Language-Specific and Universal Constraints on Tonal Alignment: The Nature of Targets and “Anchors”

Mariapaola D’Imperio

Laboratoire Parole et Langage, CNRS UMR 6057
Université de Provence, France
dimperio@lpl.univ-aix.fr

Abstract

Much recent work on tonal alignment, including its modeling, implies the existence of well-defined “targets” as well as “segmental” anchors to which the tones would be aligned. But researchers in this area are saddled with a paradox. In search for universal principles of alignment, they need to compare acoustic measurements for many languages, and actively search for effects that are specific to a single language, though a model which is itself specific to a single language is not desirable for a strong “universalist” view. However, most of the recent alignment work has been performed within the realm of acoustics, while perception and articulatory measures have been considerably neglected. After reviewing some acoustic studies, perception studies are presented which suggest that tonal targets might have a perceptual basis and might be guided by both psychoacoustic (i.e. universal) and language-specific constraints. Also, very recent articulatory data are presented which only partially support the strong hypothesis of universal articulatory constraints being the main driving force for alignment regularities.

1. Introduction

Generally, tonal alignment refers to the synchronization of tune and some specific segment or prosodic location (such as syllable onset) within the text. The elements selected in the tune can be pitch movements or pitch “targets” (i.e., the beginning and end of a tonal movement). Specifically, tonal targets can be defined along two axes, which are the temporal location (i.e., *alignment*) and the melodic value within the F_0 range of the speaker (i.e., *scaling*). Employing the notion of tonal target calls for the need to explicitly define what a target is. Intuitively, one can think of a tonal target as a maximum or minimum value within the F_0 curve. However, it is quite hard to discern a tone target in many occasions. It is possible, in fact, to hear a tune contrast without clearly discerning the peak or valley in the F_0 contour. This is the case shown in Figure 1, presenting the F_0 contour for a broad focus declarative in Neapolitan Italian. Native speakers hear a clear falling accent on the stressed syllable *Lal-* of *Lalla* at the location marked by H+L* and the arrow. But how can one exactly locate the accent peak within an F_0 plateau? It is proposed here that the investigation of tonal alignment in perception could offer a window on some peculiarities of tonal alignment in production.

Apart from the issue of what a target really is, there is yet another issue that researchers on alignment are tackling, which is the nature of the “anchor points” for alignment. In other words, if alignment is really systematic, what are the reference points which would make these regularities clearly emerge?

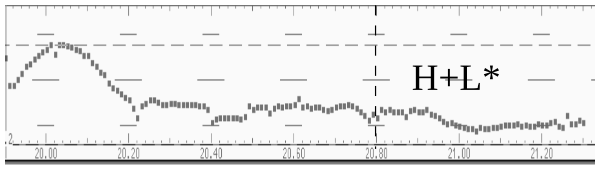

Figure 1: F_0 *curve and spectrogram for the broad focus declarative* Mamma andava a ballare da Lalla.

Also, are these anchors universal? For instance, [1] showed that rising prenuclear accents in Greek begin at the onset of the stressed syllable and peak early in the following unstressed vowel, independent of syllable composition and duration. Analogously, [12] showed that the beginning of the F_0 rise in English prenuclear accents is consistently aligned with the stressed syllable onset and that this alignment is constant under speech rate differences, suggesting that listeners might recover a certain intonation pattern by means of “invariant” alignment features. A stable anchoring of the L(ow) relative to syllable onset has also been found for Dutch rises by [3], for Mexican Spanish rises by [16] and for Mandarin by Xu (cf. [23]). In some cases, though, researchers admit that is it quite difficult to exactly locate a L tone, thus often choose to measure it at syllable onset as a matter of consistency.

At the same time, other work, more based on perception, has emphasized the importance of language-specific alignment differences for conveying pragmatic differences. For example, [21] and [15] showed that a late peak in a rising pitch accent can convey “uncertainty” or “incredulity” in American English, while an early peak cues “assertion”. Similarly, alignment of the peak has been found to be categorically perceived in German for the categories “established”, “new” and “emphatic” [11]. More recently, [9] showed that in Neapolitan Italian, the intonation patterns for statements and yes/no questions can be distinguished by a subtle alignment difference of a rise-fall pattern on the nuclear stressed syllable. In the next section, I review some studies investigating the role of factors affecting alignment in its acoustic realization, while §3 will be devoted to the perception of alignment. Finally, in §4 I shall briefly consider both cross-linguistic perceptual data as well as articulatory data bearing on the universality of alignment constraints.

