Journal of Phonetics (1994) **22**, 87–99

Influence of adjacent pitch accents on each other's perceived prominence: two contradictory effects

D. Robert Ladd, Jo Verhoeven* and Karen Jacobs†

Department of Linguistics and Human Communication Research Centre, University of Edinburgh, Edinburgh, Scotland

Received 7th July 1992, and in revised form 2nd July 1993

In a previous study by Gussenhoven & Rietveld published in the *Journal of Phonetics* in 1988, in which listeners rated the prominence of the second pitch accent in synthetic utterances containing two pitch accents, it was found that decreasing the F_0 on the *first* pitch accent peak (P_1) induced *lower* prominence ratings on the second (P_2). The present study replicates and extends this somewhat unexpected finding. Two groups of naïve subjects rating two slightly different sets of materials showed the Gussenhoven–Rietveld effect (i.e., they gave lower ratings of prominence for P_2 as the F_0 of P_1 decreased). However, they did so only when P_2 itself was below about 145 Hz in a male speaking range. For higher values of P_2 the Gussenhoven–Rietveld effect was reversed: lower prominence ratings were given to P_2 as the F_0 of P_2 *increased.* Moreover, a group of trained listeners (phoneticians and speech researchers) failed to show the Gussenhoven–Rietveld effect at any level of P_2. A broadly phonological explanation for the effect (and for its failure to occur under certain conditions) is proposed, such that, in normal range, the prominences of accents are not assessed individually, but are evaluated as a group to assess the utterance's overall degree of emphasis.

1. Introduction

1.1. *Prominence and pitch range*

One of the universally acknowledged functions of accent, at least in stress–accent languages, is to render syllables and words prominent or salient in the stream of speech. It is commonly assumed, moreover, that accentual prominence or salience is relative, and a matter of degree: some accents are more prominent than others. However, despite a substantial literature on this topic (e.g., 't Hart, 1981; Liberman & Pierrehumbert, 1984; Rietveld & Gussenhoven, 1985; Hermes & van Gestel,

* Present affiliation: Universitaire Instelling Antwerpen, Initiatief Nederlands, Universiteitsplein 1, B-2610 Wilrijk, Belgium.
† Present address; Sopot, Poland.

0095-4470/94/010087 + 13 $08.00/0

1991), there is no detailed understanding of what makes an accented word or syllable prominent, and fundamental theoretical and empirical issues over the "relative" nature of prominence remain unresolved.

In a general way, it is known that the perceived prominence of an accented syllable is affected by *pitch range*: the greater the pitch range, the more prominent the accent. Unfortunately, "pitch range" is not a well-defined concept. Specifically with respect to accented words or syllables, it can be defined in at least two ways: as the size of the F_0 excursion that accompanies the accent, or as the relative height of the F_0 peak. Opinions differ about the appropriateness of these two definitions.

The obvious attraction of defining accentual pitch range in terms of F_0 excursion is that it makes it relatively easy to normalize away from interspeaker differences of overall F_0 level (e.g., male–female differences). An actual change in pitch can be measured and directly compared to other actual changes in pitch, irrespective of speaker or utterance. This is not to minimize the difficulty of deciding on the appropriate units to use for such measurements [though Hermes & van Gestel (1991) make a fairly convincing case for using the ERB units of a critical-band scale]. The point is simply that, once the units are chosen, the actual pitch excursions can be measured and compared fairly readily.

By contrast, if we define accentual prominence in terms of peak height, we must assume that peak height is perceived relative to something else in the contour, such as the peak of another accent, the adjacent valleys, or the speaker's "baseline". Comparing accentual prominence from one utterance to another or from one speaker to another must therefore be based on theoretical assumptions about how this relative perception works, and/or on a model of how accentual peaks are scaled as a function of one or more utterance-specific or speaker-specific reference values. This then raises a number of difficulties that are in no way resolved (see Ladd, 1992, and especially Ladd, 1993, for further discussion).

Despite the theoretical difficulties with a prominence measure based on relative peak height, there is at least some empirical evidence in favor of such an approach. In a study that is important for what it failed to show, 't Hart (1981) investigated how well listeners could distinguish pitch excursions of different sizes. Working on the explicit assumption that excursion size rather than relative peak height is the essential cue to accentual prominence, 't Hart set up his experiments so as to prevent listeners from comparing peak heights, and forcing them to rely on pitch excursion alone. Under these circumstances all listeners had a great deal of difficulty discriminating excursion size differences of less than three semitones, and many listeners discriminated much less successfully than that. Rietveld & Gussenhoven (1985) showed, however, that in a prominence-rating task where listeners were able to compare peak heights most listeners could easily distinguish differences of prominence corresponding to excursion size differences of only 1.5 semitones. In the same vein, Liberman & Pierrehumbert (1984) argue that in their results perceived prominence correlates better with peak height than with excursion size.

