# TABLE OF CONTENTS

Preface ........................................ v

John J. Ohala, THE PHYSIOLOGY OF TONE ........ 1

Wayne A. Lea, SEGMENTAL AND SUPRASEGMENTAL INFLUENCES ON FUNDAMENTAL FREQUENCY CONTOURS .......... 15

James A. Matisoff, TONOGENESIS IN SOUTHEAST ASIA ........ 71

La Raw Maran, ON BECOMING A TONE LANGUAGE: A TIBETO-BURMAN MODEL OF TONOGENESIS .......... 97

William R. Leben, THE ROLE OF TONE IN SEGMENTAL PHONOLOGY ........................................... 115

Larry M. Hyman, THE ROLE OF CONSONANT TYPES IN NATURAL TONAL ASSIMILATIONS .......... 151
PREFACE

The six papers collected in this volume were presented at a Symposium on Consonant Types and Tone which took place at the University of Southern California on March 9-10, 1973. The purpose of this symposium was to bring together scholars whose previous research had focused on the effect consonants exert on pitch/tone. This topic was chosen because I sensed that there was considerable interest in it among different kinds of linguists who might wish to gather and compare notes with one another.

Initially my modest hope was to assemble a group of California linguists. However, much to my delight, a number of out of state people expressed interest in the symposium. The list of invited participants includes the following: Karen Courtenay (U.C.L.A.), Victoria A. Fromkin (U.C.L.A.), Mary R. Haas (U.C. Berkeley), Kong-On Kim (U.S.C.), Paul L. Kirk (C.S.U. Northridge), Peter Ladefoged (U.C.L.A.), Jerry Larson (Texas), Wayne A. Lea (Univac), William R. Leben (Stanford), Ian Maddieson (Indiana), La Raw Maran (Indiana), James A. Matisoff (U.C. Berkeley), Burckhard Mohr (C.S.C. Dominguez Hills), John J. Ohala (U.C. Berkeley), Manjari Ohala (U.C. Berkeley), Sven Öhman (U.C.L.A.), Beatrice T. Oshika (Speech Comm. Res. Lab.), Kenneth L. Pike (Michigan), Edward T. Purcell (U.S.C.), Russell G. Schuh (U.C.L.A.) and William S.-Y. Wang (U.C. Berkeley). A number of other guests participated in the discussion of one or more papers.

As the list of participants and the titles of the papers indicate, the question of consonant types and tone was approached from a number of directions. Phoneticians, phonologists and historical linguists, representatives of all of the major geographic areas of tone languages (Asian, Amerindian, African, European), were on hand to share the findings of their research.
Phonetically oriented, the first two papers explore the physiological and acoustic bases of pitch and their relevance to the study of tone languages. John J. Ohala's paper "The Physiology of Tone" discusses the various physiological parameters involved in controlling fundamental frequency. Ohala evaluates the various theories which have been proposed to explain why certain consonants raise or lower fundamental frequency, including vocal cord tension, aerodynamic factors, and larynx elevation. Wayne A. Lea's paper "Segmental and Suprasegmental Influences on Fundamental Frequency Contours" takes a close look at the effect of segments (consonants and vowels) and grammatical features (stress, intonation, syntactic boundaries) on fundamental frequency. Of direct concern is his report of acoustic experiments showing a raising or lowering effect of consonants on the fundamental frequency of a following vowel in English.

The second two papers address themselves to the genesis of tone in language. James A. Matisoff's paper "Tonogenesis in Southeast Asia" provides a comprehensive introduction to the birth of tone in a number of Oriental languages. Earlier consonantal contrasts have given rise to modern-day tonal contrasts, with glottal consonants affecting preceding vowels, and voiced vs. voiceless consonants affecting following vowels. La Raw Maran's paper "On Becoming a Tone Language" presents a model of tonogenesis in Tibeto-Burman. How do automatic pitch variations controlled by consonants become part of the phonology—"cognitivized" in Maran's terminology? The importance of morphology in the historical development of tone in Tibeto-Burman is forcefully argued.

The last two papers investigate the phonological parameters of consonant types and tone. William R. Leben's paper "The Role of Tone in Segmental Phonology" situates the question of consonants and tone within the general theoretical context of underlying tonal representation. Is tone segmental (e.g. a feature on vowels) or suprasegmental (e.g. a feature on morphemes and words)? After providing examples of both kinds of tonal representation, Leben suggests that the
role of consonants in tone rules can be accounted for in his framework of "melodies" and tonal mapping processes. In my paper, "The Role of Consonant Types in Natural Tonal Assimilations", I first attempt to define the notion of "natural tone rule", illustrating widely from African languages. I then provide evidence that certain kinds of consonants can exert a blocking effect in natural tone rules, while other consonants are neutral with respect to tone.

Unfortunately, two papers which were presented at the symposium could not be included in this volume. Victoria A. Fromkin's paper "The Distinctive Features of Tone" surveyed the various proposed binary distinctive feature systems for tone, their adequacy in capturing tonal processes and the relationship between consonants and tone. (For a discussion of these issues, see Fromkin's 1972 article "Tone Features and Tone Rules" in Studies in African Linguistics 3.47-76.) Jerry Larson's paper "Issues and Non-Issues in Tonology" dealt with many of the theoretical issues being discussed by tonologists, in particular, the nature of "pitch-accent" languages and their questionable distinctness from "stress" languages. (The contents of Larson's paper will be incorporated in his forthcoming University of Texas doctoral dissertation.)

There are a number of individuals I would like to thank for their assistance in organizing this symposium. First of all, I would like to thank Dean Charles G. Mayo of the Graduate School, U.S.C., for the support and encouragement he provided. Also, I would like to thank Edward Finegan, Chairman of the Linguistics Program, U.S.C., for his advice and aid. A special thanks is extended to the Los Angeles hosts, especially Peter and Jenny Ladefoged, who provided hospitality to the out-of-towners. Finally, my thanks to all the participants for making this symposium possible.

Larry M. Hyman
THE PHYSIOLOGY OF TONE

BY

JOHN J. OHALA

DEPARTMENT OF LINGUISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720

Published by the Linguistics Program
University of Southern California
Los Angeles, California 90007
O. INTRODUCTION

The sound patterns of tone, as described in the phonological literature—much of it very recent (Bien-Ming 1931, Haudricourt 1954, 1961, Wang 1967, Greenberg 1970, Chang 1972, Hyman and Schuh 1972, Cheng 1973)—exhibit many interesting asymmetries which are determined by phonetic context, for example:

1. Historically, many consonants seem to have induced tonal differences on the vowel following, but few consonants, primarily only [h] and [ʔ], seem to have caused tonal changes on the vowel preceding.*

2. Specifically, [ŋ], voiced or breathy-voiced obstruents seem to depress tone on the vowel following, whereas voiceless, particularly voiceless aspirated obstruents, [h], and implosives seem to raise tone on the following vowel. There is also some evidence that [ʔ] and [h] may raise and lower, respectively, the tone of the preceding vowel (cf. Matisoff 1973).*

3. Falling tones outnumber rising tones.

4. Sequences of low tone—high tone often change so as to reduce the pitch interval between the two tones, but this happens less often with sequences of high tone—low tone (cf. Hyman 1973).

If further research supports the claim of universality for these and the many other tendencies of tonal behavior discussed in this volume, then it is clear that these patterns can best be explained by reference to universal characteristics of the human speech production and speech recognition systems.

The primary emphasis of this paper is on the physiological mechanisms involved in the production of tone; the perception of tone is not covered extensively not because it is less important, but because less is known about it (however, see Fry 1960, Hadding-Koch and Studdert-Kennedy 1964, Lehiste 1970, van Katwijk and Govaert 1967, Klatt 1973).
1. PHYSICAL CORRELATES OF TONE

Although the primary physical correlates of what has been labelled "tone" are the level, direction, and/or rate of change of the "pitch" or fundamental frequency ($F_0$) of vibration of the vocal cords, secondary correlates may in some cases be the mode of vibration (lax, breathy, tense, creaky voice), vowel duration and quality, and manner of termination of the vowel bearing the tone, i.e. with vs. without a glottal stop (Pike 1948, Hinton 1970, Matisoff 1973). In this paper I will be primarily concerned with the physiological mechanisms that affect $F_0$.

2. DETERMINANTS OF $F_0$

The $F_0$ of voice is determined basically by two partially independent factors: (a) the state of the vocal cords and (b) the aerodynamic forces driving the vocal cords (van den Berg 1958, Ishizaka and Flanagan 1972). The state of the vocal cords includes, for example, the anterior-posterior and vertical tension of the cords and the degree of approximation of both the ligamental (the anterior 2/3) and arytenoidal (the posterior 1/3) of the vocal cords. The aerodynamic driving forces are the air pressure difference ("pressure drop") across the vocal cords and the rate of air flow through the glottis.

Both of these parameters can be varied in an active, intentional way, and in a passive, fortuitous way—fortuitous in the sense that when a speaker intentionally produces some other speech gesture, these parameters may change and thus fortuitously affect $F_0$. (This is identical to Wang and Fillmore's (1961) use of the terms "extrinsic" and "intrinsic", respectively, with respect to phonetic variation.)

2.1. Active Variation of $F_0$ by Adjustment of Vocal Cord Tension

Active, intentional changes of vocal cord tension or glottal configuration are accomplished primarily by the muscles attached to the larynx, principally the cricothyroid muscles, but also by all the adductor muscles of the larynx: the lateral cricoarytenoïd, the vocalis, and the interarytenoid muscles (Ohala 1970,
Sawashima 1970). The extrinsic laryngeal muscles, chiefly the so-called "strap" muscles of the neck, which attach either directly to the larynx (thyrohyoid and sternothyroid muscles) or indirectly via attachment to the hyoid bone, which, in turn, is connected to the larynx via ligamental attachments (sternohyoid muscles), are also active in pitch regulation, especially in lowering pitch (Ohala and Hirose 1969, Ohala 1970, Ohala 1972, Erickson and Abramson 1972). The involvement of the strap muscles in regulating pitch is obviously related to the common observation that larynx height correlates closely with pitch (Ohala 1973, Ohala 1972, Ohala and Ewan 1973), although why this happens is still disputed. Nevertheless, I view this as evidence for the existence of active laryngeal mechanisms for both the raising of pitch and the lowering of pitch. These mechanisms are the primary ones utilized for the production of intended linguistically significant pitch changes, that is, the large pitch variations characteristic of all tonal and intonational contours. There are other views on this matter, however (Lieberman 1967, 1970).

2.2. Active Variation of Pitch by Modification of Aerodynamic Factors

The pressure drop across the vocal cords can be actively increased or decreased by varying the expiratory force. Liberman (1967) claimed that this is the primary method speakers use to increase pitch on stressed, especially emphatically stressed syllables, and to let pitch fall at the end of sentences. There are serious problems with this theory (Vanderslice 1967, Kim 1968, Ohala and Ladefoged 1969, Ohala 1970, MacNeillage 1972). Electromyographic (EMG) and other evidence points instead to the activity of the laryngeal muscles being the primary force behind these and all other major linguistic pitch variations. Moreover, recently Lieberman seems to have abandoned many of the more extravagant aspects of his theory (Lieberman, Sawashima, Harris and Gay 1970).

However, Ladefoged (1967, 1971) notes that an extra expiratory pulse (which can be detected as a momentary increase in the EMG of the internal intercostal muscles) does occur during some stressed syllables and there is a momentary increase in subglottal air pressure ($P_s$) during stressed syllables (which all investiga-
tors of $P_s$ have noticed). No doubt a good part of this momentary rise in $P_s$ is due to this expiratory pulse, but part of it may be due to a momentary increase in glottal resistance which would result from the vocal cord adjustment for increased pitch and intensity. At any rate, these $P_s$ rises accompanying stress cannot account for most of the pitch changes observed during stress: the effect of $P_s$ changes on pitch has been found to be about 2.5 Hz/cm H$_2$O (Ohala and Ladefoged 1969, Hixon, Mead and Klatt 1971), so the $P_s$ rise of 1 to 5 cm H$_2$O usually found during emphatically stressed syllables could account for pitch rises of at most some 13 Hz—far less than the 50 to 100 Hz pitch changes commonly encountered on these syllables. And needless to say, the increase in $P_s$ cannot account for any pitch drops which are commonly used to manifest stress (Bolinger 1958).

It would appear, then, that the extent to which pitch is actively regulated by variations of the expiratory force is negligible.

The aerodynamic driving force can also be modified by action of the larynx: other things being equal, the volume of air flow through the glottis will vary inversely with the glottal air resistance and the pressure drop across the vocal cords will vary inversely with the volume of air flow through the glottis. But, again, the effect on pitch of this laryngeally-induced change of the aerodynamic driving force is very likely negligible for the reasons mentioned above.

2.3. Passive Variation of Pitch by Changes in Vocal Cord Tension

2.3.1. Pull on larynx by tongue. Apparently a high position of the tongue creates a slight pull on the larynx which is translated into increased vocal cord tension. This results in the widely-noted slightly higher average pitch for high vowels [i, u] and slightly lower pitch for low vowels [ə, a, ɔ], with mid vowels having a pitch intermediate between these. This explanation is disputed, however (Atkinson 1973, but see review by Ohala 1973).

For a possible (rare) instance of this effect leading to tone alternations determined by vowel height, see Wang (1970) or Mohr (1971).
This would also predict that consonants involving a high position for the body of the tongue, i.e. palatals and velars, ought also to induce a slightly higher average pitch (as opposed to labials and dentals, which do not involve the body of the tongue in the same way). The available data on this is mixed: House and Fairbanks (1953) and Mohr (1971) do show average pitch to be higher in the environment of English [g] than it is in the environment of English [b] or [d]; however the data of Lehiste and Peterson (1961) and Lea (1972) do not.

2.3.2. Vocal cord state and voicing. Halle and Stevens (1971) suggest that the vocal cords are stiff during the production of voiceless obstruents and are slack during voiced obstruents, and they provide a model of vocal cord vibration that shows why these differences are required. This, they say, accounts for the pitch variations accompanying voiced and voiceless obstruents. However, as far as I know there has as yet been no experimental verification of their model. Recent electromyographic recordings of the intrinsic laryngeal muscles (Hirose, Lisker and Abramson 1973) revealed no obvious differences in the tension of the laryngeal muscles during the production of the voiced/voiceless distinction--other than what would be expected for the abduction of the cords during the voiceless obstruents. A crucial point in Halle and Stevens' argument is their claim that the pitch of vowels preceding obstruents also show the same kind of perturbation as do the vowels following obstruents, i.e. pitch raised slightly for voiceless obstruents; lowered slightly for voiced obstruents. But I know of no hard evidence in support of this; in fact, just the opposite: pitch usually falls immediately at the beginning of any obstruent, voiced or voiceless (Lea 1972, 1973).

2.3.3. Larynx elevation and voicing. La Raw Maran (1971) posits a raised larynx for voiceless obstruents and a lowered larynx for voiced obstruents. There is some support for this in the older literature (see Ohala 1973) and in the recent work of Ewan and Krones (1973), and given that, other things being equal, larynx height is correlated with pitch, one might expect this larynx position for consonants to affect pitch on surrounding vowels.
But perhaps "other things" are not equal in these cases because Ewan and Krones did not find a consistent relation between pitch and larynx height over the small range of larynx displacements accompanying voiced and voiceless consonants. Moreover, like the Halle and Stevens hypothesis, this would also predict that consonants would perturb pitch on both preceding and following vowels, but, as noted above, the perturbation on the preceding vowels is not the same as that on the following vowels. Thus this hypothesis certainly requires further study but there is at present no overwhelming data leading us to believe larynx height causes the pitch variations associated with voiced/voiceless obstruents.

2.3.4. **Voice quality and vocal cord tension.** Ohala and Ohala (1972a,b) have speculated that the lowered pitch accompanying Hindi breathy-voiced stops could be due to a lessening of the tension of the vocal cords primarily intended to achieve lowered glottal resistance in order to produce the rapid air flow through the glottis and consequently the noisy quality of breathy-voiced phonation. However, there is no evidence for this as yet.

2.4. **Passive Variations of Pitch by Changes in Aerodynamic Conditions**

During voiced obstruents the decrease in the pressure drop across the vocal cords (due to the build up of air pressure in the oral cavity) causes the transglottal air flow and consequently the fundamental frequency of phonation to lower (cf. Ladefoged 1967:47ff). Thus, upon release of a voiced obstruent the pitch is initially very low and may gradually return to its "normal" level. Upon release of a voiceless aspirated consonant, however, the rate of air flow is initially very high, since there is momentarily little resistance to the air flow at the open glottis or in the oral cavity. Thus when the vocal cords do adduct for voicing they meet a very high rate of air flow and consequently vibrate at an initially high rate, gradually returning to their "normal" rate of vibration.

It is interesting that pitch remains perturbed for a relatively long time after the consonant release—for as much as 100
ms after the onset of voicing (Lea 1972, Löfqvist 1973). This may be due to some kind of oscillatory momentum. Or, it is also possible that the rate of transglottal air flow characteristic of the obstruent persists for a rather long time after the stop is released. Thus air flow records published by Subtelny, Kho, McCormack and Subtelny (1969) show higher-than-normal air flow persisting more than 300 ms into the vowel following a voiceless aspirated stop. Similar records but with the high air flow lasting only about 50-60 ms into the vowel have been published by Klatt, Stevens and Mead (1968) and Frøkjær-Jensen, Ludvigsen and Rischel (1971).

Implosives must cause elevated pitch by the high rate of air flow through the glottis generated by the rapidly descending larynx (Ladefoged, personal communication).

Breathy-voiced stops such as Hindi [h̥, d̥] etc. and voiced "h", [h], are complex cases. Upon release of the breathy-voiced stops and throughout the [h] the air flow is very high and thus one might expect them to raise pitch, yet they depress pitch. The reasons for this may be one or more of the following:

1. The vocal cords have to be rather closely adducted for the high rate of air flow to cause increased pitch (this is reasonable, given what we know of how the Bernoulli effect works; cf. Ladefoged 1973), and as the vocal cords are somewhat abducted during these sounds, the high air flow has little effect on pitch.

2. The vocal cord tension may be lowered in the process of abducting the vocal cords, particularly by the lateral cricoarytenoid muscles, known to play a secondary role in pitch regulation. A brief period of inhibition of the lateral cricoarytenoid muscle during [h] is evident in the EMG records presented by Ohala (1970:72).

