

The perceptual prominence of fundamental frequency peaks

C. Gussenhoven^{a)}

Department of English, University of Nijmegen, Erasmusplein 1, 6525 HT Nijmegen, The Netherlands

B. H. Repp

Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511-6695

A. Rietveld

Department of Language and Speech, University of Nijmegen, Erasmusplein 1, 6525 HT Nijmegen, The Netherlands

H. H. Rump

Institute for Perception Research, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

J. Terken^{b)}

Institute for Perception Research, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

(Received 2 April 1996; revised 10 June 1997; accepted 1 August 1997)

Five perception experiments were conducted that investigated how the perceived prominence of F_0 maxima in accented syllables in Dutch is affected by the variation of F_0 minima that is supposed to relate to variation in global pitch range. The purpose of the first two experiments was to test the predictions of a model in which the reference line is directly given by an interpolation between observable F_0 minima. The results showed that the model was inadequate, and confirmed earlier research suggesting that the reference line is calculated in a less direct way. The next three experiments investigated the role of the F_0 of the unaccented portions of speech at the beginning ('onset') and at the end ('offset') of the contour, and show that only the (low) onset is used to calibrate the reference line. The results also suggest that longer onsets affect the abstract reference more than do shorter onsets. © 1997 Acoustical Society of America. [S0001-4966(97)03711-9]

PACS numbers: 43.71.Es, 43.70.Fq [WS]

INTRODUCTION

Speakers can vary the prominence of pitch accents by varying the height of associated fundamental frequency (F_0) maxima to express different degrees of emphasis (Lieberman and Pierrehumbert, 1984; Gussenhoven and Rietveld, 1988). Likewise, listeners' prominence judgments reflect the role of F_0 variation in relation to prominence variation (e.g., Gussenhoven and Rietveld, 1988). (We use F_0 as a shorthand for the periodicity of the speech signal.)

Several studies have been conducted over the years that investigate the relation between F_0 variation and perceived prominence, in order to determine the function relating F_0 variation and perceived prominence and to develop a metric for prominence. The first relevant study was conducted by Pierrehumbert (1979). She presented listeners with utterances containing two pitch accents realized by means of F_0 peaks, and asked them to judge which one had higher pitch. It was found that peaks later in the utterance tended to be lower than peaks earlier in the utterance when they were judged as having equal pitch; in other words, if the peaks had equal frequency, the later peak was judged to have higher pitch. This finding was explained in terms of listeners' expectations about declination, the tendency for F_0 events later in the phrase or utterance to occur lower in the frequency

range than linguistically equivalent events earlier in the domain.

In subsequent studies, beginning with Rietveld and Gussenhoven (1985), listeners were asked directly to make judgments about prominence, and several aspects of the pitch contours were varied to determine how the properties of the pitch contours feed listeners' expectations, with an emphasis on the conditions under which linguistic events later in the utterance are considered equivalent to those earlier in the utterance. These studies make it clear that listeners somehow estimate the prominence of the peak on the basis of the pitch characteristics of the contour around it, but the answer as to how listeners do this remains elusive. Among other things, it has remained unclear what the relevant characteristics of the surrounding contour are: Are they neighboring F_0 maxima, or F_0 minima, or both?

One opinion is that the F_0 minima in the phrase or utterance provide a baseline which serves as a reference for the evaluation of the F_0 maxima (Fujisaki and Hirose, 1984; 't Hart *et al.*, 1990). Due to declination, this baseline tends to be lower at the end of the phrase or utterance than at the beginning. The slope of the baseline is usually expressed in terms of semitones (st) or hertz (Hz) per s. Evidence for the relevance of baseline information in perception has been provided by Terken (1991). He employed stimuli similar to those of Pierrehumbert (1979), consisting of reiterant speech containing two accented syllables ('ma MA ma ma MA ma'). Artificial pitch contours were synthesized with F_0 maxima P_1 and P_2 in the accented syllables, respectively.

^{a)}Electronic mail: c.gussenhoven@let.kun.nl

^{b)}Electronic mail: terken@ipo.tue.nl

Across stimuli, the slope of the baseline and $P1$ frequency were manipulated. Listeners went through sets of utterances, and for each set selected the utterance in which the prominence of $P2$ matched that of $P1$ most closely (“adjustment task”). Within sets, $P1$ and the slope of the baseline were held constant and only $P2$ was varied. By comparing the results of stimulus sets with and without baseline declination, it could be shown that the information contained in $F0$ minima is indeed relevant to the perceived prominence associated with $F0$ peaks. However, the results argued against the conclusion that there is a direct relation between the observable excursion size of a pitch change and its perceived prominence.

A different opinion is that $F0$ minima play no or only a minor role in determining a reference (Pierrehumbert, 1979; Liberman and Pierrehumbert, 1984; Beckman and Pierrehumbert, 1986; Ladd, 1993). These authors argue that $F0$ minima in natural utterances seldom fit nicely on a straight line, so that there would be no straightforward way for listeners to extract the baseline. Moreover, although there is evidence that the frequency of $F0$ minima varies as a function of the position of the phrase in a coherent text, (Sluijter and Terken, 1993), it appears that listeners are relatively insensitive to variations in the frequency of $F0$ minima as compared to that of $F0$ maxima (Sluijter, 1991). Hence, these authors argue that listeners determine the prominence of $F0$ peaks relative to a more abstract reference level or reference line. However, the various papers contain little concrete indication as to how listeners might determine this more abstract reference for any particular utterance.

Further problems in the modeling of the relation between $F0$ variation and prominence variation arise when the variation in overall pitch range across and within speakers is considered. The same pitch contour can be produced by speakers with different overall pitch ranges. Likewise, the same speaker can utter a sentence at different heights in his overall range. Somehow, listeners compensate for these sources of pitch variation in making prominence judgments, but it remains unclear which information listeners use to estimate the height of a pitch contour in the speaker’s overall pitch range.

Summarizing, there is convincing evidence that $F0$ minima contain information that is relevant to making judgments about the prominence associated with $F0$ maxima, but it remains unclear to what extent variations in the frequency of $F0$ minima affect prominence judgments, and also whether all $F0$ minima contribute equally. The design of the experiments in Terken (1991) was limited in two ways. In the first place, a binary comparison was made between stimuli with and without baseline declination. Second, in the stimulus set in which the rate of declination was varied, it covaried with $P1$ height. Thus additional data were needed to assess the effect of variation of baseline declination and the influence of variation in peak height separately.

The remainder of the paper is structured as follows. Section I addresses in more detail the influence of $F0$ minima on the perceived prominence of $F0$ maxima. More specifically, two experiments are described which address the question to what extent variation in the rate of baseline declination af-

fects prominence judgments, relative to the contribution of peak height variation. Among other things, the results led us to abandon the assumption that the observable baseline provides the listener with a reference line against which local $F0$ maxima are evaluated. However, they confirmed the hypothesis that both $F0$ maxima and minima are relevant to the prominence judgments. Section II describes three experiments that further explored the separate contributions of the contour beginning and the contour end to the perceived prominence of the peak(s). In Sec. III, we provide a tentative answer to the question of how the position of the contour in the overall pitch range of the speaker is estimated by the listener on the basis of the $F0$ minima around the peak, and how this estimate is used in a prominence judgment task.

I. THE RATE OF BASELINE DECLINATION

The experiments described here differ in several respects from the experiments reported in Terken (1991, 1994). First, the earlier experiments focused exclusively on male speech, and as a result, the endpoint of the declining baseline was always at a relatively low frequency. The present study attempted to rectify this situation by comparing female speech (experiment I) and male speech (experiment II). Second, a simple meaningful sentence was used instead of reiterant speech. This was expected to help subjects maintain a linguistic frame of reference, at the risk of introducing other confounding variables. Third, in experiment I two tasks were compared: a single-trial relative prominence judgment task (henceforth, “judgment task”) and a “quasi-adjustment task”, as used by Terken (1991). The judgment task, in particular, was thought to discourage a literal pitch matching strategy. Finally, a group of relatively naive subjects (i.e., not including any speech researchers) was used. In Terken’s (1991) study there seemed to be a tendency for the most experienced listeners to show the smallest effects. The segmental homogeneity of his stimuli, his adjustment paradigm, and his listeners’ experience may all have encouraged an analytic listening strategy, counteracting the adoption of the “linguistic mode” most favorable to prominence judgments.