The most that can be said with certainty at this point in the development of our understanding is that there is some relationship between perceived prominence and pitch range, but that further empirical data and further attempts at modeling the overall use of the voice pitch range in ordinary speech will be required before the relationship becomes completely clear. The present study is intended to contribute new empirical data to this line of work.

1.2. *Background to the study: the Gussenhoven–Rietveld effect*

A central empirical question, and one on which there is only a limited amount of rather conflicting evidence, is the effect of one accent on the perceived prominence of another accent in the same phrase or utterance. Given an utterance with two accents, what effect will changes on one accent have on the prominence of the other? If the second accent is a strong "contrastive stress", will increasing the prominence on the first accent make the second sound "less contrastive"?

Specifically, consider a short utterance of the form *da-DAH*$_1$*da-da-da-DAH*$_2$*-da*, where *DAH*$_1$ and *DAH*$_2$ represent accented syllables, and *da* unaccented ones. Suppose that in a given rendition of the utterance, *DAH*$_2$ has a perceived prominence of *P* units on some appropriate scale. Now let us hold the phonetic realization of *DAH*$_2$ constant, and vary the pitch range of *DAH*$_1$, as shown in Fig. 1. Let us then ask experimental listeners for judgments about the prominence of *DAH*$_2$ as the pitch range of *DAH*$_1$ varies.

There are three logically possible types of results. First, the experimental manipulation of *DAH*$_1$ might have no effect at all on the perceived prominence of *DAH*$_2$. Second, the two accents might serve as mutually defining frames or reference, so that as one goes higher the other is perceived as lower, and vice versa. Third, the two accents might contribute to some global impression of pitch range, so that as one goes higher the other is perceived as higher as well. We think that most readers will find the second possibility—which is analogous to a psychophysical "contrast effect"—the most likely of the three.

There is, however, preliminary evidence that the third possibility is what actually happens. In a study by Gussenhoven & Rietveld (1988) that largely dealt with other issues, listeners were asked in several different experiments to rate (on a 10-point scale) the prominence of accents whose pitch range had been experimentally manipulated. Not surprisingly, all experiments show a virtually linear increase in the average perceived prominence as a function of increases in the pitch range of the accent whose prominence was being rated. However, in one experiment (Section 4.1), Gussenhoven & Rietveld independently manipulated the pitch range on two accent peaks in the same utterances. Here they found that, if the pitch range on the

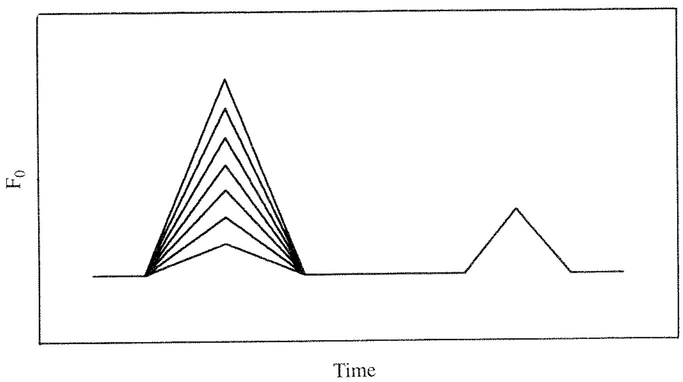

Figure 1. Schematic representation of a set of experimental two-accent F$_0$ contours in which the second accent is held constant and the first varied systematically.

