The shape of nuclear falls and their effect on the perception of pitch and prominence: peaks vs. plateaux

Rachael-Anne Knight

City University, London.

Running Head: The perception of nuclear falls.

Acknowledgements

Thanks to Laura Dilley and an anonymous reviewer for helpful comments on an earlier version of this paper. Also to Rachel Smith, Judith Broadbent and Maria Ovens for useful comments on this version, Bob Ladd for assistance at various stages of this research, and the LCS Writing Group for support.

Please contact the author at Department of Language and Communication Science, City University, Northampton Square, London, EC1V 0HB. Email: knight@city.ac.uk

ABSTRACT

This paper investigates the perceptual effect of a high plateau in the intonation contour. Plateaux are flat stretches of contour and have been observed associated with high tones in Standard Southern British English (SSBE). The hypothesis that plateaux may make the accents with which they are associated sound higher in pitch than sharp peaks of the same maximum frequency is tested experimentally. In the first experiment listeners heard pairs of resynthesized utterances where the nuclear accent differed only in shape, not frequency. They indicated which stimulus they thought contained the higher pitched accent. Results showed that plateau-shaped accents sound higher than peaks. In the second experiment the effect of a plateau on prominence relations within an utterance is investigated. Listeners heard resynthesized sentences, and compared two accents. One group indicated which accent sounded higher in pitch and the other indicated which sounded more prominent. Results again indicated that plateau-shaped accents sound higher in pitch and also more prominent; judgements of pitch and prominence were very similar to one another. The results from both experiments indicated that accent shape is a perceptually important variable, although such a fine level of detail is not taken into account by autosegmental-metrical theories of intonation.

Key Words: Intonation, perception, plateaux, pitch, prominence.

INTRODUCTION

Many theories of intonational phonology, especially those in the autosegmental-metrical tradition popularized by Pierrehumbert (e.g. 1980), represent intonation contours as strings of high (H) and low (L) pitch targets. These targets are considered to be linguistically important (analogous to phonemes), whilst the transitions between them are seen as linguistically unimportant interpolations. In recent years, much work in the field of intonation has focussed on establishing how high and low targets are aligned with the segmental string. This process usually begins by visually identifying high and low turning points in the fundamental frequency (F0) contour, on the assumption that these turning points are associated with the underlying high and low pitch targets (see Dilley 2005, esp. chapter 4 for an overview of these, and other, assumptions). The temporal alignment of the turning points is then expressed in relation to a tone-bearing unit (such as the syllable or foot) by calculating the absolute or proportional duration from the start of the unit to the turning point. More recent work by Ladd and colleagues (e.g. Arvaniti et al. 1998, Schepman et al. 2006) expresses alignment of turning points in relation to smaller units such as individual vowels and consonants.

Despite the widespread practice of visually identifying turning points in the contour it is acknowledged in the literature that turning points can be difficult to locate. Difficulties may arise because turning points are obscured by voiceless sections of the contour, or by microprosodic perturbations. Most importantly for the focus of the present work, turning points are impossible to locate when the intonation contour forms a plateau, where a pitch level is sustained rather than forming a sharp peak or valley. Figure 1 shows a high plateau associated with the word ‘Manny’ in the phrase ‘Anna came with Manny’ spoken by a female speaker of Southern British English. Even though this plateau is not completely flat it is difficult to know where exactly one should label the high tone. House and Wichmann (1996, p.3) and D’Imperio (2002, p.103) both describe the difficulties in identifying turning points when accents contain plateaux. As D’Imperio (2002) suggests, in the majority of cases where plateaux, or other effects, prevent the identification of a turning point, researchers can only make stipulative decisions about the location of underlying tonal targets. For example, Silverman and Pierrehumbert (1990, p.79) note that “where there were two alternative locations that could be construed to be the relevant F0 peak, we took the average of the two”. Often the decisions taken are not fully justified, and are not comparable across different studies, which raises a number of difficulties. If a decision about the location of a tone is only stipulative this may affect the results of studies on alignment. Choosing, for example, one end of a plateau rather than the other as the intended tonal target will affect how we believe the target to be aligned in relation to the syllable or segmental string. In turn this may affect conclusions about tonal identity, and perhaps even theoretical considerations such as whether contours are best expressed as levels or configurations. Furthermore, the very existence of plateaux raises questions about the nature of the theoretical description of an intonation contour in terms of high and low targets. If no single turning point can be found, it is more difficult to see how contours are best represented as strings of high and low targets joined by interpolations.

.

FIGURE 1 ABOUT HERE

Some studies have explicitly investigated aspects of plateaux in intonation contours. The early work in this area focused on describing how the realisation of accents is related to segmental and prosodic variables. This work was conducted on falling nuclear accents in Standard Southern British English (SSBE). House et al. (1999) made a large study of accent shape in the speech of one male SSBE speaker. House et al.'s ultimate aim was to reduce an intonation contour to a small number of points for use in speech synthesis. It was initially expected that a falling (H*+L) accent could be represented by a single point to characterize each tone. However, it became apparent that the high part of the falling accent was often realized as a sustained plateau, which it was necessary to model using two points, representing the plateau's beginning and end, if the resulting synthesized utterances were to sound natural.

House et al.'s further work investigated how the realisation of plateaux varied with segmental and prosodic factors. This work was extended by Knight (2003) to include a larger number of speakers and paralinguistic variables such as pitch span. In both House et al. (1999) and Knight (2003) plateaux were operationally defined as including all the intonation contour within 4% of an accent's maximum frequency in Hertz, as this approximates the threshold of perceptual equivalence according to a review of the literature by Rosen and Fourcin (1986) who refer to studies such as 't Hart (1981) (a fuller description and justification of this methodology is given in Knight and Nolan 2006). Both studies found that the realisation of plateaux covaries with linguistic structure. For example, plateaux are longer in duration when the syllable with which they are associated contains a sonorant onset and coda, and they stretch over proportionally more of the syllable in polysyllabic than monosyllabic feet. Wichmann et al. (1999) also describe plateau-shaped accents in SSBE, focussing on the realisation of accents in relation to discourse constraints. Plateaux (identified by a visual examination of the contour) were found to occur in paragraph-initial position, and it was suggested that plateaux may be an alternative way of achieving the peak delay that was found to be characteristic of paragraph-initial position.

It is rather unclear why speakers produce plateaux. Particularly, plateaux seem to be a feature of only certain types of accents, having been found most commonly in nuclear and paragraph-initial positions. Indeed, initial work to model (non paragraph-initial) prenuclear accents for synthesis indicated that only a single point was necessary to represent high tones (Knight 2000). The restriction of plateaux to a limited number of prosodic positions may suggest that speakers have some degree of choice over accent shape, and choose a plateau in order to signal something about prosodic structure to the listener. This view is perhaps supported by the finding of Dilley and Brown (in press) that speakers often produce peaks and valleys when imitating level tones. Alternatively, speakers may be responding to some kind of physical constraint on production that leads to plateau-shaped accents only in particular environments. A first step in deciding if speakers use plateaux to convey some kind of message (in terms of signalling prosodic structure) is to investigate whether plateaux are perceptibly different from accents of other shapes. It is unlikely that speakers will try to signal any kind of message using a distinction that is not perceptible to listeners. Therefore the aim of this paper is to address the significance of plateau-shaped accents by investigating the perceptual effect of a plateau in the intonation contour. This paper will concentrate specifically on the perceived pitch height and prominence of plateaux, as there is a range of work which demonstrates that frequency glides may be perceived differently depending on whether they are bounded by a stretch of steady frequency (a plateau). The following sections review some of this work and consider the relevance to the plateaux found in SSBE.

The perception of accent shape

There have been a variety of studies dealing, often indirectly, with the perceptual effect of accent shape. These studies span a wide timeframe, some concern speech in different languages, and some concern non-speech stimuli. What follows is an overview of these studies, beginning with those addressing accent shape in general, followed by those directly related to the perceptual effect of plateaux. This overview is followed by a discussion of the studies' relevance for the present paper.

