SEGMENTAL ANCHORING OF TONES AS A WORD-BOUNDARY CORRELATE IN ENGLISH

D.R. Ladd and Astrid Schepman
University of Edinburgh

ABSTRACT

The hypothesis that the alignment of F0 minima forms a word boundary correlate in certain phonological contexts was tested experimentally. The effect of boundary location on the alignment of F0 minima between two F0 peaks was examined in word pairs which had either an ambiguous consonant (*Jay Neeson* vs *Jane Eason*) or an ambiguous syllable (*Al Maloney* vs *Alma Lonie*). It was found that there was a strong effect of boundary location when consonants were ambiguous, with early boundaries leading to early alignment. There was no effect of boundary location when syllables were ambiguous. This finding reinforces earlier findings that the beginnings of F0 rises are consistently aligned relative to the accented syllable. Furthermore, this lawful behaviour of interpeak valleys poses challenges for existing inventories of tonal targets, which propose that interpeak valleys merely reflect sagging transitions.

1. INTRODUCTION

A number of recent studies have found that the local F0 minima and maxima related to "pitch accents" or "prominence-lending pitch movements" are quite precisely aligned with the segmental string. Studies of several European languages [1,4,11] have observed that the beginning of the F0 rise for a rising pitch accent occurs at the beginning of the accented syllable. For example, in a hypothetical sequence [taˈmasa] pronounced with a rising-falling citation-form pitch contour, the beginning of the rise (i.e. a local F0 minimum) would be aligned with the beginning of the [m] segment of the stressed syllable.

The three studies just cited find this invariant alignment only for the "beginning of the rise", and report considerable variability in the alignment of the F0 peak at the "end of the rise". More recently, however, Arvaniti, Ladd and Mennen [2] have shown for Greek that, if certain variables are duly controlled, the location of the end of the rise is actually as consistent as the beginning of the rise. They suggest that this is in keeping with autosegmental theories of intonation [6,10] in which F0 contours are thought of as strings of tones: in their view, "tonal targets" (phonetic targets corresponding to the tones) are "segmentally anchored" (consistently aligned with landmarks in the segmental string). This line of investigation has been pursued in subsequent studies of rises in English [6] and Dutch [7], with similar results (i.e. segmental anchoring of both the beginning and the end of the rise). In work in progress we are extending this general idea to falling accents as well.

Regardless of how far the notion of tonal target can profitably be pursued, though, the fact that rising accentual movements seem to begin consistently at the beginning of an accented syllable suggests that the location of the F0 minimum might be a correlate of the word boundary distinction in English pairs such *as grade A / gray da*y or *seem able / see Mabel*. Specifically, the F0 minimum should be aligned BEFORE the ambiguous consonant in *gray day* or *see Mabel* but AFTER the consonant in *grade A* or *seem able*. In fact, pairs such as these were first investigated instrumentally by Lehiste [8], who demonstrated that listeners could reliably distinguish them, and identified a variety of acoustic correlates of the distinction. In many cases she found that the distinction was cued by specific allophonic differences affecting certain segments (e.g. aspiration of /k/ distinguishing *ice cream* from *I scream*). More generally, she found segment duration to be a robust cue: "Initial (post-junctural) allophones of almost all phonemes are considerably longer than either medial or pre-junctural allophones" [8: p. 42). However, she reports no data on F0, though she says that narrow-band spectrograms were made of some of the utterances, and a reasonable inference is thus that F0 plays no role in distinguishing pairs like *grade A / gray day*. This inference is what is called into question in the present study.

Note that we do not predict that the alignment of the F0 minimum will have the general function of marking word boundaries. Since the location of F0 minima appears to be a function of how pitch accents are aligned with accented syllables, it should not serve as a general word boundary cue, but primarily as a syllable boundary cue. That is, alignment of the F0 minimum could be expected to distinguish *grade A* from *gray day*, because in that case the word boundary coincides with the boundary of an accented syllable, but in pairs like *get a board* / *get aboard* (another of Lehiste's items), the location of the word boundary should not affect the alignment of the F0 minimum. What we are looking for, therefore, are phenomena analogous to Lehiste's findings about the aspiration or non-aspiration of /k/ in *ice cream* / *I scream*: we assume that F0 alignment is conditioned by factors other than word boundary location, but in such a way that in certain cases it can serve, more or less adventitiously, as a word boundary correlate. The goal of this study is to provide evidence for this hypothesis.

