Phonology 7 (1990) 1–29
Printed in Great Britain

*Aspects of pitch realisation in Yoruba**

Bruce Connell
University of Oxford and University of Edinburgh
D. Robert Ladd
University of Edinburgh

---

0 Introduction

A great many languages of the world exhibit phenomena of FO DOWNTREND – phenomena whereby, other things being equal, the fundamental frequency (Fo) of the speaking voice declines over the course of an utterance. That much is uncontroversial; further details are either simply unknown or the subject of considerable debate. The purpose of the study reported here was to shed light on some of these unknown or uncertain matters by the controlled investigation of pitch realisation in Yoruba.

The main advantage of using a language with lexical tone as the basis of such a study is that it permits more control over the phonological influences on pitch than is the case in languages in which pitch conveys only intonational distinctions. The phonological analysis of intonation is still unclear on a number of fundamental issues, whereas the surface phonology of tone in Yoruba is fairly well established. We assume that conclusions based on our findings here will be useful for clarifying issues in intonational phonology as well.

The paper is in three parts. §1 discusses the background to the study and reviews some of the theoretical and empirical issues we wished to address. §2 reports on the procedures and results of the study itself. §3 summarises the theoretical implications of the study, particularly in the light of some unplanned findings.

1 Background

1.1 Downtrends

We use the term DOWNTREND, as in the introduction, for any overall lowering of speaking pitch during the course of an utterance. Many other terms are in use for phenomena of this general sort, their diversity often reflecting different analyses of the phenomena and/or different underlying theoretical assumptions (for some discussion see Ladd 1984). We briefly mention a few of these.

1.1.1 *Declination*. This is perhaps the most widely known term, at least in reference to European languages. We use it to refer to a gradual modification (over the course of a phrase or utterance) of the phonetic backdrop against which the phonologically specified local Fo targets are scaled – a tilting of the graph paper, to use Pierrehumbert's vivid metaphor (Pierrehumbert 1980: 63). This usage is close to the original intent of Cohen & 't Hart (1967), who, so far as we are aware, coined the term. However, more recent experimental work (beginning with Pierrehumbert 1980, followed by e.g. Poser 1984, Gussenhoven & Rietveld 1988, Pierrehumbert & Beckman 1988: ch. 3) strongly suggests that Cohen & 't Hart's original notion conflated phonetic effects with phonologically conditioned lowering phenomena such as downstep (cf. §1.1.5). In this more recent work DECLINATION tends to be reserved for the phonetic effects, and this is the usage we adopt here. We avoid the term DOWNDRIFT – another term that is also sometimes used to refer to phonetic rather than phonological lowering effects – because of its confusing range of uses in Africanist work (cf. §1.1.3).[1]

1.1.2 *Final lowering*. It is useful to distinguish declination – gradual lowering throughout a phrase or utterance – from final lowering, which, as its name suggests, is a more abrupt lowering confined to phrase and utterance ends. Such a phenomenon has been noted in many African languages. For example, Welmers (1973: 93) states that 'even where overall terracing is not present...a perceptible lowering in final position appears to be exceedingly common in discrete level systems'. Similarly, among the practical tips for fieldworkers in Pike (1948) – presumably based more on Mesoamerican than on African tone systems – we find the comment that 'pitches may be lowered non-phonemically at the ends of phrases' (1948: 57). In languages without lexical tone, final lowering is more difficult to identify, but recent work based on quantitative models of Fo suggests that it is present. For instance, Liberman & Pierrehumbert (1984) suggest that much of what has previously been ascribed to declination in English is actually the result of final lowering. They model final lowering as a substantial progressive lowering of overall pitch range during the last 250 ms or so of an utterance. The distinction between declination and final lowering in modelling downtrends is adopted by Pierrehumbert & Beckman (1988) for Japanese as well.[2]

1.1.3 *Downstep and downdrift*.[3] It has long been known in African linguistics that, in many languages, downtrends occur in tone sequences involving alternating high (H) and low (L) tones. This was formerly taken to be a predictable matter of phonetic realisation – in such languages any H preceded by L is realised at a lower pitch than an earlier H – and was commonly referred to as DOWNDRIFT. Downdrift in this sense was distinguished from DOWNSTEP, in which one H tone is realised at a lower pitch than a preceding H tone *without* any apparent conditioning factor, e.g. Efik *ọ́bọ́ng* 'mosquito', *ọ́bꜜọ́ng* 'chief' and *ọ́bọ̀ng* 'cane'.[4] More recently, however, 'downdrift' and 'downstep' in these restricted senses have generally been seen as manifestations of essentially the same phenomenon. This is reflected in e.g. Stewart's terms (1983: 70) 'automatic downstep' (for surface HLH sequences) and 'non-automatic downstep' (where no conditioning factor is present in the surface tonal string).

There are several reasons for treating these effects together. First, in some languages (e.g. Hausa: Lindau 1986) it has been shown that there are downtrends across non-alternating strings of tones (e.g. all-H), and that these downtrends are markedly less steep than those across alternating tone strings. The former (also sometimes confusingly called 'downdrift') would appear to represent declination of the phonetic backdrop in the sense discussed in §1.1.1, while the latter clearly involve phonological conditioning. Second, a link between 'automatic' and 'non-automatic' downstep has generally been recognised anyway, inasmuch as non-automatic downstep can often be traced diachronically or derivationally to automatic downstep, i.e. to the effect of a L tone that has been deleted. Whether this analysis can be extended to all cases of non-automatic downstep has been the subject of much discussion in autosegmental phonology; see e.g. Stewart (1983) and Clements & Ford (1979). Third, and perhaps most important, both involve a *localisable* phonological phenomenon with *global* phonetic effects: the place in the string where downstep occurs – the lowered H tone – may be precisely identified, but its effects persist, in the sense that subsequent H tones (within the phonological phrase) tend not to rise above the level of the lowered H. It is as if the lowering of H in the HLH sequence reflects a lowering of the 'ceiling' on tonal realisation, not merely a lowering of the second H alone.

1.1.4 *Tone terracing*. The persistence of the phonetic effect of downstep inevitably raises the issue of tone terracing. In Africanist work a dichotomy is often drawn between DISCRETE-LEVEL and TERRACED-LEVEL tonal realisation (or simply discrete-level and terraced-level languages); the terminology, if not the underlying assumption that a dichotomy is involved, is apparently due to Welmers (1959). In a discrete-level language the realisations of the tone phonemes are not supposed to trespass on each other's phonetic space, but are realised in discrete

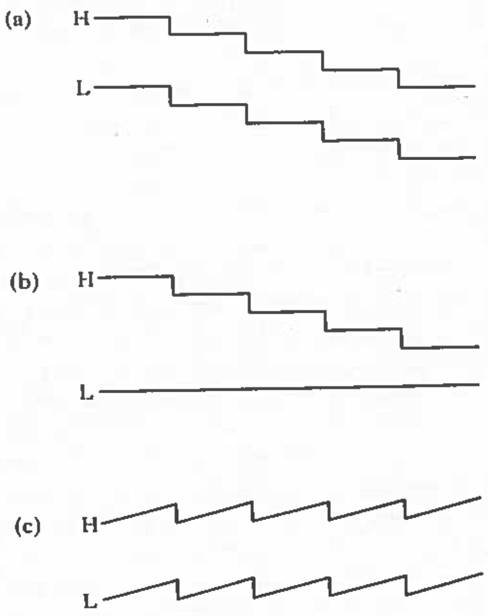

Figure 1

Terraced-level tonal realisation. (a) represents a case in which both H and L tones are affected by downstep. (b) shows only H tones affected by downstep, while (c) shows the local effect of downstep being gradually reversed, resulting in a 'hogback ridge' pattern. Either (b) or (c) would make it possible to have downstep in a 'discrete-level' system in which the phonetic spaces allotted to the different tones do not overlap.

frequency bands that remain more or less fixed throughout the utterance. In a terracing language, on the other hand, the repeated application of downstep (or, in some cases, upstep) produces frequency bands with the form of a sequence of steps or terraces, with the result that, at the end of an utterance with several downsteps, the realisation of H tone will often be phonetically lower than the utterance-initial realisation of L. This is illustrated in Fig. 1a.

However, the assumption that there is a dichotomy between these two types seems both empirically doubtful and logically unnecessary (cf. Stewart 1965, cited in Welmers 1973: 86). There is no necessary connection between terracing and the phonetic overlap of tonal realisations. One obvious way in which terracing might not create phonetic overlap would be if only H tones were subject to downstep, giving a pattern like Fig. 1b. As it happens, the effect of downstep on low tones is actually a matter of considerable uncertainty (cf. Hombert 1974; Welmers 1973: 88), which suggests that patterns of this sort may occur. Another way in which a kind of terracing might coexist with discrete tonal realisations would be for the overall lowering brought about by a downstep to reverse itself gradually, yielding not a sequence of terraces but rather (to continue the topographical metaphor) a set of 'hogback ridges' (Fig. 1c). This apparently is what Welmers (1973: 91) describes for Xhosa; there also seems to be some evidence for it in our own data (cf. §3.2.1).

1.1.5 *Phonologically conditioned downtrends in languages without lexical tone.* In the last several years the notion of downstep has been extended beyond African linguistics to apply to other languages. This extension was first proposed by Pierrehumbert (1980) for the description of English. Pierrehumbert's proposal has been widely adopted in descriptions of European intonation systems (e.g. Ladd 1983, 1990; Gussenhoven 1984: 313; van den Berg *et al.* forthcoming), although the nature of the phonological 'trigger' for European downstep remains a matter of debate (cf. Ladd 1989). The existence of downstep in Japanese has been convincingly demonstrated by Poser (1984), a finding replicated and developed in the work of Kubozono (1988, 1989) and Pierrehumbert & Beckman (1988); in Japanese, as in many African languages, downstep is triggered by a HL sequence (which occurs at accented syllables) and affects a following H.[5]

It is primarily because of the recognition that downstep occurs in many widely differing languages that a study of Yoruba downtrends may be said to have wider implications. We return to these implications in §3 of the paper.

