Phonological Conditioning of F0 Target Alignment

D. R. Ladd
University of Edinburgh, Scotland
E-mail: bob.ladd@ed.ac.uk

ABSTRACT

Recent research on the alignment between F0 “targets” (e.g. maxima and minima) and the segmental string show that F0 targets exhibit “segmental anchoring”, i.e. they co-occur with specific segmental landmarks such as the onset of a stressed syllable. Undoubtedly there are principles of speech production at work in this phenomenon, but the purpose of the paper is to demonstrate that phonological structure must also form part of our eventual explanation for segmental anchoring. In studies of both Dutch and Japanese, phenomena of alignment are observed that seem to preclude phonetic explanation (in terms of e.g. segment duration) and to require reference to elements of phonological structure (e.g. distinctions of vowel length, sub-syllabic structure).

1. INTRODUCTION

In some intuitively clear way, F0 features such as tone and accent belong with specific elements of the segmental string: Chinese tones go with syllables (or possibly syllable rhymes), English pitch accents go with stressed syllables, Japanese word accents go with a specific mora, etc. This loose “belonging together” is known in the autosegmental literature as *association*. However, it has been clear for some time that the precise temporal coordination or *alignment* of F0 events with segmental events is quite complex, and does not follow straightforwardly from the mere fact of association. Two types of complications appear quite common:

Peak delay: The peak of a pitch accent may be aligned outside (usually after) the stressed syllable with which it is intuitively associated (e.g. [24]).

Contrasts of alignment: Some languages seem to exhibit linguistically distinctive contrasts of alignment, e.g. earlier and later alignment of a rise or fall relative to the associated segmental element (Bruce [4] on Swedish, Pierrehumbert [21] on English).

Three general solutions have been proposed to deal with these complications:

The phonology-vs.-phonetics solution: phonologically there is categorical “association” of F0 events with segmental elements, but phonetically the “alignment” may vary gradiently [15,11] as a function of various phonetic factors.

The “starred tone” solution: in a pitch event that consists phonologically of more than one tone, one of the component tones can be analyzed as central or “starred” [21]; only the starred tone is associated with the relevant segmental element, while the unstarred tone merely “leads” or “trails”.

The “secondary association” solution: Pierrehumbert and Beckman [22] propose that the initial H tone of the Japanese prosodic word has a primary association to the initial boundary but has a *secondary association* to a specific mora. The notion of secondary association has since been applied to data from a number of European languages [10,9,3]. In all cases the evidence for secondary association consists principally of phonetic alignment data.

These three solutions are not entirely mutually exclusive – in particular, the phonology-vs.-phonetics solution is compatible with either of the two phonological accounts – but they do make somewhat different implicit predictions about the kinds of alignment behavior we should observe . Roughly speaking, the phonology-vs.-phonetics solution implies few constraints on possible patterns of alignment or on the kinds of factors that affect alignment; the starred tone solution suggests that there can be at most two patterns of alignment for a bitonal accent; and the secondary association solution suggests that there might be more than two patterns of alignment but that the patterns we find will be categorically rather than gradiently different from one another.

Early work on F0 alignment emphasized the interacting gradient effects of several phonetic factors (stress clash, proximity to boundaries, “time pressure” from intrinsic segment duration or speech rate, etc.) on the alignment of accent peaks [24,5,23]. These phonetic effects are certainly important for a full understanding of F0 alignment. However, a more recent line of work, beginning with a paper on Modern Greek by Arvaniti et al. [1], starts from the discovery that when factors of time pressure are excluded (i.e. when the F0 features under investigation are as far as possible from boundaries, other accented syllables, etc.), then alignment of F0 targets is consistently governed by what they call “segmental anchoring”: F0 targets are aligned in time with specific points in the segmental string. This paper summarizes some of the findings that have followed from Arvaniti et al.’s work.

2. SEGMENTAL ANCHORING

A. Basic finding

Arvaniti et al. found, unexpectedly, that the beginning and

 ISBN 1-876346-48-5

the end of rising prenuclear pitch accents in Greek are aligned in time with the end of the pretonic syllable and the beginning of the posttonic vowel, respectively. The duration of the rise is almost entirely a function of the time interval between these two segmental landmarks, i.e. determined by the intrinsic durations of the segments that happen to accompany the rise. The "scaling" (F0 level) of the beginning and the end of the rise is unaffected by its duration; thus the slope of the rise is also a function of the interval between the two segmental landmarks. This finding is difficult to reconcile with most phonetic models of pitch contours (e.g. Fujisaki's [7] or 't Hart et al.'s [12], even Taylor's [25] to some extent), which assume that slope and/or duration are the most appropriate ways of characterizing accent types, but is strongly compatible with an autosegmental description of pitch movements, in which a pitch movement is the phonetic manifestation of a sequence of "tones" aligned with the segmental string in well-defined ways. The alignment of one tone can be at least partially independent of another, even if the two tones are part of the same tonal configuration.

