

The effect of pitch span on intonational plateaux

Rachael-Anne Knight

City University London

knight@city.ac.uk

Francis Nolan

University of Cambridge

fjn1@cam.ac.uk

Previous research has indicated that the H (high) of a nuclear accent may be realized as a flat stretch of contour rather than as a single turning point. Both the duration of this plateau and its alignment within the accented syllable are affected by the segmental and prosodic structure of the utterance. The present work investigates whether a non-structural variable, namely pitch span, also affects the realization of the plateau. Speakers replicated all-sonorant utterances in different pitch spans. Results show that both the duration and alignment of the plateau vary with pitch span but in ways different from the way they vary with prosodic structure. Importantly, results also indicate that, when using a proportional measure of alignment, the end of the plateau is anchored within the syllable for each speaker and may be a marker of linguistic structure.

1 Introduction

Autosegmental Metrical theories of intonation might lead us to expect pitch targets to be realized as single turning points in the fundamental frequency (F0) contour. However, several studies (e.g. J. House, Dankovičová & Huckvale 1999, Wichmann, House & Rietveld 1999) have shown that, instead, H targets often appear as flat stretches of contour, or plateaux. These plateaux have the effect of realizing the ‘nuclear fall’ as a step down from a high stressed syllable to a following low unstressed syllable in an English polysyllabic word, as can be seen in figure 1 and figure 2 in section 2 below. This suggests a rather different strategy for the use of pitch in English than in some other languages. For instance, in connection with Swedish word accents, D. House (e.g. 1990) has shown that HL (high to low) pitch movements are only perceived as such in the spectrally steady-state portion of the vowel, whereas pitch movements that occur during rapidly changing spectral configurations are perceived as pitch levels. In the examples below, however, it seems that English is not making use of this strategy. The stressed vowel is associated with the relatively level pitch of the plateau whilst the fall itself occurs during the sonorant /l/ and the following unstressed vowel.

The plateaux that have been observed in nuclear position are affected by both the segmental structure of the syllable and the prosodic structure of the utterance. J. House et al. (1999) examine a medium-sized single-speaker database. They show that plateaux (defined as consisting of that section of the F0 contour that falls within 4% of the absolute maximum F0) are proportionally longer, and begin earlier inside the syllable and foot when the onset of the accented syllable is sonorant.

In terms of prosodic structure, House et al. (1999) find that the plateau is aligned later in the syllable when the foot contains more syllables, and Knight (2004) demonstrates that there is a significant difference between alignment in mono- and polysyllabic feet. These findings

fit well with the large body of literature on F0 peak alignment. Several studies suggest that peak alignment is affected by the lengthening of prosodic structures such as the syllable or the foot, and it seems that different causes of lengthening have different effects on peak alignment. For example, studies by Steele (1986) and Silverman & Pierrehumbert (1990), dealing with nuclear and prenuclear accents respectively, suggest that when a structural unit is lengthened by prosodic context, such as an upcoming word boundary or a stress clash, the peak is aligned earlier within that unit.

Peak alignment is also affected by lengthening induced by non-structural factors such as intrinsically long vowels. For example, Steele (1986) and Silverman & Pierrehumbert (1990) note that peaks are aligned later in a syllable that is lengthened due to a slower tempo. Other work by Ladd & Morton (1997) suggests that, similarly, peaks are aligned later in the syllable when that syllable is lengthened in an expanded pitch span (defined by Ladd 1996 as the difference between high and low targets in the speaker's range). The present study investigates how pitch span affects the alignment and the duration of intonational plateaux. It is hypothesized that plateaux will be shorter in wider pitch spans for physiological reasons. In expanded pitch spans the speaker must reach more extreme values in pitch and it is likely that it will take longer to reach these more extreme values (Xu 2002: 92). It is possible, therefore, that there will be less time available to remain at the high level and realize a plateau.

In terms of alignment, two alternative hypotheses are suggested. Firstly, it is possible that the whole plateau will be aligned later in the syllable in more expanded spans. As discussed above, peaks are aligned later when syllables are lengthened by non-structural factors such as increased pitch span. Because the plateau is defined in relation to the peak it is hypothesized that the plateau will also be aligned later in syllables lengthened by an increased pitch span.

Alternatively, given the above hypothesis concerning duration, the plateau may contract around the peak in order to allow the speaker more time in which to reach the more extreme frequencies characteristic of wider spans. This may result in the later alignment of the plateau's beginning, and the earlier alignment of its end.

2 Method

2.1 Stimulus materials

This experiment utilized recordings made for a separate experiment on pitch equivalence in intonation described in Nolan (2002). A replication task was used to ensure subjects produced utterances in different pitch spans.

Two utterances were used as stimuli, both composed of entirely sonorant material in order to minimize microprosodic effects. The utterances are (with autosegmental transcription):

- A. We were relying on a milliner
H* + L L-L%
- B. A milliner
H* L-H%

Two phoneticians, one male (the second author) and one female (a volunteer), recorded template utterances in a sound-treated booth. The male speaker would produce a token of utterance A in one of three impressionistically defined pitch spans, hereafter referred to as neutral, compressed (narrower than the neutral span) and expanded (wider than the neutral span), and the female speaker would attempt an exact replication of the token in her own tessitura. The male speaker would then produce a token of utterance B in the same pitch span as utterance A and the female speaker would again replicate it. This procedure was repeated twice more for each pitch span. The second author compared each replication pair, and the auditorily most accurate pair in each pitch span was chosen as the template.

Table 1 Measurements of template utterances of the male and female phonetician. C refers to compressed, N to neutral and E to expanded pitch span.

| Measure | Male | | | Female | | |
|------------------------|------|------|------|--------|------|------|
| | C | N | E | C | N | E |
| Frequency of peak (Hz) | 104 | 133 | 192 | 222 | 263 | 363 |
| Frequency of L (Hz) | 77 | 77 | 70 | 166 | 168 | 148 |
| Span (H/L) | 1.35 | 1.73 | 2.74 | 1.34 | 1.57 | 2.45 |
| Syllable duration (ms) | 205 | 230 | 265 | 245 | 215 | 270 |

Several measurements were made of each template utterance to ensure that the speakers were in fact using different pitch spans and to aid subsequent interpretation of the results proper. Means for each variable are shown in table 1. It can be seen that peaks (the highest F0 in the nucleus) get higher and the proportional relationship between the peak and L (the elbow in the contour, as described in section 2.4.2) increases when the wider spans are replicated. Ls are lower in the expanded span, and also in the compressed span for the female speaker. In general, the duration of the syllable increases in wider spans. However, measurements for the neutral span for the female speaker are shorter than would be expected from this generalization.

Three tokens of each template utterance were used in the experiment, giving 36 tokens (3 tokens \times 3 pitch spans \times 2 utterances \times 2 speakers) in total. These tokens were combined to form 18 conversational dyads. Within each dyad the tokens were randomized, with the restrictions that utterance A always preceded utterance B, and that the speaker and the pitch span were different for each token. The order in which the pairs were presented was also randomized.

2.2 Subjects

Twelve native speakers of British English participated in the experiment. There were 8 females and 4 males whose ages ranged from 21 years to 29 years. Speakers 3, 5, 6 and 9 were male, and in the tables below their speaker identification numbers are shown in *italics*. All were students at the University of Cambridge with some training in phonetics and intonation.