3. Most of the air flowing through the glottis during these sounds is escaping through the arytenoidal portion, not the ligamental portion of the vocal cords (Ladefoged 1973). Thus, although the average transglottal air flow may be high during breathy-voice, the air flow through
the ligamental portion may be lower than normal. As it is only or mainly the ligamental portion of the vocal cords which is vibrating in this case this lower-than-normal air flow may cause the lowered pitch.

3. SPEED OF PITCH CHANGE

Ohala and Ewan (1973) found that for a given pitch interval subjects could execute a falling pitch faster than a rising pitch. They tentatively concluded that raising pitch may involve more "effort". This may account for (a) the aspect of "downshift" which involves the lowering of successive high tones (Hyman and Schuh 1972, Hyman 1973), (b) the higher incidence of falling tones over rising tones in Chinese (Cheng 1973), and (c) the asymmetrical behavior of sequences of low-high tones vs. high-low tones (Hyman and Schuh 1972, Hyman 1973).

4. SYNCHRONIZATION OF TONE (ACCENT) AND SEGMENTS

There is evidence that at least in Swedish, and possibly Dutch, and very likely, English, pitch contours and segments need not be precisely synchronized, at least not as closely synchronized as the oral articulators are with each other (Öhman 1965, van Katwijk and Govaert 1967, Löfqvist 1973).

5. PERCEPTUAL RELEVANCE OF TONAL DIFFERENCES TO CONSONANT IDENTIFICATION

Although the pitch perturbations induced by different types of consonants are typically quite small, there is ample evidence that in languages not considered to be tonal (e.g. English and Russian), these pitch changes do have value as perceptual cues for the identification of the voicing of stops (Chistovich 1968, Haggard, Ambler and Callow 1970, Fujimura 1971).

If these supposedly small fortuitous pitch contours following consonants can be used as perceptual cues by listeners, it is a small step beyond that to suppose that eventually these small pitch differences might be taken by listeners as the major acoustic cue differentiating the lexical items formerly differ-
entiated by voicing or voice onset time. Thereafter the pitch contours after the stops would be greater and would be intentionally produced. This is undoubtedly what happened over the centuries in those cases where voicing contrasts have been lost and tonal differences on the following vowel have been left in their place (cf. Maran 1973, Matisoff 1973). This is the mechanism called "rephonologization" by Jakobson (1931).

NOTE

*A somewhat puzzling exception to this generalization may be the case of the Punjabi high tones which appear on vowels which were followed by [h] or breathy voiced stops in Middle Indo-Aryan (Arun 1961, Gill and Gleason 1969).

REFERENCES


SEGMENTAL AND SUPRASEGMENTAL INFLUENCES ON FUNDAMENTAL FREQUENCY CONTOURS

BY

WAYNE A. LEA

DEFENSE SYSTEMS DIVISION
UNIVAC PARK
ST. PAUL, MINNESOTA 55165

Published by the Linguistics Program
University of Southern California
Los Angeles, California 90007
1. INTRODUCTION

The primary physical correlate of tonal phenomena is the rate of vibration of a speaker's vocal cords, or his "voice fundamental frequency" (abbreviated $F_0$). However, voice fundamental frequency also conveys information about aspects of language structure other than tonal contrasts within words. In this paper, we consider the influences of sentence intonation, stress, and phonetic sequences on fundamental frequency values at various times during English utterances. Special attention will be directed toward the influences of vowels and consonants, and transitions between them. How stress interacts with the effects of phonetic sequences will also be discussed.

Four types of experiments have been performed by this author to establish the specific influences of sound sequences, stress, phrase structure, and sentence intonation on fundamental frequency values during various isolated words and connected discourses. Some of these influences are physiologically determined, and thus presumably language-universal, while others are learned aspects of the sound structure of specific languages. All such influences may be pertinent to explaining such tonal phenomena as the stair-cased falling of high tones throughout a long utterance, the most common consonantal contexts accompanying falling and rising tones, or the simple effects of unvoiced and voiced consonants on the likelihood of high versus low tones.

In section 2, investigations of the influences of sentence intonation and stress on $F_0$ contours are described. Fundamental frequency is shown to increase substantially at the first stressed syllables of most major constituents, so that both the constituent division of a sentence and the locations of many stressed syllables can be established from acoustic data. Other stressed syllables within constituents are also shown to be associated with inflections in $F_0$ contours.

Segmental influences interfere with the determination of constituent structure and stress from $F_0$ contours. Different consonant types such as unvoiced versus voiced obstruents are manifested not simply by the presence or absence of interruptions in the per-
iodic vibration of the vocal cords, but also by local variations in $F_0$ values during the consonants and within transitions into or from surrounding vowels. These effects were studied in detail in experiments on 326 different CVC (consonant-vowel-consonant) structures, as discussed in section 3. Primary among the conclusions from such studies are the general trends toward high and falling $F_0$ after unvoiced consonants and lower but rising $F_0$ after voiced obstruents. These effects, which are physiologically determined, may influence the relative frequencies of high versus low, and falling versus rising, tones for various consonantal contexts in tone languages.

These studies of $F_0$ contours in CVC sequences were conducted for the stressed syllables of [heCVC] nonsense utterances recorded by two native English talkers. To extend studies to other possible stress patterns, sixty pairs of English words with contrasting stress patterns were recorded by the same two talkers. The resulting interactions found between stress, phonetic sequences, and $F_0$ contours are discussed in section 4.

Conclusions and implications for continued study of tone languages are given in section 5.

2. INFLUENCES OF INTONATION AND STRESS ON $F_0$ CONTOURS

Voice fundamental frequency is known to be a close physical correlate of the percepts of intonation and stress. In this section we consider two series of experiments that demonstrate how $F_0$ changes can signal the presence of syntactic boundaries and stressed syllables. Individual vowels and consonants are also found to interfere so prominently with such suprasegmental aspects that segmental influences must also be investigated in detail.

2.1. Effects of Sentence Type and Constituent Structure on $F_0$

For decades, linguists have claimed that intonation indicates the immediate constituent structure of English sentences (Jones 1909; 1932; Pike 1945; Wells 1947). Trager and Smith, whose pitch and stress "levels" are widely used, claimed that monitoring $F_0$ makes it possible to have "solidly established objective procedures" for "the recognition of immediate constituents and parts of speech.
syntax (1951:77). Gleason (1961:169) considered intonation and
stress as "the dominant elements in the syntax-signaling system".
Study of metrical patterns in English verse also indicated strong
markings of syntactic boundaries by the prosodic features (Keyser
1969).

Transformational linguists have also recognized this syntax-
signaling role of intonation. Stockwell (1960) noted that "inton-
ation patterns are the absolutely minimal differentiators of numer-
ous utterance tokens". Lieberman (1967:314) asserted that "Inton-
ation has a central role in the transformational recognition rou-
tines that the listener must use for syntactic analysis. Intonation
provides acoustic cues that segment the speech signal into lingui-
static units suitable for syntactic analysis." Bierwisch (1965) de-
monstrated that it is possible to generate an intonation contour
(for a German sentence), if only the surface syntactic tree and re-
lated syntactic information is provided.

The overall intonation of English sentences has been charac-
terized (Armstrong and Ward 1926) in terms of two alternative
"tunes" or contours, as shown in Figure 1. Tune I has a character-
istic rising of pitch (or its acoustic correlate, F0) until the
first stressed syllable in the sentence is reached, and a falling
of the pitch from the first stressed syllable to the last. Sentence-
final intonation was said to fall dramatically (for declarative sen-
tences, exclamations, and questions with interrogative words). Tune
II is like Tune I, but is terminated by a brief rise (or leveling)
in pitch (within the last stressed syllable, if it is sentence-final;
otherwise in the final unstressed syllables). Tune II was said to
mark yes-no questions, uncertainty or indifference in expression,

Armstrong and Ward observed that complex sentences may be bro-
en up into more than one "sense group", each sense group being
marked by a Tune I or Tune II contour. They argued that different
people divide their speech into different sense groups (cf. also
Scholes 1971:38), and that the division depends upon the type of
message and whether the speech is conversational or deliberate (such
as the reading of written texts). Despite this disclaimer, the
Figure 1. Tune I and Tune II Intonation Contours
intonation contours they gave actually frequently showed a close correlation between pitch rises and the beginnings of syntactic units.

Kenneth L. Pike (1945) divided intonation contours of sentences into **primary contours** and **precontours**. A sentence may be broken into several primary contours and associated precontours. Pike characterized these contour portions as follows: "A stressed syllable constitutes the BEGINNING POINT for every primary contour; there is no primary contour without a stressed syllable, and every heavily stressed syllable begins a new contour" (p. 27). "Immediately preceding the stressed syllables of a primary contour there will oftentimes be one or more syllables which ... are unstressed. These syllables may be called PRECONTOURS ..." (p. 29).

Trager and Smith (1951) asserted a one-to-one relation between boundaries between constituents and prosodic cues. They said (cf. Lieberman 1967a:189) essentially that each linguistic unit **always** is represented by a prosodic pattern clearly present in the acoustic speech signal. Lieberman summarized their position as follows: "The suprasegmentals always provide 'acoustic' cues that tell the listener how to divide the sentence for syntactic analysis" (p. 190).

In summary, there has been a spectrum of expressed viewpoints about intonational cues for constituent structure. One of the weakest hypotheses (Armstrong and Ward 1926) is that sentences may (but need not necessarily) be divisible into parts by intonation contours associated with any arbitrary (but fairly long) sequences of syllables or words. The units need not be syntactic constituents, and indeed the individual talker may divide (or not divide) a sentence differently from time to time, and different talkers may divide utterances differently. At the other extreme, all sentences are assumed to be divisible into syntactic units by intonational (or other prosodic) cues that always occur at unit boundaries (Trager and Smith 1951; Wells 1945). The phonological principle of **invariance** (Chomsky and Miller 1963) applied to such prosodic aspects of language would imply that a syntactic boundary always has an associated acoustic (or phonetic) boundary marker manifested, and only when the syntactic boundary occurs will that acoustic mar-
ker appear (Trager and Smith 1951:51). **Linearity** (Chomsky and Miller 1963) would imply that a boundary between two syntactic units would be manifested by acoustic features at the time stretch ("pause" or such) after the time stretch associated with the last phoneme of the earlier constituent and before the time stretch associated with the phonemes of the later constituent.

Malmberg (1963:69) implicitly rejected the linearity condition for structural boundaries. He broke up utterances into "measures" on the basis of perceived intonation, yielding such divisions as the following:

The boys are ---- playing in the ---- street.
(ʊ bɔr z ar) ---- (pləɪɪŋ in ʊ) ---- (strɪt)

Each measure or group has an "accented" (stressed) syllable and zero or more unstressed syllables. The breaks he shows occur just before the stressed syllables, and not necessarily at the points in the phonemic string where structural boundaries occur. We shall see that this breakdown into groups is similar to what is obtained from automatic analysis of fundamental frequency contours. That is, strict linearity must be rejected if one is to succeed in finding acoustic cues to the syntactic breaks.

Recent experiments on $F_0$ in several forms of connected texts have confirmed the ability to detect syntactic structures from $F_0$ contours. Lea's thesis research at Purdue University (Lea 1972a, chapters 2 and 3) involved four male and two female talkers reading a five-sentence paragraph composed of only monosyllabic words, portions of three weather reports, and a portion of a news report. Also, Lea processed some excerpts from conversational interviews, bringing the total amount of speech processed to 500 seconds of speech, from nine talkers. More recently, Lea (1973a) extended such studies to include six talkers reading a paragraph of the "Rainbow Passage" (Fairbanks 1940) and ten talkers spontaneously speaking instructions (mostly questions or commands) for simulated or actual interaction with computers.

Lea performed intuitive constituent structure analyses of the reallocated surface structures (cf. Chomsky and Halle 1968) of these utterances, giving some thought to the transformational history of
each sentence, and thereby predicting where major syntactic breaks might be expected to be manifested by phonological structure. The boundaries thus predicted by this independent syntactic analysis (which resemble the types of structural boundaries Malmberg displayed) were then compared with $F_0$ contours of all the utterances. His results showed that a decrease (of about 7% or more) in $F_0$ usually occurred at the end of each major syntactic constituent, and an increase (of about 7% or more) in $F_0$ occurred near (but not necessarily coinciding with) the beginning of the following constituent.

Figure 2 illustrates the $F_0$ contour (that is, $F_0$ values, in Hertz, versus time in seconds throughout the utterance) of a typical sentence taken from a weather report. Fall-rise "valleys" (marked by vertical dotted lines) accompany the syntactically predicted boundaries (marked by arrows labelled with the categories of surrounding constituents). A computer program, based on the regular occurrence of $F_0$ valleys at constituent boundaries, correctly detected over 80% of all syntactically predicted boundaries, for all the various texts.

Of the less than 20% of "missing" boundaries, about half were due to predicted boundaries between noun phrases and following verbals (auxiliary verbs or main verbs). There is considerable evidence (e.g., contractions like "I've", subject-verb inversion rules, and the general clitic phenomena) that such noun phrase - verbal boundaries should not be expected in phonological structure. Thus, neglecting noun phrase - verbal boundaries, about 90% of all other boundaries are detected.

It is important to realize that Lea's algorithm is not locating syntactic boundaries, but rather only detecting them. Sometimes the rise in $F_0$ into a constituent may be delayed due to initial weakly stressed syllables or function words like a, of, into, from, etc. A predictable delay occurs before the bottom of the $F_0$ valley occurs, but the delayed boundaries are still considered correctly detected.

Besides this regular acoustic manifestation of boundaries between major syntactic constituents, some boundaries between minor constituents (e.g. between an adjective and a following noun) were
Figure 2. An F0 Contour (Vertical Axis Frequency; Horizontal Axis, Time), of a Sentence, with Predicted Constituent Boundaries Shown by Arrows (Labelled with Category Symbols for Surrounding Constituents), and Detected Boundaries Shown by Vertical Lines.
also detected by the fall-rise patterns in F₀. In his earliest studies, Lea (1972, chapter 3) investigated the effects of specific constituent categories (noun phrase, verb, prepositional phrase, etc.) on boundary detection. The lack of regular boundary marking between noun phrases and following verbs has already been noted. On the other hand, around 95% of all boundaries before prepositional phrases were detected by F₀ fall-rise valleys. Also, coordinate noun phrases or coordinate adjectives were always accompanied by F₀ valleys between the conjuncts. Sizes of F₀ valleys were also studied for the various syntactic categories.

Sentence boundaries were always accompanied by fall-rise F₀ contours. In fact, the rise in F₀ (around 90% change) after a sentence boundary was substantially larger than the usual rises (about 40% or less) after non-sentential constituent boundaries. In addition, sentence boundaries were usually (in over 90% of all cases) accompanied by long (35 centisecond) stretches of unvoicing. Here "sentence boundaries" refer to both boundaries between matrix (unembedded) sentences and boundaries between embedded full-clausal sentences.

These earlier results with the texts studied at Purdue University were confirmed in Lea's later studies (Lea 1973a) with other texts. However, the spontaneous sentences taken from "man-computer interactions" showed more monotonic intonation and occasional thoughtful "hesitation" pauses that could be confused with sentential pauses.

Boundaries were occasionally falsely detected at F₀ valleys that were apparently not in any way syntactically related. These resulted primarily from F₀ variations between vowels and consonants. As shown in Figure 3, sudden increases or "jumps" in F₀ occur following unvoiced consonants, and are followed by rapid fall of F₀ within a few centiseconds (hundredths of a second, equivalent to tens of milliseconds). Fundamental frequency will also decrease suddenly, or "dip", within voiced obstruents, then increase suddenly again at opening of the vocal tract, then increase gradually in the first few centiseconds of vowels (or sonorants) following such voiced obstruents. Such F₀ changes due to consonant-vowel (and
Figure 3. A Portion of an $F_0$ Contour from a Weather Report Read by GWH, Showing Sudden Increases in $F_0$ after Unvoiced Consonants (↑) and Decreases, or Dips, in $F_0$ During Voiced Consonants (↓).
vowel-consonant) transitions not only interfere with the detection of constituent boundaries; they also interact with stress patterns in the utterances, so that an \( F_0 \) fall in a vowel following an unvoiced consonant is more likely in unstressed than stressed syllables, while \( F_0 \) increases following voiced obstruents are more likely in stressed than in unstressed syllables (Lea 1972, chap. 5).

These complicated interactions between stress patterns, constituent boundaries, and phonetic transitions suggest the need for controlled experiments (which isolate and analyze variations in \( F_0 \) due to one of these variables while the others are kept fixed or controlled).

2.2. General Effects of Stress Patterns on \( F_0 \) Contours

Before considering in considerable detail the controlled studies of phonetic effects on \( F_0 \) contours, and the interactions between stress-related and phonetic transitional effects, let us briefly consider what is known about how linguistic stress affects \( F_0 \) contours.

Many studies have shown that higher \( F_0 \) is associated with stressed syllables (Bolinger 1958; Fry 1958; Lieberman 1960; Morton and Jassem 1965; Lehiste 1970; Lea 1972b). However, even better correspondence is to be found between local increases in \( F_0 \) and stress than is provided by the absolute peak (or mean) values of \( F_0 \) within stressed vowels or syllables (Bolinger 1958; Medress et al 1971; Morton and Jassem 1965). Some studies suggest that it is the presence of such \( F_0 \) changes that marks stress, not the specific magnitude of the change (Fry 1958; Morton and Jassem 1965).

Intonation studies (Armstrong and Ward 1926; Lieberman 1967; Lea 1972b, 1973a) have shown that, in connected texts and spoken sentences, \( F_0 \) will usually reach a maximum near the first stressed syllable (the so-called "head") of each breath group or clause, and will fall gradually until the last stressed syllable, after which may occur either the rapid fall of an utterance-final Tune I contour or the rise in \( F_0 \) at the end of Tune II contours (which mark "incompletion"). Figure 1 illustrated the general shapes of these basic intonation contours. Obviously, the last stressed syllable of Tune I contours will not consistently exhibit the \( F_0 \) rises
generally assumed to accompany stressed syllables. Also, unstressed syllables in the terminal rise of a Tune II contour will be accompanied by $F_0$ rises that do not mark stress. On the other hand, these studies suggest that the peak $F_0$ of the total contour will be associated with a stressed syllable.

