Segmental anchoring of pitch movements: Autosegmental association or gestural coordination?

D. Robert Ladd

Arvaniti, Ladd and Mennen (1998) reported a phenomenon of 'segmental anchoring': the beginning and end of a linguistically significant pitch movement are anchored to specific locations in segmental structure, which means that the slope and duration of the pitch movement vary according to the segmental material with which it is associated. This finding has since been replicated and extended in several languages. One possible analysis is that autosegmental tones corresponding to the beginning and end of the pitch movement show secondary association with points in structure; however, problems with this analysis have led some authors to cast doubt on the 'hypothesis' of segmental anchoring. I argue here that segmental anchoring is not a hypothesis expressed in terms of autosegmental phonology, but rather an empirical phonetic finding. The difficulty of describing segmental anchoring as secondary association does not disprove the 'hypothesis', but shows the error of using a symbolic phonological device (secondary association) to represent gradient differences of phonetic detail that should be expressed quantitatively. I propose that treating pitch movements as gestures (in the sense of Articulatory Phonology) goes some way to resolving some of the theoretical questions raised by segmental anchoring, but suggest that pitch gestures have a variety of 'domains' which are in need of empirical study before we can successfully integrate segmental anchoring into our understanding of speech production.

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 phonological literature as 'association'. However, it has been clear for some time that the precise temporal coordination or 'alignment' of F0 events with segmental events does not follow straightforwardly from the mere fact of association, and within the past decade alignment has become a major topic of investigation. Much of this recent interest has been sparked by what Ladd *et al.* (1999) called 'segmental anchoring'. This term refers to the situation, first reported by Arvaniti *et al.* (1998), in which both the beginning

Rivista di Linguistica 18.1 (2006), pp. 19-38 *(ricevuto nel settembre 2006)*

and the end of a pitch movement appear to be independently aligned with specifiable points in the segmental string. The goal of this paper is two-fold: to clarify the definition of segmental anchoring on the basis of what we know now, and to consider where the next empirical and theoretical challenges lie.

2. Segmental anchoring

2.1. Backdrop

Three studies from the early 1990s set the backdrop for the discussion here. The first two of these (Silverman & Pierrehumbert 1990 and Prieto *et al.* 1995) were based on informal observations of 'peak delay', in which accentual F0 peaks sometimes seem to be aligned phonetically after the end of the syllable with which they are phonologically associated. Both studies specifically aimed at identifying the interacting factors that affect peak delay and more generally at understanding the alignment of F0 peaks relative to accented syllables. Silverman & Pierrehumbert, working on American English, demonstrated that the alignment of an accentual pitch peak is strongly affected by the distance to the end of the accented word and by the distance to the accented syllable of the following word: the greater the distance to the word boundary and/or the next accent, the later in the accented syllable the peak is aligned. Prieto *et al.*, working on Mexican Spanish, found similar kinds of effects, but contested the importance that Silverman & Pierrehumbert had attached to upcoming boundaries and accents. They developed a quantitative model of peak alignment in which the various aspects of the segmental composition of the accented syllable itself – specifically, the duration of the onset, the vowel, and any coda – are the most influential factors. Importantly for the topic of segmental anchoring, they also noted that the beginning of the F0 rise on the accented syllable begins at the beginning of the syllable, irrespective of the alignment of the peak; Silverman & Pierrehumbert did not report the alignment of the beginning of the rise.

About the same time as these studies of 'peak delay' were being carried out, Caspers & van Heuven (1993) investigated the effect of 'time pressure' on pitch movements in Dutch. Their experiments were similar to those of Silverman & Pierrehumbert and of Prieto *et al.*, in that they manipulated phonological vowel length, speech rate, and distance to the next pitch movement and studied how these affected the phonetic properties of rising and falling pitch movements. However, their goals and assumptions were subtly different. First, they took slope, duration, and excursion size as the primary phonetic properties of pitch movements, and dealt only secondarily with alignment. More importantly, they assumed that speakers would aim to preserve the essential or linguistically significant properties of a pitch movement when the physical constraints of time pressure prevented them from realizing it fully, and their goal was to shed light on what pitch movements' essential properties are. They found a complex set of effects on slope, duration, and excursion size, apparently different for rises and falls, and did not arrive at any general principles. However, like Prieto *et al.*, they also noted that irrespective of the experimental manipulations the beginning of an accentual pitch rise was consistently aligned with the beginning of the accented syllable.

2.2. The basic finding of segmental anchoring

The first report of segmental anchoring – in the sense of temporal coordination of both the beginning and the end of a pitch movement with the segmental string – came from a study of the alignment of prenuclear rising pitch accents in Modern Greek (Arvaniti, Ladd & Mennen 1998 [henceforth ALM]). ALM originally accepted the assumption that the alignment of the peaks of these accents would vary with time pressure. They knew from earlier work (Arvaniti & Ladd 1995) that the beginning of the accentual pitch rise in Greek prenuclear accents is consistently aligned with the beginning of the accented syllable (exactly as found for Mexican Spanish by Prieto *et al.* and for Dutch by Caspers & van Heuven) and consistently scaled at the same F0 level for a given speaker. They also knew from impressionistic observations that Greek prenuclear accent peaks are frequently aligned after the end of the accented syllable. They therefore assumed that the beginning of the rise is the manifestation of an autosegmental L tone associated with the accented syllable (possibly a 'starred tone'), and they wanted to see if the alignment of the end of the rise (presumably an 'unstarred' H tone) would respond systematically to time pressure.