2.3 Procedure

Subjects were recorded individually in a sound-treated booth. Stimuli were presented through headphones, and after each utterance (A or B) had been played, an on-screen message prompted subjects to repeat it exactly, aiming to produce an intonationally equivalent utterance in their own voice. Their speech was recorded directly onto the hard-drive of an SG workstation via a high quality microphone. An investigator was present as each utterance was recorded. The investigator monitored the quality of the signal auditorily and by examining a waveform immediately after each recording. Recordings that were unsatisfactory, due to noise or to the speaker starting too late and therefore the signal being too long for the recording window, were repeated at the time of recording.

2.4 Method of analysis

2.4.1 The definition of the plateau

Firstly, it is important to decide how the plateau should be defined, as this will affect the results presented later. Initially it might seem that the work of Rossi (1971) is of relevance to this issue, as Rossi investigated the perceptual threshold for tonal variations in speech. Rossi's

aim was to find out, first, whether or not a frequency variation in a given time is perceptible and, second, how a glide affects the perceived pitch of a vowel. The experiment that is most relevant to the current investigation involved subjects comparing a static tone and a rising tone and saying whether the second tone was higher, lower or of equal pitch to the first. The results showed that even when the rise was perceptible it was not perceived in its entirety, as the perceived pitch corresponded to the pitch of the glide situated between two thirds and the end of the vowel. Although Rossi's work suggests that the pitch of glides is not perceived in entirety, and that the perceived pitch is close to the end of the glide, the results are not directly applicable to the present work. The reason for this is that Rossi focuses exclusively on rising tones, while the plateaux in question in the present work are related to falling tones. Rossi (1971: 1) states that he did not examine falling tones because their perception poses a more complex problem than the perception of rises.

Further strands of evidence that bear on the definition of the plateau are those used by House et al. (1999). House et al. (1999), citing Rosen & Fourcin (1986), state that the 4% measure used in their study approximates the range of perceptual equality for speech sounds, so listeners should be expected to hear anything in this range as being equal in pitch. The studies cited by Rosen & Fourcin (1986: 398f.), however, show that the situation is actually more complicated. Three main studies are cited, which suggest the size of the smallest differences in fundamental frequency that can be detected in speech. In the first study Klatt (1973) performed several experiments where subjects compared two stimuli in order to determine the discriminability of pitch changes. Stimuli consisted of unchanging synthetic vowels with flat F0, and also rising and falling F0 around 120 Hz. The just noticeable differences (jnds) ranged from 0.3 Hz when the F0 was flat to around 2 Hz when the F0 was falling by 30 Hz over 250 ms. In the second study Pierrehumbert (1979) used stimuli consisting of reiterant speech containing two stressed syllables in which the second syllable varied in maximum frequency. Subjects were asked to judge which of the two stressed syllables was higher in pitch. The jnd (the difference in F0 between subjects guessing which peak was higher and being able to tell reliably) for first-peak values of 121 and 151 Hz was around 12 Hz (10% and 8%, respectively). In the final study 't Hart (1981) presented pairs of four-syllable number names to Dutch listeners. In each item, a rise or fall over between one and six semitones took place on the accented syllable. Listeners were asked to judge which of the two items in the pair contained the larger pitch movement. Results indicate that on average the jnds were not smaller than 1 semitone or 6% (7 Hz for fundamentals of 120 Hz).

't Hart's (1981) study is probably the most useful for deciding the range of perceptual equality for the plateau. It uses the most natural speech of the studies cited above and examines both rises and falls over a number of different ranges. 't Hart's results suggest that listeners will hear everything within around 6% of the peak as being equal in pitch. House et al.'s (1999) 4% range is, therefore, a conservative estimate of the range of perceptual equality.

The question then concerns whether to use 6% (as suggested by 't Hart 1981) or 4% (to maintain consistency with House et al. (1999)). A comparison of some F0 contours makes the task easier. Figure 1 and figure 2 show two pitch contours spoken in a neutral pitch span taken from the experiment discussed below. In each, the position of plateaux defined as 2%, 4% and 6% of the maximum frequency are marked for comparison. In fact there is not a great deal of difference between plateaux identified according to these different measures. For the first utterance ('a milliner' taken from utterance A, 'We were relying on a milliner') the difference between 2% and 6% for the start of the plateaux is 9 Hz and 6 ms, whilst the difference for the end of the plateau is 10 Hz and 27 ms. For the second utterance (utterance B, 'A milliner') the difference between the 2% and 6% definitions is 10 Hz and 54 ms for the start of the plateau and 10 Hz and 22 ms for the end of the plateau.

In the event it was decided to use 4% as the definition for the plateau. First, because of the rate of change of the slopes in the contour, there is often no difference between plateaux defined using the different percentage measures. Second, when the different percentages DO give different results, a visual examination of contours, including the ones shown below,

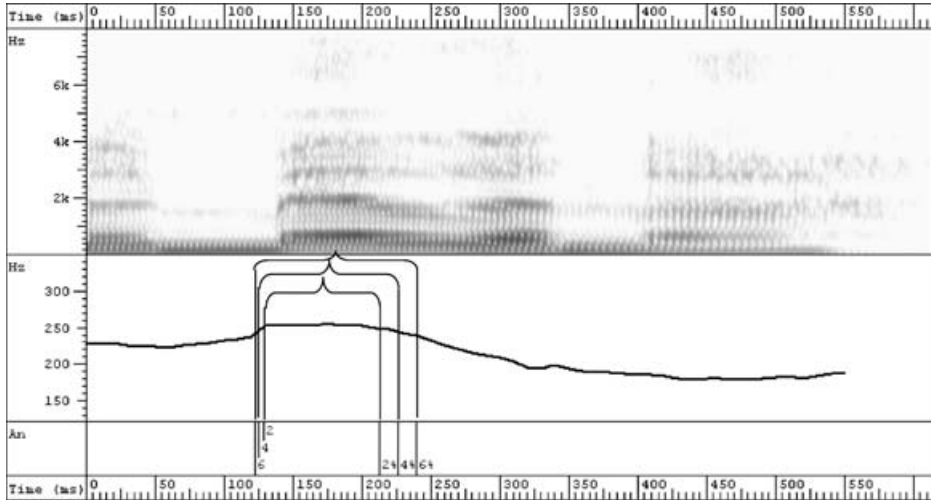


Figure 1 'A milliner' taken from the utterance 'We were relying on a milliner' spoken by a female in neutral pitch span. Nuclear plateaux, defined as 2, 4 and 6% of the maximum F0, are shown.