The assumption of Lea's constituent boundary detector is that sentences consisting of several major grammatical constituents will be broken into several Tune I- or Tune II-like subcontours, riding on the general tune for a sentence or clause. Thus, as illustrated in Figure 2, $F_0$ contours in sentences with several major constituents will have major $F_0$ changes associated with the constituent structure. We might call these rapidly rising and gradually falling $F_0$ contours "archetype constituent contours". They resemble Lieberman's (1967) unmarked and marked breath groups, and Pike's (1945) primary contours plus precontours, and other contours associated with "sense groups" in the literature.

Lea devised a general algorithm for automatically locating stressed syllables, which assumed that the $F_0$ rise that marks the beginning of the "constituent" found by his boundary detector was associated with the first stressed (or "head") syllable of the constituent. In a sense, then, the constituent boundary detector may be said to be detecting some stressed syllables (but not locating them). Lea's stress-location algorithm then located stressed HEAD syllables within the nearest portion of speech which had non-falling $F_0$ and high speech intensity for a long duration (yielding a large integral of energy within the "syllable"). Substantial decreases in energy level marked the ends or limits of this stressed syllable.

If each constituent had exactly one lexical word with a major-stressed syllable within it (as has been suggested for deep structures; cf. Chomsky 1965; Emonds 1970), one might well expect a one-one correspondence between a detected constituent boundary and the presence of a following stressed syllable. In fact, however, surface structure constituents (both as predicted syntactically and as found by Lea's boundary detector) sometimes contain more than one stressed syllable. In Lea's stress location algorithm, other stressed syllables in the constituent were assumed to be manifested
by deviations in $F_0$ above an archetype line (a straight line on a semi-log plot) from the $F_0$ peak to the $F_0$ value at the end of the constituent. Again, such stressed syllables were delimited by decreases in energy.

Lea's algorithm for stressed syllable location succeeded in locating about 85% of all syllables perceived as stressed by the majority votes of a panel of listeners (Lea 1973a). This demonstrated the regularity with which stress is marked, in connected English texts, by local increases in fundamental frequency.

2.3. Interactions Between Segmental and Suprasegmental Influences

Sentence intonation, constituent structure, stress, and phonetic content of utterances have all been shown to influence the shapes of $F_0$ contours in English sentences. To understand their individual effects more fully and quantitatively, further controlled experiments are needed (cf. Lea 1973a). Here we shall use the example sentence in Figure 2 to illustrate how a complex $F_0$ contour might be composed from the interactions of intonation, syntax, stress, and phonetic influences.

Shown in Figure 4a is the gross overall sentence intonation for a two-clause sentence. Each clause has an overall Tune I contour. (If incompletion was to be marked on the first clause, as when it is a subordinate clause, a Tune II might occur in that clause; if a yes-no question were involved, the second clause could have ended in a Tune II terminal rise in $F_0$.) The first clause has the highest $F_0$, while the large rise associated with the beginning of the second clause will normally not attain the full height of the earlier clause. There is a general trend toward decreased $F_0$ values as the utterance proceeds. (This is even true for paragraphs, in that the $F_0$ peak in later sentences rarely are as high as that of the first sentence of a paragraph (Lea 1972b).)

Next, suppose that we know the first clause consists of five constituents, and the second clause consists of four constituents. We expect each such constituent to exhibit a Tune I-like contour (or, under some conditions, a Tune II-like contour), that "rides on" or centers around the overall clause contour. Then, we might expect the general $F_0$ contour shapes of Figure 4b.
(a) Overall Sentence Intonation Contour for Two-Clause Sentence.

(b) Archetype $F_0$ Contours of Constituents Riding on the Overall Sentence Contour.

(c) Local Increases in $F_0$ Assigned Where Stressed Syllables (S) Occur, All Riding on Contours of the Constituents.

(d) $F_0$ Variations Due to Phonetic Contents Superimposed on Contours that Incorporate Stress, Constituent Structure, and Sentence Intonation Influences on $F_0$ Contours.

Figure 4. Progressive Composition of a Complex Intonation Contour from the Superposition of Effects of Sentence Type, Constituent Structure, Stress Patterns, and Phonetic Sequences. Each figure shows the contour of the prior step as a dotted curve upon which the effect of the new factor is superimposed. The sentence is that shown in Figure 2.
The inclusion of these archetype $F_0$ contours of the constituents thus begins to yield a general pattern grossly as in Figure 2.

Stress effects may then be superimposed on the archetype contours, by increasing $F_0$ in the region of stressed syllables, and reducing the relative $F_0$ below the archetype for reduced syllables. If the stressed syllables are at those time stretches marked by the dark horizontal bars and "S's" shown in Figure 4c, then the contour of Figure 4c results. The contour is beginning to resemble the $F_0$ contour of Figure 2.

Finally the effects due to phonetic sequences are introduced in Figure 4d. During unvoiced consonants, $F_0$ contours are blanked out. Immediately following unvoiced consonants, $F_0$ is assumed to start high (higher than the contour from 4c) and rapidly fall. During voiced obstruents, $F_0$ dips slightly below the contour of Figure 4c. The resulting contour in Figure 4d closely resembles the actual $F_0$ contour of the utterance shown in Figure 2.

Finer-grain adjustments in the contours could be made based on such additional aspects as relative levels of stress, syntactic subordination (Crystal 1969), vowel heights and consonantal place of articulation (Lea 1972b), etc. In general, however, it is apparent that the superimposed effects of sentence intonation, constituent structure, stress, and phonetic content give an encouragingly accurate account of the structure of $F_0$ contours. Before a more thorough analysis of $F_0$ contours can be accomplished, the individual contributions of each major factor must be more precisely specified. This is a major reason for the detailed studies of phonetic influences and stress, to be discussed in the remainder of this paper. Extensive experiments that isolate the effects of syntactic bracketing, syntactic categories, and subordination are presently being devised, to parallel those experiments which isolate phonetic effects, which will now be discussed.

3. INTRINSIC VALUES OF $F_0$ IN VOWELS AND CONSONANTS

To isolate the segmentally dictated aspects of intonation from the syntactically determined aspects, studies were conducted on $F_0$ contours in isolated "words", with a representative sampling of
English consonant-vowel-consonant sequences. Isolated bisyllabic nonsense utterances of the form [hαCVC] were recorded by two male talkers (ASH and GWH), then digitized, and processed through a computer program which determined $F_0$ values for every 10 milliseconds (called a time segment) of the speech. The resulting $F_0$ contours were analyzed and compared to digital spectrograms to establish phonetic effects due to consonant and vowel features.

A total of 326 different [hαCVC] utterances were recorded by each talker (Lea 1972b:93). All symmetric $C_1VC_2$ combinations (where $C_1 = C_2 = [p, t, k, b, d, g, f, θ, s, f, tʃ, v, ʒ, z, ɾ, m, n, or l]$) were included, for each English consonant except [ʒ, r, η, h, w, and j], which are not expected to be naturally placed in both consonantal positions by General American talkers. For each such symmetric consonantal context, twelve vowels ([i, ɪ, e, ɛ, a, ɑ, Α, o, ɔ, u, ʊ, ʌ]) were inserted as the stressed vowel, yielding 18 lists of words with 12 words per list.

In addition, asymmetric CVC combinations were recorded, in which either silence (i.e. a glottal stop) or one of the consonants not included in the symmetric lists was placed in either final or medial position, and the other position was filled by one of the six English stops. Each stop thus occurred twice in the asymmetric words of each such list, randomly paired with two of the vowels.

3.1. **Intrinsic $F_0$ Parameters Related to Single Phonemes**

Figure 5 illustrates the general forms of $F_0$ contours found for such [hαCVC] utterances.

Inspection of Figure 5 quickly demonstrates that, immediately after unvoiced consonants, $F_0$ is high and falling, while, in the same position, $F_0$ rises from the low values it has within voiced consonants, to yield rising $F_0$ at boundaries between voiced consonants and following vowels. These representative contours also show another immediate result found in all the data; namely, the lack of any rise in $F_0$ immediately before the medial unvoiced consonants. Thus unvoicing is not accompanied by high $F_0$ values on both sides. On the other hand, there is a general tendency for $F_0$ to fall at the beginning of voiced medial consonants, and rise at
(a) $F_0$ Contour with Unvoiced Consonants.

(b) $F_0$ Contour with Voiced Consonants.

Figure 5. Typical $F_0$ Contour Shapes in [hCVC] Utterances with Medial Unvoiced (a) or Voiced (b) Consonants.
the beginning of the following vowel, yielding a distinctive "valley" which marks the presence of voiced consonants.

Figure 5 also illustrates that voicing of consonants cannot always be detected simply from whether or not the fundamental frequency contour is continuous. Voicing may cease during later portions of voiced obstruents, as shown by the missing line segment in Figure 5b. This is due to reduced transglottal pressure during closure of the vocal tract, causing a cessation of vocal cord vibration. This does not occur during sonorants. In addition, the closure of the vocal tract for unvoiced consonants usually occurs before voicing ceases, so that the brief dip in $F_0$ values shown by the dashed line in Figure 5a may occur. This makes the voiced and unvoiced obstruents sometimes look quite similar, with only the duration of unvoicing and the slope of the $F_0$ contour in the following vowel as remaining cues to the phonemic state of voicing of the consonant.

These are some of the qualitative observations one can make from the $F_0$ contours of Figure 5. Quantitative observations require defining specific features of the contours that will represent the relative $F_0$ values and the rises and falls mentioned above.

In Figure 6 are shown the definitions of some important contour parameters that relate primarily to a single phoneme, which might be called "intrinsic" $F_0$ parameters. They are: the central value $CF_{V1}$ of $F_0$ in vowel $V_1$ (which is the value that was judged to be representative of the values in the center of the vowel [a]); the value $CF_{C1}$ of $F_0$ central to the consonant $C_1$; and the initial value $F(i)$ and peak value $PF_{V2}$ of $F_0$ in vowel $V_2$. The average slope of $F_0$ in $V_2$ might also be considered "intrinsic" to that vowel and relevant to falling and rising tones in tone languages. However, this utterance-final slope is probably primarily determined by the mode of list reading rather than being a correlate of vowel identity or consonantal context. Consequently, we shall not discuss here this slope in the latter part of the vowel. In section 4, however, the slope in the initial portion of a vowel will be discussed.

The final consonant in the [hasC$\hat{V}_2$,C$_2$] utterance showed no
Figure 6. Contours Illustrating Definitions of Intrinsic $F_0$ Parameters for (a) Unvoiced and (b) Voiced Consonants in [hacVC] Utterances.
systematic effects on any of the $F_0$ values within, or prior to, the stressed vowel. Comparisons of $F_0$ values in the symmetric lists (for $C_1$ and $C_2$ as stops) with those in the unsymmetric lists (which have $C_1$ a stop, and $C_2$ either $\theta$, $\eta$, or $\omega$ [that is, silence; no final consonant]) showed no substantial effects of $C_2$ on peak $F_0$ in $V_2$, initial $F_0$ in $V_2$, or other critical $F_0$ values as defined above. The studies of Mohr (1971), which include more complete variations of final consonant for a fixed $C_1$, also show that $F_0$ in vowels is essentially independent of the identity of following consonants.

In fact, it was even difficult to tell where the final consonant began without careful study of the digital spectrograms. For example, the sudden $F_0$ dip expected in voiced obstruents was not usually evident in these final consonants. Consequently, final consonants will not be discussed in the following presentations of phonetic effects.

This lack of influence from final consonants would suggest that, at least in isolated words and utterance-final syllables, final consonants might not be expected to affect the relative likelihood of high versus low, or falling versus rising, tones in tone languages.

3.2. Consonantal Influences on Intrinsic $F_0$ Parameters

Figures 7 and 8 illustrate the intrinsic $F_0$ parameters found for [hoCVC] utterances spoken by talkers ASH and GWH, respectively. The solid dots indicate the average value of the parameters for the 12 symmetric utterances with each specific consonantal context, and the vertical bars with horizontal lines as end points indicate the range of values of that parameter for the 12 stressed vowels. Thus, in Figures 7a and 8a, for example, are shown the central value $CF_{V1}$ of $F_0$ in $V_1$ and the peak value $PF_{V2}$ of $F_0$ in $V_2$, displaced horizontally as they might be in an $F_0$-versus-time plot of the utterance with unvoiced consonants. In Figures 7b and 7c, and 8b and 8c, are shown the parameters $CF_{V1}$, $CF_{C1}$, $F_1$ and $PF_{V2}$ in order as they appear in [hoCVC] utterances with voiced obstruents and sonorants.

A rise of approximately 40 Hz average in $F_0$ from $CF_{V1}$ to $PF_{V2}$ is evident in the unvoiced environments of Figures 7a and 8a. A
similar, but smaller, rise of approximately 30 Hz average is evident
in the voiced environments of Figures 7b and c and 8b and c. This
rise from C_F V_i to P_F V_2 is a cue to the relative stress levels of
the reduced [ə] and stressed V_2. The initial F_0 values P(i) in
vowels after unvoiced consonants, which are equal to or very close
to the peak F_0 values P_F V_2 in the stressed vowel, are about 30 Hz
(or 20%) higher than the initial values in vowels following voiced
consonants.

Studies (Lea 1972b:111) showed that 98% of all stops and fric-
atives spoken by ASH in [hæCVC] utterances could be correctly
categorized as unvoiced or voiced by a simple hypothesis that "if
initial F_0 in V_2 is greater than 155 Hz, C_1 is unvoiced; otherwise,
it is voiced". A similar hypothesis for talker GWH, with a divid-
ing line at 157 Hz, also yielded 98% correct categorization of un-
voiced and voiced stops and fricatives. On the other hand, the
best unvoiced/voiced separation one can achieve with a comparable
threshold value for peak F_0 is 78% for a threshold of 166 Hz for
ASH, and 90% for a threshold of 166 Hz for GWH.

These results suggest that the initial value of F_0 in V_2 is a
good cue to the unvoiced/voiced feature of C_1 (provided C_1 is re-
stricted to being a stop or fricative), and that even the peak F_0
in V_2 gives a fair indication of the state of voicing. Higher F_0
values are, as expected, generally associated with unvoiced con-
sonantal contexts. House and Fairbanks (1953:131) found that the
average F_0 value in the stressed vowels of [hæCVC] utterances was
4.47 Hz higher with unvoiced consonantal contexts than with voiced
contexts. Here we have seen that initial and peak values of F_0 in
V_2 show even larger differences (in absolute value or percentage)
than did House and Fairbanks' time-averaged values.

Another prominent trend shown in Figures 7 and 8 is graphically
illustrated in Figure 9, where average F_0 values for each talker
are superimposed for that talker's sonorants (shown by •'s) and his
voiced obstruents (shown by x's). Fundamental frequency will de-
crease, or "dip", during the closure of voiced obstruents yielding
a prominent local "valley" in F_0 contours. This dip (of about 10%)
in F_0 is "caused by the increased air pressure [during closure of
Figure 7. Average Values ($\dagger$) and Extreme Ranges of Values (I) of Intrinsic $F_0$ Parameters for Various Consonants $C_1$ in $[h\tilde{a}C_1 V_2 C_2]$ Utterances Spoken by ASH. ($C_1=C_2$)
Figure 8. Average Values (●) and Extreme Ranges of Values (I) of Intrinsic $F_0$ Parameters for Various Consonants $C_1$ in $[\text{h}aC_1\text{\text{̆}}_2C_2]$ Utterances Spoken by GWH. ($C_1 = C_2$)
Figure 9. Average F₀ Values, Taken from Figures 7 and 8, Show the Distinctive F₀ Dips that Accompany Voiced Obstruents, and the Monotonic Increases that Accompany Sonorants. The X's and •'s mark average values (for all vowels pooled), one X for each prevocalic obstruent, and one • for each prevocalic sonorant.
the oral cavity] in the upper cavities of the vocal tract above the vocal cords, and, therefore, by a reduction in the pressure differential on the cords" (Chistovich 1969:372). Thus, during a /b/ stop, F₀ is lowered as intra-oral pressure increases, and, as the stop is opened, F₀ returns within a few centiseconds to the value typical of the vowel. During an intervocalic nasal, such as /m/, there is no reduction in F₀ because the air escapes freely out through the nasal openings.

Chistovich (1969) found that, when a synthetic VCV utterance was presented to each of ten untrained subjects, and the individual subject was allowed to adjust the value of F₀ during medial labial consonants until he found the value perceived as corresponding to the transition of the consonant identify from /b/ to /m/ or vice versa, the subject set the F₀ value in the consonant at about 12 Hz lower than in the vowels. Thus, an F₀ dip of 12 Hz or more was needed in the consonant to make it sound like an obstruent, as opposed to a sonorant. In addition, Chistovich found that over 80% of all /b/ and /m/ phonemes in natural spoken utterances /babab/ and /mama/ could be correctly recognized (that is, distinguished) by using only the amount of F₀ dip during closure, thus demonstrating that "the jump in the fundamental pitch can be used as a phonemically useful cue" (Chistovich 1969:377).

Figure 10 shows the close correlation between critical F₀ values in [hæCV] utterances of the talkers. The clustering of values near a linear relation (in Figures 10a and c) shows that consonantal effects on F₀ contours are fairly consistent from talker to talker. Also apparent in Figure 10 are: (1) the clustering of F₀ values in the consonant C₁ into two classes for sonorant and non-sonorant consonants, with F₀ higher by about 13% in sonorants than in obstruents; (2) the higher peak F₀ values (by about 10% on the average) after unvoiced consonants, in contrast to voiced consonants; and (3) the higher F₀ values (by about 7% for F₁V₂) associated with [-coronal] voiced stops in comparison with the F₀ values associated with [+coronal] voiced stops. Figure 10b shows a clustering of F₀ values at onset of the stressed vowel into groups of semi-vowels, stops, and fricatives, but no correlation from talker to talker, perhaps due to difficulties in defining the onset of the vowel.
In summary of these consonantal effects on intrinsic $F_0$ parameters, we may note the following average trends, listed in order of decreasing sizes or significance:

1. Initial $F_0$ values in stressed vowels are about 20% higher when the preceding consonant is unvoiced, compared to when that consonant is voiced;

2. Central $F_0$ values in voiced obstruents are about 13% lower than in sonorants. This is produced by a physiologically-dictated dip in $F_0$ during closure of voiced obstruents, while $F_0$ values during sonorants follow along the line between $F_0$ values in surrounding vowels;

3. Peak $F_0$ values in stressed vowels are about 10% higher when preceded by unvoiced consonants than when preceded by voiced consonants;

4. Other aspects of the manner and place of articulation of the consonants preceding stressed vowels have smaller and less consistent effects on $F_0$ contours. One such effect is that peak $F_0$ values in the stressed vowels are about 7% higher after [-coronal] voiced stops than for [+coronal] voiced stops;

5. The final consonant shows no systematic effect on $F_0$ values in the stressed vowel. The state of voicing of the medial consonant has no apparent effect on the central $F_0$ value in the preceding unstressed [ə]. There is a slight tendency for higher $F_0$ values in the [ə] if the following consonant is sonorant.