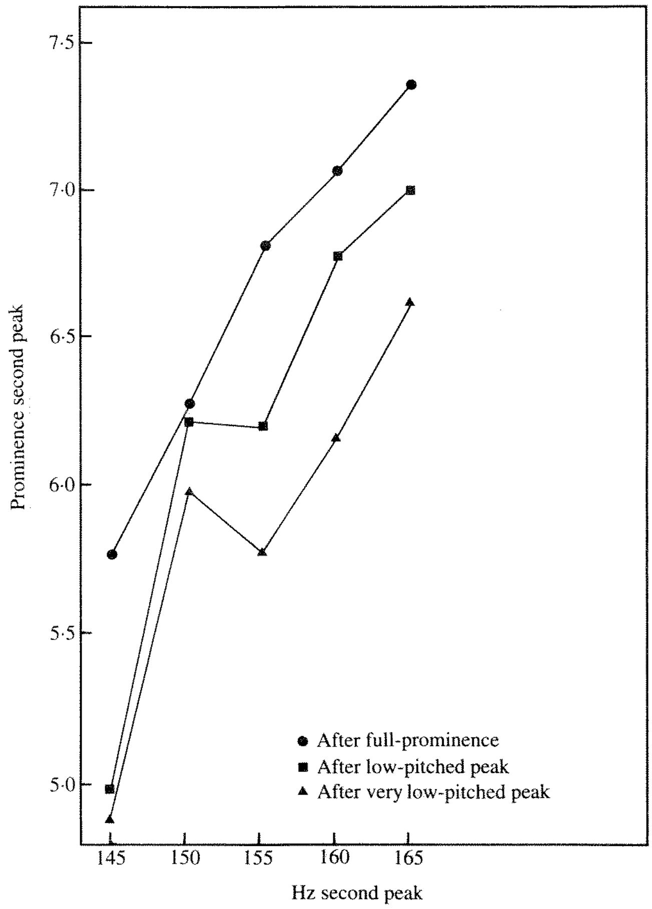

Figure 2. The Gussenhoven–Rietveld effect. For a given F_0 value of P_2, the perceived prominence of P_2 is lower when P_1 is lower. [Reproduced with permission from Gussenhoven & Rietveld (1988)].

first accent of an utterance is reduced, the perceived prominence of the *second* accent is *reduced as well.* The relevant data from Gussenhoven & Rietveld are shown in Fig. 2. In what follows we shall refer to this finding as the "Gussenhoven–Rietveld effect".

Gussenhoven & Rietveld acknowledge that this effect is unexpected, but they do not follow it up in their published paper. Moreover, since the finding was tangential to the main point of their study, it can scarcely be regarded as an established finding. However, it seems a relatively straightforward task to replicate it, and this is what is undertaken in the experiment described in the following section.

2. Experiment 1

2.1. *Method*

The general approach we took to replicating the Gussenhoven–Rietveld effect is that shown above in Fig. 1. Given a set of stimulus utterances, in which the first accented syllable (henceforth P_1) varied on a continuum of peak F_0 and the second (henceforth P_2) was held constant, we asked listeners to judge the prominence of the *second* peak. The hypothesis based on the Gussenhoven–Rietveld effect was that, as the peak F_0 on P_1 increased, the perceived prominence of P_2 would increase as well.

2.1.1. *Speech materials*

The experimental utterance was derived from a single token of the sentence *The melon is yellow*, spoken by an educated male native speaker of Standard Scottish English. This sentence was chosen because its two stressed syllables (a) do not contain any obstruents, which tend to perturb the F_0 course for non-intonational reasons, and (b) contain the same vowel with the same following consonant. This makes listeners' judgments less likely to be affected by intrinsic F_0 effects. We preferred a natural utterance to Gussenhoven & Rietveld's reiterant speech because we felt it would make the task more natural.

The utterance was recorded on professional equipment in the recording studio of the Edinburgh University Phonetics Laboratory. The utterance was digitized at 20 000 Hz and low-pass filtered at 8500 Hz. Fundamental frequency was extracted using the API command of the Interactive Laboratory System (ILS) signal processing package, and modified using the ILS-compatible FRED program written by Kim Silverman; the stimulus utterances were then resynthesized using ILS.

While the basic design of the experiment calls for a single stimulus continuum, with variable P_1 and constant P_2, we actually created two continua, with the same range of P_1 values but different values of P_2. We did this on the assumption that, if P_2 were always exactly the same, listeners attending to P_2 might become aware of this fact and would no longer give different prominence ratings to P_2. In effect, we intended that tokens from the two continua should serve as distractors for each other, and help to maintain the plausibility of the task of rating P_2's prominence.

The values chosen for P_2 were 140 Hz and 160 Hz. The contour onset was fixed at 116 Hz, the valley between the accents at 106 Hz, and the contour offset at 90 Hz (i.e., the contour had a moderately declining "baseline"). For each of the two values of P_2 there were 13 levels of P_1, ranging from 120 to 192 Hz in 6-Hz steps. There were thus 26 different stimulus types.

A test tape was created in which each stimulus type occurred three times, with all stimuli occurring randomly throughout the tape. There was a 3.5 s interval between adjacent stimuli. The tape also included a practise run consisting of 12 stimuli, and there were six additional stimuli at the very end of the main run which were intentionally disregarded in analysing the results. (This final "buffer" was included to counteract the possible effects of subjects' relaxing their attention on the last few items of the test; since the subjects had numbered answer sheets, they could readily tell when they were coming to the end of the sequence.) In all, the main run consisted of 84 $[(13 \times 2 \times 3) + 6]$ stimuli. The entire session lasted about 15 minutes.