2. Acoustics of tonal alignment

Among the major phonetic and phonological factors affecting alignment, we find: tonal environment (e.g., upcoming or preceding tones), focal structure, speech rate, phonological length, syllabic structure and segmental effects (intrinsic vowel duration, consonant voicing, etc.). Here I shall be particularly concerned with tonal and syllabic structure effects.

As to the tonal factors, [2] first showed that in Stockholm Swedish the shape of a word accent can change depending on its sentential and focus context. For instance, the copresence of a word accent and a sentence accent on the same syllable (e.g. when the syllable is sentence-final) leads to a situation of tonal crowding that can be resolved either through temporal readjustments of peak location or through melodic readjustments of the F_0 level reached by the tonal target (*undershoot*). This and other effects are accounted for in terms of "interference" between tonal commands. Interference, or "overlap" of articulatory sequences for tonal commands has also been invoked in works on English. [19] tested various hypotheses stemming from the general idea of "gestural overlap", already present in the segmental literature. For instance, when two competing tones have to be realized on the same TBU (such as a stressed syllable), the target for the first tone can be moved earlier in time to make space for the second tone to be realized. This was defined as the "tonal repulsion" hypothesis.

The presence as well as the absence of alignment alterations due to tonal crowding have also been observed for Neapolitan Italian [4, 6, 8]. Among the findings, tonal repulsion has been found to affect the alignment of the H peak in the yes/no question pitch accent (L*+H) as a consequence of the vicinity of a falling phrase accent. On the other hand, the number of syllables to the right of the nuclear accented syllable does not seem to have a clear effect in determining peak repulsion [6]. In another study [8], the absence of a repulsion effect on the peak of broad focus declarative pitch accents has been taken as evidence for a bitonal structure of the same (H+L*, instead of H* followed by a L- phrase accent). Effects of tonal crowding have also been shown for other varieties of Italian [10]. In sum, the temporal alignment of F_0 targets seems to be affected by a number of very fine phonetic and phonological variables that can potentially interfere with meaningful alignment (for pragmatic contrast) that is crucial for an effective communication.

2.1. Phonological length and syllabic structure effects

In an attempt to model the alignment of the H* of American English, [20] showed the importance of weighted effects of intra-syllabic segmental composition, syllable onset and rhyme duration. More recently, the findings in [13] suggest that vowel duration, as a consequence of the phonological length of vowels, affects peak alignment in Dutch prenuclear rises. Italian presents a length contrast that is analogous to the Dutch case, though it strictly depends on syllable structure. Open stressed syllables are in fact characterized by long vowels in Italian (though this holds true only for penultimate syllables). On the other hand, closed stressed syllables will only allow the presence of short vowels.

It had been previously observed that in Neapolitan accent peaks are closer to vowel offset when the stressed vowel occurs in a closed syllable. It is also true that the peak of Neapolitan yes/no question accent is later than that of (narrow focus) statements, everything else being equal (see Figure 2). Syllable structure effects can therefore render the relative alignment of statement peaks dangerously similar to that of question peaks. Compare the peak location in the question utterance *Vedrai il NANO dopo?* "Will you see the DWARF afterwards?" (capital letters indicate accent placement) and the peak location in the declarative *Mangerai il MANGO dopo* in Figure 3. Both

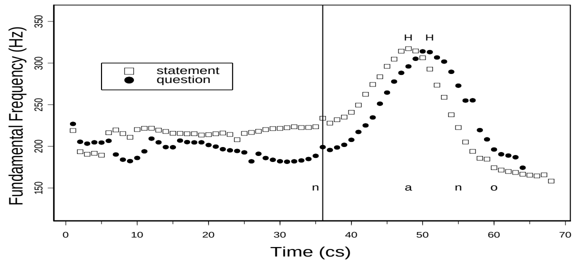

Figure 2: F_0 *traces for a narrow focus statement (open squares) and a question (filled squares)* Vedrai il NAno./? *"Will you see the dwarf" in Neapolitan Italian. The vertical bar marks the onset of the stressed vowel* a.

