Phonology Yearbook

An annual journal

1

Edited by
Colin J. Ewen and John M. Anderson

The right of the University of Cambridge to print and sell all manner of books was granted by Henry VIII in 1534. The University has printed and published continuously since 1584.

CAMBRIDGE UNIVERSITY PRESS
CAMBRIDGE
LONDON NEW YORK NEW ROCHELLE
MELBOURNE SYDNEY

*Declination : a review and some hypotheses**

D. Robert Ladd
Cornell University/University of Edinburgh

Introduction

One of the most widespread – and widely studied – properties of speech fundamental frequency (F o) is a tendency to decline gradually during the course of utterances. This tendency has been given a variety of names, of which the best known is probably DECLINATION. This is the term I shall use in this paper. The purpose of the paper is to review past work on declination as it affects the phonological and phonetic modelling of F o contours, and to outline some ideas for the empirical resolution of the issues that emerge from the discussion.

The relevance of declination for the phonological description of pitch phenomena is far from clear. To the extent that they are not merely impressionistic, observations that declination is widespread or universal are based on a statistical reduction of acoustic data – trend lines through time plots of F o values. The observation that such trend lines tend to decline could be incorporated into a theoretically grounded phonological description in at least two different ways.

1 *The statistical view.* The statistical generalisation about acoustic properties of F o may be taken to reflect nothing more than a statistical generalisation about the phonological categories of intonation systems. Irrespective of how intonational phonology is to be described, one's description must be such that (a) there are more contour types that decline than types that do not, and/or (b) contour types that do decline are used more frequently than those that do not. These generalisations have the same status as e.g. the fact that [p] occurs more commonly in English than [ž], or the fact that most syllables in most languages are CV or CVC: they are essentially irrelevant to the description of individual languages or individual contours. They imply nothing about the structure or manner of realisation of F o contours with rising trend lines; they only state that such contours are less common.

Given this approach, any explanation for the phenomenon of declination must be put in terms of why phonological *systems* are the way they are. The notion 'model of declination', in the sense of a unitary quantitative description of F o trend lines, is meaningless. F o trend lines are determined

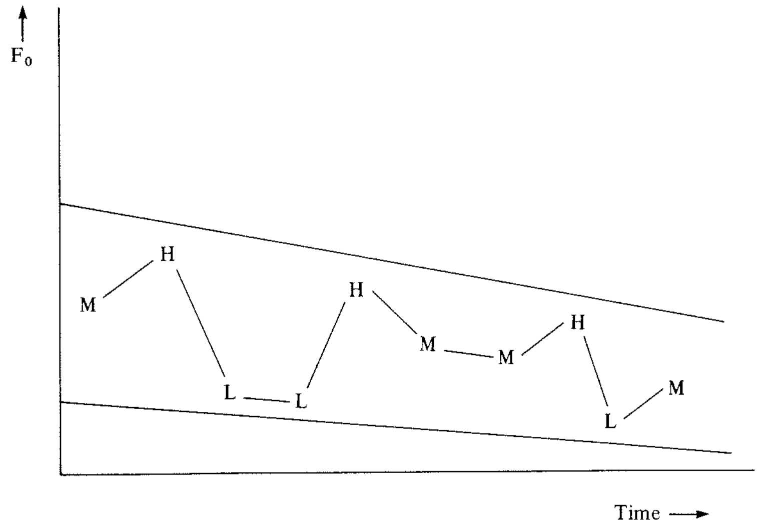


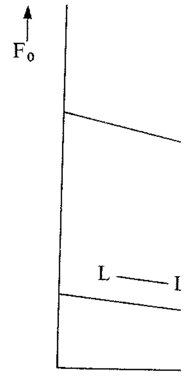
Figure 1

String of tones in a hypothetical tone language utterance. Instances of the same tone are realised at progressively lower pitch; 'declination' may be defined as the gradual decline in the phonetic frame of reference.

by phonological and phonetic specifications; declination is not something that can be factored out of those specifications, but at most a generalisation about what the specifications are like.

2 *The frame-of-reference view.* In this approach, declination is treated as a systematic modification, during the course of an utterance, of the phonetic frame of reference in which phonological elements are realised – a kind of gradually changing backdrop to local Fo events. The strongest hypothesis about the cause of this backdrop change is that it is automatic, caused by physiological consequences of the act of speaking (e.g. reduced subglottal pressure due to breathing out), but other explanations are possible as well.

This approach can be illustrated most clearly with hypothetical tone language data. Assume a language with three level tones (high, mid, and low), in which occurrences of a given tone early in an utterance are higher than occurrences of the same tone later in the utterance. This could be described by saying that the range of pitch used for realising tonal contrasts becomes gradually lower and narrower, as shown in Fig. 1; tones that are phonologically 'the same' are realised differently depending on the width of the frame of reference at the time they are realised. There are two important points to be made about such a description. One is that a rising trend line need not imply the absence of declination: the tone sequence LLMMHH, realised within the gradually narrowing frame of reference, has a rising trend line, as shown in Fig. 2, but still obviously manifests declination in the relevant sense. The other point is that there may be phonological rules that lower pitch in specific phonological environments



A hypothetical string
'declination' in the se

(e.g. 'final lowering' and
below), which mean that t
of reference itself.

A 'model of declinatio
backdrop change. The m
not the measurement of F
phonology, which makes
phonologically 'the same
trend line through a sec
relevant for discovering t
trend line through a sequ
sequence of high tones the
would be relevant only af
of downstep and final low

*

The goal of this paper is
declination that have be
evidence for and against th
and to suggest some specif
of these views.

1 Theoretical proble

By outlining the statistical
just done, I do not mean to
theoretical foundation or t

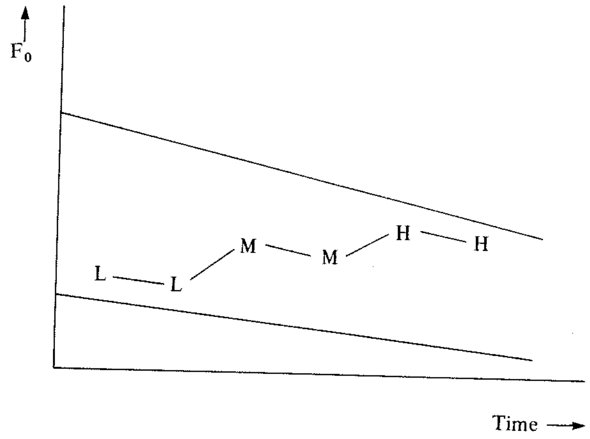


Figure 2

A hypothetical string of tones exhibiting both a rising trend line and 'declination' in the sense illustrated in Fig. 1.

(e.g. 'final lowering' and 'downstep', discussed at greater length in §3.2 below), which mean that trend lines may drop more steeply than the frame of reference itself.

A 'model of declination', given this approach, is a description of the backdrop change. The most important empirical basis of such a model is not the measurement of Fo trend lines, but a clear description of the pitch phonology, which makes it possible to determine which Fo events are phonologically 'the same'. To continue the tone language example: the trend line through a sequence of equal high tones would be directly relevant for discovering the properties of the frame of reference, but the trend line through a sequence of downstepped high tones, or through a sequence of high tones the last of which had been lowered by final lowering, would be relevant only after one had factored out the quantitative effects of downstep and final lowering.

* * *

The goal of this paper is to describe some of the specific accounts of declination that have been proposed in the literature, to review the evidence for and against the two general theoretical positions just outlined, and to suggest some specific approaches to further testing and refinement of these views.

1 Theoretical problems in past work

By outlining the statistical view and the frame-of-reference view as I have just done, I do not mean to suggest that past work is clearly based on one theoretical foundation or the other. For the most part, what one actually finds in work on declination in the European languages is that the distinction between the two points of view is seldom maintained: in particular, models which are ostensibly models of a declining component or frame of reference are actually models of average trend lines. Trend lines are simply fitted to a few particularly salient points in contours – usually the obvious local peaks or valleys – and the average characteristics of many such trend lines are taken to represent the decline of the phonetic frame of reference. Clear representatives of this general approach are Maeda (1976) and Cooper & Sorensen (1981).

