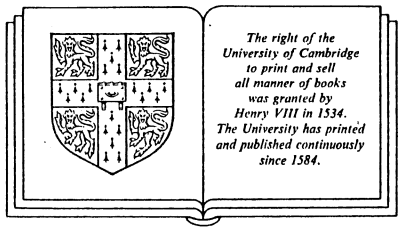
Papers in Laboratory Phonology I
Between the Grammar and Physics of Speech

EDITED BY JOHN KINGSTON

Department of Modern Languages and Linguistics, Cornell University

AND MARY E. BECKMAN

Department of Linguistics, Ohio State University



CAMBRIDGE UNIVERSITY PRESS
CAMBRIDGE
NEW YORK PORT CHESTER MELBOURNE SYDNEY

1990

5

The timing of prenuclear high accents in English

KIM E. A. SILVERMAN AND
JANET B. PIERREHUMBERT

5.1 Introduction

Speaking English means doing two things at once: saying the words, and saying the melody. The coordination between these two activities is far from arbitrary. Rather, it is determined by the stress pattern and phrasing of the utterance. Some features of the melody fall on certain stressed syllables, while others fall at the boundaries of prosodic phrases.

In this paper, we will be concerned with the timing of pitch accents, the melodic features which fall on certain stressed syllables. We shall examine the way in which prenuclear pitch accents are aligned with their associated syllables in a variety of prosodic environments, and our phonetic data will allow us to address two related issues. One of these is the degree of similarity between prenuclear and nuclear pitch accents – one of the longstanding points at which the American and British traditions of intonational analysis diverge. The second issue, which we believe supersedes the former, is the nature and status of the process by which underlying phonological forms are given their surface phonetic realization.

From a phonological point of view, the association of pitch accents with syllables can be described using autosegmental links. An example transcription is given in (1). The accents are transcribed using the system of Pierrehumbert (1980), and prosodic structure above the syllable level is omitted.[1]

(1)

[mama lem]	phoneme tier
σ σ σ	syllable level in the prosodic structure
H* H+L*	melody tier

Links between elements constrain them to overlap, as they are produced in time. For instance, the first accent (H*) and the first two phonemes ([m] and [a]) are phonologically specified to occur on the first syllable.

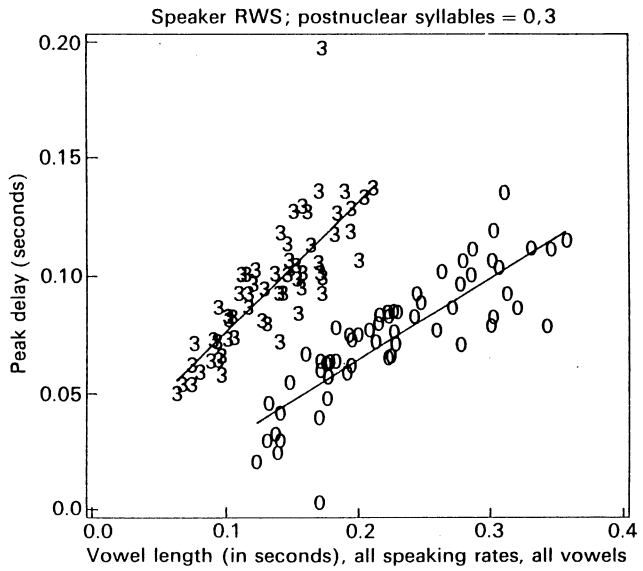


Figure 5.1 Position of nuclear peak, relative to vowel onset with either three (3) or zero (0) postnuclear syllables. From Steele (1986).

Such a transcription, in which the coordination between words and melody is mediated through the prosodic structure, is very nonspecific about exactly how segments in the two tiers are overlapped during pronunciation. And this is as it should be, because the phonological representation should capture only linguistic contrasts. Phonetic implementation rules, by contrast, have to specify how the phonological structure is realized in actual speech. Phonetic observations reveal a great deal of variation in how the realization of the pitch accents is coordinated with the realization of the speech segments. Furthermore, this variation is systematic. This point is illustrated in figure 5.1 with data from Steele (1986).[2] This study investigated the timing of the fundamental frequency peak for a nuclear H* accent,[3] as a function of speech rate and of the number of postnuclear syllables. The figure plots the peak delay (the distance of the peak from the vowel onset of the nuclear syllable) against the length of the vowel. Points labeled "0" are for utterances in which no syllables followed the nuclear stress, produced as the response in:

Did John write to Sue?
– No, he wrote to NAN.

Points labeled "3" are for utterances in which three syllables followed; these were produced as the response in:

Did John write to SUSIE Moore Lane?
– No, he wrote to NANA Moore Lane.

Regression lines summarize the trend for each of these two cases, as speech rate is varied. Points to the right within each group represent longer vowels, corresponding to utterances spoken at the slower rates. The way the 3's and 0's separate out into two distinct clouds makes clear that the peak is much earlier, relative to the total vowel duration, when no syllables follow than when three do. A related point is that speech rate interacts differently with peak alignment than phrase-final lengthening does. If they interacted in the same way, then the 3's and 0's would be located on top of each other in one cloud instead of two: it would be possible to predict the peak delay from the vowel length alone, without knowing the prosodic configuration. Clearly, this is not possible.

Here, we address two questions raised by Steele's data and other data like them. The first is: are prenuclear and nuclear accents similar or different in the tonal phonology? Theories of English intonation in the British school, such as O'Connor and Arnold (1973), use a different phonological inventory for prenuclear and nuclear accents. Within such a framework, phonetic differences between the two types of accent would be expected as a matter of course, because they would be the surface realizations of an underlying difference. Pierrehumbert (1980), on the other hand, claims that English has the same inventory for both types of accent. Furthermore, a uniform phonetic realization process is claimed to be responsible for explaining how accents in both positions are pronounced. This theory leads one to expect that data similar to those in figure 5.1 would also arise for prenuclear accents as prosodic context is varied. Silverman (1987) in fact questions this aspect of Pierrehumbert's phonology for a number of reasons. A survey of the relevant phonetic studies of F_0 contours showed a consistent tendency for nuclear peaks to be aligned much earlier in their syllables than prenuclear peaks. Alignment rules proposed by previous researchers (Mattingly 1966; Pierrehumbert 1981) have treated nuclear and prenuclear positions differently. In his own F_0 synthesis model, which incorporated Pierrehumbert's intonational phonology as one of its components, Silverman found that his phonetic implementation rules needed to know whether an accent was nuclear or prenuclear in order to compute how it should align with its associated syllable. The resultant alignment differences were crucial for the quality of the synthetic speech. Taken together, these considerations seemed to bring the underlying unity of prenuclear and nuclear accents into question.

This issue is attacked directly in our study, which examined the alignment of accent peaks in prenuclear position. Prosodic context was varied by varying the

proximity of the accent to a word boundary and by varying the number of syllables separating the prenuclear and nuclear accents. Speech rate was also varied. In general, differences similar to those in figure 5.1 emerged, supporting the hypothesis that prenuclear and nuclear accents are represented and realized by a uniform mechanism.

The second question is why differences like that in figure 5.1 occur. The observed contrast in alignment is apparently related to the fact that the "0" syllables are lengthened because of their prosodic position (i.e. utterance-final), whereas the "3" syllables are not. However, the mechanism for this relationship is unclear. Some of the alternatives include:

Invariance. The F_0 rise for a H* accent has an invariant duration for any given speech rate. The position of the peak relative to its syllable varies artifactually when independent factors (such as utterance-final lengthening) alter the segment durations without affecting the rise duration. In this account, observed differences in peak alignment when prosodic context is varied are considered to be consequences of low-level phonetic properties of speech.

Gestural overlap. The relatively early peaks observed for the "0" points occur because the articulatory gesture for the accent is interrupted by the gesture for the following Low (L) tone which marks the end of the phrase. The relation of the peak placement to the duration pattern arises indirectly, as a consequence of the fact that phrase ends in English are marked by tones as well as by lengthening. Similar effects could be found in prenuclear position when another accent occurs before the gesture for the first is completed.

Tonal repulsion. The articulatory gesture for the accent is moved earlier in time, in order for the accent and the following tone to be fully pronounced in the time available. The relation to the duration pattern arises the same way as in the overlap hypothesis.

Phonological mediation. Structural prosodic features modify the phonological representation in a way that speech rate changes do not, such as by adding extra beats to the metrical grid. The modified phonological representation causes certain syllables to be lengthened, and also gives rise to a difference in alignment of the pitch accent.

Sonority profile. The opening and closing gestures for the syllable give rise to an increase and decrease in sonority (where we define sonority loosely in terms of the overall openness of the vocal tract or the total impedance looking forward from the glottis). The sonority profile, or the time course of sonority for the syllable, differs between lengthening and nonlengthening environments because the closing gesture is more extended by prosodic lengthening than is the opening gesture. The F_0 gesture for the accent is coupled to the entire sonority profile of the syllable (not just aligned with the vowel onset, as under the invariance theory). The exact form of the coupling interacts with the different

effects of rate and prosodic lengthening on the sonority profile, and thereby yields differences in alignment.

All five of these theories predict that alignment differences will be found in prenuclear as well as nuclear position, as the prosodic context is varied. They differ among each other in their detailed predictions, and this will make it possible to compare them using data from our experiment. For example, the invariance and phonological mediation theories provide no mechanisms for varying alignment without co-varying duration. According to these accounts, differences in syllable-relative peak alignment must be either mere artifactual consequences of prosodically-induced lengthening (invariance), or else are induced at the same time by the same prosodic triggers (phonological mediation). Undershoot and tonal repulsion, on the other hand, do provide such a mechanism for varying peak placement independently of duration: contexts can be contrived in which the distance between tones varies without varying the structural factors relevant to the duration rules, and effects on peak alignment are predicted in such cases. The sonority profile theory permits contrasts in alignment pattern among syllables with the same overall duration, but only if syllable-internal details of the timing pattern vary.

In general, the data supported neither the invariance nor the phonological mediation theories. Invariance failed because different prosodic contexts at the same speech rate showed systematically different absolute distances between the vowel onset and the F_0 peak (as well as different relative positions of the peak within the syllable). Phonological mediation failed because the results suggested that alignment differences were gradient (ie. continuous) rather than discrete. A particular difficulty for the phonological account arose in one subset of the data, in which a systematic difference in alignment was not related to a difference in duration; the account offers no way of handling this case or of relating it to the rest of the data. The gestural overlap and repulsion theories were somewhat more successful, but still failed to provide a full explanation because the absolute distance between accents was less important than the right-hand prosodic features in determining peak placement. The subset of the data that caused difficulties for phonological mediation tends to support an account based on tonal repulsion, although it could be accommodated within the sonority profile approach if additional assumptions were borne out. A small investigation of syllable-internal timing in fact provided some evidence for an account based on sonority profiles. Our conclusion is that observed alignment differences arise through gestural overlap, through tonal repulsion, through coupling to the sonority profile, or (most likely) through a combination of these three mechanisms.