2.1.2. *Procedure*

Condition 1—untrained listeners. This condition of the experiment, which used untrained listeners as subjects, is the study reported in Jacobs (1990). A total of 17 subjects took part, in three separate sessions with eight, five and four subjects, respectively. Most of the subjects were Edinburgh University undergraduates and all were native speakers of English. Many of them had a little practise with IPA transcription but were in no sense trained phoneticians.

At the beginning of the session subjects were given written instructions in which they were told that the experiment dealt with "how people hear accents on words". They were told that "words may be emphasized to convey a particular meaning" and that "this [emphasis or prominence] can sometimes be indicated in writing by capitalizing or italicizing a word". They were then asked to rate the "DEGREE of prominence" (emphasis in original written instructions) of the word *yellow* in the test sentence. For each stimulus there was a separate numbered line on the response sheet, with the word "... yellow" and a 10-point scale of boxes, labeled from 1 (least prominent) to 10 (most prominent).

We intentionally followed Gussenhoven & Rietveld in using a 10-point scale, in order to make our results more comparable with theirs. We are aware that there are certain methodological difficulties with the use of numbered scales for this sort of rating task. However, these difficulties should not be exaggerated: Gussenhoven & Rietveld (1988, Section 3.2) report statistical analyses of the validity and reliability of the rating scale judgments which seem to indicate that they give a reasonably accurate picture of the relative prominence of the stimuli.

The test tape was played over loudspeakers in a small language laboratory listening room; each subject was seated at a separate desk. The tape was stopped briefly after every 12 stimuli to allow subjects time to turn to the next page of their response forms.

Condition 2—phonetically trained listeners. The first attempt at replicating the results of Jacobs (1990) (i.e., Condition 1) was carried out during summer vacation, when there were no undergraduate students available as subjects. Consequently, our subjects were mostly colleagues, teaching staff and Ph.D. or post-doctoral researchers in linguistics, phonetics, and speech technology. Almost all were well trained in IPA transcription and/or listened regularly to synthetic speech. In all other respects Condition 1 and Condition 2 were identical. A total of 16 subjects, in two separate sessions, took part in Condition 2.

2.2. *Results*

2.2.1. *Condition* 1

In computing the results we took the mean of all the prominence ratings for each of the 26 (13 × 2) distinct stimulus types. Since each stimulus type occurred three times on the test tape and there were 17 subjects, the prominence ratings for each stimulus type are based on 51 observations. The standard deviations of these pooled means averaged 1.12 scale units and showed no significant variation with different values of P_1 or P_2.

We will consider the results separately for each of the two stimulus continua, viz., with lower (140 Hz) and higher (160 Hz) values of P_2. Although (as noted above) we intended the two continua simply as distractors for each other, it turned out that they produced divergent results, as can be seen in Fig. 3. In the continuum with the lower (140 Hz) value of P_2, there is a trend consistent with the Gussenhoven–Rietveld effect: as the F_0 on P_1 increases, the perceived prominence of P_2 increases

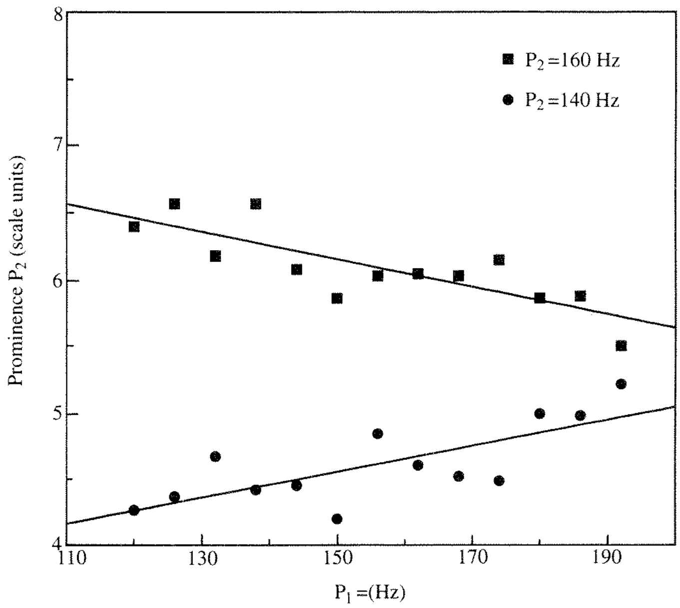

Figure 3. Mean prominence ratings for P_2 as a function of the F_0 of P_1 for two F_0 values of P_2 in Experiment 1, Condition 1 (naïve listeners). Lines are regressions over the means. The results for $P_2 = 140$ Hz seem to display the Gussenhoven–Rietveld effect, but those for $P_2 = 160$ Hz seem to show the reverse.

as well [on a simple linear regression, $F(1, 219) = 6.94$, $p < 0.01$]. One might be prepared to accept this as a replication of the Gussenhoven–Rietveld effect. However, in the continuum with the higher (160 Hz) value of P_2, no such effect can be observed. In fact, increases in the F_0 on P_1 appear to have produced a slight decline in the perceived prominence of P_2 [$F(1, 219) = 7.83$, $p < 0.01$]. The response curves for the two levels of P_2 thus converge as P_1 increases.