If the basis of the frame-of-reference view is taken seriously, as I showed in the tone language example earlier, it makes sense to model trend lines only if the points to which the trend lines are fitted are known to be phonologically equivalent. But in most of the empirical work on the European languages, the problem of phonological equivalence is ignored, and thereby as it were turned upside-down. In the absence of any clear theoretical basis for their selection of which contour points to fit, and which contour types to average together, several models of declination describe *a statistical norm* and then use it as if it were *a property of individual utterances* which can then be used for determining the phonological specifications of local F0 events. Cooper & Sorensen, for example, say that in formulating a quantitative model of declination in this way they hope to 'provide a basis from which one can parcel out the effects of declination when attempting to study other, local F0 attributes' (1981: 28).

The theoretical confusion between treating declination as a statistical norm and treating it as a phonetic property of individual utterances seems to be involved in Cooper & Sorensen's decision to use mean signed error as a basis for evaluating the fit of their model; see Pierrehumbert & Liberman (1982) for an effective critique of Cooper & Sorensen's statistical methods. The same confusion also seems to underlie the way in which those who are concerned with modelling trend lines approach the problem of declination in questions. Vaissière, for example, states that 'the declination tendency has been shown to be often suppressed in interrogative sentences' (1983: 57), meaning that questions often show a non-declining trend line. Vaissière's terminology reflects the theoretical problem: 'declination tendency' suggests a statistical norm, but to say that it is 'often suppressed' makes sense only if declination is being treated as a phonetic property of individual utterances.

The same theoretical confusion is also the reason that trend lines are often fitted by eye – i.e. by informal or intuitive criteria. Informal criteria are used because the implicit theoretical goal of the enterprise is to fit a *simply described* line – often a straight line – even though the points to be fitted seldom lie exactly along such a line. (This point has been made strongly by Fujisaki & Hirose 1982 in arguing against any trend line approach.) This obviously makes it difficult to evaluate the goodness of the fit of any model very objectively, and is certainly one of the main reasons that those concerned with modelling trend lines have found that 'calculation of the exact rate of declination is a difficult task' (Vaissière 1983: 56).

Among the descriptive
(1) Are declination lines
beginning of the utter
untransformed F0 da
1979) fit straight (and
transformed data, and
two straight line segn
formed data.
(2) Is the amount of decli
and the end of the fit
length of the utterance
for each speaker, an
utterances. The formu
that in utterances less
with the length of the
the overall drop rema
decline is less marked.
(3) Does the initial valu
constant? (This questi
be unlikely to arise i
amount of drop.) Coop
of initial value of ne
utterance, and 't Hart
or 5 semitones; yet St

It is not hard to see that
content. For example, g
bottomlines to logarithmic
in linear (untransformed)
the valleys (as in Maeda's
Cooper & Sorensen's ben
Cooper & Sorensen's findi
of a contour lies well above
that in logarithmically tran
fit to all peaks. The point i
it is difficult to choose em
different assumptions; the

Much the same is true f
declination and length of u
empirically well-groundec
utterance contours ending
speaker (Maeda 1976; Me
1984). Depending on how
course, always depending
inclusion in the model – it i
the overall drop is constan
various studies have show
on the topline depends to

Among the descriptive debates about trend lines are the following:

(1) Are declination lines straight, or do they decline more rapidly at the beginning of the utterance? Maeda (1976) fits straight bottomlines to untransformed Fo data, while 't Hart and his colleagues (e.g. 't Hart 1979) fit straight (and parallel) top- and bottomlines[1] to logarithmically transformed data, and Cooper & Sorensen (1981) fit a 'bent' topline – two straight line segments of which the first is steeper – to untransformed data.

(2) Is the amount of declination (i.e. the difference between the beginning and the end of the fitted line) constant, or does it increase with the length of the utterance? Maeda found a constant *amount* of declination for each speaker, and hence a slower *rate* of decline in longer utterances. The formula proposed by 't Hart (1979), by contrast, states that in utterances less than five seconds long the overall drop increases with the length of the utterance (in utterances longer than 5 seconds the overall drop remains constant), so that the difference in rate of decline is less marked.

(3) Does the initial value increase with length of utterance, or is it constant? (This question is related to the previous one, since it would be unlikely to arise in models like Maeda's that claim a constant amount of drop.) Cooper & Sorensen (1981) found average differences of initial value of nearly a semitone depending on the length of utterance, and 't Hart's formula predicts differences of as much as 4 or 5 semitones; yet Sternberg *et al.* (1980) found no such effect.

It is not hard to see that some of these issues are without much empirical content. For example, given that 't Hart can fit straight top- and bottomlines to logarithmically transformed data, it would be expected that in linear (untransformed) Fo data, straight lines would give a better fit to the valleys (as in Maeda's straight bottomlines) than to the peaks (hence Cooper & Sorensen's bent toplines). Or we can turn it around: given Cooper & Sorensen's finding that in linear-plotted Fo data the first peak of a contour lies well above a straight line fitted to the later peaks, it is clear that in logarithmically transformed data a straight line would give a better fit to all peaks. The point is that since all the fits are approximate anyway, it is difficult to choose empirically among different hypotheses based on different assumptions; the problem is a theoretical one.

Much the same is true for the question of the relation between rate of declination and length of utterance. One of the few generally accepted and empirically well-grounded facts in this area is that the endpoints of utterance contours ending in a fall are relatively constant for any given speaker (Maeda 1976; Menn & Boyce 1982; Liberman & Pierrehumbert 1984). Depending on how one decides to fit the bottomline, then – and of course, always depending on which utterances one has selected for inclusion in the model – it is quite possible to conclude, as Maeda did, that the overall drop is constant. With the topline there is no such constancy: various studies have shown quite clearly that the height of the last peak on the topline depends to at least some extent on the number of accents

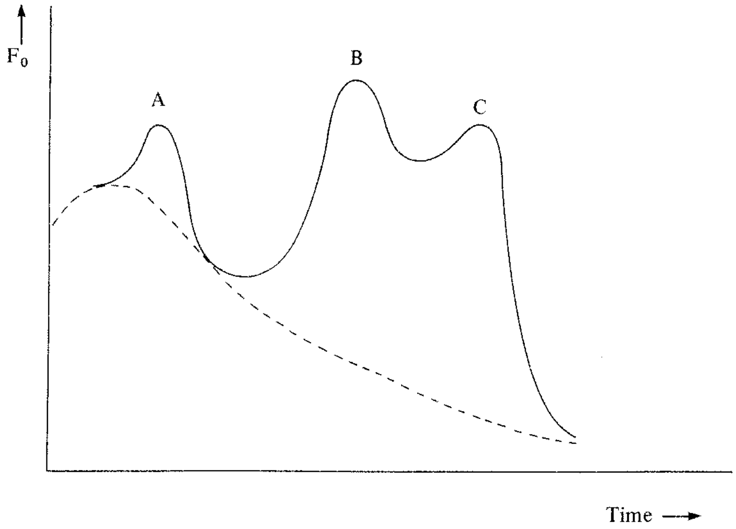


Figure 3

Declination in Fujisaki's model. The 'phrase component' (dashed line) rises briefly, then decays exponentially, in response to an impulse command at time zero. The output Fo contour (solid line) results from the superposition of an 'accent component' onto the phrase component; the accent component is a step response to accent commands of various magnitudes. Peak A results from a smaller accent command than Peak C, though the two have roughly the same Fo because A is superimposed on a much higher point in the phrase component. Conversely, Peaks B and C arise from accent commands of the same size, but differ in Fo because of the decay of the phrase component.

in the utterance (e.g. Liberman & Pierrehumbert 1984; Thorsen 1980a). Once again, then, the debate about the invariance or lack of invariance in declination depends to a considerable extent on theoretical decisions about how to fit trend lines – decisions which have seldom been rigorously made.