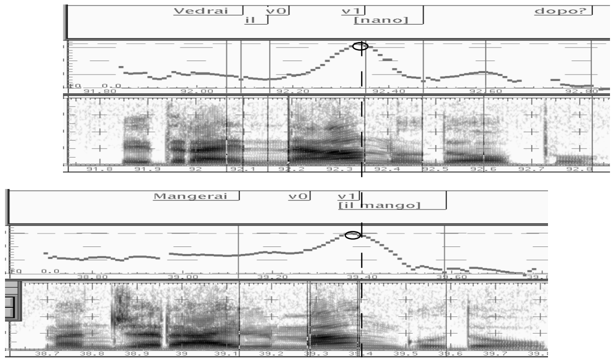

Figure 3: *Tone labels,* F_0 *curve and spectrogram for an open-syllable question (upper) and a closed-syllable statement (lower). Stressed vowel offsets are lined up at the dashed line. (From D'Imperio, 2000).*

peaks are aligned very close to stressed vowel offset (indicated by "v1", marked by the dashed line). Note, though, that the stressed vowel is shorter in the declarative, as a result of occurring within the closed syllable *man-*.

If we take F_0 alignment relative to the *right* edge of the syllable to be the main index of accent identity, then pairs such as these seem awkwardly similar. How do Neapolitan listeners deal with such a potential parsing ambiguity? The issue here is what counts as phonologically the same intonation pattern. Hence, in [5, 7] the open vs. closed syllable contrast was investigated in relation to the alignment of the rise-fall configuration of Neapolitan yes/no questions and narrow focus statements. The strong hypothesis tested was: tonal targets are timed to occur at a specific, absolute location in the utterance (e.g., at a certain temporal distance after vowel or syllable onset). This predicts that peaks will be invariantly produced at such locations, independent of vowel duration differences due to syllable structure (open vs. closed syllable) and segmental environment. Alternatively, if target alignment is proportional to the entire vowel, rhyme and/or syllable duration, target location is expected to differ accordingly. For instance, we would expect peaks to be relatively later within the stressed vowel when this is short. Therefore, one of the factors manipulated in the acoustic study in [5] was whether the accented vowel occurred in an open or in a closed syllable.

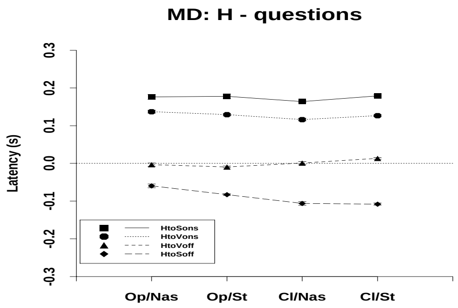

Figure 4: *Mean H peak latency from vowel onset (HtoVons), vowel offset (HtoVoff) and syllable offset (HtoSoff) for speaker MD (Op/Nas = open syll., nasal; Op/St = open syll., stop; Cl/Nas = closed syll., nasal; Cl/St = closed syll., stop). The dotted line is the reference point for each latency measurement. (From D'Imperio, 2000).*

Remember that [13] found an effect of vowel duration on peak alignment in Dutch, but alignment was measured only relative to vowel offset. An interesting picture emerges for Neapolitan when both right and left edge of the prosodic domain are elected as possible anchor points. Different latency measures were therefore compared in order to test whether they would be more sensitive to syllable structure and segmental environment. In Figure 4 such measures are shown for question H peaks for one of the speakers across all conditions. Apart from latency from stressed vowel onset (HtoVons), we find latency from stressed syllable onset (HtoSons), latency relative to stressed vowel offset (HtoVoff) and latency relative to stressed syllable offset (HtoSoff). Note that only HtoVons and HtoSons (squares and circles) are pretty constant throughout the conditions. On the other hand, when H peak latency was calculated relative to the right edge of the syllable, the Open/Closed effect was significant. Similar results were also obtained for statements and were confirmed through statistical analysis for both speakers.

