INFLUENCE OF FOCUS STRUCTURES ON TONAL TARGETS OF PITCH PEAKS

H. H. Rump
Institute for Perception Research/IPO, Eindhoven, The Netherlands

ABSTRACT

The purpose of the present study was to find out how 'equal prominence' and the peak heights of two pitch accents are related to the focus structure of an utterance. Subjects adjusted the height of one of two pitch peaks, matching the pitch contour to four different focus structures. The results suggest the existence of target values for the pitch peaks for each of the different focus conditions.

INTRODUCTION

In previous experiments involving utterances with two pitch accents (e.g. [6], [7], [8]) it appeared that the height of the second pitch peak was somewhat less than that of the first when the peaks lent equal prominence. The heights of the first and the second peaks turned out to be linearly related to each other. This was found for both nonsense and meaningful utterances, and having baselines with or without declination.

From these previous experiments, it is not clear, however, how equal prominence is related to the focus structure of the utterance in which the pitch accents occur. It may be assumed that equal prominence occurs either in the pragmatic context of the broad focus structure, in which the whole utterance is in focus, or in that of the double focus structure, in which the accented syllables are in focus separately. It may be assumed that the double focus structure is prosodically marked by higher pitch peaks, although this is not clear from the literature ([1], [2]). Neither is this clear for the previous experiments, however, since the focus structure of the test utterances was not made explicit. In the present experiments, different focus conditions were used in order to test how the heights of two pitch peaks are related to the focus structure of an utterance.

METHOD

The utterance used in the present experiments was "A'manda gaat naar 'Malta" (Amanda goes to Malta), which was spoken by a male person. It contained two accented syllables, /man/ and /mal/, with associated pitch peaks P1 and P2, respectively. The pitch accents had rising-falling pitch contours while the pitch in the unaccented syllables declined along a baseline which was a straight line in the ERB-rate frequency domain (unit: E; [4]). The starting frequency was 137 Hz, the end frequency was 100 Hz. The duration of the utterance was 1.45 s. The rate of declination was 0.7 E/s.

Two experiments were performed. In the first one, the height of P1 was fixed while the subjects adjusted the height of P2. In the second one, the height of P2 was fixed while the height of P1 was adjusted. Pitch manipulations were performed using the PSOLA method ([3]). Listeners selected the appropriate pitch contours from a prepared set of stimuli.

The task of the subjects was to adjust the height of a given pitch peak so that the resulting pitch contour would fit as close as possible one of four given focus structures. Broad focus was meant to give a neutral reading to the utterance. Single-focus conditions were elicited by asking questions which would result in contrastive readings of the target utterance so that only one of the pitch accents was in focus and the other one was not. In the double-focus condition, both accents were meant to be contrastive at the same time. The four instructions were:

- Wat zei je dat er gaat gebeuren?
(intended focus structure: broad focus)
- Gaat Jan naar Cyprus? Nee,...
(double focus)
- Gaat Jan naar Malta? Nee,...
(single focus on P1)
- Gaat Amanda naar Cyprus? Nee,...
(single focus on P2)

The instructions were printed to the computer screen while the test utterance 'Amanda gaat naar Malta' was made audible through headphones. The broad-focus and the double-focus conditions were expected to represent the 'equal-prominence' conditions. The single-focus conditions were included because they were expected to represent explicit 'different-prominence' conditions, thus providing some kind of boundaries for the equal-prominence conditions.

ADJUSTMENTS OF P2

The first group of subjects adjusted the height of P2 so that the utterance with the resulting pitch contour would be an adequate answer to the question/instruction which was written on the screen. During each trial, the height of P1 was fixed at one of three different values: 165, 183, or 202 Hz. The adjustments started at both extremes of the peak height continuum of P2 which ranged from 110 to 214 Hz, corresponding to excursion sizes of zero to 2.5 E. The range was divided in 10 steps of 0.25 E (about 1.5 st). Each adjustment was repeated twice, so that a subject completed four trials per instruction per P1 height. The order of presentation was completely random.

The ten subjects were students and research staff of the institute. They were all native speakers of Dutch, and they were not working on speech.

Results

The results averaged across all subjects are presented in Table 1. Every subject adjusted the height of P2 to be almost maximal when P2 was in single focus, and to be almost minimal, i.e. having excursion size zero, when P1 was in single focus.

The effect of Instruction on the adjusted height of P2 was highly significant ($F_{(3,27)} = 70.8$, $p < 0.001$). It is remarkable to find that the effect of P1 height on the adjusted height of P2, however, was not significant ($F_{(2,18)} = 2.94$, $p < 0.08$). The difference between the adjusted heights under the broad-focus and double-focus conditions turned out not to be significant ($F_{(1,27)} = 0.69$, $p > 0.05$), although the height of P2 tended to be greater under the double-focus than under the broad-focus condition.

Table 1. Adjusted P2 heights (Hz) under four different focus conditions and for three fixed heights of P1 (Hz).

	focus			
	P2	broad	double	P1
P1				
165	116	157	170	201
183	111	162	167	199
202	111	164	178	206

Discussion

The results for the single-focus conditions were as expected. If P2 was in focus, its height was made almost as large as possible. If P1 was in focus, the height of P2 was adjusted to be as small as possible. The latter is also in line with the theory that the last accented word in an utterance contains the nuclear accent. Focusing on the first accent, making it the nuclear accent, implies that the second accent should be deaccented, i.e. its excursion size should be zero.

