TIMING OF PITCH MOVEMENTS AND ACCENTUATION OF SYLLABLES

Dik J. Hermes

Institute for Perception Research
P.O. Box 513
NL 5600 MB Eindhoven, The Netherlands
Email: hermes@natlab.research.philips.com

ABSTRACT

In this study, the relation is investigated between the timing of a rising or falling pitch movement and the syllable it accentuates. The experiments were carried out with the five-syllable utterance /mamamamama/ provided with a rising or falling pitch movement. The timing of the pitch movement was systematically varied and subjects were asked to indicate which syllable they perceived as accented. In order to find out where in the pitch movement the cue which induces the percept of accentuation is located, the duration of the pitch movement was varied.

The results show that the percept of accentuation is induced by a change in pitch at the start of the movement. The moment at which the course of pitch has changed significantly determines which syllable is perceived as accented. If this moment lies some tens of ms before the vowel onset, the syllable preceding this vowel onset is perceived as accented. For a rise, a high accent is perceived, for a fall a low accent. If the pitch change occurs after this moment, the syllable which contains this vowel onset is perceived as accented. For the rise, a low accent is then perceived, for the fall a high accent.

1. INTRODUCTION

In various descriptions of intonation, pitch movements are divided into different categories on the basis of their position in the syllable. For example, in the description of Dutch intonation [14], two kinds of accent-lending rises are distinguished: an early rise '1' starting before the vowel onset and a late rise '3' starting after the vowel onset. In autosegmental phonology, realizations of this type of pitch movements correspond with L+H* and L*+H, respectively [12]. In this paper, the early rise will be referred to as a rise inducing a high accent, while the late rise will be referred to as a rise inducing a low accent.

For the accent-lending fall, in the description of Dutch intonation by 't Hart et al. (1990) [14], only one phonetic category of full-sized, accent-lending falls is given: the 'A'. Rietveld and Gussenhoven (1995) [13], however, distinguish two phonological categories, H*L and $^!$H*L. In the autosegmental phonology of English intonation ([12, 2]) two bi-tonal falls are distinguished, H*+L and H+L*. In this paper, the early fall will be referred to as a fall inducing a low accent, while the late fall will be referred to as a fall inducing a high accent.

As the timing of a pitch movement in an utterance shifts from left to right, there will be a point where the percept of accentuation shifts from one syllable to the next. This point in the speech signal is referred to as the *accentuation boundary*. There will also be a point where, for the rise, the percept of a high accent changes to that of a low accent and, for the fall, the percept of a low accent changes to that of a high accent. This point is referred to as *high-low boundary*.

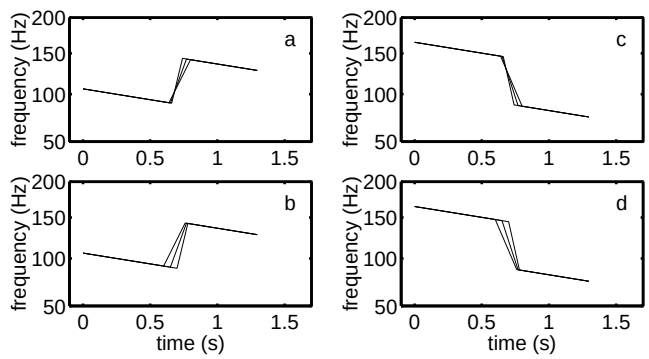

Figure 1: Three pitch movements of different durations at the accentuation boundary between two syllables. In (a) and (c) the situation is depicted if the cue for accentuation is at the onset of the pitch movement. In this case, the onset of the pitch movements at the accentuation boundary shifts at most to the left as the duration of the movement gets longer, while the offsets shift to the right. In (b) and (d) the situation is depicted if the cue for accentuation is at the offset of the pitch movement. Here, both the onset and the offset of the pitch movements at the accentuation boundary shift to the left as the duration of the movement gets longer, the onset much more so than the offset.