Thus, $F_0$ values are higher in sonorants than in obstruents. Initial and peak $F_0$ values in vowels are higher when the preceding consonant is unvoiced than if it is voiced, but are not affected significantly by following consonants.

3.3. Higher Vowels Have Higher $F_0$

Figures 11 and 12 show the effects of vowel identity on the initial and peak values of $F_0$ in vowel $V_2$, for talker ASH and talker GWH, respectively. Each line shows the average of results for a category of consonants in symmetric ($C_1 = C_2$ in [həC_1V_2C_2]) lists. Vowels $V_2$ have been ordered such that the highest vowels are closest to the extreme right and left positions, with tongue height lowering toward the center. Front vowels are on the left, and back vowels on the right. With this ordering, it is evident that both initial and peak $F_0$ increase with tongue height. This is in agreement with previous studies of time-average $F_0$ or peak $F_0$ in vowels (Lehiste 1970:68-70).
(a) Initial $F_0$ Values in Stressed Vowel

(b) Peak $F_0$ Values in Stressed Vowel

Figure 11. Effects of Vowel Identity on Initial and Peak $F_0$ Values in $V_2$ of $[haC_1V_2C_2]$ Utterances, for Talker ASH, with Consonants Pooled by Categories.
Figure 12. Effects of Vowel Identity on Initial and Peak F₀ Values in $V_2$ of [hC₁v₂C₂] Utterances, for Talker GWH, with Consonants Pooled by Categories.
Also evident in Figures 11 and 12 is the general tendency toward higher peak $F_0$ in vowels with unvoiced than with voiced consonantal contexts (on the average, 14 Hz higher for unvoiced than for voiced stops and fricatives, for ASH, and 24 Hz for GWH). Even larger differences occur between the initial $F_0$ values in vowels after voiced and unvoiced consonants. The initial values of $F_0$ after unvoiced consonants are usually the peak $F_0$ in the vowel, or very near that value. It is evident in comparing the peak $F_0$ values after unvoiced stops and fricatives, from Figures 11 and 12, with the initial $F_0$ values after voiced stops and fricatives, shown in Figures 11 and 12, that initial $F_0$ values after unvoiced consonants are considerably higher (by an average of 24 Hz for ASH, 31 Hz for GWH) than after corresponding voiced consonants.

Thus, while the tongue height during vowel articulation affects $F_0$ values significantly, the preceding consonant may have even more effect on the $F_0$ values during that vowel.

In summary, we may note that $F_0$ values in [hsCVC] utterances are most markedly affected by the stress pattern and the state of voicing of the consonant $C_1$, then progressively less affected by the height of the tongue during the vowel $V_2$ articulation, the manner of consonant $C_1$ articulation, and the place of consonant $C_1$ articulation. $C_2$ has no systematic affect on the defined intrinsic $F_0$ parameters.

4. $F_0$ VARIATIONS AT CONSONANT-VOWEL BOUNDARIES

4.1. The $F_0$ Slope is a Voicing Cue

Consonantal context and the high/low feature of vowels have been shown to affect $F_0$ values at critical times in [hsCVC] utterances. Perhaps even more significant are the effects of consonant identify (in the lists with symmetric CVC) on the transition in $F_0$ value from one phoneme to another. Such $F_0$ transitions may be involved in affecting the relative likelihood of falling and rising tones in various consonantal contexts.

The transitional $F_0$ changes are known to be of some perceptual significance. Recent research has disclosed that synthesized speech is more natural when, immediately after unvoiced consonants,
the voice fundamental frequency ($F_0$) jumps suddenly to higher values and then rapidly decreases, and, at the onset of voiced obstruents, $F_0$ dips and remains at a lower value throughout the voiced obstruents. Mattingly (1966) found that incorporating about 6% pitch rises after unvoiced segments, with a following reduction in $F_0$ of about one percent per centisecond, yielded considerably more natural-sounding synthesized speech. Haggard, Ambler, and Callow (1970) suggested that such pitch jumps may be good cues to the unvoiced character of the preceding consonantal segment. They synthesized stop-vowel syllables with falling, rising, and flat contours at the onset of voicing, and concluded that "a low rising pitch leads to perception of an initial stop as voiced, while a high falling pitch leads to perception as voiceless" (Haggard et al 1970:617).

4.2. The Parameter $\Delta_{10}F$ is a Voicing Cue

With some restrictions, this work with synthetic speech demonstrated the validity of the hypothesis that "pitch is a voicing cue," in that the pitch jumps and subsequent decrease in $F_0$ after vowel onset, and the dips during voiced consonants followed by generally rising $F_0$ after vowel onset, provide phonemic information about the unvoiced/voiced character of stop consonants. The initial rise or fall in $F_0$ after initiation of the vowel may be quantitatively characterized by the elementary $F_0$ contour parameters defined in Figure 13. Time $t_{i+10}$ is defined as the time which is 10 time segments (thus, 100 msec) after the $t_i$ of initiation of the vowel. The corresponding value of $F_0$ at that time is represented as $F(i+10)$. The rise of $F_0$ from $t_i$ to $t_{i+10}$ is then given by: $\Delta_{10}F = F(i+10) - F(i)$. For a fall in $F_0$, following unvoiced consonants, this parameter $\Delta_{10}F$ would be negative.

The choice here of 10 time segments (100 ms) was somewhat arbitrary. Haggard, Ambler and Callow (1970) used a time interval of 55 ms in their synthesis of falling and rising $F_0$ contours. Time intervals of 30 ms and 50 ms were also used in Lea's study, but showed no patterns not more systematically and vividly shown by the 100 ms parameter (Lea 1972b:100-103 and section 4.7).

The parameter $\Delta_{10}F$ measures the amount of rise (or fall, for
Figure 13. The Parameter $\Delta_{10}F$ is Defined as the Increase in $F_0$ in the First Ten Centiseconds after the Time $t_i$ of Onset of the Stressed Vowel of the [həCVC] Utterance. For decreasing $F_0$, $\Delta_{10}F$ will be negative.
negative values) in $F_0$ from its initial value in the vowel, $F(i)$, to its values 100 ms (10 "segments") later. Figure 14 compares the average values of $\Delta_{10}^F$ for talkers ASH and GWH, and vividly shows the contrasts between influences of voiced and unvoiced consonants. In Figure 14, each consonant is marked at the position whose coordinates are the average values (for the pooling of all 12 vowel environments) of $\Delta_{10}^F$ following that consonant, for the two talkers. Clearly, there is some consistent tendency from one talker to the other, since the sounds are not scattered randomly, but are clustered near the 45° line of equal $\Delta_{10}^F$'s. Also, there is a prominent tendency toward negative $\Delta_{10}^F$ (that is, a falling contour) following unvoiced consonants, and corresponding positive values (indicating a rising contour) for voiced consonants.

Haggard et al (1970:616) had contended that, since the glottis is open partially for the strident fricatives, cessation of glottal vibration will occur with them, and the $F_0$ jump and following fall of an unvoiced fricative may then be observed. Contrary to their conjecture, voiced strident fricatives in $C_1$ position of the \text{[hC₁} \text{;} C_2] utterances caused substantial rises in $F_0$ following onset of the following vowel, just as other voiced consonants did. Both strident and non-strident fricatives exhibited temporary cessation of voicing, apparently due to reduced transglottal pressure during vocal tract closure.

Lea's results (1972b:110-121) demonstrated the reliability of detecting voicing/unvoicing from the values of $\Delta_{10}^F$. For example, the simple hypothesis that "a rise in $F_0$ at vowel onset marks a preceding voiced consonant, and a fall marks an unvoiced consonant" would be correct for 99% of all consonants of ASH, and 92% for GWH, if "a rise" meant $\Delta_{10}^F > 0$ and "a fall" that $\Delta_{10}^F \leq 0$.

Thus, the rise or fall of $F_0$ after onset of the stressed vowel $V_2$ is a good cue to the state of voicing of the prevocalic consonant. It is not perfectly reliable, anymore than any other speech parameter has been for other phonetic classification efforts.

4.3. Low Vowels May Have Flatter Initial $F_0$ Slopes

The effects of vowel identity on the contour parameter $\Delta_{10}^F$,
Figure 14. Comparison of $\Delta_{10}F$ Parameters for Talkers ASH versus GWH, showing an increasing $F_0$ ($\Delta_{10}F$ Positive) following Voiced Consonants, and decreasing $F_0$ ($\Delta_{10}F$ Negative) following Unvoiced Consonants. Each symbol represents an average for the twelve vowels following the given consonant.
are shown in the plots of average $\Delta_{10}F$ values for each type of consonantal environment, as shown in Figure 15a for talker ASH, and in Figure 15b for GWH. Again the most dominant effect is related to the vowel feature high/low, for talker ASH. Low vowels spoken by talker ASH tend to yield somewhat flatter contours at vowel onset, for both unvoiced and voiced preceding consonants. While there is a slight tendency of the same form for GWH (Figure 15b), his contour rises or falls are considerably less consistent functions of vowel height.

In summary, we might conclude that tongue height in vowel articulation does affect the magnitude of $F_0$ rises and falls ($\Delta F$ parameters) after medial consonants, but not always consistently.

Despite such slight and irregular variations with vowel height and consonant articulation, the $F_0$ contour very reliably rises at stressed vowel onset for preceding voiced consonants and falls for preceding unvoiced consonants.

5. INTERACTIONS BETWEEN STRESS AND SEGMENTAL INFLUENCES ON $F_0$ CONTOURS

A natural extension of the previous studies of phonetic effects of $F_0$ contours in [hCVC] utterances would be to consider whether the $F_0$ jumps, dips, and slope parameter values occur in unstressed syllables as well. All [hCVC] words exhibit stress on the second vowel, so that studying $F_0$ contours in the CVC sections of such words has demonstrated effects when the consonant precedes a stress peak. Studying the VC sections indicated the lack of any significant effects of a word-final consonant following a stressed vowel. However, more general studies require the use of words with other stress patterns.

In this section, we consider a set of pairs of bisyllabic words with contrasting stress patterns (first syllable stressed, $\hat{V}_1 - V_2$, versus the second syllable stressed, $V_1 - \hat{V}_2$). The word pairs have been carefully selected to include a variety of consonants and vowels in various positions in the words. This will permit a study of stress effects on $F_0$ contours, and an extension of the study of phonemic effects (particularly, prevocalic consonants)
Figure 15. Effects of Vowel Identity on the Slope Parameter \( \Delta_{10}^F \) for Talkers ASH (a) and GWH (b).
on $F_0$ contour parameters for different stress patterns.

Questions to be considered here include the following: Do the $F_0$ jumps after unvoiced consonants, and the $F_0$ dips within voiced consonants, occur in both stressed and unstressed syllables? Is there any relationship between the contour slope in the initial part of a vowel and the stress on that vowel? Is $F_0$ consistently higher in stressed syllables than in unstressed syllables?

5.1. Pairs of English Bisyllabic Words with Stress Contrasts

A list of pairs of bisyllabic words has been compiled such that one word in each pair has major stress on the first syllable while the other word has major stress on the second syllable. The phonemic structures of the words in each pair were similar or identical, so that stress effects could be substantially isolated from phonemic effects. These word pairs are shown in Table I. In most cases, both words are spelled alike, and so only one spelling is given for such pairs. A few pairs have different spellings for the two words and are shown with a diagonal (/) between the words. Exactly how the list in Table I was designed was discussed by Lea (1972b, Appendix A). In brief, word pairs were arranged in terms of what consonant came immediately before the second vowel. For each such consonant before the second vowel, three or four word pairs were selected by attempting to include, in the second syllable position, both high and low vowels, front and back vowels, etc. Some consideration was given to including a variety of consonantal environments immediately after the first vowel, $V_1$, as well.

This selection of word pairs, arranged according to the manner and place of the medial consonant $C_g$, is similar to the set of lists of [heC$_1$VC$_2$] utterances grouped according to the identity of C$_1$. The list of word pairs was recorded (and $F_0$ contours obtained) for the same two talkers (ASH and GWH) who recorded the [heCVC] utterances, with special care given to removing any interactions between word order and stress contrasts (Lea 1972b:137-138).

5.2. Effects of Stress on "Intrinsic" $F_0$ Parameters

Representative $F_0$ contours in the word pairs are illustrated in Figure 16. The general tendency toward higher $F_0$ values (in both
TABLE 1. BISYLLABIC WORD PAIRS WITH CONTRASTING STRESS PATTERNS
(C₀V₁C₀C₀V₂C₀ versus C₀V₁C₀C₀V₀C₀, grouped by C₀,
where C₀ is any sequence of zero or more consonants.)

<table>
<thead>
<tr>
<th>PLACE OF ARTICULATION OF C₀</th>
<th>MANNER OF ARTICULATION OF C₀</th>
<th>LIQUIDS AND GLIDES</th>
</tr>
</thead>
<tbody>
<tr>
<td>LABIAL</td>
<td>STOPs</td>
<td></td>
</tr>
<tr>
<td>UNVOICED</td>
<td>VOICED</td>
<td></td>
</tr>
<tr>
<td>labial</td>
<td>unvoiced</td>
<td></td>
</tr>
<tr>
<td>affix</td>
<td>invite</td>
<td>commune</td>
</tr>
<tr>
<td>defect</td>
<td>convert</td>
<td>refuse</td>
</tr>
<tr>
<td>refund</td>
<td>convoy</td>
<td>perfume</td>
</tr>
<tr>
<td>transfer</td>
<td>survey</td>
<td></td>
</tr>
<tr>
<td>FRICATIVES</td>
<td>VOICED</td>
<td></td>
</tr>
<tr>
<td>labial</td>
<td>unvoiced</td>
<td></td>
</tr>
<tr>
<td>present</td>
<td>conjure</td>
<td>annex</td>
</tr>
<tr>
<td>outshoot</td>
<td>digest</td>
<td>annal/annul</td>
</tr>
<tr>
<td>surcharge</td>
<td>project</td>
<td>allay</td>
</tr>
<tr>
<td>AFFRICATES</td>
<td>NASALS</td>
<td></td>
</tr>
<tr>
<td>labial</td>
<td>unvoiced</td>
<td></td>
</tr>
<tr>
<td>annex</td>
<td>exploit</td>
<td></td>
</tr>
<tr>
<td>collectors</td>
<td>collect</td>
<td></td>
</tr>
<tr>
<td>conflict</td>
<td>billow/ below</td>
<td></td>
</tr>
<tr>
<td>VELAR</td>
<td>VOICED</td>
<td></td>
</tr>
<tr>
<td>labial</td>
<td>unvoiced</td>
<td></td>
</tr>
<tr>
<td>annex</td>
<td>escrow</td>
<td></td>
</tr>
<tr>
<td>forecast</td>
<td>increase</td>
<td></td>
</tr>
<tr>
<td>MASTER</td>
<td>UNVOICED</td>
<td></td>
</tr>
<tr>
<td>labial</td>
<td>unvoiced</td>
<td></td>
</tr>
<tr>
<td>annex</td>
<td>compress</td>
<td></td>
</tr>
<tr>
<td>forecast</td>
<td>contrast</td>
<td></td>
</tr>
<tr>
<td>labial</td>
<td>UNVOICED</td>
<td></td>
</tr>
<tr>
<td>annex</td>
<td>escrow</td>
<td></td>
</tr>
<tr>
<td>forecast</td>
<td>increase</td>
<td></td>
</tr>
</tbody>
</table>
Figure 16. Typical $F_0$ Contours in Words with Contrasting Stress Patterns, Grouped by Medial Consonant Type. Shown are contours for the words *produce, produce, compact, compact* spoken by talker ASH (a) and GWH (b).
vowels and consonants) in stressed syllables is clearly evident. However, it is also evident that the individual talker’s mode of list reading (most specifically, his rising or falling contour) can affect relative $F_0$ values in stressed and unstressed syllables. In all 120 words, talker ASH produced higher peak $F_0$ in the stressed syllable than in the unstressed syllable of the word. No such completely clean cut distinction was made by talker GWH. This is evident in the scatter plots of Figures 17a (for ASH) and 17b (for GWH), where peak $F_0$ in the second vowel is plotted versus peak $F_0$ in the first vowel. Pairs of peak $F_0$ values cluster into two completely isolatable groups for ASH, corresponding to whether the first or the second vowel is stressed. Any straight line between those two clusters would separate $\hat{V}_1 - V_2$ from $V_1 - \hat{V}_2$ stress patterns with 100% accuracy. The roughly 10% differences in $F_0$ values between the two clusters are presumably perceptible and actually comparable to $F_0$ differences that have been previously shown to be linguistically significant (Lehiste 1970; Mohr 1971). On the other hand, some overlap occurs between peak $F_0$ values for stress patterns $\hat{V}_1 - V_2$ and $V_1 - \hat{V}_2$ for talker GWH, as shown in Figure 17b.