More importantly for the point of this paper, as Arvaniti et al. [2] have pointed out, the Greek findings suggest that a starred-tone account of alignment is unable to account for significant aspects of language-specific alignment facts. Other problems with the starred-tone account have independently been discussed by Grice [8].

B. Some further developments and related findings

A considerable amount of subsequent work has shown that segmental anchoring is found in other languages as well and that segmental anchoring is relevant to speech perception. Among the relevant findings are the following.

1. *Consistent segmental anchoring under changes of speech rate in English*: English rising prenuclear accents remain anchored to segmental landmarks regardless of speech rate. Slope and duration are adjusted to keep the beginning and the end of the accentual F0 rise aligned with their respective segmental anchors as segment durations decrease or increase with rate [16].

2. *Consistent alignment of between-accent F0 valleys in English:* The F0 valley between accents is aligned with the beginning of the second accented syllable. This means that in potentially ambiguous phrases like *Norma Nelson* and *Norman Elson,* the alignment of the F0 valley is affected by the syllable membership of the ambiguous consonant. Effects on perception can be produced by manipulating the alignment of the F0 valley [18].

3. *Stress effects on alignment of "phrase accent" in Dutch*: The L of a Dutch falling-rising nuclear contour (L+H* L H%) aligns with the most prominent postnuclear syllable. In falling-rising questions like *Woon je in X?* 'Do you live in X?', where X is a town name with primary stress on the first syllable, speakers consistently align the postnuclear accentual F0 minimum according to the location of any postnuclear secondary stress: when X is stressed like [ˈdɔmələ] the F0 minimum is aligned at the beginning of the final syllable, but when X is stressed like either [ˈnɛjˌmeːxə] or [ˈhʏməˌloː] the F0 minimum is aligned in the middle of the secondary-stressed vowel [19].

4. *Consistent small differences of alignment between languages and between varieties of the same language*: In German, rising prenuclear accents are aligned consistently later than those in English and Dutch, and within German, such accents are aligned consistently later in Southern varieties than in Northern varieties. The effects are small but significant [3].

5. *Regularities in the alignment of Chinese lexical tone contours with syllables:* The end of the rising contour for Mandarin second tone is closely coordinated with the end of the syllable, regardless of speech rate and syllable composition (specifically the presence or absence of a nasal coda) [26].

6. *Perceptibility of small alignment differences:* In an Edinburgh MSc project, Chorianopoulou [6] synthesized Greek sentences with experimentally modified F0 contours. F0 rises were aligned (a) in a way consistent with Arvaniti et al.'s findings [1], and also (b) in several ways that deviate from the Arvaniti et al. means (e.g. beginning of rise 60 ms too early or 60 ms too late). Chorianopoulou obtained naturalness preferences from 22 Greek listeners in a forced-choice experiment. Overall, there was a small but definite preference for the "Arvaniti" versions of the sentences, suggesting that listeners can perceive the difference and prefer the version that matches production data. A follow-up experiment is in progress.

All the findings just cited suggest that the phenomenon of segmental anchoring is genuine and needs to be incorporated into our understanding of how F0 contours are coordinated with the segmental string. However, the reasons for the consistency of segmental anchoring are not obvious. It seems implausible that the speaker's motor plan is so precise that it can vary the placement of F0 targets in response to fine differences of secondary association. It is also difficult to reconcile the evidence for fine differences of secondary association with the apparently gradient nature of so many effects on alignment. It seems much more likely that segmental anchoring is somehow a consequence of a more general coordination between F0 changes and units of speech production such as syllables (this is the view taken by Xu [26] on the basis of his Chinese data).

At the same time, an explanation in terms of speech production mechanisms makes it difficult to account for differences between languages such as those between Greek and English or Dutch. Recall that the beginning of a prenuclear rising accent in these languages is anchored to the onset of the accented syllable, but the end of the rise is considerably later in Greek (following unstressed vowel) than in English or Dutch (following consonant). If we are looking for production-related explanations, this difference would seem to imply the existence of different articulatory mechanisms for producing rising accents in different languages. The phonological alternative (viz., that the H of a L+H accent has a different secondary association in Greek than in English or Dutch) puts the difference between the two sound systems in the phonological description, and simply leaves the question of production mechanisms open.

It seems likely that the ultimate explanation for segmental anchoring will involve some combination of phonological differences between languages and cross-linguistically different patterns of phonetic realization. The remainder of the paper discusses two specific cases that not only demonstrate the relevance of phonological structure to alignment but also show us how complex the interaction between structural and phonetic factors may be.