The aim of Rossi (1971) was to find out, firstly, whether or not a frequency variation in a given time is perceptible and, secondly, how a frequency glide affects the perceived pitch of a vowel. Subjects compared a static tone and a rising tone and indicated whether the second tone was higher, lower or of equal pitch to the first. 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. This indicates that the perceived pitch of a glide does not necessarily correspond to its end points, suggesting that in contours with sharp peaks it may not be the highest frequency that corresponds to the perceived pitch.

Niebuhr's (2003) work with German suggests that the entire intonation contour, rather than just the timing and scaling of turning points, can influence a listener's perception. Niebuhr manipulated the sentence '*Sie war mal Malerin'* ('she was once a painter') so that the accent on the first syllable of '*Malerin*' varied in two ways. The first variable was the alignment of the peak, which could be early (suggesting known information) or medial (suggesting new information). For stimuli with each of these alignments the accent shape varied so that it had a fast or slow rise, and a fast or slow fall. Listeners took part in a discrimination experiment, and an identification experiment where they decided if the sentence matched a preceding context designed to make them expect new information. The results showed that accent shape interacted with peak alignment in giving rise to perceived pitch. Specifically, peaks with a fast rise and slow fall showed more 'new' responses, whilst those with a slow rise and fast fall showed more 'known' responses. These findings suggest that the "perception of intonational contrasts depends on holistic contours rather than local features" (2003, p. 1228).

Studdert-Kennedy and Hadding (1973) came to similar conclusions to Niebuhr (2003). The word 'November' was vocoded so that the F0 was manipulated at four points (the start, peak, low turning point and end) to create 72 contours. Each contour was presented to subjects as either speech, a frequency modulated sine wave, or a pulse train. Subjects made various linguistic and auditory judgements such as whether they heard a question or a statement. For the present paper the relevant finding is that many aspects of the F0 contour engaged in trading relations. Although the terminal glide (from the low turning point to the end of the contour) was found to be a powerful predictor of listeners' responses, its effect could sometimes be overridden by earlier sections of the contour, again showing the importance of the entire contour for influencing listeners' perception.

Gosy and Terken (1994) also investigated the effect of accent shape, with a focus on Hungarian. In Hungarian the distinction between statements and questions is signalled by prosody alone rather than by the lexical or syntactic means employed by many other languages. The experiments were designed to determine which aspects of the intonation contour signal this distinction. Subjects listened to resynthesized sentences and identified each one as a statement or a question. In each case the relevant accent was a rise-fall where the rise and fall each took place over 120 ms (steep) or 240 ms (gradual). A gradual rise in the contour resulted in fewer question responses than a steep rise, whilst a gradual fall led to more question responses than a steep fall. Again this indicates that target and interpolation models may not capture all the perceptually relevant aspects of a contour.

While the experiments discussed above investigated the perceptual effect of accent shape generally, other experiments have specific relevance for the perceptual effect of a plateau in the intonation contour. 't Hart (1991) investigated the perceptual effect of accent shape in order to identify the best type of approximation to use in the IPO speech synthesis system. 't Hart observed that although the contours used in synthesis often consist of straight lines, real F0 contours are usually curved. Experiments tested whether there is any audible difference between synthetic contours created using straight lines, parabolas or flattened peaks containing 30 ms-40 ms plateaux. Thirteen highly musical subjects took part in a four interval, two alternative, forced choice task. In each trial they heard two pairs of stimuli; one pair contained identically shaped stimuli while in the other pair the stimuli were shaped differently. Subjects indicated which pair contained different items. Seven subjects perceived no differences between the different sets of stimuli and their results were not analysed further. The other six subjects were classified as accurate observers as they could perceive differences between some of the different sets. These six subjects could distinguish sharp peaks from parabolas and from flattened peaks, but could not distinguish parabolic curves from flattened peaks. 't Hart concluded that it was acceptable to use straight lines to approximate natural F0 contours in the IPO system as long as accents also contained a short flattened peak. In relation to the plateaux found in SSBE, 't Hart's results suggest that, at least for some subjects, there will be an audible difference between accents that contain a plateau and those that contain a sharp peak.

Although 't Hart's experiment indicates that, for some subjects, there is a perceptible difference between plateau and peak shaped accents, the results give little indication of the nature of this difference. Work by D'Imperio, however, begins to give an indication of the type of difference created by a plateau in the contour. D'Imperio and House (1997) investigated the effect of a difference in contour shape in Neapolitan Italian. In this dialect both questions and narrow focus statements are realized by a rise-fall accent, the peak of which is aligned around 40 ms later in questions. In perceptual experiments using resynthesized stimuli, subjects identified earlier peaks as statements and later peaks as questions. D'Imperio (2000) also demonstrated that there was a difference created by contour shape. The responses to stimuli with sharp peaks were compared with responses to stimuli where the accent contained a 45 ms plateau, in order to test the hypothesis that the number of question responses would depend on where the peak was timed relative to the beginning of the plateau. When the peak was timed at the beginning of the plateau, the stimuli containing plateaux elicited more question responses. However, when a comparison was made between results for peaks located at the end of plateaux, there was very little difference in the percentage of question responses, suggesting that "the shape of the accent peak does indeed affect the perception of target location" (D'Imperio 2002, p.104).

Nábělek et al. (1970) investigated the perceived pitch of pure tones that changed in frequency. These results further indicate the nature of the perceptual difference between a peak and plateau. Changes in frequency occurred either over the whole tone, or over only part of it with the changing section bordered by sections of steady frequency (plateaux). Subjects matched the pitch of each stimulus to the pitch of a steady tone of the same duration. In one experiment the frequency changed throughout the duration of the stimulus tone. When the tone was short, and when there was only a small frequency change, subjects matched the tone's pitch according to the mean of the initial and final frequencies. This was also the case for longer bursts and larger changes in frequency, although the perceived pitch was shifted towards the final frequency of the stimulus tone.

In a further experiment, the frequency changed over only 25% of the tone, beginning at seven different points. Therefore some tones immediately changed in frequency and were then followed by a steady state, whilst some frequency changes were preceded by a steady state, and others had a steady state both before and after the portion of the signal where the frequency was changing. When the frequency changed by 50 Hz subjects matched the pitch of the stimulus tone to that of its steady state regardless of whether the steady state occurred at the start or the end of the signal. When the frequency change was over 200 Hz the pitch was matched to either one or both of the terminal frequencies; that is to say that the tone could be perceived as having one of two different pitches.

Nábělek et al.'s (1970) experiments all involved pure tones, with rather a small number of subjects, but if the results can be generalized and extended to speech they make a number of predictions about the perceptual effect of a plateau in the intonation contour. Firstly, in a falling contour the pitch that is perceived should be that of the plateau if one is present. Secondly, the perceived pitch of a fall after a plateau will be higher than the perceived pitch of a fall that is *not* preceded by a plateau. This is because, in the latter case, the initial and final frequencies of the signal will be averaged in order to make a judgment of pitch.

The above review of the literature shows that there is often a perceptible difference between accents of different shape. If we know that this is the case, then one may wonder why, specifically, the perceptual distinction between a peak and plateau shaped accent matters. A first reason is that this type of detail will inform good speech synthesis, where we have to know about particular accent shapes in detail, rather than just that accent shape matters in general (see e.g. House et al. 1999). More importantly, because we have already gathered quite a large amount of data on the phonetic distribution of plateaux in SSBE, once we have details about how plateau are perceived we will be able to relate production and perception and begin to explain why these accents occur. In turn, this greater understanding will help us to further develop target and interpolation models of intonation.

Physiological constraints on the perception of intonation contours

The work discussed above indicates that differences in contour shape can influence perception, and that plateaux may sound higher in pitch than peaks of the same maximum frequency. These findings can be explained by examining the perceptual mechanisms by which the human ear perceives pitch. It is generally considered that there are two main perceptual mechanisms used by humans to extract pitch from a signal (see Moore 1997, especially chapter 5, for a review). Of these, the temporal mechanism is believed to be the main determiner of perceived pitch for speech sounds. When the basilar membrane vibrates in response to stimuli, a shearing motion is created between the basilar and tectorial membranes. This shearing motion displaces the stereocilia of the outer hair cells, which in turn causes the inner hair cells to fire and send information to the auditory nerve. Phase locking then occurs as the spikes in the activity of the auditory nerve occur regularly at the same phase of the stimulating waveform. Pitch is then extracted from the signal as integer multiples of the intervals between nerve firings.