To be sure, not all the issues just outlined are purely a matter of theoretical assumption. The question of initial peak Fo and overall length, for example, does seem to be a matter of conflicting data, with Sternberg *et al.*'s results disagreeing with the findings of a number of others.[2] A number of differences of procedure may help explain the disagreement, notably the fact that Sternberg *et al.*'s subjects were instructed to speak fast and in a monotone, and that their utterances were all very short. This is discussed further by Pierrehumbert (1980: 133f).

2 Some theoretically explicit accounts of declination

2.1 Frame-of-reference models

The essence of the frame-of-reference approach, as I said in the introduction, is to factor a declination component out of the phonological and phonetic specifications for Fo contours. This component need not bear any

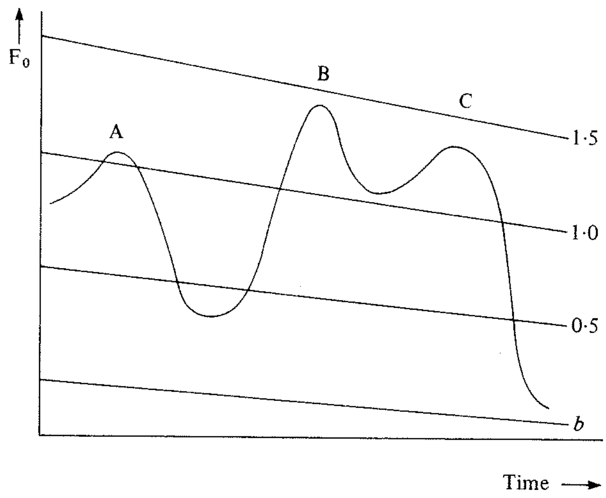


Figure 4

Declination in Pierrehumbert's model. The phonetic frame of reference for the hypothetical contour is defined by the declining baseline (*b*); the slanting grid lines show distances in 'baseline units' above *b*. Peaks A and C are at roughly the same Fo, but C is higher above *b* in baseline units and is presumed to be more prominent; Peaks B and C are roughly the same height above *b* in baseline units and hence are of equal prominence, but B has a higher actual Fo.

obvious resemblance to the gross shape of the contours themselves. At least two variants of this idea have been proposed, by Fujisaki and his coworkers (e.g. Fujisaki & Nagashima 1969; Fujisaki & Sudo 1971; Fujisaki & Hirose 1982) and by Pierrehumbert in her dissertation (1980).

Fujisaki's model posits two components, a phrase component and an accent component, which are additively combined to yield output Fo contours. The phrase component shows an initial sharp rise and then an exponential decay; accent commands of various sizes ride, so to speak, on top of this component. This is illustrated in Fig. 3. Declination can be defined as the slow decline of the phrase component.

In Pierrehumbert's model, there is no phrase component as such, but Fo targets (analogous in some respects to Fujisaki's accent commands) are scaled in proportional units above a declining x-axis or 'baseline' (cf. footnote 1). Pierrehumbert invokes the metaphor of slanted graph paper to describe the effect of scaling Fo this way (see Fig. 4). As with Fujisaki's phrase component, declination can be defined in Pierrehumbert's model in terms of the slope of the baseline.

The major interest in these models lies in the fact that they make it possible to talk about declination in utterance contours that do not actually exhibit an overall downward trend (e.g. the kind of case illustrated in Fig. 2). This makes plausible the contention that declination is caused by some sort of 'automatic' physiological mechanism, because it makes it possible to maintain that declination is present in all utterances. I will return to this point below.

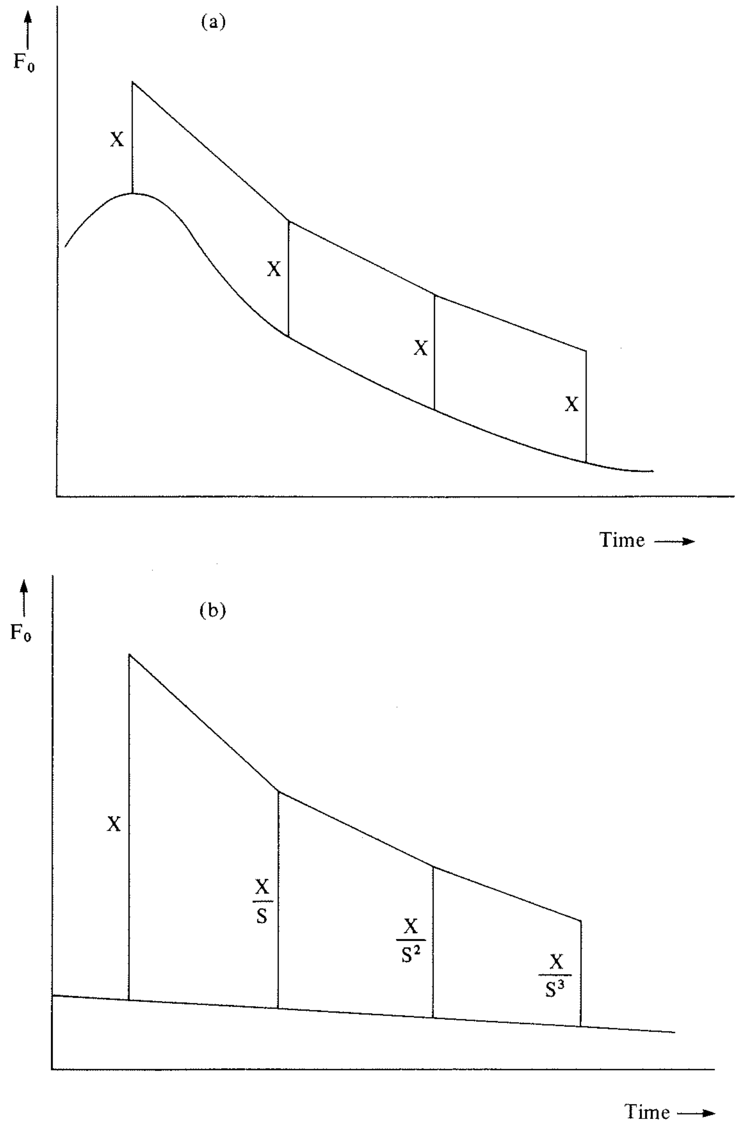
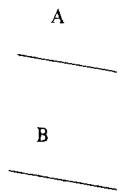


Figure 5

Exponential decay of topline in Fujisaki's and Pierrehumbert's models. Fujisaki would generate exponential topline decay by a series of identical accent commands superimposed on the exponentially decaying phrase component (Fig. 5a). Pierrehumbert would treat it as downstepping of the individual accent peaks in sequence, scaling the peaks progressively smaller amounts above the baseline (Fig. 5b).



Schematised repres
and B does not, des
Fo to the baseline,
from Cohen *et al.* (

The major empirica
which features of actu
the frame of referenc
phonological specifica
shape of many topline
in the sense that the '
means that otherwise
will yield peaks that
Pierrehumbert, on the
step', a local feature
a fixed proportion low
in the slanted graph
plotted. This differen
I should make it cle
models apart from th
theoretical aims of so
previous section of t
(Cohen & 't Hart 196
The topline and botto
reference grid, and are
phonological analysis.
that the interpretation
on where they take pl
6). However, the fact t
reference to actual co
problems of fitting tren
Cohen *et al.* argue at s
be adequately modelle
from the other), provi
concede that there ma
word is 'adequately',
empirical issue.

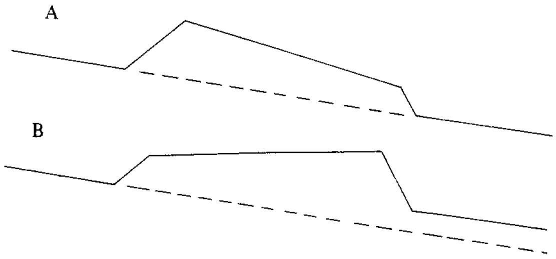


Figure 6

Schematised representation of two contours of which A sounds terminal and B does not, despite A having a smaller final fall. In A, the fall lowers Fo to the baseline, which is not the case in B. Reproduced by permission from Cohen *et al.* (1982).

