Declination "reset" and the hierarchical organization of utterances

D. Robert Ladd
Department of Linguistics, University of Edinburgh, Edinburgh EH8 9LL, United Kingdom

(Received 31 August 1987; accepted for publication 4 April 1988)

Sequences of accented fundamental frequency ($F0$) peaks ("toplines") were measured for sentences of the form A and B but C and A but B and C, where A, B, and C are separate main clauses with three (four in a replication experiment) accented words each. It was hypothesized that the differences in boundary strength (but-boundary stronger than and) would be reflected in the way declination was reset following the boundaries. There were significant differences between the toplines of the two sentence types, the most consistent being that the accent peaks immediately following the two boundaries were higher if they followed but than if they followed and. In other respects the toplines were similar, exhibiting declination within each clause and across the three clauses; corresponding clause toplines that began at different levels following different boundaries nevertheless tended to end at the same level. The fact that hierarchical organization is reflected intonationally in this way makes problems for models of $F0$ in which contours result from the interaction of a number of preplanned overall trends. Suggestions are made for the incorporation of hierarchical information into models that analyze $F0$ contours as strings of abstract targets.

PACS numbers: 43.70.Fq

INTRODUCTION

A. Two approaches to modeling F0

There are two major traditions in the description of fundamental frequency ($F0$) contours. One view, more familiar to most readers of this *Journal*, draws its theoretical inspiration from harmonic analysis; it models $F0$ by superimposing component shapes or slopes, specified over domains of various sizes, to generate the complex variability of actual contours. The other approach is based in linguistics, not acoustics; it assumes that a contour can be described like other linguistic phenomena, as a structured string of linguistic elements. This divergence of approach has recently been more fully discussed in the pages of this *Journal* by Thorsen (1985, 1986).

Specific versions of the first approach include the work of Fujisaki and his coworkers (e.g., Fujisaki and Sudo, 1971), Cooper and Sorensen (1981), Thorsen (1979, 1985, 1986), O'Shaughnessy and Allen (1983), Gårding (1983), and Vaissière (1983). These are by no means equivalent either in spirit or in detail, but they typically include at least two kinds of components: (a) local $F0$ movements or changes whose effects are confined to the immediate vicinity of accented syllables and/or boundaries, and (2) a global approximation to the shape or slope of the contour spanning a phrase or utterance.

The opposing view is exemplified by a variety of theoretical and descriptive studies that are generally more concerned with affective nuances or phonological theory than with modeling details of $F0$ (e.g., O'Connor and Arnold, 1973; Halliday, 1967; Crystal, 1969; Trager and Smith, 1951; Ladd, 1983; Liberman and Pierrehumbert, 1984). The essence of these analyses is that they do not include global shapes and slopes. To be sure, in many such models there are elements like "rise" and "fall," and elements that may span several syllables, like the "head" of the traditional British description of English intonation (e.g., O'Connor and Arnold, 1973). But for purposes of the distinction being drawn here, what counts is that *no element is superimposed on any other*: The specification of the contour is entirely sequential.[1]

The only area where there might be a sufficient basis for comparing the ability of the two approaches to model $F0$ data is in studies of *declination*, the gradual $F0$ decline often observed over the course of phrases or utterances. [For a review of the literature on declination, see Ladd (1984).] In the first approach, the downward trend in $F0$ is modeled directly as some sort of global component of the output $F0$ contour (e.g., Fujisaki and Sudo, 1971; Cooper and Sorensen, 1981). In the linguistic model proposed by Liberman and Pierrehumbert (1984), by contrast, declination arises from the repeated occurrence of *downstep*—a kind of stepwise downward adjustment of the speaking range—at each accentual movement. As it happens, however, narrowly empirical considerations would appear to provide little basis for preferring one approach to the other: Models of both sorts appear equally able to generate the course of a single declination slope. Moreover, there are broader theoretical arguments on both sides of the debate. On the one hand, Liberman and Pierrehumbert suggest that models with global trends imply unrealistic amounts of detailed advance phonetic planning; their model generates declination entirely from left to right, without the need for any sort of look ahead. On the other hand, Thorsen (1980) has shown that the overall slope of a clause contour can be used by speakers

 0001-4966/88/080530-15$00.80

of Danish as a direct intonational cue to the clause's intended pragmatic effect (complete statement, incomplete statement, question), and has consequently argued more generally that, for Danish at any rate, the left-to-right model "misses the essential aspect of the production of intonational phenomena" (1985, p. 1214).

How can we progress beyond this theoretical stalemate?

B. Implications of declination reset for the two models

It is well known that longer sentences may contain more than one intonational phrase, each with its own declination slope. It is also known that declination slopes beginning after a sentence-medial phrase boundary generally start lower than those at the beginning of a sentence; that is, when declination is "reset" in midsentence, it is generally not reset all the way. This phenomenon, referred to as *partial reset* by O'Shaughnessy and Allen (1983), has seldom been studied in any detail, and few investigators have considered the implications of partial reset for their models of declination. [Cooper and Sorensen (1981, Sec. 2.3), for example, having demonstrated the existence of partial reset, decline to attempt to incorporate it into their quantitative model.]

The only serious discussion of the theoretical implications of declination reset known to me is in Thorsen's recent work (1985, 1986). Thorsen investigated Danish texts in which shorter declination domains (clauses or short sentences) were conjoined to form a single longer domain (long sentences or paragraphs). For example (Thorsen's translations of the original Danish texts), see the following.

(a) Amanda is going away camping, and her mother is taking a course in Germany, and her father is going to hike in Lapland.

(b) Amanda is going away camping. Her mother is taking a course in Germany. Her father is going to hike in Lapland.

She found that the declination slopes across the shorter domains varied: They were steeper in the case of the complete sentences and less steep in the case of the clauses separated by commas. But, in both cases, the shorter-domain slopes appear to fit the same overall downward trend across the whole text. Thorsen concludes that we must recognize a hierarchy of textual units, and that overall $F0$ trends or shapes must be specified for the units at each level in the hierarchy. That is, the overall downward trend across the whole text reflects a preplanned, superordinate declination slope that interacts with the other component slopes and configurations into which the $F0$ contour is analyzed. Thorsen rejects an alternative model, proposed speculatively by Liberman and Pierrehumbert (1984, p. 231), in which pitch range parameters could still be set from left to right, but with the left-to-right setting operating separately for each layer in the prosodic hierarchy. She claims that this approach would have difficulty with certain constancies in her data that the superposition approach handles with no trouble (Thorsen, 1985, p. 1214).

However, the case is by no means as clear as Thorsen suggests—certainly not if we attempt to generalize beyond her Danish data. The main reason that the partial resets in Thorsen's data fit a smooth downward trend (which is what makes it attractive to model her data in terms of superimposed declination slopes) is that the hierarchical organization of her texts is relatively simple. Her texts are strings of roughly equal subtexts, coordinated by <u>and</u> or simply juxtaposed. Yet syntactic hierarchies—bracketed constituent structures—can be much more complex than that, and impressionistic observation suggests that $F0$ responds to that greater complexity. If this is true, then models like Thorsen's will have at least as much trouble with the hierarchical organization of $F0$ domains as models like Liberman and Pierrehumbert's.

Consider, for example, the following pair of sentences, which are only slightly more complex than Thorsen's originals.

(a) Amanda likes Mozart, and her mother likes Schubert, but her father likes modern jazz.

(b) Amanda likes Mozart, but her mother likes Dixieland, and her father likes modern jazz.

The most natural interpretation of these sentences appears to be one in which the <u>but</u> opposes one proposition to the two conjoined by <u>and</u>; in Cooper and Paccia-Cooper's terms (1980, Chap. 7), the <u>but</u>-boundaries are *stronger* than the <u>and</u>-boundaries. Assuming that this difference of boundary strength is reflected in the $F0$ patterns, one obvious way to model the difference, given Thorsen's approach, would be to posit an intermediate-level declination slope for the two clauses conjoined by <u>and</u>, intermediate, that is, between the slope that spans the whole sentence and the slopes on the individual clauses. Alternatively, one might posit nonlinear (e.g., "convex" and "concave") slopes for the sentence-level declination component. Either way, the simplicity with which the superimposed trend model fits Thorsen's <u>Amanda</u> sentences is lost. This is not to suggest, of course, that the approach proposed by Liberman and Pierrehumbert would find such cases simple—only that the apparent suitability of superimposed trends for modeling declination within declination may be partly a consequence of the simple cases investigated by Thorsen.

In any case, it will be apparent that the details of declination reset are of considerable theoretical interest. The present study investigated $F0$ in texts where boundary strength at reset points was systematically manipulated, with the following goals: to replicate findings of declination within declination; to show that differences of boundary strength are reflected in differential reset; to provide systematic instrumental data on differential reset that can be used to formulate more specific hypotheses and more testable models of how resetting is controlled.

I. EXPERIMENT 1

A. Method

1. General

The procedure adopted here is comparable to that of several other recent studies of sentence intonation (e.g., Thorsen, 1980, 1981; Bruce, 1982; Liberman and Pierrehumbert, 1984). Controlled speech materials were recorded under laboratory conditions, and mean $F0$ values for selected points in the resulting utterance contours—computed for each speaker separately—were taken as the basic data. In order to investigate differential reset in this way, it was necessary to construct two sets of sentences that contained boundaries of different strength but were otherwise as similar as possible. The investigation then centered on comparing the "toplines" of the two sets, i.e., the values of the $F0$ peaks associated with the major accented syllables.

