Prominence lent by rising and falling pitch movements: Testing two models

H. H. Rump and Dik J. Hermes
Institute for Perception Research/IPO, P.O. Box 513, 5600 MB Eindhoven, The Netherlands
(Received 1 March 1995; revised 13 November 1995; accepted 14 March 1996)

Two experiments are reported in which a pitch-level difference (PLD) model for prominence perception [Hermes and Rump, J. Acoust. Soc. Am. 96, 83–92 (1994)] is subjected to further tests. The model holds that the contribution of pitch to the perceived degree of prominence is proportional to the difference in pitch level between the vocalic nuclei of the accented and the previous syllable. In experiment I, the influence of stretching and compressing the utterance in time was assessed. It was found that the predictions made by the model were not fully supported by the data. An alternative model was developed according to which pitch movements resynthesized in the same register lend equal prominence when pitch levels on the upper declination lines in the stimuli are equal. These two models gave different predictions when the lower declination lines are different. This was tested in experiment II. The results which are more or less between the predictions by the two models suggest that low pitch levels play a smaller role in prominence perception than high pitch levels do. © 1996 Acoustical Society of America.

PACS numbers: 43.71.An, 43.71.Es, 43.70.Fq [RAF]

INTRODUCTION

Prominence, defined as the conspicuousness of a syllable, plays an important role in intonation. Various studies have been reported relating the relative prominence of pitch-accented syllables to the succession of pitch peaks (e.g., Bruce, 1982). A typical conclusion of one such study is, e.g., ‘‘After a focal accent, the downstepping of successive non-focal accents is a characteristic pitch pattern. This downstepping seems to be the expression of equal prominence of successive post-focal accents within the phrase. In a pre-focal position, however, up to the focal accent of a phrase (or of a whole utterance) there is typically no downstepping, but instead only a very gentle declination (if any) for successive non-focal accents’’ (Bruce and Touati, 1992, pp. 455 and 456).

Careful studies to relate equal prominence of accented syllables within one utterance to simple acoustic features of the pitch contour have not lead to a simple and plausible model (Pierrehumbert, 1979; Rietveld and Gussenhoven, 1985; Terken, 1991, 1993b, 1994; Ladd, 1993; Repp et al., 1993). This may be due to the implicit focus structure of the utterance coupled with the absence or presence of downstep, which may have been confounding factors. In addition, the proposed models deal with the relative heights of the rise–falls which are associated with the accented syllables, leaving other possible accent-lending pitch movements like isolated rises and falls aside.

Only very few studies (Hermes and Van Gestel, 1991; Hermes and Rump, 1994) have been reported dealing with the relative prominence of accented syllables in two successive utterances. In these studies, the pitch accents were always on the second syllable of the three-syllabic utterance /ma’mama/. Hermes and Rump (1994) asked subjects to adjust the excursion size of the pitch movement in the second utterance until the prominence in that utterance was equal to the prominence of the pitch-accented syllable in the first utterance. The utterances were resynthesized in the same or in different registers. The pitch movements in the two utterances, however, were always different. Each could be a rise, a fall, or a rise–fall. One of the main findings was that, lending equal prominence, a fall had a smaller excursion size than a rise or a rise–fall. Varying the temporal position of the rise only lead to different results when the rise started later than the vowel onset of the accented syllable. Varying the temporal position of the fall did not affect the results very much. In the description of Dutch intonation by ‘t Hart et al. (1990) an accent-lending rise can be either early (rise 1) or late (rise 3), in autosegmental phonology corresponding with $L\!+\!H^*$ and $L^*\!+\!H$, respectively. In the description by ‘t Hart et al. there is only one full-sized fall, corresponding with $H\!+\!L^*$ in autosegmental phonology.

The unexpected finding in the experiments by Hermes and Rump (1994) was that excursion sizes being equal, the rise and the rise–fall lent less prominence to an accented syllable than the fall. This did not depend on whether the two utterances were in the same register or in different registers. Based on these considerations, Hermes and Rump proposed a pitch-level difference (PLD) model in which the prominence lent by an accent-lending pitch movement is proportional to the difference between the pitch level$^2$ in the vocalic nucleus$^3$ of the accented syllable and that of the preceding syllable. In other words, the height of the pitch level in the accented syllable is judged relative to the height of the pitch level in the preceding unaccented syllable. The difference between the two levels is the PLD. The PLD model states that two pitch movements lend equal prominence when the PLDs are equal.

\*Electronic mail: hermes@natlab.research.philips.com; fax: +31 40 2773876.
The PLD model is illustrated in Fig. 1, in which two accent-lending pitch movements, a rise and a fall lending equal prominence, are shown on top of each other. It explains the (experimentally observed) difference in excursion size needed to match the prominence lent by the two pitch movements. “Excursion size” is defined, like in ’t Hart et al. (1990), as the distance between the lower declination line, or baseline, and the upper declination line, or topline, measured perpendicularly to the time axis. Here, in contrast to ’t Hart et al., who measured excursion sizes in semitones, the distance is measured in numbers of equivalent rectangular bandwidths (ERBs; unit E) (Patterson, 1976; Hermes and Van Gestel, 1991). Declination lines run parallel in the ERB-rate–frequency domain.