However, the phase locking mechanism may not work well if the frequency of the signal is changing too rapidly. Sek and Moore (1999) investigated the discrimination of frequency steps linked by glides of various durations. Frequency discrimination was found to worsen as the glide duration increased and the authors suggest that discrimination is better when frequency changes less rapidly. Gockel et al. (2001) went on to demonstrate a similar effect in listener's matching of frequency modulated and unmodulated tones. The author's explain this effect in terms of auditory sluggishness. Pitch estimates from portions of the signal where F0 is changing rapidly receive less weight than estimates from sections where F0 is changing slowly. This idea is referred to as "stability-sensitive weighting" (2001, p. 702).

It seems likely that auditory sluggishness and stability-sensitive weighting can, to some extent, predict the effect of different shapes in an intonation contour. In accents containing sharp peaks, the fundamental frequency is changing rapidly throughout and therefore there will be no one steady portion of the signal which can be weighted more highly than others. For plateau stimuli, on the other hand, the highest frequency is sustained and the estimates from this section will be weighted more highly than the estimates from other sections of the accent. This extra weighting given to the highest frequency in the signal should result in a plateau sounding higher in pitch than a peak.

The present paper aims to further investigate the perceptual effect of plateaux found in nuclear falls in SSBE. Two perceptual experiments are reported where it is hypothesized, on the basis of earlier work and of physiological pitch perception mechanisms, that an accent containing a plateau will sound higher in pitch than an accent containing a peak of the same maximum frequency. Experiment 1 addresses this issue by asking subjects to judge which of two syllables is higher in pitch and comparing results when accents are realised as a peak or a plateau. In Experiment 2 the accents to be compared are located within the same utterance (instead of them being located in two separate utterances as in experiment 1), and subjects are additionally asked about perceived prominence.

EXPERIMENT 1

The first experiment reported is a preliminary test of the hypothesis that a plateau in the contour affects the perception of pitch height.

Stimuli

A female phonetician with a Standard Southern British accent recorded the utterance 'Anna came with Manny'. This utterance was chosen because it contains mainly sonorant segments, and therefore microprosodic variations in pitch should be minimal. The speaker was instructed to read the sentence in three different ways, which were indicated typographically by underlining particular words.

1) (Broad focus) Anna came with Manny.

2) (Narrow focus on Anna) <u>Anna</u> came with Manny.

3) (Narrow focus on Manny) Anna came with <u>Manny</u>.

Three repetitions of each version were recorded in a random order. The purpose of recording the three versions was so that the detailed pitch patterns of each could be identified and the appropriate values used to inform resynthesis. As the aim of the second experiment was to create stimuli that could sound as though either 'Anna' or 'Manny' were high in pitch or prominent, it was important to see how this effect was created in the natural utterances.

Resynthesis A version with narrow focus on 'Anna' was used for the template utterance as informal listening tests by several phoneticians indicated that this version sounded natural when F0 modifications were made on 'Manny'. Resynthesized versions of the utterance were created in Praat using PSOLA and stylized, straight line approximations (as shown in Figure). Twelve versions were created where the maximum frequency of the nuclear accent on 'Manny' was 160 Hz, 180 Hz, 200 Hz or 210 Hz. Accents were not created with a maximum frequency of 220 Hz because this would have made the frequency of 'Manny' as high as the frequency of 'Anna'. At each maximum frequency, accents were created shaped as a sharp peak, a 50 ms plateau and a 100 ms plateau. The maximum duration of the plateau was chosen by reference to that produced by the speaker when putting narrow focus on 'Manny'. Results from Knight (2003, chapter 2) also indicate that speakers often produce a plateau of this duration when the onset and coda of the nuclear syllable are sonorant. The 50 ms length was chosen to further investigate whether the duration of a plateau might be an important factor in perception.

The peak and the end of the plateau were positioned at the end of the first vowel in 'Manny'. That is to say that in plateau stimuli, the plateau was extended backwards in time (as shown in Figure 2). This decision was taken because a given frequency later in an utterance will sound higher in pitch than the same frequency earlier in the utterance (see e.g. Pierrehumbert, 1979) because listeners tend to compensate for the expected effect of F0 declination. Therefore, creating plateau stimuli by extending them forwards in time from the peak could make plateaux sound higher in pitch simply due to their position, rather than their shape, and the method was avoided for this reason.

One consequence of this method of stimuli creation (extending the plateau backwards) is that the peak occurs rather late in the syllable. Although the lateness of the peak leads to a rise over part of the syllable, the accent created is rather different to L*+H. In particular the high tone is never later than the end of the first vowel and the pitch begins to rise from the very start of the syllable rather than remaining low and then rising sharply towards the end. In addition, a number of trained phoneticians agreed that the created stimuli sounded most similar to an H*+L accent.

The rising section of contour started at the same time (at the beginning of /m/ in 'Manny') in every stimulus. This meant that the slope and duration of the rising section varied in these stimuli, being steeper and shorter in 100 ms plateau (100 ms) than 50 ms plateau (130 ms), and steeper and shorter in 50 ms plateau than in the peak stimuli (190 ms). The section of the syllable covered by the rise was also different, rising through the onset and nucleus for the peak and 50 ms plateaux, but only through the onset for the 100 ms plateaux. In addition, the slope was steeper for stimuli containing accents of a higher maximum frequency. The falling part of the accent always took place over 156 ms. The remainder of the utterance was left unaltered from the original.

FIGURE 2 ABOUT HERE

Pairs of resynthesized stimuli were created so that each stimulus was paired with every other stimulus of the same maximum frequency. This was done in two different orders, varying the order so that the stimulus was presented as both the first and the second member of the pair. Each stimulus was also paired with itself so in total there were 36 pairs created.

Informal listening exercises by several phoneticians indicated that it was rather difficult for the listener to concentrate on the accents on ‘Manny’, when they were separated by the high accent of ‘Anna’ in the second utterance. Therefore the portion of the signal containing ‘Anna’ was removed from each stimulus leaving only two versions of ‘… came with Manny’ in each pair. Stimuli within a pair were separated by a silent interval of 600 ms.

Subjects

Subjects were six staff and students at the University of Cambridge. They were five men and one woman. Their ages ranged from 19 to 50 (mean = 30). All subjects were monolingual speakers of British English with no reported history of speech and hearing disorders. Subjects were interviewed about their musical backgrounds. Although it seems that even people with congenital amusia (‘tone-deafness’) probably process intonation contours normally in natural speech (e.g. Ayotte et al. 2002, Foxton et al. 2004) it was possible that the more artificial nature of the present task might lead to different results from musicians and non-musicians. Therefore there was a mix of musical backgrounds in each group, including subjects with no history of musical involvement, and those who had achieved a high level of proficiency (at least grade seven on the English grading system of 1 to 8) on one or more instrument.

Set-up and instructions

Subjects took part in a two-alternative forced choice task. They were told that they would hear several sound files containing two versions of '… came with Manny'. They were told that the pitch of 'came with' would be the same in every case but that they should listen to the word 'Manny' and, for each pair, decide whether the pitch was higher in the first or the second sentence. They could listen to each pair as often as they liked and then pressed a button to indicate their choice. They were told to make a decision about each pair, even if they felt they were guessing. Subjects listened to sentences over headphones in a soundproof booth. They made their responses via a computer keyboard, and responses were recorded directly to the hard drive of a Silicon Graphics workstation.

Every subject heard each pair once, so there were 36 responses for each subject (9 ordered pairs X 4 F0 levels) totalling 216 responses. Before the experiment proper subjects took part in a practice session where they heard four pairs of stimuli with accents at maximum frequencies not used in the main experiment. The purpose of the practice session was to familiarize subjects with the format of the experiment, and to check that no one had difficulties judging pitch height. No subject reported difficulties in understanding what was required of them, even though they were not given an explicit definition of pitch height.