In their experiments, ALM varied the amount of time between one prenuclear rise and the next, expecting to find that this variation would affect the extent to which the peak could be aligned past the end of the accented syllable. As it happened, however, their experimental manipulations did not seem to have any consistent effect on alignment at all, except when the pitch accent was immediately followed by another accent or a word boundary. Instead, in most cases, both the beginning and the end of the pitch rise were consistently aligned in time relative to identifiable landmarks in the segmental string. Specifically, ALM replicated Arvaniti & Ladd's finding that the rise begins at the beginning of the stressed syllable, and found to their surprise that the rise consistently ends early in the posttonic vowel. Crucially, these alignments are maintained regardless of the distance between the alignment points. That is, the duration of the rise is almost entirely a function of the time interval between the two segmental landmarks – the beginning of the stressed syllable and the beginning of the posttonic vowel. Importantly, moreover, the 'scaling' (F0 level) of the beginning and the end of the rise is unaffected by its duration; longer rises do not have greater overall pitch excursions. This means that the interval between the two segmental landmarks also determines the SLOPE of the rise. These findings are summarized in the idealized diagrams in Figure 1.

Fig. 1. Idealized diagrams of the segmental anchoring of the Greek prenuclear rise with different segmental material. In the left-hand panel the rise spans the short-duration stretch [ðit] and in the right-hand panel it spans the longer sequence [ɾɛmv], but the starting and ending F0 levels are the same in both cases.

As ALM point out, the findings summarized in Figure 1 are difficult to reconcile with earlier phonetic models of pitch contours (e.g. Fujisaki 1983, 't Hart *et al.* 1990, even Taylor 2000 to some extent), which assume that slope, duration, and excursion size are the most appropriate phonetic properties for characterizing pitch movement types, and which treat coordination with the segmental string almost as an afterthought. ALM's finding clearly suggests that slope and duration are NOT identifying characteristics of pitch movements, but rather that slope and duration depend on the scaling and alignment of 'tonal targets'. The beginnings and ends of pitch movements, that is, have phonetic characteristics of their own; these characteristics can be modified by time pressure and other phonetic factors, but what pitch movements are 'trying' to do in the absence of such factors is to align with the segmental string in a specified way, not to attain a specified slope or duration. 'Time pressure' on pitch movements is not only a matter of physical limits on the rate of pitch change, but of adapting the rate of pitch change to structurally defined constraints. To put it in the terms that motivated Caspers & van Heuven's work, the defining phonetic features of the prenuclear accentual rise in Greek are the level of its beginning and ending points – the excursion size, in a sense – and the way those beginning and ending points are coordinated with the segmental string.

2.3. Further developments and related findings

Surprising though it was to many people, the finding of segmental anchoring has since been replicated for other languages and extended in various ways. This section summarizes a number of empirical findings that complement or build on the original ALM study. The brief summaries are set out in short numbered subsections for ease of subsequent reference.

2.3.1. Segmental anchoring under changes of speech rate in English (Ladd, Faulkner, Faulkner & Schepman 1999). English rising prenuclear accents remain anchored to segmental landmarks regardless of speech rate: slope and duration are adjusted to keep beginning and end of accentual F0 rise aligned with their respective segmental anchors as segment durations decrease or increase with rate. Caspers & van Heuven found a similar effect for Dutch, though of course they did not interpret it in exactly these terms. This does not mean that speech rate never has any influence on alignment; some studies (e.g. Silverman & Pierrehumbert 1990; Ishihara 2006) have suggested that fast rate does have small effects. But on the whole it appears that rate modifications affect the slope and duration of pitch movements rather than their alignment with the segmental string.

2.3.2. Effects of phonological vowel length (tenseness) on segmental anchoring in Dutch and English (Ladd, Mennen & Schepman, 2000; Ladd, Schepman, White, Quarmby & Stackhouse, work in progress). Ladd *et al.* (2000) found that the beginning of prenuclear rising accents in Dutch is aligned as in Greek, but the alignment of the end of the rise (i.e. the peak) depends on whether the vowel is phonologically long or short (tense or lax). Specifically, the peak accompanying a long vowel is late in the vowel, but that accompanying a short vowel is late in the following consonant. This is not merely an effect of time pressure brought about by differences of vowel duration, as was assumed by Caspers & van Heuven (1993): Dutch 'long' /i/ and 'short' /ɪ/ are essentially identical in phonetic duration and differ only in vowel quality (Nooteboom 1972); nevertheless, a small difference in alignment is still found. Similar effects are found in English (Ladd *et al.* work in progress), although the definition of phonological vowel length is less clear in many varieties of English than it is in Dutch.

2.3.3. Consistent alignment of between-accent F0 valleys in English (Ladd & Schepman 2003; Dilley, Ladd & Schepman 2005). The F0 valley between accents on adjacent words 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. However, in keeping with idea that the valley and the peak are aligned independently, there is no significant effect of the consonant's syllable membership on the alignment of the following accentual peak; accentual rises are shorter and steeper in syllables that begin with a vowel (*Norman Elson*) than in those that begin with a consonant *(Norma Nelson*). This is entirely analogous to ALM's original finding for Greek.