Similar to Italian, absolute peak delay performed better than delay proportional to a section of the syllable for Mexican Spanish [16], while the opposite was true for the English model in [19] and [20]. Remember also the long/short vowel effect for Dutch. How can these discrepancies be explained? It might be that the Dutch effect, if it remains true when peaks are measured relative to vowel or syllable onset, could be interpreted as a redundant means to ensure the correct identification of the phonological contrast between long and short vowels, which might otherwise be difficult to maintain under noisy speech conditions. A vowel length contrast might be a harder contrast to maintain perceptually than the open/closed syllable opposition of Italian.

Moreover, if, as maintained by Xu (this volume), targets align with structural locations within the syllable (such as syllable onset or offset) on the basis of universal articulatory constraints, then a clear paradox emerges regarding the cross-linguistic differences found for acoustic results on alignment. The answer could lie in the fact that we are looking at the wrong "anchors". Assume that the structural anchors have a different nature. A plausible alternative to strict acoustic alignment would be that tonal targets are perceptual constructs which are dependent on psychoacoustic constraints as well as on language-specific constraints on perception. Alignment invariance will then depend on the delicate balance among these constraints. The next paragraph will present results from some perception experiments that can shed some light on this issue.

3. Perception of target alignment

Most of the perception research on alignment conducted thus far is based on the strong assumption that tonal targets are to be identified with the pitch peaks or valleys in the F_0 contour. However, there are cases in which only stipulative decisions can be made regarding tonal target position, either because a single maximum (or minimum) cannot be discerned (as in plateaus, cf. Figure 1 above), or where a perturbation is created by some consonant, or because the target is masked by the presence of a voiceless sound. The alignment of the target for such tough cases could be revealed through the study of the interaction of psychoacoustic constraints on perception. For instance, work on non-speech stimuli conducted in the seventies suggests that complex pitch configurations (such as rising or falling pitch movements) are not perceived in their entirety, and that the perceived pitch target corresponds to the F_0 value at roughly two thirds of the length of the rise or fall [14, 18]. Moreover, when a pitch movement is preceded or followed by a plateau, the perceived pitch value will correspond to the pitch within the plateau. The perceived "timing" of a tonal target had instead never been investigated.

The investigation of tonal alignment in perception could also shed light on some peculiarities of tonal alignment in production. Consider for instance the phenomenon of "peak delay". In a number of languages (American English, Italian, Mexican Spanish, Greek), the alignment of the H prenuclear accent is late relative to the syllable with which it is associated. A similar peak delay has also been found in a non-Western language, i.e. Mandarin [22]. By performing perception studies on the role of tonal alignment, one can discover whether such peculiarities of alignment patterns in production are due to specific constraints on our perceptual system. The next section will present the results of a perception experiment aimed at determining the alignment of tonal targets in plateau stimuli. In §3.2 the hypothesis of absolute alignment formulated above for production is tested in perception.

3.1. Perception of questions and statements in Neapolitan Italian: plateau versus peak stimuli

As mentioned above, [9] showed that in Neapolitan Italian, the intonation patterns for statements and questions can be distinguished by the alignment of the peak of a rise-fall pattern on the main stressed syllable, though this difference is rather small (around 40 ms). Basically, through F_0 event timing manipulations alone, the authors were able to induce a clear question and statement category with a region of ambiguity at the center of the continuum. Two of the stimuli series employed were made of flat peak stimuli, where the implicit assumption was that the peak of those stimuli would correspond to the time coordinate of the end of the rise. Consequently, we expected statements to be identified at early peak locations, just as for peak stimuli. However, peak target location for flat peak stimuli was neither explicitly defined nor tested. Surprisingly, one of these series (the "inter-rise", see Figure 5) scored a majority of question responses already at early continuum locations.