1.2 Yoruba

1.2.1 *Past work.* Yoruba is one of the major languages of West Africa, spoken by approximately 15 million people primarily in southwestern Nigeria and southeastern Benin. A standard description of the language is Bamgbose (1966); Welmers' book *African language structures* (1973) uses Yoruba illustrations extensively; recent theoretical work on Yoruba phonology includes Pulleyblank (1986) and Archangeli & Pulleyblank (1989). An exhaustive bibliography of works on Yoruba is provided by Adéwọlé (1987). To our knowledge the only other systematic instrumental studies of tone realisation in Yoruba are: (i) LaVelle (1974), which studied mostly isolated words, and concentrated on the perception of tone in stimuli with synthetic pitch contours; (ii) Hombert (1978), which is more concerned with the interaction of tone and segmental phenomena (but see also Hombert's references to his other work); (iii) Laniran (1988, 1989, in preparation), which parallels our study in many ways and which produces similar findings on certain key points. We did not become aware of Laniran's work until after our own study was essentially complete.

1.2.2 *Surface tonology.* In the surface phonology of Yoruba there are three contrasting tonemes, High (H), Mid (M) and Low (L), and in most cases each syllable is associated with one tone and *vice versa*.[6] Phonetically, H is realised as Rise after L, and L is realised as Fall after H; these effects can be chained, so that HLH is realised as H–Fall–Rise. In addition, there are widespread phenomena of elision and contraction which sometimes bring about the association of two tones with one vowel and sometimes create floating tones that are deleted postlexically (ꜜ indicates downstep):[7]

(1) a. M M H L → M M H L
| | | | → | |/ |
ẹ ja o pọ → ẹ ja pọ 'much fish'

b. H L H → HꜜH
| | | → | |
ni oni → loni 'today'

c. M L H → L H
| | | → | |
kọ iwe → kọwe 'write a book'

Together with downstep (discussed in the next paragraph), these are the only reported tone sandhi phenomena. However, the degree of confusion in works intended for reference or pedagogy (e.g. Abraham 1962; Rowlands 1969) makes it clear that many aspects of tonal realisation are not understood, and it seems likely that other phenomena remain to be discovered.

It is generally said (e.g. Courtenay 1971; Welmers 1973; Hombert 1974) that Yoruba has downstep – both 'automatic' and 'non-automatic'. Most descriptions state that H and M may be downstepped after L in sequences like HLH or MLM; in such sequences the L may be either present on the surface, or left floating through vowel elision and deleted postlexically. The following examples are taken from Rowlands (1969: 15, 29):

(2) a. H L H — M L M
| | | — | | |
ma je le 'poison' — a di ẹ 'chicken'
(realised phonetically with the last tone lower than the first)

b. H L H → HꜜH
| | | → | |
ni o ni → loni 'today' (= 1b)

c. H L M → HꜜM
| | | → | |
ni ode → lode 'outside'

However, descriptions with which we are familiar do not go beyond citing individual lexical items or short phrases, and it is unclear whether downstep may be iterated to create tone terracing. Clements (1979: 537) includes Yoruba in a list of terracing languages, whereas Welmers (1973: 109) classifies it as a discrete-level language. Furthermore, it is unclear whether downstepping affects L tones in any way. To some extent, these unclear points must be related to the theoretical shortcomings of the supposed dichotomy between terraced-level and discrete-level systems, but they also – particularly with regard to the behaviour of L – reflect the difficulties of establishing tonal data on the basis of auditory impressions in the field.

2 The instrumental study

2.1 Method

2.1.1 *General.* In order to investigate the issues discussed above, we posed the following experimental questions:

- (i) Do declination and/or final lowering occur in Yoruba? If declination occurs, is it affected by sentence length?
- (ii) Does downstep occur in Yoruba? If so:
  - a. What triggers it? Does it require a sequence HLH, or will the sequence LH alone lower the H relative to what it would be in an otherwise comparable HH sequence?
  - b. Does it induce terracing, and if so, what happens to the levels of all three tonemes in terracing?
- (iii) Is there any overall effect on pitch range in questions as opposed to statements?

With the help of a Yoruba consultant, we designed an experimental corpus of 108 sentences (described in more detail in the next section) and recorded the sentences as spoken under laboratory conditions by four educated native speakers of Yoruba, including the consultant (here designated as Speaker F). The speakers were three men (B, F and O) and one woman (A), all in their twenties or thirties, who had been resident in Edinburgh for periods ranging from a few months to a few years; except for Speaker F, who is a linguist, all were as linguistically naive as well-travelled, well-educated speakers of any language may be said to be.

We extracted Fo from the recorded utterances and identified the Fo levels at several key syllables or other target points in each utterance. The data are thus reduced to sets of mean Fo values for e.g. 'penultimate syllable in an utterance consisting of all H tones', 'L tone valley in the sequence HLH', etc. This appoach to data reduction has been taken in a large number of recent studies of Fo, e.g. Poser (1984), Kubozono (1989). More detail on the experimental procedures is given in the Appendix.

The exploratory nature of the experimental questions and the procedures will be clear. Wherever possible we formulated our questions in a way that allowed for the statistical evaluation of planned comparisons, but in some respects our goal was simply to assemble a corpus of carefully controlled data that might be used for subsequent formulation of more rigorously testable hypotheses.

2.1.2 *The corpus.* The experimental corpus consists of 9 datasets of 12 similar sentences, with each dataset representing a particular condition we wanted to investigate. The data points in our results are means based on measurements taken from all 12 sentences in a given set. The structure of the datasets is shown, with examples, in Table I:

Like-tone sentences		
1. All H: H H...H		Wọn tún gbé túwó wá 'They brought tuwo again'
2. All M: M M...M		Omo wọn ni e lo fi se oko 'It is their son that you marry'
3. All L: L L...L		Èwù ònà Àrà ò tán 'The colour of the garments on the way to Ara is dull'
Mixed-tone sentences		
Statements		
4. H H T...T	a.	Dúró|ò wọ ṣòkòtò 'Duro did not wear any trousers'
	b.	Wọn fẹ|jàmọlà tán 'They wanted to eat up the amola'
5. L H T...T	a.	Àpọ́n|ò wọ ṣòkòtò 'The bachelor did not wear any trousers'
	b.	Ọ̀báyéjẹ́|jàmọlà tán 'Tale-bearer ate up the amola'
ṣé-*questions**		
6. ṣé H H T...T	a.	Ṣé Dúró|ò wọ ṣòkòtò 'Did Duro not wear any trousers?'
	b.	Ṣé won fẹ́|jàmọlà tán 'Did they want to eat up the amola?'
7. ṣé L H T...T	a.	Ṣé àpọ́n|ò wọ ṣòkòtò 'Did the bachelor not wear any trousers?'
	b.	Ṣé Ọ̀báyejẹ|jàmọlà tán 'Did Tale-bearer eat up the amola?'
ǹjẹ́-*questions**		
8. ǹjẹ́ H H T...T	a.	Ǹjẹ́ Dúró|ò wọ ṣòkòtò (= 6a)
	b.	Ǹjẹ́ wọ́n fẹ́|jàmọlà tán (= 6b)
9. ǹjẹ́ L H T...T	a.	Ǹjẹ́ àpọ̀n|ò wọ ṣòkòtò (= 7a)
	b.	Ǹjẹ́ Ọ̀báyejẹ|jàmọlà tán (= 7b)

* formed on Statements in sets 4 and 5.

[Table I. *Structure of datasets (left) with examples (right). In the examples of datasets 4–9, the part to the right of the vertical line is the 'shared text'*]

2.1.2.1 *Like-tone sentences (datasets 1–3).* The first three datasets are LIKE-TONE sentences – 12 sentences of all H tones, 12 all M, and 12 all L – designed to look for declination and final lowering (experimental question 1). The assumption here is that e.g. the H tone at the end of an all-H sentence is phonologically the same as the H tone at the beginning, so that we might attribute any difference in Fo from beginning to end to one or the other (or both) of these phonetic effects. Some studies of declination, e.g. 't Hart (1979), suggest that pitch should be lowered further in longer sentences than in shorter ones, while others suggest no such link or are at best unclear. In order to study any possible effect of sentence length, these three datasets include sentences that are 4, 6, 8 and 10 syllables long.

2.1.2.2 *Mixed-tone sentences (datasets 4–9).* The remaining six datasets are MIXED-TONE sentences. Each sentence in a given dataset has a counterpart in the other five sets, all sharing the same second half of the sentence (usually the verb phrase) and differing only at the beginning (see Table I). In what follows we refer to the identical second half of these corresponding sentences as the 'shared text'. The mixed-tone datasets were intended to shed light on experimental questions 2 and 3 by means of several planned comparisons of mean pitch levels in the shared texts of different datasets.

The issues grouped together under experimental question 2 are in some sense the heart of the study. The beginnings of the sentences in the mixed-tone datasets differ in ways thought likely to have different effects on the realisation of subsequent tones. For example, the sentences in dataset 6 begin with the sequence HHH, while those in dataset 7 have initial HLH. If the HLH sequence triggers downstep, we would expect to find that mean values of tones are lower in the shared texts of dataset 7 than in dataset 6.[8] Several such comparisons were planned.

Finally, studies of many languages have suggested overall higher pitch range in questions, but generally without controlling for possible phonological effects on average Fo, e.g. final rise *vs.* final fall. The use of materials from a lexical tone language makes it easier to control for possible phonological effects; this is what we investigated under experimental question 3. As can be seen in Table I, datasets 4 and 5 are statements, while datasets 6–9 are questions constructed from those statements by the addition of one of two sentence-initial question particles *ṣé* and *ǹjẹ́*.[9] Combinations of these datasets were used for further planned comparisons.