We can step back from these particular alternative formulations to make some observations about how these theories relate to each other. They differ on which components of speech production access which aspects of the phonological

representation, and on how they make use of it. The phonological mediation theory represents one extreme. In this approach, the phonetic implementation process does not have direct access to the hierarchical prosodic structure. Rather, the phonetic rules can only refer to an intermediate phonological representation in which those structural characteristics that are responsible for the observed phonetic differences have been explicitly re-encoded.

The invariance theory tends in the opposite direction, by removing from the phonological representation the burden of generating differences in peak alignment. It shares with the phonological mediation theory the view that the rules responsible for producing the F_0 contour for any pitch accent do not themselves access an utterance's hierarchical structure, but it does allow the duration rules to do so. This approach thereby somewhat enhances the status of phonetic implementation processes in explaining the sound patterns of speech. Gestural overlap is similar to invariance in this regard, although it takes a less simplistic view of the relationship between the underlying tonal sequence and its surface realization in a particular utterance.

Tonal repulsion requires the accent realization rules to access the hierarchical prosodic structure in a particularly powerful and complicated way. Instead of the alignment varying according to *structural features* in the right-hand context, it requires a computation of their *detailed phonetic consequences* – an estimate of how long in absolute time units until the next tonal element – before any accent can be pronounced.

The account in which F_0 peak alignment is related to the sonority profile lies somewhere between the above extremes. The form of the sonority profile for a syllable is influenced by factors which can be directly read from the hierarchical prosodic structure, without any need for an intermediate level of representation. Accent placement is coupled to the overall shape of the sonority profile, and so the same contextual factors which give rise to prosodically-induced lengthening also determine peak placement.

In what follows, we shall consider our phonetic data in the light of the above alternative formulations. We shall present evidence that phonetic implementation rules for F_0 and duration must directly access an utterance's hierarchical prosodic structure and use this information to compute the alignment of accents with their associated syllables.

5.2 Method

5.2.1 *Materials*

Two adult speakers took part in the experiment: one male (RWS – the same speaker whose data for peaks in nuclear position are shown in figure 5.1) and one female (JBP). Each speaker produced names of the form *Ma Lemm*, *Mom Le Mann*, *Mamalie Lemonick*, and *Mama Lemonick*, with all twelve combinations of

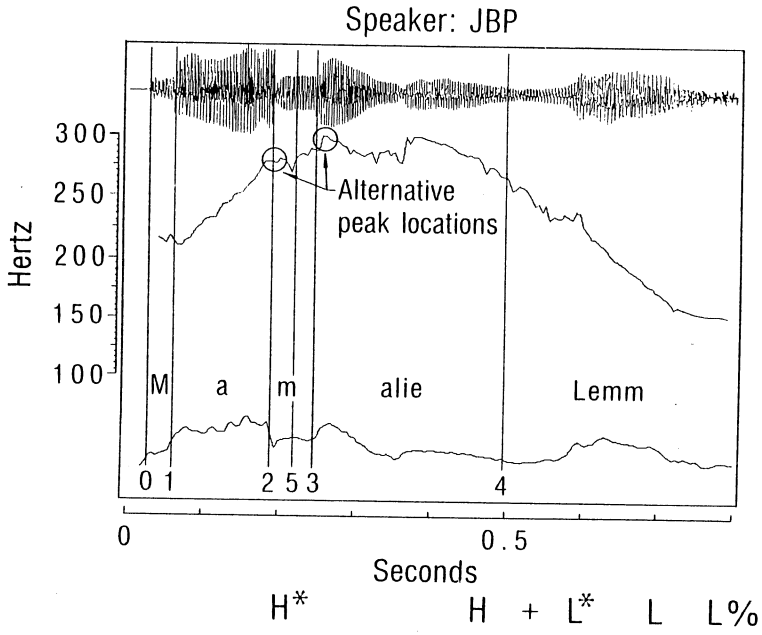


Figure 5.2 Extracted F_0 contour for *Mamalie Lemm*, spoken by JBP at the slow rate. The vertical lines numbered 0 through 4 represent measurement points for the segmental durations. Line number 5 is the chosen peak location, midway between the two possible candidates labeled with arrows.

the four first names and three surnames.[4] Names possessing entirely voiced sonorant consonants were used in order to minimize segmentally-induced perturbations on fundamental frequency. These names are chosen in order to vary the rhythmic configuration between the two syllables with word stress. The names differ in the number of syllables separating the stresses (0 to 3) and in the location of the word boundary with respect to the stresses. In some of the names (those beginning with *Mamalie* and *Mama*), the accented syllable does not immediately precede a boundary. In two (*Ma Le Mann* and *Mom Le Mann*), it is lengthened because it is the last syllable before the word boundary. In the remaining two (*Ma Lemm* and *Mom Lemm*), it is lengthened by two effects – it precedes a word boundary and it participates in a stress clash.[5] Subjects used a H* H+L* melodic pattern, which was illustrated by example to subject RWS. This pattern has a peak associated with the prenuclear stress, followed by a plateau and a step down onto the syllable with nuclear stress. The utterance ends with a low pitch, transcribed as the sequence L L%. An example is given in figure 5.2. This pattern was selected because it offers a conspicuous, measurable F_0 feature in prenuclear position, and because it seemed to encourage productions free of major phrasing breaks.

Each speaker produced, in randomized order, five repetitions of each combination in each of three speaking rates (fast, normal and slow), yielding altogether 360 utterances. Randomization was carried out in blocks, to control for practice and fatigue effects within the recording session.

5.2.2 *Measurements*

Measurements were made of the segmental durations and the F_0 maximum corresponding to the prenuclear H*, using a computer F_0 track and waveform display shown in figure 5.2. In *Ma*, *Mom*, and *Mama*, the durations of all segments were measured. In general, we believe that the time points for transitions between /m/ and vowels are quite accurate: enlargement of the relevant section of the acoustic waveform clearly revealed the boundary between the smooth glottal periods belonging to the interval of lip closure and the glottal periods showing higher frequency components (related to the formant structure) that appeared as soon as the lips were (even slightly) open. In addition, the waveform typically had a higher amplitude during the vowels (see measurement lines 1, 2 and 3 in figure 5.2). In *Mamalie*, the sequence 'alie' was measured as a unit, since the unstressed [1] could not always be so reliably segmented. The onset of the initial [l] of the surname was also somewhat problematic, especially when it was unstressed or followed an [m]. However, it was found fairly reliably by examining local perturbations in the autocorrelation and amplitude contours, by playing speech fragments, and by examining the formant transitions in computer-generated spectrograms. The time point for the F_0 maximum is probably the least reliable of the measurements. Segmental effects from the nasal and irregularities in the pitch made its location rather uncertain in many cases. Where there were two alternative locations that could be construed to be the relevant F_0 peak, we took the average between the two. The example F_0 contour in figure 5.2 illustrates one of the more uncertain cases: line 5 marks the chosen measurement point.

Of RWS's utterances, 2·8% had to be omitted from the data set because the speaker produced the wrong intonation pattern or because of articulation errors. After the initial measurements were taken and plotted, all utterances whose peak placements represented extreme outliers in the data set were individually examined. Values that were found to arise from measurement errors were corrected. This enterprise was conducted in a theoretically unbiased fashion; whenever an outlier was examined in a measurement set, another outlier which deviated from the trend in the opposite direction was also examined.

5.3 Statistical Analysis

In analyzing the data, we wished to look at the relationship between the structural characteristics of the utterances and the pitch accent alignment. A useful tool for modeling and statistically assessing such relationships is Multiple Regression.[6] Applying this to the present results, we attempt to "predict" the location of the

Table 5.1 *Coding of prosodic characteristics for all twelve names*

Name	wb	sc
Ma Lemm	1	1
Mom Lemm	1	1
Mama Lemm	0	0
Mamalie Lemm	0	0
Ma Le Mann	1	0
Mom Le Mann	1	0
Mama Le Mann	0	0
Mamalie Le Mann	0	0
Ma Lemonick	1	1
Mom Lemonick	1	1
Mama Lemonick	0	0
Mamalie Lemonick	0	0

F_0 peaks on the basis of other characteristics of the utterances. If there is a systematic relationship between the characteristics we choose and the alignment of the peaks, then the predictions will account for a significant amount of the variation in the data. The properties that we choose can be either quantitative or qualitative. In the latter case we represent a dichotomous feature as a binary variable: to represent the presence of word boundaries and stress clashes at the right-hand edge of the prenuclear stress, we have constructed two such variables; **wb** and **sc**. Each utterance receives a score of either 1 or 0 on each of these variables, depending on whether or not the corresponding feature is present. Hence for *Ma Le Mann*, which has a word boundary but no stress clash, **wb** = 1 and **sc** = 0. Table 5.1 lists these values for all twelve names.

In a similar way, the three speech rates can be encoded in a further two variables: **fast** and **slow**. The variable **fast** is assigned a value of 1 for all utterances spoken at the fast speaking rate, and 0 otherwise. Similarly, **slow** is assigned 1 for each slow utterance, and 0 otherwise. Utterances spoken at the normal speaking rate do not require a separate variable, since they are already uniquely specified by virtue of being neither fast nor slow and so having a value of 0 for both variables. The values of the variables **fast** and **slow** for each speech rate are summarized in table 5.2.

Relationships in the data can be evaluated by expressing them in algebraic (linear) equations using the above variables. We illustrate the technique here with a simple model in which we attempt to account for the peak placements in terms of speech rate alone. (The invariance hypothesis predicts that this should provide a good model of the data.) The equation is:

(2) **peak delay** = a + b.**fast** + c.**slow**
where **peak delay** is the distance from the start of each vowel to its F_0 peak, in milliseconds.