Pitch levels that are assumed to be relevant for the determination of PLDs are indicated by ellipses. The pitches averaged across the vocalic nuclei, which are defined as areas of maximum sonority, were believed to provide the relevant pitch levels. The black ellipses represent the relevant pitch levels for the rise; the open ellipses represent the relevant pitch levels for the fall. The arrows on the left-hand side indicate the PLDs for the rise and the fall. The length of the arrows may hence be seen to represent the contribution of pitch to the prominence lent by a rise or a fall.

From the traditional view of prominence perception (e.g., Cohen et al., 1982), it would follow that a rise and a fall lend equal prominence when their excursion sizes are equal, i.e., that listeners take declination into consideration. In the PLD model, however, no such compensation for short-term declination, i.e., declination over the time span of two syllables within one utterance, is assumed. In Fig. 1 the black arrows on the right-hand side indicate the excursion sizes of the rise and the fall lending equal prominence. The difference in excursion size (open arrow) is called the “shift” $S$, which can be calculated to equal twice the rate of declination $d$ times the time interval $T$ between the relevant pitch levels, so that

$$S = 2dT.$$

(1)

Shifts are measured in ERB-rate units E since $d$ is measured in $E/s$ and $T$ is measured in s. In this equation, the declination $d$ is assumed to be constant over the utterance, or at least over the syllable preceding the accented syllable and the accented syllable itself.

Hermes and Rump have shown that the predictions derived from Eq. (1) were quite accurate for different declination rates $d$. They measured $S$ as a function of $d$ and so they could calculate $T$. It was found that the measured shifts were indeed linearly related to the rate of declination. The value of $T$ appeared to be about 220 ms ($S/2d$) which corresponded more or less precisely with the distance between the vowel onsets of the syllabic nuclei of the accented syllable and the preceding syllable. Since the shift depends on both the rate of declination and the time interval between relevant pitch levels [Eq. (1)], the PLD model predicts that, if $d$ is kept constant and is not zero, a change in $T$ will also affect the PLDs. For example, an increase in $T$, achieved by slowing down the speech rate, will result in a larger shift, because, at the same time, the PLD for the fall increases, while the PLD for the rise decreases. It is not known, however, whether the relevant pitch levels are indeed anchored to the vocalic nuclei. $T$ may also be fixed to some value of about 220 ms. According to the autosegmental analysis of bitonal pitch accents, a fixed value for $T$ may be assumed. Pierrehumbert (1980, p. 77), for example, found for the $L^* + H^-$ accent that the $L^*$ tone is directly associated with the accented syllable, while the position of $H^-$ is fixed at about 200 ms after the starred tone, more or less independently of the syllabic structure. The situation in Hermes and Rump’s experiments was different from that in Pierrehumbert’s. Pierrehumbert obtained the value for a bitonal accent in which the first and not the second tone was associated with the stressed syllable, whereas in Hermes and Rump, it was the second tone which was starred, i.e., $H + L^* = $falling and $L + H^* = $rising. However, it may be that a leading unstressed tone also occurs at a fixed interval from the starred tone, like a trailing unstressed tone does. If $T$ is indeed fixed, changing the speech rate might not result in a change in the shift. The aim of the first experiment was to test this.

The exact distance between vocalic nuclei is difficult to determine. As the time interval $T$ turns out to be about the same as the time interval between the vowel onsets of the accented and the previous syllable $V$, this time interval $V$ between subsequent vowel onsets is used for predicting the shift:

$$S = 2dV.$$

(2)

The hypothesis is then that the inter-vowel-onset interval $V$ and the inter-pitch-level interval $T$ are identical, i.e., $T = V$. If $T$ is indeed anchored to the vowel onsets, shifts are predicted to be equal to $S = 2dT = 2dV$, i.e., $S$ will increase if $V$ increases, and $S$ will decrease if $V$ decreases, the nonzero declination being kept constant. Alternatively, if the $T$ is a stable interval, shifts are predicted to be independent of $V$, and hence to be equal to $S = 2dT$ [Eq. (1)], and not to $S = 2dV$ [Eq. (2)].
I. EXPERIMENT I

A. Method

In Hermes and Rump it was shown that listeners were well able to adjust prominences to make them equal if they were associated with pitch movements which were resynthesized in equal or different pitch registers. The measured shifts turned out to be almost the same, regardless of the register. Only the standard deviations of the shifts were larger for different registers, 0.2 E, than for equal registers, <0.1 E. It was concluded that, in general, the same averages would be obtained for different registers as for equal registers, and that only standard deviations would differ. Therefore, in the current experiment, the stimuli were resynthesized in the same register, since this would lead to smaller standard deviations. This was important because changing the inter-vowel-onset interval $V$ was expected to result in relatively small differences between the measured shifts under the different conditions. The predicted shifts under the different stimulus conditions were as follows: The declination rate $d$ was fixed at 1.0 E/s and the inter-vowel-onset intervals, $V$, were 150, 220, and 350 ms. The ratio between the short $V$ and the long $V$ was therefore greater than 1:2. According to Eq. (2), the predicted shift $S$ was 0.30 E for the short stimuli, 0.44 E for the original stimulus duration, and 0.70 E for the long stimuli. Shifts under the different stimulus-duration conditions were therefore expected to differ by about 0.4 E at most. Large standard deviations might obscure these small differences.