An Analysis of Variance (ANOVA) in which P_1 and P_2 were treated as independent variables seemed to confirm this convergence. There was a large main effect of the F_0 of P_2 on P_2's perceived prominence [$F(1, 180) = 467.43$, $p < 0.001$], but no main effect of P_1 [$F(12, 180) = 0.77$, ns]. However, there was a significant interaction of P_1 and P_2 [$F(12, 180) = 4.60$, $p < 0.001$], suggesting that the effect of P_1 on the perceived prominence of P_2 is different at different F_0 values of P_2.

2.2.2. *Condition* 2

Results, computed exactly as for Condition 1, are shown in Fig. 4. It can be seen that there is no convergence of the regression lines for the two different values of P_2, and that both regression lines fall (slightly and non-significantly). The standard deviations of the pooled means in Condition 2 were somewhat larger than in Condition 1, averaging 1.36 scale units as against 1.12; this difference was highly significant on a paired t-test ($t = 5.26$, $\mathrm{df} = 25$, $p < 0.0001$).

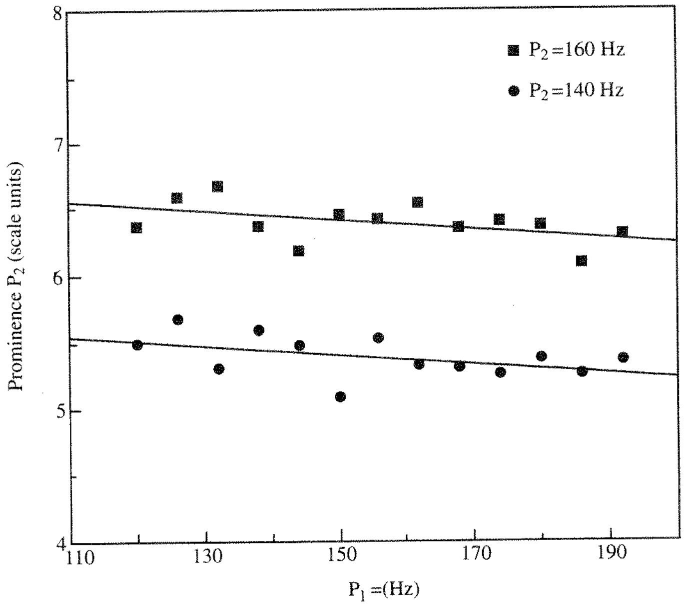

Figure 4. Results as in Fig. 3 of Experiment 1, Condition 2 (trained listeners). For neither value of P_2 is the Gussenhoven–Rietveld effect observed.

2.3. *Discussion*

It is by no means clear what to make of these findings. One defensible conclusion would be that the original Gussenhoven–Rietveld effect was simply an experimental artifact of some sort, and that the attempted replication has failed. In support of this conclusion one might cite the lack of agreement between the two continua in Condition 1, the tiny proportion of the variance accounted for by the convergence of the regression slopes (about 3% in both cases), the fact that no convergence and no Gussenhoven–Rietveld effect was found in Condition 2, and in particular the fact that the F_0 of P_2 itself clearly has a much greater effect on the perceived prominence of P_2 (that is, the perceived prominence of the 160 Hz P_2 is invariably greater than that of the 140 Hz P_2, irrespective of P_1).

However, suppose we take the apparent convergence of the two lines on Fig. 3 at face value. That is, let us attempt to explain the interaction of P_1 and P_2 observed in Condition 1. We hypothesize that in using two different values of P_2 we inadvertently introduced two distinct experimental conditions, one in which P_2 represents "normal" pitch range (140 Hz), and one in which it represents some sort of "emphatic" range (160 Hz). When P_2 is in normal range, we get the Gussenhoven–Rietveld effect: increases in the F_0 of P_1 produce increases in the perceived prominence of P_2. But when P_2 is emphatic, the Gussenhoven–Rietveld effect does not appear: instead, we get something like a psychophysical contrast effect whereby increases in the F_0 of P_1 bring about slight decreases in the perceived prominence of P_2.

It should be possible to test this hypothesis in an experiment in which both P_1 and P_2 are manipulated as independent variables. This was the goal of the second experiment, described in the next section of the paper. Before proceeding to describe that experiment, however, it is appropriate to discuss the failure of Condition 2 to replicate the interaction found in Condition 1. We think there is good reason to discount the significance of this failure and therefore good reason to assume that the interaction may be a genuine effect.