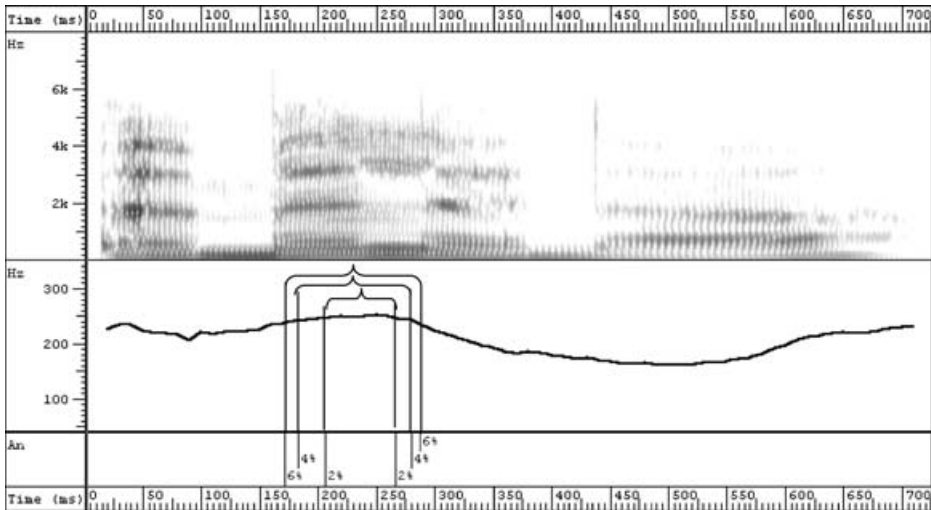


Figure 2 The utterance 'A milliner' spoken by a female in neutral pitch span. Nuclear plateaux, defined as 2, 4 and 6% of the maximum F0, are shown.

suggests that 4% usually gives the most reasonable estimate of the 'corners' of the plateau whilst 2% is too small and 6% too large. In addition this 4% measure allows for direct comparison with the results of House et al. (1999).

2.4.2 Measurements taken from replicated utterances

The measurements taken, and the relationships between these measurements, are shown in figure 3. After first identifying the maximum F0 within the utterance, the start and end points of the plateau were located using the 4% criterion described above. The elbow in the contour representing the following low tone was also located for comparison with the high reference

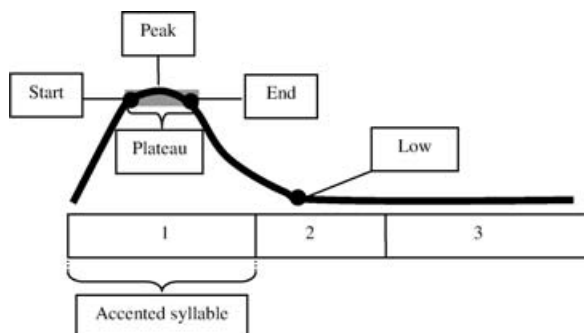


Figure 3 Schematic representation of measurements taken.

points. The low tone was generally easy to locate as it occurred towards the end of the utterance where the following contour was flat. For this reason the low tone was identified by eye. In order to check the reliability of this identification the first author checked 15% of the recorded utterances from each speaker without being able to see the original label for the low tone. In no case did the original measurement and the blind check produce results that were more than 15 ms or 10 Hz apart. The duration of the accented syllable (/mrl/) was measured and used in the calculation of alignment.

2.4.3 Statistical analysis

For most variables two types of statistical analysis were used. The first type used a repeated measures design (MANOVA). In cases where Mauchly's test indicated sphericity could be not assumed a Greenhouse-Geisser correction was used. Independent factors were: utterance (2), sex of speaker whose utterance was replicated (2), and pitch span (3). Only the results for pitch span are reported here, but results for utterance type and sex of speaker replicated are reported in detail in Knight (2003, chapter 3).¹ Whilst the MANOVA analysis treats pitch span as a nominal variable in relation to the three pitch spans defined impressionistically in the recordings, it can also be argued that, in reality, pitch span is a continuous parametric variable. When viewed in this way pitch span (defined as the frequency of H divided by the frequency of L) can be used in a bivariate correlation analysis with each of the other factors under consideration. Therefore, where appropriate, both analyses are given and compared.

3 Results

3.1 Pitch height and span

Three tests were conducted to see if subjects were in fact using different pitch spans in their replications. These tests examined the frequency of the peak and the low tone and the relationship between them. The tests were conducted separately for male and female speakers to allow for natural differences in pitch range and span. The results for females are shown in figure 4, and for males in figure 5. (The results are shown in more detail in the appendix.) As

¹ In summary, in utterance B the peak was higher and the plateau was shorter. SP (start of plateau) was timed and aligned later in utterance B but the peak and EP (end of plateau) were both timed and aligned stably, and L was timed and aligned earlier in utterance B than A. The sex of the speaker replicated did not affect the timing of any timing point and only affected the alignment of EP, which was earlier when subjects replicated the female speaker. Speakers used wider spans when replicating the female voice.

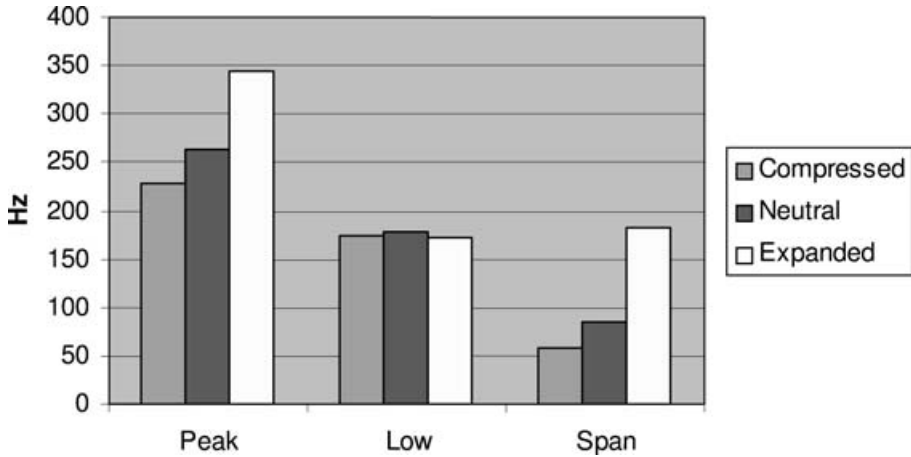


Figure 4 The frequency of the peak and low tone and the span (difference between peak and low) when replicating different pitch spans, for female subjects.

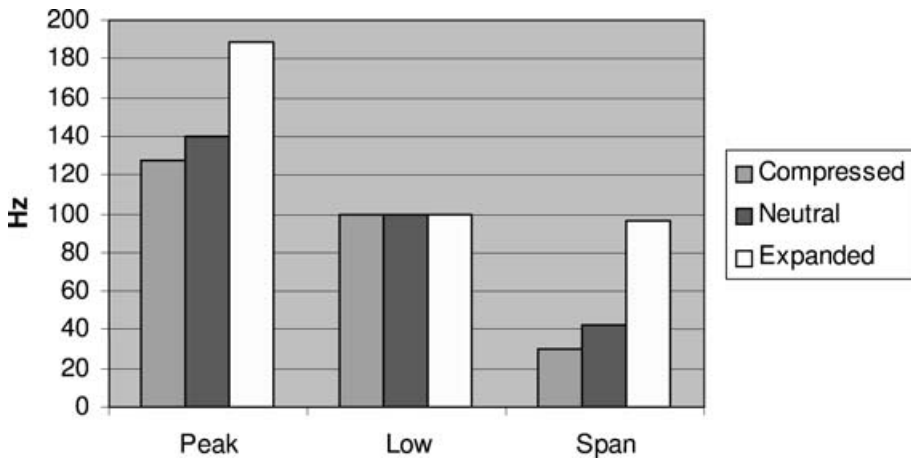


Figure 5 The frequency of the peak and low tone and the span (difference between peak and low) when replicating different pitch spans, for male subjects.

expected, the maximum frequency (the peak) was significantly higher in the more expanded pitch spans for both women ($F(2,14) = 314.8$, $p < 0.01$) and men ($F(2,6) = 45.1$, $p < 0.01$). Planned comparisons indicate that the frequency of the peak is lower in the compressed span than the neutral span for both women ($t(7) = 15.5$, $p < 0.01$) and men ($t(3) = 3.7$, $p < 0.05$) and lower in the neutral span than the expanded span for both women ($t(7) = 17.6$, $p < 0.01$) and men ($t(3) = 7.7$, $p < 0.01$).