2.1.2.2.1 *Planned comparisons.* The planned comparisons based on the mixed-tone sentences were as follows:

(1) *Datasets 4 and 5:* Comparison of these two sets was intended to show whether there was any effect of the initial LH sequence on the shared text – specifically, to see whether the occurrence of the sequence LH at the beginning of the sentences in dataset 6 caused the shared text to be lower than in the dataset 4 sentences beginning with the sequence HH. If the initial LH sequence triggers downstep, then the mean values for the tones in the shared text should be lower in dataset 5 (following LH) than in dataset 4 (following HH).

(2) *Datasets 6 and 8:* This comparison replicates the preceding one, but using questions rather than statements.

(3) *Datasets 6 and 7:* This comparison is similar to the previous two, except that here the issue is the effect of the initial sequence HLH (dataset 7) compared to HHH (dataset 6). Again, the assumption was that if the initial HLH sequence triggers downstep, then the mean values for the tones in the shared text should be lower in dataset 7 (following HLH) than in dataset 6 (following HHH).

(4) *Datasets 8 and 9:* This comparison is similar to the preceding one, except that the initial sequences are LHHH and LHLH. If initial LH alone triggers downstep (cf. comparisons 1 and 2), then this comparison should reveal the effect of repeated downstep, i.e. the shared text means in dataset 9 should be lower than in 8, reflecting the application of a second downstep. However, if initial LH is shown not to trigger downstep, then this comparison simply replicates comparison 3.

(5) *Datasets 4 and 6:* This comparison is intended to show whether there is any effect on overall pitch in questions. Specifically, if the overall pitch in questions is higher, then the means for shared text tones should be higher in set 6 than in set 4. Note that the initial sequence in both sets is HH(H).

(6) *Datasets 5 and 8:* This comparison replicates the preceding one, except that here the initial sequence in both cases is LH(H). Regardless of whether initial LH triggers downstep, this comparison should reveal any effect of overall raising in questions.

2.1.2.3 *Summary*. Datasets 1, 2 and 3 were intended to provide data relevant to experimental question (i) (presence of backdrop declination and/or final lowering). Comparisons between datasets 4 and 5, datasets 6 and 7, datasets 8 and 9 and datasets 6 and 8 were intended to permit conclusions regarding experimental question (ii) (nature of downstep trigger and effect of repeated downstep on terracing). Comparisons of datasets 4 with 6 and 5 with 8 were intended to provide data relevant to experimental question (iii) (presence of overall raising of range in questions).

2.2 Results and discussion

The results are presented in two sections. The first deals with the investigation of declination and final lowering in the like-tone sentences in datasets 1–3. The second presents the outcome of the planned comparisons based on the mixed-tone sentences in datasets 4–9. Discussion of some further observations that emerged from the data, having to do with the question of what happens to L tones in downstep and more generally with the paradox of apparent tone terracing in a 'discrete-level' language, is postponed till §3 of the paper.

2.2.1 *Like-tone sentences (datasets 1–3)*. All four speakers show very similar patterns for the like-tone sentences, which are illustrated in Figs. 2 and 3. In the all-H and all-M sentences, the pitch declines very little, while in the all-L sentences it declines rather more, especially in the final few syllables. Moreover, the pitch of the last syllable of the all-H and all-M sentences is quite steady, in many cases actually rising abruptly (but, notice, imperceptibly) at the very end, whereas in the all-L sentences, the pitch of the final syllable falls steadily and audibly throughout the syllable to a value that for any given speaker is quite consistent from utterance to utterance.[10] Because of these differences, the all-H and all-M sentences are discussed separately from the all-L sentences in the following paragraphs.

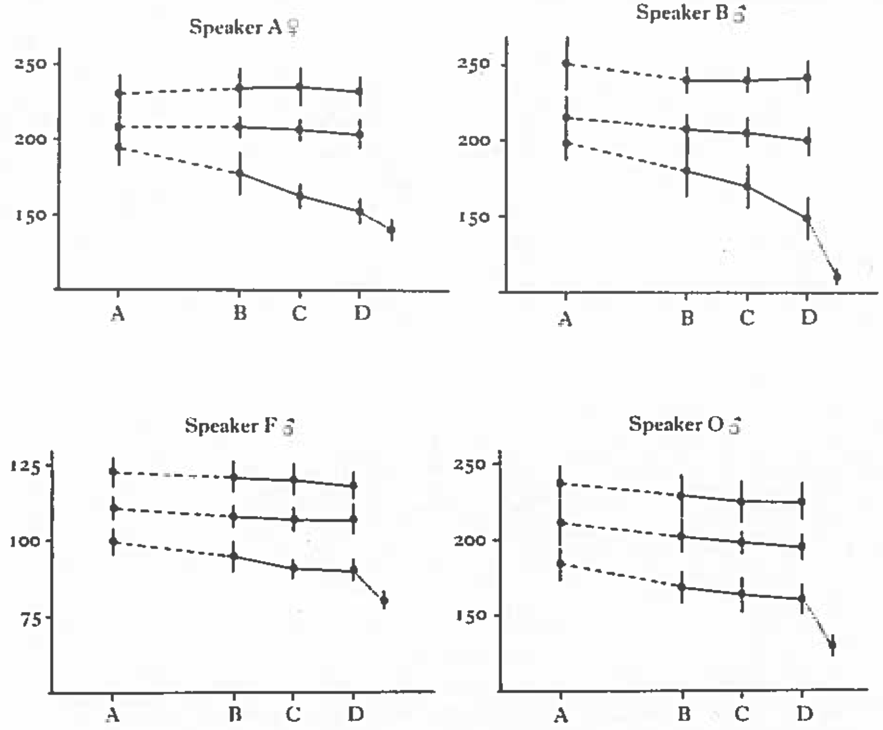

Figure 2

Mean values for like-tone sentences. A = initial value after start-up (peak reached sometimes as late as second syllable); B = antepenultimate; C = penultimate; D = final syllable. The dashed line between A and B represents a varying number of syllables, depending on sentence length. The dotted line at the end of the graphs for all-L sentences represents the drop in Fo through the final syllable. The vertical bars represent one standard deviation.

2.2.1.1 *All-H and all-M (datasets 1 and 2)*. Inspection of Fig. 2 shows a slight declination in the all-H and all-M sentences for all the speakers except speaker A, who has a slight rise across the all-H sentences. Omitting Speaker A, the amount of declination averages exactly 1 semitone. Assessing the statistical significance of this declination is difficult because it is so small that pooling the data across speakers with different pitch ranges will obscure the differences from beginning to end

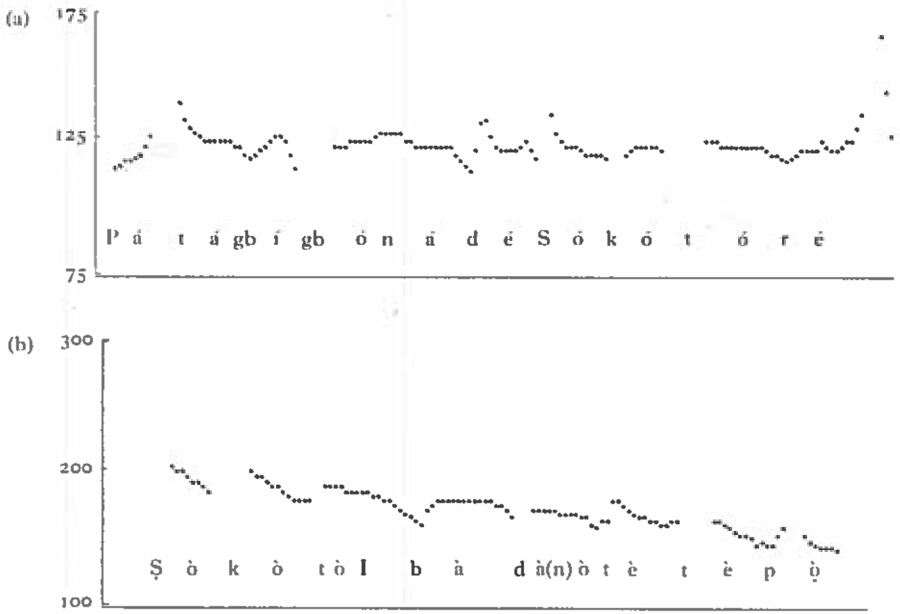

Figure 3

Sample traces for Like-tone sentences. (a) shows all-H (Speaker F), (b) shows all-L (Speaker A). The transcription underneath the pitch trace shows the approximate alignment of the pitch contour with the segmentals. The transcription is given in tone-marked Yoruba orthography, which is more or less surface phonemic, except that the letter *n* after a vowel letter indicates nasalisation of the vowel and is given here in parentheses.

of utterance. In any case the effect is too small to determine whether it is influenced by sentence length.

Even assuming that the declination itself is a real effect, it nevertheless seems fair to conclude from these results that backdrop declination does not exist in Yoruba to the degree documented by Lindau (1986) for Hausa, by Hombert (1974) for Shona, and by Poser (1984), Kubozono (1988) and Pierrehumbert & Beckman (1988) for Japanese. Our results are, however, comparable to those obtained for all-H sentences in Bambara by Mountford (1983). This suggests that backdrop declination is a matter of language-specific phonetic realisation rules.[11] This is consistent with current models that distinguish declination from phonologically conditioned pitch lowering, and with current ideas about the language-specific nature of 'phonetic realisation rules' (cf. Pierrehumbert & Beckman 1988): no theoretical considerations would prevent a model of pitch realisation in Yoruba from specifying a rate of declination different from the one used in, say, Pierrehumbert & Beckman's model of Japanese.

