

DISCOURSE CONSTRAINTS ON PEAK TIMING IN ENGLISH: EXPERIMENTAL EVIDENCE

Anne Wichmann¹ and Jill House²

¹*Department of Cultural Studies, University of Central Lancashire, Preston, UK*

²*Department of Phonetics and Linguistics, University College London, UK*

ABSTRACT

The timing of F0 peaks in accented syllables is known to be subject to a number of segmental and prosodic constraints. The experiment reported here was carried out to investigate the additional effect of topic structure on F0 peak timing. Three specially constructed English texts were read aloud by nine subjects. In the texts, each of four accented target syllables occurred in different discourse positions: paragraph/topic initial, sentence initial and sentence final. The results show that the timing of F0 peaks is consistently and significantly determined by discourse position. In sentence-final position the F0 peak is always considerably earlier than in either of the initial conditions. Sentence-initial peaks occur late in the target syllable or beyond it altogether, while in sentences which are also topic or paragraph initial, the F0 peak occurs even later.

1. INTRODUCTION

A number of recent studies have investigated the alignment of fundamental frequency (F0) contours to the segmental string. (For a survey see [1] and [2].) Some of these are concerned with identifying phonologically distinct tonal categories, but others are concerned with within-category constraints on alignment. Such studies assume that it is possible to reliably identify instances of a particular specified F0 category such as a specified pitch accent, and then examine its realisation in context. Many have investigated the alignment of tonal peaks—H* in the autosegmental-metrical framework—which are assumed to be associated to some tone-bearing unit, typically identified as the accented syllable. We make the same assumptions in this study.

Although the results of other studies are not always easy to compare, a consistent finding seems to be that F0 peak timing is subject to both segmental and prosodic effects: the make-up of the segmental string, including intrinsic vowel duration, and onset and rhyme duration, can change the alignment of the peak in relation to the accented syllable with which it is associated. Prosodic effects include the proximity of the next accented syllable and the difference between nuclear and pre-nuclear position.

These studies are all concerned with the constraints of low-level structures (segment, word, foot, intonational phrase) on intonation. However, it is well-known that intonation is also affected in a number of ways by a higher level discourse structure. The information contained in a text is not simply expressed by a sequence of sentences, but by sentences grouped together around a topic or sub-topic to make up a meaningful unit, often referred to as a discourse unit. In written texts the boundaries of such topical units are often highlighted by typographic means such as paragraph divisions, headings and sub-headings. In speech, both the internal coherence of a discourse unit and the demarcation of its boundaries can be

indicated prosodically. The feature most commonly associated with the transition to a new topic is an extra high pitch reset on the first accented syllable. Observations of naturally-occurring data [3] have revealed that the influence of topic structure also extends to the timing of such accent-lending F0 peaks. It was found that there is a tendency for the alignment of accent-lending F0 peaks to be additionally influenced by the topic structure of a text: the peak associated with the first accented syllable (Intonational Onset or IO) in topic-initial position was found to occur later in the syllable, or even outside the syllable, than an IO in sentence-initial position which was not topic-initial. The present study was designed to test these observations under experimental conditions.

2. PROCEDURE

One of the difficulties of using natural data is that each IO falls on a different syllable/word, so that different segmental constraints operate in each case. An experimental text was therefore designed in which a defined set of syllables/words were placed in different positions in the text. In this way the segmental make-up of the syllables under scrutiny was kept constant, obviating the need to take segmental timing effects into account. Three similar texts were constructed, based loosely on an existing Open University Lecture (taken from the Spoken English Corpus [4]) on the subject of “The Enlightenment in France”. Each text was approximately 380 words in length, and contained 3 to 4 paragraphs. The texts were complex, of an academic nature, and required some preparation before performance. Readers were given time to familiarise themselves with the text.

There were 10 readers (5 male, 5 female), all speakers of Southern British English (near-RP), of whom 9 were used for subsequent analysis. In order to elicit a reading style as close to a professional performance as possible, readers were instructed to read as if they were experts in their field, recording an Open University lecture. If they stumbled during the reading, they were asked to repeat the paragraph in which the mistake occurred.

Each word occurred in the course of the three texts in three different positions. These positions were coded as +/- Sentence Initial (SI), and +/- Paragraph Initial (PI):

- Sentence Final: -SI, -PI
- Sentence Initial (but paragraph medial): +SI, -PI
- Paragraph Initial: +SI, +PI

“Sentence Initial” here implies the first accented syllable in the sentence (IO), rather than absolute initial position, and “Sentence Final” is the last (i.e. nuclear) accent in the sentence. We hypothesised that readers would interpret “Paragraph Initial” position as essentially topic initial, although we were aware that there

is no precise correlation between paragraphs, an essentially typographic convention, and topics.

