THE PHONETIC INTERPRETATION OF TONE IN IGBO

Mark Liberman, J. Michael Schultz, Soonhyun Hong, Vincent Okeke

Department of Linguistics
University of Pennsylvania
Philadelphia, PA 19104-6305

ABSTRACT

Igbo, a language of the Kwa branch of the Niger-Congo family, is spoken by about 15 million people in southeastern Nigeria. Its phonology, morphology and syntax have been widely studied (e.g. [1, 2]), especially with reference to the intricate patterning of lexical tone. This paper is a preliminary study of the phonetic interpretation of Igbo tone. We use an experimental method first applied to English ([4, 5]), in which a speaker varies pitch range orthogonally with variation in tonal material, and we compare the success of different models in characterizing the interaction of tone identity, phrasal position, tone sequence, and pitch range in determining patterns of measured F0 values. From the statistical structure of these data, we draw several conclusions about Igbo tone and its phonetic interpretation.

(A shorter version of this paper was published as [3])

1. Lexical Tone in Igbo

In this section, we will introduce the system of lexical tone in Igbo, especially as it relates to the design and interpretation of the experiments we have done. We pass over in silence most of Igbo's intricate and fascinating tonal phonology, since our immediate concern is with the phonetic interpretation of the surface tonal categories. However, the basic nature of these surface categories will emerge as a crucial question.

An Igbo syllable can have one of three tonal categories, known as "high" (H), "low" (L), "mid" (M). The H and L tones occur freely, but the M tone can only occur following an H tone or another M tone. Thus there are five possible tone patterns for two-syllable words: (1) HH, (2) HL, (3) LH, (4) LL, (5) HM. Monosyllabic words are rare, and all of those that can occur in isolation have high tone.

Among the Igbo as elsewhere, speakers are free to change their pitch range, and do so for many reasons. We may ask whether tonal distinctions are nevertheless maintained—for instance, is HH in a low pitch range distinguished from LL in a higher one? Are HM and HL kept distinct? If so, how? Our experiments provide a tentative answer: contextual effects in the system of tonal interpretation largely maintain the distinctions among bitonal patterns as pitch range varies.

There are several general effects that modify the realization of Igbo tones in phrasal context. One of the most important of these is known as "downdrift," which progressively lowers H and L tones when they occur in sequence. Thus on one typical pronunciation of the phrase Ábànóbì ọ́ma, "good Abanobi," the successive minimum and maximum F0 values were as follows:

TEXT:	A	ba	no	bi	ọma
TONES:	**H**	**L**	**H**	**L**	**H**
F0:	237	178	210	158	181

A model for the phonetic interpretation of Igbo tone must obviously take account of downdrift. The term "downstep" is used to refer to a different circumstance. Although some phonologists have treated Igbo "mid" tones as a third tonal category, distinct from high and low (e.g. [6], [7]), others (e.g. [1]) have interpreted the restricted distribution of the "mid" tone to mean that it is actually just a "high" tone that happens to be "downstepped", i.e. realized at a lower pitch. On this view, HM is actually H $^{!}$ H, where the raised exclamation point marks the downstep location. This situation (or its counterpart in other tone languages) has been given various phonological interpretations:

1. the downstep marker may be reified as a separate phonological entity, as suggested by the exclamation-point diacritic;
2. downstep may be viewed as the consequence of a "floating" low tone between the two high tones, which thus triggers lowering by the more general process of downdrift, but is not otherwise realized ([8], [9]);
3. the downstepped sequence HM may be viewed as the expression of two distinct high tones, whereas an HH sequence is viewed as a single high tone spread over two syllables ([1]).

One obvious question is whether downdrift and downstep are phonetically the same, as we might expect from the second hypothesis given above. One of our experiments suggests that Igbo downdrift and downstep are phonetically different, in a way that may be interpreted to support a variant of Clark's theory in [1].

2. Design of the experiments

We will mark Igbo tone using the system of [22, 11], which expresses the concept of downstep made explicit in Clark's theory: "high" is written with an acute accent; "low" is written with a grave accent; an unmarked syllable continues the previous tone; and a repeated "high" tone mark is interpreted as "mid." In this system, as in all the similar tonal orthographies known to us, downdrift is not marked, because it is assumed to be automatic (or if variable, only expressively so).

In this notation, the materials for our first experiment consisted of the 13 words:

HH:	ísi *head*	óke *male*	íre *tongue*
HL:	ísì *odor*	ókè *boundary*	írè *to be effective*
LH:	ìsî *six*	òké *rat*	
LL:	ìsi *blindness*	òke *share*	ìre *effectiveness*
HM:	ísí *to cook*		íré *to sell*

On each trial, the subject was asked to read a word or phrase in one of three modes: addressed quietly to someone seated nearby; addressed to someone seated on the other side of a broad table, a little more than a meter away; or addressed to someone at the other end of a room, about 10 meters away. 195 utterances (five repetitions of each phrase in each mode) were elicited in random order. The subject was a man aged 45, from the village of Awo-Omamma in Imo State, about halfway between Owerri and Onitsha.

This procedure produces a good deal of pitch range variation, of the kind involved in thus "raising" or "lowering" the voice. Comparable measurement points in different tokens of the same tonal type have F0 values up to about an octave apart, which is several times larger than the difference between lexically-distinct tone categories in a given pitch range. As a result, we often find (for instance) that an initial L tone in a wide-pitch-range utterance is actually higher than an initial H tone in a narrow-pitch-range utterance.