A. Experiment I

1. Method

A female Dutch speaker spoke the sentence “Amanda gaat naar Malta” (“Amanda goes to Malta”), with pitch accents on –man– and –mal–, in a neutral fashion without special emphasis on “Amanda” or “Malta.” The speech was recorded in a recording studio using high-quality equipment. The utterance was input to a computer, digitized at 20 kHz, and low-pass filtered at 10 kHz. It was then LPC coded (24 coefficients, 10 ms frames), and $F0$ was determined using the method of subharmonic summation (Hermes, 1988). A stylized $F0$ contour was fitted to the utterance following the procedures described in ’t Hart *et al.* (1990), to obtain an estimate of the baseline declination. This gave a declination rate of 2.4 st/s for the baseline, and an end frequency of 164 Hz. Phonetic segment boundaries were measured using a waveform editing program. The total utterance was 1635 ms in duration; the critical /man/ and /mal/ syllables had dura-

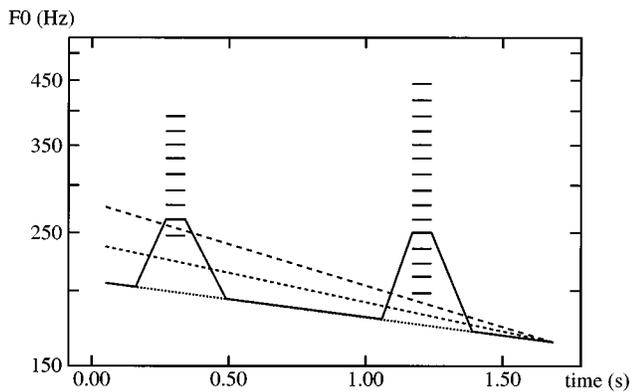


FIG. 1. Schematic illustration of stimulus parameters in experiment I (female speaker). The solid line represents an actual F_0 contour for the sentence *Amanda gaat naar Malta* ("Amanda goes to Malta") with pitch accents on *-MAN-* and *-MAL-*. The dashed lines represent the declination lines for different declination conditions. Also the ranges for the peaks of the pitch accents on *-MAN-* and *-MAL-* are indicated.

tions of 291 ms (/m/: 87 ms; /an/: 204 ms) and 268 ms (/m/: 106 ms; /al/: 162 ms), respectively. There were three voiceless gaps: for /χ/ and /t/ in "gaat," and for /t/ in "ta"; the last one occurred during the fall of P_2 .

The basic F_0 contour used in the experiment consisted of a baseline with two rise-and-fall combinations superimposed on it, for P_1 and P_2 , respectively (cf. Fig. 1). A set of 105 F_0 contours was constructed by combining 3 declination rates, 5 P_1 heights, and 7 P_2 heights. The three declination rates were 2.4 (original), 3.9, and 5.5 st/s, with a fixed end frequency of 164 Hz. The rises began 30 ms before the vowel onsets of the accented syllables and lasted 110 ms; the falls began 50 ms after vowel offset and lasted 150 ms; in between the rise and fall there was a flat top lasting 70 ms (these values were inspired by the properties of the original contour). Different P_1 and P_2 values were produced by manipulating the size of the rise-and-fall combinations (i.e., the

frequency interval between onset and offset of the rise and the fall), while keeping the duration constant.

The combinations of declination rate, P_1 height, and P_2 height were not strictly orthogonal because the range of P_1 heights necessarily had to covary to some extent with the height of the baseline, and the range of P_2 heights had to be approximately centered around each P_1 height to yield an unbiased estimate of the point of equal prominence. The precise stimulus design is shown in Table I. As can be seen in the table, 9 P_1 values were assigned to the 3 baselines in 3 overlapping sets of 5 (giving 15 combinations of P_1 and declination rate), and 15 P_2 values were assigned to the 9 P_1 values in 9 overlapping sets of 7 (giving 105 possible combinations of P_1 , declination rate and P_2). P_1 values ranged from 247 to 392 Hz in steps of approximately 1 st, and P_2 values ranged from 198 to 444 Hz, also with a 1-st step size. These parameters are illustrated in Fig. 1.

For the judgment task, the 105 stimuli and 5 fillers were recorded on digital tape five times, each time in a different random order. The fillers represented successive steps in the process of going from the natural contour to the basic artificial contour. Due to space constraints the results for the fillers will not be presented here. Each block of 110 stimuli lasted about 10 min. The interstimulus interval was about 4 s. A longer interval occurred after the 35th and the 70th trials, and a still longer interval in between blocks. The five test blocks were preceded by a practice sequence containing, in random order, the 30 stimuli in which P_2 had one of the extreme values in its 7 value range (i.e., all the first and last stimuli in the rows of Table I). Subjects were tested in two groups in a quiet listening room. The test tape was presented over earphones at a comfortable loudness level. Subjects were asked to judge for each stimulus which of the two words, "Amanda" or "Malta," was given more emphasis by the speaker. (The instructions pointed out that this was the same as judging the relative prominence of the accented syllables /man/ and /mal/.) The responses were made by writing

TABLE I. Stimulus design, showing the range of P_2 values for each P_1 value at different declination rates, and percentages of "1" responses (where "1" means that P_1 was judged to be more prominent) to each stimulus in the judgment task (Experiment I).

Decl. (st/s)	P_1 (Hz)	P_2 (Hz)														
		198	211	222	235	250	263	278	294	313	333	351	370	392	417	444
2.4	247	94	77	80	66	54	26	11								
	263		97	91	71	69	31	20	9							
	278			86	86	71	63	29	9	3						
	294				97	77	46	43	34	0	9					
	313					91	74	54	43	17	3	6				
3.7	278			94	97	80	63	49	17	11						
	294				89	80	77	51	34	6	9					
	313					94	91	77	46	26	3	0				
	333						91	83	66	31	11	3	14			
	351							97	77	49	34	6	3	3		
4.9	313					91	83	89	60	14	26	6				
	333						94	91	63	66	20	9	3			
	351							91	80	57	37	20	9	6		
	370								94	80	63	26	6	6	0	
	392									89	80	54	17	3	0	0

“1” or “2” on a prepared answer sheet. A forced choice was required on every trial. The experiment took a little over one hour, including instructions and a short break after the third block. The subjects were seven members of the research staff at IPO, four women and three men, all native speakers of Dutch, who volunteered to participate.¹ None of them was involved in speech research.

About two months later, the same subjects took part in the quasi-adjustment task, which lasted about 30 min. Subjects were tested individually in a quiet room with instructions similar to those in the judgment task. They sat in front of a computer screen and listened to the stimuli over earphones. Subjects worked through a series of runs. Within a given run, P_1 and declination were fixed and only P_2 varied; that is, a given run used the seven stimuli that had the same combination of P_1 and declination rate and differed only with respect to P_2 (one of the rows in Table I). In each run they “adjusted” the frequency of P_2 , by selecting the stimulus in the run in which the prominence of P_2 matched the prominence of P_1 most closely. The subject was first presented with the stimulus having the lowest (or highest) P_2 , and could then press any one of the digit keys between 1 and 9 on the computer keyboard to hear the next stimulus. Pressing key 5 corresponded to “replay the last stimulus,” pressing keys below 5 corresponded to selecting stimuli with lower P_2 if available (the farther away from 5 the larger the difference from the current stimulus), and pressing keys above 5 corresponded to selecting stimuli with higher P_2 than the current stimulus (again, the farther away from 5 the larger the difference). They completed 60 adjustment runs, 4 for each of the 15 combinations of P_1 height and declination rate, arranged in 4 blocks of 15 runs each. The adjustments were subsequently averaged over the 4 replications.

2. Results and Discussion

For the 105 test stimuli in the judgment task the number of “1” responses was tallied, added up across subjects, and converted into percentages. These percentages are shown in Table I. Although some subjects were not totally confident even at the extremes of the P_2 ranges, an orderly progression from “1” to “2” judgments can be seen in each of the 15 rows of the table.

Probit analysis (Finney, 1971) was used to estimate the 50% crossover point from “1” to “2” responses within each of the 15 series of 7 P_2 values, individually for each subject. In one anomalous instance, the first two percentages of a nonmonotonic response function were ignored in order to get a reasonable estimate. These estimates represent the points of equal judged prominence of P_1 and P_2 . Figure 2(a) shows these estimates, averaged across the seven subjects. Figure 3(a) gives the corresponding mean values for the adjustment task, which yielded such estimates directly. The results of the two experiments are quite similar. Despite some variability at the individual level, the overall results are remarkably orderly.

The average data show the expected $P_1 - P_2$ difference: in all cases P_2 was lower than P_1 when it was judged to be equal in prominence.² All seven subjects showed this overall effect in both tasks.

