The Effect of Pitch Span on Intonational Plateaux

Rachael-Anne Knight

Department of Linguistics
University of Cambridge, UK
rak34@cam.ac.uk

Abstract

Previous research has indicated that the H* of a nuclear accent may be realised as a flat stretch of contour rather than as a single turning point. The duration of this plateau and its alignment within the accented syllable and foot 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 realisation of the plateau. Speakers imitate all-sonorant utterances in different pitch spans. Both the duration and alignment of the plateau vary with pitch span but in ways different from their variation with prosodic structure. Results also indicate that the beginning and end of the falling movement within the contour are anchored within the syllable and foot for each speaker.

1. Introduction

Autosegmental Metrical theories of intonation may lead us to expect pitch targets to be realised as single turning points in the fundamental frequency (F0) contour. However, several studies (e.g. [1] [2] [3] [4]) have shown that on closer examination H targets often appear as flat stretches or plateaux. The proportional duration of these plateaux and their alignment within the accented syllable or nuclear foot are affected by both the segmental structure of the syllable and the prosodic structure of the utterance. [2] examine a medium-sized single-speaker database. They show that plateaux 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, [1] [2] and [3] find that the plateau is aligned later in the syllable when the foot contains more syllables. This finding fits well with the large body of literature on peak alignment (summarised in [5]). Several studies suggest that peak alignment is affected by the lengthening of prosodic structures such as the syllable or the foot. Different causes of lengthening have different effects on peak alignment. For example, studies by [6] and [7], 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 in the unit under consideration.

Peak alignment is also affected by lengthening induced by non-structural factors such as intrinsically long vowels. For example, [6] and [7] note that peaks move later in a syllable that is lengthened due to a slower tempo. Other work by [8] suggests that similarly peaks are aligned later in the syllable when the syllable is lengthened in an expanded pitch span. The present study investigates how pitch span affects the alignment and the duration of intonational plateaux.

Pitch span is defined by [9] as the difference between high and low targets in the speaker's range. When pitch span increases high targets are higher than in a normal span while low targets remain at the same level or are slightly lower than usual. Thus the distance between high and low targets increases. Although [10] classes pitch span as a prosodic variable it is also possible to think of it as a paralinguistic variable. [9] defines paralinguistic variables as those which are gradient rather that categorical in nature and in [11] pitch span is considered to be paralinguistic when it conveys something about the speaker's attitude to what they are saying. As the perceptual correlate of an increase in pitch span is an increase in emphasis, pitch span is considered to be a non-structural, paralinguistic variable for the purposes of this study.

It is hypothesized that plateaux will be shorter in wider pitch spans due to physiological causes. 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. Assuming a constant rate of change in F0 and the same constraints on when F0 starts to fall again it is possible that there will be less time available to remain at the high level and realise 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. Following [1] the plateau is defined as the range of frequencies that fall within 4% of any absolute peak and therefore it is possible that each end of the plateau will be affected in the same way as the peak itself. As discussed above, peaks are aligned later when syllables are lengthened by non-structural factors such as pitch span and it is hypothesized on this basis that the plateau's alignment may be similarly affected.

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. An imitation task is used to ensure subjects produce utterances in different pitch spans.

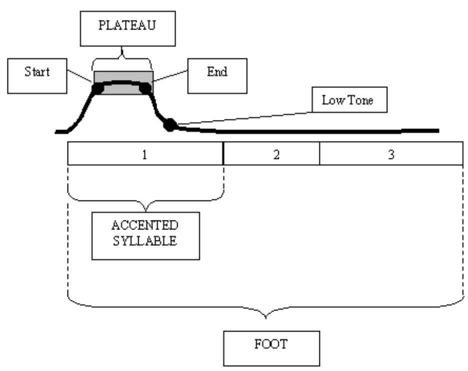

Figure 1: Measurements taken.

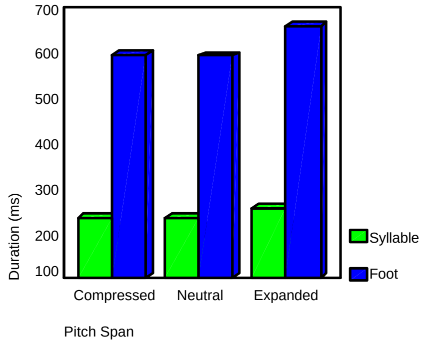

Figure 2: Duration of the accented syllable and the foot in the three pitch spans.