It is interesting that the peak value of fundamental frequency in the first vowel alone separates $V_1 - V_2$ and $V_1 - \hat{V}_2$ stress patterns of both talkers very well. A simple hypothesis (shown by the dashed lines in Figures 17a and 17b) that "peak value of $F_0$ in vowel $V_1$ is greater than 127 Hz if and only if vowel $V_1$ is stressed" yields 100% correct stress classification for ASH, and 96% correct for GWH. The use of any absolute frequency threshold (like 127 Hz) for distinguishing between stress patterns would undoubtedly be futile for talkers with different physiological structures and vocal system sizes, and is certainly not advocated here. However, the success of such an hypothesis here does illustrate that peak $F_0$ values in initial syllables are more reliably related to stress patterns than are values in the utterance-final syllables. We would expect initial syllables to be less affected by confusing influences such as the individual’s mode of list reading (his rising and falling terminal contours) and the sloppiness, reduced amplitudes, and falling subglottal pressure that prevail at the ends of utterances (particularly more extended utterances than the two-syllable ones
Figure 17. Peak $F_0$ Values in the First and Second Vowels of Bisyllabic Words, Showing Clustering Due to Stress Contrasts.
The initial values of $F_0$ in both vowels were also studied. Vowel initiation was determined from onset of formant structure in digital spectrograms, and from energy contours, as outlined previously for the [həCV]C utterances. As with peak $F_0$ values, a simple hypothesis that "initial $F_0$ value in vowel $V_1$ is greater than 120 Hz if and only if vowel $V_1$ is stressed" yielded fairly accurate stress classification (98% for ASH and 82% for GWH). These results are obviously not as good as those with peak $F_0$ values, and are influenced by the identities of previous consonants, as we have learned to expect (see Figures 7 and 8).

The central values of $F_0$ in a voiced consonant $C_q$ (immediately before the second vowel) were also considerably higher when the second syllable was stressed. Thus, a prestressed voiced consonant will have higher $F_0$, just as its associated following vowel does.

5.3. Effects of Stress on $F_0$ Transition Parameters

The above results with intrinsic $F_0$ parameters agree with previous studies showing that $F_0$ tends to be higher in stressed syllables. The effects of stress on the transitions in $F_0$ at consonant-vowel and vowel-consonant boundaries, however, apparently had not been studied until recently (Lea 1972b, chapter 5). Various hypotheses are listed in Table II which relate falling or rising $F_0$ contours at vowel onset to the stress condition of the syllable, the voicing/unvoicing of preceding consonants, and the exceptional condition where vowel $V_1$ is word-initial (as in aster, astir, etc.). Also shown in Table II, for comparison, is the hypothesis used for [həCV]C utterances, that was shown to correctly mark consonant state of voicing by the vowel-initial $F_0$ slope.

Published works that suggest that "pitch is a voicing cue" (Haggard, Ambler and Callow 1970, Chistovich 1969), and the results with [həCV]C utterances, would suggest that voiced consonants are followed by rising $F_0$ contours, while falling $F_0$ contours follow unvoiced consonants. The simple "consonantal hypothesis", listed as hypothesis I in Table II, states these claims. However, the data for the sixty word pairs showed that only about half of the
| HYPOTHESES RELATING $F_0$ CHANGES  
TO STRESS AND CONSONANT STATE OF VOICING |
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>haCVC UTTERANCES:</strong></td>
</tr>
<tr>
<td>RISING $\iff$ PREVIOUS CONSONANT IS VOICED</td>
</tr>
<tr>
<td><strong>STRESS PAIRS:</strong></td>
</tr>
<tr>
<td><strong>CONSONANT HYPOTHESES:</strong></td>
</tr>
<tr>
<td>1. RISING $\iff$ PREVIOUS CONSONANT IS VOICED</td>
</tr>
<tr>
<td>2. RISING $\iff$ PREVIOUS CONSONANT IS VOICED, OR VOWEL IS WORD-INITIAL</td>
</tr>
<tr>
<td><strong>STRESS HYPOTHESIS:</strong></td>
</tr>
<tr>
<td>3. RISING $\iff$ VOWEL IS STRESSED</td>
</tr>
<tr>
<td><strong>STRESS AND CONSONANT HYPOTHESIS:</strong></td>
</tr>
<tr>
<td>4. RISING $\iff$ VOWEL IS STRESSED, OR WORD-INITIAL, OR PREVIOUS CONSONANT IS VOICED</td>
</tr>
</tbody>
</table>
falling [rising] contours result from unvoiced [voiced] consonants. The one exception was with rising contours in second syllables. These were regularly associated with previous voiced consonants (in 100% of the syllables of ASH, and 84% of those of GWH).

Other hypotheses that weaken the implications of a rising $F_0$ contour (number 2), or that consider stress effects alone (number 3) are given in Table 4, but hypotheses based on the combination of stress and voiced environments yielded the most accurate predictions of stress and phonetic context. As hypothesis 4 shows, a falling $F_0$ contour may be evidence either of a preceding unvoiced consonant, or of an unstressed syllable (with previous unvoiced or voiced consonant). A rising $F_0$ contour at vowel onset may indicate (by hypothesis 4) either that the vowel is word-initial or preceded by a voiced consonant, or that the syllable is stressed (with a prevocalic voiced or unvoiced consonant). These reliable hypotheses are, however, predictively quite weak. All that they positively eliminate from consideration are (1) the possibility of a stressed vowel with preceding voiced consonant yielding a falling contour; and (2) the possibility of an unstressed vowel with previous unvoiced consonant yielding a rising contour.

It is evident that falling or rising $F_0$ contours at vowel onset do not simply mark either stress or state of voicing, but are functions of a complex combination of stress and phonetic context. Falling contours are associated with both unstressed vowels and unvoiced consonants. Likewise, rising contours co-occur with preceding voiced consonants or initial vowels, and also with stressed syllables. Since fundamental frequency often falls dramatically when passing from stressed syllables to following unstressed syllables, one might expect the stress-dictated $F_0$ drop to frequently overshadow any tendency for $F_0$ to rise following voiced medial consonants. The rise of $F_0$ at the beginning of a stressed syllable may likewise overshadow initial $F_0$ drops after unvoiced consonants, so that rising contours may occasionally be manifested following unvoiced consonants.
6. CONCLUSIONS AND RELATIONS TO TONAL PHENOMENA

6.1. Conclusions

Fundamental frequency is not independent of the phonetic content of an utterance, whether the language be a tone language or not. Studies reported here have shown that, immediately after unvoiced consonants in [haVC] utterances, \( F_0 \) is high and falling, while \( F_0 \) rises at vowel onset after voiced consonants. Since \( F_0 \) contours were not found to be continuous in all voiced consonants, and since the beginnings of phonemically unvoiced consonants are sometimes voiced, voicing of consonants cannot be detected simply from whether or not the \( F_0 \) contour is continuous. However, several secondary features of \( F_0 \) contours were found to provide cues to the consonant voicing/unvoicing feature. Peak \( F_0 \) in a vowel tends to be higher in the context of unvoiced consonants, and the initial \( F_0 \) value at vowel onset is even more reliably associated with consonant state of voicing. For [haC\(_1\)VC\(_2\)] utterances, if \( F_0 \) rises in the first ten centiseconds of the stressed vowel, the consonant \( C_1 \) is very likely to be voiced; conversely, a falling \( F_0 \) in the initial part of the vowel marks \( C_1 \) as unvoiced.

The rising or falling of \( F_0 \) in the initial part of a vowel is more pronounced for vowels articulated with a high tongue position. Initial and peak \( F_0 \) values in the vowel also tend to be higher for high vowels.

Manner of consonant articulation affects \( F_0 \) values primarily through the dip of \( F_0 \) during voiced obstruents, which does not occur for sonorants. Place of consonant articulation shows little or no significant systematic effects on \( F_0 \) contours.

Other stress patterns seriously complicate these interactions between phonetic effects and \( F_0 \) parameters. From the study of bisyllabic word pairs with contrasting stress patterns, it was apparent that falling or rising \( F_0 \) contours at vowel onset do not simply mark either stress or state of voicing, but are functions of a complex combination of stress and phonetic context. The results for stress pairs showed that it is unlikely that a stressed vowel preceded by a voiced consonant will yield a falling contour, and un-
likely that an unstressed vowel preceded by an unvoiced consonant will yield a rising contour. But, other combinations of stress, state of voicing, and rising or falling contours are apparently not readily distinguished from one another from $F_0$ contours alone. A rising $F_0$ contour at vowel onset may indicate either that the vowel is word-initial or preceded by a voiced consonant, or that the syllable is stressed (with either a voiced or unvoiced prevocalic consonant). A falling $F_0$ contour in the initial part of a vowel may indicate a preceding unvoiced consonant or that the syllable is unstressed.

In both studies of isolated words, it was evident that $F_0$ rises and dips of around 10% may occur at vowel-consonant and consonant-vowel boundaries. These $F_0$ changes were the main source of false boundaries in the constituent boundary detection program.

In the 800 seconds of connected texts, conversational excerpts, and computer instructions studied, over 80% of all boundaries between major syntactic constituents were accompanied by substantial (greater than 7%) fall-rise $F_0$ patterns. About half of those boundaries that were not accompanied by fall-rise $F_0$ valleys were NP-Verbal boundaries. When these were neglected, about 90% of all other boundaries between major syntactic constituents were accompanied by substantial $F_0$ valleys. Over 90% of all boundaries between sentences, whether embedded or matrix, were accompanied by long (greater than 35 csec) periods of unvoicing ("pauses"). Also, all sentence boundaries were accompanied by fall-rise valleys, and almost all sentence boundaries were followed by $F_0$ rises of 40% or more. Boundaries between matrix sentences generally showed longer unvoiced periods and larger $F_0$ rises than those between embedded S's.

The first stressed syllable of a constituent was assumed to be associated with the $F_0$ rise following a detected boundary. Other stressed syllables were assumed to be associated with local increases in $F_0$ above an archetype falling $F_0$ contour following the first stressed syllable in the constituent. An algorithm based on these assumptions successfully located 85% of all syllables perceived as stressed by the majority of a panel of listeners. The boundaries of such stressed syllables were identified with dips in speech energy.
The total $F_0$ contour of a sentence may thus be decomposed into four general parts: the overall intonation contours of clause or sentence types; the division of sentences or clauses into constituents with archetype (fast-rise, gradual-fall) contours; the local increases in $F_0$ associated with stressed syllables; and, most localized of all, the effects of vowels, consonants, and phonetic transitions on relative $F_0$ values and $F_0$ local increases and decreases.

It may be hoped that interested readers can translate many of these individual effects on $F_0$ contours into implications for the description of tone language phenomena, and implications for explaining historical development and changes in tone. Here some preliminary suggestions can be given.

6.2. Tonal Influences on Perceived Consonants

One might conjecture that tone operates in tone languages as a suprasegmental feature much like stress does in English. Then, the high $F_0$ intended on a high-tone syllabic nucleus might be expected to be accompanied by a preceding increase in $F_0$ at the vowel onset, regardless of the unvoiced or voiced state of the preceding consonant. This rising $F_0$ preceding towards the center of the high tone would thus tend to make the preceding consonant appear to be more like a voiced consonant. The listener to a sequence of unvoiced consonant followed by high tone might begin to perceive the prevocalic consonant as voiced. Historical change may then yield a decision that the preceding consonant is phonemically voiced.

Similarly, the falling $F_0$ preceding a low tone may lead to perception of the preceding consonant as unvoiced, regardless of its intended phonemic state of voicing.

Such conjectures are based on assuming that the assignment of tones is independent, or "prior to", the determination of phonemic structure. Thus, tones are assigned to a syllable and then their effects on $F_0$ contours are used to suggest possible reinterpretations of state of voicing of preceding consonants. However, alternative conjectures are possible within this same viewpoint. If the high versus low $F_0$ in the vowel center, rather than the slope of $F_0$ in the consonant-vowel transition, is assumed to affect the
perceived state of voicing of the preceding consonant, we might expect high tones to yield interpretation of the preceding consonant as unvoiced. This is because high $F_0$ usually accompanies unvoiced environments, and lower $F_0$ accompanies voiced environments. However, this alternative conjecture is less reasonable to adopt, since there is less evidence that high $F_0$ in a vowel significantly affects the perceived state of voicing of the preceding consonant than that the initial slope of $F_0$ in the vowel affects the perceived state of voicing.

Continuing the line of thought that suggests that rising [or falling] $F_0$ into a high [low] tone yields an interpretation of the preceding consonant as voiced [unvoiced], we might suggest one mechanism by which tonal phenomena might disappear in a language. Since high tones would become more associated with preceding voiced consonants, a redundancy sets in whereby the state of voicing also marks the intended tone. If tone were then dropped, the contrast would still be maintained by the state of voicing of the preceding consonant.

6.3. **Consonantal Influences on Perceived Tones**

Rather than assuming that tones are independently assigned and their effects on $F_0$ contours force reinterpretations of preceding consonants, we may reverse the process and assume that consonant sequences are first assigned and then their effects on $F_0$ contours affect the relative likelihood of interpreting the following vowel or syllabic nucleus as having a high versus low tone. Since $F_0$ rises after voiced consonants, the following tone may appear high since the slope into it is upward. Similarly, the fall after an unvoiced consonant may lead to an interpretation of the following tone as low. This set of conjectures would be based on the assumption that the perceived slope in $F_0$ into a vowel would affect the perception of the high/low tone on the following vowel (or syllabic nucleus). However, a more reasonable conjecture would be that the relative $F_0$ values **within** the vowel (and not the preceding slope into the vowel) affect its interpretation as a high versus low tone. Then, since $F_0$ is higher in vowels with preceding unvoiced consonants, and lower for preceding voiced con-
sonants, one might expect high tones to occur with unvoiced consonants and low tones to occur with voiced consonants. Larry Hyman's (1973) rules for tonal assimilation do indeed suggest that tone raising accompanies voiceless consonants, and tone lowering accompanies voiced obstruents.

Extending this latter line of thought that asserts that unvoiced consonants lead to high tones and voiced consonants lead to low tones, we might suggest one mechanism by which tonal phenomena would appear in a language which had none. Since unvoiced [versus voiced] consonants yield perceived high [versus low] F₀, high [versus low] tones soon redundantly mark the state of voicing of the preceding consonant. Then voicing contrasts might be lost, and the tonal differences would still mark the intended lexical differences (cf. Matisoff 1973, Maran 1973).

In summary, we may hypothesize that tonal contrasts may enter a language as vowels in unvoiced environments become high in tone and those in voiced environments are lowered in tone. Then, the voicing contrast may or may not be lost, so that voiced consonants or unvoiced consonants may be before both high and low tones. Later, unvoiced consonants before high tones may be perceived as voiced if the F₀ slope into the following vowel is positive, while voiced consonants before low tones appear unvoiced due to falling F₀ slopes into the vowel. Also, at this stage, the high [or low] tones after voiced [or unvoiced] consonants are redundant, and might be dropped. In passing through such genesis and exodus of tonal contrasts, the language may thus undergo increased and decreased attention to state of consonant voicing, and actual reversals of state of voicing may result. This would thus be one example of how the consonantal variety in a language may progressively decrease as the tonal system progressively complicates, and vice versa.

6.4. Other Segmental and Suprasegmental Influences on Tone

Similar patterning and historical change may be hypothesized for contrasts between falling and rising tones. If the phonetically-dictated slope in the initial portion of a vowel was extended to make phonemic (or "extrinsic") distinctions between falling ver-
sus rising tones, we might expect falling tones to usually accompany preceding unvoiced consonants and rising tones to accompany preceding voiced consonants.

Similar arguments about "tonogenesis" and "tonoexodus" might be made for any major \( F_0 \) effects due to manner of consonant articulation (such as sonorant versus obstruent contrast) or suprasegmental influences such as constituent structure and positions in total intonation contours. For example, the general tendency toward reduced \( F_0 \) on the progressively later high tones in a long utterance (so-called "tonal terracing" or "downdrifting") may be associated with the generally falling \( F_0 \) of overall Tone I contours. Also, words that occur in certain syntactic positions, such as constituent-final or sentence-final, might experience pressures to be interpreted as high-tone if \( F_0 \) is generally high in that structural position, or low-tone if \( F_0 \) is generally low in that structural position. If tone and stress contrasts both exist in a language at some point in its development, one would expect interactions which force the interpretation of one type of contrast as becoming a perceived contrast of the other type, depending upon which contrast is growing in its importance in the language at that period.

Another type of tonal phenomenon is suggested by the lack of exact time placement that was found to occur for suprasegmental phenomena such as constituent boundaries. Boundaries were found to be marked in the general region of a syntactic break, but not at the strict point in the flow of speech that corresponds to the division between the phonetic strings associated with the two surrounding constituents. In similar fashion, we might expect tone, as another suprasegmental factor, to not strictly adhere in time to a single syllable. There might be a lack of synchrony between syllables and pitch contours or tones, so that a tone might spread to, or overlap with, surrounding syllables, yielding the form of horizontal tonal assimilation hypothesized by Larry Hyman (1973). Thus, for example, a previous high tone in an utterance might spread into the first part (or, ultimately, all) of the following syllable, due to non-synchrony between tones and phonetic segments. Hyman hypothesized that voiced obstruents (due to their inherent low \( F_0 \)) may block the spread of a high tone, while voice-
less consonants (with their inherent high F₀) may block the spread of a low tone. The studies of inherent F₀ values, and perceptually-significant transitions in F₀ at consonant-vowel boundaries, as discussed in this paper, thus reinforce such hypotheses.

All of these conjectures concerning tonal phenomena as presented here in section 6 are highly speculative, but are fortunately open to explicit testing with extensive data on tonogenesis and tonoexodus from the historical linguists. The primary purpose of their inclusion here is to illustrate some ways in which knowledge of the suprasegmental and segmental influences on F₀ contours may help guide the descriptive and historical linguists to productive hypotheses about tonal phenomena.

NOTES

¹ Decreases in F₀ are sometimes the mark of stressed syllables in non-neutral semantically-marked expressions such as for some emotions and for special emphasis. For relevant studies, see Pike (1945), Bolinger (1958), or Lieberman (1967).

² Timing information for the clauses, constituents, words and phones must also be given. Lengths of constituents and clauses, for example, will depend upon the number of words or syllables, rhythm effects, and perhaps the relative importance of each constituent in the sentence structure. In Figure 4, we shall assume that such timing information is provided, and consider only how the properly timed F₀ changes accompanying events of phonetic sequence, stress patterns, and syntactic structure combine to yield the overall F₀ contours.

³ The initial value F(i) of F₀ in V₂ presents a convenient reference point for measuring the rise or fall in F₀ at the beginning of V₂, and for determining the difference between F₀ of the medial consonant C₁ and the following stressed vowel V₂. When unvoiced segments occur throughout the medial consonant, the time t₁ corresponding to F(i) is set at the time when the pitch tracker first gives a value of F₀ after unvoicing. If the F₀ contour is continuous, as in some voiced fricatives and sonorants, the time t₁ is selected as the time segment at which vowel formant structure is first evident on the digittal spectrogram, or, whenever that spectral change is not abrupt, at that time where F₀ and low-frequency energy jump suddenly from lower values within the consonant to higher values in the vowel (Lea 1972b:99). These judgments of vowel onset are based on general segmentation techniques hypothesized in other studies (cf. Fant 1960:207 and Chen 1970:131).
REFERENCES


Lea, W. A. 1973b. Use of prosodic features to segment continuous speech into sentences and phrases. Univac Park, St. Paul, Minn. (To be published under the title "An approach to syntactic recognition without phonemics" in the IEEE Transactions on Audio and Electroacoustics, June, 1973.)