Span, as stated above, is defined as the frequency of H divided by the frequency of L, and this proportional measure allows for men's and women's results to be averaged together. Results show that the span is significantly affected by the impressionistically defined pitch span being replicated ($F(2,22) = 1.030$, $p < 0.001$) and that all pitch spans are significantly different from each other.

A further measure of span was used, which was calculated by subtracting the frequency of L from the frequency of H. When calculated in this way the span increases when speakers

Table 2 Correlation between span (H/L) and duration of accented syllable for each speaker separately. Identification numbers for male speakers are shown in italics. R is Pearson's correlation coefficient and P is the significance level.

| Speaker | Duration of accented syllable | |
|----------|-------------------------------|-------|
| | R | P < |
| 1 | 0.010 | 0.954 |
| 2 | 0.492 | 0.002 |
| <i>3</i> | 0.738 | 0.001 |
| 4 | 0.645 | 0.001 |
| <i>5</i> | 0.699 | 0.001 |
| <i>6</i> | 0.451 | 0.006 |
| 7 | 0.715 | 0.001 |
| 8 | 0.595 | 0.001 |
| <i>9</i> | 0.620 | 0.001 |
| 10 | 0.565 | 0.001 |
| 11 | 0.580 | 0.001 |
| 12 | 0.468 | 0.004 |

replicate wider spans for both women ($F(2,14) = 278.4$, $p < 0.01$) and men ($F(2,6) = 34.6$, $p < 0.01$). The difference between H and L is smaller in the compressed span than the neutral span for both women ($t(7) = 9.9$, $p < 0.01$) and men ($t(3) = 3.5$, $p < 0.05$), and smaller in the neutral span than the expanded span for both women ($t(7) = 10.8$, $p < 0.01$) and men ($t(3) = 8.6$, $p < 0.01$). Results calculated in this way mirror those when span is calculated as H divided by L. In figure 4 and figure 5 this measure of span (H-L) is shown. This is because the proportional measure (H/L) gives values that are too small to be seen when plotted on the same axis as the results for the frequency of the peak and L.

For men, the low tone is not significantly affected by the pitch span used ($F(2,6) = 0.3$, $p > 0.05$) but for women the low tone *is* significantly affected by pitch span ($F(2,14) = 9.3$, $p < 0.01$). For women, planned comparisons reveal that whilst there is no difference between the frequency of L in the compressed and expanded spans ($t(7) = 1.8$, $p > 0.05$), L is higher in the neutral span than in either the compressed ($t(7) = 2.6$, $p < 0.05$) or expanded ($t(7) = 4.8$, $p < 0.01$) span. It is not clear why women and men should perform differently with regards to the frequency of L in different pitch spans. However, there is an interesting comparison with the findings of Ladd et al. (1999) for modification of F0 under changes in speech rate. Ladd et al. (1999: 1554) found that, overall, L was higher at a normal speaking rate than at either a fast or a slow rate and, therefore, it may be that deviations from the normal or neutral setting involve a lowering of L.

Correlations are not calculated for these initial tests as the results would only reflect the fact that the F0 of the peak and L are used in the calculation of span.

3.2 Duration of the accented syllable

As found by Ladd & Morton (1997), the duration of the accented syllable is significantly affected by pitch span ($F(1.23, 15.55) = 8.2$, $p < 0.01$) and this is confirmed for all but one speaker in the correlation analysis shown in table 2. As shown in figure 6 there is only a small difference between the duration of the syllable in the compressed span compared to the neutral span, which is not significant ($t(11) = 0.82$, $p > 0.05$). The difference between the durations in the neutral and expanded spans is significant, however, as the syllable is longer in the expanded than the neutral pitch span ($t(11) = 8.2$, $p < 0.01$). These figures are given in more detail in the appendix.

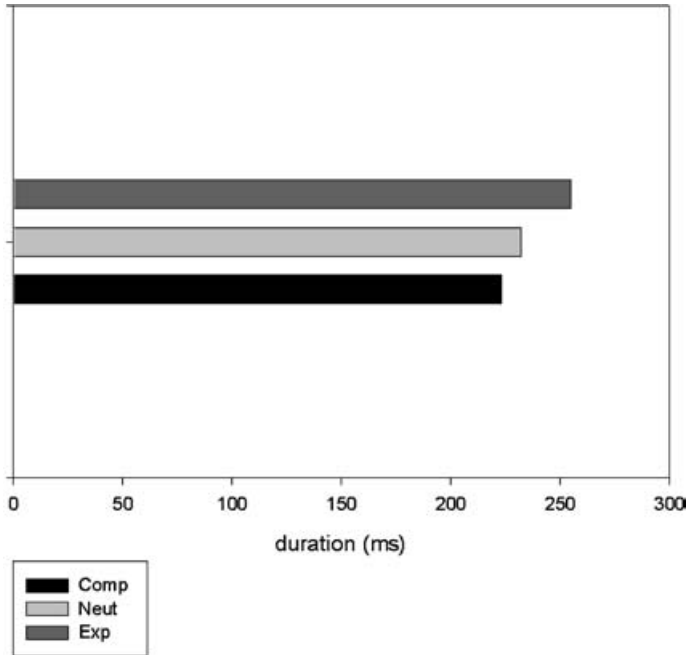


Figure 6 Syllable duration when replicating different spans, pooled over subjects.

To some extent these differences in syllable duration were present in the templates that the speakers replicated as can be seen from table 1. However, even though the female speaker had a shorter syllable in the neutral than in the other spans, subjects did not replicate this, suggesting that there may be an automatic association between expanded pitch span and longer syllable duration for most speakers.

3.2.1 Duration of the plateau

As hypothesized, the absolute duration of the plateau is significantly affected by the pitch span used ($F(1.328, 14.606) = 79.7, p < 0.01$) and this is confirmed for all but one speaker in the correlation analysis shown in table 3. The plateau is significantly longer in the compressed span than the neutral span ($t(11) = 7.6, p < 0.01$) and significantly longer in the neutral span than the expanded span ($t(11) = 6.0, p < 0.01$). It should be noted that this shortening in wider spans is not an artefact of the linear representation in Hertz. If the plateau had been defined in terms of a threshold in Hertz, it could be claimed that the shorter plateau measured for higher peaks would not be perceptually realistic, because at higher pitches equivalent pitch intervals represent larger Hertz intervals. However, this non-linearity is accommodated precisely by using the 4% range in the calculation of the plateau rather than a fixed threshold in Hertz.