2. METHOD

2.1. General Considerations

At the time of Lehiste's study instrumental data for F0 were limited to narrow-band spectrograms. It would have been extremely difficult for her to look for detailed alignment data of the sort under consideration here, even if she had had the theoretical incentive to do so. Our study depends crucially on the capabilities of modern digital F0 extraction and display packages, which make it possible to localise F0 minima and maxima with some confidence. Even with modern techniques, though, there are many obstacles to testing our hypothesis, which we discuss for the benefit of future researchers.

First, the observations we want to make require us to study only the smooth F0 contours characteristic of all-sonorant utterances. As is well known [12, 9: p.71-74], many consonants create local perturbations in F0, which would make it impossible to locate a "true" F0 minimum. This means we are restricted to test phrases that do not have obstruents in the vicinity of the word boundary. This restriction is somewhat unnatural, but is widely accepted in instrumental studies of F0 [e.g. 3].

Second and more important, we need utterances with an appropriate intonation pattern. What is needed is a contour with a clear valley or local minimum between two clear accent peaks. So-called "hat patterns" or realisations in which the first word is unaccented do not provide us with an unambiguous local F0 minimum whose alignment can be compared across conditions. This means we need to work with phrases in which it is plausible for both words to bear prominent pitch accents.

Third, we must block the tendency of most English speakers to pronounce vowel-initial words with a glottal stop (or at least some degree of glottal constriction; [cf. 8: 44f]) preceding the vowel, because this causes very prominent local lowering of F0. Like obstruent perturbations, these local dips would tend to conceal any consistent differences in the location of the F0 minimum. Unfortunately, pilot work showed that an emphatic style of speech that favours having accents on both words (which we wanted) also favours putting a glottal stop at the beginning of vowel-initial surnames (which we didn't want).

After some trial and error, we found that we could often get accents on both words, without glottal stops on the vowel-initial words, if we asked speakers to produce the test sentences quickly. Even under these conditions, however, many speakers still produce many utterances that are unusable for testing our hypothesis; either they contain glottal stops, or they fail to accent the first word, or their accents are not prominent enough to produce a clearly identifiable F0 minimum in between. As a result, our report here is based on only two speakers' productions.

2.2. Method: Details

2.2.1. Materials and Design. The test phrases were names of people, like *Joe Neeson* and *Helena Lonsdale*. The two parts of each name will be referred to as FN (first name or given name, e.g. *Joe*) and LN (last name or surname, e.g. *Neeson*). Each test name was paired with another, from which it differed in the location of the word boundary relative to the segmental string. Wherever possible, perfect minimal pairs were used, such *as Joe Neeson / Joan Eason* or *Norma Nelson / Norman Elson*. In some cases it was not possible to find perfect minimal pairs. Examples of non-minimal test pairs are *Helena Lonsdale / Helen Alonso* and *Malkah Mandelson / Malcolm Anderson*. We looked for names with which our speakers might be expected to be familiar, but some of the names are identifiably foreign (e.g. FN *Julio*, *Malkah*; LN *Iwaki*, *Miranda*) or unusual (e.g. FN *Shea*, *Kyle* LN *Neely*, *Malaney*).

There were 104 test phrases, each embedded in a carrier sentence of the form *It was [test name], not [foil name]*. Every test name had its own foil name, and we chose foil names that might plausibly be confused with the test names (e.g. *Eleanor Mullins / Annabelle Myers*). There were also 10 practice items and 64 fillers of the same form. The list was pseudo-randomly ordered, with two halves of the list separated by a 15-filler buffer and members of a pair appearing in different halves, allocated using a Latin Square. The list started with 5 fillers.

The word boundary ambiguity involved either a single consonant (*Joe Neeson / Joan Eason*) or a whole syllable (*Al Maloney / Alma Lonie*). We predicted that the alignment of the F0 minimum would be different in the former case, but not in the latter. Specifically, we predicted that the F0 minimum would be aligned later in names like *Joan Eason* than in those like *Joe Neeson*. In addition, we tested the possibility that the results would be influenced by whether the FN is potentially monosyllabic, since we expected that segment durations would be substantially different in accented monosyllabic words (e.g. we expected that the syllable *Al* in *Al Maloney* would be longer than in *Alma Lonie*, whereas we expected little effect on segment durations in pairs like *Norman Elson / Norma Nelson*) [13]. We had no basis for predicting whether this factor would affect the alignment of the F0 minimum or not.