The sentences compared were of the form A and B but C and A but B and C, where A, B, and C are clauses of roughly parallel rhythmic and syntactic structure. In what follows, these structures will be referred to as the and/but and the but/and structures, respectively. For example, see the following sentences[2]:

Allen is a stronger campaigner, and Ryan has more popular policies, but Warren has a lot more money. (and/but)

Ryan has a lot more money, but Warren is a stronger campaigner, and Allen has more popular policies (but/and).

Like the sentences just discussed in the Introduction, these sentences have a natural interpretation in which the but-boundary is stronger, as illustrated in Fig. 1. Since all three clauses can plausibly appear in all three positions in both structures, these texts make it possible to isolate the factor of boundary strength from potential confounding variables such as differences of length and semantic content.

Preliminary auditory analysis of a variety of such sentences suggested that there would be declination across each clause and across the whole sentence (as in Thorsen's study), but that (unlike Thorsen's results) the boundary strength differences would yield systematic differences in the amount of reset. No detailed hypothesis was formulated about the reset differences, because they were extremely difficult to pin down auditorily. That is, the two structures sounded different intonationally, but it was difficult to say where the difference lay; in the event, it appears that this can be attributed partly to the subtlety of the differences, and partly to variation among speakers. The hypothesis was simply that the topline patterns of the and/but and but/and structures would differ significantly.

In addition to the topline analysis, measurements were made of the contour-final low $F0$ values at the end of each separate clause (A, B, and C). In what follows, these values will be referred to as "end points" (see Fig. 2). The endpoint analysis had three purposes. Most important was to provide data relevant to the question of whether declination within declination should be modeled in terms of superimposed overall trends, irrespective of any differential reset. A model like Thorsen's makes the obvious prediction that the clause end points should show an overall downward trend

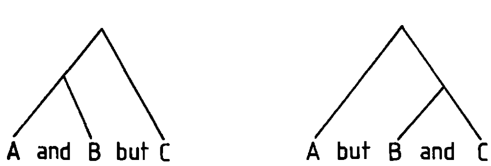

FIG. 1. General structure of the two sentence types studied in experiments 1 and 2. Clauses A, B, and C are of roughly parallel rhythmic and syntactic structure.

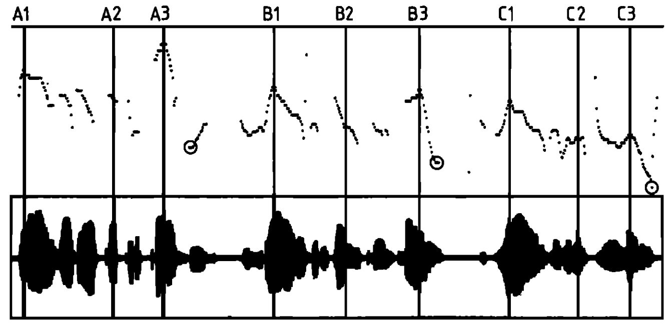

FIG. 2. Sample $F0$ contour and speech waveform from experiment 1 (speaker DM), showing how the topline points were selected. The text of the utterance (with accented syllables capitalized) is ALlen has more POPular POLicies, and WARren is a STRONGer camPAIGNer, but RYan has a LOT more MONey. The vertical cursor lines (labeled A1–C3 at the top of the figure) mark the nine topline points associated with the accented syllables. The three circled points in the $F0$ contour are the clause-final "end points." Among other things, the figure illustrates topline points taken at (a) $F0$ peak on following unstressed syllable (campaigner, point B3); (b) most level part of syllable contour that begins and ends with segmentally induced steep falls (stronger, point B2); (c) energy peak of stressed syllable in the absence of a clear $F0$ peak (lot, point C2).

across the sentence. If this is not true (e.g., if all three clause end points are scaled equally low), then the model's validity would be called into question. The other two purposes for measuring end points were: first, to replicate findings that sentence-final low end points are constant for a given speaker (Maeda, 1976; Menn and Boyce, 1982; Liberman and Pierrehumbert, 1984); and second, to test whether the scaling of end points correlated with that of the following topline peak, in the manner of the so-called fall–rise patterns in Cooper and Sorensen (1981).[3]

Finally, the durations of the two boundaries in each sentence were measured (specifically, what was measured was the combined duration of the preboundary stressed vowel, the following unstressed syllables, and the following pause). This was intended as an independent check on the assumption that the two experimental conditions did in fact involve a difference of boundary strength: There is evidence (e.g., Lehiste, 1972; Oller, 1973) that both the duration of pauses and the amount of preboundary lengthening correlate strongly with the strength of the boundaries they accompany. If boundary durations showed that speakers treated the boundaries as being of equal strength, we would expect $F0$ results comparable to Thorsen's cited above—and the experiment would obviously be judged irrelevant to the question of whether differences of boundary strength are accompanied by differential reset. Conversely, if the duration patterns showed clear evidence of different boundary strengths but there were no accompanying $F0$ differences, then doubt would be cast on the impressionistic evidence for differential reset.

2. Speakers

The speakers were four students at the University of Sussex, all male native speakers of British English in their twenties and thirties. The decision to use only male speakers was made because of the technical limitations of the $F0$ extraction programs available. Two of the speakers—one (DM) from London, the other (GH) from Liverpool—had clearly identifiable regionalisms in their speech; the other two used a more or less standard pronunciation. Speakers were paid a small sum for their participation.

Notices advertising the study stipulated that speakers should be "able to read aloud with reasonable fluency and expressiveness," and six speakers were first put through a brief screening test in which they read a paragraph of connected literary prose and several two sentence dialogues. None of the speakers was rejected on the basis of the screening test, so that the original speakers were simply the first to respond to the advertisement. However, data from two speakers (DG and XG) were not fully analyzed and are not reported below. DG failed to complete the series of recording sessions and his data were thus incomplete in any case. The decision to drop XG as well was motivated primarily by practical considerations of time, money, and computing resources available during the period of my research fellowship. The topline analyses for experiment 1 (including the ANOVA) were completed for both DG and XG, and the results do not contradict the findings reported below; the most obvious difference is that these two speaker's productions were substantially more variable (as reflected in much larger standard deviations) than the four speakers whose data are reported.

3. Speech materials

As noted above, the test sentences were of the form A and B but C or A but B and C, where A, B, and C were clauses of similar rhythmic and syntactic structure. The clauses contained one of three subjects (Ryan, Warren, and Allen) and one of three predicates (has a lot more money, has more popular policies and is a stronger campaigner); the order of the subjects and the predicates was systematically varied to help control for segmental effects on $F0$ (cf. Lehiste, 1970, pp. 71–74). In all, 18 and/but sentences and 18 but/and sentences were constructed, such that all the subjects and all the predicates occurred in all three positions (i.e., in each of the A, B, and C clauses). As a further control on segmental effects, the accented syllables to be measured were intended to contain only low and mid vowels; however, the speaker with the Liverpool accent had a high-back rounded vowel [ɷ] in place of a midcentral or midback vowel [ʌ] in the words money and governor. No attempt was made to compensate for this.

4. Recording procedure

Speakers were recruited for a series of three recording sessions, which were spread out over 6 weeks and included material from several different experiments. In each recording session, test sentences from one experiment served as filler sentences for other experiments. The materials for experiment 1 were recorded at the first recording session.

The session consisted of the 36 test sentences together with approximately 100 filler sentences of various structures. All the sentences were typewritten on slips of paper, grouped in blocks of from two to six similar sentences. (The order of the blocks was systematically varied so that no two speakers read the sentences of a given session in the same order.) The speaker was given the entire stack of slips at the beginning of the session; at the bottom of the stack were between 10 and 30 "dummy" slips, so that the speaker would not be aware when he was reading the last sentence of the session. A practice period of 20–30 sentences preceded the recording session. The whole session (including the practice period) lasted about 35–40 min.

The recordings were made in a sound-attenuated booth on high-quality equipment. Speakers controlled the tape by remote control, and were instructed to stop the tape after each sentence. They were permitted to proceed at their own pace, and were encouraged to take at least one break outside the booth during the session. They were told to repeat sentences if they stumbled or were otherwise dissatisfied with their performance, but only DM did this consistently.

Speakers were not told the specific goals of the study until after the whole series of recording sessions. In particular, their attention was never drawn to the points of syntactic difference between experimental conditions. Speakers were, of course, aware that I had advertised for speakers who could read expressively, and DM and GH knew in a very general way that I was interested in intonation.

5. Acoustical analysis

The recorded utterances were low-pass filtered at 4.5 kHz and digitized at 10 kHz. Here, $F0$ was extracted using the cepstrally based API command of the Interactive Laboratory System (ILS) software package. Measurements were made using the interactive computer graphics provided by ILS. A sample utterance contour showing some of the details of these measurements is given in Fig. 2.

(i) *Toplines*. As noted above, the investigation focused on the "toplines" of the utterance contours, i.e., on the nine $F0$ peaks—three in each clause—associated with major accented syllables. To determine toplines of this sort, it is obviously necessary to assign a single $F0$ value to each of the accented syllables under consideration. This was done according to the following criteria (see Fig. 2).

(1) If the accent was marked by a clear $F0$ peak (as perhaps half of them were), the highest $F0$ value was taken as the value of the target. This was done regardless of the alignment of the $F0$ excursion with the segmentals. For example, as can be seen in the case of campaigner in Fig. 2, $F0$ peaks were frequently aligned with the unstressed syllable following the stressed syllable, but those peak values—not the highest value of the rapidly rising $F0$ on the stressed syllable itself—were taken to represent the topline point. For theoretical justification of this procedure, see Ladd (1983, p. 729 ff.) and the references cited there.