Loudness and duration also vary in an experiment of this type, giving an independent indication of the speaker's level of vocal effort. We would like to point out that there are many other functional dimensions that are often associated with F0 effects that could be described in terms of "pitch range" variation. Examples include the speaker's overall level of arousal, the topic structure of the discourse, and the relative prominence of particular words and phrases. It should not be assumed that all such F0 effects will show the same patterns as the effects we have studied here, which are linked with the distance of an interlocutor.

We will not model the duration and amplitude effects of the range variation that we have induced, nor will we try to model all aspects of the F0 contour. We will represent the pitch of each syllable by a single value, taken automatically from the F0 time-function. We have tried various definitions for this representative value, including the value at the syllable mid-point and the average or median value; we find that the general structure of the data is the same, but the cleanest patterns result when we pick the maximum value for H tones, and the minimum value for L tones. We will discuss this point somewhat further when we consider the question of whether high tones are raised before low tones.

Like many (but apparently not all) tone languages, Igbo shows a pattern in which tones occuring later in a sequence are sometimes lowered relative to the values for the tones occuring earlier. In order to model these downtrends, we need to look at F0 patterns in longer sequences of tones. Thus our second experiment used the 13 phrases shown below, each produced six times in each of the three pitch-range modes. These phrases are personal names, or concatenations of personal names with "nà" *and*, a construction chosen because it is semantically flat, and also fails to induce the complex tone changes that occur with many syntactic combinations.

1	Díké nà Áma	HM L HH
2	Íke nà Áma	HH L HH
3	Íbè na Áma	HL L HH
4	Ńne Ụ́ba nà Íbè	HH MH L HL
5	Ụ́gwụ nà Íbè	HH L HL
6	Áma nà Íbè	HH L HL
7	Áma nà Íke	HH L HH
8	Íke nà Áma nà Ába	HH L HH L HH
9	Ọ́nụ́ ọma	HM HH
10	Ọ́lụ̀ká ọma	HLH HH
11	Ọ́nwụká ọma	HHM HH
12	Ábànóbì ọ́ma	HLHL HH
13	Ábàríkwú ọma	HLHM HH

The term *downdrift* applies to the regular lowering of H and L in alternating sequence, while *downstep* is used for the so-called "mid" tone, which is thus considered to be a lowered version of H. The materials in Experiment 2 are designed to let us characterize and compare these phenomena.

3. Results: Experiment 1

Figure 1 shows the some data from Experiment 1 as a a scatter plot in which the X axis shows the F0 value of the first syllable, and the Y axis shows the F0 value of the second syllable. The $y = x$ line is plotted as well.

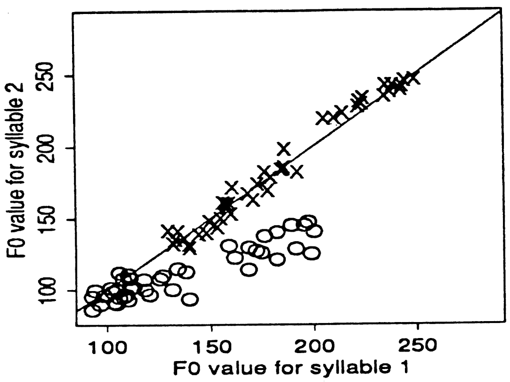

Figure 1: Igbo HH (X) LL (O) Disyllables.

We can see several things in Figure 1. First, although the points span a broad range, their relationship is quite a tight one: the correlation of the HH measurements is .988, while the correlation of the LL measurements is .905.

Second, the two H tones are generally about equal in value, but the second of the two L tones is usually lower than the first. Functionally, this helps distinguish an HH sequence in a low pitch range from an LL sequence in a high pitch range. We can express the same difference between the HH and LL data by saying that the LL trend is best fit by a line with a slope less than .5 and a non-zero intercept, while the HH data is not statistically distinguishable from $y = x$. Regression on the HH data gives a slope of 1.0 (standard error .025), and an intercept of -6.8 (s.e. 4.7). Regression on the LL data gives a slope of .44 (s.e. .032), intercept 51.9 (s.e. 4.5).

Figure 2 shows that the H/L proportion is systematically greater in the HL order than in the LH order, consistent with the fact that L is lower in the second position of LL sequences. Also, the HL data is clearly requires a non-zero intercept, while the LH data does not. Thus regression on the HL data gives a slope of .25 (s.e. .03), with an intercept of 60 (s.e. 6.1), while the LH data gives a slope of 1.3 (s.e. .08), with an intercept of -5.7 (s.e. 12). The HM data gives a slope of .92 (s.e. .03), and an intercept of -5.6 (s.e. 5.4). The fact that the HM slope is so much greater than the LL slope calls somewhat into question the interesting suggestion of Stewart ([12]) that the lowering of final low syllables is to be interpreted as "another manifestation of key lowering," a cover term in which he includes downstep.

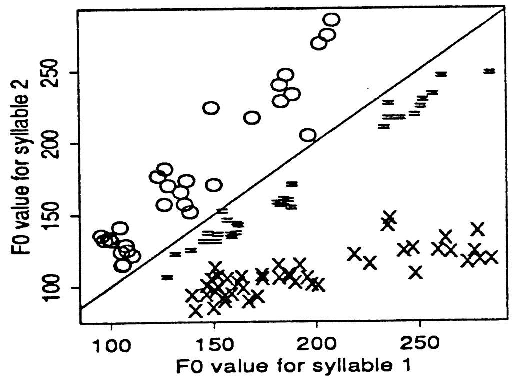

Figure 2: Igbo HL (X) LH (O) HM (=) Disyllables.