2. Method

2.1. Stimulus Materials

Two utterances composed of entirely sonorant material were used as stimuli in order to minimize microprosodic effects. They are (with tonetic stress marks):

A. ↑We were re↓↗ lying on a ↑\ milliner
B. A ↑∨ milliner

Two phoneticians, one male and one female, recorded each sentence in three different pitch spans referred to as neutral, compressed (narrower than the neutral span) and expanded (wider than the neutral span). Three tokens of each utterance were used in the experiment giving 36 (3x3x2x2) 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 randomised.

2.2. Subjects

12 native speakers of British English participated in the experiment. There were 8 females and 4 males whose ages ranged from 21-29. All were students at the University of Cambridge with some training in phonetics and intonation.

2.3. Procedure

Subjects were instructed to repeat each stimulus exactly, aiming to produce an intonationally equivalent utterance in their own voice

Stimuli were presented through headphones in a sound-treated booth. After each utterance had been played, an on-screen message prompted subjects to record an imitation in their own time. Their speech was recorded directly onto the hard-drive of an SG workstation via a high quality microphone. Recordings that were unsatisfactory were repeated at the time of recording.

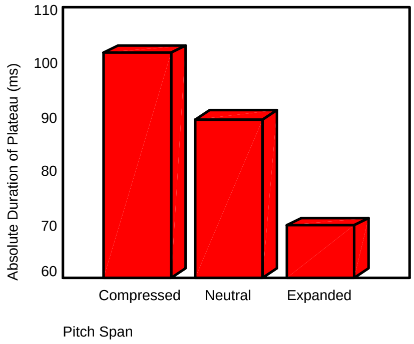

Figure 3: Absolute duration of the plateau in the three pitch spans.

2.4. Method of analysis

The measurements taken and the relationships between them are shown in Figure 1. The start and end points of the plateau were, following [1], defined as 4% below the absolute maximum frequency in the nucleus. Four percent is the range of perceptual equality for speech sounds, so listeners could be expected to hear anything in this range as being equal in pitch ([12]). The elbow in the contour representing the following low tone was also located for comparison with the high reference points.

The duration of the accented syllable and of the foot were measured. Following [1] feet are defined as in the Abercrombian [13] tradition. A foot contains one initial stressed syllable plus any following unstressed syllables. Unstressed syllables at the beginning of an utterance (before the first stressed syllable) comprise a degenerative foot. In this experiment, the foot of interest consists of the word 'milliner' and the accented syllable is /mIl/.

Analyses used the repeated measures design (MANOVA) available in SPSS. 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 (2) and pitch span (3). Space constraints mean only results for pitch span can be reported here.

3. Results

3.1. Pitch height and span

Two tests were conducted to see if subjects were in fact using different pitch spans. As expected the maximum frequency was significantly higher in the more expanded pitch spans ($F(1.016, 11.178) = 127.2$, $p<0.01$.). Planned comparisons using paired t-tests on the means of results for each pitch span indicate the frequency is lower in the compressed than the neutral range ($t(11) = 8.3$, $p<0.01$) and lower in the neutral than the expanded range ($t(11) = 12.5$ $p<0.01$).

The difference between the maximum frequency and the frequency of the low point also increases ($F(1.019, 11.205) = 109$, $p<0.01$). The difference is smaller in the compressed than the neutral range ($t(11)= 8.5$, $p<0.01$) and smaller in the neutral than the expanded range ($t(11) = 10.9$, $p<0.01$). These results indicate that the experimental paradigm works; when imitating more expanded spans high targets are higher and the difference between high and low targets is greater showing that subjects are producing a true difference in pitch span in the different conditions.

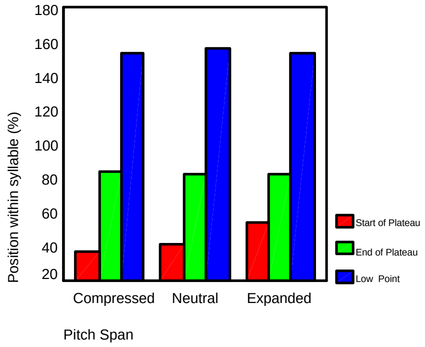

Figure 4: Alignment of the three reference points within the syllable.

3.2. Duration of the syllable and foot

Figure 2 shows that, as expected, both the duration of the syllable and of the foot are significantly affected by pitch span (syll: F(1.232, 15.552) = 8.2, p<0.01, foot: F(1.204, 13.241) = 69.3, p<0.01). There is no significant difference between the duration of the units in the compressed compared to the neutral span (syll: t(11) = 0.82, p> 0.01), foot: t(11) = 0.425, p>0.01). The difference between the durations in the neutral and expanded spans is significant however. Both the syllable (t(11) = 8.2 p<0.01) and the foot (t(11) = 10.9, p<0.01) are longer in the expanded than the neutral pitch span. .