The major empirical task, given a frame-of-reference model, is to decide which features of actual contours are to be attributed to characteristics of the frame of reference or declining phrase component and which to the phonological specification of the contour. A case in point is the exponential shape of many toplines. Fujisaki's model treats this as part of declination, in the sense that the 'phrase component' exhibits exponential decay: this means that otherwise equal 'accent commands' in a sequence of accents will yield peaks that can be fitted by an exponentially decaying topline. Pierrehumbert, on the other hand, treats the exponential decay as 'downstep', a local feature of individual accents that causes them to be scaled a fixed proportion lower than the preceding accent; declination resides only in the slanted graph paper against which the downstepped accents are plotted. This difference is illustrated in Fig. 5.

I should make it clear that by classifying Fujisaki's and Pierrehumbert's models apart from the rest, I am not playing altogether fair with the theoretical aims of some of the authors whose work I discussed in the previous section of the paper, particularly 't Hart and his colleagues (Cohen & 't Hart 1967; 't Hart & Collier 1975, 1979; Cohen *et al.* 1982). The topline and bottomline in this model are intended as a kind of abstract reference grid, and are based on an explicit (though arguably oversimplified) phonological analysis. Furthermore, Cohen *et al.* (1982) have pointed out that the interpretation of pitch movements depends not on their size, but on where they take place relative to the grid lines (this is shown in Fig. 6). However, the fact that the topline and bottomline are determined with reference to actual contour high and low points saddles them with all the problems of fitting trend lines discussed in the previous section. Specifically, Cohen *et al.* argue at some length that the topline and the bottomline can be adequately modelled as straight and parallel (and hence derivable one from the other), provided that Fo is plotted logarithmically, though they concede that there may be a problem with the initial high peaks. The key word is 'adequately', and that is what makes this a theoretical, not an empirical issue.

2.2 Fo models without declination components

To end this section, I will briefly mention a few recent models of intonation and Fo in which the idea of a unitary phenomenon of declination plays no role in the phonological description. These include the work of Thorsen (e.g. 1980a, b), Bruce and Gårding and their coworkers (e.g. Bruce & Gårding 1978; Gårding *et al.* 1982; Gårding 1983; henceforth the 'Lund model'), and Liberman & Pierrehumbert (1984; this supersedes the model of Pierrehumbert 1980 just discussed). These models are all consistent with the 'statistical' point of view outlined at the beginning of the paper.

Thorsen's model and the Lund model involve a variety of overall contour lines indicating different pragmatic and grammatical functions. No one of these lines is more basic than any of the others, and no attempt is made to reduce them all to a single phenomenon of 'declination'. Liberman & Pierrehumbert, meanwhile, have gone further in the direction of accent-by-accent specification of Fo scaling suggested by Pierrehumbert's treatment of 'downstep'. The slanted reference grid has been done away with, and *all* overall trends in contours are generated by phonological characteristics of individual accents and pairs of accents. The principal issue between these two approaches is whether overall trends in contours should be modelled in terms of overall contour lines, or solely in terms of local tonal specifications. The wider arguments for and against these two views are beyond the scope of this discussion, but I will return to this question briefly in §4.2 below.

2.3 Causes of declination

One of the most common hypotheses advanced to account for declination is that it is due to some sort of automatic mechanism, some physical consequence of the act of speaking. An early hypothesis along these lines was put forth by Lieberman (1967), who suggested that the drop in Fo was due to the drop in subglottal pressure that accompanies expiration during speech. This was part of a much grander hypothesis that subglottal pressure is the major determinant of speech Fo, a hypothesis that was quickly discredited (for a review see Ohala 1978).[3] Nonetheless, there is little doubt that subglottal pressure can affect Fo, and a gradual decline in subglottal pressure throughout the utterance could well be an important factor in declination. Various work by Collier (especially 1975, 1983; cf. also Cohen *et al.* 1982) suggests a production model in which subglottal pressure and laryngeal tension are the two major interacting components of output Fo. This is also explicitly suggested by Pierrehumbert (1980: 137f).

A more complex, but conceptually similar, proposal is Maeda's 'tracheal pull' hypothesis (1976). According to this hypothesis, the larynx is gradually lowered during an utterance or phrase as a result of its physical links to the sternum, which should be lowered by decreasing lung volume. While the physical mechanisms proposed are different, this proposal is like

Collier's in the sense that it posits two major interacting components in the production system for Fo.

A rather different explanation – though still couched in terms of physical mechanisms of speech – is suggested by Vaissière (1983), citing a paper by Ohala & Ewan (1973). According to Ohala & Ewan, it is more difficult to produce an Fo rise than a fall: the laryngeal gesture requires more effort and takes more time. This asymmetry, together with a 'laziness principle', means that speakers should tend to produce smaller rises than falls, so that over the course of an utterance the Fo would trend downwards. Unlike Collier's or Maeda's, then, this explanation for declination is based solely on laryngeal gestures – a 'one-component' model. More importantly, unlike Collier and Maeda, it does not claim to explain specific characteristics of individual utterances so much as general properties of intonation systems. Ohala (1978) argues that the tendency for Fo to trend downwards would lead listeners to expect such a decline and use it as a perceptual cue to utterance boundaries, which in turn would cause it to be incorporated into the linguistic system as a variable that speakers use with communicative intent. This explanation is consonant with Ohala's general theory of how phonetic considerations can be used to explain phonological change (e.g. Ohala 1981).

The idea of an automatic declination mechanism is basically quite attractive. It accounts for the apparent universality or near-universality of declination in compelling terms – speakers can't help it. The essential problem with any such account is to integrate it with the fact that there are non-automatic determinants of Fo as well. Collier's or Maeda's approach involves the postulation of two separate production components – in effect, one automatic and one not – whose outputs are added together or otherwise interact in some way. Ohala's approach has only a single major production component, but assumes that certain physical characteristics of the production mechanism have influenced the developments of the linguistic systems that make use of that mechanism.

A two-component explanation such as Maeda's or Collier's is obviously closely compatible with Fujisaki's and Pierrehumbert's models. The phrase component or baseline of those models would be the correlate of the 'automatic' production component and the accent component or target sequence the correlate of the 'non-automatic' (laryngeal) component. A single-component explanation such as Ohala's, since it does not directly involve automatic mechanisms in explaining the shape of actual contours, is consistent with the recent models that have abandoned the notion of automatic declination altogether – the work of Thorsen, Bruce and Gårding, and Liberman & Pierrehumbert discussed above. I emphasise that these models are merely consistent with Ohala's explanation; they do not in any sense presuppose it or follow from it.

2.4 Summary

To sum up the paper so far, I see two theoretically coherent ways to take account of the general phenomenon of declination in phonological description. One is a rigorously held two-component view, with one of the components showing some sort of automatic decline that really is automatic, i.e. non-suppressible. Physical mechanisms responsible for the automatic decline need not be fully understood for models along these lines to be valid. The clearest examples of this approach are the models of Fujisaki and Pierrehumbert. The other coherent position would be to leave declination in the realm of statistical generalisation, irrelevant for phonological description. All overall trends in actual contours – downward or otherwise – would be generated by phonological rules and specifications and assumed to be controlled by the speaker; explanation of the universality of declination would have to be in terms such as Ohala's – i.e. explanations for linguistic systems, not for acoustic outputs. The recent models of Thorsen, Bruce and Gårding, and Liberman & Pierrehumbert all exemplify this approach, albeit in rather different ways.

3 Evidence for the two theories of declination

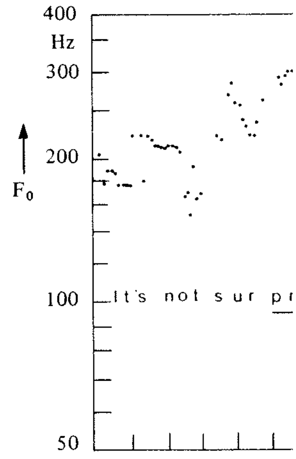
This section outlines evidence for and against the two clear theoretical positions just outlined, and considers what new kinds of evidence might be sought to test and refine them further.