TONOGENESIS IN SOUTHEAST ASIA

BY

JAMES A. MATISOFF

DEPARTMENT OF LINGUISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720

Published by the Linguistics Program
University of Southern California
Los Angeles, California 90007
In the Beginning was the Sino-Tibetan monosyllable, arrayed in its full consonantal and vocalic splendor. And the syllable was without tone and devoid of pitch. And monotony was on the face of the mora. And the Spirit of Change hovered over the segments flanking the syllabic nucleus.

And Change said, "Let the consonants guarding the vowel to the left and the right contribute some of their phonetic features to the vowel in the name of selfless intersegmental love, even if the consonants thereby be themselves diminished and lose some of their own substance. For their decay or loss will be the sacrifice through which Tone will be brought into the world, that linguists in some future time may rejoice."

And it was so. And the Language saw that it was good, and gradually began to exploit tonal differences for distinguishing utterances -- yea, even bending them to morphological ends. And the tones were fruitful and multiplied, and diffused from tongue to tongue in the Babel of Southeast Asia.

***

1.0. INTRODUCTION

The languages of Southeast Asia, some of which are fully tonal, others of which are only marginally or incipiently tonal, and some of which are not tonal at all, constitute an ideal terrain for the investigation of the mechanism of "tonogenesis".¹

This paper is organized as follows. First come some introductory remarks on the role of laryngeal final consonants and syllable-initial voicing vs. voicelessness in the generation of tonal phenomena (1.1); then a discussion of the interrelationship among monosyllabicity, intersegmental feature-sharing, and compensatory tone (1.2). In the next section we give a brief overview of the present state of our knowledge about the tonal situation at the Proto-Sino-Tibetan (PST) and Proto-Tibeto-Burman (PTB) levels (2.1), followed by some thoughts on the areal diffusion of tones in SE Asia and the utility of tone-systems for the establishment of genetic relationship among languages (2.2).
1.1. Laryngeal States and Tonal Effects

Twenty years ago the French botanist and Orientalist André Haudricourt wrote a classic article\(^2\) which addressed itself to the problem of how standard Hanoi Vietnamese acquired its six tones. This question had a vital bearing on the genetic affiliation of Vietnamese—previous scholars had held that Vietnamese belonged in the Tai family rather than in the Mon-Khmer (M-K) group,\(^3\) largely because the Tai languages are tonal while the Mon-Khmer languages are not.\(^4\) Haudricourt succeeded in demonstrating that the tones of Vietnamese were secondary developments arising from a breakdown of the system of consonantal oppositions at the beginning and the end of the Mon-Khmer syllable. The proto-language had syllables with final segments of three significant types: those ending in an open vowel or nasal (i.e. with no laryngeal final segment); those ending in voiceless spirants, \(^*s\) or \(^*s\), which had reduced to \(-h\) by pre-Vietnamese times; and those ending in some sort of stop\(^5\) which had reduced to glottal stop by the pre-Vietnamese period. In addition the language had a voiced/voiceless distinction for its syllable-initial consonants. See Figure 1.

FIGURE 1. Vietnamese A (beginning of Christian era)

---

By the sixth century, final \(-h\) and \(-?\) had disappeared, leaving in their wake a compensatory falling and rising effect (respectively) on the pitch of the preceding vowel. See Figure 2. At this point the language had a three-tone system, which apparently remained stable as long as the voiced/voiceless opposition for initial consonants remained in force. But by the 12th century, the old
FIGURE 2. Vietnamese B (sixth century)

<table>
<thead>
<tr>
<th>MID</th>
<th>FALLING</th>
<th>RISING</th>
</tr>
</thead>
<tbody>
<tr>
<td>pə</td>
<td>pə̀</td>
<td>pə́</td>
</tr>
<tr>
<td>bə</td>
<td>bə̀</td>
<td>bə́</td>
</tr>
</tbody>
</table>

THREE TONES:
exeunt laryngeal C_f's/enter "contours"

voiced series had merged with the voiceless series. The language responded to this threat to its contrastive power by doubling the number of tones from three to six; the three tones descending from syllables with *voiced initials were then distinctively lower in pitch than the three which derived from syllables with *voiceless C_i's. See Figure 3.

FIGURE 3. Vietnamese C (twelfth century)

<table>
<thead>
<tr>
<th>HIGHER</th>
<th>pə &quot;ngang&quot;</th>
<th>pə̀ &quot;họi&quot;</th>
<th>pə́ &quot;sắc&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOWER</td>
<td>pə̀ &quot;huyễn&quot;</td>
<td>pə &quot;ngã&quot;</td>
<td>pə̀ &quot;nảng&quot;</td>
</tr>
</tbody>
</table>

SIX TONES:
exeunt voiced C_i's/enter "registers"

This explanation—which has gone unchallenged by subsequent scholars—presupposes the existence of certain universal phonetic mechanisms which interrelate articulatory gestures of the larynx with the production of audible tonal effects. (a) Laryngeal C_f's affect the contour of the preceding vowel's pitch, with -h acting as a pitch depressor (i.e. leading to falling tones) and final - trong having the opposite effect (leading to rising tones). The exact physiological causes of these effects are being worked out in de-
tail by experimental phoneticians (see elsewhere in this volume), but Haudricourt's impressionistic explanation still seems generally valid: the pitch-drop before -h is due to a "relâchement brusque du larynx", while the pitch-rise before -h is caused by an "augmentation de la tension des cordes vocales". (b) Syllable-initial consonants merely affect the register of the following vowel, with voiced C₁'s provoking lower pitch and voiceless C₁'s provoking higher pitch. Again the physiological explanation for this fact involves a complex interplay of aerodynamic and articulatory factors, but one thing seems clear: this is a universal phenomenon which obtains even in languages like English which would never dream of exploiting such redundant pitch-differences for contrastive purposes.

In a 1968 talk I roughly characterized the two basic contrasting "laryngeal attitudes" as shown in Figure 4.

**FIGURE 4. Laryngeal Attitudes**

<table>
<thead>
<tr>
<th>TENSE-LARYNX SYNDROME</th>
<th>LAX-LARYNX SYNDROME</th>
</tr>
</thead>
<tbody>
<tr>
<td>higher pitch/rising contour association with -h</td>
<td>lower pitch/falling contour association with -h</td>
</tr>
<tr>
<td>voicelessness</td>
<td>voicedness, breathiness</td>
</tr>
<tr>
<td>retracted tongue-root</td>
<td>advanced tongue root</td>
</tr>
<tr>
<td>&quot;creaky&quot; laryngeal turbulence</td>
<td>&quot;rasping&quot; laryngeal turbul-</td>
</tr>
<tr>
<td>larynx tense and/or raised=</td>
<td>larynx lax and/or lowered=</td>
</tr>
<tr>
<td>reduced supraglottal cavity</td>
<td>distended supraglottal cavity</td>
</tr>
</tbody>
</table>

More recently La Raw Maran (1971) has persuasively proposed a small set of binary distinctive features which are intended to capture simultaneously not only the role of the larynx in the production of voiced obstruents, h, and glottal stop, but also the concomitant tonal effects on adjacent vowels. Similarly motivated features (spread vs. constricted glottis and slack vs. stiff cords) have been adopted by Halle and Stevens (1971), and are now being widely discussed by generative phonologists.
Despite the complexity of the simultaneous bundles of articulatory activities which go to make up the "tense" vs. "lax" syndromes, it seems clear that the syndromes as a whole do stand in a binary opposition to each other. Otherwise how are we to understand the oft-noted diachronic phenomenon of tonal "flip-flops" whereby a high tone and a low tone abruptly switch places, so that the *high becomes low, and the *low becomes high? Some sort of "alpha-reversal" of laryngeal gesture must be assumed.

Maran (1971) has noted that in Jinghpaw, a Tibeto-Burman language which he speaks natively, syllable-final stops (-p, -t, -k, -?) are voiceless under the high-tone, but voiced (-b, -d, -g, -?) under the low tone, inferring from this that it is the voicing contrast which is distinctive here, with the tonal difference being redundant. For several reasons I prefer to interpret the situation in the opposite sense. It seems to me that voicing/voicelessness has a causational effect on the tone of the adjacent vowel only in syllable-initial position. In syllable-final position the voicing or voicelessness of a consonant (at least in Tibeto-Burman) is rather the automatic, redundant consequence of a pre-existent tonal opposition, not its cause. For both -h and -? are voiceless (though in rather different ways)—yet they have opposite tonal effects in syllable-final position.

1.2. Monosyllabiciticy, Intersegmental Feature-sharing, and Compensatory tone

If the laryngeal mechanisms we have been considering are really universal, why haven't all human languages been tonal at some point in their history, like Chinese, Burmese, or Jinghpaw? Some language families seem more hospitable to the development of tones than others, and the same goes for geographic areas of the world. It is as if the seeds of tone potential required a particularly fertile soil of a certain structural type in order to take root and flourish. In particular, it appears that to become truly tonal a language must have a basically monosyllabic structure (i.e. the morphemes must be only one syllable long). Polysyllabic languages like Japanese, Swedish, or Serbo-Croatian may develop "pitch-accent" systems, but these differ from true tone-systems in many important respects.
There is something about the tightly structured nature of the syllable in monosyllabic languages which favors the shift in contrastive function from one phonological feature of the syllable to another. The Tibeto-Burman (TB) languages have always been monosyllabic. The proto-monosyllable was quite complex in structure: the initial consonant could be preceded by a variety of prefixes (or even by a sequence of two prefixes) and followed by one of four glides (\(-w-, -y-, -r-, -l-\)). The vowel could be followed by any of a number of final nasals, stops, liquids, or \(-s\), or even by a nasal or stop plus \(-s\). Written Tibetan (WT) may be taken to preserve the proto-syllable canon faithfully, with maximally complex forms like brgyad 'eight', brnyabs 'diligence', bsnyigs 'sediment'. Written Burmese (WB) syllables may have initial consonant clusters of up to three members, but no more than a single consonant in final position: mrwe 'snake', krwat 'leech', krwak 'rat'. We may symbolize the proto-syllable canon as follows:

\[(P_1) \ (P_2) \ C_1 \ (G) \ V \ (') \ (C_f) \ (S)\],

where \(P\) = prefix, \(C_1\) = initial consonant, \(G\) = glide, \(V\) = vowel, \(\prime\) = vowel length, \(C_f\) = final consonant, and \(S\) = suffixial \(-s\). When we look at the phonological changes which these richly complex syllables have undergone through time (e.g. from WT to modern colloquial Lhasa Tibetan, or from WB to modern Rangoon Burmese, or from Proto Lolo-Burmese to Lahu), we find that the different parts of the syllable have constantly been influencing each other: the prefixes affect the root-initial consonant, as do the glides; the glides also affect the vowel, as do the final consonants; the vowel itself affects the preceding and following consonants, etc. It thus makes little sense to ask questions like "What happens to the Proto-Tibeto-Burman (PTB) vowel *\(a\) in language X?" Rather we must specify the syllabic environment more precisely: "What is the PTB reflex of *\(-am\), or *\(-ak\), or *\(-wa\), or *\(-ya\), or *\(-yaŋ\) in language X?" Thus, Proto-Lolo-Burmese (PLB) *\(-a\) develops into Lahu -ö in syllables without a \(G\) or a \(C_f\); but *\(-ya\) becomes -e, and *\(-wa\) becomes -o. A nasal or stop following the PLB nuclear vowel *\(a\)- detemined different Lahu reflexes for each point of articulation. Thus PLB *\(-am\) > Lh. -ö, but *\(-an\) > -e, and *\(aŋ\) > -o; *\(-ak\) > aʊ,
but *-at > -eʔ and *-ap > -oʔ. It is for this reason that Sino-
Tibetanists traditionally lump the vocalic nucleus together with
any post-vocalic consonants the syllable may have, and refer to
this complex, well-integrated entity as the "rhyme" of the syllable.

So tightly interdependent are these neighboring vowels and
consonants, that certain phonetic features seem to have bounced
back and forth from vowel to consonant and back again through the
history of the TB languages. The fate of the PTB rhyme *
-ik in
Burmese is a good case in point. By the time Burmese was committed
to writing in the 12th century, older *-ik had become -ac (e.g.
PTB *tsik 'joint', WT tshigs, WB chac); that is, the palatality of
the vowel had been transferred to the $C_f$, so that the latter changed
from a velar stop to a palatal affricate, thereby depalatalizing
the vowel from *i to a. What is remarkable is that this develop-
ment was then completely reversed between the Old Burmese period
and the modern standard Rangoon dialect, so that words written with
-ac are now pronounced with the rhyme -Iʔ (Mod.Bs. hsIʔ 'joint').
That is, the palatality has been shifted back again from the $C_f$ to
its "original" vocalic home!

This leads us to the key question: did this complex proto-
monosyllable already carry a lexically distinctive tone? The an-
swer is far from clear at the moment.16 What does seem certain is
that, given the intimate relationship between consonantal and vo-
calic features of the TB syllable, there must have been phonetic
perturbations of the pitch of vowels due to the influence of neigh-
boring consonants throughout the history of the family. However,
as long as the consonants maintained themselves in a good state of
preservation, such pitch-differences as existed were likely to have
remained subphonemic--predictable, automatic, redundant. It was
only when the old consonantal system had decayed through cluster
simplification, losses, mergers that the daughter languages were
forced to exploit those pitch-differences for contrastive purposes.

Initial consonants "decay" rather differently from final ones.17
At the beginning of the syllable, the prefixes generally found
themselves in a weak position, sometimes fusing with the root-ini-
tial and often dropping entirely (see Matisoff 1972c). Before
their departure, however, they were likely to have affected the
voicing or voicelessness of the root-initial consonant. Thus the glottal prefix *ʔ- or *ʔe- typically devoiced a following sonant, while the nasal prefix *N- often voiced a following surd. Yet it is noteworthy that the basic TB *voiced/*voiceless opposition in root-initial position was everywhere preserved systematically, even though the phonetic nature of the contrast changed in many languages (like Burmese and Lahu) from voiced/voiceless to voiceless unaspirated/voiceless aspirated.

In syllable-final position there is a whole continuum of consonantal decay for final nasals and stops. The three-way contrast among *-m, *-n, and *-ŋ was sometimes reduced to a two-way one, but without the remaining nasals losing their point of occlusion. (This is what happened in Mandarin Chinese, where *-m and *-n merged to -n.) At more advanced stages of decay, one or more of the nasals could lose their point of articulation, so that the feature of nasality shifted back onto the vowel, yielding a new type of oral/nasal contrast for vowels. At the ultimate stage, the nasal feature disappears altogether from the syllable; but in this case the vowel quality itself has usually already been altered differentially by the particular nasal which had followed it, so that the language does not necessarily suffer a loss of contrast.21

Final stops may undergo even more finely graded degrees of attrition than the nasals. The three-way proto-contrast among *-p, *-t, and *-k could be reduced to a two-way contrast without the remaining stops losing their buccal occlusion. More radically, one or more of the stops could be reduced to *—a glottal stop might be termed the "minimal stop" from this point of view. At more advanced stages the final consonant may disappear entirely, after transferring its occlusion back onto the vowel, so that the vowel has "laryngeal constriction" or "creakiness". At a still further stage even this constriction may disappear, and the only trace of the former Cf may be a tenseness in the vowel, or some other alteration in the vowel quality.22 The smile of the Cheshire cat, fading away imperceptibly.

Looking at the TB family as a whole, we find that the details of consonantal decay differ considerably from subgroup to subgroup
and from language to language, but one important generalization holds: the better-preserved the consonantal system, the fewer the vowels and the fewer the tones; the more vestigial the consonant system, the more proliferation of vowels and tones.

2.1. Redundant and Contrastive Tone in PST and PTB

Paul K. Benedict (1972a, 1972b, 1973a) has argued persuasively that even back at the remote Proto-Sino-Tibetan period the proto-language had a "phonemic" two-way tone-contrast in non-stopped syllables (though syllables whose \( J_f \) was a stop had no distinctive tone). Benedict bases his argument mainly on evidence from Chinese on the one hand, and from certain subgroups of TB on the other: Lolo-Burmese, Karen, and Nungish. Karen is extremely aberrant from the other TB languages from the grammatical point of view (for one thing, the Karen object comes after the verb instead of before it); so much so that one is tempted to set up a higher-level taxonomic group "Tibeto-Karen" comprising Karen on the one hand and "Tibeto-Burman proper" on the other. Yet as Benedict has shown, the four tones of Karen correspond systematically to the two main non-stopped tones of Lolo-Burmese, in a simple, straightforward way. Two explanations are therefore possible: either the two-way tone-contrast must be placed at least as far back as the remote Tibeto-Karen period (and thus a fortiori at the PTB period), or else the tone-system of Lolo-Burmese somehow "diffused" into the Karen languages (see next section). Benedict rejects the diffusion hypothesis (see note 40), and goes on to show that the Lolo-Burmese/Karen/Nungish\(^{23}\) two-tone system can be systematically related to the two principal non-stopped tones of Chinese, the level tone (p'ing sheng) and the rising tone (shang sheng).\(^{24}\) He therefore projects the two-way tone-system back to the PST period itself.

A serious objection to Benedict's theory is the fact that the oldest attested TB language, Written Tibetan, shows no evidence of tonal distinctions at all. In fact some modern dialects of Tibetan, like Balti\(^{25}\), don't have tone either, or at any rate do not have fully developed tone-systems like Lolo-Burmese. (Significantly it is those dialects, like Balti and Purik, which preserve the WT syllable-initial consonants the best that have non-existent
or rudimentary tones, while those dialects, like that of Lhasa, which have a degenerate consonantism, have developed relatively complex tone-systems that are of demonstrably recent origin.26) Are we then to suppose that the original PTB two-way tonal contrast was lost in Tibetan before the language was committed to writing (around the 7th century), so that the language got along without phonemic tones for centuries, only to reacquire it in certain dialects in quite recent times? Given the cyclical nature of TB phonological developments27 this is not as far-fetched as it might sound.