Different explanations can be invoked in order to account for the "rise-shift" results based on different hypotheses regarding the perceived location of the H target. One hypothesis relates to the the "shape" of the accent. The rise-shift stimuli were

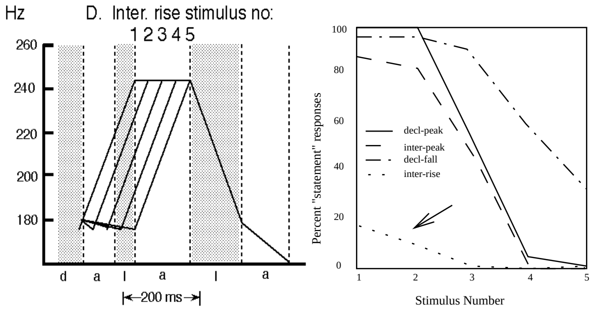

Figure 5: *Stimuli in which the rise was shifted across the stressed vowel (left) and results for all series with arrow marking the inter-rise results (from D'Imperio and House, 1997).*

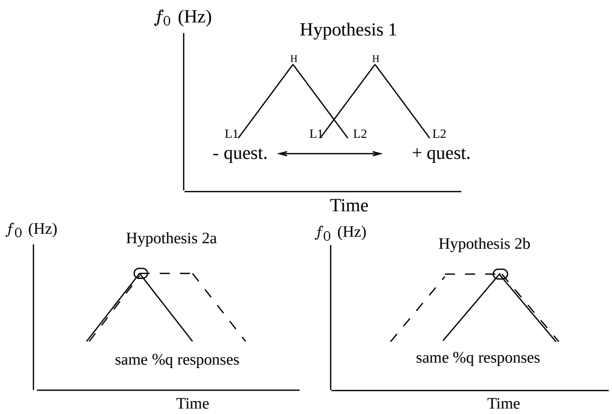

Figure 6: *Schematic representation of the hypotheses. The dotted line represents plateau stimuli.*

in fact characterized by a high F_0 plateau, which resulted from splitting the rise and the fall in stimuli 1-4 (see Figure 5, left). One obvious hypothesis would therefore be that in "plateau" stimuli the perceived target is *not located at the end of the rise*, but elsewhere. Plausible locations would be the middle of the plateau, the end of the plateau, etc. A perception experiment aiming at finding the exact location of the perceived tonal target of plateau stimuli was hence conducted by employing a comparison with canonical "peak stimuli".

Hence, in [5], I first examined the responses of the listeners to a manipulation of the three main targets of the rise-fall configuration in the time domain (hypothesis 1, see Figure 6, upper), through the use of resynthesized stimuli. It was found that such a manipulation can indeed shift the perception of a question to a statement and took this to mean that the perceived location of the rise-fall peak is shifted in the time domain from a "late" to an "early" value (see Figure 7, left), replicating the results of [9] for peak stimuli. Then I tested the hypothesis that the number of question responses would be the same for plateau stimuli and peak stimuli whose peak is timed at plateau onset (hypothesis 2a, see Figure 6, lower left). In other words, I tested whether the "target" of the rising transition of plateau stimuli is timed to occur at the end of the LH rise (which was implicitly assumed in [9]). The alternative hypothesis would be that the perceptual target location for plateau stimuli corresponds to the the offset of the plateau (hypothesis 2b, see Figure 6, lower right), i.e., the beginning of the HL fall. This was done by comparing scores

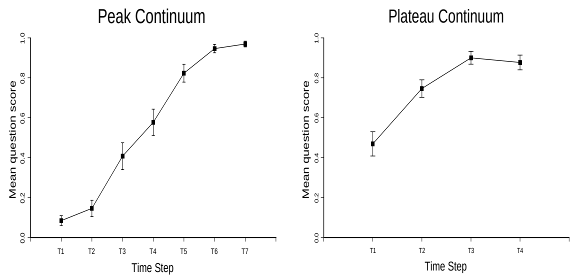

Figure 7: *Peak continuum mean scores (left) and plateau continuum scores (right) for all listeners. From D'Imperio, 2000.*

Plateau	Mean score	Peak	Mean Score
T0	0.29	T3	0.37
T1	0.38	T4	0.47
T2	0.61	T5	0.79
T3	0.86	T6	0.94
T4	0.94	T7	0.99
T5	0.95	T8	0.97

Table 1: *Mean scores for the plateau continuum and peak stimuli timed with plateau offset.*

for peak stimuli with peak timed at plateau offset.[1]

