Phonological Structure and Phonetic Form
Papers in Laboratory Phonology III

EDITED BY PATRICIA A. KEATING
Department of Linguistics, University of California Los Angeles

CAMBRIDGE
UNIVERSITY PRESS

Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
40 West 20th Street, New York, NY 10011–4211, USA
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

First published 1994

Printed in Great Britain by Woolnough Bookbinders Ltd, Irthlingborough, Northants

A catalogue record for this book is available from the British Library

Library of Congress cataloguing in publication data

Phonological structure and phonetic form/edited by Patricia A. Keating.
p. cm. – (Papers in laboratory phonology; 3)
Papers from the Third Conference in Laboratory Phonology held at UCLA.
Includes indexes.
ISBN 0 521 45237 6 (hardback)
1. Grammar, Comparative and general – Phonology – Congresses.
2. Phonetics – Congresses. I. Keating, Patricia A. II. Series.
P217.P484 1994
414–dc20 93-17718 CIP

ISBN 0 521 45237 6

TS

The Third Conference in L
teaching of Peter Ladefoge
in recognition of the role
study of sound.

4

Constraints on the gradient variability of pitch range, or, Pitch level 4 lives!

D. ROBERT LADD

4.1 The Free Gradient Variability hypothesis

One of the central assumptions of most work on intonation is that pitch range can vary gradiently to convey differences in emphasis or prominence. Indeed for most investigators this is not a "central assumption" but simply an indisputable fact: it is trivial to observe that when you raise your voice your utterance sounds more emphatic, and also – this is the gradient part – that the more you raise your voice the more emphatic it sounds. I don't propose to dispute either this fact, or the often tacit assumption that such variation in pitch range is "paralinguistic" and largely beyond the scope of phonological analysis.

However, the general observation that pitch range can vary gradiently and paralinguistically – which I don't dispute – has found its way into many theories of intonational phonology in the form of a much more specific assumption about the nature and extent of gradient variability – which I have been disputing for some years now. The assumption is this: the pitch range *on any pitch accent* can be gradiently varied to convey differences in "emphasis" or "prominence," and this variation is largely independent of, or unconstrained by, the pitch-range variation on any other part of the utterance. I will refer to this assumption as the Free Gradient Variability (FGV) hypothesis.

The FGV hypothesis is illustrated in the following two quotes, which display strikingly similar assumptions despite the differences due to the three decades of theoretical change that separate them.

> [W]hen emphasis is desired *on any part* of any utterance, several procedures can be used. . . . One can say the whole utterance, *or certain parts of it*, with greatly increased loudness and accompanying extra high, or, in some cases, extra low, pitch; this is often represented by special typography: I said JOE, not Bill. When this happens, the

whole utterance *or portion of it* is stretched out horizontally and vertically, as it were; this is then the point at which we draw the line between microlinguistics and metalinguistics: the phenomena that are segmentable were analyzed as phonemes of one kind or another; the phenomena that transcend segments are now stated to be metalinguistic, matters of style, and not part of the microlinguistic analysis. *Here, then, phonology ends.* (Trager and Smith 1951: 52; all emphasis supplied.)

The amount of difference in phonetic value between one accent and another accent which is metrically subordinated to it is continuously variable.... What controls this variation is something like 'amount of emphasis'....[I]ntonation patterns with only one pitch accent can be produced with different amounts of emphasis, with consequent variation in the height of the accent. It is not surprising that *this kind of variation also plays a role where there are several accents.*

[T]he term 'prominence' will be used to refer to the aggregate of metrical strength and emphasis, as it pertains to the control of tonal values. We will assume that *each pitch accent* has an associated prominence value, that prominence is continuously variable, and that the prominence of a metrically stronger accent is at least as great as that of a weaker accent, though not necessarily greater. We will not attempt to explain where prominence values come from, but *will leave this task to pragmaticists and semanticists.* (Pierrehumbert 1980: 39–40; all emphasis supplied.)

The most obvious problem with the FGV hypothesis, as I observed in Ladd 1990, is its falsifiability. *Unrestricted recourse to gradient pitch-range variability makes it nearly impossible to falsify quantitative models* of the phonetic realization of intonation: any observed F_0 target value that deviates from the predictions of a model can be said to have had its pitch range modified.[1] Furthermore, there is a pernicious corollary to the FGV hypothesis, which is that almost any "vertical scale" effect in the phonetic realization of intonation is automatically assumed to be a case of FGV. This actively discourages potentially fruitful investigation: because variation in the vertical scaling of pitch accents *may be* a matter of unpredictable, paralinguistic, gradient variation, phonologists generally assume that any variation in vertical scale *is* a matter of unpredictable, paralinguistic, gradient variation, and consequently do not look more closely at cases where the vertical scaling may actually be subject to more systematic constraints.

Finally, there is something paradoxical about the place of the FGV hypothesis in most work, which in a sense makes the case against it even more damning. The paradox is that, in practice, variation in prominence plays very little role in the detailed workings of quantitative models. On the one hand, such models all assume that any individual pitch accent *can* vary gradiently if the speaker chooses to vary it. On the other hand, they all adequately describe a wide range of speech data and hardly ever have to say that it *does* vary in this way. This surely makes the whole idea suspect. That is, if most of the data on which the quantitative models are based shows little or no evidence of Free Gradient Variability, then FGV is much too powerful a wild card to be included in the model as a little-used option.

The rest of the paper is devoted to presenting evidence against the hypothesis of FGV. In section 4.2 I discuss some general theoretical problems with the version of FGV embodied in Pierrehumbert's approach to the description of intonation. In section 4.3 I present some new and puzzling experimental findings that are seriously inconsistent with the assumption that pitch-range variation is interpreted accent by accent, and propose an explanation in terms of a categorical distinction between normal High tone and "Overhigh." In section 4.4 I relate this explanation to earlier proposals for the description of emphatic pitch range in English intonation.