Table 5.2 *Coding of speech rate*

Speech rate	fast	slow
slow	0	1
normal	0	0
fast	1	0

Table 5.3 *Multiple Regression coefficients for prediction of absolute peak delay (in milliseconds) on the basis of speech rate alone, from equation (2)*

Speaker	a	b	c	R^2
JBP	151	−38	56	43·6%
RWS	109	−24	22	25·9%

Multiple Regression calculates the three coefficients (a, b and c) that give the best fit to the data, and thereby enables us to assess how much of the variance in the dependent variable (in this case **peak delay**) can be accounted for by the independent variables (in this case **fast** and **slow**, which jointly represent speech rate). This proportion, known as R^2, will here always be reported as a percentage of the overall variance in the dependent variable.[7] Table 5.3 gives the coefficients and R^2 values for this model.

These results predict that (for JBP) the peak delay in utterances spoken at the middle speech rate will be *on average* 151 ms. In the fast utterances peaks will be placed *on average* 38 ms. earlier, and in the slow utterances they will occur *on average* 56 ms. later (i.e. **peak delay** will be 151 − 38 = 113 ms., and 151 + 56 = 207 ms., respectively).

We emphasized *on average* for a reason. Although the signs of the coefficients in table 5.3 indicate that the relationship between speech rate and peak placement was in the same direction for both speakers, the R^2 values in the table indicate that speech rate alone accounted for less than half of the variation in JBP's data, and only about a quarter of the variation for RWS. This model is not particularly successful because it ignores any influence of the prosodic features (the **wb** and **sc** variables were not included in the equation). We shall show that the effects of these variables were systematic, comparable in magnitude (but different in nature) to the effects of speech rate, and therefore are a necessary component of any model of peak placement.

5.4 Results

5.4.1 *Upcoming context affects peak placement*

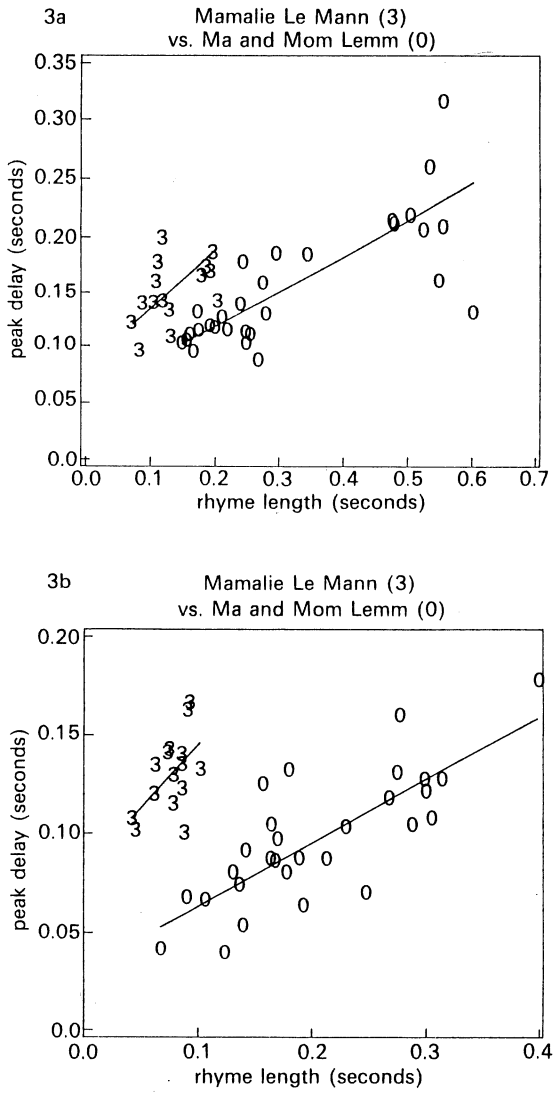
The locations of the F_0 peaks, relative to the onsets of the associated vowels, showed considerable variation in the productions of both speakers. Figures 5.3a and 5.3b show that prosodic context determines the alignment of peaks with the accented syllables, in a way that is qualitatively different from the effects of speech rate. In these figures we have plotted the data for those names that correspond most closely to the subset of Steele's data which we presented in figure 5.1. Points labeled "3" represent the utterances where three unstressed syllables separate the prenuclear and nuclear accents (i.e. *Mamalie Le Mann*), while points labeled "0" indicate no intervening syllables (*Ma Lemm* and *Mom Lemm*). The horizontal axis represents the length of the rhyme in the syllable bearing the prenuclear accent (the /om/ in *Mom* and the /a/ in *Ma* and *Mamalie*). The delay from the start of the rhyme to the F_0 peak is plotted on the vertical axis. Regression lines summarize the trend for each of the two prosodic structures.

The data exhibit a contextually-governed variation similar to that shown by Steele for nuclear pitch accents. Points for each name fan out into an elongated cloud, with the points further from the origin in each cloud representing slower utterances. The important pattern in the plots is that the separate clouds for each name are almost completely nonoverlapped. Within the cloud of points for each name, the relationship between peak delay and vowel length can largely be modeled by a line with a positive slope: when a vowel is lengthened because the utterance was spoken more slowly, the peak is correspondingly delayed. In contrast to this, the difference between the clouds shows that when a syllable is lengthened because of the upcoming prosodic context (as in *MA Lemm*), the peak is aligned much earlier in the syllable.

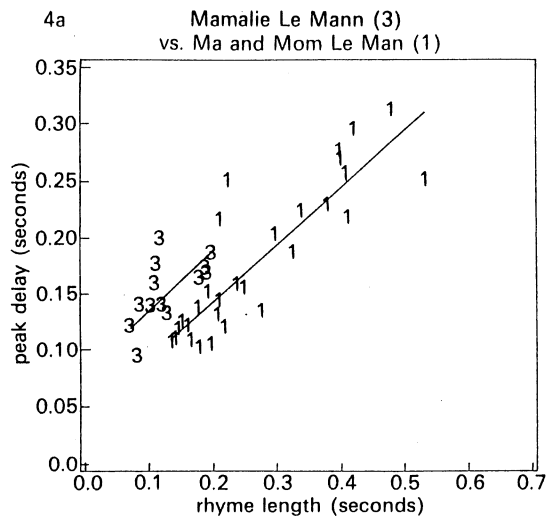
Comparisons between other names in our corpus show that the difference in peak alignment between *Mamalie Le Mann* and *Mom Lemm* arises as a result of two contributing factors. Figures 5.4a and 5.4b illustrate that one of these is word-final lengthening. The plots compare the peak alignment in *Mamalie Le Mann* ("3") with the alignment in *Ma* and *Mom Le Mann* ("1"). In the latter names, the accented syllable is word-final and therefore longer, but the peaks occur earlier in the syllable rhymes.

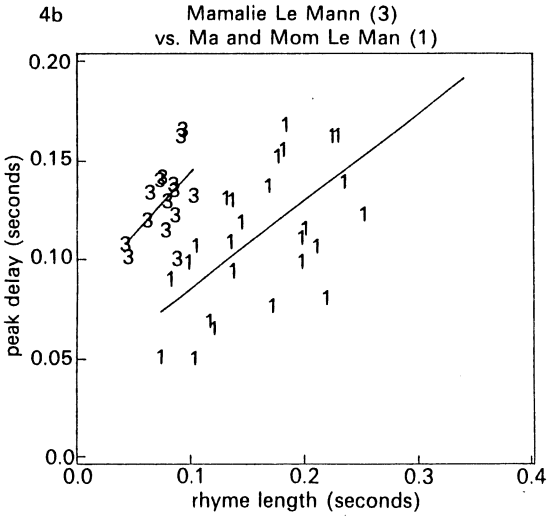
The second factor that contributes to early peak alignment is the effect of a stress clash. Figures 5.5a and 5.5b compare the data for *Ma* and *Mom Lemm*, in which the first syllable is lengthened by a stress clash as well as a word boundary, with *Ma* and *Mom Le Mann*, in which the first syllable is lengthened by a word boundary alone. In the former case we see that when a stress clash is present, syllables are even longer but peaks are again relatively earlier.

These results show a systematic effect of the right-hand prosodic context on peak alignment, but they might seem to suggest a simpler explanation; namely that



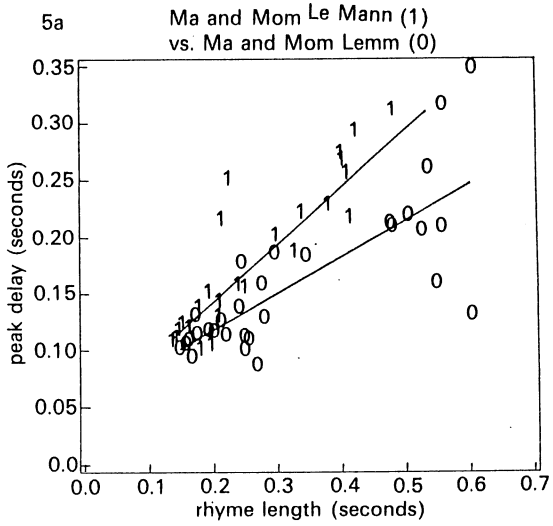
Figures 5.3a (JBP) and 5.3b (RWS) Peak delay, relative to the onset of the vowel, as a function of the length of the syllable rhyme (/a/ in *MA* and *MAmalie*, /om/ in *MOM*). "3" = *Mamalie Le Mann*, "0" = *Ma* and *Mom Lemm*.

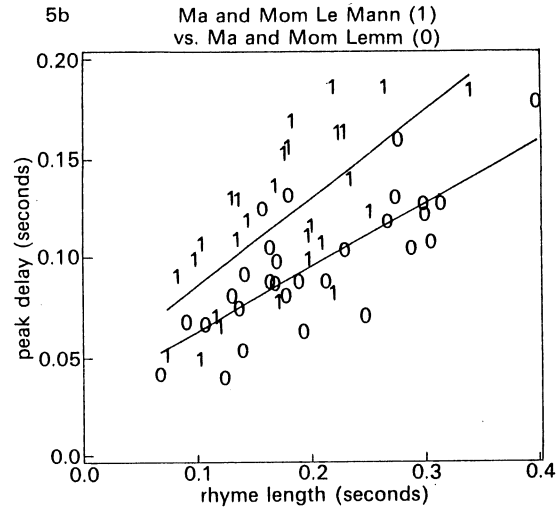




Figures 5.4a (JBP) and 5.4b (RWS) Peak delay, relative to the onset of the vowel, as a function of rhyme length. "3" = *Mamalie Le Mann*, "1" = *Ma* and *Mom Le Mann*.