The patterns in the data of Experiment 1 suggest a model in which the F0 is predicted to be $T + T\,R\,F\,D$. Here T is a factor that has one value for H and another for L; F is a factor that is < 1 for L in second position and 1 otherwise; D is a factor that is < 1 for a "downstepped high" (here mid) tone; and R is a "latent variable" representing the pitch range of a given utterance. This model yields the following expressions for predicting the syllable 2 pitch (y) from the syllable 1 pitch (x) in a given utterance:

$$\begin{array}{lrcl} \text{HH} & y & = & x \\ \text{HL} & y & = & (FL/H)x + (1-F)L \\ \text{LH} & y & = & (H/L)x \\ \text{LL} & y & = & Fx + (1-F)L \\ \text{HM} & y & = & Dx + (1-D)H \end{array}$$

The model suggests values for the L and H parameters which are quite reasonable, given the speaker's observed F0 in low pitch-range utterances. The model also predicts significant intercept terms in the LL and HL cases, where we did find them, and no intercepts in the HH and LH cases, where we didn't. Unfortunately, it predicts an intercept in the HM case as well, where we didn't find one; the predicted intercept is small, since $1 - D$ appears to be small, but this may point to a problem in the model. $(T+TRF)D$ would predict no intercept for the HM case; space does not permit further exploration.

We should point out that this model embodies some probably wrong assumptions about the nature and environment of the lowering conditioned by D, namely that (in disyllables) it applies only to the HM sequence, and not (for instance) to the L in the HL sequence.

Having decided on the structure of our model (right or wrong), we can optimize its parameters with respect to the whole data set.[1] Doing this on the data of Experiment

[1] Treating the R parameters as latent variables, we use the downhill simplex method ([13]) to perform the optimization.

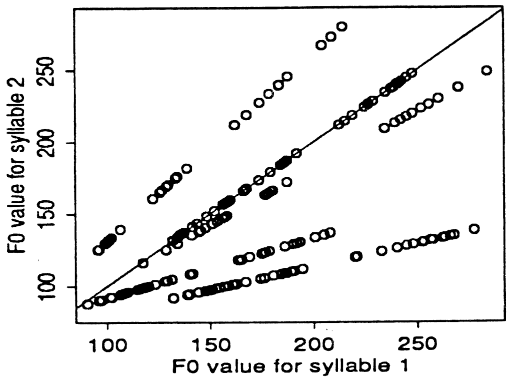

Figure 3: Model Values for Igbo Disyllables

1, assuming the $T{+}TRFD$ model with D applying only to the HM case, we get parameter estimates H=112, L=86, F=0.42, D=0.80, and an RMS error across all the data of 6.2 Hz. Figure 3 plots the result of replacing each data point with its model counterpart. In this plot, the basic features of the data set are well captured. Given space limitations, we leave it to the reader to convince herself that certain other simple functional forms for the model, such as $TRFD$, $T + R + F + D$, and so on, produce qualitatively wrong predictions. Their overall prediction error is significantly higher, but just as important, some of the key qualitative aspects of the data are misrepresented. We find it especially interesting that the all-multiplicative model makes such obviously wrong predictions, since this failure calls into question the popular practice of using a semitone scale for F0 contours.

4. Results: Experiment 2

The second experiment permits us to explore some other interconnected questions about the phonetics of Igbo tone. For instance, we can ask whether downdrift affects both H and L tones in the same way; whether downstep and downdrift are the same; and whether H is raised in front of L.

4.1. Raising of H before L

We take up the last point first, since it affects the measurements used in evaluating the other issues. In two independent studies of Yoruba, Connell and Ladd ([14]) and Laniran ([15]), suggest that H tones are raised before L. They offer various arguments, notably the fact that the final H in sequences such as HHL, HHHL, HHHHL, etc. is markedly higher than the non-final occurrences, and that this effect is not seen in sequences such as HHM, HHHM, HHHHM, etc.

There are no exactly comparable cases in Igbo, since Yoruba has a true, freely-distributed M tone, whereas Igbo's M tone is a downstepped H. Still, we ought to be able to see the effect in Igbo, *mutatis mutandis*.

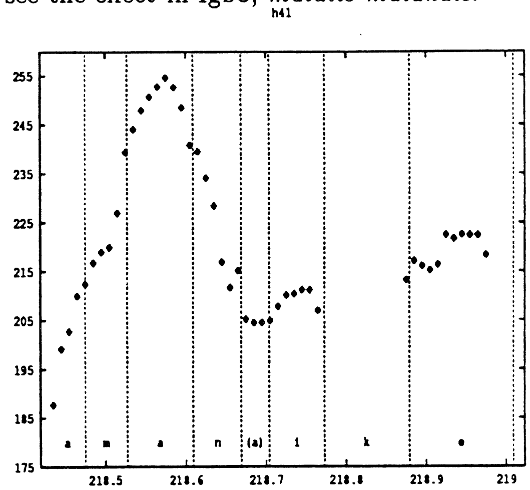

Figure 4: **Áma nà Íke** (far interlocutor)

Some examples seem to suggest that raising of H before L also holds in Igbo, as figure 4 clearly exemplifies. This figure shows the phrase "Áma nà Íke" (tone pattern HH L HH) spoken in the version with a distant interlocutor (and thus a high and broad pitch range). We can clearly see that the second syllable of "Áma" is much higher than the first.