First, note the difference between the results of the peak stimuli timed at plateau onset and plateau stimuli (see Figure 7). Plateau stimuli scored higher percentages of question responses already very early in the continuum. A much smaller difference was found, instead, when a comparison was made between the results of peak stimuli timed at plateau offset and plateau stimuli. This result suggests that the shape of the accent peak does indeed affect the perception of target location, and that the perceived target for plateau stimuli must be located within the region of the plateau offset.[2] This experiment was actually replicated with a larger stimuli series (adding a time step to the left and one to the right), and no statistical difference was found between plateau stimuli scores and scores for peak stimuli timed at plateau offset (see Table 1). Above all, this result is taken to mean that the perceived target of plateau stimuli cannot be identified with the end of the LH rise. One might object that such a difference between plateau and peak stimuli scores is due to an effect of F_0 level height within the LHL configuration. However, F_0 had relatively little effect when stimuli with peaks (and lows) characterized by a 20 Hz difference were compared, and the stimuli used for the primary (peak) continuum did actually not differ either from the highest stimuli nor from the lowest ones. The import of slope on question identification was also tested, and no effect was found.

3.2. Syllable-structure effects on perception

In §2.1, I presented the results of a production study testing acoustic alignment differences between open and closed sylla-

[1] Other hypotheses were explored in the study, but I shall concentrate only on those ones here.

[2] Since each plateau stimulus corresponds to 3 steps of peak stimuli, the first plateau stimulus (T0) is compared to peak stimulus 3 (T3), and so forth.

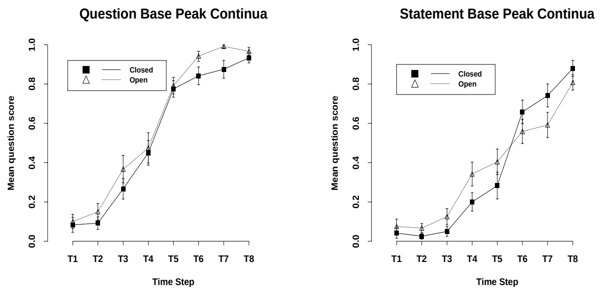

Figure 8: *Mean question scores for the open and closed question base continua (left) and for the open and closed statement base continua (right), for all subjects. (From D'Imperio, 2000).*

bles in Neapolitan Italian and claimed that the results supported the hypothesis of an invariant target alignment relative to the left edge of the syllable. As a consequence, one could ask whether relative or absolute target alignment is employed by the listener to differentiate between a question and and a statement pitch accent. If tonal targets are aligned relative to the left edge of the syllable (either syllable or vowel onset), we expect the difference in duration induced by syllable structure to have no effect on perceived target. If, on the other hand, perceived tonal target location is computed relative to the entire stressed syllable or stressed vowel duration, we expect to find such an effect.

Hence, in a subsequent experiment in [5], I tested perceived target alignment by creating two resynthesized continua for each utterance type (i.e., two for questions and two for statements) whose base would be, in one case, an utterance with a closed stressed syllable and, in the other, an utterance with an open stressed syllable. Similarly to the results for peak stimuli presented in §3.1, notice from Figure 8 an orderly increase from statement to question judgments as a function of timing. As both panels show, at early locations within the timing continuum, such as T1 and T2, statement responses were largely dominant, while at the opposite end of the continuum question responses were largely dominant.

Crossover points were also recorded for each individual function relative to the syllable structure manipulation. Such values were used as the basis for a two-way analysis of variance, with Open/Closed and Question/Statement as factors. The hypothesis tested was that differences in syllable structure would not produce a boundary shift. The statistic results confirmed the hypothesis, thus suggesting for perception what was already found for production, that is an invariant computation of target alignment from the left edge of the syllable. Once more, we find a discrepancy between our results and those for other languages. Specifically, [17] found that the presence of voiced segments in the coda (as well as in the onset) shifts the PSE (Point of Subjective Equality) in the perception of the Dutch flat hat pattern. This leads us to consider the impact of language-specific constraints on alignment perception.

4. Language-specific and universal constraints on alignment

We saw above that there is much about tonal alignment that is predictably influenced by the phonetic context. On the other, though, small differences of alignment can apparently also create clearly perceptible differences of meaning. Also, different languages appear to align tones towards a certain edge of the structural domain they are associated to, which justifies also the need to keep alignment and association separate. For instance, it appears that targets for starred tones, such as H*, in English tend to occur towards the right-edge (i.e., "right-peripheral" alignment) of the stressed syllable, while the opposite seems to be true for starred tones of Standard Swedish. [3]

Recent findings seem to reflect the contribution of language-specific intonation patterns towards our ability to perceive specific tonal configurations. Ad additional goal of [5] was to perform perceptual experiments with more than one language so that the contribution of language-specific information could be factored out.