The first experiment reported in this paper is concerned with the accentuation boundary between two syllables. This boundary was determined by varying the timing of the pitch movement within an utterance and asking subjects which syllable they perceived as accented. Moreover, this was done for three different durations of the accent-lending pitch movement. The aim was to find out where in pitch movement the cue is located which accents the syllable, i.e., at

the beginning, at its end or somewhere in between. The basic idea is that, if the pitch cue which induces the percept of accentuation is at the *onset* of a pitch movement, the location of the *onset* of the pitch movement at the accentuation boundary will be more or less independent of the duration of the pitch movement. The perceptual process which determines where the cue is located will, however, have an integration time, and this integration time may be a bit longer for the weaker stimuli. It is supposed that the onset of the longer-duration pitch movements is less abrupt and, therefore, weaker than that of the shorter pitch movements. The onset of the pitch movements at the accentuation boundary may, therefore, shift somewhat to the left for the longer pitch movements, but not to the right. The offset of the pitch movement at the accentuation boundary will accordingly shift to the right as the duration of the pitch movement gets longer. This situation is depicted in Fig. 1a for the rise, and in Fig. 1b for the fall. If, on the other hand, the cue which accentuates the syllable is at the *offset* of the pitch movement, the *offset* of the pitch movement at the accentuation boundary will be more or less independent of the duration of the pitch movement. If the perceptual process which determines the moment at which the pitch movement has been completed has an integration time, the shift of the offset of the pitch movement at the accentuation boundary will at most shift somewhat to the left as the pitch movement gets longer, but not to the right. The onset of the pitch movement at the accentuation boundary will accordingly shift to the left, the longer the duration of the pitch movement. This situation is sketched in Fig. 1c for the rise, and in Fig. 1d for the fall.

The same procedure was followed in the second experiment concerned with the high-low boundary. First, what is the position of a pitch movement when it provides a syllable with a high or low accent, and second, where in the pitch movement the cue is located which induces these percepts.

2. EXPERIMENT I

2.1. Method

Stimuli The stimuli were derived from a natural three-syllable utterance /mamama/, spoken with an accent on the second syllable. Speech modifications were carried out with pitch-synchronous-overlap-add (PSOLA) techniques. The first step was to triplicate the middle syllable, resulting in the utterance /mamamamama/ consisting of five syllables, of which the middle three were identical as to amplitude and spectral envelope. Next, the original pitch contour was replaced by a rising or a falling pitch movement superimposed on a declination line. The declination rate was about 24 Hz/s. The range of the timings of the pitch movements used in this experiment is shown in Fig. 3a and Fig. 3d for the pitch movements with duration of 120 ms. The two other durations used were 80 and 160 ms. In total there were 360 different stimuli, which were presented twice to each subject in the same random order.

Subjects and Procedure Eleven subjects took part in this experiment. Each subject did the test individually in an interactive computer session. He/she could listen to a stimulus as often as desired. The task was to indicate on which of the five syllables an accent was perceived. So, there were five response classes corresponding to whether an accent was perceived on the first, the second, the third, the fourth or the fifth syllable.

2.2. Results and Discussion

For all five response classes, distributions were calculated for the onsets of the corresponding pitch movements. The distributions are presented in Fig. 2. The upper two panels show the results for the stimuli with the pitch movement of 80 ms. The second row shows the results for the movements of 120 ms, and the third row for the movements of 160 ms. In all six panels, the five distributions can clearly be distinguished. Clear intervals can be seen where the corresponding syllable is almost always selected as the accented syllable. These regions are indicated by the numbers 1 to 5.