(2) If the accent was not marked by a clear $F0$ peak (as in the case of lot in Fig. 2), then the $F0$ value at the energy peak of the stressed syllable was taken to be the topline point.

(3) However, in some cases, especially on the syllables "pop" of popular and "strong" of stronger, there was a rapid fall in $F0$ after the onset of voicing, followed by a slight leveling off, followed by another rapid fall before the following stop consonant; in these cases, the topline point value was taken from the most level part of the syllable contour, even

though the energy peak often occurred slightly earlier or slightly later, during the rapid $F0$ falls.

On the basis of these $F0$ values, mean values for each topline point—rounded to the nearest Hz—were computed for each speaker. In computing these means and in subsequent statistical analysis, the different versions of each sentence type were averaged together as if they were repetitions of identical sentences. That is, all 18 and/but sentences for a given speaker were averaged to yield a topline for the and/but structure for that speaker, and similarly for the 18 but/and sentences. This means that the different combinations of subject and predicate in each sentence set were used only as a control on segmental $F0$ effects, not as a basis for systematic comparison of sentences within each set.

In order to assess the amount of measurement error involved in the procedures just outlined, the data for one speaker were processed twice, once at the beginning of the analysis and again at the end about 2 months later. As can be seen from Table I, the rounded means for the majority of the topline points are the same in both sets of measurements. Only at one point does the difference in the rounded mean exceed 1 Hz.

(ii) *End points.* In general, the $F0$ extraction was not very reliable for end points, and all the measurements were made "by hand" using the ILS waveform editing programs, measuring the duration of three pitch periods and computing $F0$ on that basis. In cases where voicing remained regular until the end of the last vowel and there was no obvious final rise in $F0$, the last three identifiable pitch periods were used as the basis for the computation. In cases where there was an obvious final rise (mostly restricted to DM's utterances), the three longest pitch periods were visually identified and the computation was based on those. In cases where voicing was markedly irregular (e.g., creaky voice) or absent, no value was computed; consequently, most of the end-point means are based on fewer than 18 values.

(iii) *Boundary durations.* Boundary duration measurements were made with the help of the ILS cursor program. The intervals measured began with the onset of the *last stressed vowel* preceding the boundary (viz., the stressed vowels of campaigner, money, and policies) and ended with the onset of phonation for the following but or and. This procedure was adopted, rather than simply measuring the silent interval between phrases, for two reasons.

(1) To avoid any difficulties identifying the beginning of the silent interval where the signal gradually diminishes at the end of each clause.

(2) To take account of evidence that boundary pauses are closely integrated into an overall rhythm or "foot structure" governed by the location of stressed syllables (cf. Scott, 1982) and may thus interact with the preboundary lengthening of final syllables.

TABLE I. Comparison of two sets of measurements of $F0$ topline points in the data for experiment 1, speaker TH. Each value is the average of data points from 18 utterances, rounded to the nearest whole Hz.

Topline point	and/but sentences First measurement	and/but sentences Second measurement	but/and sentences First measurement	but/and sentences Second measurement
A1	188	189	187	187
A2	149	151	149	149
A3	141	141	145	145
B1	154	154	159	159
B2	131	131	134	134
B3	141	141	136	136
C1	150	151	144	145
C2	126	127	126	126
C3	133	133	135	135

B. Results

1. Toplines

Figure 3 displays mean topline and end-point values for the two experimental conditions (and/but and but/and) for

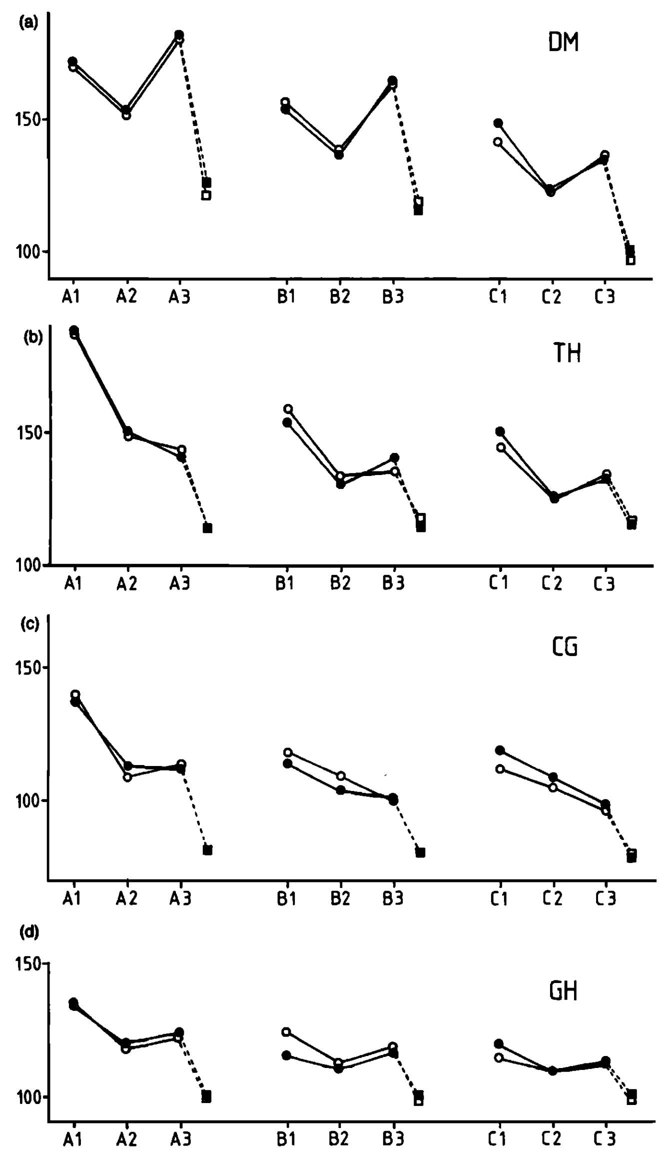

FIG. 3. Mean toplines (circles) and mean clause end points (squares) for the two sentence types in experiment 1, plotted separately for the four speakers. Filled circles and squares show the and/but condition ("[clause A] and [clause B] but [clause C]"), hollow circles and squares the but/and condition. Along the horizontal axis, topline points (accents) are numbered consecutively from left to right, separately within each clause: Thus B3 is the third accent in the second clause. The vertical axis shows $F0$ in Hz.

each speaker separately. Topline points have been numbered separately within each clause (A2, C1, etc.). To avoid cluttering up the figures, standard deviations (which range from about 12 Hz down to as little as 2 Hz) are not shown. Complete $F0$ data including standard deviations are given in Appendix A.

Table II shows the results of analyses of variance performed on each speaker's topline data: For these ANOVAs, sentence type (i.e., the experimental manipulation and/but versus but/and) was treated as a grouping factor, and the nine topline values for each utterance were treated as repeated measures, with clause (A, B, C) and accent position within each clause (1, 2, 3) treated as within-group factors. End points were not included in the ANOVA, but are discussed further below.

(i) *Declination within declination*. Inspection of Fig. 3 strongly suggests the existence of independent trends for the individual clause toplines and for the sentences as a whole. This is confirmed by the large main effects shown in the ANOVAs for clause and accent.

The trend across the whole sentence is clearly downward, in the sense that a given peak in one clause topline is almost invariably higher than the corresponding peak in the following clause and/or lower than the corresponding peak in the preceding clause (e.g., C3 is lower than B3 which is lower than A3). The trend of individual clause toplines is less clear, and seems to vary more from speaker to speaker. Some of the clause toplines (e.g., TH's clause A toplines) do show a classic declination shape, but more of them have the shape of a flattened V, with the second peak the lowest. Even these, however, might well be included as exhibiting declination, since the first peak is the highest, and since there are good reasons for suggesting that the words at the second peak (popular, a lot, stronger) have a lower degree of stress or prominence than those at the third peak (policies, money, campaigner). Only DM's pattern is genuinely different, with the third peak highest in its clause in both clause A and clause B. To the ear, it was quite plain that DM used a different tune type, with contrastive-sounding, fall–rise accents on peaks A3 and B3 where the others generally had unemphatic plain falls.

(ii) *Differential reset*. In a general way, then, findings that declination over short domains can coexist with declination over longer ones are confirmed, or at least not contradicted. We now turn to the question of whether the experimental manipulation—the difference between the and/but and but/and conditions—has produced any differences in the amount of reset. Inspection of Fig. 3 suggests that there are such differences, but the most consistent pattern is not very conspicuous and requires some discussion. Indeed, the difficulty in interpreting these results draws attention to the fact that the very definition of reset depends to a considerable extent on one's theoretical assumptions.

One obvious approach, consistent with Cooper and Sorensen's notion of "fall–rise pattern" (see footnote 3), would be to define the amount of reset as the difference between the last preboundary peak (A3 or B3) and the first postboundary peak (B1 or C1, respectively). Looking at the data this way, we see no consistency whatever: GH steps up slightly across a but-boundary and down slightly across an and-boundary. TH steps up by similar amounts across both types of boundary. DM steps down by similar amounts across both types of boundary. (This is obviously related to the shape of DM's clause toplines, but that fact in no way alleviates the problem of defining reset.) CG steps up negligibly across the first boundary of the sentence, regardless of whether it is and or but, and steps up markedly across the second boundary, more for but than for and. In short, there is no obvious characteristic that distinguishes the course of $F0$ across the and-boundary from that across the but-boundary.