2.2.1.2 *All-L* (*dataset 3*). The all-L results are a little more difficult to interpret. For Speaker O, the general pattern of declination seems the same in all 3 tones, except of course for the abrupt fall during the last syllable in the all-L sentences. For the others, however, the rate of declination seems to increase during the last two or three syllables; and even for Speaker O, it is worth noting that the amount of declination is greater in the all-L sentences than in the other two types (1 semitone in all-H, 1·5 semitones in all-M, and 2·5 semitones in all-L, *excluding* the final-syllable fall). One plausible analysis of these results would be to distinguish final lowering from backdrop declination – i.e. to say that final lowering, spanning up to two or three syllables, occurs in all-L sentences, and that it is superimposed on the small amount of backdrop declination that characterises all the sentences. However, one might equally argue that the all-L sentences, unlike the others, show a gradual decline that approximates the degree of declination found in Hausa, Japanese, etc., and that only the one phenomenon of backdrop declination – albeit at different rates – is necessary to account for the three different datasets. Final lowering, under this interpretation, would be restricted to the rapid 'baseline fall' in utterance-final L-tone syllables.

This latter explanation leaves us with the problem of explaining why declination should differ for different tones, and makes it difficult to generalise from the like-tone sequences investigated here to utterances with a mix of different tones. Yet it is more consistent with the notion of final lowering outlined in §1.1.2 – i.e. a phonetically governed lowering restricted to the last quarter-second or so of the utterance. Moreover, it is in some sense consistent with what is known about Yoruba tones in isolation. LaVelle (1974) found that Yoruba L tones in isolation are sharply falling, and that the distinction between falling and sustained pitch is an important perceptual cue to the phonological distinction between L and M tones. This could of course be analysed as final lowering (since isolated tones are final), but it may also be that even non-final L tones exert some sort of local lowering effect – greater in some speakers than in others – which would contribute to the steeper decline in all-L than in all-M and all-H sentences. That is, it may be that the notion 'rate of declination' is simply a statistical reduction of a cluster of phonetic effects, some of which are indeed global (e.g. decline in subglottal pressure), but others of which are actually the cumulative result of local lowerings (e.g. for Yoruba L-tone) that are not completely reversed on subsequent syllables. We return to this question briefly at the end of the paper.

2.2.2 *Mixed-tone sentences* (*datasets 4–9*). The following paragraphs discuss the results of the planned comparisons based on experimental questions 2 and 3. What is compared here are the mean values of the H tones and the M tones in the shared text of one dataset with the mean shared-text H and M tone values from another dataset. We had of course intended to compare L tones as well, but for reasons outlined in the Appendix this was unfortunately not possible.

		Comparison 1				*Comparison 2*			
Speaker	*Tone*	Dataset 4	Dataset 5	*t*	(df)	Dataset 6	Dataset 8	*t*	(df)
A	H	237 (9·3)	238 (15·0)	0·16	(6)	238 (4·6)	241 (8·7)	0·86	(6)
	M	208 (7·4)	206 (6·5)	0·50	(6)	203 (5·6)	205 (6·5)	0·52	(6)
B	H	249 (12·1)	233 (16·5)	2·35	(6) *a*	261 (11·4)	256 (16·5)	0·75	(8)
	M	217 (14·8)	204 (13.6)	1·71	(8)	222 (16·7)	229 (7·6)	1·27	(12)
F	H	123 (9·2)	121 (6·5)	0·62	(10)	126 (9·1)	127 (8·7)	0·28	(10)
	M	113 (6·7)	111 (3·4)	0·96	(16)	112 (5·5)	114 (3·6)	1·03	(15)
O	H	226 (11·3)	231 (14·5)	0·94	(12)	231 (8·7)	234 (14·8)	0·61	(12)
	M	192 (10·3)	198 (9·6)	1·35	(14)	208 (9·0)	203 (12·0)	1·02	(13)

[Table II. *Effect of initial LH sequence (datasets 5 and 8) on tonal realisation in subsequent shared text, compared to cases with initial HH sequence (datasets 4 and 6). In this and subsequent tables, the columns headed 'Dataset' show mean Fo value in Hz, with standard deviations in parentheses. Significance of the* t-test *is indicated by the conventional star notation* (*** $p < {\cdot}001$; ** $p < {\cdot}01$; * $p < {\cdot}05$); *a indicates that the difference approaches significance* (${\cdot}05 < p < {\cdot}10$)]

2.2.2.1 *Comparisons 1 and 2.* These compared dataset 4 with dataset 5 and 6 with 8, in order to see whether an initial LH sequence triggers downstep. We assumed that if LH alone (as distinct from HLH) triggers downstep, the shared text means would be lower in datasets 5 and 8 (which begin with an initial LH sequence) than in 4 and 6 respectively (which begin with initial HH). The results of the two comparisons, given in Table II, seem quite clearly negative: only in one case (Comparison 1, Speaker B) does the expected lowering approach significance,[12] and the other speakers show conflicting tendencies. Thus there is no evidence that the sequence LH alone triggers downstep. It is worth noting that Laniran (1988) performed a similar experiment with a similar expectation to ours, and similarly found no evidence for downstep triggered by initial LH. In fact, for the one speaker analysed in a preliminary paper, she found a raising rather than a lowering effect of initial L on following H (i.e. comparable to Speaker O's comparison 1 results in the present data).

2.2.2.2 *Comparisons 3 and 4.* These compared dataset 6 with 7, and 8 with 9, investigating the effect of an HLH sequence compared to that of an HHH sequence. (In datasets 7 and 6 these were the sentence-initial sequences; in 9 and 8 the full initial sequences were LHLH and LHHH.) Results are shown in Table III. Here the pattern is unmistakable: the sequence HLH, whether sentence initial or preceded by L, triggers downstep. The effect on the shared-text means is large and significant in many of the individual comparisons, and none of the comparisons that fail to reach significance contradicts the conclusion.

Taken together with the findings of Comparisons 1 and 2, this suggests that downstep is somehow crucially associated with the sequence HL, rather than the sequence LH. The theoretical implications of this are explored further in §3.

		Comparison 3				*Comparison 4*			
Speaker	*Tone*	Dataset 6	Dataset 7	*t*	(df)	Dataset 8	Dataset 9	*t*	(df)
A	H	238 (4·6)	216 (7·7)	6·94	(6) ***	241 (8·7)	222 (15·0)	3·10	(6) *
	M	203 (5·6)	196 (7·2)	1·88	(6)	205 (6·5)	197 (4·8)	2·25	(6) *a*
B	H	259 (11·0)	236 (17·6)	3·84	(10) **				
	M	227 (11·3)	212 (12·8)	2·78	(12) *				
F	H	126 (9·1)	121 (9·9)	1·29	(10)	127 (8·7)	123 (4·7)	1·40	(10)
	M	112 (5·5)	107 (3·0)	2·79	(16) *	114 (3·6)	110 (5·4)	2·09	(15) *a*
O	H	231 (8·7)	222 (13·6)	1·91	(11) *a*	234 (14·8)	217 (11·9)	3·02	(11) *
	M	208 (9·0)	193 (10·6)	3·30	(13) **	203 (12·0)	185 (10·9)	3·51	(14) **

[Table III. *Effect of HLH sequence on tonal realisation in subsequent shared text (datasets 7 and 9), compared to cases with HHH sequence (datasets 6 and 8). See further the caption to Table II.*

Comparison 4 was not carried out for Speaker B because a large number of his utterances in these datasets, especially dataset 8, had to be eliminated from the analysis (see Appendix); the comparison would therefore not have been based on comparable means]

		Comparison 5				*Comparison 6*			
Speaker	*Tone*	Dataset 4	Dataset 6	*t*	(df)	Dataset 5	Dataset 8	*t*	(df)
A	H	237 (9·3)	238 (4·6)	0·27	(6)	238 (15·0)	241 (8·7)	0·49	(6)
	M	208 (7·4)	203 (5·6)	1·32	(6)	206 (6·5)	205 (6·5)	0·24	(6)
B	H	249 (12·1)	262 (11·2)	2·37	(6) *a*				
	M	217 (14·8)	230 (9·8)	1·94	(8) *a*				
F	H	123 (9·2)	126 (9·1)	0·80	(10)	121 (6·5)	127 (8·7)	1·91	(10) *a*
	M	113 (6·1)	112 (5·5)	0·44	(16)	111 (3·4)	114 (3·6)	2·10	(15) *a*
O	H	226 (11·3)	231 (8·7)	1·21	(12)	231 (14·5)	234 (14·8)	0·50	(12)
	M	192 (10·3)	208 (9·0)	3·59	(13) *a*	198 (9·6)	203 (12·0)	1·03	(14)

[Table IV. *Comparison of tonal realisation in questions (datasets 6 and 8) with corresponding statements (datasets 4 and 5). See further the captions to Tables II and III*]

2.2.2.3 *Comparisons 5 and 6.* These compared dataset 4 with 6, and 5 with 8, in order to see whether questions bring about an overall raising of pitch. Results are given in Table IV. The male speakers show some tendency to have higher means in questions than in the corresponding statements; the differences are not very large, but in several instances they are significant on a one-tailed test (i.e. on the hypothesis that questions should be higher). However, the female speaker (Speaker A) shows no consistent or significant effect on either comparison.

If this result can be generalised to other languages, it suggests that previous studies showing higher overall pitch in questions may have confounded a small amount of global raising of pitch range with larger local phonological effects such as fall *vs.* rise. At the very least, our finding should make clear that controls for phonological differences, which have often been lax or non-existent in experimental studies of Fo, cannot be neglected if the results are to be used with confidence.

3 What happens in downstep?

3.1 Register

One of the important theoretical issues in much recent work on downstep has been the question of how to model the persistence of downstep's phonetic effects through subsequent tones. Broadly speaking, we may distinguish two approaches to this problem:

(i) The phonetic realisation of tones may be modelled in such a way that a persistent lowering results automatically from the repeated application of a purely local rule. For example, each tone might be defined phonologically with respect to the preceding tone in the string, and any modification of one tone through a local rule of downstep would therefore necessarily be passed on to subsequent tones.