Analysis

Subjects indicated the stimulus in which 'Manny' sounded higher in pitch. Identical pairs are excluded from the following analysis as the primary interest is in accents with different shapes. Therefore there are 24 responses for each subject totalling 144 responses. Responses were converted to percentages indicating how often the longer stretch of plateau was chosen as higher in pitch. These figures were entered firstly into a one-sample t-test where they were tested against chance (50%). Secondly, in order to establish the effects of order, frequency and the identity of members within a stimulus pair they were entered as the dependent variable into a univariate MANOVA with independent within-subjects factors of *frequency* (4 levels; 160, 180, 200, 220), *stimulus pair* (3 levels; peak + 50 ms, 50 ms+100 ms and peak+100 ms) and *order* (2 levels; longer stretch first, longer stretch second).

Results

FIGURE 3 ABOUT HERE

FIGURE 4 ABOUT HERE

Firstly, results of the t-test show that subjects choose the longer stretch of contour as sounding higher in pitch 75% of the time ($t(143)=7.16$, $p<0.01$). Results from the MANOVA indicate that there is a significant effect of *frequency* ($F(3,15)=5.44$, $p=0.01$) as shown in Figure 8, but no significant main effects of *stimulus pair* ($F(2,10)=0.84$, $p>0.05$) or *order* ($F(1,5)=0.56$, $p>0.05$). However, there is a significant interaction between frequency and order ($F(3,15)=6.05$, $p<0.01$) as shown in Figure 9, and between frequency, stimulus pair, and order ($F(6,3)=3.94$, $p<0.01$).

Discussion for experiment 1

As hypothesized, a plateau in the contour does indeed lead to an accent being perceived as higher in pitch than a peak of the same maximum frequency. This is important, because, as explained above, it demonstrates a perceptual difference between two accent shapes that would be given the same description in a target and interpolation model of intonation. This finding leads on to the suggestion that more work in the area is necessary to further investigate the phenomenon and find out if such differences are meaningful in any way. These ideas are developed further by the second experiment and in the general discussion.

It is interesting to note that there is no main effect of *stimulus pair*. This indicates that subjects' choice of the longer stretch of plateau held regardless of the members of the pair. It also raises questions about the relationship between the two lengths of plateau. On one hand the lack of difference between the peak + 50 and peak + 100 pairs indicates that there is little difference between the perceived pitch of these plateau. This is supported by scores of 77% for the peak + 50 pair and 79% for the peak + 100 pair. In terms of stability-sensitive weighting it suggests that phase locking has already occurred sufficiently for the 50 ms plateau so the extra duration in 100 ms impacts little on the equation for calculating pitch. On the other hand, if there is no difference between the two lengths of plateau then we should expect pitch in the 50 + 100 pair to be harder to judge and for this stimulus pair to be significantly different to the others. This is not the case however, although subjects only chose the 100 ms plateau 71% of the time, perhaps indicating that this choice was harder than for other pairs. It seems likely that there is very little perceptual difference between the two durations of plateau when compared to a peak, but that when the two are placed side by side the longer plateau gains a slight advantage.

The final point to be discussed is the interaction between *frequency* and *order*. As can be seen in Figure 4 it appears as though the tendency to choose the longer plateau as higher in pitch is more pronounced at higher frequencies. However, the interaction with order demonstrates that the situation is rather more complicated. The previous generalisation holds when the longer plateau is in second position. However, the reverse is true when the longer stretch is in first position i.e. the tendency to choose the longer stretch is more pronounced at lower frequencies. It is possible that this effect is due to interaction with compensation for declination. Much work (e.g. Pierrehumbert 1979, Gussenhoven and Rietveld, 1988 Gussenhoven et al. 1997) has demonstrated the existence of the Gussenhoven-Rietveld effect, where subjects comparing accents of identical frequency are increasingly more likely to choose the second accent as higher in pitch as the frequency of both increases. This phenomenon is explained by Gussenhoven et al. (1998 p.3021) and Pierrehumbert (1979 p.368), as showing that listeners expect more declination in a wide pitch range (when the first peak is higher) and therefore compensate more for declination. This seems to be the effect found here when the plateau is in second position and is further enhanced by the fact that this position also contains the longer stretch of plateau.

When the plateau is in first position it seems that the shape to some extent overrides the Gussenhoven-Rietveld effect, and compensation for declination in general, as subjects are more prepared to choose the first accent (as it has a longer stretch of high pitch) at every frequency. As frequency increases the Gussenhoven-Rietveld effect becomes stronger, leading to less pronounced choices, but compensation for declination is still unable to override the effect of a longer plateau.

It is clear that the shape of an accent does affect how its pitch is perceived, and that, as predicted, a plateau shaped accent sounds higher in pitch than a peak of the same maximum frequency. A detailed discussion and explanation of this finding will appear in the general discussion, when data from the second experiment can also be brought to bear on the matter.

EXPERIMENT 2

The introductory experiments reported above show that the shape of an accent can affect the perception of pitch height. In particular, it seems that accents are perceived as higher in pitch when they contain a plateau rather than a sharp peak. The second experiment investigates whether this finding holds true when the accents to be compared are in the same utterance, and additionally whether the result can be extended to judgements of prominence. The research question asks whether a plateau-shaped accent not only sounds higher in pitch, but also sounds more prominent than an accent realized as a sharp peak.

It seems intuitively likely that a higher sounding accent will also sound more prominent. However, the relationship between perceived pitch and prominence is not necessarily a simple one, and it is not entirely clear that an accent which sounds higher in pitch will also sound more prominent. Terken (1991) presented listeners with seven syllables of reiterant stimuli with pitch accents on the second and penultimate syllables (maMAmamamaMAma). Subjects adjusted the fundamental frequency of the second accent so that it matched the first in terms of either pitch or prominence. Results showed that listeners adjusted the second accent to lower values if they were instructed to focus on prominence than if they were instructed to focus on pitch. Terken states that "perceived prominence is related in a complex way to the range of F0 values employed" (p. 1768), and that "listeners employ different strategies in judging prominence and pitch” (p. 1773). Therefore, although the results of Experiment 1 indicate that plateaux sound higher in pitch than peaks it is not necessarily the case that plateaux will also sound more prominent.

Stimuli

The sentence ‘Anna came with Manny’, recorded with narrow focus on ‘Anna’ for experiment 1, was again used as a template. Fourteen resynthesized versions were created using PSOLA in Praat, where the frequency of the accent on ‘Manny’ varied in seven linearly equal steps between 160 Hz and 220 Hz. 220 Hz had not been used as a maximum frequency in the first experiment but informal listening experiments indicated that it sounded natural, so was included here in order to ensure linear spacing of the maximum frequencies in different versions. At each of these seven frequencies, one stimulus was created where the shape of the accent was a sharp peak and another where the accent was a 100 ms plateau. The duration of the plateau was chosen as before (with reference to the natural utterance and Knight (2003)), although this time the shorter 50 ms plateau was not included as that issue was dealt with in experiment 1.

For the creation of plateau stimuli, the plateau began just after the onset of the stressed vowel in ‘Manny’ and ended just before the offset of the same vowel (as shown in Figure 5), mirroring its position when ‘Manny’ is in focus in natural speech. For the creation of the peak stimuli, the peak was located towards the end of the stressed vowel, and therefore at the same location as the end of the plateau in plateau stimuli. Although the speaker never produced a sharp peak when there was an accent on Manny in the natural utterances, it is preferable to have the peak aligned at the same position as the end of the plateau for reasons explained in the introduction to experiment 1 above. The rising portion of the accent always began at the start of the word ‘Manny’, which mirrored the pattern found in one of the natural utterances that had narrow focus on ‘Manny’. In the other two natural utterances there was no voicing at the end of the previous segment and the high frequency began at the start of the word ‘Manny’. The remainder of the utterance was unchanged from the original narrow focus utterance.

FIGURE 5 ABOUT HERE

Subjects

Twenty-four naïve speakers of British English served as subjects. All were students at the University of Cambridge. None were bilingual or reported any speech, hearing or language difficulty. Subjects were questioned about their musical history to ensure a mix of musical abilities in each group as this factor may affect judgements of pitch and prominence. Subjects were asked about their musical abilities before the test but it is unlikely that this biased them to respond to stimuli in a musical way as they were also asked about their abilities in foreign languages. The twelve who judged pitch height were six men and six women aged between 18 and 31 (mean=21.9). Seven were proficient musicians (having gained at least grade seven on one or more instrument), four had some musical training (but had obtained a level of proficiency lower than grade seven), whilst one had never studied an instrument. The twelve who judged prominence were three men and nine women aged between 20 and 25 (mean=21.3). Seven were proficient musicians, two had some training and three had no musical training.