3.3. Duration of the plateau

Figure 3 shows that, 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). The plateau is significantly longer in the compressed than the neutral span (t(11) = 7.6, p<0.01) and significantly longer in the neutral than the expanded span (t(11) = 6.0, p<0.01).

Pitch span also significantly affects the proportion of the syllable (F(2, 22) = 80.4, p<0.01) and the foot (F(2, 22) = 123.8, p<0.01) taken up by the plateau. Pitch span affects the proportional duration of the plateau in the same way as it affects the absolute duration. The plateau takes up more of the syllable or foot in the compressed than in the neutral span (syll: t(11) = 4.0, p <0.01, foot: t(11) = 6.1, p<0.01) and it takes up more of these units in the neutral than in the expanded span (syll: t(11) = 7.4, p<0.01, foot: t(11) = 8.7, p<0.01).

3.4. Alignment of the plateau

Alignment is calculated by measuring the duration from the beginning of the syllable to the reference point in question. This duration is then divided by the duration of the prosodic unit and expressed as a percentage. The results for alignment can be seen in Figure 4.

As expected, the alignment of the beginning of the plateau is significantly affected by pitch span whether this distance is measured as a proportion of the duration of the syllable (F(2, 22) = 44.8, p<0.01) or the foot (F(2, 22) = 44.3, p <0.01 There is no significant difference between the location of the beginning of the plateau in compressed and neutral spans (syll: t(11) = 2.5 p>0.01, foot: t(11) = 2.6, p >0.01). However, in both syllables and feet, the location of the plateau is significantly earlier in neutral than in expanded spans (syll: t(11) = 7.0, p<0.01), foot: t(11) = 7.3, p<0.01.

Interestingly, the location of the end of the plateau is not significantly affected by pitch span no matter whether this distance is measured as a proportion of the syllable (F(2, 22) = 0.6, p>0.05) or foot (F(2, 22) = 1.1, p>0.05). Thus, the alignment of the end of the plateau is consistently anchored for each subject regardless of the pitch span used.

3.5. Alignment of the low tone

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 (syll: (F(2, 22) = 0.1, p>0.05, foot: (F(1.230, 13.525) = 0.2, p>0.05). Again, the location is anchored for each speaker.

4. Discussion

It is clear from the results that the experimental paradigm does result in speakers using different pitch spans. As expected, the high points are higher in expanded pitch spans, the difference between the high and low values is greater, and the syllables and feet are longer. This therefore provides the environment required to examine how the alignment of the plateau is affected by constituent lengthening due to the non-structural factor pitch span. Results for alignment are discussed only in terms of the syllable, as there is no substantial difference between these and the results for the foot.

The results for syllable and foot duration and alignment of the beginning of the plateau do not fall into distinct categories for the lower (normal and compressed) pitch spans. This is probably due to there being more scope for expanding than compressing pitch span in relation to the neutral span as compressing the span too much will sound monotonous. This suggestion is supported by the difference between the high and low targets in the different ranges. Although all three are significantly different from one another the span is 69 Hz in the neutral range, 46 Hz in compressed and 142 Hz in the expanded range. Figures for the durations follow the same pattern. These results would be explained by a constant rate of change in F0.

The results for alignment are not completely consistent with either of the alternative hypotheses. The beginning of the plateau behaves as predicted by either hypothesis. That is to say it is aligned later in the syllable when there is an expanded pitch span. The end of the plateau however behaves very differently. As Figure 4 shows, the end of the plateau is aligned at a fixed position in the syllable. This is interesting in the light of the alignment of the low tone, which is also anchored at a fixed position. As these two points are anchored, this means that the slope between them changes with pitch span, becoming steeper in the expanded spans.

As hypothesized, the plateau is shorter in absolute terms when speakers use a wider pitch span, as shown in Figure 3. The anchoring of the end of the plateau however suggests that this shortening may not be due solely to time limitations when the rise and fall are greater as was originally hypothesized. The explanation for the shortening may lie, rather, in the combination of this anchoring and later peak alignment in wider spans (such as that reported by [8]) which in turn results in the later alignment of the beginning of the plateau.

Further work [14] is in progress to investigate this issue by examining the alignment of the peak within the syllable and within the plateau.