1. Stimuli

The stimuli consisted of modified versions of the carrier (nonsense) utterance /ma’mama/, spoken by a male speaker, which was the same utterance as that used in Hermes and Rump. Manipulations of duration and pitch were performed using the pitch-synchronous overlap-and-add (PSOLA) method developed by Hamon et al. (1989), resulting in natural-sounding speech. Vowel onsets in the original utterance were determined perceptually. In order to make the short and long stimuli sound optimally natural, the utterance was shortened and lengthened non-linearly, as is illustrated in Fig. 2. The resulting values of $V$ were 150 and 350 ms for the short and the long stimuli, respectively. In the stimuli with the original syllable durations $V$ was 220 ms. The extra lengthening of the /m/ in the second syllable of the long stimuli in particular seemed to be crucial for the second syllable still to be perceived as accented. In addition, this extra lengthening helped to increase $V$ without making the stimuli sound unnatural.

An accent-lending rise and fall were effected between the baseline and the topline which ran parallel in the ERB-rate frequency domain. The pitch movements had durations of 120 ms, independent of the speech rates. The rise started 70 ms before the vowel onset of the second, accented syllable; the fall started 20 ms before the vowel onset of the accented syllable (the standard values for rise 1 and fall A in the IPO text-to-speech synthesis system; see ’t Hart et al., 1990; Terken, 1993a). The end frequency of the baseline was fixed at 75 Hz (2.64 E). The rate of declination was also the same for all stimulus durations, i.e., 1.0 E/s. This was somewhat higher than the standard declination rate of 0.7 E/s. Since declination is a factor in the PLD model, a relatively fast rate of declination would make the differences between predicted shifts as large as possible while keeping the sound of the stimuli natural. Since the rate of declination and the end frequency of the baselines were kept constant during the experiment, the starting frequencies of the baselines for the three stimulus durations were different, i.e., 93 Hz (3.18 E), 101 Hz (3.39 E), and 111 Hz (3.69 E) for the short, original, and long stimuli, respectively.

The stimuli were presented in pairs. The first stimulus will be referred to as test stimulus, the second one as comparison stimulus. The test stimulus had four excursion sizes: 0.83, 1.11, 1.38, and 1.67 E. The range of adjustable comparison stimuli, from 0 to 2.5 E (about 94 Hz or 14 semitones), was divided into 18 equidistant steps of 0.14 E.

2. Procedure

The test and comparison stimuli always had identical durations. Between the stimuli in a pair there was a silent interval of about 600 ms. The listeners matched the adjustable excursion size of the pitch movement in the comparison stimulus until its prominence matched the prominence lent by the opposite type of pitch movement in the test stimulus. Within one trial, the excursion size of the latter was fixed while the comparison stimulus had a variable excursion size. It means that after listening to a stimulus pair, the listeners could increase or decrease the excursion size of the comparison stimulus by pressing keys on a computer keyboard. They could repeat this until they found that the prominences of the accented syllable in the comparison stimulus and that in the test stimulus were equal. Their key strokes were recorded automatically. When they were satisfied by the finally ad-
justed excursion, they pressed the return key and the next stimulus pair was presented. A typical number of key strokes within a trial was nine.

For each excursion size of the test stimulus, the comparison stimulus was adjusted twice, starting from either end point of its excursion size continuum. This was done in order to counter the effect of hysteresis, a methodological bias due to which excursion sizes are made larger when the starting point is high and smaller when the starting point is low. In other words, for both rising and falling comparison stimuli, larger adjustments in excursion are made when starting at the high than at the low end of the continuum. Another way of undoing the effect of hysteresis is to take the means of adjustments of the rise and the fall, as was done in Hermes and Rump. Taking two starting points for each adjustment, however, leads to better estimates of the adjusted excursion sizes for both rises and falls. The total number of trials was 48 (3 stimulus durations×2 pitch movements in the test stimulus×4 excursion sizes×2 starting points). The order of presentation was randomized.

Nine subjects participated in the experiment. They were instructed to make the prominence (‘‘pep’’ or ‘‘schwung’’) of the two stimuli equal. All of them had taken part in earlier prominence-adjustment experiments. The consistency of their scores was determined using the criterion that the correlation between the scores under a given stimulus-duration condition had to reach the level of 0.75 (Hermes and Van Gestel, 1991). Data were excluded if subjects did not reach the level of consistency under at least two of the three conditions.

B. Results and discussion

One of the subjects mentioned that he had adjusted equal musical pitch intervals in the individual stimuli, while not paying much attention to prominence. His data were therefore omitted from any further analysis. The data of two other subjects did not reach the required level of consistency, and they were therefore also discarded.