(ii) The phonetic realisation of tones may be modelled in terms of an explicit independent construct of REGISTER, a phonetic frame of reference relative to which tones are defined. In this case downstep rules apply to the register and not directly to the first affected tone, and the persistence of the lowering effect is a consequence of the lowering of the register.

Within African linguistics this issue was raised in something like these terms by Clements (1979), who argued that register – his 'tone level frame' – must be treated as an explicit element of any description of pitch realisation. (Clements was arguing specifically against an earlier attempt by Peters 1973 to formalise the first type of model.) In the recent expansion of the notion of downstep to languages outside Africa, Pierrehumbert's original model of English downstep espoused the first approach, but subsequent work (e.g. Poser 1984; Pierrehumbert & Beckman 1988; Ladd 1990; Clements 1990) has by and large accepted the need for some sort of explicit definition of register.

This then raises the further issue of what exactly happens to the register when downstep occurs. For Clements and for Ladd, downstep is in the first instance a *lowering* of the overall register within the speaking range; whereas for Pierrehumbert & Beckman it is a *narrowing* of the register, with the neutral midpoint of the register ('reference line') remaining constant and only the high-tone line stepping down. This issue is obviously directly related to the questions of tone terracing raised in § 1 of the paper: in effect, these two models correspond to the situations depicted in Fig. 1a (Clements, Ladd) and Fig. 1b (Pierrehumbert & Beckman). Evidence that downstep affects L tones in the same way as H and M tones would thus constitute support for the Clements/Ladd register-lowering point of view; evidence that L tone was unaffected by downstep would favour the Pierrehumbert & Beckman register-narrowing model.[13]

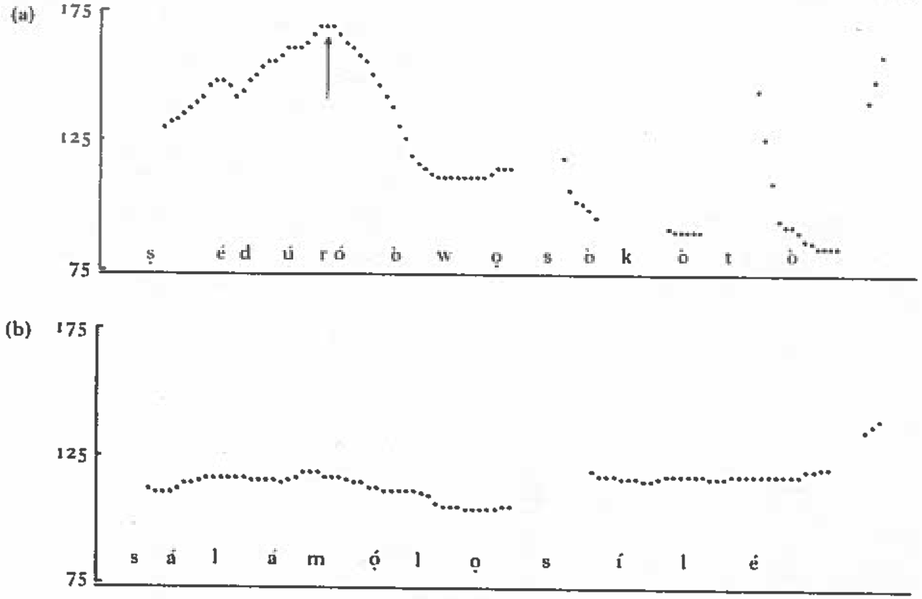

Figure 4

Local effects of tones. (a) shows raising of H (arrow) before L. (b) shows that the same effect is not present with H before M. (Both Speaker F.)

3.2 Some supplementary findings

Obviously, the fact that we were unable to get reliable data on L tones for the planned comparisons makes it difficult to say anything conclusive about the question of what happens to Ls in downstep, and more generally about the terracing patterns in Yoruba. This is not to say, however, that our data are neutral on this question. This section of the paper reports two further findings relevant to the workings of downstep in Yoruba, and, as we suggest below, to the general question of how register changes are implemented.

3.2.1 *Raising of H before L.* As we pointed out in § 1.2.2, it is difficult to reconcile the occurrence of downstep in Yoruba – with its supposedly persistent phonetic effects that should lead to tone terracing – with the language's putative classification as 'discrete-level'. However, our data provide clear evidence for a *local* effect related to downstep that may help resolve this paradox, namely the *local raising of H before L*. This can be seen, first of all, in the pitch trace for the sentence *Ṣé Dúró ò wọ ṣòkòtò* ('Did Duro not wear trousers?') in Fig. 4a. This begins with a sequence of three H-toned syllables before the L on the negative particle (*k*)*ò*, and the H immediately preceding the L is substantially higher than the first two Hs. For comparison, Fig. 4b shows the pitch trace for the utterance

	Scaling of H in question particle			
Speaker	*Before H* (Datasets 6 and 8)	*Before L* (Datasets 7 and 9)	*t*	(df)
A	239 (9·3)	261 (8·8)	6·79	(30) ***
B	276 (20·9)	304 (31·4)	3·25	(35) **
F	129 (10·8)	161 (8·3)	11·19	(44) ***
O	250 (13·6)	286 (15·3)	7·64	(39) ***

[Table V. *Local raising of H before L. See further the caption to Table II*]

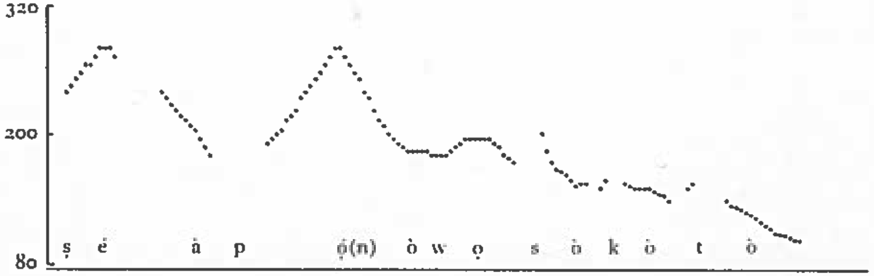

Figure 5

'Hogback ridge' effect. H before L is raised to the level of earlier H, cancelling any terracing effect due to downstep (Speaker B).

Sálámọ́ lọ sílé ('The red ant went home'), in which an initial sequence of three high tones is followed by a M tone: no such raising of the third H occurs.

Less anecdotally, we can demonstrate the raising of H before L by comparing the level of the H on the question particle (*ṣé* or *ǹjẹ́*) before HH (datasets 6 and 8) and before LH (datasets 7 and 9). The data for the four speakers are summarised in Table V. They show very clearly that for all speakers the H on either particle is substantially and significantly higher before L than before H.

Local raising of H before L, which is also reported by Laniran (1988), goes a considerable way toward explaining the apparent contradiction between the existence of downstep and the classification of Yoruba as a discrete-level (non-terracing) language. If H is raised before L, then in a local sequence HLH there would be a conspicuous difference between the first H and the second. Actual register lowering – i.e. persistent effects of downstep that would lead to terracing – might be slight or even non-existent, but field observers would note the existence of downstep on the basis of the local effects. Some register lowering, of course, does take place in Yoruba downstep (cf. comparisons 3 and 4, §2.2.2.2). But in certain cases register lowering seems to be effectively cancelled by local H-raising, giving rise to the 'hogback ridge' pattern referred to in §1.1.4 (Fig. 1c). This is seen in Fig. 5, in which the H on *àpọ́n* – which should be downstepped by the preceding HL – is raised before the following L on *ò* all the way to the level of the first H on *ṣé*. We have no systematic basis for studying this phenomenon in the present corpus, but it seems worth further investigation.

	Scaling of L early in utterance			
	in all-L sentence	*in the sequence HLH*		
Speaker	(Dataset 3)	(Dataset 7)	*t*	(df)
A	195 (11·9)	175 (8·7)	4·07	(18) ***
B	199 (10·9)	171 (18·8)	3·49	(14) **
F	100 (4·1)	96 (5·2)	2·06	(21) *a*
O	185 (10·1)	162 (9·2)	5·32	(18) ***

[Table VI. *Lowering of L in the sequence HLH. See further the caption to Table II*]

3.2.2 *Downstepping of L?* The other phenomenon for which our corpus provides some unplanned evidence is the lowering of L between Hs. This is seen in the (L)HLH sequences at the beginning of datasets 7 and 9 (question particle followed by statement beginning LH). The L tones in this context are substantially lower than the initial L of the all-L sentences, as can be seen from Table VI. Note that datasets 7 and 9 are questions, whereas the all-L sentences are statements, so that in the absence of any lowering effect we might actually expect the L in datasets 7 and 9 to be higher rather than lower. Taking this effect together with the raising of H before L just noted, it might appear that in HL sequences the contrast between them is exaggerated in both directions by raising the H and lowering the L. Unfortunately we have no basis in our corpus for seeing whether L-lowering applies in *any* HL sequence (i.e. irrespective of the following tone) or whether it is strictly an effect on L tone *between* Hs.

An alternative possible interpretation of this lowering of L, especially if it can be shown to occur in any HL sequence and not just before H, is that it is not a local effect at all, but the manifestation of downstep. That is, it might be taken as evidence that the register modification involved in downstep 'occurs' in a HL sequence *before* the L. Exactly that has been suggested for Japanese by Poser (1984), and evidence for Poser's suggestion is presented by Pierrehumbert & Beckman (1988: 87ff). Our data could readily be taken as further evidence of this proposal.

If this interpretation is accepted, it has a couple of interesting theoretical consequences. First, it makes it plausible that downstep in Japanese and Yoruba are essentially the same phenomenon. Second, it makes it clear that downstep does apply to L tones as well as to H and M – something which, as we saw in the first half of the paper, has been a point of considerable uncertainty in African linguistics. This, as we pointed out in §3.1, would make it difficult to accept Pierrehumbert & Beckman's phonetic model of downstep as register *narrowing*, and would provide support for the proposals of Clements (1979, 1990) or Ladd (1987, 1990) that downstep involves register *lowering*.