There were thus three factors in the experimental design, which we refer to as Boundary (Early - Late), Ambiguity (Consonant - Syllable) and FNlength (Short - Long). Boundary was a within-items factor; the other two were between-items. Data were analysed using 2 x 2 x 2 ANOVAs, with items as the random factor. Comparisons were planned of the effect of boundary in the ambiguous consonant condition and ambiguous syllable condition separately, averaging over the two FNlength conditions (simple effect of Boundary). Although Boundary was a within-items factor in the design, it was a between-items factor in the analyses to avoid the pairwise loss of data if one member of the pair was missing. Note that this analysis is less sensitive and therefore does not bias the results in favour of supporting the hypothesis.

2.2.2. Speakers. Results are reported for two speakers, both female undergraduates at Edinburgh University, one with an English accent (EF) and one with a Scottish accent (SF). A number of other speakers were recorded but their recordings were not analysed for the various reasons discussed in 2.1.

2.2.3. Procedure. Speakers were recorded on professional equipment in a sound-attenuated booth in the Phonetics Laboratory of the Edinburgh University Linguistics Department. The recordings were made direct to disk at a sampling rate of 16 kHz. The test sentences were displayed one at a time on a computer screen. Speakers pressed the space bar on the keyboard to call up the next sentence. Speakers were instructed to read the sentences fluently and to repeat any sentence if they stumbled. As discussed above, they were asked to read the sentences at a fairly fast rate. Subjects were given feedback about reading speed or technical problems after the 10 practice items. The session lasted about 20 minutes.

2.2.4. Measurements. Pitch was extracted using 10 ms windows on ESPS xwaves software and measurements were made on simultaneous displays of speech wave, wide-band spectrogram and F0 tracks. Three points were marked in each speech file:

- L: the location of the F0 minimum between the FN and LN of each test phrase;
- C1: the beginning of the "target consonant", i.e. consonant immediately preceding the stressed vowel of the LN. In the ambiguous consonant condition, the target consonant is the ambiguous consonant, and in the ambiguous syllable condition it is the onset of the accented syllable of the LN;
- V1: the end of the target consonant, i.e. the beginning of the accented vowel of the LN.

From these three points we calculated two dependent variables for L, namely its distance from both C1 and V1. The third dependent variable was the duration of the target consonant, V1 minus C1.

3. RESULTS AND DISCUSSION

Datapoints were discarded: 1) when speakers showed glottal stops or glottalisation at vowel-onset targets, judged from the spectrogram; 2) when speakers deaccented either the FN or the LN of the test name, or when they used a "flat hat" contour; 3) when speakers mispronounced a name in a way that affected the experimental comparisons (e.g. put stress on a syllable other than the one intended). For speakers EF and SF we discarded 15.5 and 18.4 % of utterances respectively. Figure 1 shows the means of the valid cases for each of the eight cells in the design for each speaker separately. Analyses for SF and EF will now be reported in turn. Note that unreported effects did not reach significance.

The ANOVA for the alignment of L relative to V1 showed that speaker SF had a main effect of Boundary, $F(1,83) = 6.18$, $p = .015$, and an interaction between Boundary and Ambiguity, $F(1,83) = 17.49$, $p < .0001$. The simple effects analyses showed that Boundary had a significant effect when the consonant was ambiguous, $F(1,39) = 26.651$, $p < .0001$, but not when the syllable was ambiguous, $F(1,41) = 1.48$, $p > .2$.

The ANOVA for the alignment of L relative to C1 for SF showed that the effect of Boundary approached significance, $F(1,83) = 3.14$, $p = .081$, as did the interaction between Boundary and Ambiguity, $F(1,83) = 3.37$, $p = .070$. The simple effect of Boundary was significant in the ambiguous consonant condition, $F(1,39) = 8.082$, $p = .007$, but not in the ambiguous syllable condition, $F(1,41) < 1$.

The ANOVA for the duration of the target consonant (C1 to V1) showed that SF had a main effect of Ambiguity, $F(1,83) = 5.49$, $p = .022$, and an interaction between Ambiguity and Boundary, $F(1,83) = 22.15$, $p < .0001$. The simple effect of Boundary was significant in the ambiguous consonant condition, $F(1,39) = 28.714$, $p < .0001$, and also in the ambiguous syllable condition, $F(1.41) = 4.952$, $p < .032$.