The words which were used for the analysis are as follows (the relevant accented syllable is highlighted):

carTESian;

COMmon;

COMPENDium;

enLIGHTenment.

The four target words were incorporated into the text as the following example illustrates:

Paragraph Initial:

*The **Enlightenment** and its ideas are nowhere more evident than in the great Encyclopaedia published in 28 huge folio volumes between 1751 and 1772.*

Sentence Initial:

*The **Enlightenment** was now, although often thought of as a unified system of thought, in fact becoming increasingly diverse.*

Sentence Final:

*The movement they created has become famous as the **Enlightenment**.*

If a reader repeated a paragraph and unwittingly produced a target utterance twice, we included only the first version. Some target utterances had to be excluded from analysis, either because an F0 trace could not be extracted or because the reader failed to accent the appropriate syllable. A total of six data points were omitted in this way, leaving 102 for the final analysis.

The texts were recorded under anechoic conditions onto DAT tape using a Sony 1000 ES DAT recorder, and a B&K 2231 sound level meter fitted with a 4165 microphone cartridge.

F0 analysis was carried out using the LSI (Loughborough Sound Images) speech processing system. The signal was sampled at 10 kHz; the pitch analysis option ‘cepstrum’ was used with a frame advance of 5 ms. The timing of the F0 peak was calculated as a percentage of the total duration of the accented syllable. Syllables were measured respecting the principle of maximum onset. Some contours did not have an obvious peak but flattened out to a plateau. The F0 value at the beginning of the plateau was taken as the target value.

3. RESULTS

Subjects provided a good reading performance in the style we had hoped to elicit, although one male subject’s recording was excluded from further analysis because in a number of target sentence-final utterances he used an L* accent (a rising tone), categorically different from, and therefore not comparable to the final H*L (falling tone) used by the other speakers. Although reading speeds varied considerably from subject to subject, the performances were highly proficient, clearly reflecting the structure and meaning of the text, and in a varied, expressive style.

The positions of the F0 peaks were defined as their relative locations (expressed as a percentage) in or beyond the interval stretching from the start of the syllable until its end. An analysis of variance was carried out on these positions as a function of three

	Paragraph Initial	Sentence Initial	Sentence Final
F0 peak position (%)	115.7	105.2	62.2
F0 peak (Hz)	314	285	189
Syllable duration (ms)	206	207	217

Table 1. Mean values of F0 peak position (expressed as % of the syllable duration), F0-peak (Hz) and Syllable durations (ms) as a function of discourse position (pooled data).

fixed factors: speakers (9 levels), word (4 levels) and discourse position (3 levels); the factor ‘speaker’ was considered to be a fixed factor as the speakers had been selected by one of the authors on the basis of the criterion ‘good and experienced reader’. Three main effects turned out to be significant at the 1%-level: speaker ($F_{9,41} = 4.60$, $p = 0.001$), word ($F_{3,41} = 28.50$, $p = 0.001$) and discourse position ($F_{2,41} = 88.58$, $p = 0.001$); two two-way interactions were significant: word condition: ($F_{6,41} = 3.97$, $p = 0.003$) and speaker condition ($F_{14,41} = 2.55$, $p = 0.010$) The relative positions of the pitch peak as a function of discourse position are: 116%, 105% and 62%, with the highest value for the paragraph initial position and the lowest for sentence final. Post-hoc comparisons (Tukey’s HSD) showed significant differences ($p < 0.05$) between all discourse positions. As the interactions were ordinal, the order of the F0 peaks as a function of discourse position did not change at the different levels of the factors ‘word’ and ‘speaker’.

Discourse position also affected syllable duration (see Table 1). However, no difference in syllable duration was found between paragraph-initial and sentence-initial positions, while sentence-final position showed the expected lengthening effect. There was further effect of discourse position on the height of the pitch peak: $F_{2,41} = 133.51$, $p < 0.01$. Post-hoc comparisons (Tukey’s HSD) showed significant differences between all discourse positions at the 5%-level.

These results mean that the height and timing of the F0 peak were consistently and significantly determined by discourse position. Sentence-initial peaks occurred late in the target syllable or beyond it altogether, while paragraph-initial peaks were later still (see Figure 1). Paragraph-initial peaks were also predictably higher than sentence-initial peaks. In sentence-final position the F0 peak was always considerably earlier and lower than in either of the initial conditions.