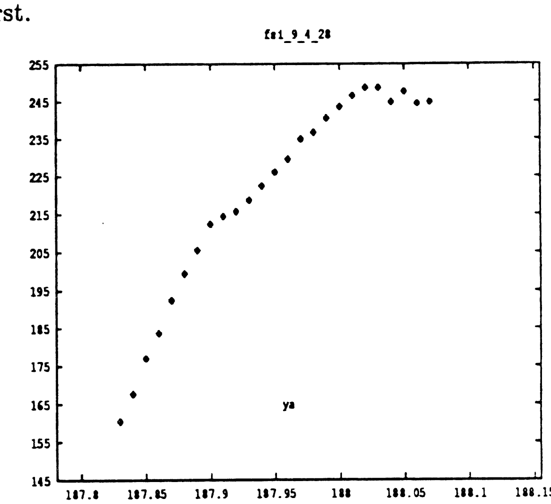

Figure 5: **yá** (FSI 9.4)

However, we are not convinced that this fact should be interpreted in terms of a rule of high-before-low raising. Essentially the same pattern for "Áma" would be possible in the absence of a following L tone. Indeed, the presence of more than one H syllable is by no means required. For instance, the F0 track (shown in figure 5) for the isolated monosyllable "yá" (*he*) shows a rising contour that is rather similar to the rising pattern seen on "Áma" in figure 4, and quite similar rising patterns could be seen in utterances that begin with three H syllables.

In general, the starting F0 for utterance-initial H-tone stretches in Igbo is variable, but almost always lower than the peak F0, which is usually not reached until near the end of the sequence of H-tone syllables. It is not clear what governs the variability in this matter, but stretches of different numbers of H syllables show qualitatively similar patterns, whether in isolation, before L, or before M.

Furthermore, this behavior seems to be a particular case of a much more general non-equivalence of F0 values realizing adjacent equal tones. Similar phenomena in Yoruba are extensively discussed by Laniran. In our much more limited examination of the Igbo case, these phenomena seem consistent with a system in which a block of contigous "same-toned" syllables (or other tone-bearing units) have just the same number and type of F0 targets that a single syllable in the same context would have. This view is a sort of phonetic version of the phonological principle known as the Obligatory Contour Principle, or OCP. The phonological OCP principle has the consequence that a stretch of segments with the same value of a phonological feature must represent a single feature spread over the whole sequence. Our phonetic version of this principle suggests that a tone spread over a stretch of tone-bearing units will only be interpreted phonetically one time, regardless of the nature and quantity of the units that it is associated with (abstracting away from coarticulatory effects that may arise when targets are crowded too close together, or physical interactions with other on-going articulations). This hypothesis contrasts with systems in which each tone-bearing unit in such a block is subject independently to F0 interpretation.

Laniran's treatment of Yoruba is intermediate between such a "phonetic OCP" system and a system in which each tone-bearing unit is given independent F0 interpretation. The raising of the last H in a sequence when an L tone follow is one key case in which additional targets seem to be required. This is not the place to examine this matter in any detail, but it seems to us that the Yoruba examples might be amenable to a treatment in which the H-tone targets in the sequences HL, HHL, HHHL, etc. have the same number and the same timing principles regardless of the count of H tones.

Note that we accept Laniran's argument that the H tone target is raised in Yoruba before L vs. M. However, our Igbo data provide quite a strong argument against the view that a similar process operates there. This argument arises from comparing the relationship of the two H tone targets in HLHL sequences and in HLHM sequences. If H were raised before L to anything like the same extent that Laniran suggests for the case of Yoruba, the relationships should be significantly different in the two cases.

Figure 6 presents a scatter plot of this relationship in our data. As this figure suggests, the two cases are not statistically distinguishable. This makes it rather unlikely that Igbo has a general rule of H raising before L, and suggests that the apparent existence of such an effect in examples like figure 4 must be explained in terms of other F0-interpretation principles.

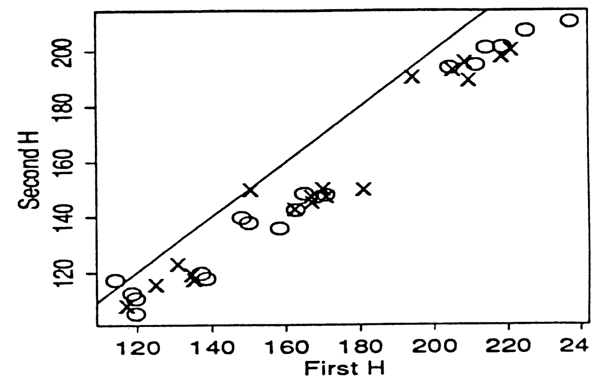

Figure 6: H/H Relationship in HLHL (O) vs. HLHM (X)

Applying this assumption, we will treat patterns such as HLH, HHLHH, and HLLHH in the same way for purposes of evaluating the scaling of downdrift, choosing the F0 minima and maxima (wherever they occur) as the points of reference.

4.2. Effect of downdrift on H and L

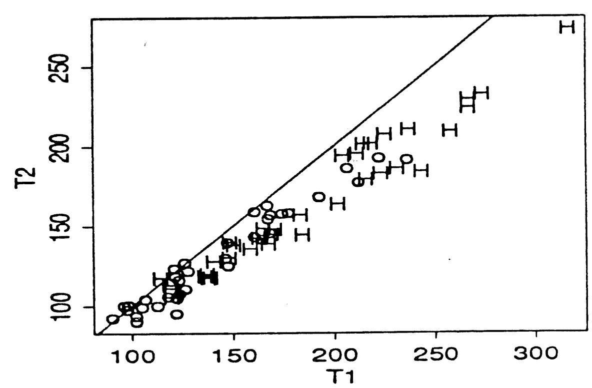

Figure 7: Downdrift on L (o) and H (H)

Does downdrift affect H and L tones in the same way? The answer appears to be "yes" in Igbo. Figure 7 shows successive L tones in an HLHL sequence are clearly lowered. Furthermore, the relationship between pairs of downdrifting H tones in an HLHL sequence (plotted with character "H"), and the relationship between pairs of downdrifting L tones further along in the same sequence (plotted with character "o"), appear to be the same. Naturally H tones are higher than L tones, other things equal, and so the L tone pairs tend to be drawn from a lower region of the distribution—but the distribution seems to be the pretty much the same for both kinds of tone pairs.