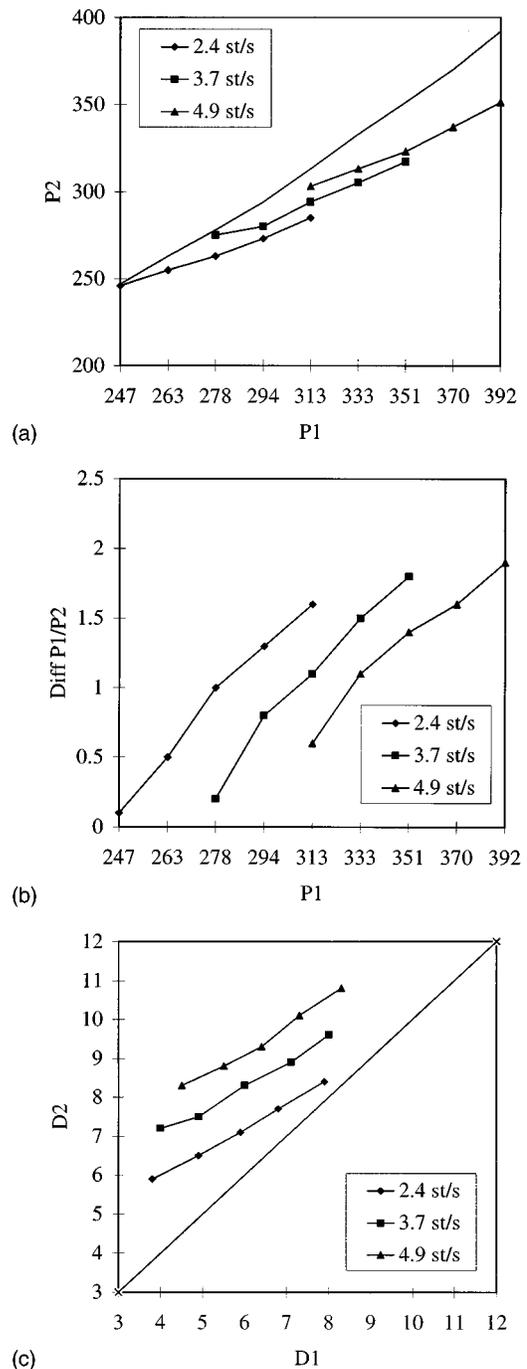


FIG. 2. Mean values of P_2 giving the same prominence as P_1 , as a function of P_1 height, for different declination rates (experiment I, judgment task). (a) Peak values for P_2 against P_1 ; the solid diagonal-like line represents the $P_2 = P_1$ curve; (b) Differences between peak values for P_1 and P_2 (Diff P_1/P_2 , in semitones) against P_1 peak values; (c) Excursion size D_2 (distance between peak P_2 and the interpolated baseline) against excursion size D_1 .

Figures 2 and 3 also show that the discrepancy between P_1 and P_2 (Diff P_1/P_2) increased with P_1 height [this was true both when expressed in Hz and st, see Figs. 2(b) and 3(b)]. The consistency of this effect was assessed by fitting straight lines to the five equal-prominence P_2 values obtained in each declination condition, separately for each subject (there were no systematic nonlinearities in any subject's data), and by examining the slopes of these lines. Nearly all

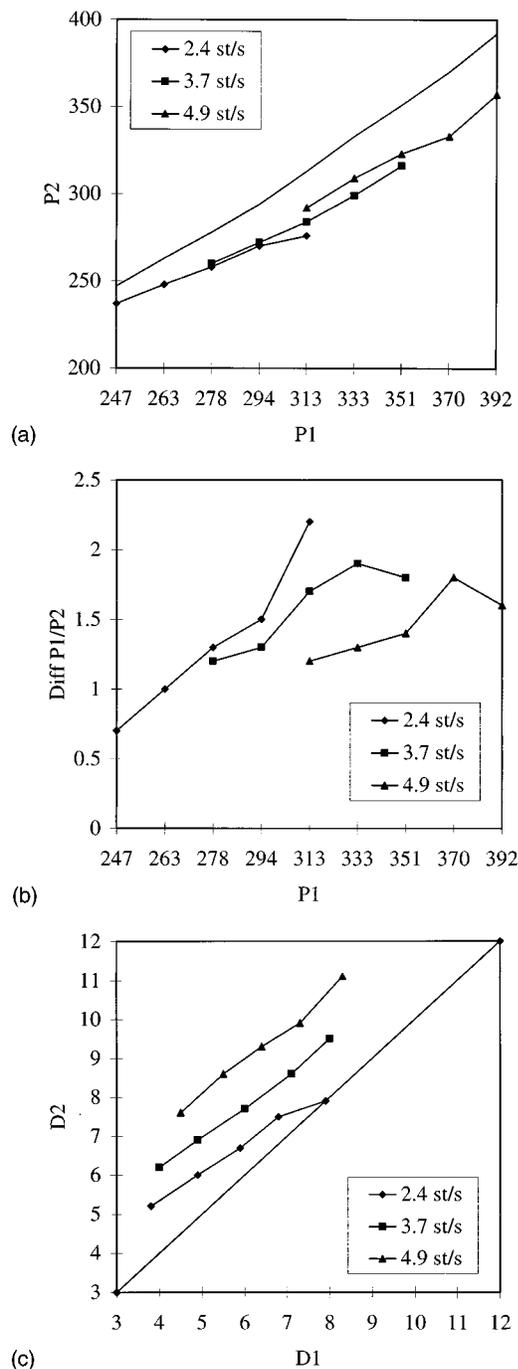


FIG. 3. Mean values of P_2 giving the same prominence as P_1 , as a function of P_1 height, for different declination rates (experiment I, adjustment task). (a) Peak values for P_2 against P_1 ; the solid diagonal-like line represents the $P_2 = P_1$ curve; (b) Differences between peak values for P_1 and P_2 (Diff P_1/P_2 , in semitones) against P_1 peak values; (c) Excursion size D_2 (distance between peak P_2 and the interpolated baseline) against excursion size D_1 .

were less than 1. The increase in the $P_1 - P_2$ difference with P_1 height was statistically reliable, as shown by a sign test ($z = 5.71$, $p < 0.001$). The slopes tended to be steeper in the adjustment task, but this task difference fell short of significance [$F(1,6) = 5.6$, $p < 0.06$].

Third, the ANOVA on the slopes of the lines fitted to the data points revealed a significant main effect of declination

rate [$F(2,12) = 4.3$, $p < 0.04$]: There was a significant trend for the slopes to get steeper as the declination rate increased. In other words, the increase in the $P_1 - P_2$ difference with P_1 height was more pronounced with low than with high declination rate. This effect did not interact with task either. However, the data suggest that the slope differences were essentially restricted to the adjustment task.

Finally, Figs. 2(c) and 3(c) show that the excursion sizes D_1 and D_2 , i.e., the distances between the peaks and the interpolated baselines, varied at different rates for P_1 and P_2 . This confirms earlier findings that equal prominence is not simply a matter of equal excursion sizes.

Our primary interest is in the effect of rate of declination on the relation between P_1 and P_2 . This effect was relatively small, as may be seen in Figs. 2(b) and 3(b) by connecting equal P_1 's: Diff P_1/P_2 decreased slightly when the baseline became steeper. That is, for a given P_1 , P_2 should be slightly higher in stimuli with steeper declination than in stimuli with less steep declination, in order to maintain equal prominence. The significance of this effect was tested in three ANOVAs on the P_2 estimates: two for the two sets of P_1 values that were shared by two declination conditions (278, 294, 313 Hz, and 313, 333, 351 Hz, respectively), and one for the single P_1 value that was shared by all three (313 Hz). The declination effect was significant in all three analyses [low versus medium declination rate: $F(1,6) = 13.4$, $p < 0.02$; medium versus high declination rate: $F(1,6) = 56.9$, $p < 0.0004$; all three: $F(2,12) = 17.6$, $p < 0.0004$], and there was no significant interaction with task or with P_1 height. Of course, there were highly significant main effects of P_1 height in the first two analyses, but the interactions of that effect with task were nonsignificant. The results will be further discussed in combination with those of experiment II.

B. Experiment II

1. Methods

A male Dutch speaker spoke the same sentence as in experiment I, "Amanda gaat naar Malta." The speech was digitized at 20 kHz and low-pass filtered at 10 kHz. The utterance was LPC coded (32 coefficients, 10-ms frames), and the F_0 contour was determined. This contour was then subjected to stylization, which was straightforward as the unaccented syllables could be fit easily by a single straight (i.e., exponentially) declining line. The declination rate was 3.9 st/s, which is almost exactly the rate predicted by the Dutch synthesis model (3.8 st/s), and the end frequency was 102 Hz. In contrast to the source utterance of experiment I, P_2 was much lower than P_1 in the natural utterance used to make the stimuli for experiment II. However, this was considered irrelevant, as F_0 was changed in the experimental stimuli. The total utterance was about 1410 ms in duration; the critical /m/ and /mal/ syllables had durations of about 230 ms (/m/: 80 ms; /an/: 150 ms) and 270 ms (/m/: 80 ms; /al/: 190 ms), respectively.

As in experiment I, a set of 150 F_0 contours was constructed by combining 3 declination rates, 5 P_1 heights, and 7 P_2 heights, selected from a range of 9 P_1 and 15 P_2 values. The two accent peaks were modeled on those of the

TABLE II. Range of $P2$ values for each $P1$ value in experiment II.

Decl. (st/s)	$P1$ (Hz)		$P2$ (Hz)																	
2.4	150	119	126	133	141	150	159	168												
	159		126	133	141	150	159	168	177											
	168			133	141	150	159	168	177	189										
	177				141	150	159	168	177	189	200									
	189					150	159	168	177	189	200	211								
3.9	168			133	141	150	159	168	177	189										
	177				141	150	159	168	177	189	200									
	189					150	159	168	177	189	200	211								
	200						159	168	177	189	200	211	225							
	211							168	177	189	200	211	225	238						
5.5	189					150	159	168	177	189	200	211								
	200						159	168	177	189	200	211	225							
	211							168	177	189	200	211	225	238						
	225								177	189	200	211	225	238	250					
	238									189	200	211	225	238	250	267				

female utterance of experiment I, and were similarly aligned with the segmental structure: the durations of rise, top, and fall portions were 110, 70, and 150 ms, respectively, for both peaks, with the rise starting 30 ms prior to vowel onset. The $F0$ of the top portion declined at the baseline rate, so that the maximum $F0$ occurred at the end of the rise. The temporal distance between $P1$ and $P2$ was 650 ms. $P1$ heights ranged from 150 to 238 Hz, and $P2$ heights from 119 to 267 Hz, in steps of approximately 1 st. The precise $P1$ and $P2$ combinations are presented in Table II. The three declination rates chosen were 2.4, 3.9 (original), and 5.5 st/s. The basic set up of the stimulus set is similar to that of experiment I (cf. Fig. 1).