Another possible definition of reset that might reveal consistent differences between experimental conditions would be to look only at the postboundary peaks (B1 and C1), and to compare the peak that follows but with the one that follows and in the same experimental condition. This is the approach most consistent with a theory (like Thorsen's) that seeks to fit individual declination slopes to an overall trend across the sentence, and it makes somewhat more sense of the data than the approach based on Cooper and Sorensen. For all speakers, the post-but peak (B1) is higher than the post-and peak (C1) in the but/and condition. In the and/but condition, however, there is no clear pattern; some of the speakers do have the post-but peak (C1) higher than the post-and peak (B1), but others do not. Moreover, even if

TABLE II. F values from ANOVA on experiment 1 topline data. Results marked *a* approach significance ($0.05 < p < 0.10$). The asterisks indicate the level of significance.

Main effects	*df*	DM	TH	CG	GH
Sentence type					
(and/but, but/and)	1	0.42 n.s.	0.02 n.s.	0.25 n.s.	0.51 n.s.
Clause (A, B, C)	2	630.30***	391.75***	101.83***	190.81***
Accent (1, 2, 3)	2	289.17***	864.98***	262.74***	280.85***
Interactions					
Clause × S type	2	1.70 n.s.	1.18 n.s.	5.30**	11.49***
Accent × S type	2	0.34 n.s.	0.50 n.s.	0.18 n.s.	2.57 *a*
Clause × accent					
× S type	4	2.42 *a*	6.65**	3.74**	12.74***
Clause × accent	4	29.63***	74.21***	27.04***	25.35***

there were a completely consistent result here (i.e., post-but peak always higher than post-and peak, regardless of experimental condition), it would still not be very obvious how to use overall trends across the sentence to model such specific comparisons between points in the contour (cf. the suggestions made in the Introduction).

The most consistent pattern emerges when we define reset only in terms of the postboundary peaks (B1 and C1), and compare *across experimental conditions* at the same position in the sentence—that is, compare the height of B1 in and/but with B1 in but/and, and of C1 in and/but with C1 in but/and. Looking at the data in this way, we see that in every case the postboundary peak is higher after the but-boundary than after the and-boundary at the same position in the sentence. There is no comparable difference—indeed, virtually no difference at all—between experimental conditions on peak A1. Nor is there any consistent effect on the second and third peaks in any of the clauses, though certain marked differences do occur on these peaks in individual speakers. That is, the most consistent difference between the two experimental conditions appears to be concentrated on points B1 and C1. This pattern can be seen very clearly in GH's data in Fig. 3, but it applies to a greater or lesser extent to all speakers in both conditions.

If the most significant differences between the two experimental conditions are in fact concentrated on the two postboundary peaks in the way just outlined, this should be revealed in the pattern of interactions in the ANOVA. Specifically, we would expect to find some clause-by-sentence-type interaction (because the differential reset would affect the clause means), and an even clearer clause-by-accent-by-sentence-type interaction (because of the effect on the first accent of clauses B and C, but not A), but no accent-by-sentence-type interaction at all (because the effects of sentence type on accents B1 and C1 would tend to cancel each other out in the means). This is exactly the pattern of results in the ANOVA, as can be seen from Table II.

2. End points

The most conspicuous fact about the end points is that, except in DM's data, they show no obvious effect of experimental condition, nor any downward trend. This can be seen in Fig. 3 (for details, see Appendix B). It would be problematical to incorporate this finding into a description that models the downward trend across the whole sentence in terms of an overarching declination slope, because such a model would appear to predict a downward trend across the end points as well. But the data from TH, CG, and GH show that it is possible to have a clause-final fall to sentence-final baseline level without interrupting the long-domain trend manifested in the toplines.

DM, however, does behave in accordance with the most obvious predictions of the superimposed trend models: His end points show an overall downward trend across the sentence. Conceivably this is related to the fact that he mostly used fall–rises (in the standard sense, not the Cooper–Sorensen sense) at peaks A3 and B3, and to the fact that his A3 and B3 peaks were higher than A1 and B1, respectively. Moreover, DM's end points also seem to show an effect of experimental condition: They are higher before an and-boundary than before a but-boundary. This is the opposite of what would be predicted if we define reset in terms of the Cooper–Sorensen notion of boundary-spanning "fall–rise pattern": Given that the postboundary peak is higher after but than after and, we would expect that the preboundary valley should be higher as well.

3. Boundary durations

Boundary durations, defined as discussed above, are shown for the four speakers individually in Table III. In order to assess the significance of these data, an ANOVA was performed with sentence type (and/but versus but/and) as a grouping factor and conjunction type (and or but) as a within-group factor. Results (given in Table IV) show that but-boundaries are significantly longer for GH and TH, and longer for CG when one takes into account the effect of position in the sentence. This seems to confirm that the differences between the and/but and but/and structures may be treated as involving differences of boundary strength.

However, it can also be seen that DM's boundary durations show no significant effect of boundary strength. It is noteworthy that DM also showed the least difference in his toplines. By contrast, GH, whose topline effects were very striking, also shows large and significant differences of boundary duration.

C. Discussion

There are two noteworthy facts about the pattern of differential reset observed in this experiment.

First is the fact that the reset differences appear to be concentrated on peaks B1 and C1. This causes particular problems for Liberman and Pierrehumbert's model, and lesser problems for a number of other models of declination. The problem is that one would expect to find that greater height on the initial peak of a declination domain was propagated down the slope, as it were, all the way to its last peak (e.g., if B1 is higher, then B2 and B3 should also be higher). This does not appear to be the case anywhere in the topline data except for CG's clause C toplines. Moreover, to the extent that the final peaks of each clause are unaffected by

TABLE III. Mean boundary pause durations (in ms) and standard deviations for experiment 1. For all cases except GH's and/but sentences (one of which was inadvertently omitted by the speaker), $n = 18$. See the text for a description of where the intervals began and ended.

		and/but		but/and	
Speaker	Boundary	Mean	s.d.	Mean	s.d.
DM	and	665	121	643	78
	but	683	91	666	115
TH	and	475	77	476	105
	but	611	114	554	119
CG	and	763	144	828	113
	but	892	91	811	103
GH	and	537	160*	477	153
	but	666	171*	671	221

*$n = 17$.

TABLE IV. *F* values for ANOVA on pause durations, experiment 1. The asterisks indicate the level of significance.

	df	DM	TH	CG	GH
Sentence type (and/but versus but/and)	1	0.84 n.s.	1.45 n.s.	0.12 n.s.	0.76 n.s.
Boundary (and versus but)	1	0.58 n.s.	16.58***	3.40 *a*	9.63**
Boundary×S type	1	0.01 n.s.	1.25 n.s.	5.70*	0.41 n.s.

the experimental manipulation—as is strongly the case for GH and DM, and somewhat true for CG—then it appears that the declination "knows where it is going." This is easy to model in Cooper and Sorensen's model, but problematical if declination is described in terms of a downstep mechanism with no look ahead. The only obvious solution given Liberman and Pierrehumbert's approach is to postulate the existence of one or more *independent* parameters that interact with downstep to affect the scaling of each topline point; for example, one might postulate "initial raising," analogous to Liberman and Pierrehumbert's final lowering, as the mechanism behind the higher reset after but boundaries.

The other noteworthy feature of the results is the fact that the differential reset is best defined by comparing across experimental conditions. This means that the size of a given reset must in effect be evaluated relative to something that is not there. The most plausible way of doing this is to treat the postboundary peak as a *target* of some sort, calibrated relative to the range of the speaker's voice. This notion of target is entirely in keeping with a model of intonation in which contours are generated as sequences or strings of local "intonational events," but is more difficult to express naturally in terms of superimposed overall trends. I have described elsewhere (Ladd, 1983) how phonological features might be used to provide abstract invariant characterizations of intonational targets. Ultimately, it may be possible to map combinations of such features onto $F0$ target values relative to a speaker's voice range, in much the same way as features like front/back or rounded/unrounded, used to distinguish segmental phonemes, can be mapped onto target values for formants. (A model loosely based on this theoretical approach, which successfully generates natural-sounding intonation in speech synthesis, is outlined in Ladd, 1988; forthcoming.) Obviously, however, the details are not very clear, and the question is left for future research.

II. EXPERIMENT 2

Both problems just discussed seem to hinge on the question of how peak values are determined at the beginning of a declination domain. In virtually all models, the first value in a declination slope is either an anchor point (one of the points with reference to which a trend line is drawn; cf. 't Hart, 1979; Maeda, 1976; Cooper and Sorensen, 1981) or a kind of "initialization" of the function that generates the rest of the slope (Fujisaki and Sudo, 1971; Liberman and Pierrehumbert, 1984). Either way, there is strong presumption of *dependency* between the first peak and the other peaks in the declination domain. Yet, as the foregoing discussion makes clear, experiment 1 suggests that the first peak can be quite independent of the rest of the slope. Consequently, in experiment 2 the length of the individual clauses in the test sentences was modified, since length of sentence has often been reported (e.g., 't Hart, 1979; Cooper and Sorensen, 1981) to have some effect on the course of declination.

A. Method

Experiment 2 was meant primarily as a replication of experiment 1, with the sentences modified by the addition of a fourth accented word to each clause in the test sentences. This modification was achieved by changing the subjects Allen, Warren, and Ryan to Governor Allen, Senator Warren, and Congressman Ryan. In all other respects, the set of test sentences was identical to that used for experiment 1, though different filler sentences were used. Recording and analysis procedures were also identical except that boundary durations were not measured. The recordings were made about 3 weeks after those for experiment 1. Results are reported for the same four speakers as in experiment 1.