4.3. Downstep vs. downdrift

Is downstep (the lowering of M tones after H) the same as downdrift (the lowering of the second H in an HLH sequence)? Our data suggest that it is not. We take downdrift data from sentences 2, 3, 5, 6, 7, 8, 10, 12 and 13 of experiment 2, using only the first downdrift in sentence 8. We take downstep data from the HM words of experiment 1, and from sentences 1, 9, and 11 of experiment 2. As before, the F0 relations are remarkably homogeneous and well controlled: $r = .978$ for the downdrifts, and .980 for the downsteps. The slopes and intercepts derived from regression, and the standard errors of the estimates, are given below:

	Intercept	std. err.	Slope	std. err.	N
HLH:	7.65	2.74	0.824	0.014	162
HM:	-2.30	4.14	0.924	0.021	85

It is pretty clear that the slopes are different in the two cases, and indeed it seems that downdrift imposes roughly twice as much lowering as downstep does. Thus we can provisionally reject the hypothesis that (from a phonetic point of view) downstep is "the same thing" as downdrift.

This is a surprise, since there have been taken to be good typological, historical and phonological reasons to equate the two processes. What to do? We see three paths: to accept that the processes are distinct; to salvage their phonological equivalence by excusing their phonetic distinctness on some independent grounds; or to take the view that downdrift is two units of downstep. This last move strikes us as the most interesting one. To make its content clearer, we return to our simple-minded model $T + TRFD$. If the D parameter is to be used to model either downstep or downdrift, it will have to take on a sequence of successively lower values as we accumulate lowerings in a phrase like HLHLHLH or HMMMM. The obvious way to do this is to write the formula as $T+TRFD^N$, where N starts at 0 and increments by 1 for each unit of lowering. Then we might increment N in several different ways, among them:

Tones:	H	L	H	L	H	L	H	H	M
N.1:	0	1	1	2	2	3	3	0	1
N.2:	0	0	1	1	2	2	3	0	1
N.3:	0	1	2	3	4	5	6	0	1

The first two are traditional "downstep equals downdrift" theories, differing in whether the lowering occurs on the L tone or on the H tone. Phonologists who have explicitly considered the issue have generally opted for N.1 (e.g. [1, 7]). The last idea (labelled N.3) is an example of a way of counting that makes a downdrift worth two downsteps—it says, basically, that N counts the distinct (in the OCP sense) tones in the string. As far as we know, it has not previously been suggested.

Note that all three of these ways of counting N predict that H and L tones "see" downdrift in the same way. They differ in the predictions they make about the relationship between downdrift and downstep, and the specific relationship between H and L tones (although the last point can only be explored given independent constraints on the functional form of the model, and on the other model parameters). Production scaling experiments, like those discussed in this paper, offer a good opportunity to compare such hypotheses in a quantitative way. Space does not permit a consideration of this comparison here, but it represents an interesting example of how phonetic evidence can be brought to bear on what has been taken to be a phonological issue.

As we noted earlier, there are several different ideas in the literature about the nature of downstep (i.e. "mid" tone) in Igbo. Some authors treat downstep as a case of downdrift in which the L tone between two H tones is "floating" (i.e. unassociated with any segmental material), and as a result is not realized phonetically except by virtue of causing the H tone that follows to be lowered. Others argue against this treatment, and suggest that the "mid" tone is either a third phonological category, or else simply an independent H tone, which is interpreted phonetically at a lower pitch value than the H tone that precedes. This point of view (suggested notably in Clark ([1]) is represented graphically in figure 8:

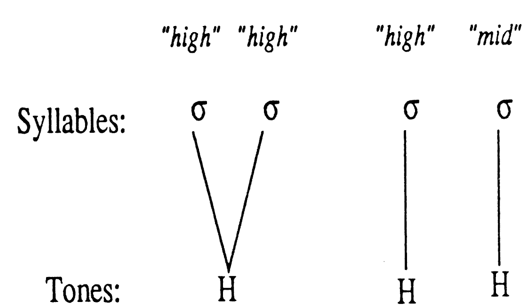

Figure 8: Clark's theory of the Igbo mid tone

Clark's arguments against the "floating low tone" theory of Igbo downstep are entirely phonological in character. She points to the regular appearance of a downstep at the boundary between cyclic domains when we would expect the tones on each side to be simply H; and also the regular disappearance of a downstep in environments that can be simply characterized in terms of a rule "delete the middle of three adjacent H tones." She argues that both these classes of phenomena require unmotivated complexities to state if downstep in Igbo is treated in terms of the presence of a floating L tone.

Clark is still able to maintain the identity of downstep and downdrift:[2] she expresses them as a rule of REGISTER LOWERING, which says:

> Lower the high (and low) pitch registers at the juncture between a high tone and any following tone within the same phrase.

Our evidence with repect to this question is purely phonetic, and somewhat indirect. First, we observe that downstep and downdrift do not lower an H tone by the same amount; thus the floating L tone theory, which treats downstep as downdrift of H tones across an invisible L, must provide some excuse for this difference in values. Second, we observe that L tones are downdrifted in exactly the same manner, quantitatively, as H tones. Finally, we observe that downstep lowers successive H tones by about half as much as downdrift does. All of this comes together nicely if each successive new tone, whether H or L, causes the whole tonal system to "deflate" by a fixed amount. Then both downstep and downdrift are just the tick of passing tones.