2.3.4. Regularities in the alignment of Chinese lexical tone contours with syllables (Xu 1998, 1999; Xu & Wang 2001; and many other works by Xu and his colleagues). Xu's extensive body of work on alignment in Chinese, which is entirely independent of our own, has yielded a number of findings consistent with the work that builds directly on ALM. The clearest example is his finding that 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). Note, however, that Xu's interpretation of his data is rather different; I return to this point in section 3.1. below.

2.3.5. Phonological factors in the alignment of Japanese accentual H (Ishihara 2003, 2006). The end (or peak) of the F0 rise signalling a word-initial accent in Tokyo Japanese is consistently aligned at the beginning of the 'moraic' part (in the sense of Hayes 1989) of the *second* mora, regardless of whether the second mora is a separate CV syllable or only the second half of a long (CVN or CVV) syllable. However, adjustments to segment durations in the different types of syllable mean that it is more difficult to distinguish between structural and physical effects than in the case of Dutch long and short vowels (section 2.3.2); Ishihara speaks of 'mutual synchronization' of the duration of the segments and the pitch movements.

2.3.6. Consistent differences between nuclear and prenuclear accents (Schepman, Lickley & Ladd 2006; Face 2002; Ladd *et al.* work in progress). In Dutch, Spanish and English, F0 peaks are aligned earlier in nuclear accents than in prenuclear accents, while the leading F0 valley is unaffected. Caspers & van Heuven found essentially the same effect, if we equate 'prenuclear' with their 'Type 1 rise' occurring on its own and 'nuclear' with their sequence of 'Type 1 rise' and 'Type A fall' occurring on the same word. Assuming that this equation is valid, we can say that Caspers & van Heuven interpret the earlier alignment of nuclear peaks as a case of 'time pressure': the peak of the Type 1 rise is pushed earlier by the immediately following Type A fall. A very similar explanation, with 'L phrase accent' (Grice *et al.* 2000) in place of 'Type A fall', is proposed by Schepman *et al.*

2.3.7. Consistent small differences of alignment between languages and between varieties of the same language (Atterer & Ladd 2004; Arvaniti & Garding forthcoming). 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, and they crucially affect both the beginning and the end of the rise (the L tone and the H tone, in autosegmental terms). Similar differences are reported for Southern California and Minnesota varieties of American English by Arvaniti & Garding. Our data on British English (Ladd *et al.* work in progress) suggests that Scottish speakers align pitch accents slightly later than Southern English speakers. Findings of this sort have important implications for the interpretation of segmental anchoring, as we shall see presently.

2.4. Interpreting segmental anchoring

ALM interpreted their original finding as evidence for an autosegmental conception of pitch movements, in which movements are analyzed as sequences of tones (e.g. L*+H). They suggested that INDIVIDUAL TONES (not the pitch accent as a whole) would be anchored to specific places in structure, such as the left edge of the stressed syllable, or an unstressed vowel. This interpretation was strengthened by Ladd, Mennen & Schepman's (2000) finding that phonological vowel length affects alignment in Dutch, because one obvious explanation for that finding is that the end of the pitch movement is aligned with the right edge of the syllable. (Dutch syllables with a long vowel and a single following intervocalic consonant are invariably open; syllables with a short vowel are arguably closed by the following consonant.) Indeed, Xu repeatedly cites this finding by Ladd *et al.* (2000) in support of his assumption that syllable boundaries are universally relevant for the alignment of pitch movements, though Xu emphatically rejects the autosegmental analysis.

However, there are problems with the autosegmental interpretation of segmental anchoring, which have become more apparent with recent work. In an autosegmental model, the obvious way to think about specific patterns of alignment is in terms of the 'secondary association' of tones (Pierrehumbert & Beckman 1988; Gussenhoven 2000; Grice, Ladd & Arvaniti 2000; and especially Prieto, D'Imperio & Gili-Fivela 2005). For example, in the case of the Greek prenuclear accents studied by ALM, the basic association is between the L+H accent AS A WHOLE and the accented SYLLABLE, but if we want to express the details of the independent alignment of the L and the H in our autosegmental representation, we might say that the L is secondarily associated with the left edge of the syllable, and the H is secondarily associated with the following vowel. This is shown in Figure 2.

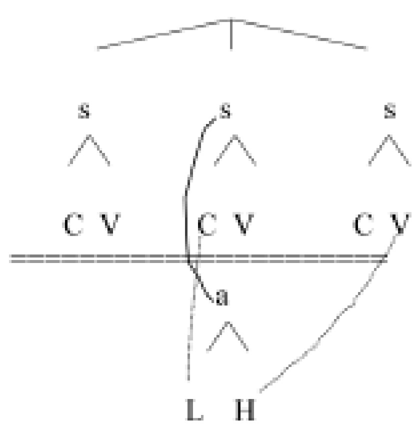

Fig. 2. Hypothetical autosegmental representation of the primary association of an accent to a syllable and the secondary association of the individual tones L and H to individual segments. See Pierrehumbert & Beckman (1988:128) for a similar representation of the secondary association of phrasal tones to specific moras in Japanese.