4. CONCLUSION

The findings in the experiment described above confirm the tendencies previously observed in natural, unconstrained data, namely that in addition to any segmental or prosodic constraints there is also a discourse effect on the alignment of F0 peaks. The experimental data show first of all a very marked difference between the timing of sentence-final pitch accents and the timing of those in sentence-initial position. In addition we find that speakers distinguish between degrees of initiality: the F0 peak of an intonational onset at the beginning of a topic or paragraph occurs later (further to the right) in relation to the syllable as a whole than if the syllable is sentence initial but not topic or paragraph initial.

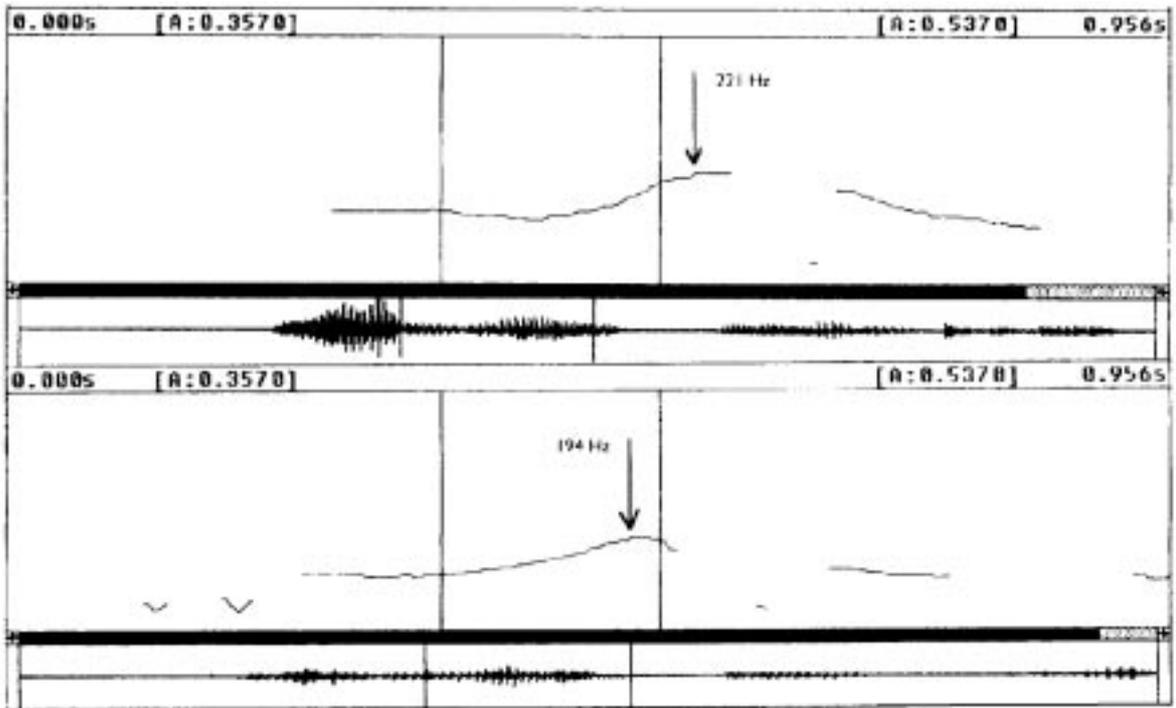


Figure 1: The F0 contour of the target word “enLIGHTenment” in (i) paragraph-initial position (upper panel) and (ii) sentence-initial position (lower panel). The cursors enclose the accented syllable.