4.2 The FGV hypothesis in Pierrehumbert's model

In order to focus the argument more specifically on current work, it will be useful to frame the discussion in terms of the model of intonational phonology developed by Pierrehumbert and her colleagues (e.g. Pierrehumbert 1980, 1981; Liberman and Pierrehumbert 1984; Beckman and Pierrehumbert 1986, 1992). I assume familiarity with the basic ideas of this model: the idea that a pitch contour is phonologically a string of tones, aligned in certain well-defined ways with the segmental string (Bruce 1977); that the only tones are H (high) and L (low), organized into "pitch accents" and arranged in various specified sequences; and that the F_0 targets corresponding to the tones are determined by phonetic realization rules, such that one H tone need not have the same F_0 as another. What I wish to focus on is Pierrehumbert's treatment of the relative height of tonal targets in pitch accents.

In Pierrehumbert's model, such differences of relative height effectively arise in only two different ways. First, one accent can be *downstepped* relative to an immediately preceding one. Recognition of the existence of downstep in English intonation, and the proposal of a quantitative model for it, was one of the important contributions of Pierrehumbert's thesis. In Pierrehumbert's conception, downstep is a phonetic realization rule applying to certain sequences of tones within a single phrase. Since downstep is phonologically conditioned (triggered by the occurrence of certain tonal sequences), it is categorically either present or absent. It is the only such categorical effect on vertical scale that Pierrehumbert's model recognizes.

The other way in which vertical scaling can be modified is by gradient modification of the overall pitch range. The quantitative details have evolved since Pierrehumbert's (1980) dissertation but the underlying theory has not. If two pitch accents within a phrase are not in a downstep relationship but have different peak levels, they are assumed to have different degrees of "prominence." If two phrases have similar accent patterns but the overall level of one is different from that of the other, they are assumed to have different "initial pitch-range settings." If two utterance contours are identical but for overall range, they too are assumed to have different pitch-range settings, reflecting the speaker's choice of different degress of "overall emphasis," different discourse organization or paragraph structure, etc. With one exception, the different degrees of prominence, emphasis, initial range, and so on are (a) assumed to be paralinguistic, and hence outside the realm of phonology, and (b) modeled as effects on a single parameter in the quantitative phonetic realization model. (The one exception is that metrical strength – which of course is phonological, not paralinguistic – is assumed to contribute, along with paralinguistic emphasis, to the prominence of individual accents. I will return to this point at the very end of the paper.)

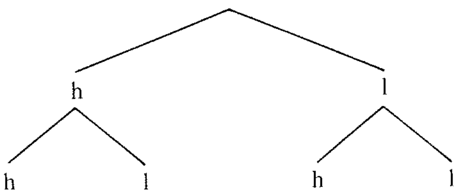
In short, vertical scale effects in Pierrehumbert's model, unless they involve phrase-internal downstep, are assumed to be a matter of FGV. This assumption has a number of unfortunate consequences, of which I will briefly discuss two. For more detail on these two issues see Ladd (1993).

4.2.1 Nested downstep

First, consider Beckman and Pierrehumbert's decision to ignore what might be called "nested downstep" in their intonational phonology. It is well established that F_0 downtrends can be nested, so that for example a sentence consisting of three distinct intonational phrases can show downtrends within each phrase and an overarching downtrend across the three phrases. Since the work of Pierrehumbert (1980), as just noted, it has been widely accepted that downtrending pitch contours within short phrases are the result of downstep – accent-by-accent lowering of the pitch register. However, there is good evidence that the downtrends from phrase to phrase also involve stepwise register lowering (e.g. Van den Berg *et al.* 1992; Monaghan 1988, 1991).

In order to express this similarity between accent-by-accent and phrase-by-phrase register shifts, I have elsewhere (Ladd 1988, 1990, 1993) proposed that downstep is a high–low phonological relation between two constituents in a prosodic tree, comparable to the weak–strong and strong–weak relations familiar from metrical phonology. For example:

(1)



H* H* L H* H* L L%

We WOULD have CALLED but there WASn't a PHONE.

As can be seen from this example, the constituents in such a downstep relation can be either terminal elements (individual pitch accents within a phrase) or nonterminal elements (phrases within a larger domain). That is, I see the existence of nested downstep as evidence about the phonological nature of downstep, and my analysis treats downstep – at any level in the prosodic hierarchy – in a uniform way.

In Pierrehumbert's phonological analysis, as just noted above, downstep is a phonetic-realization rule that only applies to certain tonal sequences *phrase-internally*. Obviously, by this definition, downstep is not something that can occur from one phrase to the next. Yet since phrase-to-phrase downstep, or something looking very much like it, manifestly does occur, it must be handled in a different way – and the only other way is as a reflection of FGV. Specifically, according to Beckman and Pierrehumbert (1986) the pitch range for each "intermediate phrase" is selected *independently* according to general discourse principles, and these "phrasal manipulations of overall pitch range *mimic catathesis* [= accent-to-accent downstep]" (299–300, emphasis supplied). The similarity of the accent-to-accent and phrase-to-phrase downtrends is thus ascribed to "mimicry," but why the one should mimic the other is left unexplained.

4.2.2 Pitch-range expansion experiments

Another problem for Pierrehumbert's model is the well-established finding that the relative height of pitch targets is preserved when overall utterance pitch range is experimentally modified. Pierrehumbert (1980), Bruce (1982), Liberman and Pierrehumbert (1984), and Pierrehumbert and Beckman (1988) all include reports of experiments in which specific intonation contours were uttered in varying overall ranges. In all of these experiments, two types of intonation-related variables were manipulated. First, the test utterances involved differences of emphasis, discourse structure, phrasing, length, etc. – differences that affect the height of accent peaks *relative to each other*. Second, each of the test utterances was pronounced in two or more overall pitch ranges – which affects the height of all the accent peaks in an utterance *relative to the speaker's voice range*. In every case, the two manipulations of the contour can be distinguished quite clearly in the experimental results. The patterns of relative F_0 within contours – the patterns that signal relative prominence, discourse, status, etc. – remain extraordinarily constant, while the overall range varies from just a few semitones to (in some cases) a few octaves.