Setup and Instructions

Subjects heard each of the fourteen resynthesized stimuli twice in a pseudo-random order so that no frequency or shape occurred in two successive presentations. They completed a two-alternative forced choice task, and for each stimulus compared the accents on 'Anna' and 'Manny'. Each group heard the stimuli in the same order, but one group was asked which accent sounded higher in pitch whilst the other group was asked which accent sounded more prominent. They could hear each stimulus as often as they liked, and they registered their choice by pressing one of two labelled buttons on a computer keyboard; 'a' for 'Anna' and 'm' for 'Manny'. Before the experiment proper, the subjects took part in a trial run using frequencies and shapes of contour not used in the main experiment. The purpose of the trial was to familiarize the subjects with the test procedure, and to ensure that they felt able to judge pitch and prominence. No subject experienced any difficulty making their judgement despite being given no explicit definition of pitch height or prominence.

Analysis

Each subject (24) heard each stimulus (14) twice resulting in 672 responses for analysis. Subjects responded by indicating whether 'Anna' or 'Manny' sounded either higher in pitch or more prominent. These responses were coded as 1 for a response for 'Anna' and 2 for a response for 'Manny'. Responses were entered as the dependent variable into a three factor, univariate mixed design MANOVA. Within subjects independent variables were maximum fundamental *frequency* of accent (7 levels; 160 Hz-220 Hz) and *shape* of accent (2 levels; peak or 100 ms plateau). Where Mauchly's test indicated that the assumption of sphericity was violated, a Greenhouse Geisser correction was used. The between groups variable was *task* as one group judged height and the other prominence. Comparisons for frequency are calculated automatically from the MANOVA data using estimated marginal means with a Bonferonni correction for multiple comparisons.

Results

FIGURE 6 ABOUT HERE

FIGURE 7ABOUT HERE

As figures 6 (for pitch height) and 7 (for prominence) show, there was no significant effect of *task* (F(1,22)=3.263, p>0.05), indicating that, in this task at least, pitch height and prominence are judged in the same way. Task did not interact with any other factor. Therefore the results from the two groups are analysed together.

Results indicate that, as expected, the maximum fundamental *frequency* of the accent on 'Manny' is a significant factor (F(3.182, 70) = 112.04, p<0.01). In general, subjects are more likely to choose 'Manny' as the higher, or most prominent, accent when the maximum frequency of the accent is higher. Comparisons indicate that there is no difference between responses to 160 Hz , 170 and 180 Hz accents, or between 200 Hz and 210 Hz, but all other frequencies are significantly different from each other. An examination of figure 14 indicates that 'Manny' begins to sound higher in pitch than 'Anna' at around 190 Hz for plateau stimuli and 200 Hz for peak stimuli. As the accent on ‘Anna’ had a maximum F0 of 206 Hz the listeners seem to be compensating for expected declination in each case.

Shape is, as hypothesized, also a significant factor ($F(1,22) = 41.875$, $p<0.01$). Subjects were more likely to chose ‘Manny’ if the accent associated with this word is a plateau rather than a peak. This indicates that the plateau in the contour does indeed make the accent sound both higher in pitch and more prominent. It can be seen, from figures 6 and 7, that a plateau in the contour makes an accent sound about 10 Hz higher than an equivalent peak in the middle frequencies. However there is a significant interaction between *frequency* and *shape* ($F(3.904, 85.894)=5.282$ $p<0.01$), which warrants discussion. An examination of the mean responses indicates that a plateau is always more likely to cause ‘Manny’ to be chosen as higher or more prominent except at the lowest frequency (160 Hz) when there is no difference between the two shapes. The difference between the two shapes becomes increasingly wide at 170 Hz, 180 Hz, 190 Hz and 200Hz, and then narrows again for the final two frequency steps (210 Hz and 220 Hz). It seems likely that at 160 Hz the frequency of ‘Manny’ is so low that it is difficult for accent shape to override the tendency to hear ‘Anna’ as higher in pitch.

Discussion for Experiment 2

The results from experiment 2 indicate that accent shape affects perceived pitch and prominence, with the plateau-shaped accent sounding higher in pitch and more prominent than a peak of the same maximum frequency. It seems that in the middle frequency ranges, a plateau is perceived similarly to a peak about 10 Hz higher, but there is less of a difference at the lowest frequency, where accent shape cannot override frequency. It is interesting to note that there was no significant difference in the way listeners responded when they were asked to judge pitch or prominence. This is in contrast to the results of Terken (1991), discussed above, which indicated that the two were judged in different ways. It seems likely that the differences in the stimuli may account for these different results. Firstly Terken used reiterant stimuli whereas the stimuli employed in the present experiment are resynthesized natural utterances. Secondly Terken acknowledges that the timing of some accents may have been somewhat unnatural whereas the timing of the accents in the present experiment was controlled to be a natural as possible. It is possible that the more natural task and stimuli in the current experiment contributed to the different results. In real listening conditions listeners may, therefore, judge pitch and prominence in a similar fashion.

GENERAL DISCUSSION

The experiments reported above investigated the perceptual effect of plateaux and peaks associated with nuclear falls in SSB English. Experiment 1 was an introductory experiment designed to see if there was any effect of accent shape on perceived pitch. Results showed that the shape of the accent did affect which stimulus sounded higher in pitch, and generally the accent containing the longer stretch of contour was chosen as higher. Frequency was also significant but interacted with order suggesting that compensation for declination was also an important factor, which can be overridden to some extent by shape. Experiment 2 extended the findings of the introductory experiment to judgments about both pitch height and prominence within a single utterance. The results showed that there is a significant effect of accent shape, with plateaux sounding higher and more prominent than sharp peaks at all frequencies except the lowest.

The results both confirm and extend results from earlier studies. They confirm that accent shape is perceptible by listeners, and secondly that a plateau is perceived as higher than a peak, as is suggested in the work of Nábělek et al. (1970). Furthermore the experiments reported here extend this finding to speech sounds, and to English, and demonstrate that accent shape also affects perceived prominence.

The finding that a plateau sounds higher in pitch than a peak can most likely be explained, as above, by the effect of stability-sensitive weighting as the pitch perception mechanism is more influenced by steady portions of the signal (Gockel et al. 2001). An alternative explanation may be that listeners pay attention to the area under the curve and judge greater areas to sound higher or more prominent. Segerup and Nolan (2005), for example, demonstrate a cue-trading relationship between the timing and height of a fall in Gothenburg Swedish. They show that for the area under the curve within the stressed vowel, the 'pitch integral', seems to be a good model for speaker's judgements of the word accent distinction in this dialect. For the present experiment, however, there is perhaps some additional support for stability-sensitive weighting, rather than for the pitch integral, in the comparisons between 100 ms and 50 ms plateau in the first experiment. The results suggest these two accent shapes are not perceived differently, although the area under the curve of a 100 ms plateau would be greater than that of a 50 ms plateau. In terms of stability-sensitive weighting, phase-locking to a 50 ms portion of steady frequency already greatly influences the perception of pitch, so little is added when the plateau is extended.

The general results further suggest that, as phoneticians and phonologists, we may need to be more aware of the effect of accent shape in our transcription practices when labelling tonal targets, and the implications this has for theory. For example, when a plateau occurs in an intonation contour, we must be aware that this will create a different perceptual effect from a sharp peak, even when the maximum F0 is the same in each case. However, it is unclear to what extent accent shape can affect meaning. In this experiment subjects were only asked about height and prominence, neither of which is a linguistic task. It would be useful to perform an experiment similar to that of Lieberman and Pierrehumbert (1984), where listeners choose if 'Anna' or 'Manny' is new information. Unfortunately this is impossible in British English as listeners will interpret 'Manny' as new information when it is associated with any accent, however low (as demonstrated by informal pilots for Experiment 2 above). It is therefore impossible to decide from the present results whether linguistic meaning can be altered by accent shape but it seems that paralinguistic meaning, in terms of degree of emphasis, can be affected.