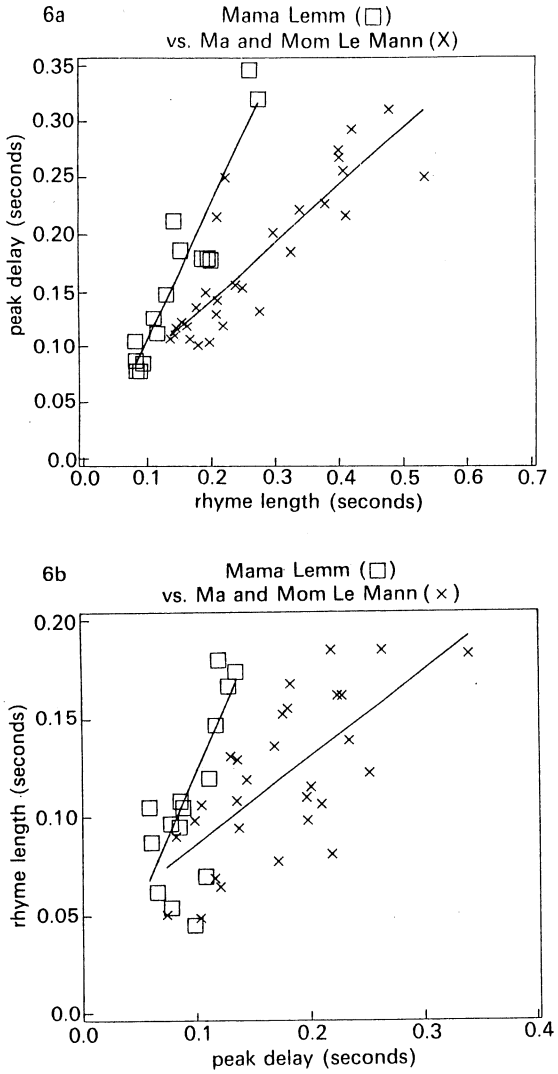
rather than stress clashes or word boundaries, the relevant contextual feature is merely the number of following unstressed syllables. This simpler model is not supported by the data, however. Figures 6.6a and 6.6b compare peak alignment in *Mama Lemm* (□) with that in *Ma* and *Mom Le Mann* (×). All of these





Figures 5.5a (JBP) and 5.5b (RWS) Peak delay, relative to the onset of the syllable rhyme, as a function of rhyme length. "1" = *Ma* and *Mom Le Mann*, "0" = *Ma* and *Mom Lemm*.

utterances have only one intervening unstressed syllable, yet for both speakers the peaks are aligned earlier in the latter names, where the syllable bearing the prenuclear accent directly precedes a word boundary. Hence the differences in figures 5.6a and 5.6b are another manifestation of the same effect that was illustrated in figures 5.4a and 5.4b.



Figures 5.6a (JBP) and 5.6b (RWS) Peak delay in syllables followed by only one unstressed syllable. □ = *Mama Lemm,* x = *Ma* and *Mom Le Mann.*

5.4.2 *A statistical model of peak alignment*

The above figures all show a tendency for peak delay differences between names to be larger at the slower rates, when the rhymes are longer. We have found that it is not the absolute peak delay, but rather the peak placement in proportion to

Table 5.4 *Regression coefficients for the proportional placement of F_0 peaks, according to equation (3)*

Speaker	a	Prosodic context b	c	Speech rate d	e	R^2
JBP	1·223	−0·487	−0·170	0·072	−0·212	64·2%
RWS	1·432	−0·774	−0·191	0·219	−0·038	62·9%

the syllable rhyme length, that exhibits the most regular patterns.[8] Our statistical analysis reflects this by modeling peak proportions (peak delay divided by the duration of the associated syllable rhyme).[9] With the peak delays expressed as proportions in this way, we had far greater success in predicting the data on the basis of prosodic context and speech rate. The regression equation was:

(3) **peak proportion** = a+b.**wb**+c.**sc**+d.**fast**+e.**slow**
where **peak proportion** is the peak delay divided by the rhyme length

and the results are summarized in table 5.4. The R^2 values show that for both speakers this model accounts for nearly $\frac{2}{3}$ of the variance in the data. It captures the relationships illustrated in the previous figures, and shows (by the signs of the coefficients) that all of the effects work in the same direction for both speakers. The equations mean that in the absence of a word boundary or a stress clash, the peak occurs past the end of the rhyme, with a minor adjustment of the offset dependent on the speech rate. In the presence of a word boundary, the peak is moved earlier by a relatively large fixed proportion of the rhyme (0·487 for JBP, 0·774 for RWS), and it is moved even earlier by a fixed proportion (0·170 for JBP, 0·191 for RWS) in the presence of a stress clash. Both **wb** and **sc** are necessary in the model; dropping either of them significantly worsens the fit. The fact that the peaks are past the end of their associated syllable in utterances like *Mamalie Le Mann* is represented by the coefficient "a" being greater than 1. In fast speech, syllables are shorter and peaks occur proportionately later, so the "d" values are positive. Similarly, slow speech causes peaks to occur earlier in their syllables, so the "e" values are negative.

This model assumes that the effects of prosodic context on the proportional peak alignment are independent of speaking rate. For example, in the case of JBP a word boundary will decrease the peak proportion by nearly half of the rhyme-length (0·487) at all rates, regardless of its rate-dependent initial value. We can test this assumption by adding a set of variables to the equations in order to express each aspect of the possible interactions, as shown in table 5.5. When we repeat the analyses with these terms in the equations, we find that they add a very small,

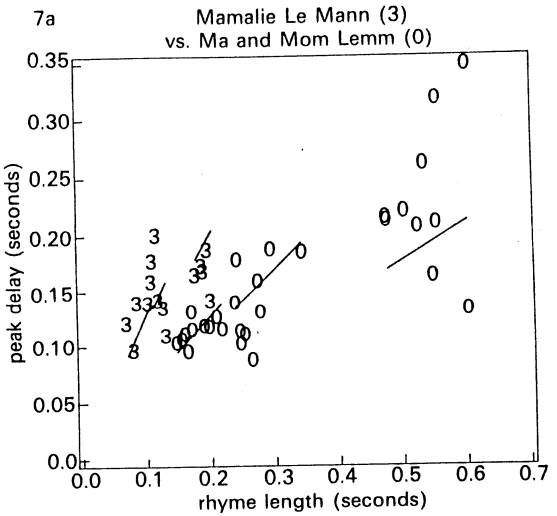
Table 5.5 *Variables carrying all aspects of an interaction between prosody and speech rate*

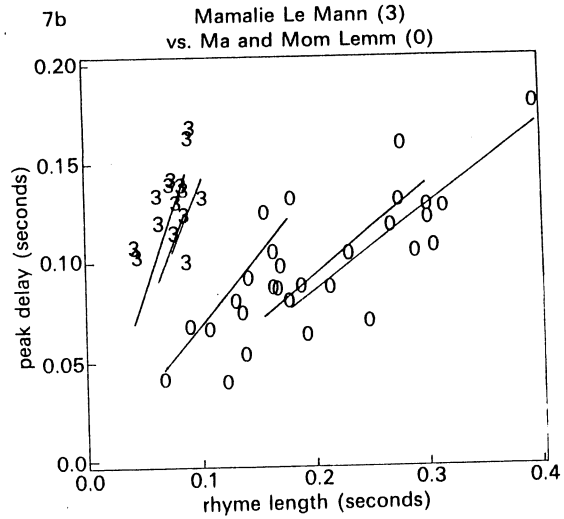
Interaction term	Interpretation
wb × fast	Adjustment to the word-boundary effect when speech rate is fast
wb × slow	Adjustment to the word-boundary effect when speech rate is slow
sc × fast	Adjustment to the stress-clash effect when speech rate is fast
sc × slow	Adjustment to the stress-clash effect when speech rate is slow

statistically barely significant amount to the variance already accounted for by the model of equation (3) (for JBP a further 1·6%, for RWS an extra 2·1%[10]). We further find that unlike the main effects of prosody and speech rate that we showed in table 5.5, the directions of the interactions are not consistent across the two speakers. For example, the largest interaction for JBP (accounting for 1·3% of the variance) was that the word-boundary effect was decreased in magnitude by 0·175 in the slow speaking rate. For RWS, however, in the slow speech rate the word-boundary effect was changed in the opposite direction: it *increased* by 0·254. Since the interactions were inconsistent across the two speakers and in any case did not clearly reach statistical significance, we believe that the data confirm that the effects of right-hand prosodic context on peak alignment (expressed as a proportion of the syllable rhyme) bear no systematic relationship to speech rate.

To summarize so far, it was the proportional alignment of F_0 peaks with their associated syllables, rather than the absolute distance in time, that exhibited rule-governed behavior. The effects of prosodic context and speech rate are consistent across both speakers in the way they influence peak placement, and they operate independently of each other. This simple additive model accounts for nearly $\frac{2}{3}$ of the variance in the data. In figures 5.7a and 5.7b the data showing the combined prosodic effects are replotted (from figures 5.3a and 5.3b) along with the values predicted by this model.

The R^2 values yield an intuitively tractable but stringent evaluation of our model. Statistically, even if we had only been able to account for as little as 8% of the variance in the data our result would still have been highly significant ($F_{(4,175)} = 3{\cdot}44$, $p < 0{\cdot}01$). Multiple regression models do not often achieve R^2 values as high as those presented here. Nevertheless we may ask why the fit was not even better. Three extraneous sources of variation in the data were: (i) uncertainty in measuring the precise peak locations in the presence of segmentally-induced perturbations, as mentioned earlier; (ii) slight differences between *Ma* and *Mom* for RWS (peaks placed in an earlier proportion of the rhyme for *Mom*),





Figures 5.7a (JBP) and 5.7b (RWS) Peak delay, relative to the onset of the vowel, as a function of vowel length: observed versus predicted values. "3" = *Mamalie Le Mann*, "0" = *Ma* and *Mom Lemm*. The lines represent the proportions predicted by equation (3) for each rate, multiplied by the corresponding actual rhyme durations.

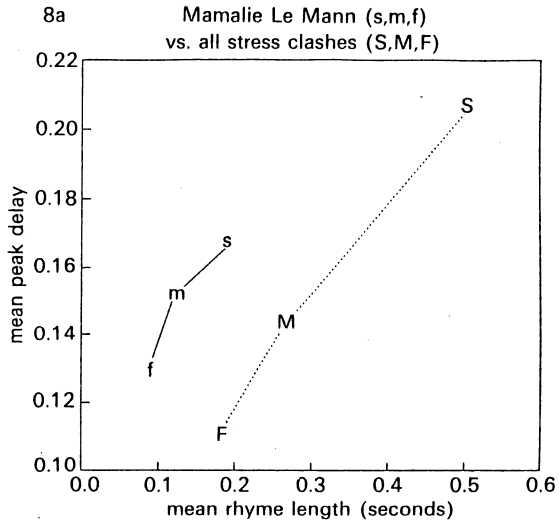
and (iii) JBP alternated between two different strategies for producing the slow condition: sometimes she would increase the amount of prosodic lengthening while retaining the proportional peak placement, and other times she would increase the syllable durations without changing the peak delay. This latter strategy was particularly evident when there was no adjacent prosodic lengthening trigger, such as in *Mamalie Le Mann*.