The lengthening of the test sentences was intended to shed light on the following two questions related to the scaling of domain-initial topline points.

(1) Where, if anywhere, is the increased length of the clause declination domains reflected in differences of $F0$, relative to the results of experiment 1: in the first topline peak? in the last topline peak? in the end point? The literature provides a basis for a wide range of contradictory expectations. On the basis of 't Hart's model and results from Cooper and Sorensen (1981), we might expect that the first peak would be scaled rather higher in the longer clauses, but that the final peak would be unaffected. Results from Sternberg *et al.* (1980), Liberman and Pierrehumbert (1984), and Thorsen (1980, 1981) would lead us to expect little or no difference at the first peak, but a lower final peak in the longer sentences. Maeda (1976), finally, would lead us to predict that clause length should have no effect either on the beginning or the end of the topline.

(2) Assuming that the differential reset effect from experiment 1 is replicated, which accents does it affect? Does it apply only to the first postboundary accent (Governor, Senator, Congressman), only to the more strongly stressed second accent (Allen, Warren, Ryan), or does it apply to both? If it applied only to the first or only to the second, then the tentative conclusion from experiment 1—that boundary strength can affect individual topline points independently of the course of $F0$ over the rest of the declination domain—would be very much strengthened. If it applied only to the second, it would also provide further evidence of the relevance of hierarchical organization for $F0$, since the two top-

line points of the subject noun phrases are within a constituent of each declination domain. If, on the other hand, it applied to both, and especially if it applied more to the first than to the second, the earlier conclusion would be weakened; such a result might be taken as evidence that, to at least some extent, differential reset *is* propagated down the declination slope.

The two foregoing aspects of experiment 2 were largely exploratory. The only specific prediction that was made with some confidence about the results of experiment 2 was that the sentence-final end points would remain constant for each speaker: Constancy of end points has been found in a number of studies (Liberman and Pierrehumbert, 1984; Menn and Boyce, 1982; Maeda, 1976), and this constant value has been used as a speaker-specific "baseline" in a variety of $F0$ models (e.g., Liberman and Pierrehumbert, 1984; Ladd *et al.*, 1985; Ladd, 1987).

B. Results and discussion

1. Toplines

Mean toplines for the two experimental conditions (and/but and but/and) are shown separately for each speaker in Fig. 4. Complete $F0$ data, with standard deviations, are given in Appendix A. As in experiment 1, the data points are numbered separately within each clause; the four

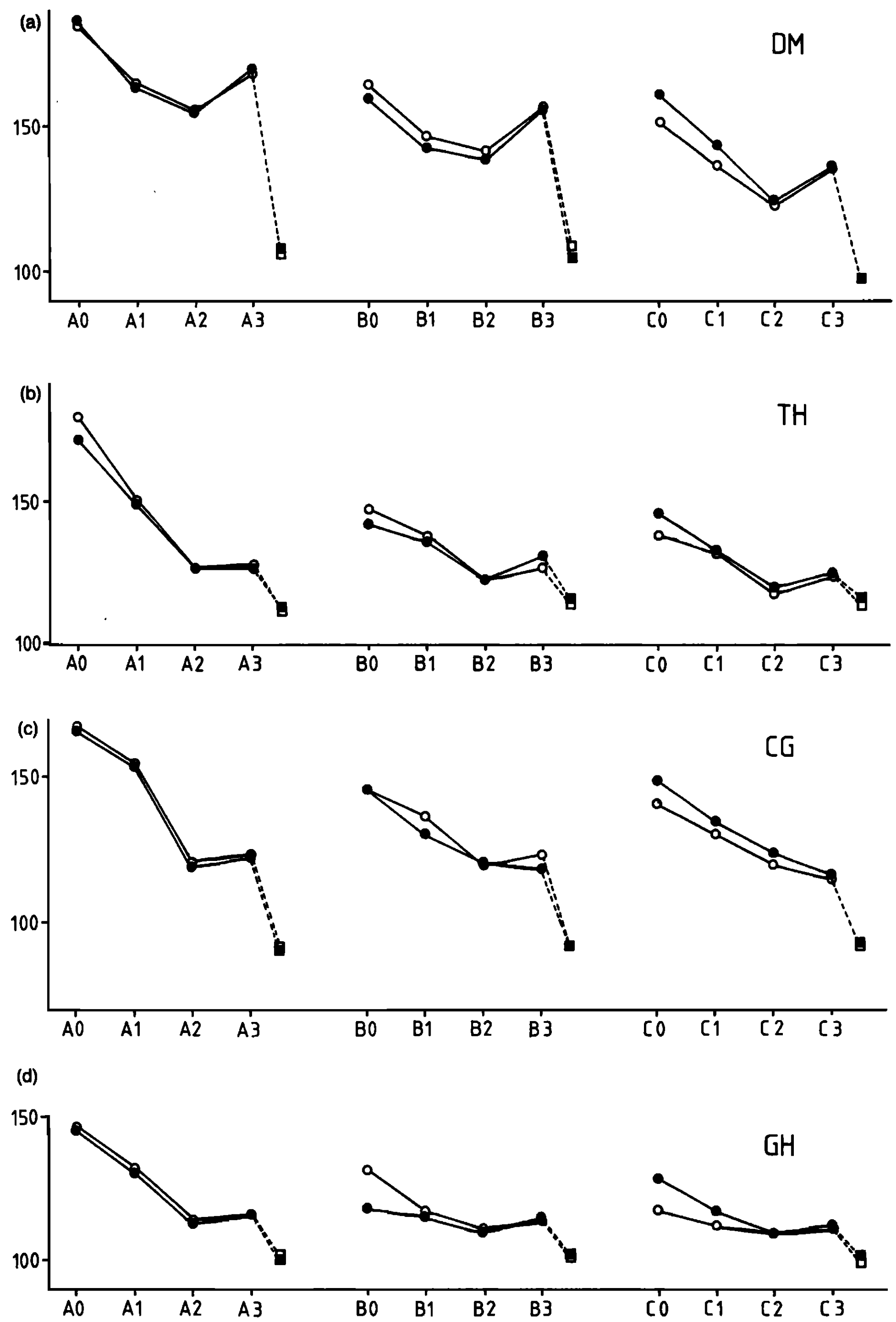

FIG. 4. Mean toplines (circles) and mean clause end points (squares) for the two sentence types in experiment 2. The first accent in each clause is numbered 0 to facilitate comparison with the results of experiment 1. For more detail, see the legend to Fig. 3.

topline points in each clause are numbered 0–3, rather than 1–4, to facilitate comparison with the results of experiment 1. Table V displays the results of ANOVAs carried out as in experiment 1.

In the following respects, the results of experiment 2 replicate those of experiment 1.

(1) Each point in clause toplines B and C is generally lower than the corresponding point in the preceding clause topline, suggesting a downward trend across the entire utterance.

(2) The section of each clause topline formed by points 1, 2, 3 is roughly like a flattened V, with point 2 lower than 1 or 3 in most cases; together with (1), this suggests the existence of independent trends for the clause toplines and the whole sentence.

(3) The clause toplines begin higher after a but-boundary than an and-boundary, comparing across experimental conditions at the same position in the utterance (but see further on this point below). The pattern of main effects and interactions in the ANOVAs is comparable to that for experiment 1, except that in experiment 2 there are also significant clause × sentence-type interactions. This difference is presumably related to point (2) in the following paragraph.

With regard to the increase in the length of the clause toplines from three to four accents, the two experimental questions yield the following answers.

(1) In general, relative to the data of experiment 1, the first point of each clause is scaled somewhat higher (i.e., A0, B0, and C0 of experiment 2 are higher than A1, B1, C1 of experiment 1), while the last point of each clause (A3, B3, C3) is unaffected. This is seen very clearly in GH's data and to a lesser extent in DM's. This strengthens the general conclusion of experiment 1 that changes in the scaling of an initial topline point are not necessarily propagated all the way down the topline. However, detailed comparison (see Appendix A) also shows that the overall level of TH's and CG's data is higher in experiment 2 than in experiment 1. This obviously complicates the direct comparison between experiments, and will be discussed further in Sec. II B 2.

(2) Although the differential reset shows up most strongly on the first postboundary topline points B0 and C0, there are also smaller effects on the second accents of each clause (B1 and C1). That is, the differential reset does appear to be propagated along the slope to a greater extent than in experiment 1. Possibly the difference between the two experiments is due to the fact that in this experiment the second-position accents (B1 and C1) are on the syntactic head of the subject noun phrase and, therefore, more strongly stressed, while in experiment 1, the second-position accents in question were the lowest pitched, more weakly stressed B2 and C2. In any case, the data clearly provide no support for the hypothesis that the effect is strongly localized on a single accent. At the same time, however, there is no evidence that the differences on B0 and C0 are propagated all the way down to B3 and C3; as in experiment 1, there is no consistent effect of experimental condition on the scaling of B3 and C3. Further research is needed to determine how domain-initial differences affect subsequent points in the declination domain.

