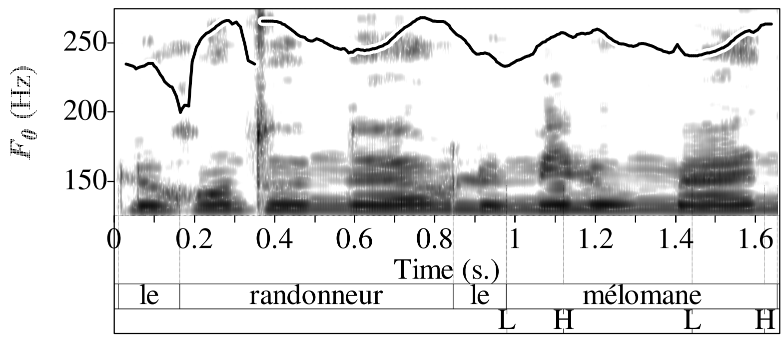


Figure 6. Boxplots of H latencies relative to Max velocity of primary articulator for Italian statements (upper panel, left: labials; right: apicals) and to Max trajectory of the primary articulator for Italian questions (lower panel, left: labials; right: apicals), for both closed and open syllables and across fast and normal speech rates (MaxVLlip = peak velocity, lower lip y; MavVTapex = peak velocity, tongue apex y; MaxLlip = peak trajectory, lower lip y; MaxTapex = peak trajectory, tongue apex y).

a)



b)

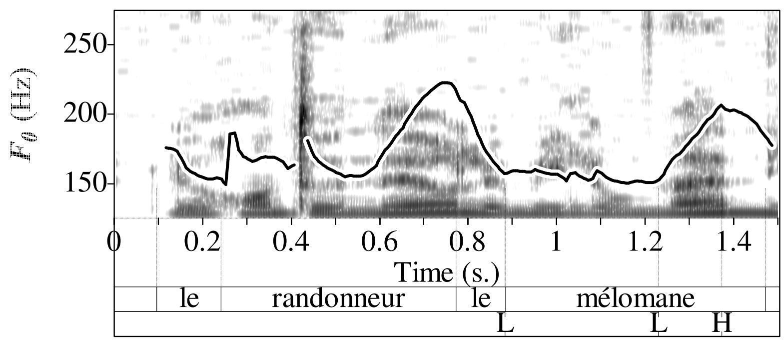


Figure 7. The French accentual phrase *le mélomane* ‘the music lover’ (a) with an optional early rise (a LHLH pattern) and (b) without the early rise (a LLH pattern). The complete utterance is *Le randonneur, le mélomane et la forestière s’étaient disputés avec eux* ‘The hiker, the music lover and the forester had been arguing with them.’ (example from Welby 2003).

Table 1. Table of French corpus.

	Position	Segmentation	No. of Items	Total
[la.na.mi.lo]	Initial	one word	6	25
		two word	19	
	Medial	one word	18	26
		two word	8	
[ma.ny.my.la]	Initial	one word	7	27
		two word	20	
	Medial	one word	18	25
		two word	7	
[no.ma.lo.my]	Initial	one word	8	27
		two word	19	
	Medial	one word	16	25
		two word	9	