The analysis of the results of the six remaining subjects showed that the effect of hysteresis was significant ($F_{1,3} = 42.5, p < 0.008$). The adjusted excursion sizes were about $0.14\ E$ larger when adjustments started at the high end of the continuum than when adjustments started at the low end of the continuum. The magnitude of the effect was equal to about one step in the excursion-size continuum. Since we wanted to counter the effect of hysteresis, the results were pooled across starting points.

Figure 3 shows the adjusted excursion sizes averaged across the six subjects. Panel (a) shows the results for the short stimuli, panel (b) for the stimuli having the original duration, and panel (c) for the long stimuli. The horizontal axes give the excursion sizes of the rise; the vertical axes give the excursion sizes of the fall. The mean adjusted excursion sizes of rises, for which the test stimulus contained a fall, are indicated by open circles and horizontal standard-deviation bars, those of falls, for which the test stimulus contained a rise, are indicated by black circles and vertical standard-deviation bars. The standard-deviation bars indicate the combined effect of intrasubject variance and intersubject variance.

As in the experiments reported by Hermes and Rump, the results displayed a trend toward the average in the stimulus range. This bias often occurs in adjustment experiments. It can be very strong in difficult tasks. In the extreme, listeners will select a random stimulus from the whole stimulus range. The average of these adjustments would be at the average of the range. Another explanation for this trend may be that subjects, after having listened to many stimuli, have built up a kind of ‘‘standard’’ stimulus of average excursion size. At low attention levels or in difficult tasks, the adjustments of the subject would be drawn into the direction of this standard stimulus. Note that, in the present experiment, by the way the data are presented, this trend toward the average is in the vertical direction for the rises and in the horizontal direction for the falls. As a result, the regression lines for rises and falls do not coincide. Therefore, in order to correct for this trend toward the average, the data for rises and falls are combined, resulting in a better estimate of the shift.

The solid line in each of the panels of Fig. 3 indicates equal excursion sizes of the pitch movements. The distance
between the solid and the dashed line, measured perpendicularly to the horizontal axis, is the predicted shift: $S = 2dV$ [Eq. (2)]. The dashed line in each of the panels therefore indicates the positions of the measurement points that would be predicted if the time interval between relevant pitch levels $T$ and the time interval between subsequent vowel onsets $V$ were to be identical. It can be seen that the majority of the measurement points lie along the dashed lines. A repeated-measures ANOVA showed that the effect of $V$ on $S$ was significant ($F_{1(2,10)} = 25.3$, $p < 0.001$). As predicted by the PLD model, a decrease in $V$ caused a decrease in the shift, and an increase in $V$ caused an increase in the shift.

The shifts for the individual subjects are shown in Table I. A longer $V$ resulted in a larger shift in the case of almost every individual subject. In the case of subject 4, the effect of $V$ on the adjusted excursion sizes was marginal with respect to the shortening of the stimuli, and indeed there was no effect with respect to the lengthening of the stimuli and his shifts were consequently more or less constant. Mean shifts and their standard deviations were calculated from the distances between the individual data points and the diagonal, measured perpendicularly to the horizontal axis, in each of the panels in Fig. 3. The predicted shifts were based on Eq. (2).

The mean shift calculated for the short stimuli turned out to be somewhat larger than predicted. The mean shift calculated for the long stimuli were considerably smaller than predicted (see also panel c in Fig. 3). The values of $T$, the time interval between relevant pitch levels, which could be derived from the measured shifts ($T = S/2d$), were about 170, 225, and 270 ms for the short, normal, and long stimuli, respectively. The corresponding values of $V$ were 150, 220, and 350 ms. $T$ was therefore 20 ms longer than $V$ in the short stimulus, whereas it was 80 ms shorter than $V$ in the long stimulus. (In the case of the original stimuli, the finding of Hermes and Rump that $T$ and $V$ were about equal was replicated.) The decrease in $T$ was therefore only 15 ms smaller than the decrease predicted by Eq. (2) (55 instead of 70 ms), but the increase in $T$ was 85 ms smaller than the predicted increase (45 instead of 130 ms). This means that the predictions derived from the PLD model were not fully supported if it is assumed that the distance between the relevant pitch levels stretches and shrinks with the distance between the

![Figure 4](image)

FIG. 4. As in Fig. 1, it can be seen that the height of the high pitch level in the unaccented syllable before the fall (open ellipsis on the left) and the high pitch level after the rise (hatched ellipses on the right) are about equal when the PLDs (arrows on the left) are equal. $T$ indicates the time interval between the pitch levels in subsequent syllables, $T'$ indicates the time interval between the high pitch levels in the first syllable, i.e., before the fall, and in the third syllable, i.e., after the rise. As in Fig. 1, the excursion sizes (black arrows on the right) of the rise and the fall then differ by the amount $S$ (shift, open arrow on the right).