3.3 Conclusion

Summarising the various effects just discussed, it seems clear that neighbouring H and L tones exert local influences on each other. In §2.2.1.2 we also noted the possibility that local L-tone effects are involved in declination. However, given the limitations of our corpus we cannot hope to shed any light on the systematic relationship between these effects, much less to ascribe them correctly to phonological rules or phonetic realisation processes. Further theoretical and empirical work is needed: we suggest that careful instrumental studies of the behaviour of L tones in downtrends would be most worthwhile.

APPENDIX

Recording and analysis procedure

1 Recording

Speakers read the sentences from computer-printed slips of paper. At the beginning of the recording session they were given a stack of slips that included (a) several practice sentences, (b) the 108 sentences of the corpus, in one of two pseudo-random orders designed to avoid repetitions of lexical items or experimental conditions and (c) several additional sentences at the end, so that the final sentence or sentences of the actual corpus would not be marked by any discourse-final Fo effects.

There were two serious problems using visual presentation of the experimental sentences. One is that ordinary Yoruba orthography makes no systematic use of tone diacritics, which tend to be written only when required in context to avoid ambiguity. The other problem is that literate Yorubas, like many educated Africans, are more accustomed to reading a colonial language (in this case English) than their native language. (The same difficulty, with French as the colonial language, is reported by Mountford in his 1983 study of Bambara.) Fully tone-marked Yoruba text would therefore apparently strike most Yoruba readers roughly the way most phoneticians would be struck by a phonetic transcription of their own language. Unfortunately, however, some tone-marking was essential, since our experimental sentences involved some tonal minimal sets, and since in any case the nature of the reading task provided fewer contextual cues than would be present in connected prose. We note in passing that the minimal sets, which seemed like an ideal way to control for segmental differences, in fact turned out to be a major methodological headache. We strongly advise future investigators of languages where tonal ambiguities are not adequately resolved in conventional orthography to avoid minimal sets, unless these are quite specifically the focus of interest.

We tackled these problems by including English glosses on the slips with the Yoruba sentences. The Yoruba sentences were minimally tone-marked according to the intuitions of Speaker F. Speakers were instructed to read silently and understand the Yoruba sentence, using the English if necessary to resolve tonal ambiguities, before speaking. (In practice they occasionally also consulted with Speaker F, who remained in the recording booth during all the sessions.) The tapes of all the recording sessions were listened to afterwards by Speaker F, and utterances that he judged seriously unnatural or which he found to contain tones other than those intended were excluded from subsequent analysis. While recognising the many potential objections to these procedures, we feel that they represent a reasonable compromise between the conflicting goals of naturalness and experimental control.

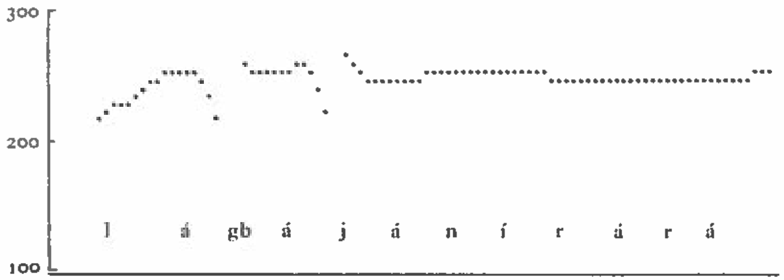

Figure 6

Fo remaining unchanged across several syllables (Speaker A).

2 Acoustical analysis

The utterances were digitised at 10 kHz and low-pass filtered at 5 kHz; Fo was then extracted using the cepstrally-based API command of the ILS (Interactive Laboratory System) software package. Subsequent work was done interactively for each utterance separately, using the graphic display routines of ILS, plus the ILS-compatible FRED Fo display routine (Silverman 1986).

The basic experimental approach, as discussed in the body of the paper, was to identify the Fo levels of a certain number of key syllables and/or presumed tonal targets in each sentence. Since syllables are not single instants in time and since Fo turning points are not always aligned with syllables in ways one might expect, this procedure required us to make various decisions about where exactly, relative to the segmental sequence, to measure Fo. In this section we discuss these decisions in some detail in the interests of helping to develop standard procedures for such investigations.

(a) *Sequences of syllables with the same phonological tone.* In some cases the extracted Fo remained completely unchanging over stretches of several syllables (Fig. 6), in which case there was no question about the Fo value to be recorded for any given syllable in that stretch. Often, however, because of segmental perturbations or for other reasons, the extracted Fo was irregular, or changed during the course of a syllable and/or from one syllable to the next. In these cases, if a syllable was predominantly at one level, that level was chosen as its Fo level (Fig. 7); if the syllable was predominantly rising or falling but with a 'shoulder' or levelling-off in the middle, the level of the shoulder was chosen (Fig. 8); if the syllable was rising or falling without any obvious shoulder, the level of the syllable's energy peak was chosen (Fig. 9); in doubtful cases where Fo had been extracted for only a few analysis frames, no value was recorded (Fig. 10).

(b) *Sequences of tones involving well-defined Fo peaks or valleys.* For the most

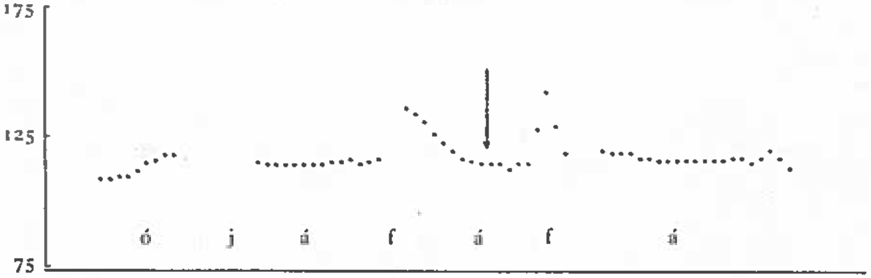

Figure 7

Fo changing across a syllable, but predominantly at one level (arrow), giving a basis for measurement (Speaker F).

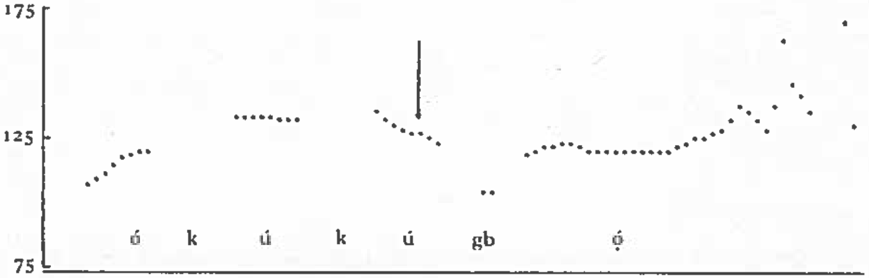

Figure 8

Fo changing across a syllable, but with a 'shoulder' (arrow) giving a basis for measurement (Speaker F).

part these were HLH and LHL sequences. It is well known that in Yoruba H is realised as a rise after L and L is realised as a fall after H. In effect, what this means is that in HLH and LHL sequences, the Fo target representing the H in a rise or the L in a fall is aligned at the *end* of the syllable with which it is associated. Subject to the qualifications noted in the next subsection, what we did in these cases was to take Fo valleys as the value of L tones and Fo peaks as the value of H tones, regardless of where they were aligned with respect to the segmentals (Figs. 11 and 12).

(c) *Segmental perturbations*. All the speakers, but especially Speaker A, exhibited marked Fo dips at voiced and/or labial-velar obstruents (Fig. 13). For the most part these could be (and were) ignored, but in a HLH sequence like *ṣé àpọ́n* (cf. Fig. 5) they affect the level of the Fo valley which would otherwise, in accordance with the preceding subsection, have been taken as the value of the L tone. In such cases we attempted to identify the beginning and/or the end of the obstruent dip and to record that value as the Fo value for the valley, but this was not always possible, and there is no doubt that these cases are a source of variability in the data.

(d) *Sequences of tones involving transitional syllables between a H and a L*. For the most part these were sequences HLL (i.e. H–Fall–L) and LHH (i.e. L–Rise–H). In these cases it was generally impossible to identify an endpoint to the 'fall' or the 'rise' independent of the following L or H; the Fo on the middle syllable was simply transition from the pitch of the first syllable to that of the third. In these cases no attempt was made to record an Fo value for the middle syllable (Fig. 13).

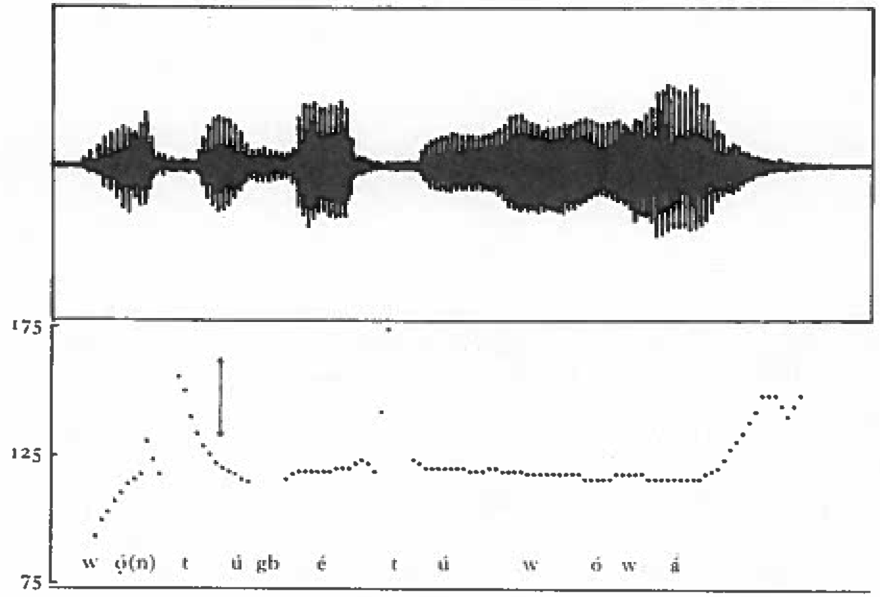

Figure 9

Fo changing across a syllable, with no obvious basis for measurement in the pitch curve. In these cases the pitch value coinciding with the energy peak of the syllable (arrow) was chosen as the pitch value for the syllable (Speaker F).