The discovery of this constancy was another of the important contributions of Pierrehumbert's (1980) dissertation, and its role in establishing the significance of target levels in intonational phonology should not be underestimated. The fact that Pierrehumbert's original findings (which were based on English) have been replicated not only in English but also in Swedish and Japanese should guarantee them a central role in our theorizing about the control of pitch range. Yet for the standard Pierrehumbert analysis, incorporating the assumption of FGV, these results now pose a problem.

The problem is that the constancy of F_0 relationships when pitch range is modified is found not only in cases of phrase-internal downstep, but also in other cases involving accentual prominence, phrase-to-phrase relationships, and so on. As we saw, according to Pierrehumbert only the downstepping relationship within a phrase reflects a linguistic effect on vertical scale; everything else – including both relationships between phrases and nondownstepping relationships within phrases – reflects paralinguistic modifications. The constant patterns that emerge in the experimental data are therefore merely the consequence of consecutive paralinguistic choices within an utterance. It is, in theory, only a remarkable coincidence that all these choices bear the same relation to one another whether the voice is lowered or raised; Beckman and Pierrehumbert (1992) are able to suggest only that speakers somehow adopt a "uniform strategy" for dealing with such tasks.

4.2.3 *An alternative to FGV*

As manifested in Pierrehumbert's analysis of intonational phonology, then, the FGV hypothesis leads us to the conclusion that various quantitative regularities observed in production data from several languages are the result of unexplained mimicry of one contour by another, or of unexpected similarities in the way experimental subjects approach certain kinds of utterances. For a theory as ambitious and as productive of new insights as Pierrehumbert's, this is surely unsatisfactory.

But there is an obvious alternative. This is to assume that only the *overall* modifications of pitch range are gradient and paralinguistic, and that the relative height of accents within phrases and of phrases within sentences is part of the *linguistic* specification of the contour – i.e. part of intonational phonology. This requires us to give up the idea that almost anything to do with vertical scaling is gradient, paralinguistic, and therefore safe to ignore, but it permits us to treat nested downstep as nested downstep, and to make straightforward sense of constant relative F_0 under range expansion rather than be forced to describe it as a curious coincidence. Experimental evidence for this alternative view is presented in the next section of the paper.

4.3 The limits of Free Gradient Variability

4.3.1 The Gussenhoven–Rietveld effect

The story begins with a perceptual effect discovered more or less accidentally by Gussenhoven and Rietveld (1988). In a set of experiments designed to test various hypotheses about the implementation of declination, they asked listeners to judge the *prominence* of pitch accents in stimulus sentences. The sentences were "reiterant" nonsense utterances of the form da-DAH-da-da-da-DAH-da, i.e. seven-syllable utterances with two accent peaks; the two peaks are henceforth referred to as P1 and P2. Various acoustic parameters were manipulated, in particular the F_0 on P2; the listeners' task was to judge the prominence of P2. One of the central findings is that in any given stimulus continuum the average listener ratings of P2's prominence correlate very well with P2's F_0. A typical graph is shown in figure 4.1. This result is scarcely surprising, and is entirely consistent with the FGV hypothesis. However, Gussenhoven and Rietveld's experiments also shed light on what happens to the perceived prominence of P2 when we manipulate the acoustic properties of *P1*, and this is what is of interest here.

Suppose that P2 is held constant but the F_0 on P1 is raised or lowered, as in figure 4.2. What will subjects say about the prominence of P2 as a function of the modification of P1? Pretheoretically, one could imagine three possible types of effects. First, there could be no effect whatsoever: a given pitch level on P2 signals prominence level *p*, and the fact that the prominence on some neighbouring accent changes is irrelevant. This would be the strongest possible confirmation of the hypothesis of FGV, since according to that hypothesis the prominence on each accent can be modified independently. Second, there might be some sort of syntagmatic comparison, similar to so-called "contrast effects" in psychophysics: if P1

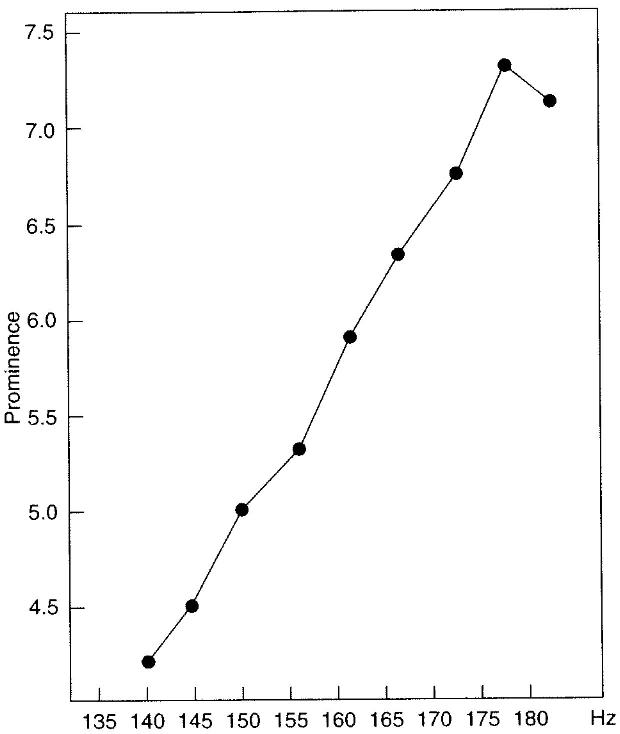


Figure 4.1. Typical results in Gussenhoven and Rietveld's experiment, showing the close correlation between P2's perceived prominence (y-axis) and F_0 on P2 (x-axis).

is made more prominent by increasing its F_0, then the prominence of a given P2 will be correspondingly reduced. I imagine that this is the effect most phonologists and phoneticians would predict if they were forced to think about it; note that the existence of some such comparison of accent peaks in context, though it might make for problems of quantitative detail, would not seriously undermine the FGV hypothesis.