### Table I. Individual shifts ($S$) in experiment I, the shifts predicted by the PLD model [Eq. (2)], and the mean observed shifts with their standard deviations. Measures in number of ERBs.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Stimulus duration</th>
<th>Obs.</th>
<th>Pred. [Eq. (2)]</th>
<th>s.d.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>short</td>
<td>0.295</td>
<td>0.300</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td>original</td>
<td>0.399</td>
<td>0.440</td>
<td>0.093</td>
</tr>
<tr>
<td></td>
<td>long</td>
<td>0.530</td>
<td>0.700</td>
<td>0.540</td>
</tr>
<tr>
<td>2</td>
<td>short</td>
<td>0.365</td>
<td>0.334</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td>original</td>
<td>0.547</td>
<td>0.453</td>
<td>0.093</td>
</tr>
<tr>
<td></td>
<td>long</td>
<td>0.590</td>
<td>0.540</td>
<td>0.054</td>
</tr>
<tr>
<td>3</td>
<td>short</td>
<td>0.156</td>
<td>0.356</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td>original</td>
<td>0.278</td>
<td>0.530</td>
<td>0.093</td>
</tr>
<tr>
<td></td>
<td>long</td>
<td>0.399</td>
<td>0.634</td>
<td>0.564</td>
</tr>
<tr>
<td>4</td>
<td>short</td>
<td>0.503</td>
<td>0.503</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td>original</td>
<td>0.521</td>
<td>0.521</td>
<td>0.093</td>
</tr>
<tr>
<td></td>
<td>long</td>
<td>0.521</td>
<td>0.564</td>
<td>0.055</td>
</tr>
<tr>
<td>5</td>
<td>short</td>
<td>0.356</td>
<td>0.356</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td>original</td>
<td>0.530</td>
<td>0.530</td>
<td>0.093</td>
</tr>
<tr>
<td></td>
<td>long</td>
<td>0.634</td>
<td>0.540</td>
<td>0.054</td>
</tr>
<tr>
<td>6</td>
<td>short</td>
<td>0.330</td>
<td>0.334</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td>original</td>
<td>0.443</td>
<td>0.453</td>
<td>0.093</td>
</tr>
<tr>
<td></td>
<td>long</td>
<td>0.700</td>
<td>0.540</td>
<td>0.054</td>
</tr>
</tbody>
</table>

### 1. A different listening strategy: Pitch levels of the same heights

So, it may still be possible to explain the results of the experiment within the framework of the PLD model. Based on the comments by some of the subjects as well as for theoretical reasons, however, an alternative strategy will be presented to account for the discrepancy between predicted and observed shifts. Given that the task of adjusting prominence was a rather difficult one, some subjects mentioned that they may have been following the strategy of adjusting the high pitch levels in a stimulus pair so as to be equal. In the following section, we will present a quantitative analysis of the shifts predicted by the strategy of matching high pitch levels, which may give a better explanation of the present findings. In experiment II, the new predictions will be contrasted with those based on the PLD model.

If we compare the predicted outcome of this strategy to the predictions of the PLD model, the following occurs. According to the PLD model, listeners adjust the PLD of the comparison stimulus so as to make it match the PLD of the test stimulus. This is illustrated in Fig. 4, in which the PLDs are indicated by the black arrows on the left. When the PLDs are the same, the excursion sizes, indicated by the black arrows on the right, differ by the amount represented by $S$, indicated by the open arrow, also on the right.

It can be seen that high pitch levels on the toplines of the stimuli turn out to be roughly equal when the PLDs are equal (open ellipsis on the left and hatched ellipsis on the right). Since the task of matching PLDs may have been quite complex, an alternative would be that listeners adjusted the indi-
lected high pitch levels, viz., in the first syllable, i.e., before the fall, and in the third syllable, i.e., after the rise, so as to make them equal. Note that the first and the last syllable are unaccented.

A linguistic reason why matching high pitch levels may have been a valid strategy can be found in the experiments by Kutik et al. (1983). They found that listeners matched the toplines of a main clause before and after a parenthetical clause which had different lengths. In our experiments, listeners may have interpreted the stimuli in a pair as being the parts of one single utterance containing two pitch-accented syllables, especially in the case of a rise followed by a fall. The pause would than be the parenthetical.

Adjusting high pitch levels so as to make them equal may have been even more plausible as to be a strategy due to the actual experimental setup. In the present experiments, differences in prominence were brought about by changing the height of the topline relative to the fixed baseline, because it is known from perception experiments that variation of high pitch levels is perceptually more salient for the perception of intonation than variation of low pitch levels (e.g., Sluijter and Terken, 1993). Since the excursion size was controlled by changing the height of the topline relative to the fixed baseline, changes in the height of high pitch levels were the only cue to changes in prominence. This might suggest that subjects had adjusted pitch levels on the toplines to become equal.

In the current experiment, these two strategies, i.e., the matching of PLDs and the matching of high pitch levels, were not conflicting. Actually, both strategies lead to the same result. The next experiment was designed in such a way that different strategies would lead to different results.