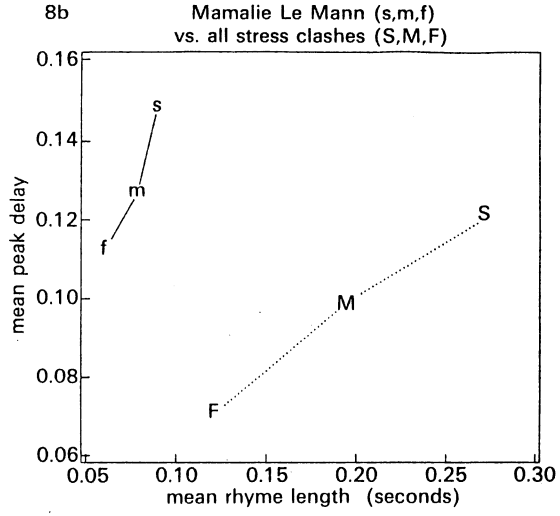
5.4.3 *A related model of duration*

How do the influences of right-hand context on peak alignment arise? They seem to be the result of a combination of two different phonetic consequences of the prosodic structure. One of these is that the absolute time, and not just the relative time, between the vowel onset and the F_0 peak is decreased when a word boundary or stress clash follows. Figures 5.8a and 5.8b show the average peak delay as a function of speech rate for the combined set of *Ma Lemm*, *Mom Lemm*, *Ma Lemonick* and *Mom Lemonick* (i.e. those utterances with both a word boundary and a stress clash at the right-hand edge of the prenuclear syllable), compared with *Mamalie Le Mann* (which has neither of these features). In all three of the speech rates, RWS placed his peaks earlier – significantly closer to the vowel onset – in the former set of names. JBP's peaks showed a slightly more complicated pattern: her peaks in the former set of names were earlier than those in the latter set in the fast rate; in the normal rate the mean peak location was also earlier, though this difference did not reach statistical significance, and in the slow rate they were significantly later.[11] Overall, the tendency was that for the most part the right-hand prosodic context alters the F_0 trajectories by pushing the F_0 peaks to the left.

The second means by which the right-hand context affects peak placement is indirect, via lengthening of the segmental material. Conspicuously, those environments in which peaks are early correspond to those environments in which prosodic lengthening occurs. This leads us to suspect that peak placement is related to the temporal structure: whatever it is that triggers prosodic lengthening may also affect proportional peak location. Two pieces of evidence lend support to this hypothesis.

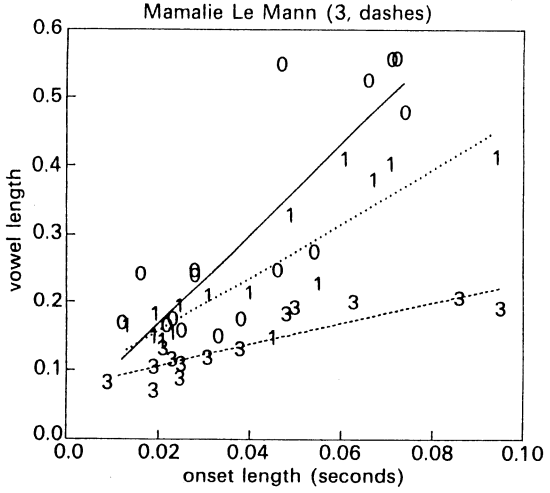
Firstly, as figures 5.8a and 5.8b indicate, the durations of the syllable rhymes show the same ranking as the peak alignment: earlier peaks tend to occur in longer syllables. Figures 5.9a and 5.9b present this generalization in a different way for three of the names: *Ma Lemm* (word boundary plus stress clash), *Ma Le Mann* (word boundary but no stress clash), and *Mamalie Le Mann* (no word boundary and no stress clash). The horizontal axis is the duration of the word-initial /m/, and the vertical axis is the length of the /a/.[12] What we see is that for both speakers the vowels are longer, relative to their preceding syllable-initial consonants, in precisely the same contexts where peaks are earlier. The scatter in the plots is due to inconsistencies in the length of the initial consonant – speakers



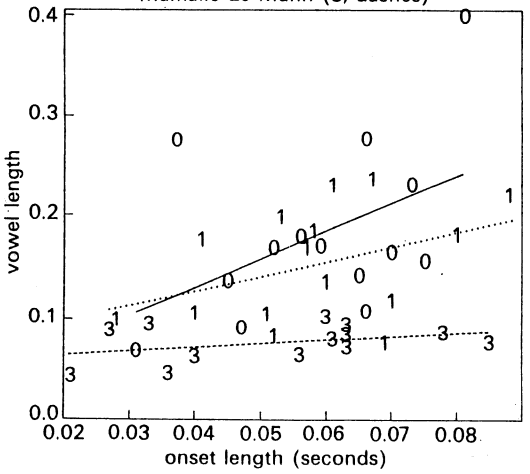


Figures 5.8a (JBP) and 5.8b (RWS) Peak delay versus vowel length, averaged within each speech rate. Lower case letters represent *Mamalie Le Mann* ("s" = slow, "m" = medium, "f" = fast), upper case letters represent the group of *Ma Lemm*, *Mom Lemm*, *Ma Lemonick* and *Mom Lemonick* ("S" = slow, "M" = medium, "F" = fast).

9a Ma Lemm (0, solid line), Ma Le Mann (1, dots), Mamalie Le Mann (3, dashes)



9b Ma Lemm (0, solid line), Ma Le Mann (1, dots), Mamalie Le Mann (3, dashes)



Figures 5.9a (JBP) and 5.9b (RWS) Vowel length as a function of the length of the preceding /m/. "0" = *Ma Lemm*, "1" = *Ma Le Mann*, "3" = *Mamalie Le Mann*.

do not seem to exert such fine control here. Nevertheless, as the fitted regression lines show, the influence of right-hand prosodic context on vowel length emerges through the noise and is quite consistent.

A better way to assess this ranking is to apply exactly the same multiple regression model to the rhyme lengths as we did to the peak proportions. This

Table 5.6 *Regression coefficients for rhyme lengths in milliseconds*

		Prosodic context		Speech rate		
Speaker	a	b	c	d	e	R^2
JBP	106	129	51	−55	152	83·8%
RWS	81	92	24	−45	42	74·8%

allows us to use the data from all twelve names, rather than selecting only three of them, and does away with the need to rely on the noisy data of the word-initial /m/ durations. If peak alignment arises from prosodic lengthening, then the same factors that were shown via equation (3) to account for the earlier placement of F_0 peaks in the syllable rhymes should explain as much or even more of the variance in the rhyme lengths.

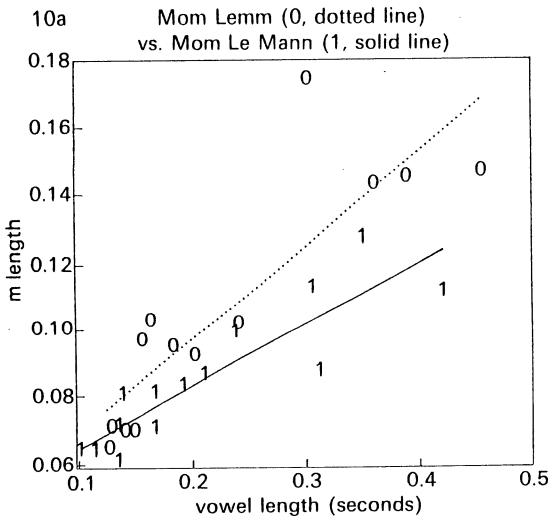
The model is:

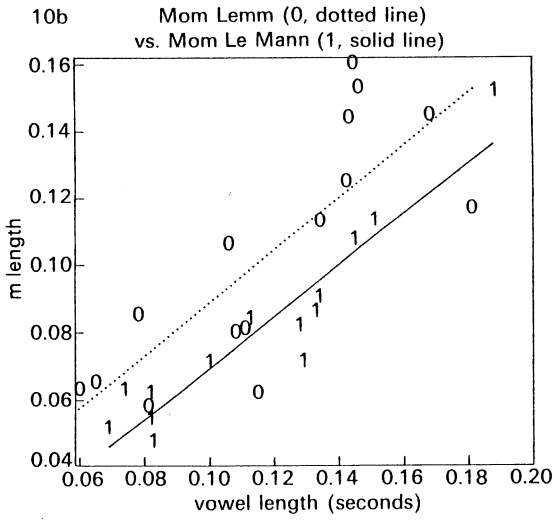
(4) **rhyme length** = a + b.**wb** + c.**sc** + d.**fast** + e.**slow**

and the results are summarized in table 5.6. The most relevant information in the table is the prosodic effects and the amount of variance explained.[13] Word boundaries and stress clashes lengthen syllable rhymes, by the number of milliseconds in columns "b" and "c" (the reader will recall that **wb** and **sc** only have values of 0 or 1, and thereby act as binary switches for the effects whose magnitude is represented by "b" and "c"), across all speech rates. The very high R^2 values show that syllable rhyme durations conform to this pattern even more consistently than the peak placement data.

Note that according to this model, the amount of lengthening applied to a syllable preceding a word boundary or a stress clash is a constant, in milliseconds, in all three speech rates. In reality, it is more likely that the magnitude of the prosodic effects expressed in such absolute time units would be rate-dependent. In other words, there may be statistically significant interactions between the two factors. Analyses of the rhyme durations with the same interaction terms as in table 5.5 indeed did show such interactions. For both speakers, the effects were larger in slow speech and smaller in fast speech, and this rate-dependence explained significantly more of the variance: 92·4% (as opposed to 83·8%) for JBP and 82·1% (as opposed to 74·8%) for RWS.[14]

The second piece of evidence that peak placement is related to prosodic lengthening sheds more light on the mechanism by which this lengthening occurs. In figures 5.10a and 5.10b, the duration of the final /m/ in *Mom* is plotted as a function of the vowel length for *Mom Lemm* and *Mom Le Mann*. What we see is that when the rhyme is lengthened by stress clash in *Mom Lemm*, the last part of





Figures 5.10a (JBP) and 5.10b (RWS) Duration of the final /m/ in "Mom," as a function of the duration of the vowel. "0" = *Mom Lemm*, "1" = *Mom Le Mann*.

the rhyme is lengthened the most. Thus the cloud of points for *Mom Lemm* is higher than in the case of *Mom Le Mann*.