It seems likely that the anchoring of the end of the plateau contributes to how the utterance's linguistic structure is signalled. [2] and [3] show that the alignment of the end of the plateau varies with syllable structure. Unpublished research by the present author indicates that the end of the plateau may be earlier in the foot for some speakers when there is an upcoming word boundary. This suggests that the alignment of the end of the plateau is signalling something about the linguistic structure of the utterance, perhaps indicating whether there are more syllables or words to come.

The obvious question concerns why it should be the end of the plateau that is anchored and used to signal structure. This is the sort of role that is traditionally assigned to the pitch peak; however, it can clearly not be the peak that is responsible for the effects found here. By definition the end of the plateau is 4% down from any absolute peak. It is very rare for the end of the plateau and the peak to be one and the same, although it does happen (occasionally, for example, when the syllable onset is voiceless). In addition, previous studies such as [8] suggest that the peak is not anchored but is aligned later in wider pitch spans.

It seems plausible that it is not so much the end of the plateau but the beginning of the fall that is anchored. The end of the plateau after all marks the first point at which the pitch falls outside the range of perceptual equality. The critical factor then may be the shift from 'no change' to 'change', a point in the signal that is likely to be extremely salient. The alignment then of this perceptual marker can signal to the listener where they are in the utterance and whether they should expect more syllables in the foot.

5. Summary and Conclusion

This paper investigated whether the duration and alignment of intonational plateaux are affected by 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 and foot with wide 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 although an explanation based solely on time constraints is not fully supported. The predictions for alignment were only partially correct. The beginning of the plateau was, as hypothesized, aligned later in an expanded pitch span; but the end of the plateau (and also the following low point) was firmly anchored inside the syllable and foot 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 changes in F0 from relatively stable to rapidly changing, or vice versa.

6. Acknowledgments

I am grateful to Sarah Hawkins and Francis Nolan for their help while I was writing this paper. I would like to thank the members of the Linguistics Department Prosody Supervision Group for their assistance and for providing the data and also an anonymous reviewer of an earlier version of this paper.

7. References

[1] House, J.; Dancovicova, J.; Huckvale. M., 1999. Intonational Modelling in ProSynth: An integrated prosodic approach to speech synthesis. In *Proceedings of the XIV ICPhS Vol 3 University of California, Berkeley, CA,* 2343-2346

[2] House, J.; Dancovicova, J.; Huckvale. M., 1999. Intonation Modelling in ProSynth. Poster presented at XIV ICPhS, San Francisco

[3] Ogden, R.; Hawkins, S.; House, J.; M. Huckvale; Local, J.; Carter, P.; Dancovicova, J.; Heid S., 2000. ProSynth: an integrated prosodic approach to device independent, natural-sounding speech synthesis. *Computer Speech and Language,* 14, 177-210

[4] Wichmann, A.; House, J.; 1999. Discourse constraints on peak timing in English: Experimental Evidence. In *Proceedings of the XIV. ICPhS, San Francisco*, 1765-1768

[5] House, J.; Wichmann, A., 1996. Investigating peak timing in naturally-occurring speech: from segmental constraints to discourse structure. In *Speech, Hearing and Language: work in progress. Volume 9,* 99-117. UCL

[6] 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

[7] Silverman, K.; Pierrehumbert, J. 1990. The timing of prenuclear high accents in English. In *Papers in Laboratory Phonology I: Between the grammar and Physics of Speech*, Kingston, J.; Beckman, M. eds. Cambridge: CUP, 72-106

[8] Ladd, R.; Morton. R., 1997. The perception of intonational emphasis: continuous or categorical? *Journal of Phonetics, 25,* 313-342.

[9] Ladd, R., 1986. *Intonational phonology*. Cambridge: CUP

[10] Bruce, G., 1990. Alignment and composition of tonal accents: comments on Silverman and Pierrehumbert's Paper. In *Papers in Laboratory Phonology I: Between the grammar and Physics of Speech*, Kingston, J.; Beckman, M. eds. Cambridge: CUP, 107-114

[11] Crystal, D.; Quirk, R., 1964. *Systems of prosodic and paralinguistic features in English.* The Hague: Mouton

[12] Rosen, S.; Fourcin, A., 1986. Frequency selectivity and the perception of speech. In *Frequency Selectivity in Hearing.* Moore, B. ed. 1986. London: Academic Press

[13] Abercrombie, D., 1965. *Studies in Phonetics and Linguistics.* London: OUP

[14] http://kiri.ling.cam.ac.uk/~rachael/research/aix.html