Only an adjustment task was used in experiment II, as the two tasks in experiment I had yielded rather similar results. Instructions and testing procedure were identical to those in the adjustment task of experiment I. The subjects were ten members of the research staff at IPO, five women and five men, all native speakers of Dutch, who volunteered to participate. None of them was involved in speech research, and none had participated in experiment I.

2. Results

The results are shown in Fig. 4. As in experiment Ib, a strong effect of initial $P2$ height was noted: When the first stimulus in an experimental run contained the highest value of $P2$, adjustments were much higher than when the run started with the lowest $P2$ value. The data presented represent an average over both types of adjustment runs.

Again, $P2$ was always adjusted to a lower value than $P1$, which replicates the basic $P2 < P1$ effect [cf. Fig. 4(a)]. The effect was shown throughout by all but two subjects, who adjusted $P2$ to values higher than $P1$ when $P1$ was above 200 Hz. Furthermore, the majority of the subjects increased the $P1 - P2$ difference (expressed in st) with increasing $P1$, as can be seen by connecting equal $P1$'s in Fig. 4(b), although the effect was smaller and less orderly than in experiment I. A combined analysis with the slopes of the regression lines of individual subjects for the adjustment data

of experiment I did not yield a significant difference between experiments [$F(1,15) = 2.20$, $p = 0.16$], and overall the slopes were again less than 1 [$F(1,15) = 17.01$, $p < 0.001$]. Unlike that found in experiment I, the difference in slopes across declination conditions was not significant [$F(2,18) = 1.30$, $p = 0.30$].

As in experiment I, the effect of principal interest, that of rate of declination on the relation between $P1$ and $P2$, was relatively small [cf. Fig. 4(a)], and it was only marginally significant in ANOVAs on the $P2$ responses to the $P1$ values shared by the different declination conditions [low versus medium: $F(1,9) = 4.86$, $p < 0.06$; medium versus high: $F(1,9) = 7.77$, $p < 0.03$; all three: $F(2,18) = 3.10$, $p < 0.07$]. Also, as in experiment I, for each separate declination condition, the distance between $P1$ and the interpolated baseline ($D1$) covered a different range of values than the distance between $P2$ and the interpolated baseline ($D2$) [cf. Fig. 4(c)].

3. Discussion

First, the results replicate the well-known finding that $P2$ should be lower than $P1$ to give an impression of equal prominence, i.e., the declination effect. Second, the results confirm earlier findings that the $P1 - P2$ difference required for equal prominence increases as $P1$ increases, although the effect is smaller here than in Terken (1991), which matched the set-up of the current experiments most closely. Third, the effect of declination, which is of main interest here, is relatively small: For a given $P1$, a higher declination rate requires a small increase in $P2$ frequency to maintain equal prominence.

The finding that an increase in the rate of declination requires a small increase in $P2$ frequency to maintain equal prominence is in contrast with the findings of Terken (1991), where the effect of the presence of declination was to lower $P2$ relative to the "no declination"-condition. One possible explanation for this discrepancy is in terms of the different methods used. The current experiments differed from those in Terken (1991) with respect to both stimulus properties and

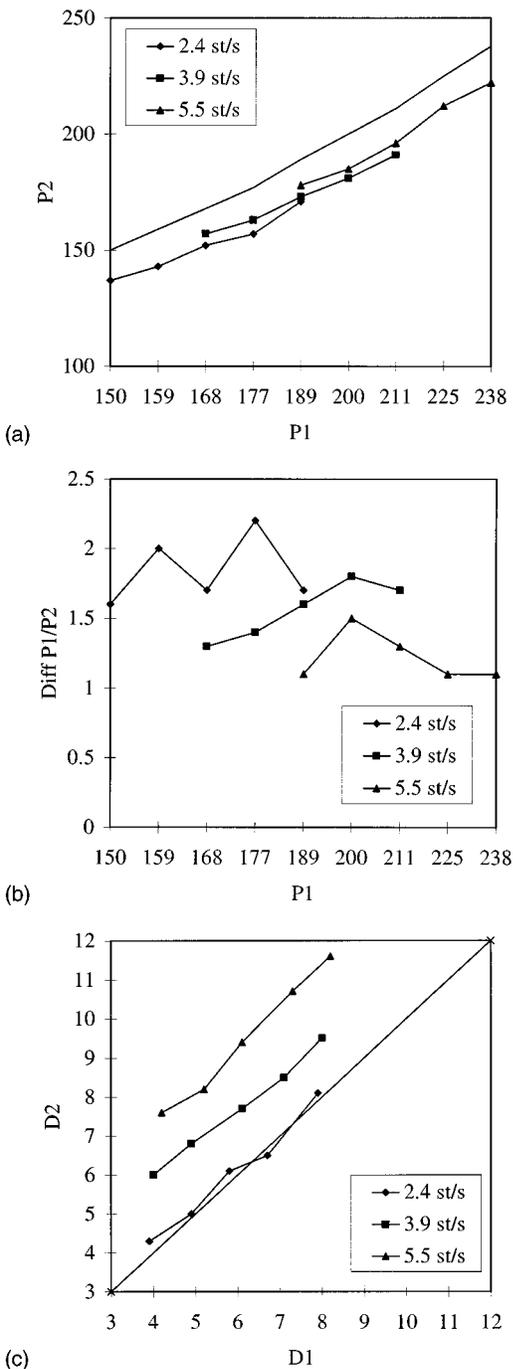


FIG. 4. Mean values of $P2$ giving the same prominence as $P1$, as a function of $P1$ height, for different declination rates (experiment II, adjustment task). (a) Peak values for $P2$ against $P1$; the solid diagonal-like line represents the $P2 = P1$ curve; (b) Differences between peak values for $P1$ and $P2$ (Diff $P1/P2$, in semitones) against $P1$ peak values; (c) Excursion size $D2$ (distance between peak $P2$ and the interpolated baseline) against excursion size $D1$.

presentation methods. In particular, whereas in the current experiments the stimuli representing the different declination conditions were presented in a random order in the same experiment, in the experiments described in Terken (1991) they were presented in separate experiments. This may have induced a different, more analytic mode of listening in the “flat baseline” condition than was used for the more natural “declination” condition, and $P2$ may have been adjusted to

higher values in the “flat baseline” condition than would have been the case under a more synthetic listening mode. Thus the two conditions in Terken (1991) may have provided qualitatively different outcomes, so that the comparison between the conditions of Terken (1991) would not be not relevant to our present concerns.

The results of experiment I and experiment II can be summarized as follows. If $P1$ is raised while keeping the baseline constant (so that the distance between the peak and the baseline increases), the difference between $P1$ and $P2$ that is required for equal prominence increases (a finding that was obtained only in experiment I). Second, if the declination becomes steeper while keeping $P1$ fixed (so that the distance between the peak and the baseline decreases), the $P1 - P2$ difference required for equal prominence decreases. Thus, there would appear to be a tendency for the difference between $P1$ and $P2$ to be proportionally related to the distance between $P1$ and the baseline, regardless of the position of the contour in the overall range of the speaker.

Our results argue against the hypothesis that prominence is related in a direct way to the distance between the peaks and the overt baseline, a conclusion that concurs with those of Pierrehumbert (1980:128) and Ladd (1993), who assume the reference is abstract. At the same time, the results provide evidence for the relevance of the $F0$ minima in the contour to the perceived prominence of $F0$ peaks. Therefore, we need to explore how listeners construct an abstract reference line on the basis of the physical characteristics of the contour. To this end, the role of $F0$ minima was investigated in a more piecemeal fashion. These investigations are described in Sec. II.

II. ANCHORING THE REFERENCE LINE

A logical step to take, one that emerges from the conclusions of Sec. I, is to test for the effect of the $F0$ minima at different locations in the contour and try to establish their effects independently of each other. In experiment III, we tested for the effect of the final low pitch. In experiment IV, we did the same for the beginning of the contour. Finally, in experiment V we again looked at the influence of the beginning of the contour in order to replicate the results of experiment IV. In addition, with experiment V we returned to the issue of the declination effect, and investigated it as a function of the length of the utterance before the pitch peak.

A. Experiment III: Offset height

The purpose of experiment III was to test the relevance of the low end point of the contour to the perceived prominence of the peak or peaks in it. While the results of experiments I and II suggest that a raising of the beginning of the contour reduces the prominence of peaks, no information has been obtained about a possible effect of the contour end. The depth of the downward trajectory at the end of a falling contour might well determine the perceived prominence of the final peak, or of all peaks, in the contour. Also, this effect may depend on how close the final peak is to the end of the contour.