However, as was pointed out by Atterer & Ladd (2004), if we use secondary association to represent the fine phonetic detail of alignment it leads to a rapid proliferation of distinct phonological representations for subtly different variations of phonetic detail between languages or between language varieties. For example, to express the differences among English, Northern German, and Southern German (section 2.3.7.), we would have to associate the initial L tone of the rise with the left edge of the accented syllable (English), the onset consonant of the accented syllable (Northern German), and the left edge of the accented syllable nucleus (Southern German); we would also have to posit similar differences for the association of the H tone. This seems to make the secondary association analysis less plausible, given that we are apparently dealing with 'the same' basic intonational phenomenon in these different languages or language varieties.

Moreover, even if we ignore the fundamental implausibility of assigning different phonological representations to closely related (and phonetically similar) phenomena in closely related languages, we must acknowledge that such an analysis fails to express two important facts: first, that there appears to be a CONTINUUM of alignment possibilities from English to Southern German; and second, that differences in the alignment of the L and the H are NOT entirely independent of one another, but that the earlier or later alignment somehow applies to the pitch movement as a whole. The first point calls to mind any number of recent demonstrations that some language-specific phonetic detail needs to be described quantitatively, not symbolically (e.g. Zsiga 1997). The second point suggests that the pitch movement as a whole does have some kind of phonetic unity, as maintained by Xu and as assumed by earlier models such as the one developed for synthesizing Dutch speech by 't Hart *et al.* (1990).

These considerations lead us to a somewhat broader view of what segmental anchoring actually is. The central features of this broader view are that it does not depend on identifying the beginnings and ends of pitch movements as autosegmental 'tones', and does not depend on equating phonetic alignment with phonological association. Recall that what was unexpected about ALM's original result – and what has been replicated in repeatedly since then – is the fact that the slope and duration of the pitch movements they studied depend quite precisely on the accompanying segmental structure. In their key experiment, what ALM found was (i) a very strong correlation between the duration of a segmentally defined interval (beginning of stressed syllable to beginning of following unstressed vowel) and the duration of the associated pitch rise, and (ii) a lack of any effect of rise duration on the F0 levels of the beginning and end of the rise. What they did not find was a precise temporal coordination between the endpoints of the pitch movements and PHONOLOGICALLY definable points in the segmental string: the beginning of the rise was aligned on average few milliseconds BEFORE the beginning of the accented syllable, and the end of the rise was aligned 10 or 15 ms AFTER THE ONSET OF the following posttonic vowel. ALM ignored these small quantitative details in their early discussions of their findings, but Atterer & Ladd's work – and the subsequent work by Arvaniti & Garding and by Ladd *et al.* (work in progress) – suggests that ultimately they must be taken into account. The relevant level of description of segmental anchoring is quantitative, not symbolic.

Nevertheless, some authors appear to have taken the autosegmental interpretation as an integral part of the notion of segmental anchoring, and have treated segmental anchoring as a hypothesis about how tones can be associated with the segmental string. Consider, for example, the finding by Prieto & Torreira (2004) that the alignment of H accent peaks in Spanish is affected by whether the accented syllable is open (e.g. CV) or closed (e.g. CVC). Specifically, the peak is aligned at the end of the vowel in an open syllable and some distance into the coda consonant in a closed syllable. If we interpret segmental anchoring as a hypothesis based on autosegmental phonology, then Prieto & Torreira's results pose a problem, because there is no obvious autosegmental representation that allows us to treat 'end of vowel in CV syllable' and 'middle of coda in CVC syllable' in a unified way. (Compare this to Xu's findings on Chinese (sec. 2.3.4), where 'end of syllable' applies equally well to CV and CVC syllables.) From a slightly broader point of view, however, Prieto & Torreira's finding can be viewed as further evidence for segmental anchoring: what they have found is that the duration of a pitch rise depends on the associated segmental material (specifically, whether the syllable is open or closed), and that the F0 levels of the beginning and end of the rise are unaffected by this dependence. These facts are entirely in keeping with ALM's original finding, and with the findings summarized in section 2.3. The difficulty of giving an autosegmental interpretation for the details of their results does not cast doubt on the phenomenon of segmental anchoring, but only on its autosegmental interpretation.

In my view, then, segmental anchoring – properly understood – is not a phonological hypothesis but an empirical finding, which can be summarized in two principal points:

1. In all languages studied so far, the duration of a pitch movement is strongly correlated with the duration of the associated segmental material, while the amount of F0 change (the F0 excursion) is unaffected by such differences. This applies whether the differences in duration are brought about by changes in rate, by intrinsic segment durations, by different syllable structures, or in other ways. In some important sense, the beginnings and ends of pitch movements represent production targets, in a two-dimensional space defined by pitch level and alignment with the segmental string.

2. The details of the correlation in point 1 – the precise alignment of F0 movements with the associated stressed syllable – can vary from language to language and from variety to variety within the same language. These differences can include (i) simple differences of phasing (e.g. 'the same' F0 movement can be aligned earlier or later, as in Northern German vs. Southern German, section 2.3.7.), or (ii) differences of basic duration (e.g. the F0 movement can have longer or shorter duration relative to 'the same' syllable structure: in Dutch (section 2.3.2) and Greek (section 2.1) the beginning of a rising prenuclear accent is aligned in the same way, but the end of the rise is later in Greek than in Dutch), or (iii) differences in the way alignment is affected by syllable structure (e.g. in Chinese the end of a rise is always aligned with the end of the syllable (sec. 2.3.4), whereas in Spanish the alignment of the end of the rise depends on whether the syllable is closed (Prieto & Torreira 2004), and in Japanese the consistent alignment of F0 peaks with accented syllables also entails adjustments in segment duration (sec. 2.3.5).