Many other modern TB languages lack well-developed tonal systems, including most members of the huge and ramified Kuki-Chin-Naga family,28 as well as the Barish or Bodo-Garo group. Significantly these languages are spoken at the Western extremity of the TB family, in Assam and Western Burma. Here Benedict is willing to use areal diffusion as an explanation, accounting for the lack of tones as being due to the influence of the non-tonal languages (Indo-European and other) with which these Westerners came in contact.

A particular problem is posed by the extremely important Jinghpaw language (Kachinic group of TB). Although Jinghpaw (Jg.) is quite close to Lolo-Burmese as far as the number of shared cognates is concerned, it is very hard to relate the Jg. tones systematically to those of LB—except, paradoxically, in stopped syllables.29

From the foregoing it should be obvious that we are still far from being able to give a clearcut answer to the question "Did the PST or PTB proto-syllable carry a contrastive tone?" Indeed, I personally believe that the question is rather meaningless when posed in these terms. For I view the whole process of tone-birth and tone-decay as a cyclical one, that has no beginning and has no end. A language or language-family that has a predisposition (in the sense of 2.1 above) to develop tones will indulge this predisposition at certain points in its history, but not in others, depending on the total vowel-consonant dynamics of the syllable at a given point in time. Thus we may imagine a hypothetical language at Stage A: it is monosyllabic, but the number of possible sylla-
bles is very large, since there is a rich system of syllable-initi-
tial and -final consonants. Grammatical information is carried by
a number of non-syllabic affixes attached to both ends of the syll-
able. Different syllables have different pitches, but the lan-
guage can afford to ignore this fact, since it is having no trouble
keeping its utterances apart.

Time passes, and the language enters a new phase, Stage B:
its initial- and final-consonantal systems are breaking down. Af-
fixes are dropping or being absorbed into their root-morphemes.
Homophony rears its ugly head. In desperation the language casts
about for ways to protect its contrasts. Although each morpheme
is still monosyllabic, the language now creates bisyllabic or even
trisyllabic compounds in order to disambiguate homophones or near-
homophones, so that the word is no longer monosyllabic. At the
same time, "analytical" ways of signalling grammatical relationships
are found. Instead of, e.g., a causative prefix s-, the language
might use a separate auxiliary verb meaning "make" or "send on an
errand" to convey the concept of causation. Meanwhile the number
of vowels has increased and lexically contrastive tones have arisen,
exploiting the previously redundant pitch-differences among syllables.

More time passes, and the language enters Stage C. Human lazi-
ness being what it is, some of the syllables in compounds are ten-
ding more and more to be pronounced laxly, slurred over. Vowels
are losing their stress all over the place, and being reduced to
shwa. These unstressed syllables also lose their tone, and tend
increasingly to hitch themselves onto the adjacent syllable in the
compound. The compounds are becoming "opaque", unanalyzable by
the native speaker (cf. Eng. housewife > hussy). The same sort
of thing is happening to grammatical morphemes like particles and
auxiliary verbs; instead of maintaining their identity as separate
words, they are fusing themselves with root-morphemes (cf. English
gonna, wanna, oughta, etc.). The language is becoming synthetic
again, and developing all kinds of new consonant clusters due to
the fusion of once-separate syllables. Most of the old affixes
left over from Stage A have long since disappeared, making way for
a new crop, though enough of the old crop still remain to confuse
the picture. The nouveau riche consonantism of the syllable is making it less and less necessary to use the tones for contrastive purposes. Vowel-contrasts are weakening in certain areas. The language is becoming monosyllabic again.

And so it goes. Plus ça change, plus c'est la même chose.

2.2. The Areal Diffusability of Tones and the "Southeast Asian Tonbund".

Generations of scholars have puzzled over the genetic interrelationships of the hundreds of languages spoken in mainland and insular Southeast Asia. This is not the place to attempt to recapitulate the various arguments that have been advanced to justify one or another classificatory scheme. Let us rather accept as a basis for discussion the classification worked out by Benedict during the thirty years he has been studying the languages of the area (see especially Benedict 1972a and 1973c). According to his scheme there are only three great linguistic superstocks in the area1: Sino-Tibetan (ST), Austro-Thai (AT), and Austro-Asiatic (AA). See Figure 5.

FIGURE 5. The Three Superstocks

(a) SINO-TIBETAN [monosyllabic; tonal]

Sinitic [Chinese]

Tibeto-Karen

Karenic

Tibeto-Burman

Kuki-Chin-
Naga

Himalay-
ish

Kachinic

Lolo-
Burmese

Homeland: eastern Tibet/western Szechwan, Yunnan; headwaters of the Yangtze, Brahmaputra, Irrawaddy, and Mekong Rivers
The indigenous inhabitants of mainland SEA are thought to have been the AA peoples. At a very early date the Austronesian branch of the AT peoples pushed southward, eventually leaving the mainland and settling on the island chains of the South Pacific. Later came the Tai peoples, whose southward invasion split the Mon-Khmer speech community in two. Some Tai communities remained behind in China, as have the Miao-Yao peoples until very recent times. The last intruders were the Tibeto-Karen peoples, who fanned out southward into Assam and Burma, and in very recent times as far as Thailand and Laos.

Of these three linguistic stocks, only Sino-Tibetan is thought to have been "intrinsically tonal" (with the qualifications expressed above in section 2.1). Proto-AT, as reconstructed by
Benedict (1973c) was devoid of tone, and had polysyllabic (often trisyllabic) root-morphemes. This polysyllabic structure is still characteristic of the Austronesian (AN) branch, and AN has remained without true tones to the present day. The Tai and Miao-Yao (M-Y) branches, however, have become monosyllabic, and have developed complex tonal systems of the Sino-Tibetan type. Proto-AA had what one might call a "sesquisyllabic" structure, with morphemes that were "a syllable and a half" in length. That is, the prevocalic consonant was often preceded by a "pre-initial" consonant, as in the modern Cambodian words psaa 'market', tkiam 'jaw', ckar 'dog', knaok 'peacock'. Unlike the ST prefixes, which tended to be unstable and easily lost, these pre-initials are well-preserved in Mon-Khmer. The Mon-Khmer languages have not quite developed true tone-systems in the ST sense, but rather an intermediate sort of two-way articulatory opposition in which pitch-difference plays a role but is not the only distinguishing factor. This phenomenon has been termed "register" (Henderson 1952). Syllables in the "high" or "head" register have a creaky pharyngealized quality, are pronounced with a tense larynx and retracted tongue-root, and are relatively high in pitch. Syllables in the "low" or "chest" register have a breathy laryngealized, "sepulchral" quality, are pronounced with a lax larynx and an advanced tongue-root, and are relatively low in pitch. See Figure 4 above. Other differences in vowel quality (i.e. tongue-higher vs. tongue-lower, tongue-fronter vs. backer, or monophthongal vs. diphthongal) also accompany the register difference. In fact, the perturbations in vowel quality have been so great, and the number of distinct vocalic nuclei has multiplied to such an extent in these languages that the simplest "phonemic solution" is to recognize these latter phonetic differences as the distinctive features distinguishing the high vs. low registers. The pitch difference is secondary—the languages are not truly tonal in the ST sense. Perhaps we could say that the Mon-Khmer languages escaped the fate of becoming tone languages by the expedient of multiplying their vocalic nuclei. It is perhaps no accident that these "halfway tonal" languages also have a syllabic structure intermediate between the truly monosyllabic ST and the truly polysyllabic AA types.
If the genetic picture outlined above is at all accurate, we must still offer an explanation for the acquisition of true tonal systems by the Tai and Miao-Yao languages (which derive from the atonal Austro-Thai parent stock), as well as by Vietnamese (from the only semi-tonal Austro-Asiatic stock). (While we're at it, we should also account for the fact that many western Austronesian languages (like Javanese) have acquired register systems.) The only reasonable explanation, given our genetic framework, is to assume that the acquisition of true tone systems by these originally atonal languages was activated or catalyzed by intimate cultural contact with languages which already had true tone systems: the "areal diffusion" hypothesis.

Given the complicated migrations and meanderings of these many peoples crisscrossing back and forth across Southeast Asia, we may be sure that all three logically possible contact situations occurred abundantly over the centuries: (a) AA / AT; (b) AA / ST; (c) AT / ST.  

As the language of the people who have been culturally dominant in East Asia for millennia, Chinese has exerted a powerful effect on the lexicon and phonology of the languages with which it has come in contact. Haudricourt (1954a), drawing on the work of earlier scholars like Henri Maspero, showed that in lexical items which Chinese has in common with Tai and Vietnamese (through borrowing in one direction or another), the tones systematically correspond: where Chinese has level tone (p'ing sheng), Vietnamese has tones ngang or huyễn, and Tai has tone "A" (unmarked in the writing system); where Chinese has departing (=falling) tone (ch'ü sheng), Vietnamese has tones hôi or ngã, and Tai has tone "B" (marked with the first tonal marker in the writing system); where Chinese has rising tone (shang sheng), Vietnamese has tones sác or nang, and Tai has tone "C" (marked with the second tonal marker in the writing system). See Figure 6.

In order for Tai, Miao-Yao, and Vietnamese to have become susceptible to tonal influence from Chinese, something must have happened to their internal structure to make them more "tone-prone". We must assume that phonological interinfluencing on the "segmental"
FIGURE 6. Sino-Xenic Tone Correspondences

<table>
<thead>
<tr>
<th>CHINESE</th>
<th>平 (level)</th>
<th>去 (falling)</th>
<th>上 (rising)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VIETNAMESE</td>
<td>ngang/huyền</td>
<td>hôi/ngã</td>
<td>sắc/năng</td>
</tr>
<tr>
<td>TAI</td>
<td>A (unmarked)</td>
<td>B (¹)</td>
<td>C (²)</td>
</tr>
</tbody>
</table>

level (i.e. involving consonants and vowels) must have preceded the tonal influence. First of all, these non-ST languages had to become truly monosyllabic (through the loss of affixes, reduction of unstressed syllables in compounds, etc.). Then, they had to suffer disastrous mergers in their consonantal systems in order to motivate their recourse to tones to maintain lexical contrastiveness. Haudricourt (1946a, 1961) has shown how widespread disruptions of the voiced/voiceless opposition in syllable-initial position must have swept through all the language families of SEA in the early centuries of the present millennium. Two main tendencies were at work: the devoicing of previously voiced stop initials, and the voicing of previously voiceless nasals and other sonorants. Standard Thai is a typical example, with the old *voiced series becoming voiceless aspirated (merging with the old *voiceless aspirated series) and the old *voiceless sonorants becoming voiced (merging with the old *voiced sonorants).

It seems likely that the development of true tones in Vietnamese was precipitated not only by influence from Chinese, but also from Siamese as well. This indicates that Tai (and Miao-Yao) acquired their tone systems from Chinese before Vietnamese did; that is, the ST > AT influence preceded the ST-cum-AT > AA influence.

The development of register systems in some Austronesian languages may be viewed as due to AA > AT substratal influence (the "Austro-linkage") at the geographical fringes of the true-tone diffusional area.

***

It should by now be apparent that tonal similarities— even
regular tonal correspondences—are not to be taken uncritically as evidence for genetic relationship among languages. Indeed, tonal criteria are not even sufficient to establish genetic subgroupings for languages which are already known to be genetically related. A striking proof of this is the fact that some modern dialects of Tibetan are truly tonal while others are not. Yet these are dialects of one and the same language, more closely related to each other than to any other language. Not only may tones be readily acquired by diffusion (provided that the acquiring language has been made sensitized or "tone-prone"); they may also be lost through contact with non-tonal languages (as in the case of some western subgroups of TB [cf. 2.1 above]).

For truly is it said, "The Language gave, and the Language hath taken away—blessed be the name of the Language" [Job 1.21].

FOOTNOTES

1This paper may be viewed as an introduction to the several articles and reviews on Tibeto Burman tones that I have written over the past five years (see References). Despite the fact that this material is easily accessible, non-specialists might find it useful to have the main motivations of this line of research presented here in one place in relatively non-technical fashion.

The term "tonogenesis" was first used, to my knowledge, in my 1970 article "Glottal dissimilation and the Lahu high-rising tone: a tonogenetic case-study".


3See 2.2 below.

4Most M-K languages have "register" systems rather than "true" tonal distinctions. See below, loc. cit.

5Haudricourt does not commit himself as to the exact nature of these stops, symbolizing them by **X**.

6Throughout the rest of this paper we use the symbols "Ci" and "Cf" for "syllable-initial consonants" and "syllable-final consonants", respectively.

7The diacritics over the vowels are those used to indicate the six tones in modern Vietnamese orthography. The words ngang, huyên, etc. are the native names for the tones.

8Haudricourt's term is "inflexion".
Haudricourt uses the words "hauteur" or "registre" for this concept. The word "register" has a different, technical sense when used to describe the two-way tonality opposition characteristic of Cambodian and the other Mon-Khmer languages. See 2.2 below.

Which my colleague John Ohala has tried to make me understand on several occasions.

William Ewan has carried out experiments which confirm this for English (personal communication); see also Lea (1973).


For a fascinating treatment of the relationship of the tongue-root to laryngeal activity in the production of tonal effects see Gregerson (1973).

See Matisoff (1972b).


The number of contrasts in a pitch-accent system is minimal (usually simply high-pitch vs. low-pitch), with no more than one syllable of each morpheme being specified for high pitch in the underlying structure. The pitches of the other syllables are typically predictable from their position in the word, or indeed from the whole grammatical construction that the word participates in. That is, the pitch contrast has a "low functional load" in distinguishing individual syllables paradigmatically.

This seems to hold for African languages as well. Those languages which have developed the most elaborate tone systems (e.g. Bamileke) are also monosyllabic (personal communications, March 1973).

Benedict wants to set up a two-way tone contrast in non-stepped syllables way back at the Proto-Sino-Tibetan period. For a brief discussion and some references, see 2.1 below.

I cannot resist observing that dental decay is no more prevalent than velar or labial decay in our family.

What Maran (1971) calls "depletion of final consonants".

This is what happened in Lahu, as we indicated above (*am > o, *an > e, *an > o).

All of these stages are attested in one or another Loloish language. See Matisoff (1972b).

Nungish is a minor TB group that shows special affinities both for LB and for Kachinic.
The third Chinese non-stopped tone, the "going tone" (ch'ū-sheng) has been demonstrated to be of relatively recent origin. See Haudricourt (1954b) and Downer (1959).


See Sedláček (1960).

See the discussion of the Burmese reflexes of the PTB *-ik rhyme, 1.2 above, and the remarks on the "tonal cycle" later in this section.

It is possible that more Kuki-Chin languages will be found to have real tone systems once they have been better recorded by modern linguists. Those Kuki-Chin languages which do have several tones (see e.g. Henderson 1968) exploit them extensively in productive morphological processes, which makes them look suspiciously recent in origin.

See Matisoff (1973d).

Instances of this process abound in the world's languages. In some American English dialects where pin and pen are homophonous, the words are replaced by the compound forms "stick-pin" /strı́kprın/ and "ink-pen" /ɪŋkprın/, respectively.

As a more exotic example, we may take the Galitsianer dialect of Yiddish, where the vowels u and i have merged, along with the spirants s and š. The words for foot and fish (standard Yiddish fus and fiš) are both pronounced /fiʃ/. Speakers of this dialect responded by creating jocular compounds whose second members were the Russian words for 'foot' and 'fish': fis-noge (< Russ. nogá 'foot') vs. fis-ribe (< Russ. rýba!)

Leaving out the fantastically complex and archaic linguistic area of New Guinea, which is now under intensive investigation by Professor Stephan Wurm and his associates at Australian National University.

An interesting Austronesian people are the Chams, who remigrated back to the mainland (Vietnam) after having lived for centuries in the islands near Malaya.

Though AN morphemes now typically have only two syllables, not three.

The reduction of the trisyllabic proto-root occurred differently in Tai and M-Y. Tai usually dropped the beginning of the root (cf. Siamese taa, Malay mata 'eye'), while M-Y dropped the end.

According to Huffman (1970), standard Cambodian has no fewer than 31 vocalic nuclei.

Not unlike those physically weak animal species, like gerbils, whose chosen evolutionary defense against extinction is the ability to proliferate their kind rapidly.
Benedict has discussed the AA/AT contact relationship, which he calls the "Austro-linkage", in Benedict (1973b). In the AA/ST area, Shorto (1973) has assembled an impressive number of Mon-Khmer etymologies for widespread ST roots. In Matisoff (1973a) I discussed the probable M-K source for the velar "animal prefix" in Lolo-Burmese. The AT/ST interaction has been intensively studied by Benedict (1967, part 3; 1972a; 1973). Many ST words for items of material culture and technology (including objects related to writing and the calendrical signs of the zodiac) can now be shown to have an AT source. Recent archaeological findings (Chang 1963, Gorman 1971) confirm a high level of material culture in the non-Chinese neolithic denizens of northern Southeast Asia.

See 1.1 above.

It is noteworthy that Japanese, despite centuries of massive lexical borrowing from Chinese, has never shown any signs of preserving lexical tone contrasts in these borrowed items. The intrinsic polysyllabicity of Japanese has resisted any such development. (Also the geographic isolation of the Japanese islands from the mainland must have been an inhibitory factor.)

Benedict's rejection of the diffusional explanation for the close correspondence between the tones of the distantly related Karenic and Lolo-Burmese languages (2.1 above) is therefore open to debate.

We might refer to this process as "tonoexodus" (Lea 1973).

REFERENCES


Benedict, Paul K. 1973a. Tibeto-Burman tones, with a note on

Benedict, Paul K. 1973b. Austro-Thai and Austroasiatic: the
Austro-linkage. Ms. Paper presented at the First International
Conference on Austroasiatic Linguistics, Honolulu.


Brown, Marvin. 1965. From ancient Thai to modern dialects. Bang-
kok: Social Science Association of Thailand.

nal of American Linguistics 33.2, Part II. Published simul-
taneously by Mouton and Co., The Hague.

Chang, Kun. 1953. On the tone system of the Miao and Yao lan-

Fifth Sino-Tibetan Conference, Ann Arbor, Michigan.

Chang, Kwang-chih. 1963. The archaeology of ancient China. Yale
University Press.