1. Methods

A male speaker recorded the sentence *Dat geblaas de hele dag van die schapen daar* (“that bleating all day of those sheep there”) in a neutral, somewhat monotonous fashion, without placing particular emphasis on any of the words. The whole utterance had a duration of 2215 ms; the word *geblaas* had a duration of 480 ms, and the word *schapen* 440 ms. The syllable *blaas* had a duration of 294 ms ([a:]: 132 ms); the syllable *scha* had a duration of 293 ms ([a:]: 142 ms). The utterance was digitized at a sampling rate of 10 kHz, with the LP filter set at 5 kHz, LPC analyzed (18 coefficients, 10-ms frames). Our utterance served as the basis of three syntactically exclamatory sentences. After splicing off the portion of the speech file corresponding to *van die schapen daar*, the shortened speech file was used to create a set of one-peak stimuli, while the original speech file was used to create a set of one-peak stimuli as well as a set of two-peak stimuli. These three carrier sentences are listed below, with their durations; the capitalized syllables indicate the locations of the *F0* peaks in the artificial contours.

- (1) Dat geBLAAT de hele dag! (1255 ms) “that bleating all day”
- (2) Dat geBLAAT de hele dag van die schapen daar! (2215 ms)
- (3) Dat geBLAAT de hele dag van die SCHAPen daar! (2215 ms) “that bleating the whole day of those sheep there”

The artificial pitch contours contained a level baseline of 100 Hz. Peaks had flanks of 100 ms, and were aligned such that the *F0* maxima coincided with the midpoint in the accented vowel. Three values for the peaks were used: 140, 150, and 160 Hz. In the two-peak contour the values of *P1* and *P2* were the same. After the last peak, the pitch fell to a low point, to which we refer as the “offset.” The offset was also varied in three steps, 95, 87, and 80 Hz. Offsets and peaks were crossed, which led to nine stimuli per sentence (see Fig. 5).

Two audiotapes were prepared on each of which the 27 stimuli appeared three times in a random order. The tapes were presented through headphones in a language laboratory to 20 subjects, about half of whom listened to one version and the rest to the second version. All our subjects were students at the University of Nijmegen, who were paid a small fee for their services. They were given a written instruction sheet along with the response form.

Their task was to rate one of the accented syllables in each stimulus for general prominence. On their score sheets, the corresponding sentence was printed with the syllable to be scored printed in capitals. This syllable was always associated with a pitch peak in the stimuli. For each stimulus, they gave their judgment by putting a tick on a 100-mm scale, which had the words “little emphasis” printed on the left and the words “much emphasis” on the right. After 15 stimuli an anchor stimulus was included with one accented syllable. The prominence of this anchor stimulus was shown on the score sheets as being on the mid-point of the scale. The stimuli were preceded by ten test trials; the first two and the last two stimuli were fillers. The stimuli were presented

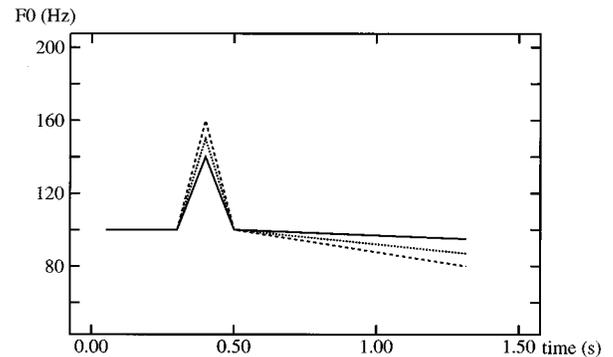


FIG. 5. Schematic representation of *F0* contours for stimuli in experiment III. The solid line represents an actual contour for the sentence *dat geblaas de hele dag* (“that bleating all day”) with a pitch accent on *-BLAAT-*. Dashed lines represent alternative *F0* peaks for the pitch accent and *F0* trajectories for the offset.

in blocks of ten, with an ISI of 6 s, and intervals of 10 s between blocks.

2. Results

The three sentences differed in the length of the post-peak stretch, i.e., the part following the target accent on “geblaas,” and in the presence in the post-peak stretch of a pitch peak on “schapen.” Figure 6 gives the mean prominence ratings for the different combinations of offset and peak height for the three sentences. As can be seen, for all sentences the perceived prominence increased as the peaks increased in fundamental frequency. However, for no sentence does there appear to be an effect of offset height. This is confirmed by two analyses of variance (repeated measures). One analysis was performed on the scores for *geblaas* in sentences 1 and 2, averaged over the three replications, with the fixed within-subjects factors “Peak Height,” “Offset Height,” and “Distance to Offset.” The other analysis was performed on the scores of sentences 2 and 3, averaged over the three replications, with the fixed within-subjects factors “Peak Height,” “Offset Height,” and “Number of Peaks.” In both analyses, only the factor “Peak Height” was significant: $F(2,60) = 38.67$ and $F(2,60) = 27.87$, respectively; Huynh-Feldt corrected *p* value < 0.05 .

3. Discussion

The perceived prominence of an *F0* peak appears to be independent of the pitch of the final low end point of the contour. This finding may seem surprising, because the final low point would appear to be such an obvious candidate as a cue to the speaker’s overall pitch range. It has been found to be a relatively invariant value for speakers in laboratory speech (Lieberman and Pierrehumbert, 1984), and so might well have served as an anchor point for the speaker’s pitch range (assuming it is calculated afresh for each stimulus). The fact that it is not rules out models of prominence perception which depend on a reference line—whether overt or abstract, whether horizontal or descending—which is determined by the contour end. With hindsight, however, it is not implausible that the final low pitch should not be used as an

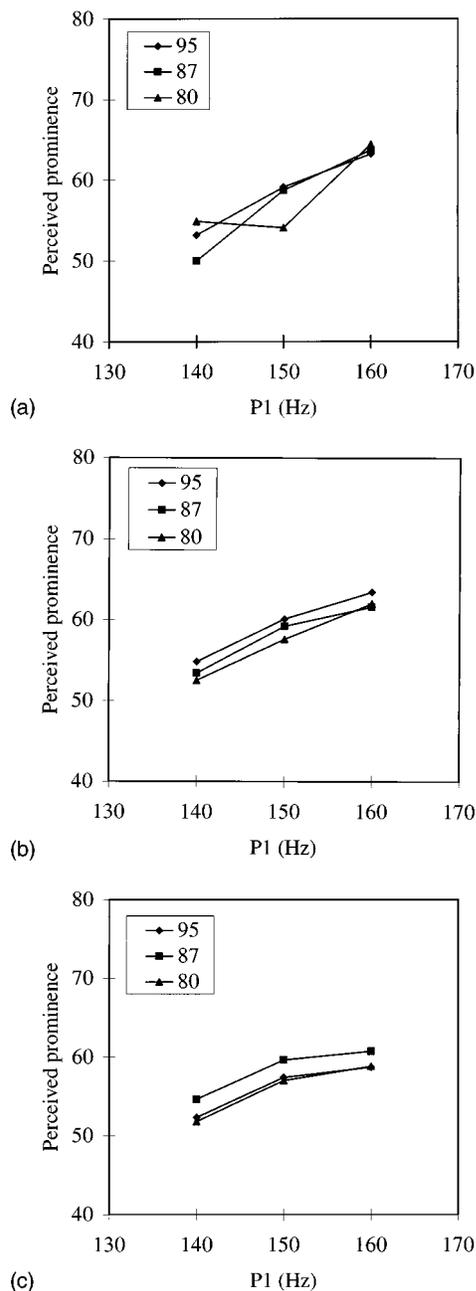


FIG. 6. Mean perceived prominence ratings for *geBLAAT* (“bleating”) on a 100-mm scale as a function of Peak Height and Offset frequency (experiment III). (a) Sentence 1, with short post-peak stretch: *dat geBLAAT de hele dag*; (b) Sentence 2 with long-post-peak stretch: *dat geBLAAT de hele dag van die schapen daar*; (c) Sentence 3 with long post-peak stretch containing an accent on *SCHApEn* (“sheep”): *dat geBLAAT de hele dag van die SCHApEn daar*.

anchor point for pitch range. Final low pitch serves as a cue to finality. Although the distinction between final phrases and nonfinal phrases largely correlates with contour type, with a boundary high tone signalling nonfinality and a boundary low tone signalling finality (Geluykens and Swerts, 1994), at least some of the variation found by Swerts *et al.* (1994) between final and nonfinal phrases in spontaneous dialogue could be attributed to the occurrence of a “half-fall” versus a fall to the bottom of the pitch range. Moreover, there is evidence that the final fall is used by speakers

to signal the end of their turn at talk, with deeper falls being more likely to lead to turn taking by the listener (Beattie *et al.*, 1982). Thus the results of experiment III can be explained by assuming that the final low pitch is itself evaluated on the basis of the listener’s knowledge of the speaker’s pitch range. Obviously, in this scenario, the overall pitch range needs to be known to the listener before the degree of finality of the speaker’s utterance can be evaluated on the basis of the final low pitch.