The issue of how to label a high plateau in turn raises further theoretical questions about the relationship between tones and turning points. If a plateau is seen simply as a differently shaped high tone then, for the purposes of alignment studies, the question still remains about where within a plateau the underlying high tone should be marked. A definitive answer to this question is beyond the scope of the current paper. However there is some evidence (Knight 2003) that the end of the plateau, rather than the beginning, is more stably aligned in relation to the syllable and the foot, so may perhaps be a more likely candidate for the location of the intended high tone. This is further supported by the work of D'Imperio (2002) discussed above, which indicates that perceptual results are more consistent if responses to plateau stimuli are compared to peak stimuli where the peak is aligned with the end of a plateau.

Related to these questions is the issue of how much choice and control a speaker has over the shape, and other aspects, of the intonation contour they produce. A strong hypothesis would suggest that the speaker has no control of the accent shape that they produce, only some control over the type of accent (such as a rise rather than a fall) and gross aspects of pitch height and span. This would suggest that a speaker only produces plateau accents due to the effects of segmental and prosodic structure. There is some evidence for this point of view in House et al. (1999) and Knight (2004) where it seems that plateaux are longer when more voiced, sonorant material is available over which an accent can be realized. However, the very fact that plateaux are found in these optimal conditions, when nothing disturbs the intonation contour, may suggest that a more plateau-shaped accent is always the speaker's intended accent shape, but that this intention is usually disrupted by sections of voicelessness and other microprosodic perturbations. Under this view the speaker still has little choice about the shape because of physiological constraints, but produces a plateau when possible. A definitive answer to this question must await a more detailed understanding of how intended targets are implemented phonetically.

An alternative argument would suggest that a speaker has a greater degree of choice over the accent shape that they produce, and therefore accent shape could be manipulated in order to influence the listener's perception of the message. The fact that a plateau makes an accent sound higher might suggest that a plateau could be used as an alternative to a higher peak in order to draw the listener's attention to a particular syllable. There is some evidence for this argument in earlier work describing plateaux. Wichmann et al. (1999) demonstrate that plateaux occur in paragraph-initial position; a position normally associated with high frequency, presumably so that the speaker can draw the listener's attention to a change in topic. The work of House et al. (1999) suggests that plateaux occur in nuclear position, where again a speaker may wish to focus the listener's attention on a particular syllable. Taken together these findings might suggest that a plateau-shaped accent could be produced as an alternative to a peak-shaped accent with a higher F0.

Gussenhoven (2002) notes that one phonetic variable can be used as a substitute for another under the umbrella of various biological codes. Ohala (e.g. 1994) proposes a "frequency code" which states that higher frequencies are associated with questions in human communication. Gussenhoven (2002) proposes two further codes; the "effort code", which states that high frequency suggests the listener should pay attention because of the speaker's added effort, and the "production phase" code, which states that high frequency is associated with the beginning of phrases when subglottal pressure is high.

Gussenhoven further suggests that speakers can create the perceptual effect of high pitch without physically producing a higher F0. For example, delayed peaks will sound higher in pitch than peaks that are not delayed, as the listener compensates for expected declination. He further suggests that delayed peaks are used by speakers as substitute variables for higher peaks, and are indeed found in the same environments as higher peaks. For example both higher and delayed peaks are found paragraph-initially (production code), in questions (frequency code) and in narrow focus (effort code).

It is possible that a plateau can also be used as a substitute variable for high F0 as it is found in nuclear and paragraph-initial positions. A speaker could use their implicit knowledge about stability-sensitive weighting in order to create an accent with the same perceptual effect as a higher frequency peak. At this stage this suggestion is speculative, but further work is planned to test it by looking at production in a more natural setting (using the IViE corpus (http://www.phon.ox.ac.uk/~esther/ivyweb/)) to further investigate where plateaux occur.

CONCLUSIONS

The experiments reported above have shown that a plateau in an intonation contour sounds both higher in pitch and more prominent than a peak with the same maximum fundamental frequency. For these stimuli, no significant difference exists between judgements involving pitch and prominence. The finding that peaks and plateaux are perceived differently is in line with earlier studies with non-speech stimuli and the effect can be attributed to stability-sensitive weighting, whereby phase locking in the auditory system works more efficiently when a frequency is sustained. It is suggested that the differences in shape between accents must be taken into account when transcribing intonation contours in terms of high and low pitch targets. Finally, it is suggested that plateaux may be used by speakers as substitutes for higher peaks.

REFERENCES

ARVANITI, A., LADD, D. R. & MENNEN, I. (1998). Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics* ,26,3-25.

AYOTTE, J., PERETZ, I. & HYDE, K. (2002) Congenital amusia. A group study of adults afflicted with a music-specific disorder. *Brain, ,125*, 238-251.

DILLEY, L. (2005). The *Phonetics and Phonology of Tonal Systems*, Unpublished PhD dissertation, Massachusetts Institute of Technology.

DILLEY, L. and M. Brown. (In press). Effects of pitch range variation on F0 extrema in an imitation task. *Journal of Phonetics*.

D'IMPERIO, M. (2000). *The role of perception in defining tonal targets and their alignment*. Unpublished PhD thesis, The Ohio State University.

D'IMPERIO, M. (2002). Language Specific and Universal Constraints on Tonal Alignment: The nature of Targets and "Anchors". In B. Bel, & I. Marlien, (Eds.) *Proceedings of the 1st International Conference on Speech Prosody* (pp.101-106). Aix-en-Provence, France.

D'IMPERIO, M. & HOUSE, D. (1997). Perception of questions and statements in Neapolitan Italian. *Proceedings of Eurospeech '97,* 1, 251-254, Rhodes.

FOXTON, J.M., DEAN, J.L., GEE, R., PERETZ, I. & GRIFFITHS, T.D. (2004). Characterization of deficits in pitch perception underlying 'tone deafness'. *Brain, 127*, 801-810.

GOCKEL, H., MOORE, B., & CARLYON, R. (2001). Influence of rate of change of frequency on the overall pitch of frequency-modulated tones. *Journal of the Acoustical Society of America*, 109, 2, 701-712.

GOSY, M. & TERKEN, J. (1994). "Question marking in Hungarian: timing and height of pitch peaks." *Journal of the Acoustical Society of America*, 22, 269-281.

GUSSENHOVEN, C. (2002). Intonation and Interpretation: Phonetics and Phonology. In B. Bel, & I. Marlien (eds.) *Proceedings of the 1st International Conference on Speech Prosody* (pp. 47-57). Aix-en-Provence, France.

GUSSENHOVEN, C., B.H. REPP, A. RIETVELD, W.H. RUMP & J. TERKEN (1997). The perceptual prominence of fundamental frequency peaks. *Journal of the Acoustical Society of America 102*, 3009-3022.

GUSSENHOVEN, C. & RIETVELD, A. (1988). Fundamental frequency declination in Dutch: testing three hypotheses. *Journal of Phonetics* 16, 355-369.

't HART, J. (1981). Differential sensitivity to pitch distance, particularly in speech. *Journal of the Acoustical Society of America* **69**, 811—821.

't HART, J. (1991). F0 stylization in speech: Straight lines versus parabolas. Journal *of the Acoustical Society of America* 90(6), 3368-3371.

HOUSE, J. DANKOVIČOVÁ, J. & HUCKVALE. M. (1999). Intonational Modelling in ProSynth. Poster presented at *XIVth International Congress of Phonetic Sciences, University of California, Berkeley, CA.*

HOUSE, J. & WICHMANN, A. (1996). Investigating peak timing in naturally-occurring speech: from segmental constraints to discourse structure. *Speech, Hearing and Language: work in progress,* 9, 99-117. University College London.

KNIGHT, R-A. (2003). *Peaks and Plateaux: The production and perception of intonational high targets in English*. Unpublished PhD thesis, University of Cambridge, UK.