These then are the two ways that right-hand prosodic context influences peak placement: when a syllable is lengthened by an upcoming word boundary or stress clash, then the peak is moved to the left. At the same time, the syllable is

lengthened in such a way that its right-hand edge is moved to the right, relative to the location of the accent peak.

However, this coincidence between durational effects and effects on peak delay is not perfect. As well as being subject to the influence of prosodic triggers located immediately to the right of the syllable bearing the prenuclear accent, RWS's peaks occurred later when more unstressed syllables separated that syllable from the lengthening trigger. To evaluate the consistency of this effect statistically, we can replace the dichotomous **sc** variable in the regression equations with a quasi-gradient variable, which we call **gradsc**. This variable ranks the names according to the proximity (in syllables) of the upcoming nuclear pitch accent in the surname. It has a value of 1 for *Ma Lemm* and *Mom Lemm*, which constitute the stress clash case expressed by **sc**, and a value of 0 in *Mamalie Le Mann*, where the greatest number of syllables intervenes between the two accents. In all other names **gradsc** has a value of $\frac{1}{3}$ or $\frac{2}{3}$, according to whether there are two or one intervening syllables, respectively. If the effect of the stress clash on peak alignment is indeed a systematically gradient phenomenon, then replacing **sc** by **gradsc** in equation (3) should significantly improve how well the model fits the data. This is precisely the result we obtain: the amount of variance explained increased significantly from 62·9% to 69·3%.[15] This effect is not reflected in the durations of the syllables: replacing **sc** by **gradsc** in equation (4), which modeled the rhyme lengths, decreased the fit of the model instead of increasing it. Consequently the longer-range influence on the peak proportions seems to have been caused by the peaks being pushed to the left, as if they were somewhat repelled from the upcoming nuclear accent. Unlike the influence of a stress clash on rhyme durations, this tonal repulsion is a gradient phenomenon that extends with decreasing magnitude over a number of intervening unstressed syllables, and it occurs without any concomitant lengthening of the accented syllable.

5.4.4 *Summary of the results*

Both speech rate and right-hand prosodic context influence F_0 peak placement, but they do so in qualitatively different ways. When a syllable is lengthened from being spoken more slowly, the peak will occur corresponding later. In contrast, when the lengthening is induced by the right-hand prosodic context, the later part of the syllable undergoes disproportionately more lengthening and at the same time the peak will occur earlier in the syllable rhyme. In addition to this length-related effect, for one of the speakers a leftward push on the prenuclear peak is exerted by the upcoming nuclear pitch accent; this extends over several syllables, and has no concomitant influence on duration.

5.5 Discussion

5.5.1 *Nuclear versus prenuclear peaks*

The data from the present experiment, when compared with that of Steele (1986), establish clear parallels between prenuclear and nuclear H* pitch peaks. Prenuclear peaks, like nuclear peaks, are aligned in proportion to the duration of the associated syllable, rather than a fixed distance into the vowel. In both positions, speech rate and right-hand prosodic context have different effects on peak location. Also, in both positions peaks are aligned later when there are more unstressed syllables following the syllable bearing the accent.

In addition to establishing these parallels, the current results provide further insight into the relevant contextual features and the mechanism by which peaks are aligned. It is not primarily the presence or number of following unstressed syllables *per se*, but rather the amount of prosodic lengthening that most directly determines where peaks will be placed. Consequently any contextual feature which induces prosodic lengthening will align peaks earlier in their syllables, be it a stress clash, word boundary, or (as found by Steele) utterance-final lengthening. If, on the other hand, a syllable's duration is varied by other factors, such as changing the speech rate, or substituting a vowel with a different intrinsic duration (Steele, 1986), then the peak delay will shift in such a way that its relationship to the overall syllable duration will be maintained. In section 5.5.2.4, below, we will return to the obvious question of exactly how prosodic lengthening might be unique.

One difference remains between our data on prenuclear H* accents and Steele's data for the same accents in nuclear position; namely that peaks are absolutely earlier if the words that bear them are in nuclear position than if they are prenuclear. However it may still be possible to explain this difference in the framework of a single set of tonal implementation rules.[16] Phrase-final nuclear syllables, which have the earliest peaks by far (compare the "0" points in figure 5.1 with those in figure 5.3b), undergo the greatest amount of prosodic lengthening and so we would expect peaks on these syllables to be correspondingly earlier. In addition, in these cases there is a Low (L) tone immediately following the H*, on the same syllable. Even in Steele's data for non-phrase-final nuclear accents the H* is still followed by a Low. If tonal repulsion is a factor in peak alignment, this would exert an extra left-ward push on the H* that would be absent in the case of prenuclear accents.

5.5.2 *Mechanisms for similarity*

5.5.2.1 INVARIANT RISE-TIME

At first glance one might be tempted to assume that the durational structure alone, in combination with a rate-dependent but otherwise invariant F_0 rise time, would account for the data. Such a view would be predicted, for example, within the

framework of the Dutch school of intonation (e.g. 't Hart and Collier, 1975; de Pijper, 1983). According to such a model, the absolute peak delay for all utterances spoken at a particular speech rate is constant; any apparent earlier peak placement wholly arises from prosodically-induced lengthening, and so is an artifact of our analysing peak proportions rather than absolute peak delays. However, closer inspection of the data shows that this model is insufficient to explain a number of the patterns. Figures 5.8a and 5.8b showed absolute peak delays within each speech rate in lengthening versus nonlengthening environments. For RWS, peak delays were shorter (i.e. closer to the vowel onset) in prosodically lengthened syllables in all three speech rates. For JBP the peak delays were also shorter in the lengthened syllables spoken at the fast rate, but were nearly the same in the normal rate, and longer in the slow rate. These differences within each rate between prosodically lengthened and relatively unlengthened syllables run counter to the prediction of the invariance hypothesis. The figures illustrated how within each utterance type the mean delays for all three speech rates are positioned along lines representing prosodically-determined proportions. The one pattern that did *not* occur was the very pattern required by an invariant rise-time: constant absolute peak delay for all utterances within each speech rate (i.e. no shift in the vertical axis despite a rightward shift on the horizontal axis).

Another pattern that contradicts the possibility of an invariant rise-time is the one in RWS's data that we described earlier, whereby his peaks were later in absolute time when more syllables separated the accented syllable from the lengthening trigger, while the accented syllable's duration was not correspondingly adjusted. If the rise to the prenuclear peak had an invariant duration, then the number of following unstressed syllables could not exhibit any such relationship with the peak delay in the syllable bearing the prenuclear accent.

5.5.2.2 GESTURAL OVERLAP AND TONAL REPULSION

A less simplistic phonetic explanation for the variation in peak alignment is possible if we allow temporal overlap of the underlying gestures for the prenuclear and nuclear pitch accents in those cases where there is insufficient intervening unstressed material separating them. Such an explanation is in the spirit of Bruce's description of Swedish intonation (1977), and is not unlike Browman and Goldstein's approach to coarticulation (this volume). If we consider the present data in this light, we notice that in the nuclear accent (H+L*) F_0 steps down onto the nuclear syllable from an immediately-preceding higher level. This movement must begin before that syllable, and so it is possible that it would overlap the prenuclear rise (the movement for the prenuclear H*) when the two accents are juxtaposed.

It is difficult to generate testable predictions of this hypothesis without making some extra assumptions about how the laryngeal gestures interact when they

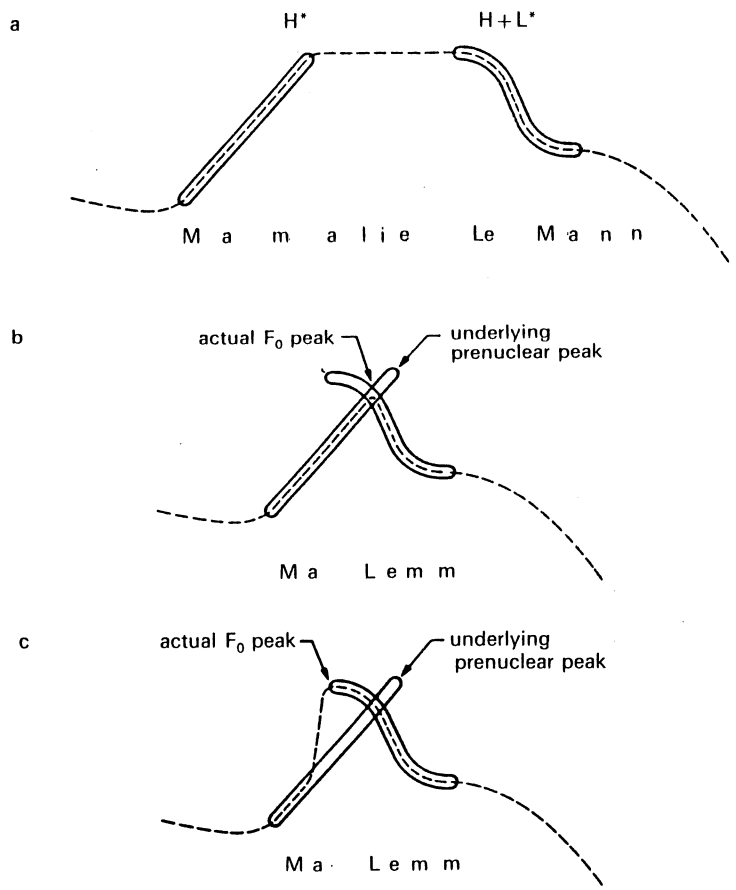


Figure 5.11 Schematic representation of underlying accent gestures for H* and H+L*, and possible resultant F_0 contours, in nonoverlapped (a) and overlapped (b and c) cases.

overlap. Two possibilities are illustrated in figure 5.11. The upper diagram (5.11a) schematizes the gestures for the prenuclear and nuclear accents realized on *Mamalie Le Mann*, with sufficient intervening material to separate them. The lower two diagrams concern the case for *Ma Lemm*, where the gestures overlap each other. In 5.11b, the F_0 contour (the dashed line) follows the minimum of the values specified by the two gestures. In 5.11c the second gesture wins out over the first. In both, the observed F_0 peak is displaced to the left of its "underlying" location.