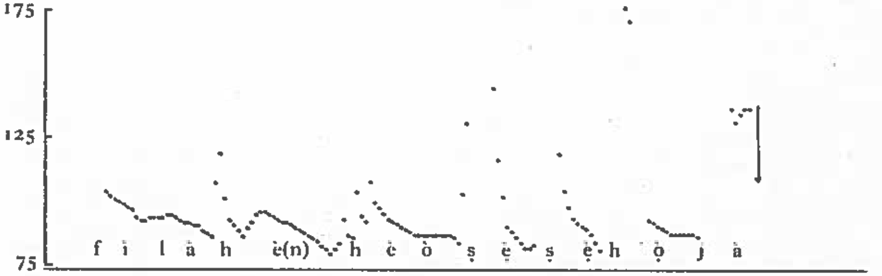

Figure 10

Example of syllable (arrow) where the pitch extraction was deemed too unreliable for a value to be taken (Speaker F).

3 Statistical analysis

The planned comparisons with which we explored experimental questions 2 and 3 required us to assess the difference between mean values of tones in the 'shared texts' of various pairs of datasets, as discussed in the body of the paper.

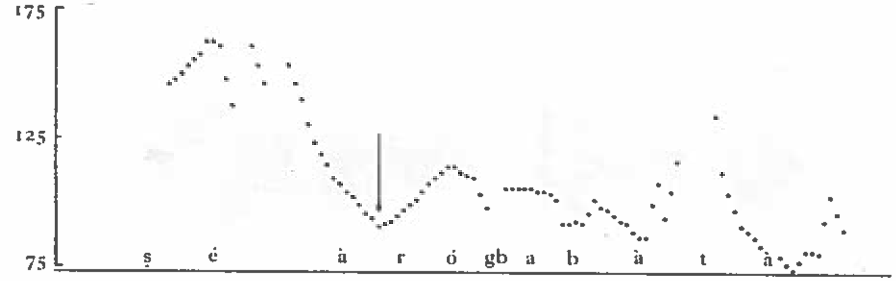

Figure 11

Fo value at valley (arrow) taken as the value of the L in a HL (phonetic High–Fall) sequence (Speaker F).

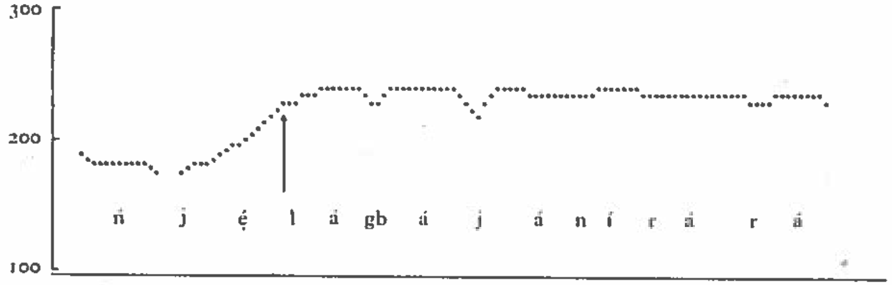

Figure 12

Fo value at peak (arrow) taken as the value of the H in a LH (phonetic Low–Rise) sequence (Speaker A).

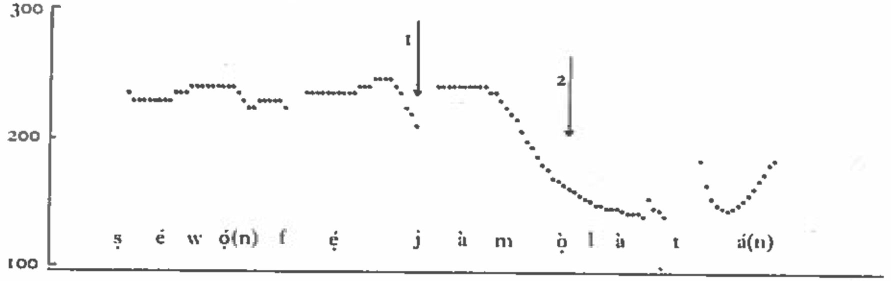

Figure 13

Fo dip caused by voiced obstruent (arrow 1); syllable whose tonal value was omitted as being purely transitional (arrow 2) (Speaker A).

As the basis for these comparisons, we took the means of all the H tone measurements and all the M tone measurements in the shared texts of all the sentences in each dataset. (We had intended to compute averages for L as well, but were unable to do so for reasons explained below.) In computing these means, we treated the pitch of *each* shared-text syllable as a data point. For example, the shared text *ní rárá* was intended to contribute 3 measurements to

the mean for H tone; the shared text *jàmọ̀là tán* was intended to contribute 1 measurement to the mean for H and 3 to the mean for L. This has the potential drawback that more than one value for any given mean may come from a single utterance, thereby exaggerating the effect of any abnormally high or low utterance. To compensate for this, in determining the significance levels in *t*-tests we took the number of *sentences* that had contributed to the mean, rather than the number of syllables, as the basis of the number of degrees of freedom. For example, in each Speaker B's datasets 4 and 5 we measured 9 H tones, but these came from only 4 sentences. The degrees of freedom for comparing these two sets were therefore 6 (= 4+4−2). All *t*-tests were two-tailed. While many objections could obviously be raised to this procedure, it does at least incorporate various provisions that work in favour of conservative data interpretation, i.e. against the too-easy acceptance of our hypotheses.

This intended design met with several problems. First, the exact number of measurements on which each mean is based varies rather considerably from speaker to speaker. There were two reasons for this: first, we were not able to use every speaker's production of every sentence in the corpus in our data analysis; second, in some cases we were not able to assign a value for a syllable even if the utterance as a whole was usable, either because of elision (especially common in Speaker A's data) or because, as mentioned above, some syllables had pitch contours that appeared to be pure transition (e.g. the syllable *-mọ̀-* in the shared text *jàmọ̀là tán*, Fig. 13).

As noted above, we had to abandon the comparison of L tone means. There were two main problems. First, many of the L syllables in the corpus were in sentences that speakers tended to omit or to pronounce with the wrong tones, so that the L means would have been based on fewer measurements than the H and M means. Second, many of the L tone syllables that were correctly produced were either (a) within the last few syllables of their respective utterances, and therefore presumably subject to the rapid declination or final lowering discovered in dataset 3, and/or (b) flanked by obstruents, and therefore subject to substantial Fo perturbations (a good example of both problems is the shared text *gba bàtà*).

Finally, the shared text *jàmọ̀là tán* (Fig. 13) provides further illustration of the problems we faced in comparing tones: the sentence-final H-toned word *tán* was subject to something like 'total downstep' (i.e. lowering all the way to the level of the preceding L) in all the speakers' renditions. Excluding it from the H-tone mean would have meant reducing the number of cases on which the mean was based, while including it would have meant drastically increasing the variance of the mean, either choice limiting our ability to draw useful statistical inferences. In this case we chose to exclude it, but we have no way of knowing whether this or other similar choices were always the most appropriate.

NOTES

* The research reported here represents an instance of an activity that is increasingly marginalised under the rubric of 'curiosity-driven research'. (In fact, it originated in a bet between the authors, which both concede neither won.) As such it was not supported by any research grant. We are grateful that the freedom to engage in such research, even in the present academic and political climate in Britain, has not yet entirely vanished. More directly, we wish to record our gratitude to Lawrence Olúfẹ́mi Adéwọlé (at the time a PhD student in the

Department of Linguistics at Edinburgh University, now of Ọbáfẹ́mi Awólọ́wọ̀ University in Ife, Nigeria) for his careful and patient help in creating the corpus and for recruiting Yoruba speakers for the recordings. He steered us away from many pitfalls and in a very real sense made the whole study possible. Our presentation of the study owes much to comments by two anonymous referees for *Phonology*, and by audiences at the BAAP colloquium in Dublin (April 1988) and the 6th International Phonology Meeting in Krems (July 1988). We take full and equal responsibility for the paper in its final form; our names are given in alphabetical order.

[1] In distinguishing downtrends due to phonetic realisation rules from phonologically conditioned effects, we nevertheless remain agnostic about the various physical explanations that have been proposed for the backdrop decline in Fo. These include gradual decline in subglottal pressure (Cohen *et al.* 1982), gradual downward pull on the trachea with decreased lung volume (Maeda 1976), and greater ease of producing Fo falls than Fo rises (Ohala & Ewan 1973: 345). We do note, however, that some of these explanations (e.g. Maeda's) appear to follow Cohen & 't Hart in conflating declination in the narrow sense with downstep and other phonological factors.

[2] Whether final lowering can be related to the abrupt fall in subglottal pressure that accompanies the end of phonation (cf. Lieberman 1967) is an open question. Liberman & Pierrehumbert (1984: 219) consider the connection to be unlikely, but to us it seems at least a plausible speculation. Pierrehumbert & Beckman's model treats final lowering as a matter of phonetic realisation, but the authors suggest elsewhere (e.g. Beckman & Pierrehumbert 1986: 302–305) that the *amount* of final lowering may be used for 'paralinguistic' purposes in indicating the boundaries of various-sized chunks in discourse structure. We do not pursue this matter further here.

[3] Thanks to John Stewart and Larry Hyman for discussion of this section of the paper.

[4] Because of such lexical contrasts, early phonemic descriptions of African languages with non-automatic downstep sometimes recognised this downstepped H as a toneme in its own right (usually as a mid tone), distinct from non-downstepped H. However, beginning with an important paper by Winston (1960), it has been customary to treat downstepped H as resulting from the application of some sort of locally triggered phonological process. The local trigger has sometimes been seen as a separate downstep phoneme (often written as a raised exclamation mark before the affected H tone), though more recently it has often been viewed simply as a floating L tone that does not appear in the surface tonal string.