Our version of Clark's REGISTER LOWERING rule says something like

> Increment the exponent of the D parameter whenever a new tone occurs.

It is equivalent, but more palatable psychologically, to maintain the current D value by successive multiplications, and reset it (partly or completely) at phrasal boundaries. Note that this formulation of the "deflation" principle differs from Clark's, in that hers still maintains the phonetic equivalence of downstep and downdrift. However, our formulation is consistent with Clark's phonological analysis, and inconsistent with the floating-L-tone analysis. In that sense, our findings support her position.

There is a small remaining problem: is H lowered after a phrase-initial L? Emenanjo ([16]) says that it is not, and Hyman and Schuh ([17]) claim that a dialect difference exists on this point. Since the amount of lowering involved would not be very great, and pitch values are also affected by overall pitch range, vowel type, etc., it is not clear how to tell if lowering occurs. Note that according to our model, the predicted effect (the change in F0 value of the first H tone in a phrase due to the presence or absence of an preceding L tone) is about 20% of the first H tone's difference from the base of the H register. In conversational speech for a typical male speaker, this difference might about 20 or 30 Hz., so that the predicted effect will be in the range of 4 to 6 Hz, which could easily be missed.

We do not believe that our data provides any way to check this question. If it turns out that H is not "deflated" at all following initial L, then our rule would have to be modified to exclude this case.

A more fundamental question is: why are there no downstepped low tones? Perhaps the lowered final L tones should be analyzed as a kind of downstep, but even if this is true, our treatment of L and H is not at all parallel. The floating-low-tone account of downstep offers a reason for this asymmetry, but at too great a phonological and phonetic cost. In effect, our account (in common with Clark's) requires that high tones are exempt from OCP restrictions in certain cases, while low tones never are.

There are many possible treatments for this state of affairs, among them a metrical account such as that offered in [2], [18]: H tones might be exempt from OCP violations just in case they are the heads of separate tonal "feet." Our experimental evidence does not bear directly on such explanations, but just helps to pose more clearly the problem that they aim to solve.

5. Summary of Experimental Conclusions

From the intricate and well-controlled patterns that emerge when pitch range is varied against tonal sequence in Igbo, we derive a number of tentative conclusions.

First, the scaling of Igbo tones requires a model that is neither multiplicative nor additive; we can describe it by saying that increasing pitch range adds to a basic H or L tone value, in units that are proportional to the basic value of the tone type in question. This produces relationships among tone values in sequence that are qualitatively different from those that would arise in purely multiplicative or additive models. Second, Igbo L tones appear to be lowered in final position. Third, Igbo H tone does not appear to be raised before L (as opposed to before M). Fourth, Igbo downdrifted H and L tones seem to behave identically. Fifth, downstep and downdrift are quantitatively distinct: downdrift imposes a significantly greater degree of lowering. This difference is consistent

[2] The main reason to maintain this equivalence seems to be a typological one. Downstep and downdrift, both very common but not ubiquitous, apparently have an implicational relationship—two-tone languages with downstep always have downdrift, but not vice versa.

with treating downstep and downdrift as two symptoms of a process that deflates both high and low pitch registers by a constant amount every time a phonologically-distinct tone is encountered.

These conclusions must be considered tentative for a number of reasons. We have looked at data from only one speaker; the task represents only one (artificial) style of speech; there are other ways to describe and explain each bit of evidence we have presented, specifically alternative statements of the environments and alternative functional forms for modeling. Still, we feel that our approach reveals things about the tonal system that would not come out without pitch range variation. We try to maximize (rather than minimize) such variation, and then use the rich statistical structure of the resulting data to distinguish among alternative hypotheses about the nature of the underlying system.

6. Tone and Intonation

The phonetic realization of lexical tone, in Igbo and in general, deserves careful consideration on its own merits. Aside from the intrinsic interest of the phenomena, we may cite both the theoretical importance of the phonological questions that arise, and also the methodological benefits of treating the phonetic interpretation of F0, which is relatively easy to measure, and behaves in a delightfully lawful way in controlled experiments. In considering the difficult problems of intonational analysis in natural speech, studies of lexical tone take on an additional significance.

One of the central questions of intonational research is whether intonation has a phonology, and if so, how we can decide what its categories and relations should be. The phenomenon of *word constancy* means that lexical tones divide into fairly clear surface phonological categories, much like other distinctions in lexical phonology. Despite their many other disagreements, the general run of humanity (linguists among them) generally agree on what tokens count as instances of the same lexical type. In plain language, any dolt can recognize the word "dog" when he hears it. The categories of lexical phonology, whether tonal or not, inherit a considerable amount of this comforting constancy. We may not agree on the feature content of an English /g/ or an Igbo mid tone, but we usually know one when we hear it. By contrast, there is relatively little agreement (among ordinary folk and linguists alike) about what utterance tokens count as instances of the same intonational type. As a result, we are likely to disagree both on how a "high rise" should be analyzed, and also on when we have heard one. This may be a fact about language (namely that intonation is not based on the same sort of categories as lexical phonology), or a fact about consciousness (namely that intonational categories are not accessible to reflection, whereas words are—for some reason perhaps connected to the phenomenon of reference).

Despite these difficulties, many linguists have analyzed intonation in terms of categories just like those used for the analysis of lexical tone, although several extensive traditions exist that are largely or completely non-phonological (e.g. [19]). In developing and evaluating tonal theories of intonation, it is natural to turn to instrumental analysis to bolster our categorically-weak perceptions, and to hope that patterns of objective measurements will help validate the postulated distinctions. Lexical categories of tone provide a phonological anchor point for studies of the interplay of tone, structure, and rhetorical or stylistic modulation. Looking at such patterns in a language with lexical tone, we see what an intonational language might be like if its pitch contours were properly analyzed as the interpretation of similar phonological structures.