The strategy of equating high pitch levels can be formulated quantitatively in the following way. As shown in Fig. 4, it is most likely that the listeners chose the high pitch level in the first syllable of the stimulus containing the fall and the high pitch level in the third syllable of the stimulus containing the rise. The shift \( S \) is then determined by the rate of declination, \( d \), and by the time interval between the relevant pitch levels in the first and the third syllable, \( T' \), or

\[
S = dT'.
\]

If we want to predict the shifts that would have resulted in experiment I, we may assume, as we did in the PLD model, that the time interval between relevant pitch levels is of the same order as the time interval between the onsets of the vowels containing the relevant pitch levels. This time interval between the first and the third syllable will be referred to as \( V' \). Measurements showed that \( V' \) was about 470 ms, which turned out to be nearly twice the time interval between the vowel onsets of the first and the second syllable (\( V: 220 \) ms). The newly predicted shifts based on the matching of high pitch levels will then be

\[
S = dV',
\]

instead of \( S = 2dV \), but since \( 2V' \) are almost equal, the two predictions are only marginally different.

The predictions derived from Eq. (4) are shown in Table II, together with the shifts measured in experiment I and the shifts predicted by the PLD model [Eq. (2)]. It can be seen that the observed shifts are between the predictions derived from the PLD model and the predictions derived from Eq. (4), at least in the case of the short and original stimuli. In the case of the long stimuli, however, both predicted values of the shift are too large. It is not clear why this has been the case. For both strategies, ad hoc explanations may be given.

1) In terms of the PLD model, as mentioned before, it may be assumed that the contribution of pitch to prominence is not proportional to the pitch at points fixed relative to the vowel onsets but that a pitch level early in the vowel of the accented nucleus is compared with a pitch level late in the vowel of the preceding syllable. In other words, \( T \) increases with an increase of \( V \), but the increment is somewhat smaller.

2) In terms of the pitch-matching strategy, this means that the listeners matched a lower than predicted pitch of the third syllable to the pitch of the first. It may be assumed that, due to its relatively long duration, the high pitch level of the accented second syllable had become a better candidate for the match with the high pitch level in the first syllable than the high pitch level in the unaccented third syllable. In other words, \( T \) changes due to the fact that a different pitch level becomes more relevant.

It is thus argued that listeners may have adopted the strategy of matching high pitch levels rather than matching pitch-level differences. We will refer to this new strategy as the high-level (H) model. In experiment II, the PLD model and the H model were compared by using stimuli with different baselines.

II. EXPERIMENT II

From the discussion of experiment I it has become clear that the PLD model and the H model both make rather similar predictions if the low pitch levels are on baselines with equal endpoints. According to the PLD model, pitch levels on the baseline and pitch levels on the topline are equally important, because the listeners need information about low and high pitch levels in order to be able to determine the PLDs. According to the H model, however, only pitch levels on the toplines in the test and comparison stimuli determine the finally adjusted excursion sizes. If the baselines in the test and comparison stimuli are different, the PLD model and
the H model make different predictions about the resulting shifts. In experiment II we tested whether low pitch levels contributed to the perceived degree of prominence.

A. Method

1. Stimuli, procedure, and subjects

As in experiment I, the stimuli consisted of PSOLA-manipulated versions of the nonsense utterance /ma m'ma ma/. The prominence lent by a rise was compared with the prominence lent by a fall. The stimuli were resynthesized in the same low register. The slopes of the declination lines were kept constant at 0.7 E/s. Unlike in experiment I, the end frequencies of the baseline in the test and comparison stimuli were chosen to be different. Two new excursion-size continua were constructed, in which the end frequency of the baseline was shifted up or down by 11 Hz, about 0.33 E or 2.4 semitones, relative to the formerly used end frequency of 75 Hz, resulting in end frequencies of 64 and 86 Hz. (Since the baselines ran parallel in the ERB-rate frequency domain, the difference of 11 Hz between the end frequencies amounted to a difference of 0.33 E between the baselines, which means that differences between the begin frequencies were slightly larger than 11 Hz.) In the remainder of the paper we will refer to stimuli with rises as R and stimuli with falls as F. The end frequency of the baseline will be indicated by specifying the Hz value in subscript after the pitch movement. For example, R_{75} means that the (virtual) end frequency of the baseline in the stimulus containing the rise was 75 Hz.

Stimuli with a shifted baseline were always compared with stimuli having the original baseline with the end frequency of 75 Hz. If the listeners had attended mainly to the high pitch levels in the unaccented syllables before the fall and after the rise, the resulting shift would have been different from the shift predicted by the PLD model. If the relevant high pitch levels, in the first syllable, before the fall, and in the third syllable, after the rise, had been made equal, the predicted shift would have been

\[ S = dV' \pm 0.33. \]  

Since \( dV' \) is equal to 0.33 E under the present stimulus conditions, the H model [Eq. (5)] predicts shifts of either 0.00 or 0.66 E. The predictions by the H model are illustrated in Fig. 5. Equation (5) predicts a shift zero when a rise is compared with a fall having a lower baseline, since high pitch levels are equal when excursion sizes are equal (comparisons between a rise with baseline A and a fall with baseline O, and between a rise with baseline O and a fall with baseline B). Equation (5) predicts a shift of 0.66 E when a rise is compared with a fall having a higher baseline (comparisons between a rise with baseline O and a fall with baseline A and between a rise with baseline B and a fall with baseline O).