EF's ANOVA for the measure L to V1, revealed a main effect of Boundary, $F(1,86) = 20.42$, $p < .0001$, a main effect of Ambiguity $F(1,86) = 32$, $p < .0001$, and an interaction between Ambiguity and Boundary, $F(1,86) = 15.92$, $p < .0001$. The three-way interaction approached significance, $F(1,86) = 3.76$, $p < .056$. The simple effect of Boundary was significant in the ambiguous consonant condition, $F(1,43) = 34.380$, $p < .0001$, but not in the ambiguous syllable condition, $F(1,40) = 1.144$, $p > .7$.

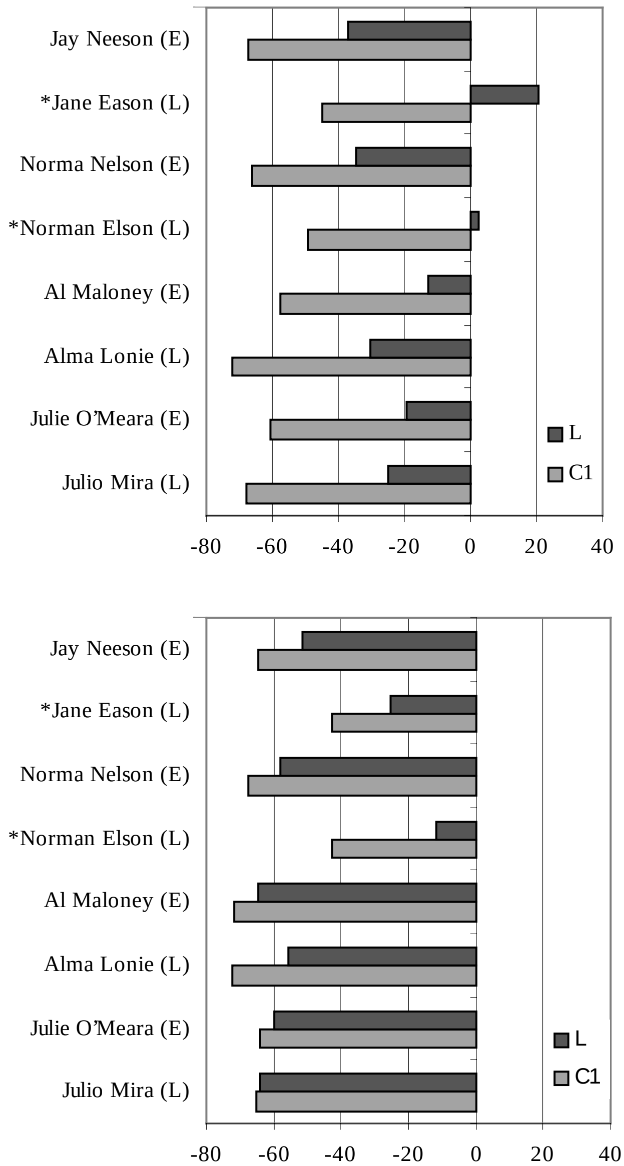

Figure 1. Cell means for the Scottish speaker (top panel) and English speaker (bottom panel). V1 is set at zero. Light bars show the duration of the test consonant, and dark bars show the location of L relative to V1, both in ms. Condition labels are indicated by representative test names, with boundary conditions, Early (E) and Late (L) marked. Conditions marked with an asterisk were predicted to have late alignment.

For the measure L to C1, EF's ANOVA showed a main effect of Ambiguity, $F(1,86) = 5.45$, $p = .022$, and a main effect of Boundary that approached significance, $F(1,86) = 3.17$, $p = .079$. The Ambiguity x Boundary interaction also approached significance, $F(1,86) = 2.66$, $p = .107$, as did the three-way interaction, $F(1,86) = 2.75$, $p = .101$. The simple effect of Boundary was significant in the ambiguous consonant condition, $F(1,43) = 4.192$, $p = .047$, but not in the ambiguous syllable condition, $F(1,40) = .197$, $p > .6$.

For the measure C1 to V1 (the duration of the target consonant), EF showed a main effect of Ambiguity, $F(1,86) = 25.39$, $p < .0001$, a main effect of Boundary, $F(1,86) = 16.80$, $p < .0001$, and an interaction between Boundary and Ambiguity, $F(1,86) = 19.61$, $p < .0001$. The simple effect of Boundary was significant in the ambiguous consonant condition, $F(1,43) = 62.618$, $p < .0001$, but not in the ambiguous syllable condition, $F(1,40) < 1$.