Finally, there is the remaining logically possible effect, which is that increasing the F_0 on P1 would *increase* the perceived prominence of *P2*, while lowering P1 would *decrease* it. That is, instead of some sort of psychophysical contrast effect, there would be a sort of global effect of raising the F_0 on any accent that would affect the prominence on all accents. This seems fairly unlikely; it is certainly difficult to imagine how one might reconcile such a finding with the FGV hypothesis, because it would appear to make it impossible to increase the prominence of an individual accent relative to the prominence of its neighbors. However, the accidental

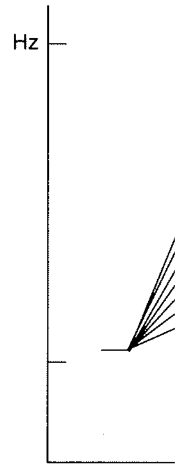


Figure 4.2. Schematic repres
held constant and P1 is syste

discovery made by Gus
the F_0 of P1 have this
For reasons not re
Rietveld's experiments
which P1 had different
peak F_0 values. One of
figure 4.1 – may be reg
accent with an F_0 excu
smaller than the F_0 ex
was reduced relative to
respectively. Taking t
normal version therefc
comparisons of the so
involved variations in i
Gussenhoven and Ri
speaking, what they fo

(2a)

DAH

da da da dā

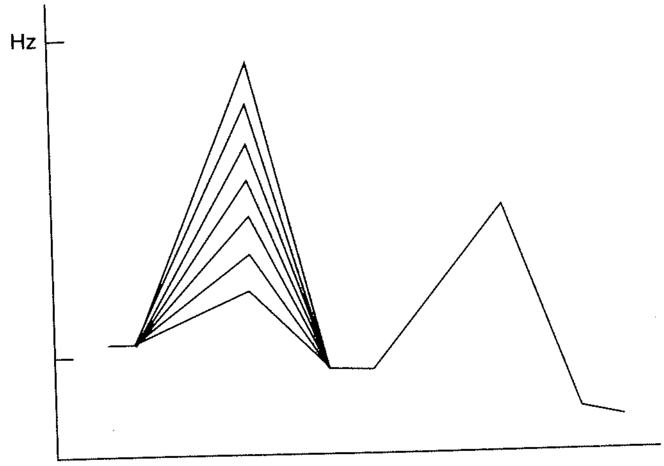


Figure 4.2. Schematic representation of a set of experimental two-peak contours in which P2 is held constant and P1 is systematically varied.

discovery made by Gussenhoven and Rietveld was precisely that changes in the F_0 of P1 have this global effect on the prominence of P2.

For reasons not relevant to the discussion here, Gussenhoven and Rietveld's experiments involved five different stimulus continua, in each of which P1 had different acoustic properties, and P2 had the same range of peak F_0 values. One of the continua – the one for which results are shown in figure 4.1 – may be regarded as having a "normal" P1; in it, P1 had a pitch accent with an F_0 excursion that at each step in the continuum was slightly smaller than the F_0 excursion on P2. In two other continua, the F_0 of P1 was reduced relative to this "normal" version, to "Low" and "Very Low" respectively. Taking these two reduced-P1 continua together with the normal version therefore provides us with a continuum of experimental comparisons of the sort sketched in figure 4.2. (The last two continua involved variations in intensity and will not be discussed here.)

Gussenhoven and Rietveld's results seem quite unambiguous. Informally speaking, what they found is that in stimulus pairs like

(2a)

	DAH			
DAH				
da		da da da		da

(2b)

			DAH	
	DAH			
da		da da da		da

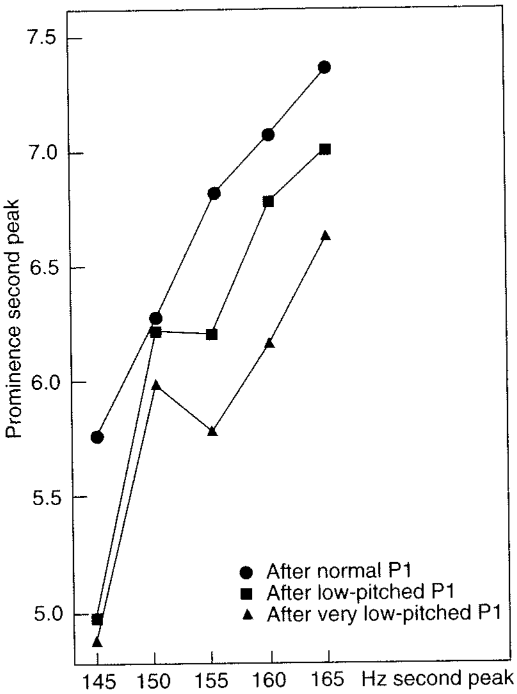


Figure 4.3. The Gussenhoven–Rietveld effect. For a given F_0 value of P2, the perceived prominence of P2 is lower when P1 is lower.

listeners judge the prominence of P2 in (2a), where P1 is relatively low, to be *lower* than the prominence of P2 in (2b), where P1 is somewhat higher. As we move from the continuum with "Very Low P1" through that with "Low P1" to that with "Normal P1," the perceived prominence for any given P2 *steadily increases*. That is, the peaks of F_0 on P1 and P2 do not function independently, nor do they set up a psychophysical contrast effect: rather, the perceived prominence of *P2* appears to correlate with the F_0 on *P1*. This finding, which I refer to as the Gussenhoven–Rietveld effect, is shown in figure 4.3. Gussenhoven and Rietveld acknowledge that this effect is somewhat puzzling but do not really pursue the matter further.