2. End points

In most respects, as can be seen from Fig. 4, the end-point data from experiment 2 replicate the results of experiment 1: There is no effect of experimental condition; the sentence-final end-points are very consistent for a given speaker; and there is no significant decline across the clause end points except in DM's data. However, there is one important respect in which these data confound expectations: The speaker-specific values, though consistent within each experiment, differ significantly (and considerably) between experiments for both CG and TH. (This can be seen most clearly by comparing the data from the two experiments in Appendices A and B.) Note that CG has higher values for experiment 2 and TH for experiment 1, so that this is unlikely to be an effect of the difference in sentence length between the two experiments. If this finding can be replicated, it would mean abandoning the idea that the baseline can be seen as a constant speaker characteristic. It would not, of course, rule out the weaker but no less interesting hypothesis that the baseline, though it may vary from day to day or change gradually over time, is constant enough under constant conditions to serve as some sort of speaker reference value for $F0$ scaling.

TABLE V. *F* values from ANOVA on experiment 2 topline data. The asterisks indicate the level of significance.

Main effects	*df*	DM	TH	CG	GH
Sentence type (and/but, but/and)	1	0.15 n.s.	0.03 n.s.	0.10 n.s.	0.03 n.s.
Clause (A, B, C)	2	1025.80***	315.94***	95.04***	255.95***
Accent (0, 1, 2, 3)	3	484.13***	819.13***	406.24***	1064.07***
Interactions					
Clause × S type	2	19.26***	10.65***	7.64**	26.00***
Accent × S type	3	1.34 n.s.	1.50 n.s.	0.81 n.s.	1.87 n.s.
Clause × accent × S type	6	3.13**	4.26***	1.31 n.s.	23.70***
Clause × accent	6	20.02***	59.60***	38.18***	84.16***

In any case, the difference in end points suggests that it may not be valid to compare topline data from one experiment to another. Possibly some sort of normalization could be developed in which topline values could be expressed relative to end-point values—for both CG and TH it is obvious that higher end points are associated with higher topline values—but, except in this very general way, the data here provide insufficient basis for formulating such a normalization. This matter is also left for future research.

III. EXPERIMENT 3

Throughout the foregoing discussion, we have assumed that the differences between experimental conditions could be attributed to the differences of boundary strength between the but/and and and/but structure. There is, however, another possible explanation that could not be controlled for, given the speech materials for experiments 1 and 2: Conceivably, the results reflect some local effect of the word but itself rather than the difference in hierarchical structure. Such a local effect might have been either essentially phonetic (e.g., raised $F0$ following the final stop of but; note that and was pronounced with at most a weakly articulated final stop) or functionally determined (e.g., some sort of "contrastive stress" after but). In order to test this explanation, a small follow-up experiment was conducted.

A. Method

Sentences were designed in which the final topline section was always immediately preceded by but, but the hierarchical structure differed between the two experimental conditions. Pairs of sentences of the following sort were used.

(1a) Their dog was ten years old, and had a lame leg and a white muzzle, but still kept a fierce watch.

(1b) They had a ten-year-old dog, who had a lame leg and a white muzzle, but still kept a fierce watch.

Sentences of the first type approximate the and/but condition of experiments 1 and 2—i.e., the stronger boundary is the one preceding but—and these are referred to here as the and/but sentences. Those of the second type have the stronger boundary at the first major break: The phrase introduced by but is entirely within the relative clause introduced by who. These will be referred to as the who/but sentences. If the results of experiment 1 and 2 are attributable to some local effect of but (or if, more generally, they are not attributable to differences of boundary strength), then we would not expect any difference between the and/but and who/but toplines. If, on the other hand, the difference in boundary strength is the determining factor in the earlier results, then we should observe a difference in the amount of reset at the beginning of the third clause even though it is preceded by but in both experimental conditions.

Four such pairs of sentences were designed. Unlike experiments 1 and 2, where there were no exact repetitions of any sentence, in this experiment speakers read each sentence a total of three times. The results for each experimental condition are, therefore, based on 12 utterances, three repetitions each of one member of the four pairs. The use of four different sentence pairs was intended as a control on segmental effects. Note that the peak following but is in every case on the word still.

The recordings were made about 3 weeks after those for experiment 2. As in the other two experiments, the test sentences were recorded in sessions consisting of over 100 sentences, arranged in short blocks of related sentences in systematically varying orders. The blocks of test sentences consisted of only three sentences, from three different pairs, in order to minimize awareness of the fact that sentences were repeated. The same five speakers were recorded as in experiment 2, and data from the same four were analyzed.

Here, $F0$ was extracted, topline points were determined, and mean values were calculated in exactly the same way as in experiments 1 and 2. Since the goal of the experiment was to provide a control on the conclusions of experiments 1 and 2 and not to explore the toplines of these new sentences in their own right, measurements were made of only four topline points. For the sentence pair just given, these four points were: . . . had a lame leg and a white muzzle, but still kept a fierce watch. The other three sentence pairs, showing the topline points measured, were the following.[4]

(2a) Her watch was an antique, and had a scratched face and a bent minute hand, but still kept perfect time.

(2b) She had an antique watch, that had a scratched face and a bent minute hand, but still kept perfect time.

(3a) The farmer down the road was retired, and no longer did much work, but still kept a flock of chickens.

(3b) Down the road lived a retired farmer, who no longer did much work, but still kept a flock of chickens.

(4a) The car was running again, and made it back to the campsite, but still kept burning oil.

(4b) We were concerned about the car, which made it back to the campsite, but still kept burning oil.

B. Results and discussion

Results are given in Fig. 5 in the same format as for experiments 1 and 2. The four topline points are simply numbered 1–4. As a first test of the hypothesis that boundary strength had an effect on the reset at the beginning of the third phrase, t tests were performed for each speaker separately on the difference between the means for point 3 in the two conditions. Only for TH was there a significant difference.

However, inspection of Fig. 5 shows several other large differences between the two experimental conditions—different differences for each speaker—so it was decided to perform an ANOVA on each speaker's data, treating the four topline points as repeated measures. The results are given in Table VI. The existence of sentence type by accent interactions for three of the four speakers, together with the fact that only one speaker shows a main effect for sentence type, suggests that the differences in hierarchical structure *are* reflected in localizable differences of $F0$. To give a crude summary based on Fig. 5, these differences (for everyone but TH) involve raising some or all of the topline on the *second* phrase in the who/but condition, rather than, as originally predicted, raising the beginning of the *third* phrase in the and/but condition. Though diverging from the specific prediction, this result is still entirely consistent with the more

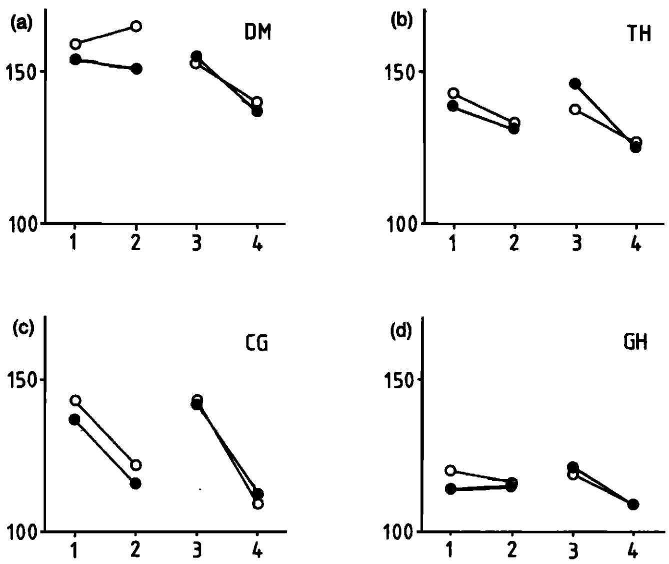

FIG. 5. Partial toplines for the two sentence types in experiment 3. Filled circles show the and/but condition, hollow circles the who/but condition. Along the horizontal axis, points 1 and 2 are the first and last accents of the second clause and points 3 and 4 are the first and last accents of the third clause. Point 3 is always on the word still in the different texts.

general hypothesis that boundary strength, not the presence of but, was the key factor in the results of experiments 1 and 2. The fact that the individual speakers' toplines differ so much in detail, even while all reflecting boundary strength in some way, is certainly worth further investigation, but is beyond the scope of the present discussion.

IV. SUMMARY AND GENERAL DISCUSSION

The experimental results reported here clearly confirm the impressionistic data discussed in the Introduction, and point to the conclusion that the hierarchical organization of texts is reflected in some detail in the course of $F0$. Experiments 1 and 2 confirm findings of "partial reset" or "declination within declination" in sentences consisting of three intonational phrases. However, they also show that the sentence's hierarchical structure—the way the individual clauses (intonational phrases) are bracketed together—has an effect on the *amount* of reset: Clause-initial accent peaks are higher following a stronger boundary. This effect is to at least some extent localized on the clause-initial peak, so that clause-final peaks are largely unaffected by the experimental manipulation of hierarchical organization. Experiment 3 shows that these conclusions cannot be attributed to the fact that the stronger boundaries in experiments 1 and 2 are always followed by but, and provides further evidence that hierarchical organization has a complex set of effects on $F0$.

As pointed out throughout the article, this general conclusion makes problems for both of the general approaches to modeling $F0$ sketched in the Introduction. *Pace* Thorsen, I see nothing in the data to suggest that a model like hers somehow captures more of "the essential aspect" of this general phenomenon than Liberman and Pierrehumbert's approach. In my view, in fact, linear models like Liberman and Pierrehumbert's are potentially better able to cope with the kinds of effects found here than models that analyze contours into superimposed component $F0$ trends and configurations. In this final section, I offer some speculations about the most appropriate way to incorporate hierarchical structure into a linear model of intonational phonology.