For the broad-focus and double-focus conditions there were some individual differences. Some subjects adjusted P2 to be higher under the double-focus than under the broad-focus condition, while others adjusted P2 to be lower under the double-focus than under the broad-focus condition. This may explain why the difference between the two conditions was not significant. This may have been due to the fact that it is very difficult to interpret a neutral reading if you hear the same utterance again and again. The resulting annoyance may then have resulted in a non-neutral reading with a relatively high P2.

ADJUSTMENTS OF P1

The same utterance was tested with a second group of subjects. Again ten subjects participated, meeting the same selection criteria as above. They now adjusted the height of the first pitch accent, P1. The different values of P2 height were 143, 160, and 179 Hz. The P1 height continuum ranged from 131 to 267 Hz, corresponding to excursion sizes of zero to 3 E in twelve steps which were equidistant in E (0.25 E or about 1.5 st). The instructions were the same as the ones in experiment I. The order of presentation was again completely random.

Results

The results averaged across the subjects are listed in Table 2. The effect of Instruction was again highly significant ($F_{(3,27)} = 58.0$, $p < 0.01$). Again, unexpectedly the height of the fixed peak (P2) did not systematically influence the adjusted height of the other peak (P1). P1 height was adjusted to about the maximum value when P1 was in focus. If P2 was in focus, however, the adjusted height of P1 was on average more than 15 Hz above the minimum, resulting in an excursion size of about 2 st.

The double-focus condition resulted in significantly higher adjusted P1 values than the broad-focus condition ($F_{(1,27)} = 5.03$, $p < 0.05$).

Table 2. Adjusted P1 heights (Hz) under four different focus conditions and for three fixed heights of P2 (Hz).

	focus			
P2	broad	double	P1	
P2				
143	149	175	210	250
160	141	174	209	250
179	150	174	219	249

Discussion

If P1 was in focus, its height was adjusted to be as large as possible. Some of the subjects complained that they could not manipulate the height of P2, so that the resulting pitch contour was not optimal, P1 still sounding as a prenuclear accent.

If P2 was in focus, the excursion size of P1 was still about 2 st, so that the resulting peak height was only slightly below the average peak height of P2. This means that the excursion size of the pitch accent on the prenuclear accent may be larger than zero although it is deaccented. This is sometimes called a thematic or rhythmical accent, not lending much prominence to the word containing the accented syllable.

In the broad-focus condition, the average excursion sizes of P1 and P2 were about equal. In the double-focus condition, P1 was much higher than P2. This was true for the results of almost every single subject.

GENERAL DISCUSSION

The results show that the focus structure was crucial for the finally adjusted peak heights, and that the height of the other, fixed pitch peak had hardly any influence. In other words, it was mainly the focus structures that determined the resulting overall pitch contours. The results for the single-focus conditions were as expected in both experiments. The peak heights were large when the accented syllable was in the focused word, and they were small when the word containing the target syllable was explicitly out of focus (deaccented).

The difference between the broad-focus and double-focus conditions was most marked for P1. It was adjusted to significantly higher values for the double-foucs than for the broad-focus condition. For P2, this was found only as a tendency.

Unlike the previous experiments on prominence, the height of the fixed peak of one pitch peak had no systematic influence on the adjusted height of the other. This may indicate that when the pragmatics of prominence are involved, just one peak height represents a target value, which should be reached in order to obtain the appropriate pitch contour. This conclusion is supported by the findings reported in [5], where it was found that the pitch measured at certain points in the pitch contour is quite constant not only for a given speaker but also for a given instruction.

When we combine the results of the two experiments into only one pitch contour per focus condition, it is found that under the broad-focus condition the topline, connecting P1 and P2, and the baseline turn out to be about parallel. Under the double-focus condition, however, the topline turns out to be much steeper than the baseline.

It is not clear yet, however, whether listeners will be able to recognize an intended focus structure when they hear the pitch contour which is created using the pitch values obtained from the present experiments. This will be tested in a follow-up experiment.

REFERENCES

[1] Bartels, C., and Kingston, J. (1994), "Salient pitch cues in the perception of contrastive focus", in: Focus & Natural language processing, Proc. of a conference in celebration of the 10th anniversary of the Journal of Semantics, Meinhard-Schwebda, Germany, vol. 1, Intonation and Syntax, pp. 1-10.

[2] Batliner, A. (1994), "Prosody, Focus, and Focal Structure: Some Remarks on Methodology", in: Focus & Natural language processing, Proc. of a conference in celebration of the 10th anniversary of the Journal of Semantics, Meinhard-Schwebda, Germany, vol. 1, Intonation and Syntax, pp. 11-28.

[3] Hamon, C., Moulines, E., and Charpentier, F. (1989), "A diphone synthesis system based on time domain prosodic modifications of speech", *Proc. ICASSP-89*, pp. 238-241.

[4] Hermes, D. J., and Van Gestel, J. (1991), "The frequency scale of speech intonation", *J. Acoust. Soc. Am.*, vol. 90, 97-102.

[5] Ladd, D. R., and Terken, J. M. B. (1995), "Modelling intra- and inter-speaker pitch range variation", *Proc. ICPhS-95*, Stockholm, Sweden.

[6] Pierrehumbert, J. (1979), "The perception of fundamental frequency declination", *J. Acoust. Soc. Am.*, vol. 66, 363-369.

[7] Repp, B. H., Rump, H. H., and Terken, J. M. B. (1993), "Relative perceptual prominence of fundamental frequency peaks in the presence of declination", *IPO Annual Progress Report*, vol. 28, 59-62.

[8] Terken, J. M. B. (1991), "Fundamental frequency and perceived prominence of accented syllables", *J. Acoust. Soc. Am.*, vol. 89, 1768-1776.