In the remainder of the paper I take these empirical findings as established, and consider their significance for our understanding of timing and temporal coordination of gestures in speech production.

3. Interpreting segmental anchoring

As we have seen, ALM interpreted their original finding of segmental anchoring in terms of an autosegmental analysis in which a pitch movement is the phonetic manifestation of a sequence of tones aligned with the segmental string in well-defined ways. This interpretation is consistent with certain aspects of the phenomenon, notably the fact that the alignment of one tone can be at least partially independent of the other; the relative independence of the two targets makes problems for the model proposed in Xu's work (section 2.3.4), in which whole F0 movements are aligned with whole syllables. However, we have also seen that there are reasons not to base our understanding of segmental anchoring too firmly on the autosegmental interpretation. In particular, we need to take account of the fact that the details of segmental anchoring must be expressed quantitatively, not symbolically, and that there is at least some evidence for recognizing the unity – at some level of description – of pitch movements like those under consideration here. I think we can make most sense of everything we know about segmental anchoring if we treat pitch movements as 'gestures', in the sense in which the term is used in models of speech production, such as Articulatory Phonology (e.g. Browman & Goldstein 1986, and much work since then). In this section I consider some of the implications of the gestural perspective.

3.1. Pitch movements as gestures

One immediate advantage of treating pitch movements as gestures is that doing so gives us a basis for taking account of the fine differences of phonetic detail attested in the work summarized in sections 2.2 and 2.3. Here a useful analogy can be drawn to voice onset time (VOT), which, like F0/segment alignment, is phonetically a matter of coordination between laryngeal and supralaryngeal gestures. Several features of VOT are relevant to our theoretical understanding of alignment. First, we know that languages can differ in the number of categories they distinguish by VOT: many languages have a single contrast between earlier and later VOT (usually described as voiced / voiceless), but some have no contrast (e.g. many Australian languages), and others have a three-way contrast of early, mid, and late VOT (usually described as voiced / voiceless unaspirated / voiceless aspirated). In the same way, some languages clearly have contrasts of alignment, while others seem not to. Second, we also know that the phonetic detail of any given language is poorly predicted by the phonological structure: given two languages with an early-late VOT contrast, the same VOT value may manifest the 'early' category in one language and the 'late' category in the other. Even in languages with no VOT contrasts we can still make language-specific phonetic generalizations about VOT. Thus it should come as no surprise that, for example, prenuclear H peaks can be aligned differently in Greek and in Dutch. Finally, we know that there do not appear to be favored phonetic 'slots' for VOT (say, +15 ms for voiceless unaspirated and +60 ms for voiceless aspirated); rather, the very careful cross-linguistic study of VOT by Cho & Ladefoged (1999) makes clear that average VOT for a given category in a given language can take on any of a continuum of values. We do not yet have comparable data for F0/segment alignment, but the data we do have shows no evidence that there are favored patterns. All of these considerations point to the conclusion that the fine phonetic detail of segmental anchoring is not a matter of secondary association after all, but of quantitative language-specific phonetic detail in the realization of phonological categories.

Given the findings of segmental anchoring, however, there would appear to be an obvious disadvantage of treating pitch movements as gestures, namely that it would appear to make it necessary to treat pitch movements as phonological units, as proposed by Xu; it would seem difficult to express the well-documented independence of the beginning and the end of pitch movements. However, here again an analogy may be enlightening: the 'unity' of pitch movements can be compared to the unity of a diphthong. The phonology of diphthongs (like other complex segments such as affricates and prenasalized stops) is a perennial conundrum, but at some phonetic level we are dealing with a specified movement from one set of formant values to another. These sets of values can be seen, at some level of phonetic specification, as targets, and the movement from one target to another in any specific context will be affected by speech rate, prosodic structure, surrounding consonants, and so on. The details are beyond the scope of this short paper, but two things are clear: first, the unitary nature of the articulatory gestures involved in a diphthong does not make the starting and ending targets irrelevant to the diphthong's phonological characterization; and second, there is no reason to assume that the time-course of the diphthong gesture is somehow universally locked to some other specific articulatory event or events – in fact, some recent work (notably Geumann & Hiller 1996; Scobbie Turk & Hewlett 1999) suggests that the temporal details of the formant movements in a diphthong can have language-specific phonological significance.

If we draw an analogy between pitch movements and diphthong gestures, we can see two useful things more clearly. First, we can see that there is no contradiction between treating the pitch movement as a phonetic gesture and describing it phonologically in terms of its endpoints (e.g. L+H). Second, we can acknowledge the phonetic unity of pitch movements without also assuming (as Xu and others have tended to do) that pitch movements are coordinated with their associated syllables in a deterministic way. In his work, Xu has started with the indisputable fact that segmental syllables and pitch movements are COORDINATED IN TIME, and has elaborated it into a model in which in many cases the beginnings or ends of pitch movements and syllables are expected to be SIMULTANEOUS. This has required him, among other things, to gloss over fine differences of alignment detail like those between Greek and Dutch prenuclear accents. For example, Xu & Wang (2001) emphasize that "tones are tightly aligned with host syllables despite the variations" (328) and appear to suggest that variations from simultaneity are all to be explained in biomechanical terms. Yet the 'variations' are highly systematic and language-specific, and the analogy suggested above between F0/segment alignment and VOT shows that there is no reason to expect only a single pattern of coordination between pitch movements and syllables. I believe we must find a way of accommodating these variations in our theoretical understanding.