3. PHONOLOGICAL EFFECTS IN ALIGNMENT

A. Phonological vowel length in Dutch

Perhaps the clearest demonstration of the relevance of phonological factors in alignment comes from a study by Ladd, Mennen and Schepman [17]. They showed that the beginning of prenuclear rising accents in both Dutch and English is aligned as in Greek, but the alignment of the end (i.e. peak) of the rise depends on whether the vowel is phonologically long or short (tense or lax): the peak accompanying a long vowel is late in the vowel, but that accompanying a short vowel is in the middle of the following consonant. (The published results in [17] deal only with Dutch. A similar phenomenon is apparently present in English but the details are considerably more complex because the phonology of vowel length is more complex in English than in either Dutch or German.)

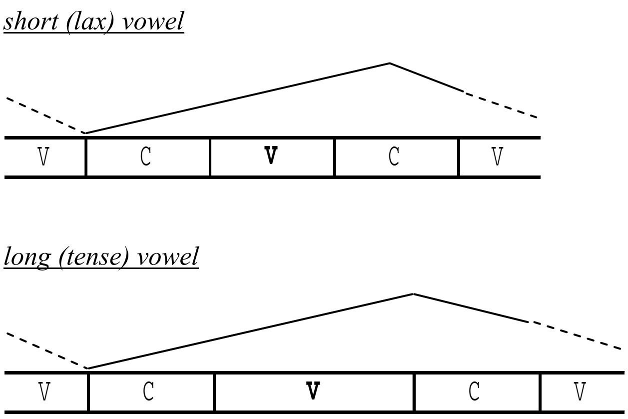

Figure 1. A possible explanation for the effect of vowel length on alignment in [17]: the duration of the F0 rise is roughly constant but the accompanying segments differ.

There are two fairly obvious possible explanations for the difference in alignment, one based on low-level phonetic factors and one on a presumed difference of phonological structure. If it turns out that the duration of the rise is essentially constant, then the difference between long and short vowels will arise from the fact that at the end of the rise the speaker is still articulating a long vowel but has finished articulating a short vowel. This is shown in Fig. 1. However, there is a plausible structure-based alternative explanation, namely that the end of the rise is seeking to align with the right edge of the syllable. A single word-medial consonant following a long vowel in Dutch will always be syllabified as the onset of the next syllable, whereas a single word-medial consonant following a short vowel may be regarded as ambisyllabic.

A fact about Dutch makes it possible to distinguish between these two possible explanations. It is known [20] that Dutch “long” /i/ and “short” /ɪ/ are essentially identical in phonetic duration and differ only in quality (and also, of course, in the structural and distributional properties that distinguish “long” and “short” vowels generally). This means that if the “phonetic” explanation is correct there should be no difference in alignment between /i/ and /ɪ/. If the structural explanation is correct, then we should still see a difference in alignment. The latter is exactly what Ladd et al. found: there was a small (11 ms) but statistically significant difference of alignment between the two cases, meaning that a simple duration-based explanation is not adequate. However, it is worth noting that the size of the alignment difference was smaller than in the case of other long-short vowel pairs, which suggests that considerations of “time pressure” are still relevant, interacting with structural factors.

B. Sub-syllabic structure in Japanese

In another study, presented at this conference, Ishihara [14] finds effects on alignment in Tokyo Japanese that similarly seem to require reference to phonological structure and not merely to phonetic duration. Ishihara compares the alignment of first-syllable accent peaks in words beginning CVCV, CVNCV, and CVVCV, where “N” is the moraic nasal and VV indicates a two-mora vowel or diphthong. In all cases the consonants (C) are nasals. Comparing CVCV and CVNCV, Ishihara finds that the F0 peak is aligned before the nasal in CV|NCV but after the nasal in CVC|V. He also finds that the peak is aligned partway through a long vowel. i.e. CV|VCV (in all cases the | indicates the alignment of the peak). This pattern of results, he argues, suggests that the peak is aligned at the beginning of the “moraic” part (in the sense of Hayes [13]) of the second mora. However, a duration-based solution cannot be ruled out, because Ishihara also finds that the vowel preceding a moraic nasal is considerably longer than the vowel before a CV mora, and a long vowel is more than twice as long as a short vowel. This means that the duration of the F0 rise is much more similar in these three cases than symbolic diagrams like CV|NCV would suggest.

ISBN 1-876346-48-5 © 2003 UAB

4. CONCLUSION

Overall, the work reviewed here points to the relevance of phonological structure in conditioning the patterns of F0-segmental alignment. At the same time, both the studies discussed in more detail in section 3 show how complex the phenomenon is, and show that it is not a matter of choosing between structural and durational explanations, but of reconciling and integrating them. I believe it will be some time before we make sense of all the interactions.