The results for the duration between each end of the plateau and the peak mirror the results for the entire plateau. The absolute duration between the start of the plateau (SP) and the peak is shorter in an expanded span than a neutral span, and shorter in a neutral than a compressed span ($F(2,22) = 16.39, p < 0.01$). The absolute duration between the peak and the end of the plateau (EP) is shorter in an expanded span than a neutral span, and shorter in a neutral span than a compressed span ($F(2,22) = 31.69, p < 0.01$). The correlation analysis shows that for five speakers the SP-Peak duration is not correlated with pitch span. For four of these speakers, though, the overall plateau duration is correlated with span because of

Table 3 Correlation of the pitch span with the duration of the plateau, the duration from SP to peak, and from peak to EP, for each speaker separately. Identification numbers for male speakers are shown in italics. R is Pearson's correlation coefficient and P is the significance level.

| Speaker | Plateau duration | | SP-peak duration | | Peak-EP duration | |
|---------|------------------|-------|------------------|-------|------------------|-------|
| | R | P< | R | P< | R | P< |
| 1 | -0.534 | 0.001 | -0.240 | 0.159 | -0.511 | 0.001 |
| 2 | -0.583 | 0.001 | -0.178 | 0.298 | -0.523 | 0.001 |
| 3 | -0.648 | 0.001 | -0.501 | 0.002 | -0.330 | 0.049 |
| 4 | -0.626 | 0.001 | -0.225 | 0.188 | -0.547 | 0.001 |
| 5 | -0.512 | 0.001 | -0.450 | 0.006 | -0.383 | 0.021 |
| 6 | -0.571 | 0.001 | -0.189 | 0.269 | -0.583 | 0.001 |
| 7 | -0.217 | 0.204 | -0.050 | 0.774 | -0.153 | 0.372 |
| 8 | -0.488 | 0.003 | -0.356 | 0.033 | -0.388 | 0.019 |
| 9 | -0.600 | 0.001 | -0.561 | 0.001 | -0.398 | 0.016 |
| 10 | -0.725 | 0.001 | -0.399 | 0.016 | -0.518 | 0.001 |
| 11 | -0.543 | 0.001 | -0.353 | 0.035 | -0.364 | 0.029 |
| 12 | -0.627 | 0.001 | -0.527 | 0.001 | -0.354 | 0.034 |

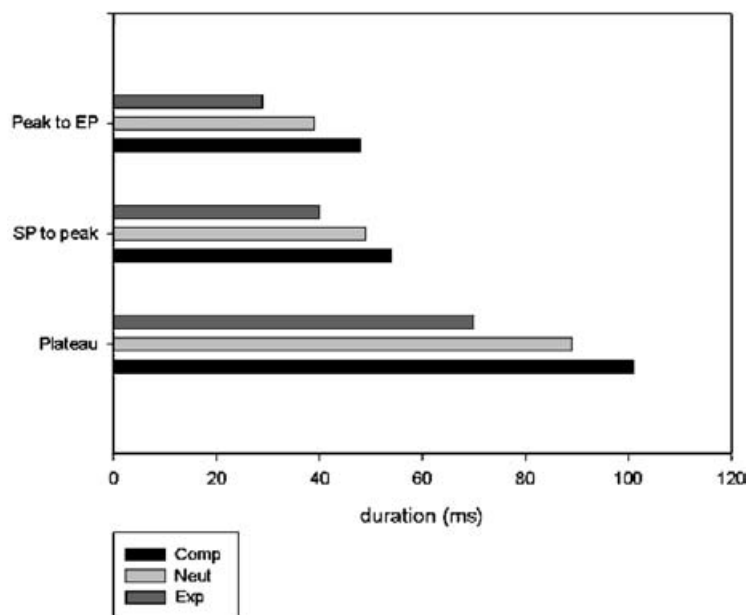


Figure 7 Duration of the whole plateau, from SP to peak, and from peak to EP, pooled across speakers for each pitch span.

the significant correlation between Peak-EP duration and span. For the remaining speaker (speaker 7) there is no correlation between pitch span and any of the measures of plateau duration and plateaux are around 90 ms in all pitch spans. The overall results for the duration of the plateau are shown in figure 7.

Pitch span also significantly affects the proportion of the accented syllable taken up by the plateau ($F(2,22) = 80.4$, $p < 0.01$). Pitch span affects the PROPORTIONAL duration of the

Table 4 Correlation of pitch span and proportional duration of plateau for each speaker separately. Identification numbers for male speakers are shown in italics. R is Pearson's correlation coefficient and P is the significance level.

| Speaker | Proportional duration of plateau | |
|----------|----------------------------------|-------|
| | R | P < |
| 1 | -0.568 | 0.001 |
| 2 | -0.621 | 0.001 |
| <i>3</i> | -0.670 | 0.001 |
| 4 | -0.640 | 0.001 |
| <i>5</i> | -0.579 | 0.001 |
| <i>6</i> | -0.618 | 0.001 |
| 7 | -0.541 | 0.001 |
| 8 | -0.599 | 0.001 |
| <i>9</i> | -0.691 | 0.001 |
| 10 | -0.769 | 0.001 |
| 11 | -0.565 | 0.001 |
| 12 | -0.657 | 0.001 |

plateau in the same way as it affects the ABSOLUTE duration. The plateau takes up more of the syllable in the compressed span than in the neutral span ($t(11) = 4.0$, $p < 0.01$), and it takes up more of the syllable in the neutral span than in the expanded span ($t(11) = 7.4$, $p < 0.01$). The correlation results shown in table 4 support the results of the MANOVA for each speaker.

3.3 Alignment

Alignment is initially calculated in a proportional manner, by measuring the duration between the beginning of the accented syllable and the pitch reference point in question (SP, peak, EP or low tone), and expressing this duration as a percentage of syllable duration.

As was also found by Ladd & Morton (1997), the alignment of the peak is significantly affected by pitch span ($F(2,22) = 0.90$, $p < 0.01$). The peak is earlier in the syllable when the pitch span is compressed or neutral than when it is expanded. There is no difference in alignment between compressed and neutral spans. Although this effect of an earlier peak in narrower spans is found when averaging over all speakers in the repeated measures analysis, the correlation analysis, shown in table 5, shows that the effect is present for only seven speakers. For the other five speakers there is little effect of pitch span on the alignment of the peak.

As expected, the alignment of the beginning of the plateau is also significantly affected by pitch span ($F(2,22) = 44.8$, $p < 0.01$). This is confirmed for 10 speakers by the correlation analysis shown in table 5. There is no significant difference between the alignment of the beginning of the plateau in compressed and neutral spans ($t(11) = 2.5$, $p > 0.05$). However, the alignment of the beginning of the plateau is significantly earlier in neutral than in expanded spans ($t(11) = 7.0$, $p < 0.01$).

The most interesting and most important finding is that the alignment of the end of the plateau is NOT significantly affected by pitch span ($F(2,22) = 0.6$, $p > 0.05$) and the correlation analysis in table 5 shows this to be true for 11 speakers. Thus, the alignment of the end of the plateau is consistently anchored for each of these 11 speakers regardless of the pitch span used. In addition, the alignment of the low tone follows the same pattern as the alignment of the end of the plateau: it is not affected by the pitch span ($F(2,22) = 0.1$, $p > 0.05$). Again, the location is anchored for each speaker as confirmed by the correlation analysis shown in table 5.