3.1 Evidence for a declination component

The following evidence appears to support the notion of automatic declination of the sort that could be modelled in a 'two component' approach:

(1) Perceptual evidence. Pierrehumbert (1979) created a set of synthetic stimuli with two high accent peaks, with the second peak ranging from considerably lower to considerably higher than the first, and asked subjects to judge which peak was higher. Oversimplifying somewhat, her findings were that listeners rated the two peaks as equally high when the second was actually slightly lower. This evidence could be interpreted as showing that listeners expect the automatic decline in pitch and correct for it in perception, so that they hear as equal two peaks that are phonologically – but not acoustically – equivalent.

(2) Production evidence. Collier (1983) found that in long stretches of syllables without pitch accents, there was a small but noticeable drop in pitch, and that when these stretches followed a single early pitch accent (as shown in Fig. 7), there was good correlation of Fo with subglottal pressure. By most accounts of intonational phonology, there would be no accent commands or other phonological elements occurring in such stretches – i.e. the pitch in such stretches would be



Fo curve of the Engli
money' with only on
slow decline of the Fo
of an abstract 'declin

phonologically 'the
a second mechanisn
accentless stretch o
with subglottal pre
'second mechanism
in subglottal pressu
(3) Explanatory power
vincingly answered,
to 'because speakers
developments of lin

3.2 Evidence against

The arguments in favo
intonational phonology
contention that no sucl
Specific cases include tl
(1) The need for some
least some of the e
must be accounted
accent commands,
several kinds of evi
(a) In many of the t
which lower high

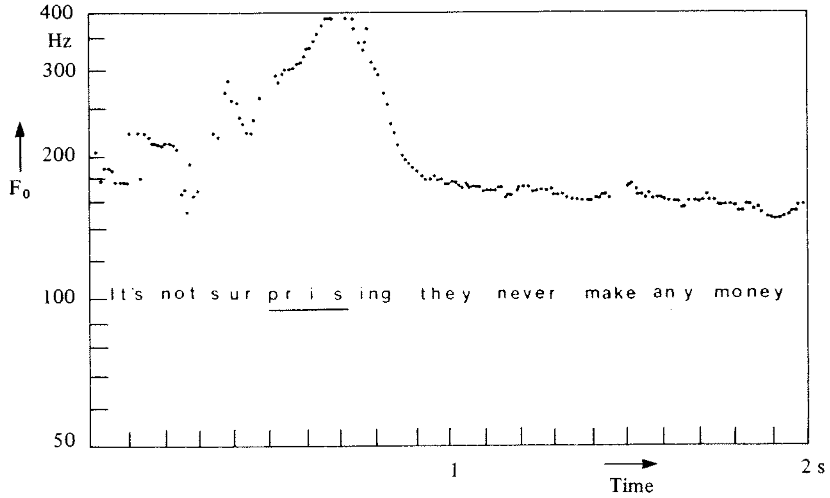


Figure 7

Fo curve of the English utterance 'It's not sur<u>pris</u>ing they never make any money' with only one pitch accent. Cohen *et al.* (1982) suggest that the slow decline of the Fo after the early accent peak directly reflects the course of an abstract 'declination line'. Reproduced by permission.

phonologically 'the same'. The fact that it actually declines suggests a second mechanism at work. (However, Collier notes that when the accentless stretch occurs between two accents, the correlation of Fo with subglottal pressure is not very good, which suggests that the 'second mechanism' may be more complicated than a simple decline in subglottal pressure.)

(3) Explanatory power. The question 'Why declination?' is more convincingly answered, as I noted above, by an explanation that amounts to 'because speakers can't help it', than by one that involves long-term developments of linguistic systems.

3.2 Evidence against a declination component

The arguments in favour of abandoning declination as a component of intonational phonology are based largely on Occam's Razor – i.e. on the contention that no such component is needed to account for the data. Specific cases include the following:

(1) The need for some sort of downstep rule. It seems quite clear that at least some of the exponential decay that often characterises toplines must be accounted for by phonological rules affecting peak scaling or accent commands, and not by any automatic component. There are several kinds of evidence, positive and negative, for this:

(a) In many of the tone languages of Africa there are phonological rules which lower high tones following low tones ('downdrift' and 'downstep'). This means that the topline of an utterance consisting of a sequence HLHLHLH will be markedly steeper than an utterance consisting only of high tones. (This has been demonstrated instrumentally for Hausa by Lindau ms.) The conditioning factor here is clearly phonological, not automatic.

(b) It has been suggested that similar distinctions between stepping and non-stepping contours exist in non-tone languages, e.g. the distinction between the 'high head' and the 'stepping head' of O'Connor & Arnold (1961). This has been reanalysed in terms of sequences of downstepped *vs.* non-downstepped accents by Pierrehumbert (1980); see also Ladd (1983). Again, a phonological distinction is involved here, not an automatic drop in pitch.

(c) The contours regarded by Pierrehumbert as involving downstep have toplines whose overall drop is probably greater than could be accounted for by any of the automatic mechanisms that have been proposed (cf. Ohala 1978: 31f; Cooper & Sorensen 1981: 160).

Given all this evidence for phonological rules specifying various overall downward trends, simplicity would argue for the incorporation of *all* such trends into phonological rules unless there were clear evidence against doing so.

(2) The need for a final lowering rule. The argument here is like the preceding one. There is considerable evidence in tone languages for rules that lower the final tone of a phrase or utterance relative to the preceding tone of the same phonological level (Pike 1948: 28). Quantitative modelling of stepping contours in English and Danish appears to provide evidence for the same kind of rule in non-tone languages (Liberman & Pierrehumbert 1984). Obviously, if the final lowering rule and other phonological rules, which are needed in any case, can account for all of the apparent downward trend, the declination component is superfluous.

Note that in short utterances, final lowering could mimic the effect of an overall downward trend. Liberman & Pierrehumbert (1984) appear to regard final lowering as the explanation for the perceptual results of Pierrehumbert (1979).

(3) Difficulty with the notion 'automatic'. The boundary between 'speaker-controlled' and 'automatic' is so difficult to draw that one is led to suspect a spurious distinction. The declination component or frame of reference in a two-component model must be controlled by the speaker to at least some extent: in particular, it must regularly be reset – at phrase boundaries and at the beginnings of utterances – or else it would quickly decay away to nothing. If the speaker can control the resetting of the phrase component, why not also its decline? In Fujisaki's model this problem is avoided to some extent, because the phrase component is conceived of as an impulse response; the speaker controls the location of impulses (according to grammatical or other criteria) and then the response characteristics of the system take over.

Even this solution is not entirely satisfactory, however, because there

are no constraints c
impulses. It would
'suppression' of de
succession of impuls
plenty of evidence th
is subject to linguisti
nature of the constrai
85ff, 120ff; Vaissière
of whether we are
backdrop decline at a
separate components
overall trends, but b
in some sense part of
in §4.2 below.

4 Testing opposing

It thus appears that the
positions outlined at the
must be brought to bear
final section of the paper
be used to decide betwe

4.1 Time dependence

One obvious testable dif
declination in Fujisaki's
This entails the predicti
peaks should correlate w
example, one would p
capitalised syllables in (
(a) 'FANny and AlexA
(b) FRAN LANDers h
The basis of this predict
A small experiment
Pierrehumbert (1983), w
accents on the Fo of th
be conclusive, because t
small, depending on
presupposed.[4] That is, L
disprove a specific tw
component, but is not
model.
Moreover, data from
& Sorensen (1981), appe
quite strongly. Cooper

are no constraints on either the magnitude or the location of the impulses. It would appear mathematically possible to simulate the 'suppression' of declination simply by providing a fairly quick succession of impulses to the phrase component. Moreover, there is plenty of evidence that the *amount* of resetting after a phrase boundary is subject to linguistic constraints of one sort or another, although the nature of the constraints is far from clear (cf. Cooper & Sorensen 1981: 85ff, 120ff; Vaissière 1983: 57f; Bruce 1982). This raises the question of whether we are dealing with a phonological component and a backdrop decline at all, or whether it makes more sense to think of two separate components, one specifying local F o events and one specifying overall trends, but both governed by linguistic specifications and thus in some sense part of intonational phonology. I will return to this point in §4.2 below.