4.1. Universals of tonal perception

The results of a perception experiment [5] in which American English subjects listened to the same stimuli used for the experiment presented in §3 showed that, when it comes to identifying a Neapolitan question vs. a statement, American English listeners (who had never learned Neapolitan Italian before) are capable to accomplish the task, and can be trained for such a purpose rather rapidly. That is, American listeners can employ a subtle alignment contrast, similar but not identical to the one exploited by their intonation system.

However, a later crossover for the peak continuum was found for Americans (between stimuli 5 and 6) which might be due to the fact that the American listeners were responding in terms of the alignment characteristics of the early-rise (L+H*) vs. late-rise (L*+H) contrast proper to their intonational phonology. In fact, the H target for English L*+H is aligned much later than for the Neapolitan L*+H. That is, in American English, the peak is usually timed beyond the stressed syllable boundary, while the L valley is generally located within the stressed vowel. Late category boundary was also found in [15], which investigated the perception of L+H* vs. L*+H in American English.

On the other hand, like for Italian listeners, plateau stimuli were identified mostly as questions, which lends additional support to the idea that this effect has a "universal" nature. That is, the wider "peak" had the effect of displacing perceived target location for the rise-fall in the time dimension. As for Neapolitans, questions scores for plateau stimuli and peak stimuli timed at plateau offset were not different. Such an effect cannot be explained in terms of parsing a language-specific tonal structure, since obviously American English and Italian do not share the same phonological system neither the same phonetic implementation mechanisms.

4.2. Articulatory evidence for universal constraints: articulatory "landmarks"?

One cannot exclude the fact that it might be not possible to demonstrate a universal role of tonal alignment in any randomly chosen language, because, for instance, alignment has a different role in stress-accent languages (where an accent peak or valley marks the location of a metrically strong syllable) vs. other language types. The same is true of many other phonetic commonalities between languages, from coarticulation to prosodic cues to phrasing. Nevertheless, [23] has recently underlined the

[3] Actually, in Standard Swedish there seems to be sociolinguistic pressure enforcing this kind of alignment, since non-standard dialects align starred tones later.

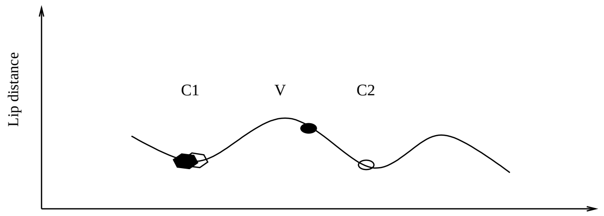

Figure 9: *Schematic representation of Euclidean lip distance with tone targets for question and statement LH rises (filled polygon = statement L; empty polygon = question L; filled circle = statement H; empty circle = question H.*

role of articulatory (hence universal) constraints in determining cross-linguistic alignment regularities. However, articulatory data (currently being analyzed) for Neapolitan seem to go counter to the hypothesis of a strict synchronization of tonal targets and syllable boundaries (onset or offset). F_0 targets were labeled relative to the measured Euclidean distance of the lips for CVCV vs CVCCV target words (in which the consonants were all bilabials) for questions and statements and three different speech rates. Preliminary results appear to show a difference between the implementation of the two rises (L+H* and L*+H) in Neapolitan. While the initial L target is indeed synchronized with syllable onset (i.e., minimum lip distance in C1) in both categories, the H target is instead quite consistently reached at syllable offset (i.e., minimum lip distance for C2) only for the L*+H of questions. Note that there is no difference in F_0 height between question and statement H peaks. A schematic representation is shown in Figure 9. An additional hypothesis might then be that, rather than being blindly determined by articulatory constraints, tonal targets are synchronized with articulatory events on the basis of linguistic contrast.

5. Conclusions

Aligning tones and recovering this alignment on the part of humans involves a number of characteristics which may be captured by a universal description which awaits for each language a language-specific implementation. Research is needed to find the universal thread binding the language-specific data, by shedding light on the the interaction between universal (perceptual and articulatory) and language-specific (phonological and sociolinguistic) constraints.