Table 5 Correlation between pitch span and alignment of intonational reference points (peak, SP, EP and L), for each speaker separately. Identification numbers for male speakers are shown in italics. R is Pearson's correlation coefficient and P is the significance level.

| Speaker | Peak align | | SP align | | EP align | | L align | |
|---------|------------|-------|----------|-------|----------|-------|---------|-------|
| | R | P< | R | P< | R | P< | R | P< |
| 1 | 0.280 | 0.098 | 0.476 | 0.003 | -0.025 | 0.883 | 0.198 | 0.247 |
| 2 | 0.597 | 0.001 | 0.747 | 0.001 | 0.007 | 0.969 | 0.243 | 0.154 |
| 3 | -0.015 | 0.931 | 0.188 | 0.272 | -0.124 | 0.470 | -0.253 | 0.136 |
| 4 | 0.386 | 0.020 | 0.726 | 0.001 | -0.311 | 0.065 | -0.110 | 0.522 |
| 5 | 0.508 | 0.002 | 0.666 | 0.001 | 0.034 | 0.843 | -0.135 | 0.434 |
| 6 | 0.407 | 0.014 | 0.623 | 0.001 | -0.018 | 0.918 | -0.291 | 0.085 |
| 7 | 0.330 | 0.049 | 0.713 | 0.001 | -0.016 | 0.926 | -0.80 | 0.643 |
| 8 | -0.056 | 0.745 | 0.282 | 0.096 | -0.303 | 0.072 | -0.003 | 0.986 |
| 9 | 0.536 | 0.001 | 0.798 | 0.001 | 0.196 | 0.253 | -0.151 | 0.380 |
| 10 | 0.591 | 0.001 | 0.818 | 0.001 | 0.365 | 0.029 | 0.255 | 0.134 |
| 11 | 0.042 | 0.807 | 0.575 | 0.001 | -0.288 | 0.088 | -0.097 | 0.573 |
| 12 | 0.113 | 0.511 | 0.607 | 0.001 | -0.229 | 0.179 | 0.041 | 0.813 |

3.4 Timing

The above results for alignment, following House et al. (1999), use a proportional measure where the reference point in question is presented as a relative duration into the syllable. However, alignment may also be calculated using an absolute measure of the relationship between intonational reference points and the segmental string. Although this absolute relationship is used by many researchers, it can be argued that it is implausible that speakers align intonational targets in an absolute fashion. Under increased speech rate for example, as syllables shorten, an absolute alignment could lead to targets occurring outside the syllable in question. Indeed Ladd et al. (1999: 1549) show that as speech rate decreases, peaks are located further from the end of the stressed vowel, suggesting that high targets may not be absolutely aligned with segmental landmarks. However, Ladd et al. (1999: 1550) propose that segmental anchoring is more stable if the alignment of the peak is expressed as a PROPORTION of the duration from the offset of the stressed vowel to the onset of the unstressed vowel.

Nevertheless, for reasons of comparability with previous research, the results below use an absolute measure, namely the duration in milliseconds from the beginning of the syllable (/mil/) to each intonation reference point. In the discussions that follow the proportional measure will be referred to as 'alignment' whilst, for clarity, the absolute measure will be referred to as 'timing'.

The timing of the peak (the time in milliseconds between the beginning of the syllable and the peak) is significantly affected by the pitch span used ($F(2,22) = 58.06$, $p < 0.01$). The peak is timed earlier in compressed and neutral spans than in the expanded span, and the correlation results in table 6 show this to be true for 10 speakers. The start of the plateau is closer to the beginning of the syllable when a compressed span or neutral span is used than when an expanded span is used ($F(2,22) = 78.5$, $p < 0.01$), and the correlation analysis shows this is true for all speakers. The end of the plateau is nearer to the beginning of the syllable in the compressed span and in the neutral span than in the expanded span ($F(2,22) = 27.25$, $p < 0.01$), and the correlation results show that this is true for seven speakers. The low tone is timed earlier in compressed and neutral spans than in expanded spans ($F(2,22) = 21.49$, $p < 0.01$), and again the correlation analysis shows this to be the case for seven speakers (although some of these speakers are different from those for whom EP timing is affected by pitch span).

Table 6 Correlation of the pitch span with timing of intonation reference points (peak, SP, EP and L), for each subject separately. Identification numbers for male speakers are shown in italics. R is Pearson's correlation coefficient and P is the significance level.

| Speaker | Peak timing | | SP timing | | EP timing | | L timing | |
|---------|-------------|-------|-----------|-------|-----------|-------|----------|-------|
| | R | P< | R | P< | R | P< | R | P< |
| 1 | 0.733 | 0.001 | 0.721 | 0.001 | 0.625 | 0.001 | 0.609 | 0.001 |
| 2 | 0.682 | 0.001 | 0.743 | 0.001 | 0.414 | 0.012 | 0.683 | 0.001 |
| 3 | 0.196 | 0.251 | 0.336 | 0.045 | 0.135 | 0.432 | 0.267 | 0.115 |
| 4 | 0.742 | 0.001 | 0.771 | 0.001 | 0.373 | 0.025 | 0.403 | 0.015 |
| 5 | 0.743 | 0.001 | 0.776 | 0.001 | 0.446 | 0.006 | 0.186 | 0.277 |
| 6 | 0.461 | 0.005 | 0.663 | 0.001 | 0.166 | 0.332 | -0.036 | 0.834 |
| 7 | 0.676 | 0.001 | 0.775 | 0.001 | 0.665 | 0.001 | 0.381 | 0.022 |
| 8 | 0.279 | 0.099 | 0.535 | 0.001 | 0.064 | 0.709 | 0.336 | 0.045 |
| 9 | 0.756 | 0.001 | 0.853 | 0.001 | 0.641 | 0.001 | 0.326 | 0.052 |
| 10 | 0.716 | 0.001 | 0.855 | 0.001 | 0.634 | 0.001 | 0.778 | 0.001 |
| 11 | 0.485 | 0.003 | 0.677 | 0.001 | 0.305 | 0.070 | 0.544 | 0.001 |
| 12 | 0.387 | 0.020 | 0.675 | 0.001 | 0.115 | 0.506 | 0.310 | 0.065 |

4 Discussion

4.1 Initial results

It is clear from the results that the experimental paradigm does indeed lead speakers to use different pitch spans. As expected, in expanded pitch spans the peak is higher in frequency, the difference between the high and low values is greater, and the syllables are longer. This experiment therefore provides the environment required to examine how the plateau is affected by constituent lengthening due to the non-structural factor of pitch span.

4.2 Increased rate of change of the fall

Although all the turning points are timed later in the expanded pitch span it is clear from figure 9 below that the time interval between the peak and the low tone remains constant and does not increase when the pitch span is greater. This is confirmed by calculating the duration in milliseconds between the timing of L and the timing of the peak. There is no significant effect of pitch span on the time taken to fall ($F(2,22) = 86.8$, $p > 0.05$), and this is confirmed by the lack of a correlation between the duration of the fall and the pitch span for 10 speakers, as shown in table 7.