4.3.2 A possible account of the Gussenhoven–Rietveld effect

A possible explanation for the Gussenhoven–Rietveld effect, consistent with the idea that gradient variability of pitch range is actually severely constrained, would be as follows. First suppose that all the contours investigated by Gussenhoven and Rietveld are instances of "nondown-

stepped" P2, i.e. all in
P1 and P2. (This is nc
Rietveld themselves aı
downstep.) Suppose
prominence of P2 is nc
a function of some so
order to increase the
increase the pitch rar
differently, overall in
primarily to the nucl
prominence of P2 can
the peak F_0 on *either*
unitary impression c
perceived prominence
This explanation is
prominence of each
other pitch accents. It
on the basis of Gusse
was an inadvertent by
investigated more clos
that a theory of pitch

4.3.3 Replicatin

In an experiment don
dissertation in the L
(1990) carried out a
Rietveld effect. The b
in which *P2* was helc
were still asked, as in
prominence on *P2*.[3]
different values of P2,
we thought it likely t
that P2 was always th
Though we intend
other, it turned out tl
continuum with the lc
with the Gussenhove
perceived prominence
noisy, one might be
Gussenhoven–Rietvel
(160 Hz) value of P2,

stepped" P2, i.e. all instances of a single phonological relationship between P1 and P2. (This is not an unreasonable supposition, as Gussenhoven and Rietveld themselves are at pains to point out that they are not dealing with downstep.) Suppose further that, at least in such cases, the perceived prominence of P2 is not purely a function of the peak F_0 *on P2*, but is rather a function of some sort of *overall* (utterance-level) pitch range. That is, in order to increase the prominence on the nuclear accent, it is sufficient to increase the pitch range on the phrase as a whole. (To put it somewhat differently, overall increases in pitch range are felt by listeners to apply primarily to the nuclear accent.) If this is the case, then the perceived prominence of P2 can be increased or decreased by increasing or decreasing the peak F_0 on *either* pitch accent: the F_0 on both peaks contributes to a unitary impression of phrasal pitch range, which in turn affects the perceived prominence of P2.[2]

This explanation is obviously deeply incompatible with the view that the prominence of each pitch accent is gradiently variable independently of other pitch accents. It would be unwise, though, to go too far in theorizing on the basis of Gussenhoven and Rietveld's results alone. Since the effect was an inadvertent by-product of their study, it needs to be replicated and investigated more closely before we consider it to be one of the phenomena that a theory of pitch range and prominence should be able to account for.

4.3.3 Replicating and extending the Gussenhoven–Rietveld effect

In an experiment done under my direction for an Undergraduate Honours dissertation in the Linguistics Department at Edinburgh, Karen Jacobs (1990) carried out a systematic attempt to replicate the Gussenhoven–Rietveld effect. The basic idea of the experiment was to create a continuum in which *P2* was held constant and *P1* was varied, but in which listeners were still asked, as in Gussenhoven and Rietveld's experiment, to rate the prominence on *P2*.[3] In fact, however, we used two such continua with different values of P2, mixing the stimuli randomly on the test tape, because we thought it likely that otherwise listeners would rapidly become aware that P2 was always the same.

Though we intended the two continua simply as distractors for each other, it turned out that they produced puzzlingly divergent results. In the continuum with the lower (140 Hz) value of P2, there is a trend consistent with the Gussenhoven–Rietveld effect: as the F_0 on P1 increases, the perceived prominence of P2 increases as well. While the data are rather noisy, 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 P2, no such effect can be observed. If anything, increases in the F_0 on P1 produce a slight *decline* in the perceived prominence of P2, so that the result curves for the two levels of P2 converge as P1 increases. This is shown in figure 4.4.

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, the generally noisy data, and in particular (because it is entirely consistent with the notion of FGV), the fact that the largest effect on the perceived prominence of P2 is the F_0 level of P2 itself.

However, one might at least consider taking seriously the apparent convergence of the two curves in figure 4.4, and conclude that something interesting is going on. Specifically, suppose that in using two different values of P2 we inadvertently introduced two distinct experimental conditions, one in which P2 represents normal High tone, and one in which it represents some sort of "Overhigh" or emphatic tone. When P2 is normal High, we get the Gussenhoven–Rietveld effect: increases in the F_0 of P1 produce increases in the perceived prominence of P2. But when P2 is Overhigh, the Gussenhoven–Rietveld effect does not appear; instead, we get something like a psychophysical contrast effect whereby increases in the F_0 of P1 bring about slight decreases in the perceived prominence of P2. In

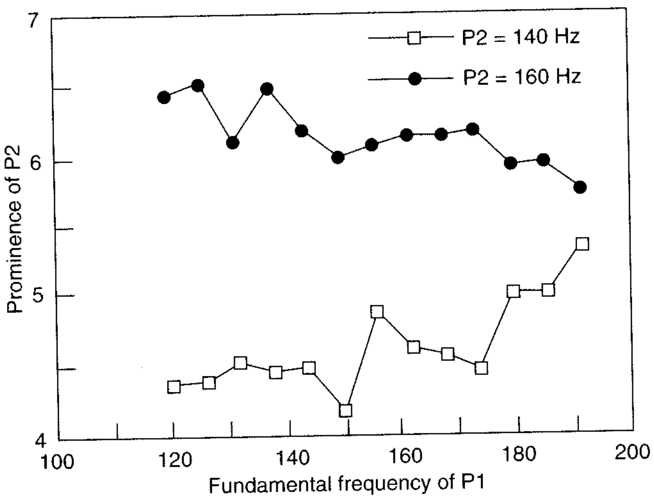


Figure 4.4. Results of the experiment by Jacobs. The lower curve seems to display the Gussenhoven–Rietveld effect (see Figure 4.3), but the upper one seems to show the reverse.

statistical terms, we have an interaction: the effect of P1 on the perceived prominence of P2 is different for different F_0 values of P2.

Extending the explanation offered in the previous section for the Gussenhoven–Rietveld effect, we might suggest that when P2 is normal High, P1 and P2 are in a fixed phonological relationship that does not permit of gradient modification except as applied to the contour as a whole. When P2 is Overhigh, on the other hand, gradient pitch-range effects can apply to it independently, and the listener evaluates the prominence of an Overhigh P2 at least partly on the basis of a direct comparison with P1. In any case it is clear that the attempt to replicate the Gussenhoven–Rietveld effect has – like Gussenhoven and Rietveld's study itself – led unexpectedly to a puzzling result, which itself needs replicating.