6. Acknowledgments

The author wishes to thank K. Munhall and N. Nguyen for the analysis of the articulatory data presented in §4.2.

7. References

[1] A. Arvaniti; D.R. Ladd; I. Mennen, 1998. Stability of tonal alignment: The case of Greek prenuclear accents. *Journal of Phonetics*, 26:3–25.

[2] G. Bruce, 1977. *Swedish Word Accents in Sentence Perspective*. Lund: Gleerups.

[3] J. Caspers; V. J. van Heuven, 1993. Effects of time pressure on the phonetic realisation of the Dutch accent-lending pitch rise and fall. *Phonetica*, 50:161–171.

[4] M. D'Imperio, 1995. Timing differences between prenuclear and nuclear pitch accents in Italian. *JASA*, 98(5):2894.

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

[6] M. D'Imperio, 2001a. Focus and tonal structure in Neapolitan Italian. *Speech Communication*, 33(4):339–356.

[7] M. D'Imperio, 2001b. Tonal alignment, scaling and slope in italian question and statement tunes. In *Proceedings of Eurospeech '01*, Aalborg, Denmark.

[8] M. D'Imperio, 2002. Italian Intonation: An Overview and some Questions. *Probus*, 14(1):37–69.

[9] M. D'Imperio; D. House, 1997. Perception of questions and statements in Neapolitan Italian. In *Proceedings of Eurospeech'97*, volume 1, pages 251–254, Rhodes, Greece.

[10] M. Grice; M. D'Imperio; M. Savino; C. Avesani, in press. Towards a strategy for labelling varieties of Italian. In S.-A. Jun, editor, *Prosodic Typology and Transcription: A Unified Approach*. Oxford: OUP press.

[11] K. Kohler, 1987. Categorical pitch perception. In *Proceedings of the XIth International Congress of Phonetic Sciences*, volume 5, pages 331–333, Tallin, Estonia.

[12] D. R. Ladd; D. Faulkner; H. Faulkner; A. Schepman, 1999. Constant "segmental anchoring" of f_0 movements under changes in speech rate. *JASA*, 106(3):1543–1554.

[13] D. R. Ladd; I. Mennen; A. Schepman, 2000. Phonological conditioning of peak alignment in rising pitch accents in Dutch. *JASA*, 107(5):2685–2695.

[14] I. V. Nábělek; I. J. Hirsh, 1969. On the discrimination of frequency transitions. *JASA*, 45:1510–1519.

[15] J. B. Pierrehumbert; S. Steele, 1987. How many rise-fall-rise contours? In *Proceedings of the XIth International Congress of Phonetic Sciences*, Tallin, Estonia.

[16] P. Prieto; J. P. H. van Santen; J. Hirschberg, 1995. Tonal alignment patterns in Spanish. *Journal of Phonetics*, 23:429–451.

[17] T. Rietveld; C. Gussenhoven, 1995. Aligning pitch targets in speech synthesis: effects of syllable structure. *Journal of Phonetics*, 23:375–385.

[18] M. Rossi, 1971. Le seuil de glissando ou seuil de perception des variations tonales pour les sons de la parole. *Phonetica*, 23:1–33.

[19] K. Silverman; J. B. Pierrehumbert, 1990. The timing of prenuclear high accents in English. In J. Kingston and M. E. Beckman, editors, *Papers in Laboratory Phonology I*, pages 71–106. Cambridge: Cambridge University Press.

[20] J. P. H. van Santen; J. Hirschberg, 1994. Segmental effects on timing and height of pitch contours. In *Proceedings of the International Conference on Spoken Language Processing*, volume 2, pages 719–722, Yokohama, Japan.

[21] G. Ward; J. Hirschberg, 1985. Implicating uncertainty: The pragmatics of fall-rise intonation. *Language*, 61:747–776.

[22] Y. Xu, 2001. Fundamental frequency peak delay in Mandarin. *Phonetica*, 58:26–52.

[23] Yi Xu, 2002. Articulatory constraints on tonal alignment. In *Proceedings of SP2002*, Aix-en-Provence, France.