4 Testing opposing claims

It thus appears that there is evidence both for and against the two clear positions outlined at the beginning of this part of the paper; further data must be brought to bear on the choice between the two approaches. This final section of the paper outlines the kind of empirical evidence that might be used to decide between the two views.

4.1 Time dependence

One obvious testable difference between the two depends on the fact that declination in Fujisaki's or Pierrehumbert's model is a function of time. This entails the prediction that the difference in F o between two accent peaks should correlate with the length of the interval between them. For example, one would predict a greater difference in F o between the capitalised syllables in (a) than in (b):

(a) 'FANny and AlexANDer' was at the MOVies.

(b) FRAN LANDers has MOVED.

The basis of this prediction is shown in Fig. 8.

A small experiment along these lines was reported by Liberman & Pierrehumbert (1983), which found no effect of the interval between two accents on the F o of the second accent. However, their results may not be conclusive, because the size of the predicted effect could be extremely small, depending on the model of 'automatic' declination being presupposed.[4] That is, Liberman & Pierrehumbert's experiment may well disprove a specific two-component model with a large declination component, but is not necessarily evidence against *any* two-component model.

Moreover, data from a rather different experiment, reported in Cooper & Sorensen (1981), appear to confirm the prediction of time-dependence quite strongly. Cooper & Sorensen had subjects read the same set of

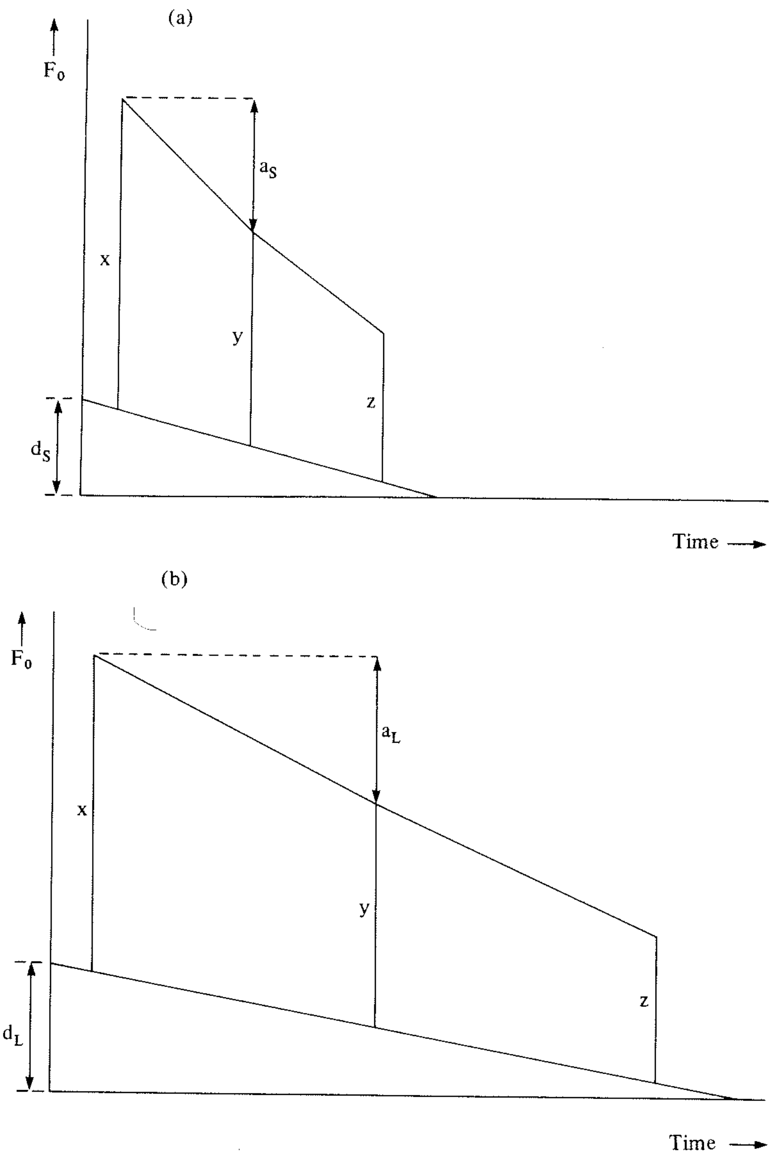
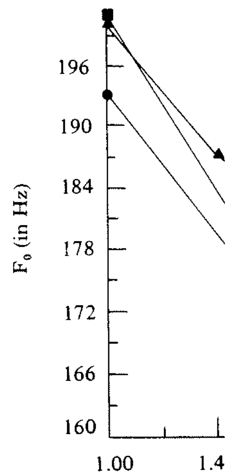


Figure 8

Topline differences based on length of interval between accents. If the declining baseline starts higher in the longer utterance than in the shorter utterance (i.e. if $d_L > d_S$) and both baselines end at the same height, and if the three accents in each utterance are scaled x, y and z units above the baseline, then accent y will be lower relative to accent x in the longer utterance than in the shorter utterance (i.e. $a_L > a_S$), because the baseline will have declined more during the longer interval.



Mean toplines obtaine
effects of speaking rate

sentences at three diffe
compared the toplines fc
Fig. 9. In addition to ce
differences in the 'shape
normal-rate topline, the
topline and lower in the
predicted by a Fujisaki
governing the peaks w
decayed exponentially as
would still be subtended
hence be higher than no
would already have deca
the second peak, which
in Fig. 10.

4.2 Tone languages

A second approach to de
the decline of an indep
phonetic studies of tone la
(e.g. the better-known le
identity or non-identity
rigorously controlled, an
gically comparable cases.

For example, one cou
sequences of high tones

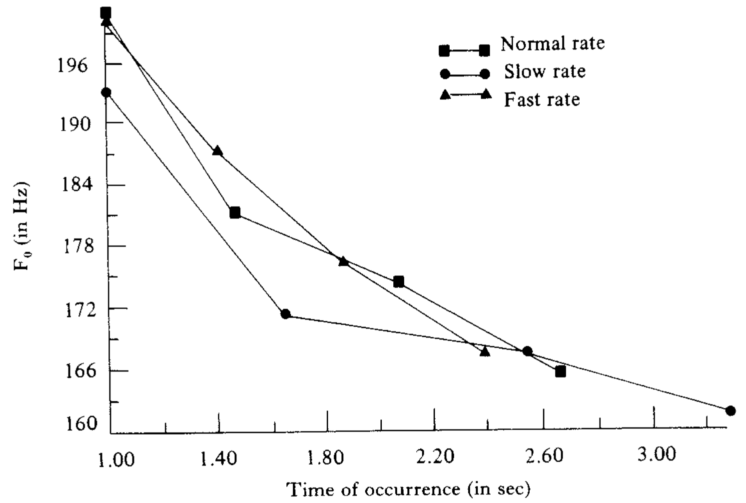


Figure 9
Mean toplines obtained by Cooper & Sorensen in their experiment on the effects of speaking rate on declination. Reproduced by permission.

sentences at three different speaking rates (slow, normal, fast), and compared the toplines for the three rates. Average toplines are shown in Fig. 9. In addition to certain differences of overall level, there are clear differences in the 'shape' of these toplines: most notably, relative to the normal-rate topline, the second accent peak is higher in the fast-rate topline and lower in the slow-rate one. This is exactly what would be predicted by a Fujisaki-style model in which the accent commands governing the peaks were overlaid on a declination component that decayed exponentially as a function of time. In fast speech, the second peak would still be subtended by the 'hump' of the declination component and hence be higher than normal; in slow speech the declination component would already have decayed away to a fairly gradual level by the time of the second peak, which would consequently be lower. This is illustrated in Fig. 10.

4.2 Tone languages

A second approach to deciding whether or not to attribute declination to the decline of an independent component would be to make detailed phonetic studies of tone languages whose tonal phonology is well understood (e.g. the better-known level tone languages of Africa). The phonological identity or non-identity of points in contours could be much more rigorously controlled, and mean values could be obtained from phonologically comparable cases.