3.2. The domain of pitch gestures

In the previous section, I suggested analogies to segmental phenomena – VOT and diphthongs – that may make it easier to think about the implications of treating segmental anchoring from a gestural perspective. In the end, though, clear analogies will only get us so far. We need further input from two areas of research before we can make real progress in understanding the phenomenon as a whole.

First, we need a better general understanding of gestural coordination in speech production, because segmental anchoring seems to imply lookahead. In a simple model of pitch movements like the one used in 'recipes' for synthetic speech by 't Hart and his colleagues in the 1970s and 1980s ('t Hart & Collier 1975; 't Hart, Cohen & Collier 1990), the production mechanism only needs a starting point and a specification of the pitch excursion and the slope (e.g. 'starting now, raise F0 by 5 semitones at a rate of 40 semitones per second'). By the same token, even in the rather more complex model of gestural coordination proposed in 'Articulatory Phonology' (AP), gestural duration is specified largely by factors intrinsic to individual gestures (such as 'stiffness') and to factors that affect all gestures equally (such as speaking rate). A more complicated model will be required to deal with segmental anchoring, because in order to know how fast to raise the pitch, the model needs to anticipate how long it will take to get to the specified finishing point. In terms comparable to those used by 't Hart *et al.*, the specification of the pitch excursion would be something like raise F0 by 5 semitones, starting now and finishing at the onset of the next vowel. In AP terms, we might model such effects by fine-tuning the stiffness of the pitch-changing gestures – but there is as yet no obvious mechanism by which the stiffness of one gesture could be affected by the duration of some other.

The second area where we need further research – and possibly better theoretical understanding – involves the domains to which pitch features apply in the phonology. In the strictly autosegmental interpretation of segmental anchoring, each tone can be idealized as a single timeless event associated with a single point in structure; 'domains' are irrelevant. But as soon as we acknowledge the existence of effects of time pressure and the fact that pitch gestures have duration that is coordinated with the duration of subparts of the segmental string, then we need to consider the domains within which time pressure applies. Are we concerned only with syllables, as maintained by Xu? Or are units like word and foot relevant as well? There is evidence to suggest that we will have to invoke domains of several sizes in order to explain a variety of alignment effects. I briefly outline two groups of preliminary findings here that illustrate this point.

3.2.1. Word-domain and sentence-domain effects: In our study of English and Scottish alignment (Ladd *et al.*, work in progress), we accidentally discovered a difference of alignment between sentences containing a prenuclear and a nuclear accent and sentences containing only a nuclear accent. In a first set of recordings of speakers of Southern British English and Scottish English, we placed the test words in nuclear-accent position in short sentences like *Their sister's name is Nell* or *They sacrificed the lamb*. Though the sentences were short, all contained one or occasionally two prenuclear accents (viz. the accents on *sister* and *sacrificed* in the examples just given.) For various reasons which are not relevant here, we later recorded a similar set of materials in which the test words were in nuclear position in even shorter sentences that were not expected to have any prenuclear accent at all, such as *He called Nell* or *We ate lamb*. To our surprise, we found a small but consistent difference in the alignment of the nuclear accent peaks in the two conditions: in the one-accent condition, the peak was aligned slightly later (on average 34 ms after vowel onset) than in the two-accent condition (on average 24 ms after vowel onset). (These means are based on all text and speaker conditions, but the difference was consistently present within conditions as well).

As just noted, we recorded the second set of materials for reasons unrelated to alignment, and we had no cause to think that we would find any difference of alignment at all. If anything, one might have expected that the alignment would be later in the *two*-accent condition, because of 'time pressure' from the prenuclear accent. We should still consider the possibility that it was a fluke, due to the fact that the two sets of recordings used two different groups of speakers. However, preliminary data from materials recorded by Caterina Petrone and myself suggest that an effect similar to that found in English is present in Italian as well; a fuller experiment is in preparation. If the effect is indeed genuine, it appears that the mere length of the sentence is enough to cause the alignment to drift: the longer the sentence, the earlier the peak of the final accent.

The idea that the effect just sketched might be due to the place of the accent in the sentence as a whole is strengthened by a finding to be reported by Ishihara (2006): the alignment of the F0 peak of the lexical accent in Japanese is progressively slightly earlier, relative to the accented syllable, the later the accented syllable is located in the word. When the lexical accent is on the first syllable in a CVCVCVCV word, the peak is aligned with the BEGINNING OF THE VOWEL in the following syllable; when the accent is on the second syllable, the peak is aligned EARLY IN THE ONSET CONSONANT of the following syllable; when the accent is on the third syllable, the peak is aligned DURING THE ACCENTED VOWEL ITSELF. Again, the details of the alignment seem to depend not only the structure of the accented syllable, but its place in a larger prosodic domain.

The effects reported in the section are reminiscent of the effects on segmental duration reported by Lehiste (1972) and subsequently investigated by others, including Turk & Shattuck-Hufnagel (2000). Broadly speaking, the longer a word, the shorter the segments it contains. The mechanisms underlying this finding are obscure, but the same mechanisms might be involved in the effects of word length on peak alignment as well. This is a potentially fruitful area for future research.