We suspect (partly on the basis of post-experiment conversations) that the more sophisticated subjects in Condition 2 had too much non-linguistic awareness of the phonetic properties of synthetically manipulated utterances and responded to the stimuli in a less natural way than the original naïve subjects. The subjects also tried far more actively than the naïve subjects to guess what the experiment was about, and at least a few of them succeeded. Differences between trained and naïve listeners have been found in other studies of the perception of prosodic properties, e.g., Boves *et al.* (1984); in the present case, these differences show up in closer analysis of the results.

For example, among the trained listeners, there were five who only ever used two or three levels of the rating scale; only one of the naïve subjects used such a narrow part of the scale. Two of these trained subjects showed no significant differences between their prominence ratings for low P_2 and high P_2, but the other three seemed to be making very fine discriminations: one trained subject used only points 5 and 6 on the scale, rating low P_2 as 5 in every instance but one, and rating high P_2 as 5 or 6 about equally—a highly significant difference.

Moreover, as noted above, the variance of the pooled means was significantly larger in Condition 2 than in Condition 1, which means that the trained listeners, as a group, responded less consistently than the naïve ones. Looking at the results for the naïve subjects individually, we find that all but four clearly exhibit the pattern of converging regression lines shown by the group as a whole. (Regression lines for the remaining four are approximately parallel.) Of the other 13 subjects, in one case the converging lines both have positive slope (i.e., this subject exhibits the Gussenhoven–Rietveld effect for both values of P_2), and in four cases both have a negative slope (i.e., these subjects do not exhibit the Gussenhoven–Rietveld effect at all), but even in these cases the convergence is clear. Among the trained subjects, by contrast, most show roughly parallel regression lines, with slopes ranging from gently positive (the Gussenhoven–Rietveld effect) to steeply negative; clear convergence is present only in four cases.

3. Experiment 2

Experiment 2 was an attempt to replicate and further investigate the different results obtained for different values of P_2 in Condition 1 of Experiment 1. Whereas in Experiment 1 we had included the second P_2 continuum purely for methodological reasons, the design of Experiment 2 took both P_1 and P_2 as independent variables, with 11 levels of P_1 and four of P_2. Our minimal prediction was that, if the Gussenhoven–Rietveld effect applies at "normal" values of P_2 but is reversed when P_2 is emphatic, then we should observe a statistical interaction between the two independent variables in their effect on the perceived prominence of P_2. More specifically, we predicted that the perceived prominence of the lowest P_2 should increase with P_1, that of the highest P_2 should decrease with P_1, and that of the intermediate values of P_2 should show a shift between these two extremes.

TABLE I. F_0 values (in Hz) used in synthesizing the contours for Experiment 2

Onset	P_1	Valley	P_2	Offset
	186			
	180			
	174			
	168		167	
	162		156	
116	156		145	90
	150		134	
	144			
	138			
	132			
	126	106		

3.1. *Method*

Speech materials were produced in exactly the same way as in Experiment 1, and the general procedures were unchanged. The only difference lay in the details of the stimuli themselves. In Experiment 1, we had stimuli involving two values of P_2 and 13 values of P_1. In Experiment 2, as just noted, we increased the number of values of P_2 to four and reduced the number of values of P_1 to 11, for a total of 44 different stimulus types. The values of P_1 and P_2 used in creating the stimuli are shown in Table I.

In order to increase the number of judgments per stimulus type we created two test tapes with the same stimuli in different orders. Each test tape contained two tokens of each stimulus, plus six "filler" stimuli at the end as in Experiment 1, for a total of 94 stimuli per tape. Subjects heard both test tapes with a short break in between; at some experimental sessions one tape was presented first and at others the other tape was presented first. Each subject therefore gave four judgments on each stimulus type.

A total of 22 subjects participated in the experiment, again mostly Edinburgh University undergraduates and all native speakers of English. This time subjects were paid a small sum for their participation. None of the subjects for Experiment 2 had participated in Experiment 1. The experimental sessions lasted about 25 minutes.

3.2. *Results*

As in Experiment 1, all subjects' prominence ratings for each stimulus type were averaged together, so that the mean prominence ratings are based on 88 separate judgements. The results are shown graphically in Fig. 5. Impressionistically they are consistent with the results of Condition 1 of Experiment 1. That is, for the lowest of the four values of P_2 (134 Hz), it appears that the perceived prominence of P_2 *increases* slightly as P_1 increases: this is the Gussenhoven–Rietveld effect. For all three higher values of P_2 (145, 156 and 167 Hz), as P_1 increases the perceived prominence of P_2 *dcreases* slightly; this is the "contrast effect".