An additional reason for assuming that the final low end point is a poor anchor for the reference line is that it varies much less under changes in register, i.e., the position of the contour in the speaker’s overall range, than other aspects of the contour (Ladd and Terken, 1995; Mozziconacci, 1995). Thus information situated elsewhere in the contour would seem preferable to anchor the reference line.

B. Experiment IV: Onset height

The purpose of experiment IV was to test the relevance of the initial pitch of the contour to the perceived prominence of the peak or peaks in it. We use the term “onset” to refer to the stretch between the contour beginning and the rise towards the accent peak. The pitch of the onset might serve as a cue to the contour’s reference line, with higher onsets signalling higher reference lines, which might consequently lead to less perceived prominence of the peak. Since any effect of the onset might depend on its length, we also varied the distance between the contour beginning and the peak.

1. Methods

We used the following sentences, obtained from the original utterance used in experiment III, as the basis of our stimuli:

- (1) *Dat geBLAAT de hele dag!* (1255 ms) “that BLEATING all day”
- (2) *Dat geblaat de hele dag van die SCHApEn daar!* (2215 ms) “that bleating all day of those SHEEP there”

The peaks were varied in three steps, with peak values of 140, 150, and 160 Hz, as before. These steps were crossed with four onset heights, 100, 108, 116, and 122 Hz. After the last peak, all contours descended to a low pitch of 87 Hz. The set-up is shown schematically in Fig. 7. The same 20 raters who participated in experiment III, were asked to judge the prominence of *geblaat* in the short sentence and *schapen* in the long sentence. In other respects, methods were similar to those in experiment III.

2. Results

Figure 8 gives the average ratings for the different combinations of onset and peak height for the different sentences. An analysis of variance (repeated measures) was performed on the data, with the fixed within-subject factors “Peak Height,” “Onset Height,” and “Sentence.” “Peak Height” and “Sentence” were significant: $F(2,38) = 50.42$, $p < 0.01$ and $F(1,19) = 35.04$, $p < 0.01$ (Huynh–Feldt corrected). Thus as expected, variation in peak height affected perceived prominence. In addition, the peak on *geBLAAT* in

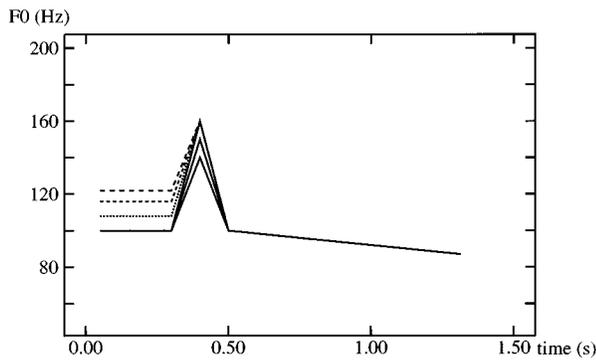


FIG. 7. Schematic representation of F_0 contours for stimuli in experiments IV and V. The solid line represents the contour for the sentence *dat geblaas de hele dag* (“that bleating all day”) with the three peaks for the pitch accent on *-BLAAT-*. Dashed lines represent the four trajectories for the onset with which the three peaks were crossed.

the sentence with the short onset was on average less prominent than that on *SCHApEn* in the sentence with the long onset [60.3 vs 71.8, cf. Fig. 8(a) and (b), respectively]. This main effect may point to an effect of onset length, but it may also be due to influences not controlled for in the experiment, like the relative durations or loudness levels of the syllables *blaas* and *scha-*. What is of interest in these data is the interaction between “Sentence” and “Onset Height:”

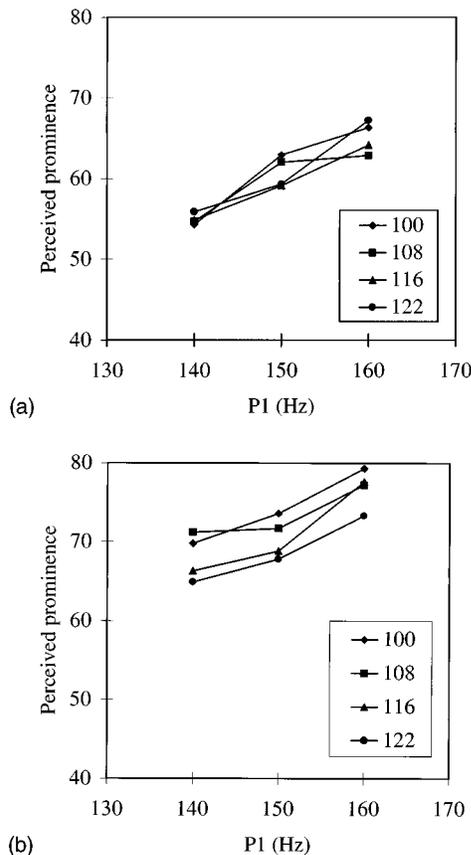


FIG. 8. Mean perceived prominence ratings on a 100-mm scale for *geBLAAT* (“bleating”) as a function of Peak Height and Onset Height, for different sentence lengths. (a) Short sentence, with short onset (Sentence 1): *dat geBLAAT de hele dag*; (b) Long sentence, with long onset (Sentence 2): *dat geblaas de hele dag van die SCHApEn daar*.

$F(3,57) = 3.95, p = 0.018$. The perceived prominence of the peak after the long onset decreased when the onset was raised. That is, the higher the onset, the less prominence the raters perceived on the peak. In the utterance with the short onset, there is no such effect of onset height.

3. Discussion

The results of Experiment IV show that the height of the onset may affect the perceived prominence of the peak. The way in which the perceived prominence depends on the onset, with higher onsets reducing the prominence of the peak, suggests that the listener uses the onset as a cue to the contour’s reference line. However, it would not appear as if it is literally the starting point of the contour, to the exclusion of the rest of the onset, that is relevant. An interpretation of the initial pitch of the contour as an anchor point for a reference line is inconsistent with our finding that the stimuli with the short contour showed no onset effect. The results rather suggest that a longer stretch of speech is needed in order to estimate the reference line. This is not surprising: the pitch of an initial portion of, say, 50 ms is probably an unreliable basis for making such an estimate. Frequently, such a brief stretch represents the rather variable rising slope up to a pitch peak whose exact beginning will be influenced by such entirely incidental factors as the presence of an unstressed syllable before the peak, the presence of a voiced onset in the accented syllable, or even whether the syllable nucleus consists of a long or a short vowel. The aim of experiment V, which we report in the next section, was to see if this reference line descends as a function of time.

C. Experiment V: Onset length, or the declination effect revisited

The purpose of experiment V was, first, to replicate the “Onset Height” effect found in experiment IV, and second, to establish whether there is a difference in prominence between peaks following short onsets and peaks following long onsets. If so, this would support the hypothesis that the prominence of peaks is estimated on the basis of a reference line which declines as a function of time. The results of Pierrehumbert (1979), which showed that the perceived pitch of a later peak is greater than that of an earlier peak, suggest that this is in fact so. However, in that experiment, the timing of the peak was confounded with its serial position, and although Gussenhoven and Rietveld (1988) found a declination effect in one-peak contours, we felt that the declination effect needed to be replicated with one-peak contours in order to establish that it is elapsed time, rather than preceding intonational structure, that causes it. If it is found that later peaks have greater perceived prominence than earlier peaks even after unaccented, level-pitched stretches of speech, then the case for time dependence becomes stronger. In addition, we wanted to test whether the effect of onset height that was found in experiment IV could be replicated. Specifically, we wanted to see if the two effects, the declination effect and the onset effect, could be shown to be independently active in the same perception experiment.

1. Methods

In order to produce the sentence with the short onset while controlling the segmental composition of the accented words, we excised the fragment *de hele dag*, which was exactly 500 ms long, from the speech file containing the sentence *Dat geblaas de hele dag van die schapen daar* (sentence 2 in experiment IV). In this way we obtained two sentences, which served as the basis of our stimuli:

- (1) *Dat geblaas van die SCHAPen daar* (1715 ms) “that bleating of those SHEEP there”
- (2) *Dat geblaas de hele dag van die SCHAPen daar* (2215 ms) “that bleating all day of those SHEEP there”

The peaks in the artificial contours were varied in three steps (140, 150, 160 Hz), the onsets in four steps (100, 108, 116, 122 Hz), while the offset was 87 Hz in all stimuli, exactly as in experiment IV. The 24 stimuli were presented to 30 subjects from the same population as before. In other respects, the methods were the same as those in experiments III and IV.

2. Results

Figure 9 gives the mean prominence ratings for the different combinations of onset and peak height for the different sentences. An analysis of variance (repeated measures) was carried out on the scores averaged over the three replications, with three fixed within-subject factors: “Peak Height,” “Onset Height,” and “Onset Length;” the factor “Subjects” was regarded as a random factor. All three main effects were significant: “Peak Height” [$F(2,58)=63.4$, $p<0.01$], “Onset Height” [$F(3,87)=16.1$, $p<0.01$], and “Onset Length” [$F(1,29)=20.2$, $p<0.01$]. Also, the interaction effect between “Onset Length” and “Onset Height” was significant [$F(3,87)=9.5$, $p<0.01$].