The newly predicted shifts are either zero or more than twice the size of the shift predicted by the PLD model, depending on the heights of the baselines in the test and comparison stimuli. The shift predicted by the PLD model was assumed to be independent of the different baselines of the stimuli, and to depend only on the height of the topline relative to the baseline within each individual stimulus. The shift would therefore always be 2dT or about 0.31 E.

The experimental procedure was identical to the procedure used in experiment I. The total number of adjustments per subject was 64 (four end-frequency combinations: 64–75, 75–64, 86–75, and 75–86; two pitch movements on the test stimulus; four excursion sizes of the test stimulus; two starting points in the excursion-size continuum). The order of presentation was randomized.

Nine subjects participated in the experiment. All except one had taken part in the previous experiments. The same consistency criterion was applied as in experiment I. A subject’s data were included in the analysis if the score was consistent for at least three of the four end-frequency combinations.

B. Results and discussion

The data of two subjects had to be omitted from the analysis because they did not show the required consistency. The same subject as in experiment I had made musical intervals equal. His data were therefore also omitted. The analysis of the results of the six remaining subjects showed that the effect of the starting points in the excursion-size continuum (hysteresis) was much greater in the present experiment than in experiment I: 0.25 E (experiment I: 0.14 E). It was of the same order of magnitude for all comparisons. The results are shown in Fig. 6. As in Fig. 3, the adjusted excursion sizes were averaged across the starting points and across the subjects. Panels (a)–(d) show the results obtained for the four comparisons. The open circles having horizontal standard-deviation bars again refer to the adjusted excursions of the rise; the black circles having vertical standard-deviation bars refer to the adjusted excursions of the fall. The solid line in each panel indicates the position of the measurement points.
if equal excursion sizes were to lend equal prominence. The dashed line in each panel indicates the shift predicted by the PLD model. The distance between the solid and the dashed line measured perpendicularly to the horizontal axis is therefore 0.31 E, and is the same in all panels. The dotted lines indicate the shifts predicted by the H model. Therefore, the distance between the solid and the dotted lines depends on the combination of the baselines in the pair of stimuli, the shifts being about zero in panels (b) and (c), in which the diagonal and the dotted line coincide, and 0.66 E in panels (a) and (d).

As in experiment I, the regression lines for the adjusted excursion sizes of rises and falls were slightly different due to the effect of a trend toward the average explained above. The observed shifts for the individual subjects are presented in Table III. The results of all the subjects showed the same trends. A repeated-measures ANOVA across subjects showed that different end frequencies of the baselines had a significant effect on $S$ ($F_{(3,15)} = 112.9$, $p < 0.001$). Mean shifts and their standard deviations were calculated from the distances between the individual data points and the diagonal, measured perpendicularly to the horizontal axis, in each of the panels in Fig. 6. The shifts predicted by the PLD model and the H model are also displayed. As predicted by the H model, the shifts were relatively large in the comparisons in which the rise had a lower baseline than the fall ($R_{64} \text{ vs } F_{75}$ and $R_{75} \text{ vs } F_{86}$), and the shifts were relatively small in the comparisons in which the rise had a higher baseline than the fall ($R_{75} \text{ vs } F_{64}$ and $R_{86} \text{ vs } F_{75}$).

The results were found to be closer to the values predicted by the H model [Eq. (5)] than to the values predicted by the PLD model [Eq. (2)]. This suggests that matching high pitch levels played a significant role when subjects adjusted prominence. But the fact that the measurements were between the two predictions indicates that the influence of the low pitch levels was significant, too.

### III. GENERAL DISCUSSION AND CONCLUSIONS

In experiment I it was shown that a change in the time interval between successive vowel onsets had a significant influence on the adjusted excursion sizes of the rise and the fall. The shift, indicating the difference between the excursion sizes of a rise and a fall lending equal prominence, decreased when the stimuli were shortened, and increased when the stimuli were lengthened. When these results are interpreted in terms of the PLD model this means that a change in the time interval between vowel onsets indeed resulted in a change in the time interval between pitch levels which were assumed to be relevant for the determination of the PLDs. The hypothesis that $T$ is a fixed time interval is therefore not supported.

Although the measured shifts were close to the values predicted by the PLD model, they did not fully support the model. An alternative explanation was proposed, to the effect that listeners attended to high pitch levels located in the unaccented syllables of the stimuli when adjusting prominence so as to make it equal (H model). The reanalysis of the results of experiment I showed that the predictions made by the H model matched the results fairly well, at least as far as the short and original stimulus durations are considered.

In experiment II, we found that the listeners tended to attend to high pitch levels, making them about equal when adjusting prominence. The role of low pitch levels turned out to be significant, too, which means that both strategies, the one following the PLD model and the one following the H model, played a role when the subjects performed the adjustment task.

In summary, if the baselines are equal, the strategy following the H model leads to the same adjustments as the strategy following the PLD model. If the baselines are different, however, the strategies lead to different adjustments.