As I noted above, for Liberman and Pierrehumbert, the main mechanism of declination is downstep, a stepwise lowering of some parameter in the model that implements intonational phonology in actual $F0$ values. Downstep is viewed in their model as a linear relation between two adjacent accents, the second being scaled relative to the level of the first. I believe that the results reported here could be accommodated in a somewhat more powerful model in which downstep is a relationship between constituents in a metrical tree ("metrical" in the sense of Liberman and Prince, 1977, and much work since then; for present purposes the metrical tree can be regarded as largely equivalent to a syntactic tree). The downstep relationship between two adjacent accents could be diagramed as shown in Fig. 6(a). A succession of downsteps across several accents could be represented by a repeatedly right-branching tree of the sort shown in Fig. 6(b).

Proposals for such a metrical representation of downstep in African tone languages (where the phenomenon is widespread) have been made independently by Clements (1983) and by Huang (1980). These proposals have been criticized by Pierrehumbert (1980) on the grounds that they make it possible to represent *nonlocal dependencies*, that is, dependencies between nonadjacent items in the phonologi-

TABLE VI. F values from ANOVA on experiment 3 topline data. The asterisks indicate the level of significance.

Main effects	*df*	DM	TH	CG	GH
Sentence type					
(and/but, who/but)	1	8.00**	0.04 n.s.	1.42 n.s.	1.58 n.s.
Accent (1–4)	3	26.56***	39.40***	74.58***	23.25***
Interactions					
S type × accent	3	3.41*	4.75**	1.29 n.s.	2.78*

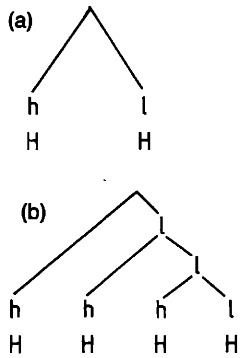

FIG. 6. Abstract phonological structures representing downstep between adjacent high tones [downstep is the source of most "declination" according to Pierrehumbert (1980) and much subsequent work]. The high tones (which represent the high peaks that manifest many pitch accents in English) are symbolized by upper-case H; the relative height of the nodes in the branching structure (high versus low) is symbolized by lower-case h and l. (a) The second high tone is downstepped relative to the first and (b) shows a string of progressively downstepping high tones. For more detail on these structures and how they are mapped onto actual $F0$ values, see Ladd (1988).

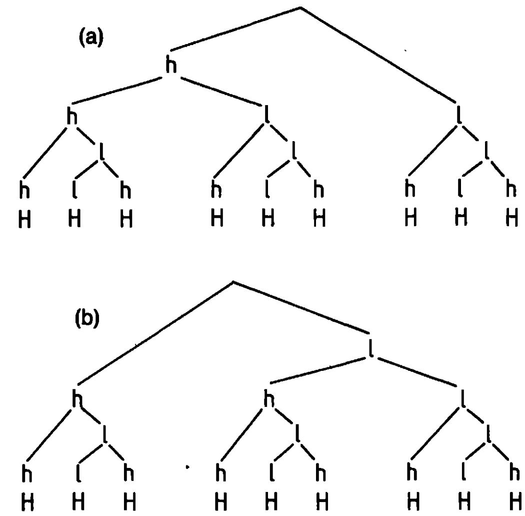

FIG. 7. Abstract phonological structures based on the simple ones in Fig. 6, representing possible topline patterns in the sentences of experiment 1. The different branching structure at the top of the trees reflects the differences between the and/but (a) and but/and (b) sentences. Within the subtrees for the individual clauses, the pattern of high (h) and low (l) nodes shown here will make the first accent the highest and the second the lowest, with the third intermediate, as is generally the case in the utterances of GH and TH. Different patterns would be needed to represent CG's and DM's typical clause toplines.

cal string. (Nonlocal dependencies are commonplace in syntax—for example, the dependency of main verb on subject in The girls who won the swimming competition come from Glasgow versus The girl who won the swimming competition comes from Glasgow—but they are less familiar in phonology and especially in acoustic phonetics. A major theoretical point of Pierrehumbert's thesis is that there are no nonlocal dependencies in intonation.) I would argue that the ability to represent nonlocal dependencies is precisely why a metrical representation is needed: There is by now ample evidence of nonlocal dependencies in $F0$ target scaling [e.g., the work of Kutik *et al.* (1983) on the recovery of toplines after a parenthetical], of hierarchically based differences of accent peak scaling in Japanese noun-phrase compounds (Kubozono, 1985, 1988), and of hierarchically based look ahead in the scaling of topline targets at the beginning of declination domains (Ladd and Johnson, in press).

One way to interpret the results reported here is in terms of a nonlocal dependency between the first topline peak of one clause and the first topline peak of some other clause. As a first approximation, we might suggest structures like those shown in Fig. 7 to represent the results of experiment 1. In the but/and condition [Fig. 7(b)], point B1 is downstepped relative to point A1, and point C1 is downstepped relative to point B1: The three points are the highest of their respective domains and the three domains are organized in a right-branching tree like the one in Fig. 6(b). In the left-branching and/but condition [Fig. 7(a)], on the other hand, both point B1 and point C1 are downstepped relative to point A1. This would predict that in the but/and condition, point B1 should be higher than point C1 (which was true for all speakers in both experiments 1 and 2), while in the and/but condition, point B1 and point C1 should be about the same (which is consistent with the fact that some speakers had C1 higher than B1 and some had B1 higher than C1). Note also that the downtrending V-shaped individual clause toplines of experiment 1 are represented in Fig. 7 by nesting a low–high constituent within a high–low one.

Obviously, the proposal illustrated in Fig. 7 is extremely speculative and leaves a good many problems unsolved and details undiscussed. (For example, it provides no obvious way to deal with the constancy of the clause-final peaks, especially C3.) But it suggests that it is indeed possible to incorporate detailed hierarchical information into an essentially linear model of intonational phonology. Declination reset, far from providing evidence for a model that decomposes $F0$ contours into superimposed components, seems to work in ways that are entirely compatible with the formal representations of current phonological theories.

ACKNOWLEDGMENTS

The work reported here was carried out while I was a Visiting Research Fellow at the Laboratory of Experimental Psychology at the University of Sussex. I am indebted to Steve Isard and Kim Silverman for useful discussions and to Ruta Noreika and Alan Richomme for assistance with the analysis.

APPENDIX A

TABLE AI. Mean topline point $F0$ values and standard deviations for experiments 1 and 2. The experimental conditions (rows) are labeled, e.g., 1AB for "experiment 1, and/but condition."

		A0	A1	A2	A3	B0	B1	B2	B3	C0	C1	C2	C3
DM	1AB	···	172 (6.9)	154 (10.0)	182 (10.5)	···	154 (4.2)	137 (6.8)	165 (11.0)	···	149 (6.7)	124 (10.1)	135 (6.1)
	1BA	···	170 (4.5)	152 (7.9)	180 (10.2)	···	157 (5.7)	139 (7.9)	163 (9.2)	···	142 (9.1)	123 (8.1)	137 (6.4)
	2AB	187 (7.5)	163 (4.7)	155 (5.5)	170 (9.4)	160 (8.1)	143 (5.4)	139 (6.1)	156 (8.6)	161 (4.9)	144 (6.7)	125 (6.1)	137 (5.0)
	2BA	185 (8.4)	165 (5.9)	156 (6.0)	168 (12.3)	165 (7.2)	147 (6.6)	142 (6.6)	157 (6.6)	152 (6.2)	137 (5.0)	123 (4.5)	136 (6.4)
TH	1AB	···	189 (11.4)	151 (8.2)	141 (6.5)	···	154 (6.4)	131 (2.6)	141 (8.2)	···	151 (4.7)	127 (4.0)	133 (3.8)
	1BA	···	187 (7.9)	149 (6.2)	144 (8.4)	···	159 (4.2)	134 (6.0)	136 (5.4)	···	145 (3.1)	126 (2.7)	135 (5.2)
	2AB	171 (7.9)	149 (7.2)	127 (2.3)	127 (3.5)	142 (4.4)	136 (4.9)	123 (4.0)	131 (4.9)	146 (4.8)	133 (3.7)	120 (4.6)	125 (5.5)
	2BA	179 (11.5)	150 (8.9)	127 (3.7)	128 (4.4)	147 (4.9)	138 (4.3)	123 (5.5)	127 (6.0)	138 (5.4)	132 (3.5)	118 (6.5)	124 (4.5)
CG	1AB	···	137 (9.5)	113 (9.3)	112 (7.6)	···	114 (5.2)	104 (5.8)	101 (3.9)	···	119 (5.9)	109 (8.5)	99 (5.3)
	1BA	···	140 (7.6)	109 (6.7)	114 (4.7)	···	118 (7.9)	109 (6.0)	100 (6.0)	···	112 (4.6)	105 (7.9)	96 (4.8)
	2AB	166 (9.8)	154 (6.9)	119 (7.9)	123 (4.4)	146 (6.5)	131 (7.8)	121 (7.2)	119 (3.1)	149 (7.4)	135 (8.1)	124 (5.7)	117 (4.3)
	2BA	168 (8.5)	155 (9.7)	121 (11.2)	124 (4.1)	146 (6.4)	137 (8.6)	120 (8.0)	124 (6.2)	141 (7.0)	131 (7.2)	120 (7.6)	115 (3.4)
GH	1AB	···	134 (3.7)	120 (5.0)	124 (2.9)	···	115 (3.2)	110 (3.5)	116 (4.4)	···	119 (3.8)	109 (2.6)	113 (4.3)
	1BA	···	135 (4.8)	118 (4.1)	122 (3.6)	···	124 (4.8)	112 (3.6)	118 (4.8)	···	114 (3.3)	109 (3.2)	112 (2.2)
	2AB	145 (5.1)	130 (5.5)	113 (2.8)	116 (2.8)	118 (3.1)	115 (4.2)	110 (3.4)	115 (3.4)	128 (5.3)	117 (1.9)	109 (2.4)	112 (3.3)
	2BA	146 (4.5)	132 (5.7)	114 (4.4)	116 (3.1)	131 (5.4)	117 (4.2)	111 (3.4)	114 (3.3)	117 (2.6)	112 (3.1)	109 (3.9)	111 (2.8)

APPENDIX B

TABLE BI. Mean clause end-point $F0$ values and standard deviations for experiments 1 and 2. DM's means are each based on 18 values; most of the other means are based on between 10 and 16 values. The experimental conditions (rows) are labeled, e.g., 1AB for "experiment 1, and/but condition."