This approach is rendered more difficult by the fact that lexical tone languages are typologically diverse and individually complex. There are many open questions (and perhaps more yet to be asked) about their phonological structure. The phonetic interpretation of lexical tone in phrasal context is still mostly *terra incognita*, especially when the many sources of rhetorical and stylistic variation are considered.

Some of the liveliest theoretical problems in intonational studies today concern the nature of phrasal downtrends in F0, and the status of what seem to be "mid" F0 values. These are questions that many lexical tone languages also bring to the fore, as we have seen—although it must be kept in mind that the facts and their explanations are apparently somewhat varied across languages (cf. [20]).

Let us continue to use the term "downdrift" to describe successive lowering in sequences of H and L tones, and the term "downstep" describe the middle case in a situation in which there are three distinctive tone levels after H, but only two after L. Let us add the term "F0 decay" to describe a gradual decay in F0 in sequences of like tones, whether H or L, such as has been reported for Luo in [24]. Then from a simple descriptive point of view, tone languages with two basic tone levels may have downdrift, downstep and F0 decay (e.g. Luo); downdrift and downstep but not F0 decay (e.g. Igbo, Efik); downdrift only (e.g. Hausa); or none of these (e.g. Vai). Languages with more than two basic tone levels may have downdrift (e.g. Yoruba) or lack it (e.g. Nupe). Those languages with downdrift may suspend it in utterances with a certain pragmatic force, typically described as "yes/no question" (e.g. Hausa),[3] or they may not (e.g. Igbo). Downdrift-

[3] Apparently such suspension in languages with both downdrift

suspension languages may also have the addition of a final high tone (Hausa), a final low tone (e.g. Kolokuma Ijọ), or no extra final tone (e.g. Efik) to phrases in which downdrift is suspended. Vai, which has no downdrift, also adds a final high tone in yes/no questions.

The treatment of such phenomena in lexical tone languages has been mirrored in various treatments of certain phenomena in English. We might mention two cases:

- the treatment of "stepping" contours in which a succession of accented syllables are realized on successively lower pitch values, with or without intervening dips;
- the treatment of the final mid-level F0 target in the contours variously called "vocative," "warning/calling", "stylized," etc.

The stepping contours have often been treated as cases of downdrift; their presence or absence might arise by using the "switch" that lets certain tone languages suspend downdrift to express (some pragmatic signal described as) a yes/no question. Here the common existence of downdrift in two-tone languages, and its suspendability in some of them, give us license to treat somewhat similar intonational phenomena in an analogous way. However, English then seems to require a rather fine granularity of downdrift suspension—in some proposals, every pitch accent is marked for presence or absence of a catathesis (i.e. downdrift) feature. This sort of freedom has never been reported for any lexical tone language—downdrift (if present) is automatic in sequences of H and L tones, and can apparently be suspended only on a phrase-by-phrase basis, whereas downstep applies only to H tones following other H tones. This does not prove the proposed descriptions of English are wrong, but it does remove the "blessing" of similarity to tone language phenomena.

What about the final mid-level F0 target in English "vocative" contours? The fact that it sounds something like an Igbo "mid" tone helped to license by analogy its treatment in [21] as a high tone downstepped by a floating low tone (not otherwise realized) associated with the previous pitch accent. If it were true that downstep in all terrace-tone languages is always an expression of floating low tones, then such an analysis for English would gain plausibility. If data in other terrace-tone languages turn out, like Igbo, to disconfirm the floating-low-tone theory of downstep, this treatment loses plausibility. The final vocative target in English might still be an Igbo-like downstepped high, but to maintain the analogy, any H phrase accent should be downstepped after any pitch accent ending in H. The only way to avoid this would be to analyze the donwstepped H phrasal tones as independent, while non-downstepped H phrasal tones (e.g. those in rising or high level contours) are just spread from the tone in the fore-going pitch accent. While not to be dismissed out of hand, such an analysis seems odd. Basically, English HM contours (such as the vocative) are just not distributionally similar to the similar-sounding sequences in Igbo.

If (as discussed under the heading of downdrift) we add a catathesis feature to every pitch accent (or even every tone) in English, we make it easy to treat final target in the vocative contour as a downstepped H—but the price is a high one. This move is informationally equivalent to doubling the language's tonal inventory, and will often be redundant with independently-needed marking of emphasis and pitch-range changes.

As an alternative, we might observe that English vocative contours also sound somewhat like H M sequences in languages like Yoruba, where the mid tone is a genuine independent category. Thus one might attempt an analysis of the vocative contour using an independent mid tone category. The fact that tone languages can easily have three or four paradigmatically distinct categories of tonal targets helps license this line of investigation. Of course, introducing a new tonal distinction raises the question of its distribution—if it occurs freely, then the multiplicity of resulting distinctions must be motivated. If its distribution is restricted to those cases where we feel we need it, then we need to explain this.

Of course, it is quite possible that the English vocative contour is neither like HM in Igbo nor like HM in Yoruba, since there are plenty of other models available among lexical tone languages. In analyzing English, we must deal one way or another with the fact (noted in [25]) that final falling contours seem to fall into at least three classes:

- ordinary terminal falls, where the endpoint is at the bottom of the speaker's pitch range;
- non-terminal falls, where the endpoint is somewhat above the bottom of the speaker's pitch range;
- vocative contours, where the endpoint is lower than the peak, but seems higher than either of the other two cases.