In order to determine the robustness of the apparent interaction, therefore, Jo Verhoeven, Karen Jacobs and I did a much larger study, involving nine levels of P1 and four levels of P2. The results, in figure 4.5, show very clearly that the interaction discovered by Jacobs is replicable. The data are far less noisy because more subjects were used, and the picture seems unmistakable. For the lowest of the four values of P2, the perceived prominence of P2 *increases* as P1 increases: this is the Gussenhoven–Rietveld effect. For all three higher values of P2, as P1 increases the perceived prominence of P2 *decreases* slightly; this is the psychophysical contrast effect. An analysis of variance on the results suggests that this difference is real: despite the massive main effect of P2 on perceived prominence, the interaction with P1 is also statistically significant.

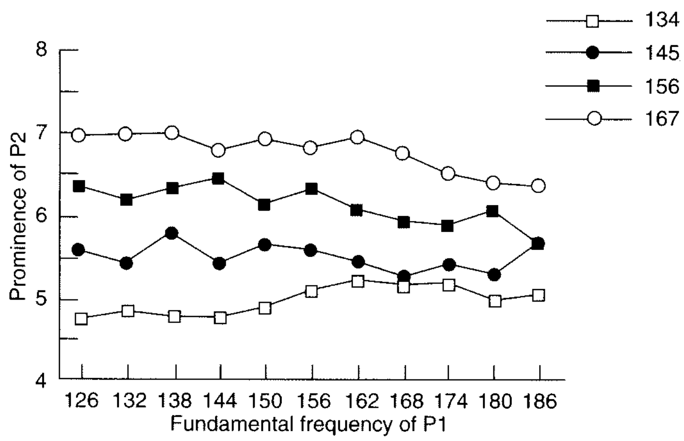


Figure 4.5. Results of the experiment by Ladd, Verhoeven, and Jacobs. The lowest curve shows the Gussenhoven–Rietveld effect (see Figure 4.3), but the three upper curves show the reverse.

4.3.4 Summary

These experimental results can be summarized as follows: if one presents listeners with an utterance containing two accent peaks, in which *both peaks* are of *moderate height*, one can produce an increase or decrease in the perceived prominence of the second or nuclear peak by increasing or decreasing the F_0 on *either* peak. If, however, the second or nuclear peak is very high, then increases in the perceived prominence of the second or nuclear peak must be produced in two different ways: either by increasing the F_0 on the already very high second peak, or by *decreasing* the F_0 on the first peak.

The proposed explanation for these findings takes the form of three theoretical conjectures:

1. Gradient modification of pitch range can be a property either of phrases or of individual accents.
2. When it is a property of the phrase, it affects the perceived prominence of the phrase's nuclear accent, *irrespective of where the gradient variability is phonetically manifested.* This implies that within the phrase there is only a limited range of possible *phonological* vertical-scale relationships (e.g. downstepped, nondownstepped) between the nuclear accent and any prenuclear ones.
3. Gradient pitch-range variability can be a property of an individual accent only when the accent is both (a) nuclear, and (b) Overhigh.

4.4 Overhigh tone?

The weak spot in the account just sketched is obviously the notion of Overhigh tone. How can a "very high" peak be distinguished from a peak that is "moderately high"? The very use of such terms seems to cry out for an analysis in terms of gradient variability of a single underlying category High tone – as in the standard FGV view. In this final section of the paper I wish to explore the possibility of Overhigh tone in greater depth.

First of all, it is worth stressing that the idea is not *a priori* ridiculous. We know that there are many languages, especially in Africa, in which categorically distinct levels of lexical tone are extracted from the continuum of the speaking range. We even know that some such languages (e.g. Chaga, McHugh 1990) have a distinction between a lexical High tone and a contextually raised "Overhigh" tone. That is, human listeners are in principle capable of putting a distinction between "moderately high" and "very high" to phonological use. The suggestion being made here is that that is exactly what they are doing in European intonation systems.

The idea of Overhigh tone in English intonational phonology is not new, of course, having originated in the work of Kenneth Pike in the early 1940s (published as Pike 1945). Pike's original analysis of English intonation involved four phonologically distinct levels, which we may call Low, Mid, High, and Overhigh. This idea was promptly taken over by Wells (1945), Trager and Smith (1951), and others, and – with the pitch levels treated as "phonemes" – it became the standard post-Bloomfieldian analysis of intonation in a variety of languages. Pike originally numbered the four levels from 4 at the bottom to 1 at the top, but in the standard version the numbering was reversed, so that pitch level 4 was Overhigh. This is the usage implied in my alternative title, and the one I will continue with here.

The four-level analysis was the subject of a fundamental critique by Bolinger (especially 1951), which led to the so-called levels-versus-configurations debate that simmered unresolved for roughly thirty years. Bolinger argued that, since sequences of pitch levels like 21, 31, and 41 are in theory phonemically distinct, they should be categorically distinct semantically as well – or even semantically unrelated. In fact, of course, all three seem to be instances of a falling contour, a single broad category with an identifiable (if hard-to-state) common element of meaning. Consequently, Bolinger argued, any representation in which the three are phonemically distinct is misleading.

In place of phonemic levels, Bolinger proposed that the units of intonation are pitch "configurations" like *fall* and *rise* – pitch accents, in the analysis subsequently developed in Bolinger 1958. More importantly for the issue under discussion here, Bolinger also claimed that the three putative variants 21, 31, and 41 are just arbitrarily selected steps on a gradient continuum of emphasis or finality. Pitch range, he said, can vary gradiently to reflect gradients of meaning; different "pitch levels" are simply the result of gradient variation of range on different pitch accents.