For example, one could determine the slope of trend lines through sequences of high tones such as HHHHH, in which no phonological

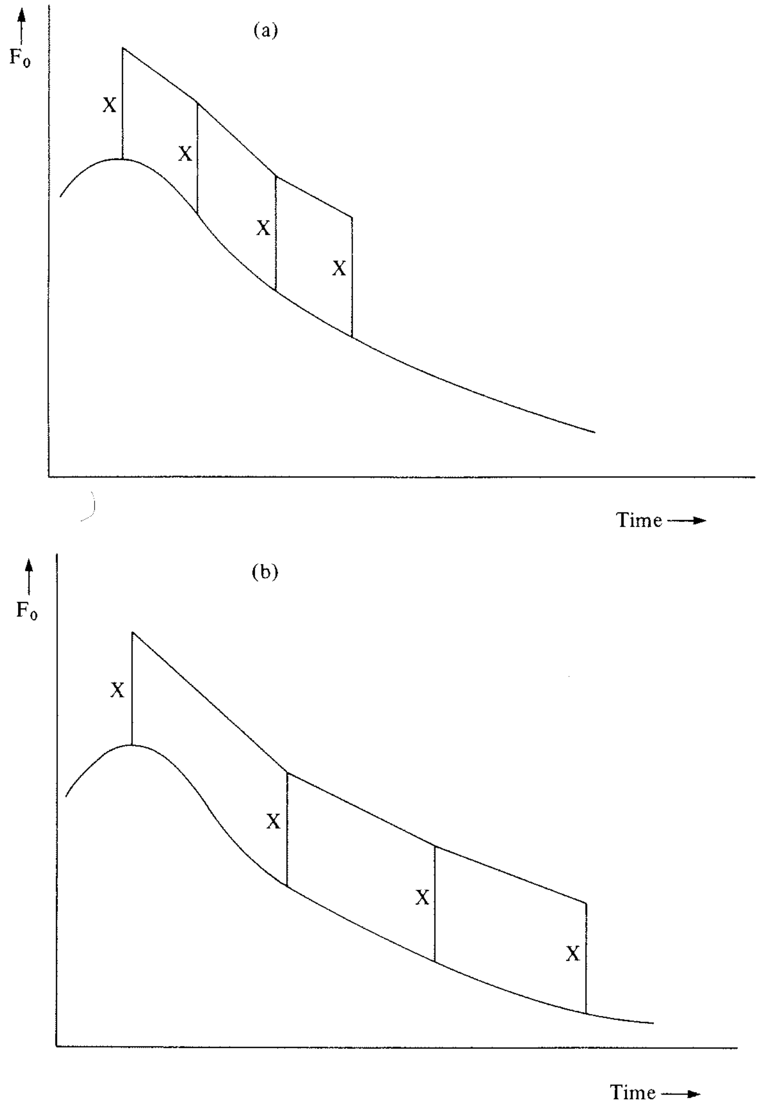


Figure 10

A possible explanation for Cooper & Sorensen's results shown in Fig. 9. Identical accent commands are superimposed on different points of a Fujisaki-style phrase component, yielding toplines of different shapes (Fig. 10a: fast speech; Fig. 10b: slow speech).

lowering rules (such as c
trend lines (as has been
that would be strong ev
component. If such tre
function only of time an
linguistic factors), that
either case, the behaviou
be incorporated into the
such as downstep and fi
decline, it would have t
tones subject to those r

A third possible outc
trend line depend on
statement) – would com
would be to assume an
the statement of the tor
would apply in question
possibility would be to
would then no longer be
certain overall characte
to the approach taken b

If studies based on
backdrop decline, then i
is caused by universal
backdrop decline into
languages.

If studies based on tor
interest would naturally
to statistical explanati
characteristic of declarat
might well be attribute
phonological rules of d
peaks. This would lend
in the sense that unrela
of Fo (lexical tone *vs.* int
rules yielding similar pl
would be comparable to
phonology, e.g. why ma

If, finally, studies bas
backdrop decline based
than statements, *all else*
richer two-component r
simply decline automat
but had linguistically m
theoretical difficulties
components are subject
Pierrehumbert (1980:

lowering rules (such as downstep) applied. If there were no decline in such trend lines (as has been suggested in some sources, e.g. Hombert 1974), that would be strong evidence against the idea of a backdrop declination component. If such trend lines showed a relatively invariant drop (i.e. a function only of time and/or duration of utterance, not correlated with any linguistic factors), that would be strong evidence in favour of the idea. In either case, the behaviour of the backdrop component – if any – could then be incorporated into the phonetic description of known phonological rules such as downstep and final lowering; if there were an invariant backdrop decline, it would have to be factored out of data containing sequences of tones subject to those rules.

A third possible outcome of such a study – that the properties of the trend line depend on phrase-level linguistic factors (e.g. question *vs.* statement) – would complicate the picture rather considerably. One solution would be to assume an invariant declination component, and complicate the statement of the tonal phonology so that different phonological rules would apply in questions and statements; this seems undesirable. Another possibility would be to enrich the role of the second component (which would then no longer be just a 'declination' component) so that it specified certain overall characteristics of different sentence-types. This is similar to the approach taken by the Lund model.

If studies based on tone languages gave evidence of an invariant backdrop decline, then it would be logical – given the idea that declination is caused by universal physiological constraints – to incorporate that backdrop decline into quantitative models of intonation in non-tone languages.

If studies based on tone languages gave no evidence of backdrop decline, interest would naturally shift to one-component models of intonation, and to statistical explanations for declination. The declining trend lines characteristic of declarative utterances read aloud in the European languages might well be attributed, as Liberman & Pierrehumbert have argued, to phonological rules of downstep and final lowering applying to the high peaks. This would lend interesting support to the statistical point of view, in the sense that unrelated languages, with different linguistic functions of F o (lexical tone *vs.* intonation), would be seen to have similar *phonological* rules yielding similar phonetic outputs. The task of explaining such a fact would be comparable to that of explaining universal tendencies in segmental phonology, e.g. why many languages have final-devoicing rules.

If, finally, studies based on tone languages gave evidence of differential backdrop decline based on sentence-type (e.g. if questions declined less than statements, *all else being equal*), then there would be evidence for a richer two-component model, in which the backdrop component did not simply decline automatically (as in Fujisaki's or Pierrehumbert's model), but had linguistically meaningful properties (as in the Lund model). The theoretical difficulties with a two-component model in which both components are subject to meaningful variation have been explored by Pierrehumbert (1980: 208ff). The basic problem is that, unless the components are highly constrained, there are too many degrees of freedom in interpreting a given contour. However, Pierrehumbert's solution – rich tonal component and invariant declination component – is not the only one possible; evidence from tone language studies might require consideration of a model with two components, both severely constrained.

*   *   *

The main point of this paper has been to argue that declination is properly a phonological problem, and can never be satisfactorily solved by the application of statistical discovery procedures. Much past work is seriously flawed by the failure to formulate clear hypotheses based on phonological models of Fo. The possibilities for future research proposed in this section are suggested as an alternative to further attempts to model Fo trend lines. They also show how declination inevitably raises one of the central issues in intonational phonology today: how an adequate general theory can incorporate both local and global properties of Fo.

NOTES

* This paper was written while I was a Visiting Research Fellow at the Laboratory of Experimental Psychology of the University of Sussex. Thanks to Steve Isard for much useful discussion of the issues treated here.

[1] I will observe a terminological distinction here between 'bottomline' – a trend line fitted to Fo valleys – and 'baseline' – an abstract bottom of the range in a theoretically coherent frame-of-reference model.

[2] Including several who are not concerned with characterising declination in terms of a trend line model, e.g. Liberman & Pierrehumbert (1984); Thorsen (1980b).

[3] It has been pointed out to me that this summary of Lieberman's thesis is inaccurate. The only intonational phenomenon that Lieberman took to be an automatic consequence of falling subglottal pressure was the drop in Fo at the very end of an utterance (the last 150–200 msec.) – not any drop during the whole utterance. He did, however, claim that subglottal pressure is the major determinant of Fo throughout the utterance, and minimised the role of the larynx.