These rather simple models are of course not the only possible ways that overlapping underlying accent gestures can be implemented. For example,

Silverman (1987) took a hybrid approach in which a preceding accent (the prenuclear H*, in the current example) will partly override the first, nonaligned tone of a following bitonal accent (in this case the H of the nuclear H+L*) but will itself yield to the aligned tone (in this case the L* of the H+L*) in those (probably few) cases when there is a great deal of overlap. An extra constraint was that the nonaligned tone could be effectively shortened but never completely elided by this process, so that the shortest addressable section of the H immediately adjacent to the L in H+L* would be maintained by the phonetic realization rules.

We have no way to directly estimate the times for the gestures corresponding to the two accents. However, according to the gestural overlap hypothesis, the relation of the underlying melodic gestures to the segmental gestures is not being varied. This permits us to derive the prediction that the F_0 peak alignment can be computed as a function of the distance in time between the accented syllables.

This prediction is shared by the tonal repulsion hypothesis, although it arises differently there. Tonal repulsion means that the entire gesture for the first accent will be shifted earlier when the next accent follows too closely in time. (There is an implicit functional assumption that the timing of the gesture must be adjusted so that the gesture can be carried out with sufficient completeness.) The overlap and the repulsion models differ in principle in their predictions about the F_0 values at the peak and during the rise. However, this difference is difficult to evaluate in practice, because of our lack of understanding of gestural overlap. In figure 5.11b for example, the F_0 peak is lowered by the overlap, but in 5.11c it is not. Accordingly, we confine our attention to the relation between peak delay and temporal separation between the accented syllables, and thereby treat the two theories together.

In the model we have already developed to account for peak alignment, in which we showed that peak proportions exhibited more rule-governed regularities (behaved in a more regular rule-governed fashion) than absolute peak delays, we accounted for 64%–70% of the variation in the data. The model predicted peak placement on the basis of the prosodic variables **wb** and **gradsc**. If early peak placement is solely the result of gestural overlap or tonal repulsion, then the success of the model was entirely due to the tendency for **wb** and **gradsc** to carry gross information about the distance between the accents. If this is true, then we should be able to account for an even greater percentage of the variation in the peak alignments (or at worst a comparable amount) on the basis of the actual temporal separation of the accents.

In the absence of the relevant articulatory data concerning the precise starting-points of the prenuclear and nuclear gestures, it seems both reasonable and consistent with the overlap/repulsion hypotheses to assume that the temporal distance between the accents is closely correlated with the distance between the onsets of the accented syllables. Consequently we selected out those utterances

Table 5.7 *Regression coefficients for proportional placement of F_0 peaks in names ending with Lemm and Lemonick, according to equation (5)*

Speaker	Prosodic context a	b	c	Speech rate d	e	R^2
JBP	1·293	−0·350	−0·375	0·043	−0·187	67·1%
RWS	1·528	−0·216	−0·862	0·248	−0·016	65·9%

ending in *Lemm* and *Lemonick* (i.e. two thirds of the total corpus), because the onset of the /l/ in these surnames was the start of the nuclear syllable. The inter-accent distance was thus estimated by calculating for these utterances the distance between the onset of the vowel bearing the prenuclear accent and the onset of the nuclear /l/.[17] We shall call this variable **IAdist**.

The model of peak proportions against which we are comparing the tonal undershoot hypothesis is

(5) **peak proportion** = a + b.**wb** + c.**gradsc** + d.**fast** + e.**slow**

For comparing this model with one representing the overlap/repulsion hypotheses, we must restrict our analysis to those names for which we also have **IAdist** values. Table 5.7 gives the coefficients derived from a multiple regression of this model using this subset of the utterances. If the hypotheses are to account for the data, then the absolute peak delays should be a rate-dependent function of the inter-accent distance:

(6) **peak delay** = a + b.**fast**.**IAdist** + c.**slow**.**IAdist** + d.**fast** + e.**slow** + f.**IAdist**

This can be rewritten in the form:

peak delay = (b.fast + c.slow + f).IAdist + (d.fast + c.slow + a)

This latter form makes clearer that peak delay is predicted as a simple linear function of **IAdist**, where the coefficient and constant vary according to the speech rate. Therefore the results are presented in table 5.8 separately for each rate, in the form:

(7) **peak delay** = g.**IAdist** + h

These regressions explain a statistically significant proportion of the peak delay variation (for JBP, $F_{(5,114)} = 21{\cdot}301$, $p \ll 0{\cdot}001$; for RWS, $F_{(5,111)} = 12{\cdot}476$, $p \ll 0{\cdot}001$). But at the same time they are significantly less successful than our own model was in its account of the peak proportions (in a test of the difference

Table 5.8 *Regression coefficients from analyses using equation (6), expressed in the form of equation (7), for all names except those ending in Le Mann*

Speaker	fast g	h	mid g	h	slow g	h	R^2
JBP	0·116	83·2	0·169	96·3	0·099	152·2	48·3%
RWS	0·210	46·5	−0·003	99·7	0·277	39·7	36·0%

between the two models: for JBP, $z = 2{\cdot}352$, $p < 0{\cdot}01$; for RWS, $z = 3{\cdot}490$, $p < 0{\cdot}001$).[18]

The partial success of this model can be explained by a number of factors. First **IAdist** indirectly carries information about prosodic structure, which was more finely encoded in model (3). Second, model (6) permits the absolute peak delay to be greater when the number of syllables between the accent locations is increased. Third, it is able to incorporate the gradient stress clash effect observed in RWS's data.

The results of model (6) do not allow us to conclusively reject gestural overlap and tonal repulsion as factors in explaining peak alignment. Nevertheless given the significantly lower overall success of model (6), as compared to model (3), it seems unlikely that they provide a complete explanation. Model (3) has less parameters, and at the same time is more successful in predicting the observed data, and so in principle we prefer it.

5.5.2.3 EXTRA BEATS

The reader may be tempted to apply the framework of Selkirk (1984) to explain the data we have described. This represents, in our opinion, the most plausible candidate for an explanation of how the contextual effects might be phonologically mediated. Within this framework, phonological phrase boundaries are encoded in the metrical grid using silent timing units; the number of timing units introduced depends on the strength of the boundary involved, and on whether a stress clash needs to be resolved. Units (or at least, a sufficiently large string of such units) may be realized as a pause; typically, however, they borrow segmental material from the phrase-final syllable, and are accordingly produced using syllable lengthening. The effects of prosody on peak proportions might be taken to follow from this description, if we assume that the pitch accent stays with the grid alignment corresponding to its original metrical association, and does not propagate to following silent beats.

We do not believe that this explanation is correct. In general, any effects that might be described by such phonological mediation could be equally well described by permitting the phonetic implementation to have direct access to a

hierarchical phonological structure. Encoding the prosodic differences into an intermediate representation with silent beats does not appear to advance the explanation. More specifically, the very same pattern that we have described for RWS, showing a gradient effect on tonal alignment that extends over a number of syllables, requires that extra beats be inserted elsewhere than just at the boundaries. The absence of any related increase in durations presents extra difficulties for such an account. Furthermore, the beats must somehow be increased or decreased in quantity according to the proximity of the upcoming accent, since RWS's peak alignments definitely showed this pattern. The rules for parcelling out the extra beats look quantitative, rather than qualitative. We conclude that the computation of extra beats unnecessarily complicates the explanation of the data: the tonal alignment rules can just as easily refer directly to the underlying metrical structure.

5.5.2.4 SONORITY PROFILE

In the introduction, our fifth conjecture related peak alignment differences to differences in the sonority profile of the syllable. This hypothesis is suggested by the well-known observation that syllable rhymes are more affected by prosodic lengthening than syllable onsets. Figures 5.10a and 5.10b refined this observation by showing a differential effect within the rhyme of *Mom*.

In order for coupling to the sonority profile to explain our data, it must shift the peak left-ward when the end of the syllable is lengthened by a factor on the right. Figure 5.12 illustrates a computation which has this result. The sonority profile of the syllable is schematized as an upslope and a downslope, and the proportional peak placement is computed from the ratio of the two. The top half shows the non-lengthened case, where the F_0 rise reaches its peak at the end of the syllable's duration. (Actually, in our data for both speakers the peak was even past the end of the syllable in these cases.) The bottom half of the figure shows the prosodic lengthening as having applied more to the downslope than to the upslope, and the F_0 peak occurring proportionately as well as absolutely earlier. This coupling provides a single mechanism for peak placement in prenuclear and nuclear position: in both cases peaks occur earlier when a lengthening trigger occurs to the right. This mechanism accounts at the same time for the two different effects of right-hand prosodic context that we described in section 5.4.3: the tendency for peaks to move to the left before a lengthening trigger arises because that trigger alters the sonority profile in such a way that the proportional placement of the peak within the syllable is reduced.

Of course the bit of algebra in figure 5.12 is not especially plausible as a model of speech production. Much further work would be necessary to formulate this model in a way that is rooted in articulatory coordination, and to evaluate it rigorously.

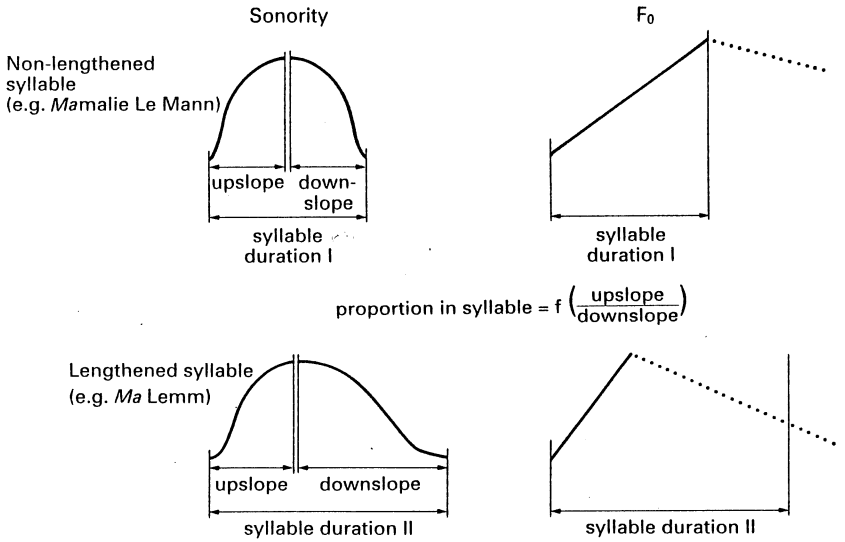


Figure 5.12 Stylized sonority profiles and corresponding F_0 trajectories for an H* accent realized on nonlengthened (upper half) and lengthened (lower half) instances of the syllable /ma/.