5. ACKNOWLEDGEMENTS

Parts of this work were supported by the UK Economic and Social Research Council under grants R000-23-5614 and R000-23-7447 to Edinburgh University.

6. REFERENCES

[1] Arvaniti, A., Ladd, D. R., Mennen, I. (1998). Stability of Tonal Alignment: the case of Greek prenuclear accents. *Journal of Phonetics* 26: 3-25.

[2] Arvaniti, A., Ladd, D. R., Mennen, I (2000). What is a starred tone? Evidence from Greek. In M. Broe and J.B.Pierrehumbert (eds.), *Papers in Laboratory Phonology V,* Cambridge University Press.

[3] Atterer, M., and Ladd, D. R. (forthcoming) On the phonetics and phonology of "segmental anchoring" of F0: evidence from German. To appear, *Journal of Phonetics.*

[4] Bruce, G. (1977). *Swedish Word Accents in Sentence Perspective.* Lund: Gleerup.

[5] Caspers, J., and van Heuven, V. (1993). Effects of time pressure on the phonetic realization of the Dutch accent-lending pitch rise and fall. *Phonetica* 50, 161-171.

[6] Chorianopoulou, E. (2002). Evaluating prosody prediction in synthesis with respect to Modern Greek prenuclear accents. MSc dissertation, Edinburgh Univ.

[7] Fujisaki, H. (1983). Dynamic characteristics of voice fundamental frequency in speech and singing. In Peter F. MacNeilage, (ed.), *The Production of Speech*. New York: Springer-Verlag, 39-55.

[8] Grice, M. (1995). Leading Tones and Downstep in English. *Phonology* 12: 183-233.

[9] Grice, M., Ladd, D. R., Arvaniti, A. (2000). On the place of "phrase accents" in intonational phonology. *Phonology* 17, 143-186.

[10] Gussenhoven, C. (2000). The boundary tones are coming: on the nonperipheral realization of boundary tones. In M. Broe and J. Pierrehumbert (eds.), *Papers in Laboratory Phonology V* (Cambridge: Cambridge University Press), pp. 132-151.

[11] Gussenhoven, C. (2003). On tune-text relations. In proceedings of this conference.

[12] 't Hart, J., Collier, R., Cohen, A. (1990). *A perceptual study of intonation: An experimental-phonetic approach*. Cambridge : Cambridge University Press.

[13] Hayes, B. (1989). Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20, 253-306.

[14] Ishihara, Takeshi (2003). A phonological effect on tonal alignment in Tokyo Japanese. In proceedings of this conference.

[15] Ladd, D. R. (1983). Phonological Features of Intonational Peaks. *Language* 59: 721-759.

[16] Ladd, D. R., Faulkner, D., Faulkner, H., Schepman, A. (1999). Constant "segmental anchoring" of F0 movements under changes in speech rate. *JASA* 106, 1543-1554.

[17] Ladd, D. R., Mennen, I., Schepman, A. (2000). Phonological conditioning of peak alignment of rising pitch accents in Dutch. *JASA* 107, 2685-2696.

[18] Ladd, D. R., and Schepman, A. (2003). "Sagging transitions" between high pitch accents in English: experimental evidence. *Journal of Phonetics* 31, 81-112.

[19] Lickley, R., Schepman, A., Ladd, D. R. (2002). Lab speech is real speech: the case of Dutch falling-rising questions. Paper presented at BAAP Colloquium, Newcastle. Written version in preparation, to be submitted to *Language and Speech*.

[20] Nooteboom, S. G., and Slis, I. H. (1972). The phonetic feature of vowel length in Dutch. *Language and Speech* 15, 301-316.

[21] Pierrehumbert, J. B. (1980). *The Phonology and Phonetics of English Intonation*. PhD Dissertation, MIT.

[22] Pierrehumbert, J. B., and Beckman, M.E. (1988). *Japanese Tone Structure.* Cambridge MA: MIT Press.

[23] Prieto, P., van Santen, J., Hirschberg, J. (1995). Tonal alignment patterns in Spanish. *Journal of Phonetics,* 23, 429-451.

[24] Silverman, K., and Pierrehumbert, J. B. (1990). The timing of prenuclear high accents in English. In J. Kingston and M. Beckman (Eds.) *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech* (Cambridge: Cambridge University Press), pp. 71-106.

[25] Taylor, P. A. (2000). Analysis and synthesis of intonation using the Tilt model. *JASA* 107, 1697-1714.

[26] Xu, Y. (1998) Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica* 55, 179-203.