The fall from H to L takes, on average, around 200 ms, even though in the expanded span the fall is much greater as the peak is higher in expanded span and L is stable for males, and slightly lower for females. This means that the rate of change must increase in wider spans. This is illustrated in table 8, which shows the mean rate of fall from the peak to the L in semitones per second. This finding may be related to the way in which English treats pitch contours under time pressure. As discussed by Ladd (1996: 132–136) and Grønnum (1991) there are at least two ways in which languages can choose to modify pitch gestures when there is only a small amount of segmental material available for voicing. In relation to falling pitch gestures, either the slope can change so that the same pitch target is reached (compression) or the slope can be maintained so that the gesture finishes at a higher value (truncation). When viewed in these terms it is clear that English (at least Southern British varieties) is a compressing language. For example, Grabe (1998) demonstrates that in the word *sheafer*, the slope of a fall will be shallower than in the word *shift* as *sheafer* contains more voiced material

Table 7 Correlation between pitch span and the time taken to fall from peak to L, for each speaker separately. Identification numbers for male speakers are shown in italics. R is Pearson's correlation coefficient and P is the significance level.

| Speaker | Duration of fall (ms) | |
|----------|-----------------------|-------|
| | R | P < |
| 1 | 0.290 | 0.044 |
| 2 | 0.250 | 0.141 |
| <i>3</i> | 0.024 | 0.891 |
| 4 | 0.001 | 0.996 |
| <i>5</i> | -0.181 | 0.291 |
| <i>6</i> | -0.461 | 0.005 |
| 7 | 0.116 | 0.500 |
| 8 | 0.264 | 0.120 |
| <i>9</i> | -0.260 | 0.125 |
| 10 | -0.006 | 0.973 |
| 11 | 0.154 | 0.370 |
| 12 | 0.155 | 0.366 |

Table 8 Mean rate of fall in semitones per second for each speaker in each pitch span. Identification numbers for male speakers are shown in italics.

| Speaker | Rate of change in semitones per second | | |
|----------|--|---------|----------|
| | Compressed | Neutral | Expanded |
| 1 | 34 | 41 | 77 |
| 2 | 20 | 27 | 49 |
| <i>3</i> | 36 | 50 | 89 |
| 4 | 21 | 31 | 50 |
| <i>5</i> | 22 | 28 | 63 |
| <i>6</i> | 23 | 36 | 74 |
| 7 | 28 | 36 | 39 |
| 8 | 23 | 37 | 53 |
| <i>9</i> | 22 | 29 | 54 |
| 10 | 20 | 35 | 63 |
| 11 | 22 | 32 | 56 |
| 12 | 26 | 33 | 57 |

over which the gesture can be realized. Nevertheless, the L is scaled at the same frequency in both words. The results described above for pitch span add extra support to the description of English as a compressing language. As well as the amount of segmental material available, a second factor that can affect pitch dynamics is the amount of tonal material to be produced. Thus, having to make a larger falling pitch movement (in a wider pitch span) over the same amount of material also results in a change of the slope of the fall.

4.3 Relationship between timing and alignment

As we have seen, for the same reference point there may be differences between the results for timing (an absolute measure) and those for alignment (a proportional measure). We have seen that all the reference points are timed later (further from the start of the syllable) when the expanded span is used. However, despite this difference in TIMING in expanded spans, we have also seen that the ALIGNMENT of EP and L is stable regardless of the pitch span

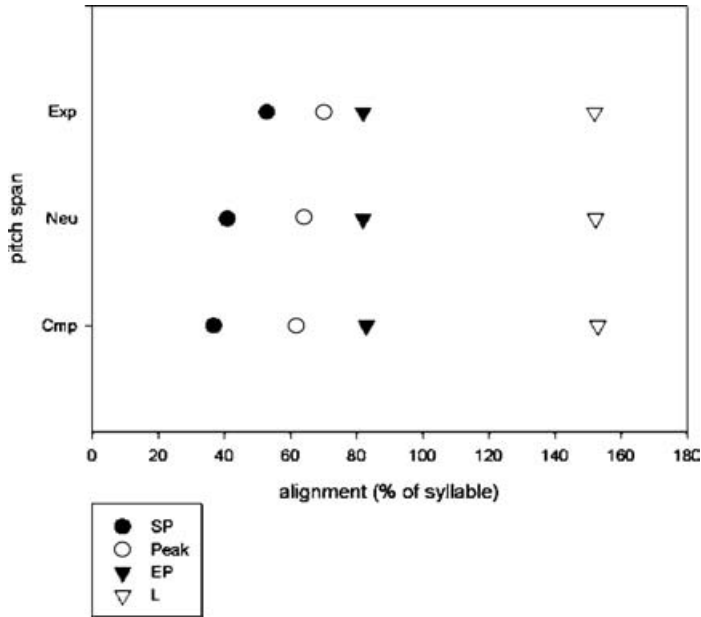


Figure 8 The alignment of the SP, peak, EP and low tone in different pitch spans, pooled across speakers.

used. Logically this must mean that the syllable has lengthened in such a way that this precise alignment comes about, and we have seen in the results for duration that the syllable is indeed longer in expanded spans. Crucially, however, SP and the peak are not stably aligned even though they show the same tendency as EP and L to be later in absolute terms. This finding suggests that different parts of the syllable are lengthened more than others in the expanded span. This differential lengthening results in the intonational reference points having similar patterns of timing but different patterns of alignment. The fact that the alignment of EP and L is stable, but that their timing is not, may suggest that speakers intend to place these reference points in proportional, rather than absolute, relation to the syllable.

4.4 Relation to initial hypotheses

As hypothesized, the plateau is shorter in absolute terms when speakers use a wider pitch span. The results for alignment, however, are not completely consistent with either of the alternative hypotheses. The beginning of the plateau behaves as predicted by both hypotheses. That is to say, it is timed and aligned later in the syllable when there is an expanded pitch span and is affected in the same way as the peak by this non-structural variable. In terms of absolute timing, the end of the plateau behaves as expected in that it is later in the expanded span. In terms of proportional alignment, however, it behaves very differently. As figure 8 shows, the end of the plateau is aligned at a fixed position in the accented syllable. The fact that it is unaffected by the non-structural variable pitch span may suggest that it is a marker of linguistic structure. House et al. (1999) and Knight (2004) show that the alignment of the end of the plateau varies with syllable structure, and Knight (2003: chapter 4) indicates that the end of the plateau may be earlier in the foot for some speakers before a word boundary. Taken together, these results suggest that the alignment of the end of the plateau is signalling something about the linguistic structure of the utterance.

The obvious question concerns why the end of the plateau, rather than the peak or the start of the plateau, should be anchored and used to signal structure. It seems plausible that this point is important not so much in its role as the end of the plateau but as the beginning of

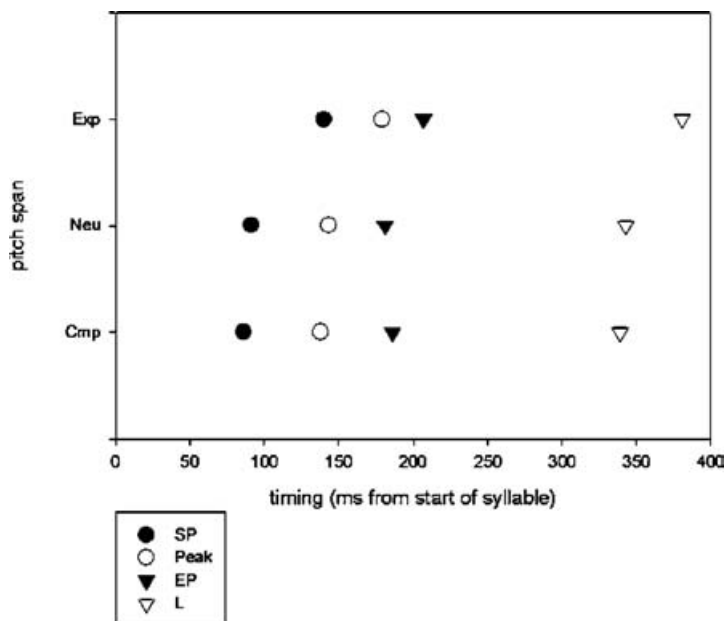


Figure 9 The timing of SP, peak, EP and low tone in different pitch spans, pooled across speakers.

the fall. The end of the plateau, after all, marks the first point at which the pitch falls outside the range of perceptual equality, and it is at this point that a listener can tell that a fall is taking place. This episode of rapid change in the signal is likely to be extremely salient and may in fact be the point where the rate of change of F0 becomes noticeable. The alignment of this perceptually salient pitch event, then, can signal to listeners what point in the utterance has been reached, and whether or not they should expect more syllables or words to follow.