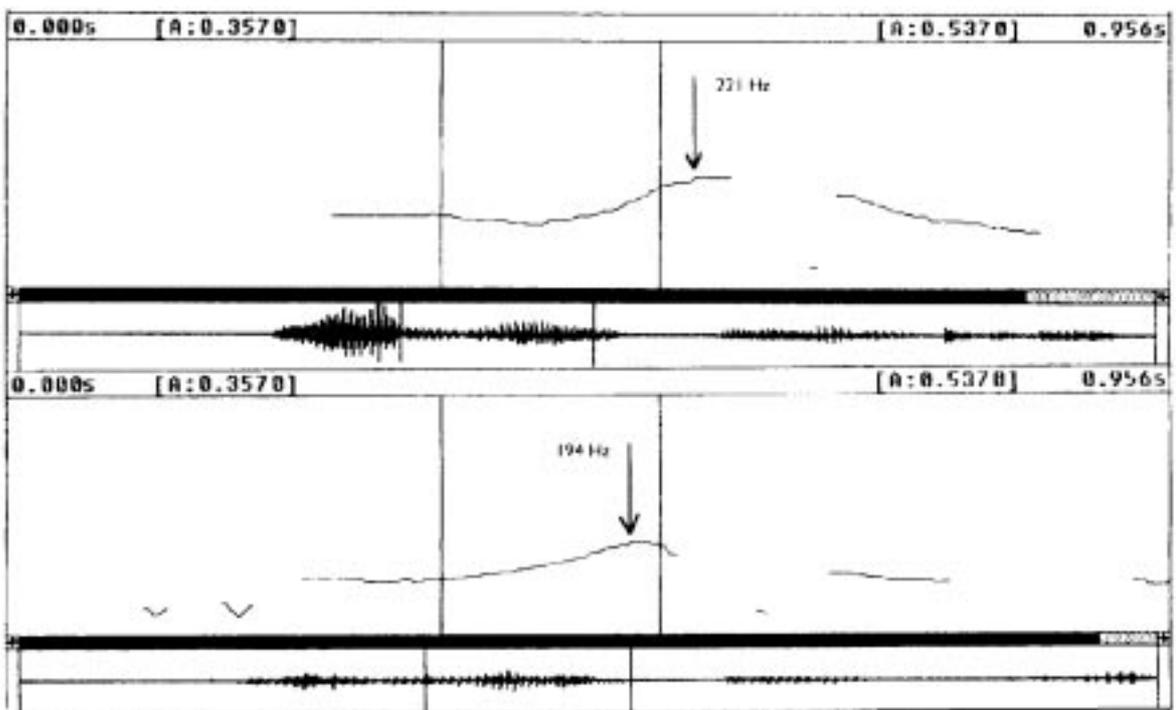


Figure 2: The F0 contour of the target word “COMPENDIUM” uttered in (i) paragraph-initial position (upper panel) and (ii) sentence-initial position (lower panel). The cursors enclose the accented syllable.

5. DISCUSSION

The results of this study suggest that peak timing in relation to an associated accented syllable is affected by discourse structure. Paragraph- or topic-initiality exerts a strong rightward push on the F0 peak of an intonational onset, even to the extent of causing it to occur beyond the accented syllable itself. However, in our experiment we found a number of cases where the readers did not show the clear effect we were looking for, but instead appeared to choose an alternative strategy, illustrated in Figure 2.

In Figure 2 we see that the F0 maximum is reached at a similar point in both the sentence-initial and paragraph-initial conditions, but that in the latter there is a longer plateau, which has the effect of delaying the point at which the contour begins to fall again. This may simply be an alternative way of achieving the effect of topic or paragraph initiality. An alternative interpretation might be that the underlying strategy is the same, namely to delay the falling contour, but that this is achieved in different ways: either the F0 peak is delayed, thus inherently delaying the starting point of the fall, or only the fall itself is delayed, resulting in a plateau between the F0 peak and the fall. Perception tests in which the timing of both peak and fall are manipulated independently might shed more light on this.

Our results reflect those reported earlier on the differences between nuclear and non-nuclear accent timing [5], and support the view that the timing of accents is context-dependent. This context-dependent variation is not limited, however, to a simple distinction between initial and final. It has already been observed that different timing of utterance-final falls can create different degrees of perceived finality. Our observations suggest that the greater perceived ‘finality’ of an early peak in a nucleus has its counterpart in the discursively greater ‘initiality’ of a late peak in an intonational onset.

ACKNOWLEDGEMENTS

We are grateful to Toni Rietveld for his valuable contribution, including the statistical analysis, to this paper.

REFERENCES

- [1] House, J. & A. Wichmann 1996 Investigating peak timing in naturally-occurring speech: from segmental constraints to discourse structure. In *Speech, Hearing and Language: work in progress*, Volume 9. University College London, 99–117.
- [2] Ladd, D.R. 1996 *Intonational Phonology* Cambridge: CUP
- [3] Wichmann, A., J. House & T. Rietveld 1997 Peak displacement and topic structure. In Botinis, Kouroupetroglou & Carayiannis (eds.) *Proceedings of the ESCA workshop on Intonation: Theory, Models and Applications*. Athens, 329–332.
- [4] Knowles, G., L. Taylor, B.J. Williams 1996. *Spoken English Corpus*. London: Longman.
- [5] Silverman, K. & J. Pierrehumbert 1990. The timing of prenuclear high accents in English. Kingston, J. & M. Beckman (eds.), *Papers in Laboratory Phonology I*. Cambridge: CUP, 72–106.