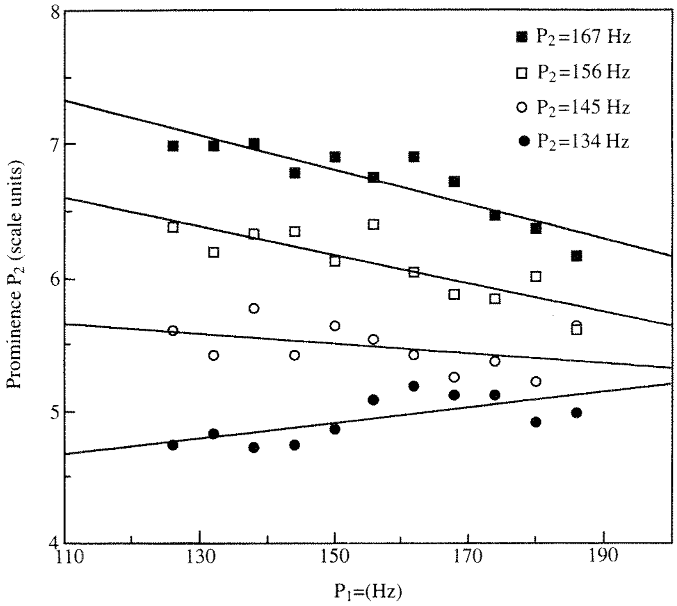

Figure 5. Results as in Fig. 3 of Experiment 2. The results for $P_2 = 134$ Hz show the Gussenhoven–Rietveld effect. Those for $P_2 = 156$ and 167 Hz show the reverse. Those for $P_2 = 145$ Hz do not show a significant positive or negative slope.

A more detailed statistical evaluation suggests that $P_2 = 145$ Hz was actually an intermediate case: the upward slope of the regression line for $P_2 = 134$ Hz was statistically significant ($p < 0.05$), as were the downward slopes of those for $P_2 = 156$ and 167 Hz ($p < 0.001$), but the slope of the line for $P_2 = 145$ Hz was not significantly different from zero. An ANOVA revealed the predicted interaction between the independent variables P_1 and P_2 in their effect on the perceived prominence of P_2 [$F(13, 4200) = 1.906$, $p < 0.005$]. In addition, there were main effects for both P_1 [$F(10, 4170) = 2.815$, $p < 0.005$] and, conspicuously, P_2 [$F(3, 4170) = 276.605$, $p < 0.0001$].

It thus appears that the Gussenhoven–Rietveld effect does occur in our materials, but that it applies only at moderate levels of P_2, as was suggested by the results of Experiment 1. The Gussenhoven–Rietveld effect is reversed at higher levels of P_2.

4. General discussion

The results of the two experiments can be summarized as follows. When the pitch range on an accent peak P_2 is moderate or "normal", then increases in the pitch range of a preceding accent P_1 will cause P_2 to be perceived as more prominent. This is the original Gussenhoven–Rietveld effect. We consider that we have replicated this effect, subject to the provision that the pitch range on P_2 must be "normal", and subject also to the apparent finding that trained listeners may hear stimuli differently from naïve subjects. When the pitch range on P_2 is high or "emphatic", then increases in the pitch range of P_1 will cause P_2 to be perceived as *less* prominent. That is, the Gussenhoven–Rietveld effect is reversed when P_2 is emphatic. This reversal—accidentally discovered in Experiment 1 and systematically replicated in Experiment 2—is the central new empirical finding reported here.

This finding raises at least two obvious questions. First, if the occurrence or non-occurrence of the Gussenhoven–Rietveld effect depends on the pitch range of P_2, on what basis does the listener decide whether P_2 is "normal" or "emphatic"? Is there some categorical threshold below which a peak counts as normal and above which it counts as emphatic? To what extent are medial valleys and utterance-initial and utterance-final F_0 values involved in setting this threshold? Is it the size of the pitch excursion from medial valley to P_2 that makes P_2 emphatic, or the height of P_2 above some "baseline"? Or is it simply that listeners place the boundary on the basis of the overall range of F_0 observed in the experimental stimuli—in much the same way as boundary placement in classical categorical perception experiments is known to depend on the range of stimuli presented (cf. e.g., Repp & Liberman, 1987: 94f). One can imagine a variety of experiments that might help to establish answers to these questions, but we make no specific predictions here.

The second general question raised by our results is more substantial: why does the Gussenhoven–Rietveld effect occur at all, and why does it fail to occur when accents have wider pitch range? Is the explanation to be sought, broadly speaking, in phonology or in psychophysics? We offer the following speculative answer.