[4] If the amount of declination is claimed to be constant regardless of utterance length, for example, little or no difference would be predicted.

REFERENCES

Bruce, Gösta (1982). Textual aspects of prosody in Swedish. *Phonetica* **39**. 274–287.

Bruce, Gösta & Eva Gårding (1978). A prosodic typology for Swedish dialects. In E. Gårding, G. Bruce & R. Bannert (eds.) *Nordic Prosody*. Lund: Gleerup. 219–228.

Cohen, Antonie & Johan 't Hart (1967). On the anatomy of intonation. *Lingua* **19**. 177–192.

Cohen, Antonie, René Collier & Johan 't Hart (1982). Declination: construct or intrinsic feature of speech pitch? *Phonetica* **39**. 254–273.

Collier, René (1975). Physiological correlates of intonation patterns. *JASA* **58**. 249–255.

Collier, René (1983). Physiological explanations of Fo declination. In A. Cohen & M. P. R. van den Broecke (eds.) *Abstracts of the 10th International Congress of Phonetic Sciences*. Dordrecht: Foris. 440.

Cooper, William E. & Joh
production. Heidelberg:
Cutler, Anne & D. Rober
Heidelberg: Springer.
Fujisaki, Hiroya & Keikicl
voice fundamental freque
In *Preprints of Papers, W*
Linguists, Tokyo. 57–70.
Fujisaki, Hiroya & S. Nag
connected speech. *Annua*
28. 53–60.
Fujisaki, Hiroya & H. Sudc
Japanese. *Proceedings of*
Gårding, Eva (1983). A g
11–25.
Gårding, Eva, Antonis Bot
Greek and French intor
137–152.
't Hart, Johan (1979). Expl
Perceptie Onderzoek, An
't Hart, Johan & René Colli
Journal of Phonetics **3**. 2
't Hart, Johan & René Colli
in Dutch. *Proceedings*
Copenhagen. 395–402.
Hombert, Jean-Marie (19
significance for a theory
Ladd, D. Robert (1983). P
Liberman, Mark & Janet
Prosody: normal and abn
Liberman, Mark & Janet Pi
in pitch range and length
structure. Cambridge: M
Lieberman, Philip (1967)
Press.
Lindau, Mona (ms). Tone
Maeda, Shinji (1976). *A*
dissertation, MIT.
Menn, Lise & Suzanne Boy
Language and Speech **25**
O'Connor, J. D. & G. F.
Longman.
Ohala, John (1978). Produ
New York: Academic P
Ohala, John (1981). The
Parasession on Language
Ohala, John & William G.
Pierrehumbert, Janet (197
JASA **66**. 363–369.
Pierrehumbert, Janet (198
dissertation, MIT.
Pierrehumbert, Janet & M
of the voice (review of
690–692.

Cooper, William E. & John M. Sorensen (1981). *Fundamental frequency in sentence production*. Heidelberg: Springer.

Cutler, Anne & D. Robert Ladd (eds.) (1983). *Prosody: models and measurements*. Heidelberg: Springer.

Fujisaki, Hiroya & Keikichi Hirose (1982). Modelling the dynamic characteristics of voice fundamental frequency with applications to analysis and synthesis of intonation. In *Preprints of Papers, Working Group on Intonation, 13th International Congress of Linguists, Tokyo*. 57–70.

Fujisaki, Hiroya & S. Nagashima (1969). A model for synthesis of pitch contours of connected speech. *Annual Report, Engineering Research Institute, University of Tokyo* **28**. 53–60.

Fujisaki, Hiroya & H. Sudo (1971). Synthesis by rule of prosodic features of connected Japanese. *Proceedings of 7th International Congress on Acoustics*, vol. 3. 133–136.

Gårding, Eva (1983). A generative model of intonation. In Cutler & Ladd (1983). 11–25.

Gårding, Eva, Antonis Botinis & Paul Touati (1982). A comparative study of Swedish, Greek and French intonation. *Working Papers in Linguistics, Lund University* **22**. 137–152.

't Hart, Johan (1979). Explorations in automatic stylization of Fo curves. *Instituut voor Perceptie Onderzoek, Annual Progress Report* **14**. 61–65.

't Hart, Johan & René Collier (1975). Integrating different levels of intonation analysis. *Journal of Phonetics* **3**. 235–255.

't Hart, Johan & René Collier (1979). On the interaction of accentuation and intonation in Dutch. *Proceedings of the 9th International Congress of Phonetic Sciences, Copenhagen*. 395–402.

Hombert, Jean-Marie (1974). Universals of downdrift: their phonetic basis and significance for a theory of tone. *Studies in African Linguistics*, Supplement **5**.

Ladd, D. Robert (1983). Phonological features of intonational peaks. *Lg* **59**. 721–759.

Liberman, Mark & Janet Pierrehumbert (1983). Paper presented at the conference *Prosody: normal and abnormal*, Zürich, April 1983.

Liberman, Mark & Janet Pierrehumbert (1984). Intonational invariance under changes in pitch range and length. In Mark Aronoff & Richard Oehrle (eds.) *Language sound structure*. Cambridge: MIT Press.

Lieberman, Philip (1967). *Intonation, perception, and language*. Cambridge: MIT Press.

Lindau, Mona (ms). Tone and intonation in Hausa. University of Lund.

Maeda, Shinji (1976). *A characterization of American English intonation*. PhD dissertation, MIT.

Menn, Lise & Suzanne Boyce (1982). Fundamental frequency and discourse structure. *Language and Speech* **25**. 341–383.

O'Connor, J. D. & G. F. Arnold (1961). *Intonation of colloquial English*. London: Longman.

Ohala, John (1978). Production of tone. In V. Fromkin (ed.) *Tone: a linguistic survey*. New York: Academic Press. 5–39.

Ohala, John (1981). The listener as a source of sound change. *Papers from the Parasession on Language and Behavior, Chicago Linguistic Society*. 178–204.

Ohala, John & William G. Ewan (1973). Speed of pitch change. *JASA* **53**. 345.

Pierrehumbert, Janet (1979). The perception of fundamental frequency declination. *JASA* **66**. 363–369.

Pierrehumbert, Janet (1980). *The phonology and phonetics of English intonation*. PhD dissertation, MIT.

Pierrehumbert, Janet & Mark Liberman (1982). Modeling the fundamental frequency of the voice (review of Cooper & Sorensen 1981). *Contemporary Psychology* **27**. 690–692.

Pike, Kenneth L. (1948). *Tone languages*. Ann Arbor: University of Michigan Press.
Sternberg, Saul, C. E. Wright, R. L. Knoll & S. Monsell (1980). Motor programs in rapid speech: additional evidence. In R. A. Cole (ed.) *Perception and production of fluent speech*. Hillsdale, New Jersey: Lawrence Erlbaum Associates. 507–534.
Thorsen, Nina (1980a). Intonation contours and stress group patterns in declarative sentences of varying length in ASC Danish. *Annual Report, Institute of Phonetics, University of Copenhagen* **14**. 1–29.
Thorsen, Nina (1980b). A study of the perception of sentence intonation – evidence from Danish. *JASA* **67**. 1014–1030.
Vaissière, Jacqueline (1983). Language-independent prosodic features. In Cutler & Ladd (1983). 53–66.

Vowel sys
typology :

Roger Lass
University of Cape

1 Fragestellung

Considering all the w
in the past half-centu
1969; Crothers 1978),
all theories of vocalic o
there are even attemp
seem to hold (from J
sophisticated treatme
etc.).

But I suggest that
empirical – in the forn
have been proposed s
critical survey anyhov
on matters such as
normalisation, and th
judice. In particular, d
of the most recent larg
of typology has been
'phonologically'.

This has led to ir
advance, and some rat
'pretheoretical' backg
'phonetic' initial app
different and more de
phonological ones. An
and basic phonologica
stem from the fact t
properly called 'vowel
problematical.

The central theoret
(i) The relation betv
do we really me
language (for inst
(ii) The status of long