[5] One might adopt Poser's term 'catathesis' for all such phonologically conditioned lowerings, and restrict 'downstep' to its original Africanist usage or abandon it altogether. However, we feel that the broad similarities among all these effects in different languages warrant the extension of downstep – as in Pierrehumbert's original proposal. Moreover, we feel that the status of English as the international language of science militates against replacing a well-known and transparently interpretable English term by an obscure classical neologism.

[6] Akinlabi (1985) and Pulleyblank (1986) argue that underlying lexical representations can be substantially underspecified tonally, but the characterisation just given is adequate for the input to the postlexical phonology, which is of course what is at issue in this paper. In certain contexts (e.g. certain one-syllable verbs) the contrast between M and L may sometimes be neutralised, and there are also certain preferred sequences of tones in two-syllable nouns and various cooccurrence restrictions between tones and segmentals (e.g. most nouns with H on the first syllable begin with a consonant). For the most part, however, all three tonemes can occur quite freely in all parts of an utterance. This is methodologically important for our purposes; cf. the problems encountered by Mountford (1983) in constructing all-H and all-L sentences in Bambara.

[7] Examples are taken from Rowlands (1969: 34, 29). For more detail see Archangeli & Pulleyblank (1989) or Akinlabi & Oyabede (1987). In some of our sentences elision was spelled in (e.g. *jàmọ̀là* for *je àmọ̀là*), but the speakers, especially Speaker A, also made other elisions that were not spelled in.

[8] This procedure presupposes that downstep and other such phenomena represent the *overall* lowering of register, with effects that persist throughout a phrase or utterance. As discussed above, this assumption seems to be uncontroversial, but we wish to make it explicit here, as we will show in §3.2.1 that at least in Yoruba it may not be the whole story.

[9] The pragmatic difference between the two question particles is subtle, and it is beyond the scope of this article to discuss this problem. Our consultant judged all the questions in sets 6–9 to be pragmatically possible.

[10] The consistent final low pitch at the end of the all-L sentences is comparable to the speaker-specific constant 'baseline' pitch that has been found in many studies of European intonation systems (e.g. Maeda 1976; Menn & Boyce 1983; Liberman & Pierrehumbert 1984; Ladd 1988). As for the final rise at the end of the all-H and all-M sentences, its presence or absence appears to be a point of variation between speakers: speakers F and B exhibit this strongly and consistently; speaker O is rather inconsistent; speaker A shows this characteristic hardly at all.

[11] In this connection it is worth mentioning anecdotal evidence from Yasuko Madsen (personal communication). Madsen used her own native pronunciation of Japanese words for ear-training when she taught phonetics in Nigeria; she reports that students whose native language was Yoruba consistently perceived declination in phonologically H-tone stretches, and transcribed it as a gradual drop in pitch, whereas those whose native language was Hausa did not perceive the declination and transcribed the pitch 'correctly' as all H-tone.

[12] For details of how we computed statistical significance see the Appendix.

[13] It would be possible to model downstepping Ls in the Pierrehumbert & Beckman model, but only in one of two ways, both of which seem counterintuitive: (i) L tone, like H, could be scaled above the reference line. This is in fact what Pierrehumbert & Beckman propose for Japanese, but their proposal is motivated by various effects of prominence on the scaling of Ls, which would not be especially relevant in Yoruba. Indeed, if all L tones were maximally prominent, then they would all (by Pierrehumbert & Beckman's 1988: 188 proposal for Japanese) be scaled on the reference line anyway, producing a picture like Fig. 1b. In order to produce downstepping Ls, the Ls would have to be consistently less than maximally prominent. (ii) L tone could be scaled below the reference line. This is what Liberman & Pierrehumbert (1984: 207ff) propose for English (it is not clear whether Pierrehumbert & Beckman 1988: 181 see their proposal for Japanese L tone as superseding the Liberman & Pierrehumbert model). This would make the computation of a downstepping 'L-tone line' mathematically independent of the downstepping of the H-tone line, and would thus divide what is presumably a single phenomenon into two.

REFERENCES

Abraham, R. C. (1962). *Dictionary of modern Yoruba*. London: Hodder & Stoughton.

Adéwọlé, L. O. (1987). *The Yoruba language: published works and doctoral dissertations 1843–1986*. Hamburg: Helmut Buske.

Akinlabi, A. M. (1985). *Underspecification of tones in Yoruba*. PhD dissertation, University of Ibadan.

Akinlabi, A. & F. Oyabade (1987). Lexical and postlexical rule application: vowel deletion in Yoruba. *Journal of West African Languages* **17**. 23–42.

Archangeli, D. & D. Pulleyblank (1989). Vowel harmony in Yoruba. *LI* **20**. 173–217.

Bamgbose, A. (1966). *A grammar of Yoruba*. Cambridge: Cambridge University Press.

Beckman, M. & J. Pierrehumbert (1986). Intonation structure in Japanese and English. *Phonology Yearbook* **3**. 255–309.

Berg, R. van der, A. Rietveld & C. Gussenhoven (forthcoming). Downstep in Dutch: implications for a model. In G. J. Docherty & D. R. Ladd (eds.) *Papers in laboratory phonology II*. Cambridge: Cambridge University Press.

Clements, G. N. (1979). The description of terraced-level tone languages. *Lg* **55**. 536–558.

Clements, G. N. (1990). The status of register in intonation: comments on the paper by Ladd and by Inkelas and Leben. In Kingston & Beckman (1990). 58–71.

Clements, G. N. & K. C. Ford (1979). Downstep in Kikuyu. *LI* **10**. 179–210.

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

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

Courtenay, K. (1971). Yoruba: a terraced-level language with three tonemes. *Studies in African Linguistics* **2**. 239–255.

Gussenhoven, C. (1984). *On the grammar and semantics of sentence accents*. Dordrecht: Foris.

Gussenhoven, C. & A. Rietveld (1988). Fundamental frequency in Dutch: testing three hypotheses. *JPh* **16**. 355–369.

Hart, J. 't (1979). Naar automatisch genereeren van toonhoogte-contouren voor tamelijk lange stukken spraak. *IPO Technical Report* **353**.

Hombert, J.-M. (1974). Universals of downdrift: their phonetic basis and significance for a theory of tone. *Studies in African Linguistics*. Supplement 5. 169–183.

Hombert, J.-M. (1978). Consonant types, vowel quality, and tone. In V. Fromkin (ed.) *Tone: a linguistic survey*. New York: Academic Press. 77–111.

Kingston, J. & M. Beckman (eds.) (1990). *Papers in laboratory phonology*. Cambridge: Cambridge University Press.

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

Kubozono, H. (1989). Syntactic and rhythmic effects on downstep in Japanese. *Phonology* **6**. 39–67.

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

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

Ladd, D. R. (1987). A phonological model of intonation for use in speech synthesis. In J. Laver & M. Jack (eds.) *Proceedings of the European Conference on Speech Technology*. Vol. 2. Edinburgh: CEP Associates. 21–24.

Ladd, D. R. (1988). Declination 'reset' and the hierarchical organization of utterances. *JASA* **84**. 530–544.

Ladd, D. R. (1989). Review of Pierrehumbert & Beckman (1988). *JL* **25**. 519–526.

Ladd, D. R. (1990). Metrical representation of pitch register. In Kingston & Beckman (1990). 35–57.

Laniran, Y. (1988). Modeling intonation in a tone language. Ms, Cornell University.

Laniran, Y. (1989). Phonetic implementation rules for tones in Yoruba. *JASA* **85**. S99.

Laniran, Y. (in preparation). *Intonation in tone languages: the Yoruba example*. PhD dissertation, Cornell University.

LaVelle, C. R. (1974). An experimental study of Yoruba tone. *Studies in African Linguistics*. Supplement 5. 185–194.

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

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

Lindau, M. (1986). Testing a model of intonation in a tone language. *JASA* **80**. 757–764.

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

Menn, L. & S. Boyce (1983). Fundamental frequency and discourse structure. *Language and Speech* **25**. 341–383.

Mountford, K. (1983). *Bambara declarative sentence intonation*. PhD dissertation, Indiana University.

Ohala, J. J. & W. Ewan (1973). Speed of pitch change. *JASA* **53**. 345.

Peters, A. M. (1973). A new formalization of downdrift. *Studies in African Linguistics* **4**. 139–154.

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

Pierrehumbert, J. & M. Beckman (1988). *Japanese tone structure*. Cambridge, Mass.: MIT Press.

Pike, K. L. (1948). *Tone languages*. Ann Arbor: University of Michigan Press.

Poser, W. (1984). *The phonetics and phonology of tone and intonation in Japanese*. PhD dissertation, MIT.

Pulleyblank, D. (1986). *Tone in Lexical Phonology*. Dordrecht: Reidel.

Rowlands, E. C. (1969). *Yoruba*. (Teach Yourself Series.) Bristol: Hodder & Stoughton.

Silverman, K. (1986). FRED: an interactive graphics program to modify frequency contours in resynthesized speech. *Behavior Research Methods, Instruments and Computers* **18**. 395–397.

Stewart, J. M. (1965). The typology of the Twi tone system. Preprint of part of abortive *Bulletin of the Institute of African Studies* **1**. 1–27.

Stewart, J. M. (1983). Downstep and floating low tones in Adioukrou. *Journal of African Languages and Linguistics* **5**. 57–78.

Welmers, W. (1959). Tonemics, morphotonemics, and tonal morphemes. *General Linguistics* **4**. 1–9.

Welmers, W. (1973). *African language structures*. Berkeley: University of California Press.

Winston, F. D. D. (1960). The 'mid' tone in Efik. *African Language Studies* **1**. 185–192.