The above analyses clearly show a pattern in line with our predictions: Both speakers showed effects of Boundary on the alignment of the local F0 minimum relative to V1 and C. Early boundaries (e.g. *Jay Neeson*) led to earlier alignment of the local pitch minimum than late boundaries (e.g. *Jane Eason*), but only when the consonant was ambiguous, not when the syllable was ambiguous (e.g. *Al Maloney* vs *Alma Lonie*). The factor FNlength did not have an effect and did not interact with other factors, showing that potential monosyllabicity did not have a significant impact on alignment or vowel length.

Note that the Scottish speaker showed later overall alignment than the English speaker. Although this may be due to individual speaker differences, impressionistic observation of both accents suggests that this is a general accent difference. However, no reliable comparative data currently exist. Crucially, though, the effects of our experimental factors were consistent across both speakers, despite the overall differences in alignment.

A further complication is that the durations of the target consonants show effects of boundary when consonants are ambiguous, and no effect (EF) or a smaller effect (SF) when syllables are ambiguous, following the same pattern as the alignment data. This is in keeping with Lehiste's original findings. Therefore, we need to consider the possibility that the effects on alignment are artefacts of the differences in consonant duration. However, this account is difficult to uphold. For instance, for SF, the duration of the target consonant in the early boundary condition is 67 ms when the boundary is early, and 47 ms when the boundary is late. Thus, the difference in consonant duration is 20 ms, but the alignment difference for these conditions is 40 ms, or twice the amount of the consonant effect. Analogous remarks apply to speaker EF.

4. SUMMARY AND PROSPECT

For the limited set of speakers and conditions studied in this paper, we have demonstrated that the precise location of the F0 minimum between two accented words may be a correlate of the location of the word boundary. As predicted, the effect is limited to those cases where the word-boundary ambiguity involves the syllable membership of a potential onset consonant in a word-initial stressed syllable. The location of the F0 minimum is thus essentially a correlate of syllable boundaries, not word boundaries.

The most obvious follow-up to the present study would be to determine whether the acoustic correlate can be used as a perceptual cue by listeners. That is, given a string such as [nɔːmənɛlsn], with the segmental durations adjusted so as to be ambiguous between the interpretations *Norman Elson* and *Norma Nelson*, can we induce one percept or the other by manipulating the alignment of the F0 minimum? We plan to do this experiment in the near future, as part of our broader project on the alignment of F0 targets in English and Dutch.

The more general issue raised by the experiment reported here involves the phonological status of the F0 minimum. In the orthodox Pierrehumbert analysis of the contours under discussion here, the F0 minimum represents merely the lowest point reached by a "sagging transition" between the two accent peaks, and does not correspond to any phonologically specified target (e.g. a L tone). However, the fact that this point is so precisely aligned requires us at the very least to entertain the hypothesis that there is such a tonal target. This, too, is a topic to be investigated further as part of our project.

ACKNOWLEDGEMENT

This work is part of the project "Alignment of Fundamental Frequency Targets in English and Dutch", supported by the UK Economic and Social Research Council. We gratefully acknowledge this support.

REFERENCES

[1] Arvaniti, A. and D.R. Ladd (1995). Tonal Alignment and the Representation of Accentual Targets. Proceedings of ICPhS 13 (Stockholm), vol 4: 220-223.

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

[3] Bruce, G. (1977). *Swedish word accents in sentence perspective.* Lund: Gleerup.

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

[5] Ladd, D. R. (1996). *Intonational Phonology*. Cambridge: CUP

[6] Ladd, D. R., D. S. Faulkner, H. Faulkner, and A. Schepman (submitted). Constant "segmental anchoring" of F0 movements under changes in speech rate. Submitted to *JASA*, November 1998; under revision.

[7] Ladd, D. R., I. Mennen and A. Schepman (in preparation). Effects of phonological vowel length on F0 alignment in Dutch.

[8] Lehiste, I. (1960). An Acoustic-Phonetics Study of Internal Open Juncture. Supplement to vol. 5 of *Phonetica*.

[9] Lehiste, I. (1970). *Suprasegmentals*. MIT Press, Cambridge.

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

[11] Prieto, P., J. Van Santen and J. Hirshberg (1995). Tonal alignment patterns in Spanish, *Journal of Phonetics*, 23: 429 - 451.

[12] Silverman K. (1986). F0 segmental cues depend on intonation - The case of the rise after voiced stops. *Phonetica*, 43: 76-91

[13] Turk, A. E. and J.R. Sawusch (1997). The domain of accentual lengthening in American English. *Journal of Phonetics,* 25: 25-41.