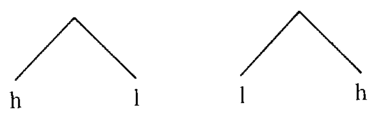
Bolinger's insistence on the primacy of pitch configurations and the irrelevance of levels now appears overstated: Bruce (1977) and Pierrehumbert (1980), and several others since them, have provided clear evidence of remarkable invariance of pitch level at certain points in contours. Moreover, Pierrehumbert (1980) showed that Bolinger's theoretical objections to levels can be met, so long as pitch accents are recognized as units at some level of analysis, and if the number of phonologically distinct levels is reduced from four to two (H and L). But Bolinger's views on gradient variability have been incorporated more or less intact into the theoretical consensus that has been built on the foundation of Bruce's and Pierrehumbert's pioneering work. Once pitch accents are analyzed as sequences of H and L tones, then the actual F_0 values in a given pitch accent can be analyzed in terms of the realization of the Hs and Ls on a vertical scale, specified in a separate, essentially orthogonal part of the phonological description. In the new theoretical consensus, the parameters that are manipulated in this orthogonal part of the description are gradient.

As I said at the beginning of the paper, it cannot be denied that certain vertical scale effects – at least those that affect whole utterances – are gradiently variable in essentially the way that Bolinger and the new theoretical consensus presuppose. In my view, however, most of the factors that govern the relative height of accents *within* a phrase or utterance are phonological, and hence categorical rather than gradient. Where I disagree with the new consensus, in other words, is in positing distinctions of relative pitch range that – like downstep – are orthogonal to the basic tone distinctions *but not gradient*. Among these distinctions is the one proposed here between normal High and Overhigh tone.

The proposal for Overhigh tone was foreshadowed in my early critique of Pierrehumbert's intonational phonology (Ladd 1983), in which I proposed that nuclear accents might display a categorical feature "raised peak." As I noted at the time, the raised-peak proposal was essentially a restatement of what was involved in the distinction between pitch levels 3 and 4 in the four-level analyses. As such it was incompatible with the FGV hypothesis, and it was simply dismissed by Beckman and Pierrehumbert (1986: 307), who reiterated their belief that all such differences of vertical scale are gradient, and suggested that my proposal was based on a "misinterpretation" of the experimental findings discussed in section 4.2.2 above. However, the data and theoretical considerations presented here suggest that the notion of raised peak or Overhigh tone is at least as plausible as unrestricted FGV.

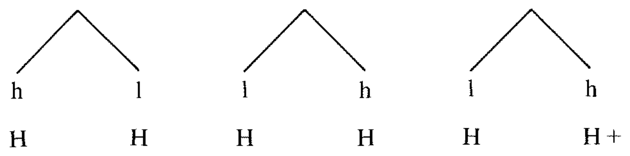
Overhigh tone fits into my relational analysis of downstep (see section 4.2.1 above) as follows. The basic claim of that analysis is that there are only two distinct phonological relations between a prenuclear and a nuclear accent, namely downstepped and nondownstepped:

(3)



One of the difficulties with this view, however, is that it provides no distinct representation for what appear to be two subcases of nondownstepped – one in which the nuclear accent is approximately at the same level as the prenuclear accent, and one in which the nuclear accent is clearly upstepped. I would now suggest that in the nondownstepped case, it is possible for the H tone of the nuclear accent to be replaced by an Overhigh (H+) tone, yielding a distinct upstep. This means that relative to a prenuclear accent peak, the peak of a nuclear falling accent can be distinctively lower, roughly the same height, or higher. Graphically:

(4)



I believe this three-way distinction is the basis for the three types of nuclear fall – 21, 31, and 41 – posited in the original Pikean analyses of English intonation. Perhaps, in other words, pitch level 4 lives.[4]

While I have framed the discussion here at least in part as a critique of Pierrehumbert's intonational phonology, it should be noted that in Pierrehumbert's original observations (quoted at length above) we can find the seeds of the analysis just proposed. Recall that Pierrehumbert sees the prominence of an accent as "the aggregate of metrical strength and emphasis"; she specifically notes that "the prominence of a metrically stronger accent is at least as great as that of a weaker accent, though not necessarily greater." We might say that Pierrehumbert is implicitly distinguishing only two kinds of cases: those where the metrically stronger accent is *not* more prominent than the weaker one (the "level 3" cases), and those where it is (the "level 4" cases). (The level 2 cases are of course downstepped and treated entirely differently by Pierrehumbert.) If we remove the large facultative element of gradient emphasis from this view, we are left with a distinction very much like the one proposed here between normal High and Overhigh tone.

4.5 Conclusion

In revising this paper for publication I have deliberately refrained from modifying the proposal for Overhigh tone, in order not to pull the rug out from under Hayes's excellent critique (chapter 5). However, I should note that I find his interpretation of Overhigh as "gesturally reinforced High" quite plausible and intuitively appealing; more generally, I think we may draw considerable insight from his suggestion that "the beast knows the grammar." On the whole I think our analyses differ little in their practical consequences for the proposed explanation of the experimental data discussed here.

However, I think there remains an issue between us, namely whether the distinction between normal and "gesturally reinforced" is categorical. Hayes appears to suggest that it is not: in his view, as in Bolinger's, the beast is always active in the production of pitch accents, and the phonetic variability of pitch accents results from the extent of the beast's activity. I incline to an alternative view, namely that the presence or absence of "gestural reinforcement" is an all-or-none matter, though of course if gestural reinforcement is present its extent is gradient. In effect, the beast may simply sleep through certain pitch accents, and reinforce only those in which it has some special involvement. At this point I see little basis for determining which of these views is correct.

In any case, the central point of the descriptive proposals I have made here and elsewhere is that the Bruce–Pierrehumbert approach to intonational phonology must be enriched with a notion of categorical distinctions of pitch range. We need to get rid of the idea that any distinction that is orthogonal to the basic opposition between High and Low tones is *ipso facto* gradient: both gradient factors and categorical ones play a role in the vertical scaling of any given tone. Once this idea is accepted, I believe that we will be in a much better position to understand downstep, emphasis, and intonational cues to textual organization generally. Perhaps more importantly, a great many conceptual problems with pitch range will effectively disappear.