3.2.2. Difference between monosyllabic and disyllabic nuclear-accented words: The other effect that emerges from the study of English nuclear accents is that accent peaks are aligned earlier when the accented syllable is sentence-final and later when it is sentence-penultimate. This is based on two specific comparisons, both showing the same effect. In the first set of recordings (the 'two-accent' sentences), we had both monosyllabic and disyllabic test words (e.g. *Their sister's name is Nell* and *The economy's based on mining*); in the second set of recordings (the 'one-accent' sentences), the test words were all monosyllabic but in some cases they were followed by an unaccented (post-nuclear) word like *now* or *there* (e.g. *Phone Mel* and *Phone Mel now*). In both cases the peak is aligned later in the disyllabic/penultimate cases (on average 38 ms after vowel onset) than in the monosyllabic/final cases (on average 21 ms after vowel onset). Yet the stressed vowel with which the accent peak co-occurs is LONGER in the sentence-final case (on average 153 ms.) and SHORTER when it is sentence-penultimate (on average 111 ms). That is, relative to the stressed vowel, the alignment is very much later in the disyllabic/penultimate cases than in the monosyllabic/final cases.

If we assume that the stressed syllable is the domain of the accent, we have no way at all to explain why the alignment gets later as the stressed vowel gets shorter. However, if we assume that both the vowel and the peak alignment are independently responding to the gestural coordination within the larger unit of the disyllabic word or foot, then the results make perfect sense: in the disyllabic/penultimate case, the vowel is shortened to make room for the extra segmental material in the foot, but at the same time the peak drifts later because the extra segmental material reduces the time pressure from the upcoming end of the sentence. The consequence is that the peak alignment, relative to the stressed vowel, looks rather different in the two cases.

4. Conclusion

All the empirical evidence discussed in this paper points to the conclusion that there is a genuine phenomenon of 'segmental anchoring', whereby the duration of pitch movements in speech is finely adjusted to the duration of the accompanying segmental material. At the same time, the full range of empirical findings we have considered make the autosegmental explanation for segmental anchoring originally proposed by Arvaniti *et al.* (1998) less attractive than it seemed at first. In one way or another, it appears that we must acknowledge that pitch movements are 'gestures' in the sense of Articulatory Phonology, and seek to explain segmental anchoring in terms of a more general theory of gestural coordination. However, as the preliminary evidence discussed in Section 3 shows, it is not straightforward to incorporate F0/segment coordination into Articulatory Phonology. A considerable range of empirical and theoretical modeling work remains to be done.

Acknowledgements

Sections 1 and 2 of this paper appeared in a somewhat different form in a festschrift for Sieb Nooteboom (Ladd 2004), while a much abridged version of the paper as a whole was presented at the 2nd PaPI conference in Barcelona in June 2005. I am grateful to my many collaborators and students for their contributions to the work summarized in Section 2. I also thank Astrid Schepman and Rebekah Stackhouse for their contributions to Section 3.2, and Marianne Pouplier, Doris Mücke, and Pilar Prieto for discussion. The data reported in Section 3.2 were collected as part of a project supported by a Small Grant from the British Academy.

Address of Author:

D. Robert Ladd, Edinburgh University bob@ling.ed.ac.uk

Bibliographical References

ARVANITI Amalia & Gina GARDING. Forthcoming. Dialectal variation in the rising accents of American English. To appear in COLE Jennifer & José I. HUALDE (eds.). *Papers in Laboratory Phonology IX.* Berlin: Mouton de Gruyter.

ARVANITI Amalia & Robert D. LADD 1995. Tonal alignment and the representation of accentual targets. In *Proceedings of the 13th International Congress of Phonetic Sciences.* Stockholm, vol. 4. 220-223.

ARVANITI Amalia, D. Robert LADD & Ineke MENNEN 1998. Stability of Tonal Alignment: the case of Greek prenuclear accents. *Journal of Phonetics* 26. 3-25.

ATTERER Michaela & D. Robert LADD 2004. On the phonetics and phonology of "segmental anchoring" of F0: evidence from German. *Journal of Phonetics* 32. 177-197.

BROWMAN Catherine & Louis GOLDSTEIN 1986. Towards an articulatory phonology. *Phonology Yearbook* 3. 219-252.

CASPERS Johanneke & Vincent 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 & Peter LADEFOGED 1999. Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics* 27. 207-229.

DILLEY Laura, D. Robert LADD & Astrid SCHEPMAN 2005. Alignment of L and H in bitonal pitch accents: testing two hypotheses. *Journal of Phonetics* 33. 115-119

FACE Timothy L. 2002. *Intonational marking of contrastive focus in Madrid Spanish.* Munich: Lincom Europa.

FUJISAKI Hiroya 1983. Dynamic characteristics of voice fundamental frequency in speech and singing. In MACNEILAGE Peter F. (ed.) 1983. *The Production of Speech*. New York: Springer-Verlag. 39-55.

GEUMANN Anja & Marcus HILLER 1996. Diphthong dynamics in Swabian. *Journal of the Acoustical Society of America* 100. 2687.

GRICE Martine, D. Robert LADD & Amalia ARVANITI 2000. On the place of "phrase accents" in intonational phonology. *Phonology* 17. 143-186.