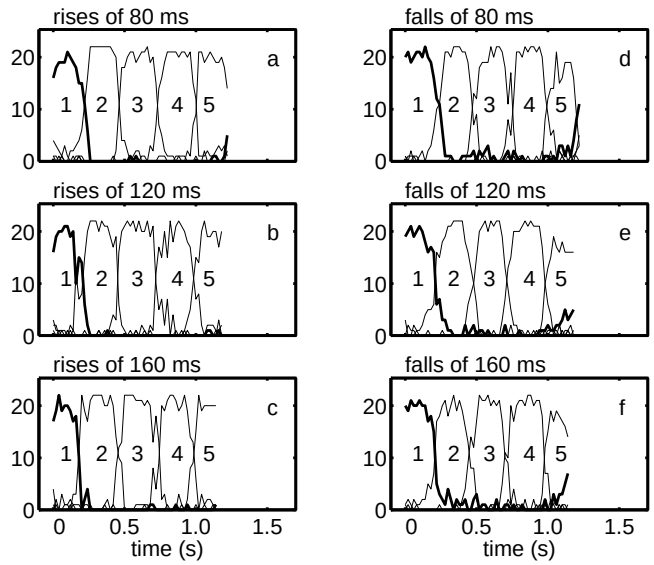

Figure 2: Frequency distributions of the onset times of the pitch movements accenting one of the five syllables. The numbers in the plots indicate the syllable which, according to these distributions, is most often indicated as accented. The thick continuous line is the distribution of the onset times of the pitch movements accenting the first syllable.

The points of interest are the locations in the speech signal where the percept of accentuation changes from one syllable to the next. These are the category boundaries in the continuum of onset times before which a pitch movement accents one syllable and after which it accents the next. These are the moments in the speech signal where the distributions as shown in Fig. 2 cross. These cross points were calculated after smoothing the distributions. The pitch movements at these accentuation boundaries are presented in Fig. 3b and Fig. 3e. As can be seen the situation at the accentuation boundaries is as depicted in Fig. 1a for the rise and Fig. 1c for the fall. This shows that *the cue of accentuation is at the onset of the pitch movement*.

The relation with the segmental structure can be derived from the lowest panel representing the oscillogram of the stimulus monotonized at 100 Hz. It appears that the accentuation boundaries are loc-

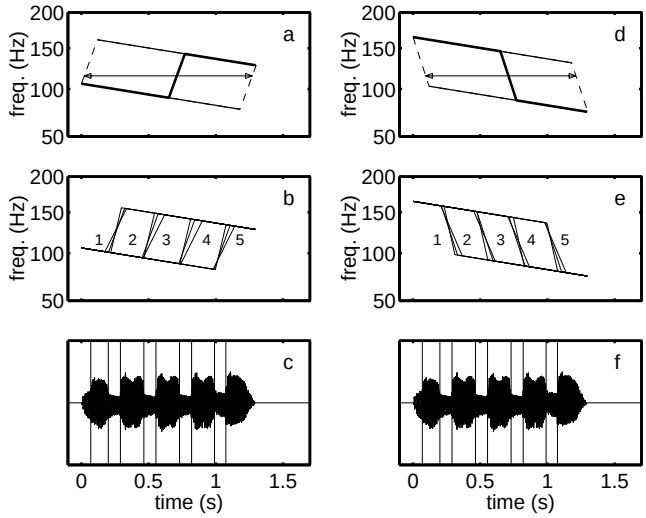

Figure 3: Stimulus range (a) and (b), pitch contours at accentuation boundaries (c) and (d), and oscillogram (e) and (f) of the stimulus monotonized at 100 Hz.

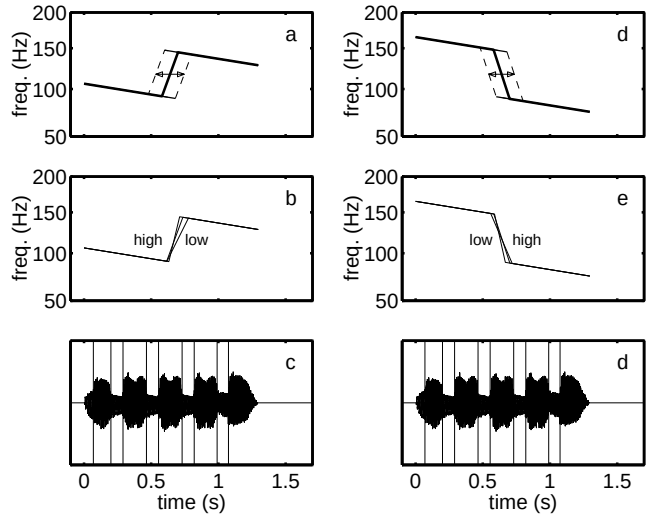

Figure 4: Stimulus range (a) and (b), pitch contours at high-low boundary of the third syllable (c) and (d) and stimulus oscillogram (e) and (f). Only a limited range of stimuli is presented in the range where the third syllable is accented.

ated at the end of the vowel /a/, somewhere in the neighborhood of the syllable onset (cf. [5, 16].