### Table III

<table>
<thead>
<tr>
<th>Comparison</th>
<th>$R_{64}F_{75}$</th>
<th>$R_{75}F_{64}$</th>
<th>$R_{64}F_{75}$</th>
<th>$R_{75}F_{86}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pred. [Eq. (2)]</td>
<td>0.330</td>
<td>0.330</td>
<td>0.330</td>
<td>0.330</td>
</tr>
<tr>
<td>Pred. [Eq. (5)]</td>
<td>0.660</td>
<td>0.000</td>
<td>0.000</td>
<td>0.660</td>
</tr>
<tr>
<td>Obs.</td>
<td>0.554</td>
<td>0.091</td>
<td>0.013</td>
<td>0.560</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.112</td>
<td>0.169</td>
<td>0.087</td>
<td>0.100</td>
</tr>
</tbody>
</table>
which makes the task more difficult. This is illustrated by larger effect of hysteresis in the case of slightly different baselines, as in experiment II.

It was already mentioned in Footnote 5 that it is obvious that the H model does not apply when stimuli are in different registers, as in experiment I by Hermes and Rump. In those cases the low pitch levels in each individual stimulus necessarily provide some kind of reference for judging the height of the high pitch levels. As mentioned earlier, however, the standard deviations under the different-register conditions were much larger than under the equal-register conditions. But also here, the fall significantly lent more prominence than the rise and the rise–fall.

In their second experiment, Hermes and Rump used stimuli which were in the same register. They found that falls lend more prominence than rises and rise–falls. Furthermore, varying the timing of the falls did not affect the prominence they lent as compared to the prominence lent by a rise–fall. Early rises lent the same prominence as a rise–fall and for rises starting later than the vowel onsets, the prominence was reduced or became difficult to compare to the prominence lent by a rise–fall. In these experiments, comparisons always included a rise–fall with only one high pitch level on the topline, viz., the level of the vowel of the accented syllable. It is difficult to say how the strategy as given by the H model will have affected these results.

In conclusion, the role of high and low pitch levels in the perception of prominence of pitch-accented syllables is not clear yet. Under the same-register condition, high pitch levels seem to contribute more to prominence than low pitch levels do, although the role of the latter is still significant. As mentioned before, in the case of stimuli in the same register, the strategy of matching high pitch levels is a plausible strategy, too, according to Kutik et al. (1983). This strategy, according to Kutik, then implicates that the two stimuli presented as a pair are considered by the subjects as constituents of one large phrase. Redoing the experiments with stimuli in different registers may therefore be more conclusive, although it has to be taken into account that standard deviations will be much larger in that case. A reason for the larger standard deviations seems to be the fact that the strategy of matching high pitch levels, i.e., making use of an additional cue for equal prominence, is no longer possible.

ACKNOWLEDGMENTS

We would like to thank Cynthia Grover, Jacques Terken, Bob Ladd, and René Collier for their most valuable comments on earlier versions of this paper. Also, the comments of John Kingston are greatly appreciated.

1Whether a pitch movement is accent lending is determined by its position in the syllable and by whether it is abrupt or gradual; although this is largely language specific (for Dutch, see ‘t Hart et al., 1990). In pitch-accent languages like Dutch or English, a syllable is called “pitch-accented” if it is prominent relative to its neighbors mainly due to the presence of an accent-lending pitch movement, which is mostly accompanied by a longer syllable duration, higher amplitude, and/or better vowel quality. Changes in pitch are normally perceptually more salient than changes in one of the other accent-lending factors. Speakers may vary the degree of prominence of accented syllables for linguistic (e.g., focus) or paralinguistic (e.g., emphasis, emotions) reasons. Accent as used in the present paper should not be confused with lexical stress. A stressed syllable can or cannot be marked by an accent-lending pitch movement. Beckman (1986) presents evidence that longer duration, higher amplitude, and/or better vowel quality are acoustically correlated with stress, at least in English (and in Dutch). As Pierrehumbert (1980; also Beckman and Pierrehumbert, 1986) makes clear, these properties also correlate with the presence of a pitch accent, but only so because the pitch accent lines up with a stressed syllable. 2According to House (1990), pitch movements are only perceived as such if they start after the vowel onset, if they have a minimum duration of about 100 ms, and if they take place in spectrally stable signals like vowels. Otherwise, rises are “perceptually recoded” to high pitch levels and falls to low pitch levels.

3The vocalic nucleus is defined as the area of maximum sonority of the syllable. It may also be called the syllabic nucleus. Note, however, that the vocalic nucleus is not the same as the P center of the syllable (e.g., Pomponio-Marschall, 1990). The former is necessarily in the vocalic part of a syllable, whereas the P center is defined as the perceptual moment of occurrence of the syllable, which may be even before the onset of the vocalic part of the syllable.

4The hypothesis that listeners focused on the high pitch level in the first syllable of the stimulus containing the fall and on the high pitch level in the accented second syllable of the stimulus containing the rise does not lead to the predicted predictions. The predicted shift would then be equal to 3dT. That means that it would be only about half as large as the shift found in experiment I and in the experiments by Hermes and Rump.

5Note that, in this extreme form, the H model claims that listeners do not use any information about low pitch levels when adjusting prominences to be equal. From earlier experiments with stimuli in different registers (Hermes and Van Gestel, 1991; Hermes and Rump, 1994), however, it is obvious that listeners do use information about baselines when determining the prominence lent by pitch movements.

References