Notes

The experiment by Verhoeven, Jacobs and myself reported in section 4.3.3 forms part of the research program of the Human Communication Research Centre (HCRC). The support of the UK Economic and Social Research Council (ESRC), which provides funding for HCRC, is gratefully acknowledged.

1 The most egregious example of this known to me comes from Cooper and Sorenson (1981). One of the sentences on which they tested their model of the declining F_0 "topline" is *The CAT in the GARAGE ran SWIFTLY UNDERNEATH the CAR* (where the capitalized words are the ones in which F_0 values constituting the "topline" were measured); in the experimental data, the measured peaks on *garage* and *underneath* were significantly lower than predicted, and that on *swiftly* substantially higher. They explain these deviations away as follows:

> It seems likely that . . . *swiftly* was responsible for the perturbation . . . . Since this word is an Adverb, it probably received more stress than a non-Adverb at the same sentence location [references omitted]. In addition, it seems reasonable that the extra focus given to the Adverb might cause a defocusing of the neighboring key words . . . . In short, the focused Adverb pulls up on the topline; to compensate, a lowering of the topline occurs just after the focus, creating the observed zigzag pattern. The present rationale is admittedly ad hoc, but such proposals seem useful at this rudimentary

stage of F_0 research in sentence contexts, in order to suggest directions for independent further testing. (1981: 70–71)

2 This proposal is broadly consistent with the findings recently reported by Terken (1991), though Terken's procedures (and his theoretical assumptions) are so different from Gussenhoven and Rietveld's that it is difficult to compare their findings in detail.

3 By and large we followed procedures similar to Gussenhoven and Rietveld's. Perhaps the biggest difference was that, instead of a reiterant nonsense utterance, we used a natural utterance of the sentence *The melon was yellow*, resynthesized with different F_0 contours. Listeners were asked to rate the prominence on the word *yellow*. Full details of this experiment and the follow-up are reported in Ladd, Verhoeven, and Jacobs (forthcoming).

4 It would appear that it is also possible to have Overhigh tone on the prenuclear accent in a downstepping phrase; in line with the Gussenhoven–Rietveld effect, this seems to add finality to the entire phrase rather than adding emphasis to the prenuclear accent, as in:

(5) DID

I NO

T

h l

H+ H

(i.e. I did not)

References

Beckman, M. and J. Pierrehumbert. 1986. Intonational structure in Japanese and English. *Phonology Yearbook* 3: 255–309.

1992. Comments on chapters 14 and 15. In G. J. Docherty and D.R. Ladd (eds.) *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*. Cambridge: University Press; 387–397.

Van den Berg, R., C. Gussenhoven and T. Rietveld. 1992. Downstep in Dutch: Implications for a model. In G.J. Docherty and D. R. Ladd (eds.) *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*. Cambridge University Press, 335–359.

Bolinger, D. 1951 Intonation: Levels versus configurations. *Word* 7: 199–210.

1958. A theory of pitch accent in English. *Word* 14: 109–149.

Bruce, G. 1977. *Swedish Word Accents in Sentence Perspective*. Lund: Gleerup.

1982. Developing the Swedish intonation model. *Working Papers 22, Department of Linguistics, Lund University*: 51–116.

Cooper, W. and J. Sorensen. 1981. *Fundamental frequency in sentence production.* Heidelberg: Springer-Verlag.

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

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. Undergraduate Honours dissertation, University of Edinburgh.

Ladd, D.R. 1983. Phonological features of intonational peaks. *Language* 59: 721–759.

1987. A phonological model of intonation for use in speech synthesis by rule. In *Proceedings of the European Conference on Speech Technology*. Edinburgh: CEP Associates, 21–24.

1988. Declination "reset" and the hierarchical organization of utterances. *Journal of the Acoustical Society of America* 84: 530–544.

1990. Metrical representation of pitch register. In J. Kingston and M. Beckman (eds.) *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*. Cambridge: University Press, 35–37.

1993. In defense of a metrical theory of intonational downstep. In H. van der Hulst and K. Snider (eds.) *The Phonology of Tone: The Representation of Tonal Register*. Berlin, New York: Mouton de Gruyter, 109–132.

Ladd, D. R, J. Verhoeven, and K. Jacobs (forthcoming). Influence of adjacent pitch accents on each other's perceived prominence: Two contradictory effects. To appear in *Journal of Phonetics*.

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

McHugh, B.D. 1990. The phrasal cycle in Kivunjo Chaga tonology. In S. Inkelas and D. Zec (eds.) *The Phonology–Syntax Connection*. Chicago: University of Chicago Press, 217–242.

Monaghan, A.I.C. 1988. Generating intonation in the absence of essential information. In W. Ainsworth and J. Holmes (eds.) *Speech 88: Proceedings of the 7th FASE Symposium*. Edinburgh: Institute of Acoustics, 1249–1256.

1991. Intonation in a text-to-speech conversion system. Ph.D dissertation, Edinburgh University.

Pierrehumbert, J. 1980. The phonology and phonetics of English intonation. Ph.D dissertation, MIT.

1981. Synthesizing intonation. *Journal of the Acoustical Society of America* 70: 985–995.

Pierrehumbert, J. and M. Beckman. 1988. *Japanese Tone Structure*. Cambridge, MA: MIT Press.

Pike, K.L. 1945. *The Intonation of American English*. Ann Arbor: University of Michigan Press.

Terken, J. 1991. Fundamental frequency and perceived prominence of accented syllables. *Journal of the Acoustical Society of America* 89(4): 1768–1776.

Trager, G.L. and H.L. S
Battenburg Press. (I
Washington.)

Wells, R. 1945. The pitc

Trager, G.L. and H.L. Smith. 1951. *An Outline of English Structure*. Norman, OK: Battenburg Press. (Reprinted 1957 by American Council of Learned Societies, Washington.)

Wells, R. 1945. The pitch phonemes of English. *Language* 21: 27–40.