5 Summary and conclusion

This paper investigated whether the duration and alignment of intonational plateaux are affected by the non-structural variable of pitch span. In terms of duration it was hypothesized that plateaux would be shorter in wide pitch spans due to time limitations when the rise and fall are greater. Two alternative hypotheses were suggested for alignment; either the whole plateau would be later in the syllable in expanded pitch spans (in line with results for peak alignment under non-structural changes in lengthening) or the plateau would contract around the peak due to time limitations. The prediction for duration was borne out, and in terms of timing all the intonational reference points occurred later in the expanded span. The predictions for alignment however, were only partially borne out. The beginning of the plateau was, as hypothesized, aligned later in an expanded pitch span. However, the end of the plateau (and also the following low point) was firmly anchored inside the syllable for each speaker. This suggests that what signals linguistic structure may be not so much the high and low turning points themselves, but rather the perceptually salient rapid changes in F0 such as those delimited by the end of a high intonational plateau and the following low turning point.

Acknowledgements

We would like to thank the members of the University of Cambridge Linguistics Department Prosody Supervision Group for their assistance and for providing the data. We would also like to thank Sarah

Hawkins, Bob Ladd and Gösta Bruce for very detailed and helpful comments on earlier versions of this paper. Thanks also to Rachel Smith for her translation work and for recording the female template utterance, and to Maria Ovens for proof-reading.

Appendix: Figures for individual speakers

The following information is given for the benefit of other researchers. It includes the syllable length, and F0 of the peak and low tone in the three different pitch spans.

Table A1 Average F0 of the peak and low tone and syllable duration for each speaker in each pitch span.

| Speaker | Compressed | Neutral | Expanded |
|----------------|------------|---------|----------|
| PEAK FO (Hz) | | | |
| 1 | 218 | 251 | 328 |
| 2 | 234 | 268 | 356 |
| 3 | 119 | 139 | 206 |
| 4 | 257 | 286 | 356 |
| 5 | 138 | 146 | 187 |
| 6 | 118 | 136 | 187 |
| 7 | 229 | 273 | 353 |
| 8 | 226 | 253 | 311 |
| 9 | 132 | 138 | 177 |
| 10 | 223 | 261 | 344 |
| 11 | 220 | 262 | 364 |
| 12 | 228 | 257 | 336 |
| LOW FO (Hz) | | | |
| 1 | 166 | 167 | 154 |
| 2 | 181 | 184 | 176 |
| 3 | 84 | 84 | 80 |
| 4 | 200 | 200 | 198 |
| 5 | 113 | 113 | 113 |
| 6 | 91 | 92 | 96 |
| 7 | 169 | 173 | 168 |
| 8 | 175 | 184 | 176 |
| 9 | 106 | 106 | 107 |
| 10 | 170 | 171 | 157 |
| 11 | 166 | 178 | 169 |
| 12 | 167 | 169 | 167 |
| SYLL DUR. (ms) | | | |
| 1 | 243 | 359 | 290 |
| 2 | 228 | 220 | 253 |
| 3 | 243 | 241 | 276 |
| 4 | 224 | 221 | 261 |
| 5 | 233 | 239 | 263 |
| 6 | 210 | 217 | 225 |
| 7 | 225 | 236 | 278 |
| 8 | 203 | 198 | 230 |
| 9 | 225 | 230 | 261 |
| 10 | 213 | 208 | 234 |
| 11 | 236 | 227 | 283 |
| 12 | 198 | 188 | 208 |

References

- GRABE, E. (1998). Pitch accent realization in English and German. *Journal of Phonetics* **26**, 129–143.
- GRØNNUM, N. (1991). Prosodic parameters in a variety of regional Danish standard languages, with a view towards Swedish and German. *Phonetica* **47**, 188–214.
- HART, J. (1981). Differential sensitivity to pitch distance, particularly in speech. *Journal of the Acoustical Society of America* **69**, 811–821.
- HOUSE, D. (1990). *Tonal Perception in Speech*. Lund: Lund University Press.
- HOUSE, J., DANKOVIČOVÁ, J. & HUCKVALE, M. (1999). Intonational Modelling in ProSynth: An integrated prosodic approach to speech synthesis. *Proceedings XIVth International Congress of Phonetic Sciences* **3**, 2343–2346.
- KLATT, D. (1973). Discrimination of fundamental frequency contours in synthetic speech: implications for models of pitch perception. *Journal of the Acoustical Society of America* **53**, 8–15.
- KNIGHT, R.-A. (2003). Peaks and plateaux: the production and perception of intonational high targets in English. Ph.D. thesis, University of Cambridge.
- KNIGHT, R.-A. (2004). The realisation of intonational plateaux: effects of foot structure. In Astruc, L. & Richards, M. (eds.), *Cambridge Occasional Papers in Linguistics* **1**, 157–164.
- LADD, D. R. (1996). *Intonational Phonology*. Cambridge: Cambridge University Press.
- LADD, D. R., FAULKNER, D., FAULKNER, H. & SCHEPMAN, A. (1999). Constant ‘segmental anchoring’ of F0 movements under changes in speech rate. *Journal of the Acoustical Society of America* **106**, 1543–1554.
- LADD, D. R. & MORTON, R. (1997). The perception of intonational emphasis: continuous or categorical? *Journal of Phonetics* **25**, 313–342.
- NOLAN, F. (2002). Intonation in speaker identification: an experiment on pitch alignment features. *Forensic Linguistics* **9**, 1–21.
- PIERREHUMBERT, J. (1979). The perception of fundamental frequency declination. *Journal of the Acoustical Society of America* **66**, 363–369.
- ROSEN, S. & FOURCIN, A. (1986). Frequency selectivity and the perception of speech. In Moore, B. (ed.), *Frequency Selectivity in Hearing*, 373–488. London: Academic Press.
- ROSSI, M. (1971). Le seuil de glissando ou seuil de perception des variations tonales pour les sons de la parole. *Phonetica* **23**, 1–33.
- SILVERMAN, K. & PIERREHUMBERT, J. (1990). The timing of prenuclear high accents in English. In Kingston, J. & Beckman, M. (eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, 72–106. Cambridge: Cambridge University Press.
- STEELE, S. (1986). Nuclear accent F0 peak location: effects of rate, vowel, and number of following syllables. *Journal of the Acoustical Society* (Supplement 1) **80**, s51.
- WICHMANN, A., HOUSE, J. & RIETVELD, T. (1999). Discourse constraints on peak timing in English: experimental evidence. *Proceedings of XIVth International Congress of Phonetic Sciences* **3**, 1765–1768.
- XU, Y. (2002). Articulatory constraints and tonal alignment. In Bel, B. & Marlien, I. (eds.), *Proceedings of the 1st International Conference on Speech Prosody*, Aix-en-Provence, France, 91–100.