3. EXPERIMENT II

In this experiment, the main point of interest is when a high and when a low accent is perceived. Subjects were asked to indicate, whether they perceived a high or a low accent. Some informal tests showed that, in general, subjects could do this task easily, although perhaps not as easily as the task to indicate which syllable they perceived as accented.

3.1. Method

Stimuli The range of the timing of the pitch movements was restricted to where the third syllable was accented, 480 ms to 680 ms. This range is shown for the 120 ms rise in Fig. 4a and for the 120 ms fall in Fig. 4d. Again, there were rises and falls of three durations: 80, 120 and 160 ms. In total there were 66 stimuli.

Subjects and Procedure The same eleven subjects took part as in the previous experiment. Their task now was to indicate whether they heard a high or a low accent. The series of 66 stimuli was presented twice in the same random order.

3.2. Results and Discussion

The pitch movements at the high-low boundary are presented in Fig. 4b and Fig. 4e. The picture is again as shown in Fig. 1a and Fig. 1c. So, here too the location of the onset of the movement determines whether a high or low accent is perceived. It appears that the high-low boundary of the rise is located 20 ms after the vowel onset, that of the fall 65 ms after the vowel onset.

4. GENERAL DISCUSSION

The main conclusion of these experiments is that both the cue that determines which syllable is perceived as accented and the cue indicating whether a low or a high accent is perceived are located at the onset of the pitch movement. In all the results, the onset of the pitch movement at the category boundaries remained about constant, or shifted somewhat to the left as the pitch movement became longer. The offsets of the pitch movement at both the accentuation boundary and the high-low boundary shifted to the right as the pitch movements got longer. This situation was sketched in Fig. 1a for the rise and Fig. 1c for the fall. This virtually excludes the possibility that the location of the offset of the pitch movement provides the cue for accentuation and the high-low distinction between accents.

It is interesting to compare these results with results obtained by House [10]. He found that our sensitivity for pitch changes is reduced in speech segments of significant spectral change. The vowel onset is one such segment. So, it is assumed that some tens of ms before the vowel onset our sensitivity to pitch change is reduced. For early rises and falls, the onset of the movement is located in this interval of reduced sensitivity. Therefore, not this change of pitch at the onset of the movement will be perceived, but rather, after the vowel onset, a jump will be perceived to a pitch level different from what is expected on the basis of only declination. So it will be this jump which induces the percept of accentuation. Still according to House (1990), if the change in the course of pitch occurs after the vowel onset, so after the interval of reduced sensitivity to pitch change, listeners naturally perceive the change itself which then provides them with the cue for accentuation.

Only a few other studies have been carried out on the accentuation boundaries between two syllables. The oldest one is the study by Van Katwijk and Govaert (1967) [15], who also varied the posi-

tion of a rising or a falling pitch movement. Their main conclusions were, first, that the rise was more effective in accenting a syllable than a fall and, second, that falls exert their prominence lending force later than rises. These conclusions do not quite agree with what is found here, but it can be shown that this is due to differences in the stimuli used by Van Katwijk and Govaert [8, 4]. Also differences with results found by Hasegawa and Hata [7] can be attributed to differences in the stimuli. The results found here do agree with those of some other studies [3, 9], and they are at least in qualitative agreement with psychophysical findings [1].

In the experiments carried out so far, only separate rises or separate falls are taken into account. In many languages, accent is lent by a combination of a rise and a fall, the rise-fall. At least in Dutch intonation, the location of these rise-falls is such that the rise starts about 50 ms before the vowel onset, while the fall starts about 80 ms after the vowel onset. In terms of the results of this paper, this shows that these rise-falls supply two cues for accentuation, both lending a high accent to the syllable. This may explain why the rise-fall is an effective cue for accentuation. In other languages, e.g. Hungarian [6] and German [11], the situation may be more complicated.