In addition to the effect of “Peak Height,” which caused higher peaks to be judged as more prominent, the factor “Onset Height” was significant: When the onset was raised, the prominence of the peak decreased. This replicates the result of experiment IV. The factor “Onset Length,” too, had the predicted effect: the later peak had greater perceived prominence (61.4 for the short onset versus 63.9 for the long onset). The effects are shown in Fig. 9. The interaction effect between “Onset Length” and “Onset Height” means that raising the onset did not have the same effect for short and long onsets. Inspection of Fig. 9(a) and (b) suggests that the depressing effect of raised onsets on the perceived prominence of the peak was more pronounced for long onsets than for short onsets: for onset heights 100, 108, 116, and 122 Hz, respectively, mean prominence ratings were obtained of 64.3, 63.3, 58.2, and 60.0 for the short onset and 66.7, 64.6, 64.1, and 60.5 for the long onset.

3. Discussion

The results of experiment V replicated those of experiment IV. First, as we expected, higher peaks are perceived to be more prominent. Second, onset height affects the prominence judgments for an F_0 peak, for onsets that were longer than 400 ms. We interpret this result as the effect of raising

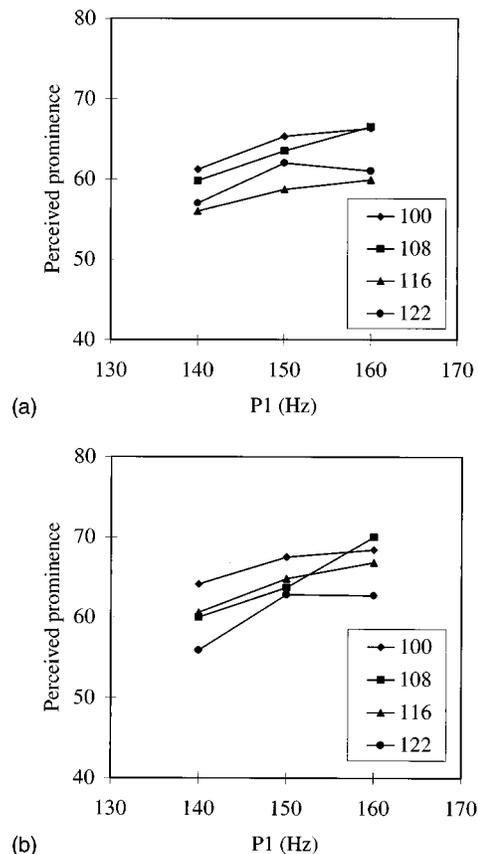


FIG. 9. Mean perceived prominence ratings on a 100-mm scale for *SCHAPen* (“sheep”) as a function of Peak Height and Onset Height, for different onset lengths. (a) Short onset (Sentence 1): *dat geblaas van die SCHAPen daar*; (b) Long onset (Sentence 2): *dat geblaas de hele dag van die SCHAPen daar*.

the reference line that listeners use to estimate the prominence of the peak. This reference line can be determined better as the listener has more information about the onset pitch.

An additional finding is that the length of the onset makes a difference. When the onset is longer, the perceived prominence of the peak is greater: 61.4 for short onsets versus 63.9 for long onsets. The finding that distance from onset is relevant can be interpreted as a replication of the declination effect reported for English by Pierrehumbert (1979). However, in her pitch range experiment, the effect of declination was shown in two-peak stimuli of the same length: The second peak had greater perceived pitch than the first when F_0 values of the peaks were the same (for the 71-Hz pitch range, cf. Pierrehumbert’s Fig. 3). The results of our experiment V show that the declination effect is purely length dependent.

An interaction effect that might have been expected is that between “Peak Height” and “Onset Length.” In experiments I and II it was found that a higher P_1 required a larger physical difference between P_1 and P_2 in order to obtain equal prominence. From this it can be concluded that greater P_1 prominence is correlated with a more steeply declining abstract reference line. Accordingly, we might expect the difference in perceived prominence between the low peaks in the short stimuli and the low peaks in the long

stimuli to be smaller than that between the high peaks in the short stimuli and the high peaks in the long stimuli. It should be borne in mind, however, that the effect of "Onset Length" was itself rather small. Further data are needed to demonstrate an even subtler interaction between onset length and peak height.

The unexpected interaction between "Onset Length" and "Onset Height" implies that the effect of a short onset is not the same as that of a long onset. While the results are extremely orderly, with prominence increasing with "Peak Height," decreasing with "Onset Height," and increasing with "Onset Length," the detailed results show that the prominence judgments for peaks in the short stimuli with the highest onset (122 Hz) are higher than expected. Possibly, the reversal of the "Onset Height" effect in the short stimuli is due to listeners' uncertainty about the category of the onset pitch. In the descriptions of Dutch intonation, a low onset contrasts with a high onset ('t Hart *et al.*, 1990; Gussenhoven, 1988), with a high onset causing the contour to sound more lively and thus likely to evoke greater perceived prominence for the F_0 peaks.³ Conceivably, our highest onset in the short sentence was ambiguous between a categorically low and a categorically high onset.

III. GENERAL DISCUSSION AND CONCLUSION

We have reported five experiments that sought to answer the question how listeners determine the prominence of F_0 peaks in one-peak and two-peak intonation contours. The following main findings were obtained:

- (1) Prominence relationships in the female and male pitch ranges are perceived in similar ways (experiments I and II).
- (2) For two peaks P_1 and P_2 , with P_2 occurring later in the utterance than P_1 , P_2 should be lower than P_1 in absolute frequency in order to be perceived as having the same prominence as P_1 (experiments I and II).
- (3) A single peak P_1 with a fixed frequency is perceived as more prominent if it occurs later in the utterance (experiment V).
- (4) A change in the distance between P_1 and the observable baseline by increasing the rate of baseline declination does not require a concomitant change in the distance between P_2 and the baseline in order to preserve equal perceived prominence (experiments I and II).
- (5) The height of the contour's final low pitch has no effect on the perceived prominence of the peaks in the contour (experiment III).
- (6) Increasing the frequency of P_1 while keeping the baseline fixed requires a small increase in the difference between P_1 and P_2 in order to maintain equal prominence; this (small) effect is more pronounced for the female than for the male speaker (experiments I and II).
- (7) Increasing the rate of baseline declination (i.e., raising the onset) while keeping P_1 fixed requires a small decrease in the difference between P_1 and P_2 in order to maintain equal prominence; this effect applies both to the female and male speaker, but again it is a small effect (experiments I and II).
- (8) Raising the onset for a given P_1 leads to a reduction of the perceived prominence for P_1 (experiments IV and V)
- (9) The effect of onset height on perceived prominence depends on the amount of onset information. Raising the onset affects the perceived prominence for a peak only with long onsets (over 400 ms) (experiments IV and V).

With these results, we can now address the question of how the perceived prominence of F_0 peaks relates to characteristics of the F_0 contour.

The first factor affecting perceived prominence is the position of the F_0 peak in the utterance: peaks in utterances with long onsets are judged to be more prominent than peaks in utterances with short onsets (finding 3). Likewise, in utterances containing two peaks, the second peak should be lower than the first peak in order to be perceived as having the same prominence (finding 2). These findings provide further evidence for Pierrehumbert's (1979) proposal that listeners expect F_0 to decline in the course of the utterance as a compensation for actual F_0 declination. It may be captured by postulating a reference line that declines as a function of time. The fact that actual F_0 in the stimuli in experiments IV and V did not decline, implies that the reference line is abstract rather than being induced by the observable F_0 declination rate.

The second factor affecting perceived prominence is related to the size of the F_0 change underlying the peak. This size can be altered in two ways. One is by altering the height of the peak while keeping its base fixed. The other is by altering the base while keeping the peak fixed. In experiments IV and V, both modifications were found to influence perceived prominence: if the size of the change is reduced by lowering the peak or raising the onset, prominence will decrease; if the size of the change is increased by raising the peak or lowering the onset, prominence will also increase (provided the onset is long enough to give rise to a stable pitch percept; finding 7).

The results of experiments I, II, IV, and V [and also the results in Terken (1991)] suggest that it is not the actual size of the change, measured as the distance between the peak and the observable baseline, which determines the degree of perceived prominence. For instance, in experiments I and II the actual size for P_1 could be changed rather drastically by changing the slope of the baseline, while only small adjustments in the size of the change underlying P_2 were needed in order to maintain equal prominence (cf. Figs. 2, 3, and 4). Instead, it would appear that what counts is the distance to the abstract reference line as postulated above, which determines the prominence associated with a peak. This was also the central assumption in the model proposed by Ladd (1993).

In Ladd's (1993) model, the perceived prominence of P_1 and P_2 is determined by their distance to the abstract reference line. Ladd assumes that the slope of the abstract reference line is fixed regardless of the contour's position in the speaker's overall pitch range; furthermore, he assumes that equal prominence is obtained when the pitch maxima have equal distance to the reference line.