5.6 Conclusion

The original notions of nuclear accents being different and distinct from prenuclear accents arose partly on descriptive grounds and partly from considerations of intonational function. By allowing a phonology of intonation and a set of phonetic implementation rules, we have an intermediate level of prosodic representation that intervenes between form and function, and at the same time enables simpler and more comprehensive models of surface phonetic realizations. Our data support parallel phonological and phonetic treatment of nuclear and prenuclear accents, and give further insight into the contextual factors affecting peak alignment in both positions. As we discussed in section 5.1, a remaining difference between the positions (i.e. that H* peaks are aligned earlier if the accents are in nuclear position) might be explained in terms of greater lengthening on nuclear syllables. Alternatively, one might wish to argue that when H* accents occur in nuclear position they are repelled to the left by the closeness of the immediately following L tone. But either way, the question of whether nuclear pitch accents are distinct from prenuclear accents becomes superseded in the light of the current framework. We believe that the question becomes the more general one of what information about an utterance's abstract phonological structure, on which autosegmental tiers, is accessed by the tonal implementation rules.

Further research will be needed to determine the mechanism by which the alignment patterns arise. Simple models based on conjectures that the F_0 rise is invariant or that phonological rules insert extra beats are not supported by the data. There is some evidence for overlap of gestures and/or tonal repulsion, but these alone are inadequate to fully explain the observed alignment patterns. Coupling of the F_0 trajectory to the syllable's sonority profile also seems to be involved.

We have sampled and tested a number of approaches to describing how words and melody might be coordinated, ranging from articulatory through to more abstract phonological levels of explanation. For each alternative we have attempted to be explicit about how it could generate the surface phonetic structure, and we have tried to derive corresponding quantitative predictions. To the extent that we have succeeded, we have been able to bring experimental methodologies to bear on the interplay of phonological representation and phonetic implementation.

Notes

1 See Pierrehumbert and Beckman (1988) for a more complete exposition of autosegmental association in relation to prosodic structure.

2 Here and subsequently in this chapter we shall refer to Steele's (1986) oral presentation of the data. For a full description and discussion, the reader is referred to Steele and Silverman (in preparation).

3 A nuclear accent is normally defined as the last, and typically most salient, pitch accent in an intonational phrase.

4 Lexical stress on the polysyllabic names was as follows: *MAma*; *MAmalie* (similar to the name *ROsalie*); *Le MANN*; and *LEmonick*.

5 For expositional purposes, we will say that a stressed syllable is longer when it clashes with an upcoming stress. Obviously, it would be equally possible to say that it is shorter when it is separated by unstressed material from any upcoming stress. Either way of putting it is tantamount to positing a foot isochrony effect, provided that each foot includes a stressed syllable and all unstressed syllables up to the next stress, without regard to the presence of word boundaries.

6 Cohen and Cohen (1975) is a useful reference for multiple regression, which we relied on extensively.

7 Mathematically, R^2 is simply the squared coefficient of the correlation between the predicted and observed values of the dependent variable.

8 Analyses in the form of equation (3) in which the dependent variable was **peak delay** rather than **peak proportion** yielded R^2 values of 44·5% for JBP and 31·5% for RWS, both being significantly poorer fits than those given in table 5.4.

9 Plots similar to those in figures 5.3 to 5.6 indicated that the data for *Mom* most resembled the rest of the corpus when the final /m/ was included. Some small differences still remained – particularly for RWS – which are partly responsible for the variance not explained by the model developed below.

10 Significance tests on the semipartial R^2s for the set of interaction terms yielded for JBP $F_{(4,171)} = 1{\cdot}9867$, $p = 0{\cdot}099$; and for RWS $F_{(4,166)} = 2{\cdot}4465$, $p = 0{\cdot}048$.

11 These differences were measured with t-tests for unrelated samples, rather than with paired comparisons in an analysis of variance, because the latter method would have pooled the error terms across all cells. This pooling would not have been justified since the peak delay data did not exhibit homogeneity of variance: the greater spread at the slow rate would have swamped out the differences in the other rates. The results for the fast, normal and slow rates, respectively, were: for RWS $t_{13} = -4{\cdot}673$, $p < 0{\cdot}001$; $t_{13} = -2{\cdot}021$, $p < 0{\cdot}05$; $t_{13} = -1{\cdot}989$, $p < 0{\cdot}05$, and for JBP $t_{13} = -1{\cdot}976$, $p < 0{\cdot}05$; $t_{13} = -0{\cdot}677$, $p = 0{\cdot}3$; $t_{13} = 2{\cdot}017$, $p < 0{\cdot}05$.

12 It is well established that prosodic lengthening affects syllable rhymes much more than syllable onsets. This can be seen in the results of, for example, Klatt (1976) and Nakatani and Schaffer (1978).

13 The reader may note that the signs of the coefficients in this table are the opposite of those in equation (3) and table 5.6. This is as it should be, because in this case the figures refer to changes in duration, rather than changes in syllable-relative peak-placement.

14 Significance tests for the amount of variance explained by the set of interaction terms yielded for JBP $F_{(4,171)} = 47{\cdot}8786$, $p < 0{\cdot}0001$; and for RWS $F_{(4,166)} = 16{\cdot}7692$, $p < 0{\cdot}0001$.

15 Replacing **sc** by **gradsc** also increased how well the model fitted JBP's data, although the effect was not large enough, relative to the noise, to reach statistical significance. A test of the difference between the two dependent semipartial correlation coefficients on the peak proportions after partialling out the variance due to speech rate and word boundaries yielded for RWS: $t_{172} = 6{\cdot}771$, $p < 0{\cdot}0001$; and for JBP: $t_{177} = 1{\cdot}234$, $p = 0{\cdot}110$.

16 Note, however, that this of itself does not do away with the distinction between nuclear and prenuclear as prosodic categories. We believe that the distinction exists in prosodic organization, but do not find evidence for it in the inventory of English pitch accents or the phonetic rules for pronouncing them.

17 By measuring from the prenuclear vowel onset, rather than from the initial consonant, we were able to exclude from the data the extraneous noise due to variability in the utterance-initial /m/, and yet maintain a measurement point that was temporally associated with the onset of the prenuclear accent gesture. We used the /l/-onset for the nuclear syllable because the vowel onset was not available.

18 The bivariate correlations between the observed values and those predicted by the two models were adjusted according to the sample size and the number of independent variables in each model, according to Cohen's *Shrunken* R^2, and then compared as independent product-moment correlation coefficients using Fisher's z' transformation. Note that the r's were in this case not completely independent, because peak delays are correlated with peak proportions. This makes the test conservative: it is likely to underestimate the difference between the two models.

References

Browman, C. P. and L. Goldstein. (this volume). Tiers in articulatory phonology, with some implications for casual speech.

Bruce, G. 1977. *Swedish Word Accents in Sentence Perspective* (*Travaux de l'Institut de Linguistique de Lund*). Lund: CWK Gleerup.

Cohen, J. and P. Cohen. 1975. *Applied Multiple Regression Correlation Analysis for the Behavioral Sciences*. Hillsdale NJ: Lawrence Erlbaum.

't Hart, J. and R. Collier, 1975. Integrating different levels of intonation analysis. *Journal of Phonetics* 3: 235–255.

Klatt, D. H. 1976. Linguistic uses of segmental duration in English: acoustic and perceptual evidence. *Journal of the Acoustical Society of America* 59: 1208–1221.
Mattingly, I. G. 1966. Synthesis by rule of prosodic features. *Language and Speech* 9: 1–13.
Nakatani, L. H. and J. Schaffer. 1978. Hearing words without words: prosodic cues for word perception. *Journal of the Acoustical Society of America* 63: 234–244.
O'Connor, J. D. and G. R. Arnold. 1973. *Intonation of Colloquial English*. 2nd edition. London: Longman.
Pierrehumbert, J. B. 1980. The phonology and phonetics of English intonation. Ph.D. dissertation, MIT.
1981. Synthesizing intonation. *Journal of Acoustical Society of America* 70: 985–995.
Pierrehumbert, J. B. and M. E. Beckman. 1988. *Japanese Tone Structure*. LI Monograph Series, Cambridge, MA: MIT Press.
de Pijper, J. R. 1983. *Modelling British English Intonation*. Dordrecht: Foris.
Selkirk, E. O. 1984. *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge, MA: MIT Press.
Silverman, K. E. A. 1987. The structure and processing of fundamental frequency contours. Ph.D. dissertation, Cambridge University.
Steele, S. A. 1986. Nuclear accent F_0 peak location: effects of rate, vowel, and number of following syllables. *Journal of the Acoustical Society, Supplement 1*, 80; s51.
Steele, S. A. and K. E. A. Silverman. (in preparation). Alignment of nuclear pitch accents: measurements and a model. MS.

6

Alignment and composition of tonal accents: comments on Silverman and Pierrehumbert's paper

GÖSTA BRUCE

6.1 Introduction

The paper by Kim Silverman and Janet Pierrehumbert specifically addresses the timing of immediately prenuclear high accents in English with respect to the following nuclear accent. A more general topic of the paper is the coordination between rhythmical structure and tonal (or accentual) structure in human, spoken language. Another general issue coupled to these topics and approached in the paper is the division of labor between phonology and phonetics.

I will begin my discussion by listing a number of factors that may determine the timing of tonal peaks (or other points in the tonal structure) relative to segmental references as part of the rhythmical make-up of an utterance. The list is not meant to be exhaustive but is intended to show that there are a fair number of factors involved, some of which we know affect accent timing and others that are likely to do so, although experimental evidence is still lacking. There is also probably some overlapping among certain categories in my taxonomy.

1. Tonal composition (phonological analysis of pitch accents – whether analyzed as mono– or bitonal, linked or unlinked tones, targets or gestures – can influence the results of a phonetic analysis).
2. Prosodic context
   a. Boundaries (word, phrase, utterance, etc.)
   b. Rhythmical organization (rhythmical grouping, e.g. stress clash)
   c. Focus (prefocal, focal, postfocal position)
   d. Tonal environment (tonal interaction within and between successive pitch accents, e.g. tonal crowding)
   e. Pitch range (local or global, e.g. differences in degree of overall emphasis due to degree of involvement)
   f. Global intonation (e.g. absence/presence of downdrift due to interrogative /declarative structure)
3. Segmental context (e.g. differences in intrinsic vowel length)
4. Speaking rate (fast, normal, slow tempo)