Speaker		A	B	C
DM	1AB	126.4 (12.5)	116.4 (10.4)	101.1 (6.1)
	1BA	121.6(11.5)	119.8 (12.0)	97.4 (3.1)
	2AB	108.4 (5.3)	104.7 (4.2)	98.2 (4.1)
	2BA	106.4 (3.9)	108.9 (5.3)	98.0 (3.5)
TH	1AB	114.4 (4.7)	115.3 (5.9)	115.5 (4.3)
	1BA	114.7 (6.2)	118.3 (4.8)	117.3 (4.2)
	2AB	114.4 (4.9)	115.8 (3.8)	116.9 (3.1)
	2BA	112.1 (5.8)	113.7 (6.3)	113.9 (4.3)
CG	1AB	81.3 (2.0)	79.9 (2.6)	78.8 (4.0)
	1BA	81.6 (3.5)	80.5 (3.1)	80.0 (2.8)
	2AB	91.1 (2.7)	92.3 (2.9)	93.5 (2.8)
	2BA	91.5 (3.0)	92.3 (2.4)	92.2 (2.9)
GH	1AB	100.4 (4.4)	100.1 (3.5)	100.4 (2.3)
	1BA	98.6 (3.4)	97.8 (1.6)	98.0 (2.3)
	2AB	100.2 (5.0)	101.6 (2.8)	102.0 (2.6)
	2BA	101.5 (3.8)	100.6 (3.1)	99.4 (3.0)

[1] Inevitably, there are models that fall in the middle of this dichotomy, the most obvious being the work of 't Hart and his colleagues (especially 't Hart and Collier, 1975) and Pierrehumbert (especially Pierrehumbert, 1980). They argue strongly for describing contours as strings of linguistic elements, but nevertheless retain a "declination" component that has global effects over a phrase or utterance. This hybrid nature is presumably the basis of Thorsen's objection (1985, footnote 1) to my classification of 't Hart and Collier's model as essentially linear.

[2] The recordings were made in February and March 1984, when the American primary elections were much in the news and the sentences were thus vaguely topical. The speakers were told that these sentences were about an American election campaign.

[3] Cooper and Sorensen's usage of the term "fall–rise" is unlike the standard usage in linguistic descriptions of English intonation; in the latter usage the fall *and the rise* all precede the boundary pause and no reference at all is made to any following peak.

[4] The phrase no longer was sometimes accented on the first word and sometimes on the second.

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

Clements, G. N. (**1983**). "The hierarchical representation of tone features," in *Current Approaches to African Linguistics*, edited by I. Dihoff (Foris, Dordrecht), Vol. 1.

Cooper, W., and Paccia-Cooper, J. (**1980**). *Syntax and Speech* (Harvard U. P., Cambridge, MA).

Cooper, W., and Sorensen, J. (**1981**). *Fundamental Frequency in Sentence Production* (Springer, Heidelberg).

Crystal, D. (**1969**). *Prosodic Systems and Intonation in English* (Cambridge U. P., Cambridge, U.K.).

Fujisaki, H., and Sudo, H. (**1971**). "A generative model for the prosody of connected speech in Japanese," Annual Rep., Engineering Research Institute, University of Tokyo **30**, 75–80.

Gårding, E. (**1983**). "A generative model of intonation," in *Prosody: Models and Measurements*, edited by A. Cutler and D. R. Ladd (Springer, Heidelberg).

Halliday, M. A. K. (**1967**). *Intonation and Grammar in British English* (Mouton, The Hague).

't Hart, J. (**1979**). *Naar automatisch genereeren van toonhoogte-contouren voor tamelijk lange stukken spraak.* IPO Tech. Rep. No. 353, Eindhoven.

't Hart, J., and Collier, R. (**1975**). "Integrating different levels of intonation analysis," J. Phon. **3**, 235–255.

Huang, C.-T. J. (**1980**). "The metrical structure of terraced-level tones," in *NELS 11*, edited by J. Jensen (Cahiers Linguistiques d'Ottawa Vol. 9), Department of Linguistics, University of Ottawa, Ottawa, Canada, pp. 257–270.

Kubozono, H. (**1985**). "On the syntax and prosody of Japanese compounds," Work in Progess Vol. 18, Department of Linguistics, University of Edinburgh, pp. 60–87.

Kubozono, H. (**1988**). "The organization of Japanese prosody," Ph.D. thesis (University of Edinburgh, Edinburgh).

Kutik, E., Cooper, W. E., and Boyce, S. (**1983**). "Declination of fundamental frequency in speakers' production of parenthetical and main clauses," J. Acoust. Soc. Am. **73**, 1731–1738.

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

Ladd, D. R. (**1984**). "Declination: a review and some hypotheses," Phonology Yearbook **1**, 53–74.

Ladd, D. R. (**1987**). "A model of intonational phonology for use in speech synthesis by rule," in *Proceedings of the European Conference on Speech Technology*, Edinburgh, September 1987, edited by J. Laver and M. Jack (CEP Consultants, Edinburgh), Vol. 2, pp. 21–24.

Ladd, D. R. (**1988**). "Metrical representation of pitch register," to appear in *Papers in Laboratory Phonology I*, edited by J. Kingston and M. Beckman (Cambridge U.P., Cambridge, U.K.).

Ladd, D. R., and Johnson, C. (**1988**). " 'Metrical' factors in the scaling of sentence-initial accent peaks" (to appear in Phonetica).

Ladd, D. R., Silverman, K., Tolkmitt, F., Bergmann, G., and Scherer, K. R. (**1985**). "Evidence for the independent function of intonation contour, pitch range, and voice quality," J. Acoust. Soc. Am. **78**, 435–444.

Lehiste, I. (**1970**). *Suprasegmentals* (MIT, Cambridge, MA).

Lehiste, I. (**1972**). "The timing of utterances and linguistic boundaries," J. Acoust. Soc. Am. **51**, 2018–2024.

Liberman, M., and Pierrehumbert, J. (**1984**). "Intonational invariance under changes in pitch range and length," in *Language Sound Structure*, edited by M. Aronoff and R. Oerhle (MIT, Cambridge, MA), pp. 157–233.

Liberman, M., and Prince, A. (**1977**). "On stress and linguistic rhythm," Linguistic Inquiry **8**, 249–336.

Maeda, S. (**1976**). "A characterization of American English intonation,"Ph.D. dissertation (MIT, Cambridge, MA).

Menn, L., and Boyce, S. (**1982**). "Fundamental frequency and discourse structure," Lang. Speech **25**, 341–383.

O'Connor, J. D., and Arnold, G. F. (**1973**). *Intonation of Colloquial English* (Longman, London), 2nd ed.

Oller, D. K. (**1973**). "The effect of position in utterance on speech segment duration in English," J. Acoust. Soc. Am. **54**, 1235–1247.

O'Shaughnessy, D., and Allen, J. (**1983**). "Linguistic modality effects on fundamental frequency in speech," J. Acoust. Soc. Am. **74**, 1155–1171.

Pierrehumbert, J. (**1980**). "The phonology and phonetics of English intonation," Ph.D. dissertation (MIT, Cambridge, MA).

Scott, D. R. (**1982**). "Duration as a cue to the perception of phrase boundary," J. Acoust. Soc. Am. **71**, 996–1007.

Sternberg, S., Wright, C. E., Knoll, R. L., and Monsell, S. (**1980**). "Motor programs in rapid speech: additional evidence," in *Perception and Production of Fluent Speech*, edited by R. A. Cole (Lawrence Erlbaum Associates, Hillsdale, NJ), pp. 507–534.

Thorsen, N. (**1979**). "Interpreting raw fundamental frequency tracings of Danish," Phonetica **36**, 57–58.

Thorsen, N. (**1980**). "Intonation contours and stress group patterns in declarative sentences of varying length in ASC Danish," Annual Rep., Institute of Phonetics, University of Copenhagen **14**, 1–29.

Thorsen, N. (**1981**). "Intonation contours and stress group patterns in declarative sentences of varying length in ASC Danish—Supplementary data," Annual Rep., Institute of Phonetics, University of Copenhagen **15**, 13–47.

Thorsen, N. (**1985**). "Intonation and text in Standard Danish," J. Acoust. Soc. Am. **77**, 1205–1216.

Thorsen, N. (**1986**). "Sentence intonation in textual context—Supplementary data," J. Acoust. Soc. Am. **80**, 1041–1047.

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

Vaissière, J. (**1983**). "Language-independent prosodic features," in *Prosody: Models and Measurements*, edited by A. Cutler and D. R. Ladd (Springer, Heidelberg), pp. 53–66.