We could deal with these observations in many ways, using one, two or three phonological categories, with a wide choice of intensional and extensional definitions in each

and downstep affects only downdrift, and leaves downstep in place (e.g. in Efik). If this is true, then it is a bit of a problem for theories under which downstep and downdrift are the same thing. However, no such case has been modeled quantitatively as far as we are aware, and it is often reported in such cases that the pitch range as a whole is raised and narrowed, so it seems wise to reserve judgment.

case. We could treat all three cases as gradient scaling of a single phonological category; we could call the vocative target an Igbo-like downstepped H (or a Yoruba-like M), and distinguish the other two cases as gradient scaling of a single phonological category; we could call the vocative and non-terminal targets M, and distinguish them gradiently (as in [26]); we could treat all three cases as different phonological categories of fall (as in [25]); we could treat the non-terminal fall as a phonetic variant of fall-rise; and so on. All of these treatments (and more) might be licensed by appeal to the facts of some lexical tone languages.

No doubt tone languages will continue to be a source of inspiration to students of intonation. The careful comparative study of languages with lexical tone should give us a sense of what phonological and phonetic resources are available for the analysis of intonational phenomena, whether intonation contours turn out to be homologies or merely analogies of lexical tone contours. However, at the present stage of development of our knowledge, we should be careful not to accept superficial analogies too easily, whether from tone to intonation or from one tone language to another. Interest in universal principles needs to be tempered by respect for the facts of each language, and by willingness to recognize that the number of carefully-modeled tone systems is small compared to the apparent diversity of human potential in this area.

References

1. Clark, M. M., *The Tonal System of Igbo*, Foris Publications, Dordrecht, 1990.
2. Manfredi, V., *Agbo and Ehugbo: Igbo Linguistic Consciousness, its Origins and Limits*, Harvard University PhD Thesis, 1991.
3. Liberman, M. Y., Schultz, J. M., Hong, S., and Okeke, V., "The Phonetics of Igbo Tone," pp. 743-746, ICSLP 92, Alberta, 1992.
4. Liberman, M. Y., and Pierrehumbert, J. B., "A Metric for the Height of Certain Pitch Peaks in English," JASA 66 (S1), 1979, S130.
5. Liberman, M. Y., and Pierrehumbert, J. B., "Intonational Invariance under Changes in Pitch Range and Length," pp. 157-234 in M. Aronoff and R. Oehrle, eds., *Language Sound Structure*, MIT Press, Cambridge, 1984.
6. Carrell, P. L., *A Transformational Grammar of Igbo*, West African Language Monographs, no. 8, The University Press, Cambridge, 1970.
7. Goldsmith, J., *Autosegmental Phonology*, MIT PhD Thesis, 1976.
8. Clements, G. N., and Ford, K. C., "Kukuyu Tone Shift and its Synchronic Consequences," *Linguistic Inquiry* 10:179-210, 1979.
9. Pulleyblank, D., *Tone in Lexical Phonology*, Reidel, Dordrecht, 1986.
10. Welmers, W.E., and Welmers, B. F., *Igbo: a Learner's Dictionary*. UCLA, Los Angeles, 1968.
11. Nwachukwu, P. A., *Towards an Igbo Literary Standard*, Kegan Paul International, London, 1983.
12. Stewart, J. M., "Niger-Congo, Kwa," pp. 179-212 in Thomas Sebeok, ed. *Current Trends in Linguistics*, v. 7, Mouton, The Hague, 1971.
13. Press, W. H., Flannery, B.P., Teukolsky, S. A., Vetterling, W. T., *Numerical Recipes*, Cambridge University Press, 1988.
14. Connell, B., Ladd, D. R., "Aspects of Pitch Realization in Yoruba," pp. 1-30, Phonology 7, Cambridge University Press, 1990.
15. Laniran, Y., *Intonation in Tone Languages: The Phonetic Implementation of Tones in Yorùbá*, Cornell University PhD Thesis, 1992.
16. Emenanjọ, E. N., *Elements of Modern Igbo Grammar*, Oxford University Press, Ibadan, 1978.
17. Hyman, L. M., and Schuh, R. G., "Universals of Tone Rules: Evidence from West Africa." *Linguistic Inquiry* 5:81-115, 1974.
18. Manfredi, V., "The Limits of Downstep in Agbo Sentence-Prosody," IRCS Workshop on Prosody in Natural Speech, August 1992.
19. 't Hart, J., Collier, R., and Cohen, A., *A Perceptual Study of Intonation: an Experimental-Phonetic Approach to Speech Melody*, Cambridge University Press, 1990.
20. Dunstan, E., ed., *Twelve Nigerian Languages*, Africana Publishing Corporation, New York, 1969.
21. Pierrehumbert, J. B., *The Phonology and Phonetics of English Intonation*, M.I.T. PhD Thesis, 1980.
22. Welmers, W. E., *A Grammar of Vai*, University of California Publications v. 84, University of California Press, 1976.
23. Elimelech, B., *A Tonal Grammar of Etsakọ*, University of California Publications v.87, University of California Press, 1978.
24. Tucker, A. N., and Creider, C. A., "Downdrift and Downstep in Luo," pp. 125-134 in R. K. Herbert, Ed., *Proceedings of the Sixth Conference on African Linguistics*, OSU Working Papers in Linguistics no. 20, 1975.
25. Pike, K., *The Intonation of American English*, University of Michigan Press, 1945.
26. McLemore, C. A., "Prosodic Variation across Discourse Types", *Proceedings of 1992 IRCS Prosody Workshop*, IRCS Technical Report, University of Pennsylvania, 1992.