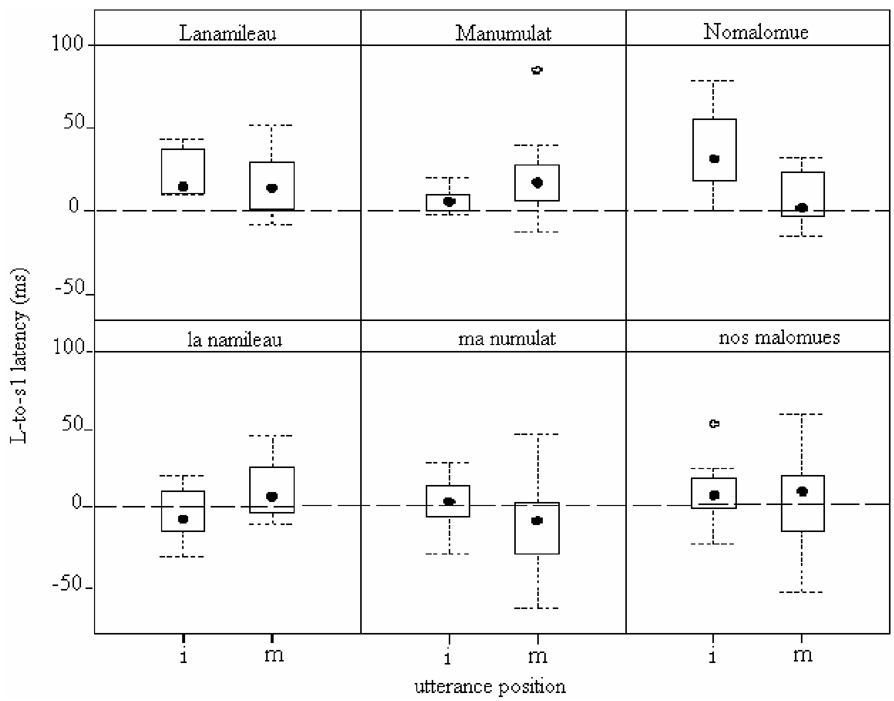


Figure 8. Latency to L tone (early elbow) from acoustic syllable onset by word type and position of target word in phrase: (i)nitial, (m)edial, for both one-word (upper panel) and two-word (lower panel) segmentation, for French.

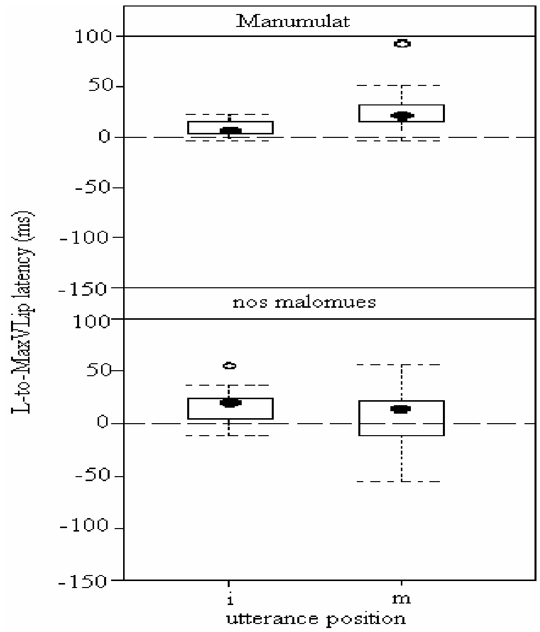


Figure 9. Latency to L tone (early elbow) from point of maximum velocity of lower lip_y separated by word type and position of target word in phrase (i)nitial, (m)edial, for both one-word (upper panel) and two-word (lower panel) segmentation, for French.

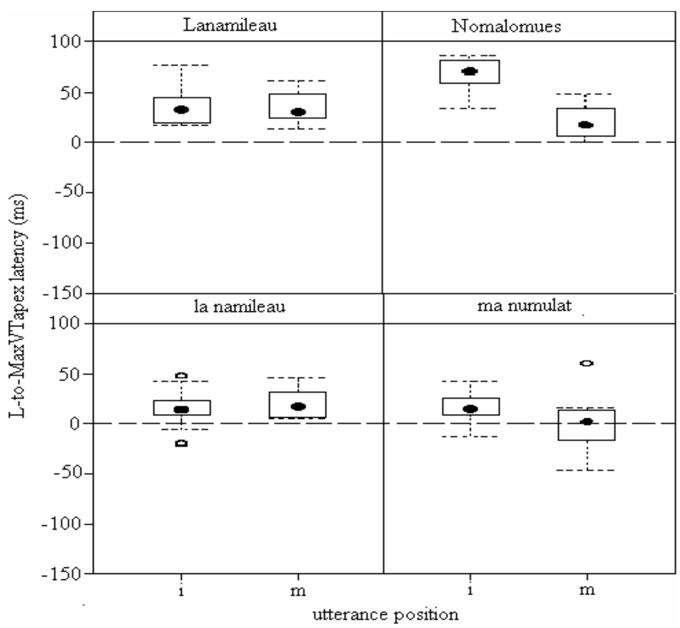


Figure 10. Latency to L tone (early elbow) from point of maximum velocity of tongue apex_y separated by word type and position of target word in phrase: (i)nitial, (m)edial, for both one-word (upper panel) and two-word (lower panel) segmentation, for French.