KNIGHT, R-A. (2004). The realisation of intonational plateaux – Effects of foot structure. In L. Astruc and M. Richards (eds.) *Cambridge Occasional Papers in Linguistics,* **1**, (pp. 157-164), Cambridge: Department of Linguistics, University of Cambridge.

KNIGHT, R-A. (2000) Prosynth report, Unpublished manuscript, University College London.

KNIGHT, R-A. & NOLAN, F. (2006). The effect of pitch span on intonational plateaux. *Journal of the International Phonetics Association*, 36, 1, 1-28.

LIEBERMAN, M. AND PIERREHUMBERT, J. (1984). Intonational Invariance under changes of pitch range and length. In ARNOFF, A. AND R. OEHRLE *Language Sound Structure*, MIT Press, Cambridge, Massachusetts, 157-233.

MOORE, B. (1997). *An introduction to the psychology of hearing*. San Diego: Academic Press (4th edition).

NÁBĚLEK, I. V., NÁBĚLEK A., K. & HIRSH, I. J. (1970). Pitch of Tone Bursts of Changing Frequency. *Journal of the Acoustical Society of America*, 48(2), 536-553.

NIEBUHR, O. (2003). Perceptual study of timing variables in F0 peaks. *International Congress of Phonetic Sciences*, Barcelona, 1225-1228.

OHALA, J. (1994). The frequency code underlies the sound-symbolic use of voice pitch. In L. Hinton, J. Nichols, & J. Ohala. *Sound Symbolism* (pp. 325-347). Cambridge: Cambridge University Press.

PIERREHUMBERT, J. (1979). The perception of fundamental freqeuncy declination. *Journal of the Acoustical Society of* America, 66(2), 363-369.

PIERREHUMBERT, J. (1980). The *Phonology and Phonetics of English Intonation*. Michigan: MIT Press.

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). "La perception des glissandos descendants dans les contours prosodiques." *Phonetica* 35: 11-40.

SCHEPMAN, A., LICKLEY, R.J. & LADD, D.R. (2006). Effects of vowel length and right context on the Alignment of Dutch nuclear accents. *Journal of Phonetics*, 34,1-26.

SEGERUP M. & NOLAN F. (2006). Gothenburg Swedish word accents: a case of cue trading? In: G. Bruce & M. Horne (eds.), Nordic Prosody: Proceedings of the IXth Conference, 225-233.

SEK, A. & MOORE, B. (1999). Discrimination of frequency steps linked by glides of various durations. *Journal of the Acoustical Society of America,* 106, 1, 352-359.

STUDDERT-KENNEDY, M.& HADDING, K. (1973). Auditory and Linguistic Processes in the Perception of Intonation Contours. *Language and Speech* 16(4), 293-313.

SILVERMAN, K. & PIERREHUMBERT, J. (1990). The timing of prenuclear high accents in English. In J. Kingston & M. Beckman. (eds.), *Papers in Laboratory Phonology I: Between the grammar and physics of speech* (pp. 72-106). Cambridge: Cambridge University Press.

TERKEN, J. (1991). Fundamental frequency and perceived prominence of accented syllables. *Journal of the Acoustical Society of America,* 89, 4, 1768- 1776.

WICHMANN, A. HOUSE, J. & RIETVELD, T. (1999). Discourse constraints on peak timing in English: Experimental Evidence. *Proceedings of. XIVth International Congress of Phonetic Sciences, San Francisco,* 1765-1768.

Figure Titles

Figure 1 An example of a high plateau in the intonation contour for the phrase 'Anna came with Manny', with narrow focus on 'Manny'

Figure 2 Schematic representation of 'Manny' accents in experiment 1. The solid line represents peak stimuli, the dashed line shows 50 ms plateau stimuli, and the dotted line shows 100 ms plateau stimuli.

Figure 3 Responses indicating the longer stretch of plateau is higher, pooled across orders and stimulus pairs.

Figure 4 Responses indicating the longer stretch of plateau is higher, pooled across stimulus pairs.

Figure 5 Schematic representation of the two accents used in Experiment 2, showing alignment with segmental landmarks. In the second accent the dotted line represents plateau stimuli and the solid line represents peak stimuli.

Figure 6 % of responses for the second accent, at each frequency, for subjects who judged pitch height

Figure 7 % of responses for the second accent, at each frequency, for subjects who judged prominence

Figures

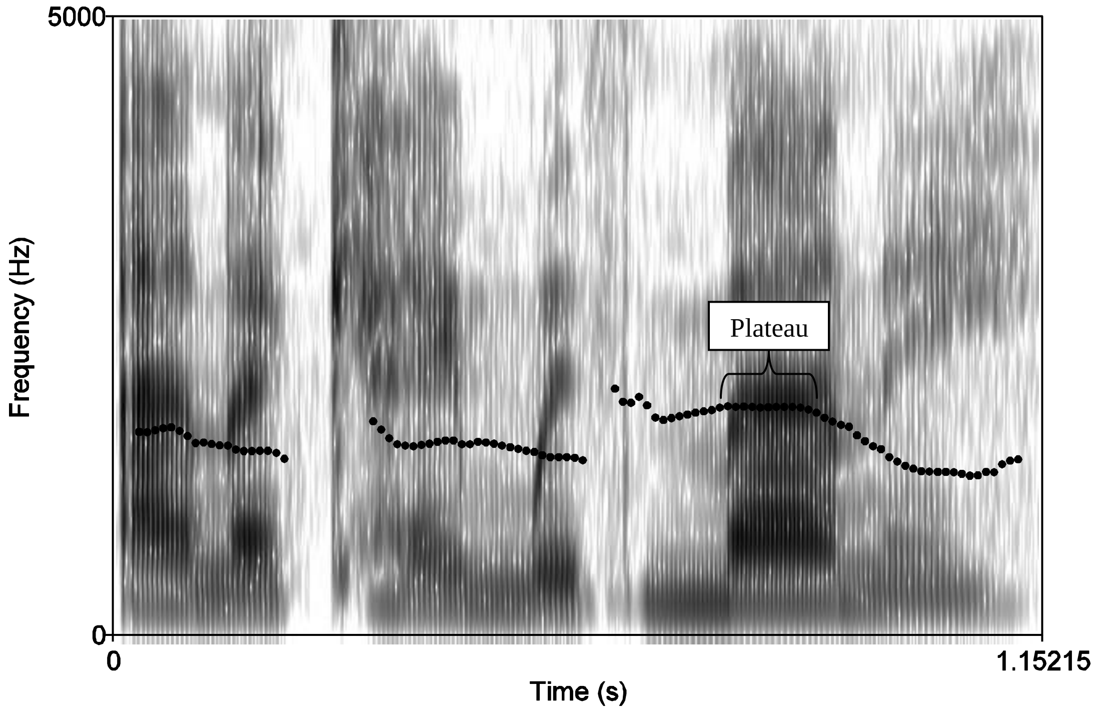

Figure 1 An example of a high plateau in the intonation contour for the phrase 'Anna came with Manny', with narrow focus on 'Manny'

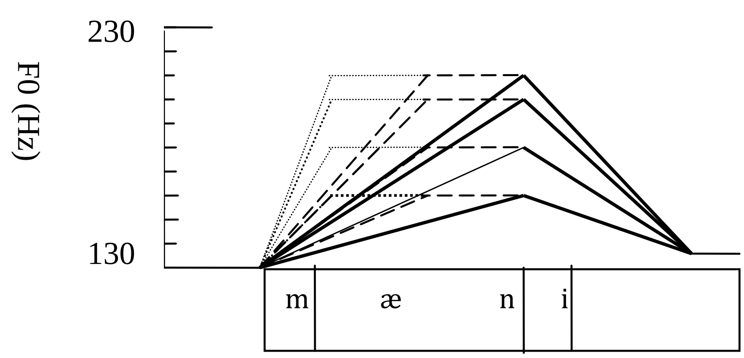

Figure 2 Schematic representation of 'Manny' accents in experiment 1. The solid line represents peak stimuli, the dashed line shows 50 ms plateau stimuli, and the dotted line shows 100 ms plateau stimuli.