4.1. Conclusions

Combining all the evidence, it appears that the moment at which the change in the course of pitch at the onset of a pitch movement is perceived determines which syllable is perceived as accented. If this moment occurs some tens of ms before the vowel onset of the syllable, the syllable preceding the syllable with this vowel onset is perceived as accented. When it comes later, the syllable with this vowel onset itself is perceived as accented. If the change in the course of pitch at the onset of the pitch movement occurs before or at the vowel onset, a high accent is perceived for the rise and a low accent for the fall. If the change in the course of pitch at the onset of the pitch movement occurs after the vowel onset, a low accent is perceived for the rise and a high accent for the fall. (Later experiments have yielded some evidence that not the vowel onset, but its presumed perceptual correlate, the perceptual moment of occurrence of the syllable, or so-called P-center, is a better candidate for this anchor point.)

Combined with the results obtained by House (1990), it appears that the perception of a pitch accent can arise from two different mechanisms: first, the perception of a *change in the course of pitch* can induce an accent, and second, an accent can be brought about by a *pitch level jump* between syllable nuclei. The perception of a pitch change gives rise to the perception of an accent on the syllable with the rhyme in which the pitch change is perceived. For a rise this induces a low accent, while for the fall it is high. If a pitch jump is perceived, the latter of the two syllables between which the jump takes place is perceived as accented. In this case the rise induces a high accent, while the fall induces a low accent.

5. ACKNOWLEDGMENTS

I would like to thank Frédéric Beaugendre and David Keyson for their critical review of previous versions of this paper.

References

[1] d'Alessandro C. and Castellengo M. The pitch of short-duration vibrato tones. *Journal of the Acoustical society of America*, 95:1617–1630, 1994.

[2] Beckman M.E. and Pierrehumbert J.B. Intonational structure in japanese and english. *Phonology Yearbook*, 3:255–309, 1986.

[3] Bruce G. *Swedish word accents in sentence perspective*. Gleerup, Lund, Sweden, 1977.

[4] Beaugendre F. and Hermes D.J. De la relation entre le timing des mouvements mélodiques et l'accentuation des syllables. In *Journées d'Études de la Parole*, page forthc., Avignon, France, 10-14 June 1996.

[5] Caspers J. and Van Heuven V.J. Perception of low-anchoring versus high-anchoring of dutch accent-lending pitch rises. In *Proceedings of the European Speech Communication Association (ESCA) Workshop on Prosody*, pages 188–191, Lund, Sweden, 27-29 September 1993.

[6] Gósy M. and Terken J. Question marking in hungarian: timing and height of pitch peaks. *Journal of Phonetics*, 22:269–281, 1994.

[7] Hasegawa Y. and Hata K. Fundamental frequency as an acoustic cue to accent perception. *Language and Speech*, 35:87–98, 1992.

[8] Hermes D.J. *Timing of pitch movements and accentuation of syllables, IPO manuscript 1143*. Institutute for Perception Research / IPO, Eindhoven, 1996.

[9] Hill D.R. and Reid N.A. An experiment on the perception of intonation features. *International Studies on Man-Machine Studies*, 9:337–347, 1977.

[10] House D. *Tonal perception in speech*. Lund University Press, Lund, Sweden, 1990.

[11] Kohler K.J. Prosody in speech synthesis: the interplay between basic research and tts application. *Journal of Phonetics*, 19:121–138, 1991.

[12] Pierrehumbert J. *The phonology and phonetics of English intonation*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 1980.

[13] Rietveld T. and Gussenhoven C. The influence of syllable composition on the alignment of pitch targets. *Journal of Phonetics*, 23:375–385, 1995.

[14] 't Hart J., Collier R. and Cohen A. *A perceptual study of intonation: An experimental-phonetic approach to speech melody*. Cambridge University Press, Cambridge, UK, 1990.

[15] Van Katwijk A. and Govaert G. Prominence as a function of the location of pitch movement. *IPO Annual Progress Report*, 2:115–117, 1967.

[16] Van Santen J.P.H. and Hirschberg J. Segmental effects on timing and height of pitch contours. In *Proceedings of the International Conference on Spoken Language Processing (ICSLP)*, pages 719–722, Yokohama, Japan, 18-22 September 1994.