GUSSENHOVEN Carlos 2000. The boundary tones are coming: on the nonperipheral realization of boundary tones. In BROE Michael & Janet PIERREHUMBERT (eds.) 2000. *Papers in Laboratory Phonology V*. Cambridge: Cambridge University Press. 132-151.

HAYES Bruce 1989. Compensatory Lengthening in Moraic Phonology. *Linguistic Inquiry* 20. 253-306.

ISHIHARA Takeshi 2003. A phonological effect on tonal alignment in Tokyo Japanese. *Proceedings of 15th ICPhS*. Barcelona 1. 615-618.

ISHIHARA Takeshi 2006. *Tonal alignment in Tokyo Japanese.* University of Edinburgh. PhD Dissertation.

LADD D. Robert 1996. *Intonational Phonology*. Cambridge: Cambridge University Press.

LADD D. Robert 2004. Segmental anchoring of pitch movements: autosegmental phonology or speech production? In QUENÉ Hugo & Vincent VAN HEUVEN (eds.) 2004. *On Speech and Language: Studies for Sieb G. Nooteboom.* Utrecht: Netherlands Graduate School of Linguistics (LOT). 123-131.

LADD D. Robert, Dan FAULKNER, Hanneke FAULKNER & Astrid SCHEPMAN 1999. Constant "segmental anchoring" of F0 movements under changes in speech rate. *Journal of the Acoustical Society of America* 106. 1543-1554.

LADD D. Robert, Ineke MENNEN & Astrid SCHEPMAN 2000. Phonological conditioning of peak alignment of rising pitch accents in Dutch. *Journal of the Acoustical Society of America* 107. 2685-2696.

LADD D. Robert & Astrid SCHEPMAN 2003. "Sagging transitions" between high pitch accents in English: experimental evidence. *Journal of Phonetics* 31. 81-112.

LADD D. Robert, Astrid SCHEPMAN, Laurence WHITE, Louise May QUARMBY & Rebekah STACKHOUSE (work in progress). On alignment of pitch accents in Scottish and Southern British English.

LEHISTE Ilse 1972. The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America* 51. 2018-2024.

LICKLEY Robin, Astrid SCHEPMAN & D. Robert LADD 2005. Alignment of "phrase accent" lows in Dutch falling-rising questions: Theoretical and methodological implications. *Language and Speech* 48. 157-183.

NOOTEBOOM Sieb G. 1972. *Production and Perception of Vowel Duration: A Study of Durational Properties of Vowels in Dutch.* University of Utrecht. PhD Dissertation.

PERKELL Joseph & Melanie MATTHIES 1992. Temporal measures of anticipatory labial coarticulation for the vowel /u/: Within- and cross-subject variability. *Journal of the Acoustical Society of America* 91. 2911-2925.

PIERREHUMBERT Janet 1980. *The Phonology and Phonetics of English Intonation*. PhD Dissertation. MIT.

PIERREHUMBERT Janet & Mary E. BECKMAN 1988. *Japanese Tone Structure.* Cambridge MA: MIT Press.

PRIETO Pilar, Mariapaola D'IMPERIO & Barbara GILI FIVELA 2005. Pitch accent alignment in Romance: primary and secondary associations with metrical structure. *Language and Speech* 48. 359-396.

PRIETO Pilar & Francisco TORREIRA 2004. Anchoring effects in Spanish. Paper presented at European Science Foundation International Conference on Tone and Intonation in Europe (TIE), Santorini, Greece.

PRIETO Pilar, Jan VAN SANTEN & Julia HIRSCHBERG 1995. Tonal alignment patterns in Spanish. *Journal of Phonetics* 23. 429-451.

SCHEPMAN Astrid, Robin LICKLEY & D. Robert LADD 2006. Effects of vowel length and "right context" on the alignment of Dutch nuclear accents. *Journal of Phonetics* 34. 1-28.

SCOBBIE James M., Alice E. TURK & Nigel HEWLETT 1999. Morphemes, phonetics, and lexical items: The case of the Scottish Vowel Length Rule. *Proceedings of the 14th International Congress of Phonetic Sciences*, San Francisco, vol. 2. 1617-1620.

SILVERMAN Kim & Janet PIERREHUMBERT 1990. The timing of prenuclear high accents in English. In KINGSTONE John & Mary BECKMAN (eds.) 1990. *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. Cambridge: Cambridge University Press. 71-106.

TAYLOR Paul A. 2000. Analysis and synthesis of intonation using the Tilt model. *Journal of the Acoustical Society of America* 107. 1697-1714.

'T HART Johan & René COLLIER 1975. Integrating different levels of intonation analysis. *Journal of Phonetics* 3. 235-255.

'T HART Johan, René COLLIER & Antonie COHEN 1990. *A perceptual study of intonation: An experimental-phonetic approach*. Cambridge: Cambridge University Press.

TURK Alice & Stefanie SHATTUCK-HUFNAGEL 2000. Word-boundary-related duration patterns in English. *Journal of Phonetics* 28. 397-440.

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

XU Yi 1999. Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics* 27. 55-105.

XU Yi & Emily Q. WANG 2001. Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication* 33. 319-337.

ZSIGA Elizabeth 1997. Features, gestures, and Igbo vowel assimilation: An approach to the phonology/phonetics mapping. *Language* 73. 227-274.