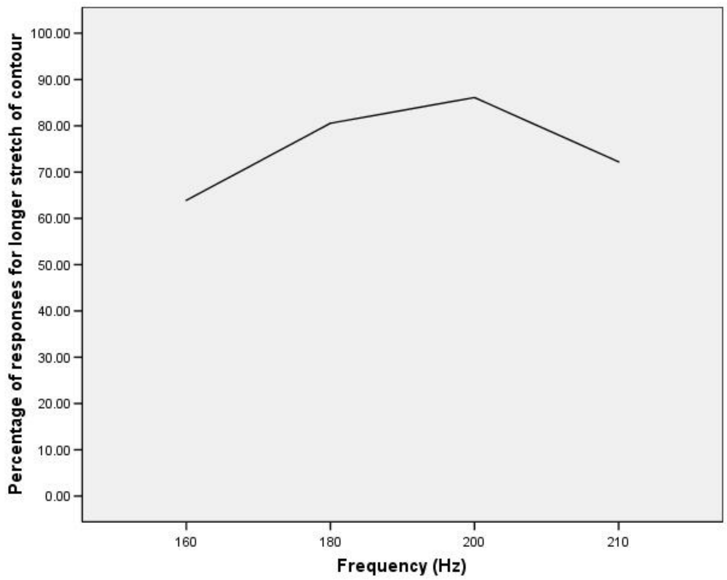

Figure 3 Responses indicating the longer stretch of plateau is higher, pooled across orders and stimulus pairs.

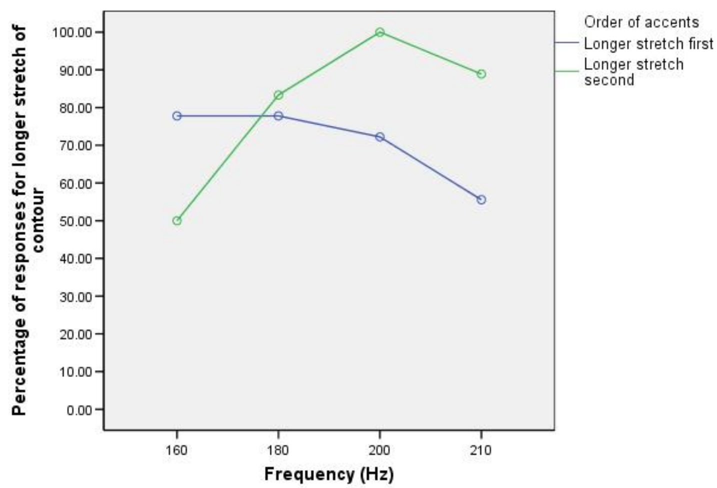

Figure 4Responses indicating the longer stretch of plateau is higher, pooled across stimulus pairs.

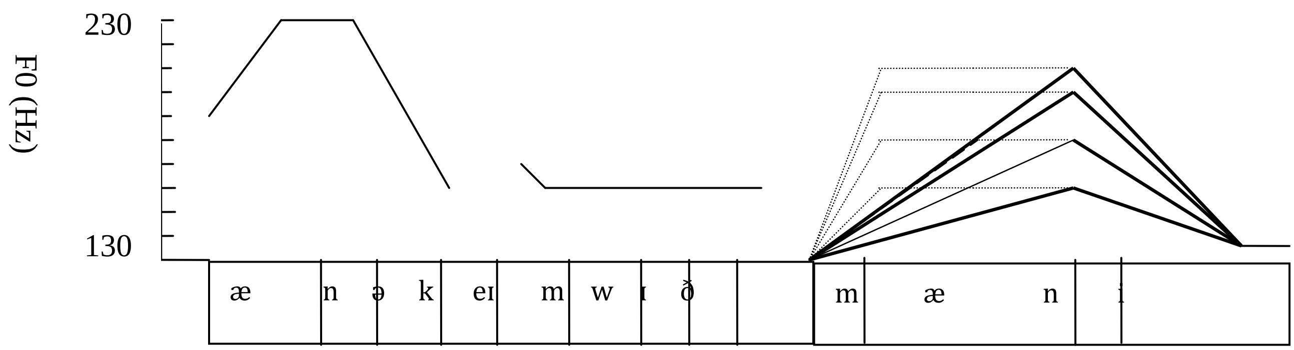

Figure 5 Schematic representation of the two accents used in Experiment 2, showing alignment with segmental landmarks. In the second accent the dotted line represents plateau stimuli and the solid line represents peak stimuli.

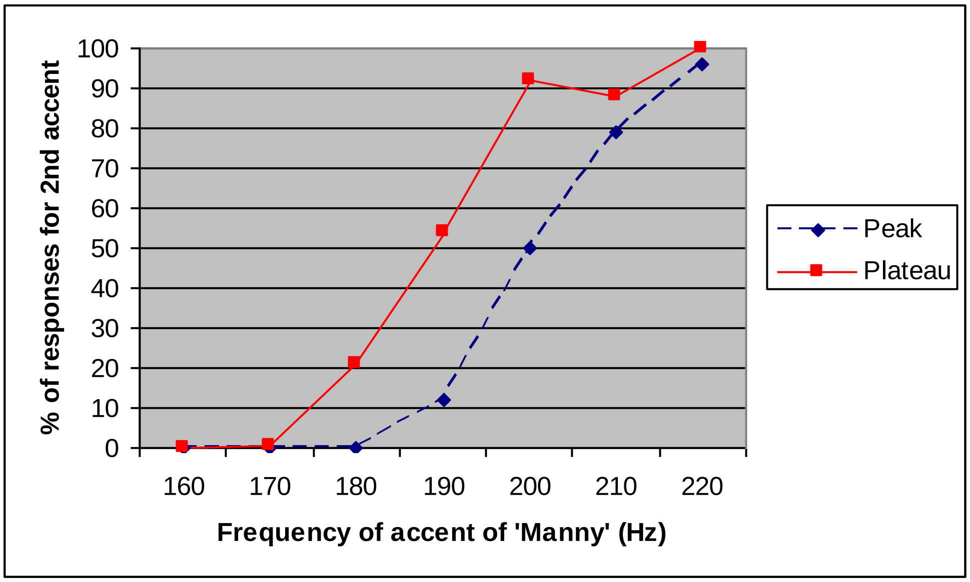

Figure 6 % of responses for the second accent, at each frequency, for subjects who judged pitch height

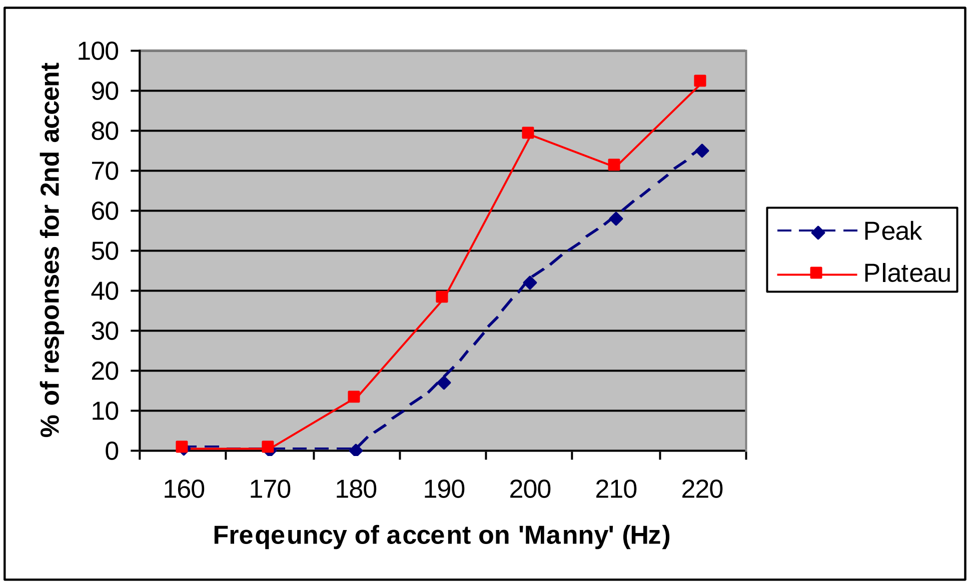

Figure 7 % of responses for the second accent, at each frequency, for subjects who judged prominence