We propose that when accents are within normal non-emphatic pitch range—whatever exactly that turns out to mean—then their prominence is *not assessed individually*. Rather, the pitch range *of the utterance as a whole* (which is computed as some function of the pitch range of the individual accents) conveys an *overall* degree of emphasis or arousal. This claim is consistent with, and indeed is a corollary of, Ladd's claim (forthcoming) that pitch and prominence relations between accents are restricted to a few, phonologically specified, categorical distinctions, rather than being continuously variable.

For example, when P_1 is very low and P_2 is non-emphatic, P_2 is interpreted as phonologically strong and phonologically non-downstepped, while *the utterance* is interpreted as having an overall low pitch range, which is reflected in *lower* prominence ratings for P_2. When P_1 is rather higher, but both P_1 and P_2 are still non-emphatic, then P_2 is still interpreted as phonologically strong and phonologically non-downstepped, but the utterance is interpreted as having an overall higher pitch range. This is then reflected in *higher* prominence ratings for P_2.

When emphatic pitch range is used, on the other hand, we speculate that this is a paralinguistic signal to override normal phonologically specified prominence relations, and to interpret pitch range on the basis of "every accent for itself". Specifically, if P_2 is emphatic, then a low P_1 is interpreted as having been reduced to set the scene for the expanded pitch range on P_2. Therefore lowering P_1 *increases* the perceived prominence of the emphatic P_2. A higher P_1, on the other hand, is interpreted as not having been so reduced, and consequently it downplays the emphasis on P_2 and causes its perceived prominence to be lower. This is the effect we observe at higher F_0 levels of P_2.

This is also the result obtained for the trained listeners in Condition 2 of Experiment 1. If this interpretation is correct, then the difference between the naïve and trained listeners is that the trained listeners are better able to evaluate "every accent for itself" regardless of the overall pitch range of the utterance. Under this interpretation, the Gussenhoven–Rietveld effect is a kind of auditory illusion, to which trained listeners are less susceptible. Whether this explanation is correct, it is at the very least plausible that phonetic training might have this kind of effect.

In any event, we submit that our results constitute substantial evidence for the existence of the Gussenhoven–Rietveld effect, and we believe that this effect must be taken into account in designing future experiments on "declination", the relative prominence of accents, and the like. However, a convincing explanation for the effect—and for its reversal—must await further research.

The partial support of the Economic and Social Research Council UK (ESRC) is gratefully acknowledged: Experiment 2 was part of the research program of the ESRC-funded Human Communication Research Centre (HCRC). We thank Norman Dryden and Irene Macleod for their assistance with producing the stimuli and analysing the results.

References

Boves, L., ten Have, B. L. & Vieregge, W. H. (1984) Automatic transcription of intonation in Dutch. In *Intonation, accent, and rhythm* (D. Gibbon & H. Richter, editors), pp. 20–45. Berlin: de Gruyter.

Gussenhoven, C. & Rietveld, T. (1988) Fundamental frequency declination in Dutch: testing three hypotheses, *Journal of Phonetics*, **16**, 355–369.

Hermes, D. & van Gestel, J. (1991) The frequency scale of speech intonation, *Journal of the Acoustic Society of America*, **90**, 97–102.

Jacobs, K. (1990) On the relationship between fundamental frequency of the initial accent peak and perceived prominence of the second accent peak, in two-peak utterances. Honours dissertation, Department of Linguistics, Edinburgh University.

Ladd, D. R. (1990) Metrical representation of pitch register. In *Papers in laboratory phonology I* (J. Kingston & M. Beckman, editors), pp. 35–57. Cambridge: Cambridge University Press.

Ladd, D. R. (1992) An Introduction to Intonational Phonology. In *Papers in laboratory phonology II* (G. J. Docherty & D. R. Ladd, editors), pp. 321–334. Cambridge: Cambridge University Press.

Ladd, D. R. (1993) On the theoretical status of "the baseline" in modelling intonation. *Language and Speech*, **36**, 435–451.

Ladd, D. R. (forthcoming) Constraints on the gradient variability of pitch range. In *Papers in laboratory phonology III* (P. Keating, editor). Cambridge: Cambridge University Press.

Liberman, M. & Pierrehumbert, J. (1984) Intonational invariance under changes in pitch range and length. In *Language sound structure* (M. Aronoff & R. Oehrle, editors), pp. 157–233. Cambridge, MA: MIT Press.

Repp, B. & Liberman, A. M. (1987) Phonetic category boundaries are flexible. In *Categorical perception: the groundwork of cognition* (S. Harnad, editor), pp. 89–112. Cambridge: Cambridge University Press.

Rietveld, A. C. M. & Gussenhoven, C. (1985) On the relation between pitch excursion size and prominence, *Journal of Phonetics*, **13**, 299–308.

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