The latter assumption is contradicted by the current findings in two ways. First, it predicts that raising or lowering the reference line up or down the frequency axis (for instance, by raising or lowering the onset) leaves the peaks equidistant, no matter what metric (Hz, st, or ERB rate) is employed, and that therefore declination rate should not affect the $P1-P2$ difference giving equal prominence. The current results show that there is an effect of rate of baseline declination on the $P1-P2$ difference, even though it is relatively small, so that this prediction of Ladd's model is refuted. Second, it predicts that an increase in the height of $P1$ should cause an equal increase in the height of an equally prominent $P2$: The distance between $P2$ and the reference line ($D2$) should still increase by the same amount as that between $P1$ and the reference line ($D1$) when $P1$ is raised. The results of experiments I and II show that this is not the case. Rather, it appears that, if the distance between the peak and the abstract reference line ($D'1$) is large, the difference between $P1$ and $P2$ needed to maintain the impression of equal prominence is larger than when $D'1$ is small. The finding that the size of the $P1$ change ($D'1$) affects the difference between $P1$ and $P2$ required to obtain equal prominence is not compatible with the hypothesis that the degree of perceived prominence is a direct function of the distance between the peak and the abstract baseline. However, as observed in the discussion of experiment V, these results can quite naturally be interpreted to mean that the slope of the abstract reference line is steeper as the $P1$ prominence is higher. Equivalently, we could assume a fixed slope for the reference line, in combination with an isoprominence line through $P1$ and $P2$ which declines more steeply as the prominence of $P1$ is higher, in line with Pierrehumbert (1980: Chap. 3).

From these observations the contours of a model for explaining prominence perception can be derived. Perceived prominence appears to be related to the distance between the peak and an abstract reference line, which declines at a rate which is independent of the observable baseline. Increases in the distance between $P1$ and the abstract baseline lead the listener to expect the reference line to descend at a faster rate. We cannot at this point provide quantitative estimates of the different effects; further and more fine-grained experiments should provide the required data.

A further finding which deserves attention is the fact that the effect of onset height was found only in the condition in which the distance between the beginning of the contour and the peak was at least two words long. Apparently, the listener requires a certain amount of speech to be able to make a reliable estimate of the onset pitch. This has two interesting implications. In the first place, it means that the initial one or two syllables apparently do not contain sufficient information to accurately estimate onset height, in contrast with what would be expected on the basis of many current theories. In the second place, it implies that in utterances with short onsets, where the listener apparently is unable to estimate onset height, the abstract baseline is not anchored at all in the actual contour. At present we cannot answer the question how the position of the contour in the speaker's overall range is estimated in these cases; it seems likely that, lacking suf-

ficient information, the listener assumes a default range. This would also apply in cases where there is no onset information at all, such as in the case where the first syllable is accented.

In summary, it appears that perceived prominence of accent peaks is determined mainly by the distance between the peak and an abstract reference line, which declines as a function of time and is anchored in the onset of the utterance (provided that the onset is long enough to enable the listener to obtain a reliable estimate of onset height). In addition, there are clear indications that greater $P1$ prominence is associated with more steeply declining reference lines.

ACKNOWLEDGMENTS

Experiment I was conducted while Bruno Repp was a Visiting Research Fellow at the Institute for Perception Research (IPO). The support of the Technical University Eindhoven and of the University of Nijmegen in the preparation and running of the experiments is gratefully acknowledged. Special thanks are due to Leo Vogten for his cheerful and effective help with various technical problems in the case of experiments I and II and to Rob van den Berg for his contribution to the successful completion of experiments III, IV, and V. We thank the colleagues and students who served as subjects in the rather tedious experiments. We are indebted to Mary Beckman, Bob Ladd, and an anonymous reviewer for helpful comments on draft versions.

¹Two additional subjects were tested, one in the judgment task only and the other in both tasks. The former almost always judged $P2$ to be more prominent than $P1$; the later produced extremely variable results. Their data were excluded. One of the authors (B. Repp), though not a speaker of Dutch, took part in both tests as a pilot subject and produced results consistent with the subject averages presented below.

²In anticipation of this effect, the range of $P2$ values assigned to each $P1$ value had been centered around a value slightly below $P1$. It might be argued that this asymmetry in stimulus design actually caused the effect. This seems unlikely, however, in view of the randomization of all stimuli in the judgment task, and the consistent presence of a $P1-P2$ difference in earlier studies. The same difference was obtained on those trials in the adjustment task where the subjects started with the highest $P2$ value, although it was smaller.

³A reviewer has pointed out to us that the reduced prominence on *scha-* in the long onset condition may be due to the listener's interpretation of the preceding *-blaat* as having a H^* pitch accent. Under that view, the prominence of $P2$ might be heard as reduced either because there is an accent-to-accent declination effect, or a phrase-to-phrase declination effect, as opposed to a purely time-dependent declination effect. There are two reasons why this alternative account is improbable. The interpretation of the level onset pitch as containing a H^* pitch accent should be most likely when the onset is highest. However, in the short condition, it is precisely the highest onset that shows a reversal of the reduction of the perceived prominence on $P2$, a finding that cannot be squared with this alternative explanation. Also, the alternative explanation might lead one to expect a more abrupt effect of "Onset Height" rather than the gradual effect that is observed.

- Beattie, G. W., Cutler, A., and Pearson, M. (1982). "Why is Mrs. Thatcher interrupted so often?," *Nature* (London) **300**, 744–747.
- Beckman, M., and Pierrehumbert, J. (1986). "Intonational structure in English and Japanese," *Phonology Yearbook* **3**, 255–310.
- Finney, D. J. (1971). *Probit Analysis* (Cambridge U.P., New York).
- Fujisaki, H., and Hirose, K. (1984). "Analysis of voice fundamental frequency contours for declarative sentences of Japanese," *J. Acoust. Soc. Jpn. (E)* **5**, 233–242.
- Geluykens, R., and Swerts, M. (1994). "Prosodic cues to discourse boundaries in experimental dialogues," *Speech Commun.* **15**, 69–77.
- Gussenhoven, C. (1988). "Adequacy in intonation analysis: The case of Dutch," in *Autosegmental Studies on Pitch Accent*, edited by H. van der Hulst and N. Smith (Dordrecht, Cincinnati), pp. 95–121.
- Gussenhoven, C., and Rietveld, A. C. M. (1988). "Fundamental frequency declination in Dutch: testing three hypotheses," *J. Phonetics* **16**, 355–369.
- 't Hart, J., Collier, R., and Cohen, A. (1990). *A Perceptual Study of Intonation: An Experimental-Phonetic Approach to Speech Melody* (Cambridge U.P., U.K.).
- Hermes, D. J. (1988). "Measurement of pitch by subharmonic summation," *J. Acoust. Soc. Am.* **83**, 257–264.
- Ladd, D. R. (1993). "On the theoretical status of 'The Baseline' in modeling intonation," *Lang. Speech* **36**, 435–451.
- Ladd, D. R., and Terken, J. M. B. (1995). "Modelling intra- and interspeaker pitch range variation," in *Proceedings of XIIIth International Congress of Phonetic Sciences, Vol. II*, edited by K. Elenius and P. Brandnerud (Stockholm University, Stockholm), pp. 386–389.
- Liberman, M., and Pierrehumbert, J. (1984). "Intonational invariance under changes in pitch range and length," in *Language Sound and Structure*, edited by M. Aronoff and R. Oehrle (MIT, Cambridge, MA), pp. 157–233.
- Mozziconacci, S. J. L. (1995). "Pitch variations and emotions in speech," in *Proceedings of XIIIth International Congress of Phonetic Sciences, Vol. I*, edited by K. Elenius and P. Brandnerud (Stockholm University, Stockholm), pp. 178–181.
- Pierrehumbert, J. B. (1979). "The perception of fundamental frequency declination," *J. Acoust. Soc. Am.* **66**, 363–369.
- Pierrehumbert, J. B. (1980). *The Phonology and Phonetics of English Intonation* (MIT, Cambridge, MA).
- Rietveld, A. C. M., and Gussenhoven, C. (1985). "On the relation between pitch excursion size and prominence," *J. Phonetics* **13**, 299–308.
- Rump, H. H., and Collier, R. (1996). "Focus conditions and the prominence of pitch-accented syllables," *Lang. Speech* **39**, 1–17.
- Sluijter, A. M. C. (1991). "Een perceptieve evaluatie van een model voor alinea-intonatie met synthetische spraak" (A perceptual evaluation of a model for paragraph intonation with synthetic speech). Internal Report 801. Institute for Perception Research, Eindhoven, The Netherlands.
- Sluijter, A. M. C., and Terken, J. (1993). "Beyond sentence prosody: Paragraph intonation in Dutch," *Phonetica* **50**, 180–188.
- Swerts, M., Collier, R., and Terken, J. (1994). "Prosodic predictors of discourse finality in spontaneous monologues," *Speech Commun.* **15**, 79–90.
- Terken, J. (1991). "Fundamental frequency and perceived prominence of accented syllables," *J. Acoust. Soc. Am.* **89**, 1768–1776.
- Terken, J. (1994). "Fundamental frequency and perceived prominence of accented syllables II: Nonfinal syllables," *J. Acoust. Soc. Am.* **95**, 3662–3665.