Downer, Gordon. 1959. Derivation by tone-change in classical
Chinese. Bulletin of the School of Oriental and African Stud-
ies 22.258-290.

Downer, Gordon. 1963. Chinese, Thai, and Miao-Yao. In Shorto,
H. L. (ed.), Linguistic comparison in Southeast Asia and the

Gorman, Chester. 1971. The Hoabinhian and after: subsistence
patterns in Southeast Asia during the late Pleistocene and
early Recent periods. World Archaeology 2.3.

Paper presented at the First International Conference on

Hale, Austin and Kenneth L. Pike. 1970. Tone systems of Tibeto-
Burman languages of Nepal. Occasional Papers of the Wolfen-
den Society on Tibeto-Burman Linguistics (OPWSTBL), Vol. III.
Urbana, Illinois.

Halle, Morris and K. N. Stevens. 1971. A note on laryngeal fea-

Haudricourt, André-Georges. 1946a. (With André Martinet). Prop-
gagation phonétique ou évolution phonologique? Assourdisse-
ment et sonorisation d'occlusives dans l'Asie du sud-est.
Bulletin de la Société de Linguistique de Paris 43.82-92.

Haudricourt, André-Georges. 1946b. Restitution du karen commun.
Bulletin de la Société de Linguistique de Paris 42.1.

Haudricourt, André-Georges. 1954a. De l'origine des tons en viet-

Haudricourt, André-Georges. 1954b. Comment reconstruire le chi-


ON BECOMING A TONE LANGUAGE
A TIBETO-BURMAN MODEL OF TONOGENESIS

BY

LA RAW MARAN

DEPARTMENT OF LINGUISTICS
INDIANA UNIVERSITY
BLOOMINGTON, INDIANA 47401

Published by the Linguistics Program
University of Southern California
Los Angeles, California 90007
0. INTRODUCTION

An attempt to provide an historical and comparative overview of Tibeto-Burman linguistic relations is confronted by a central obstacle, namely the absence of written history of most of the members of the language group. This fact has played havoc insofar as the segmental domain of language structure is being considered with the objective of demonstrating the stage-by-stage development of differences. The extent to which the historical processes of segment clipping and syllable shrinking have exerted their combined forces has left us with a variety of yet unanswered questions which are most vital to the concept of Tibeto-Burman (henceforth, TB) as a genetic assemblage. For instance, one is baffled by the position of Jinghpaw-Kachin in the larger scheme of linguistic relations; however, it is not that its relatedness to TB is held in question, rather it is the precise manner of its relationship that has lacked adequate understanding.

In the related area of the comparative study of the tonal phenomena, there may exist some reasons for optimism. The generalizing term of tonogenesis has already been coined (Matisoff 1970), and a convenient handle to cast a general and comparative overview of TB may have been obtained. In this paper we shall consider some aspects of this. In principle, this state of affairs is perhaps due to two reasons: one, tonal systems, contra segmental and syntactic systems, tend to be simpler, and two, the missing-parts problem which affect segmental entities extensively has not been applied to prosodic entities in kind. Thus, from a wide enough purview of TB problems one may be more likely to approach matters of general significance and insight in tonal studies than in the alternate area of segmental ramifications. Predictably, then, the objective of this brief paper is to outline a model of the origin of Tibeto-Burman tonal systems.*

1. THE PROBLEMS

There is growing literature on developmental aspects of tone in TB from pretonal origins and the direction of these works has followed, in a general sense, the position of Haudricourt (1954); one may cite in this instance the recent works of Lehman (1970),
Matisoff (1971, 1973), and Maran (1969, 1971a,b). Maran in the
works cited goes to the extent of suggesting that in the languages
he dealt with, the onset of the tonal grammar may have occurred in
historically recent times. Perhaps the general and compelling
evidence supporting a developmental theory of tones in TB is the
coincidence between the reanalysis and depletion of the finals on
the one hand, and the increasing dependence upon tonal features
for lexical contrast on the other. While the segmental features
specifying finals are gradually lost, the tonal features of sylla-
bles appear to gain ascendance as purveyors of cognitive cues of
contrast.

To summarize, the current feeling among many TB specialists
is that tones developed in these languages, and that the circum-
stances of development point to the interplay of phonetic influ-
ences emanating from the non-vocalic context of the syllable and
largely bearing upon the nucleus—the vowel. The position held by
Maran (1971a) as regards this phonetic interplay may be described
as a view where the nonvocalic environment, especially the final,
causes at some point the assimilation of certain phonetic features
to the nucleus where the deposited features become functional as
phonetic features of tone. In this view, it will be crucial to
show that the correct set of phonetic features which become func-
tional as tonal features on the nucleus did exist in the final at
one time. Thus, his works propose to show that in certain TB lan-
guages where tonal characteristics have not yet become "phonemic",
there exists an arrangement where the redundant phonetic features
of tone are directly recoverable from the array of existing finals.
He offers the example of the Eastern or Gauri dialect of Jinghpaw
as such a case in point.

At the level of phonetic data it is true that Gauri tone oper-
ates in a situation which might be characterized as "full redundan-
cy". If the final obstruents are phonetically p, t, k, the prece-
ding nucleus will have a high-checked tone, and where the finals are
b, d, g, a low-checked tone. The high and low checked tones also
occur with the glottal stop, and the open high/low tones with the
finals h and x (a glottal glide) respectively. It appears that
this final paradigm shows a natural class with two internal divi-
sions, the natural class being nonvocalicness and the internal division being the association with the high and low tones. For final consonants, the division rests on voicing; however, as the glottal stop and the glottal glide are not articulatorily "voiceable", the suggestion was made that the features of raised and lowered F account for this.

This position of Maran (1971a) has recently come under critical review by Matisoff (to appear), where two questions are raised. Matisoff's first question deals with the phonetic characteristics of the glottal stop and glide associated with the low tone and which Maran has described as [+lowered F] in order to avoid the contradictory use of the feature Voice. He implies that such a feature may well be phonetically implausible. In reply, it might be pointed out that the co-occurrence of the said finals and the low tone is a linguistic fact in Jinghpaw, Maru and certain other languages. One may take issue with the phonetic features used to describe the phenomenon, but the fact remains that there has been no compatibility problem in the co-occurrence of vowels and the said finals. Furthermore, the ambience of the vowel + laryngeal final sequence has produced the properties of low tone that are fully comparable to, say, the vowel + voiced nonlaryngeal final sequences. The phonetic implications of these finals will be dealt with in a forthcoming paper by Maran and Ohala; we shall for the time being say nothing on this matter here.

The second question is the more interesting as well as difficult one; if we grant that Gauri Jinghpaw has a system of complementation between its phonetic finals and tones at this stage in history, does it follow that things were generally so earlier in history? In order that all phonemic tonal systems might be historically derived from depleted finals, such will have to be clearly the case in Standard Burmese, for example. The possibility that Maran's position might entail such an implication is termed "unnatural" and "freakish" by Matisoff (to appear).

This question concerning the historical status of the tonal-segmental complementation displayed by Gauri Jinghpaw has the effect of bringing out two important issues which my earlier views as well as Matisoff's have so far failed to note and these issues
are as follows. First, we failed to distinguish between linguistic events which lead to the development of a system of tones and the events of change which occur after a language has become tonal. In other words, a model of historical development of tones is not itself a model of a tonal system, and this fact we seriously failed to notice. The task of a model of historical development of tones is to lay down phonetically and phonologically plausible points which may be observed in pretonal or nonphonemically tonal systems and to show how from this stage ensuing events may produce a system of tones; the historical backdrop itself is not to be attributed with having all the delineations of tonal structure which are preserved in the segmental units. After all, a model of the sort outlined in my earlier work simply attempts to provide the phonetic material which existed and the possible mechanisms of transfer from segmental to tonal stages. Thus, the tonal structure of modern Standard Burmese is not expected to relate directly to the pretonal finals, for instance. This fact puts a very heavy burden on the illustrative use of intermediate stages in the developmental model. One must grant the possibility that the stage of full complementation between Gauri phonetic finals and tones may be a consequence of normalization due to internal reasons rather than a natural and outstanding stage in the progression towards the development of a tonal system. In other words, we have failed to account for the fact that there may well be a series of conceptually and developmentally crucial intermediate stages rather than the single Jinghpaw-type stage. We must in either case give an account of how finals and tonal features come to display the kind of complementation that we find in Jinghpaw. This leads us into the second issue.

Underlying the notion of pretonal basis of tones in Burmese or Jinghpaw in my previous work is the model of an intermediate stage where the phonetic tonal features and the finals are in some set form of complementation. In Gauri, the complementation is complete, which means that either the finals or the tonal features are in a state of full redundancy. Let us call this the full redundancy intermediate stage. If Jinghpaw occupies this stage today, it seems highly unlikely that things were so several centuries ago. It is more likely that this state is itself a product of the
gradual process of change, that is, vis-à-vis other precedent stages. If this is the assumption, what then is the plausible series of intermediate stages which preceded the full redundancy stage and the stages which are yet to follow this full redundancy? In the model to follow, we shall try to make some observations regarding these questions.

2. A TIBETO-BURMAN MODEL OF TONOGENESIS

The model that is offered here is, as the name suggests, a genetic model. Furthermore, the assumptions and assertions made in connection with it are obtained from the developmental data of Burmese and Jinghpaw which have already been reported elsewhere (Maran 1971a). This brings about two factors upon which depends the efficacy of the model and these are: one, the assumption that both Burmese and Jinghpaw acquired tonal features through linguistic change, and two, that whatever conclusions may be drawn from the developmental data of these languages will in principle apply to the histories of the other tonal TB languages as well. We begin with a brief description of the historical settings, the comparative and the internal, or language-specific.

2.1. The Comparative Setting

Tibeto-Burman languages show two interesting tendencies: one, the comparatively disparate distribution of phonetic final consonants, and two, the simplicity or complexity of tonal systems. As regards the first, the distributions range from fairly comprehensive lists of Rawang and Tansar to the null set of C-finals in the Standard Burmese (SB), Lisu, Lahu, and Akha. Maran (1968) observed that on the face of it, the generalization that the complexity of finals is inversely related to the complexity of tonal systems seems to hold with few exceptions. In other words, the relatively complex tonal systems are associated with those members of the family which have lost the greatest number of finals. This putative causal tie has led to his dissertation (1971a) which provided a fairly detailed analysis of the correlation.

There is now a second dimension which has been added to the earlier observations about the finals and this has to do with the
utilization by lexically productive morphology of the finals as pair-wise contrastive units. For instance, the Jinghpaw morphology relies heavily on the obstruent vs. continuant phonetic contrast of the finals. Its finals have accordingly become a set of ordered pairs, p/m, t/n, k/ng and ʔ/h; there are no liquids l, r which its close relative Rawang has retained. Rawang and Tansar (Morse 1963 and personal communication) have retained p/m, t/n, k/ng, ʔ, r, l but seem to have no productive morphology. Written Burmese surely has far more finals than Central Burmese for instance, but it is only in the case of Northern Burmese (NB) that we find any measure of productive morphology being retained, and here the same form of pairwise contrastiveness is found. For instance, in addition to pairs like tsut/tsun 'insert by force/be bulging', thit/thin 'to mark as by cutting a notch/be marked', lik/ling 'to roll up/be rolled as in a ball', there is another pair of alternating consonants, the ʔ and h. (The open-high tone in NB shows a trace of postvocalic aspiration which I consider vestigial final h). Examples of this alternation include pwaʔ/pwah 'to grow or increase in size/be growing or multiplying', kawʔ/kawh 'to bend/be bent', cwaʔ/cwah 'to rise as anger or contempt/be boastful', sawʔ/sawh 'to play or fidget/be playful', etc.

The above characteristics may be summarized as follows in order to obtain a general view of the comparative setting as regards the states of finals. The table in (1) features (a) several Tibeto-Burman languages, (b) their phonetic finals, and (c) the existence of morphologically motivated pair-wise contrastive obstruent/continuant pairs.

<table>
<thead>
<tr>
<th></th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>i. Rawang, Tansar</td>
<td>p, t, k, m, n, ng, ʔ, r, l</td>
<td>none</td>
<td></td>
</tr>
<tr>
<td>ii. Jinghpaw, NB</td>
<td>p, t, k, m, n, ng, ʔ, h</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>iii. SB, Lisu, Lahu</td>
<td>t, k, n, ng, ʔ, h</td>
<td>yes</td>
<td></td>
</tr>
</tbody>
</table>

This survey appears to show three things: first, if the assumption is made that most TB languages lose finals, then JP and NB seem to occupy the intermediate stage in the ranking; second, the
formation of pair-wise contrastive finals develops for a purpose, namely morphology; and third, this linguistic purpose which contrasts obstructed finals against continuant ones has the effect of producing a form of normalization of the finals. This pair-wise symmetric distribution of finals in JP and NB is significant not only because the phenomenon is clearly part of the general reanalysis of finals but also because in JP the same linguistic process has apparently led to the \( p/b, t/d, k/g \) phonetic contrast in the finals.

2.2. **The Internal Setting**

In the nature of dialectal difference is to be found vital information which sheds light on the internal historical perspective. Robert Morse (personal communication) has recently informed me, for instance, that the TB languages of Tansar and Shangke which are spoken in the extreme northwestern hills of Burma are for all practical purposes two very closely related languages, or perhaps even dialects of a language with a functional degree of mutual intelligibility and sustained communication, except in the matter of finals and reliance on tones; Tansar has an extensive inventory of phonetic finals including the oral consonants and the liquids \( r, l \) and the glottal stop. However, Shangke has none; instead, the function of cognitive contrast of forms is carried entirely by tonal features. This situation may represent an extreme case, but the point is clearly made here that internal to a language may be found such disparity.

We have already described (in Maran 1971a) an analogous situation in the three major dialects of Burmese. NB has lost the finals \( p/m \), and whatever their phonetic import, the finals \( c/ny \) of written Burmese as well. It has the morphologically induced obstructed/continuant paradigm in the finals retained. In the next stage, Central Burmese has only the \( k/ng \) and \( v \) finals left; when we come to SB, \( v \) remains the lone nonvocalic final. And yet, all three dialects are modern versions of spoken Burmese and continue to exist side by side. There is something in the nature of historical or chronological sequence of final reanalysis and loss which is displayed in these dialects.
A nonstandard dialect of Jinghpaw spoken in the extreme southeast, an offshoot of the dialect of the Lahtaw clan, has apparently lost the finals m and ng, as these examples illustrate:

<table>
<thead>
<tr>
<th>Standard Jinghpaw</th>
<th>Southeastern JP</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>mam</td>
<td>ma</td>
<td>'paddy'</td>
</tr>
<tr>
<td>kawng</td>
<td>kaw</td>
<td>'tusk'</td>
</tr>
<tr>
<td>nang</td>
<td>na</td>
<td>'you (pr.)'</td>
</tr>
<tr>
<td>kang</td>
<td>ka</td>
<td>'be tight'</td>
</tr>
<tr>
<td>manam</td>
<td>mana</td>
<td>'be odorous'</td>
</tr>
</tbody>
</table>

The purpose of this brief illustration is to show that within the conventional concept of a language may be found dialectal communities which bear the evidence of continuing change, a striation which ranges from fairly conservative stages of final depletion to fully advanced ones. Given this type of situation, one possible way to construct a theory of the development of tone from an atonal beginning is to look into the sort of internal striation we have touched upon where gradual syllable simplification is apparent, and to perceive in it the transfer of the basic cognitive function of contrast of forms from segmental to prosodic systems of cues. Such, in essence, is the tonogenetic model of Tibeto-Burman.

We shall argue, therefore, that a sequence of stages which is both plausible comparatively as well as in terms of internal striation may also be presumed to represent the set of events which leads to the development of tones as a logical outcome.

2.3. The Model

We shall construct this model in three parts; the first part will deal with a hypothesized set of events leading to the development of tonal systems. This will be in substance a strategy inspired by the observations pertaining to the comparative and internal settings. The second part of the model concerns the mechanism of cognitivization of prosodic features. In the last section of this part we shall give some remarks about prosodic influence which derives from initials.

2.4. Internal Stages

Idealistically, or in the abstract, a language which proceeded
from an atonal stage in history to gradually acquire a system of
tones may be presumed to pass through the following stages.

I. The onset point in history marked by atonality of the
basic contrast of forms in the lexicon. If there were
pitch characteristics attendant upon the spoken forms
of basic lexical units, these would not be utilized by
the cognitive process as the basic property of lexical
contrast. One assumes that pretonal TB languages were
in this stage. Garo may in a loose sense still be at
this stage.

II. The onset of final reanalysis and realignment into a
pair-wise, contrastive paradigm motivated by internal
necessities of the linguistic process. The attendant
pitch characteristics may well assume a paradigmatic
pattern at this stage but will have no significant cog-
nitive function associated with them.

III. The full redundancy stage, the reanalysis and realign-
ment of finals having reached an epoch. This is the
intermediate stage alluded to in earlier models (Maran
1971a, for instance). Jinghpaw and NB are examples of
languages which have reached this stage.

IV. Advanced depletion of finals or their reanalysis and
the partial carriage of cognitive function of contrast
by prosodic features. Tonal phenomena in clear ascend-
ance. Illustrative languages are Central Burmese and
possibly the nonstandard dialects of Jinghpaw which are
located to the south of the standard JP.

V. Tonal features having been cognitivized become lexical
property; the finals, if any are left, are largely dys-
functional. Examples of this stage are SB, Lisu, Lahu
and Akha.

These complete our hypothetical set of stages. We will notice
that the three dialects of Burmese, NB, CB and SB occupy respec-
tively the third, fourth and fifth stages, whereas archaic Burmese
may be assumed to occupy stage I. It is not unlikely that some
(presently) unknown dialect in the peripheries of NB may yet be
found to be in Stage II; it seems that between Katha and Bhamo
there exists a vast richness of Burmese dialects which remain un-
tapped.

2.5. Cognitivization

This is not so much a specific stage as an event when the
basic system of cognitive contrast is transferred from a segmental
system of cues to a prosodic system. This event means, among other
things, that it is now far simpler to conduct the cognitive process
of perceiving and analyzing linguistic data largely in terms of prosodic features of tone.

The notion of such cognitivization or elevation of surface phonetic facts to cognitive function is in itself an intriguing view of linguistic change. If our model is at all plausible, we are talking about change in terms of moments and situations when the linguistic process, without prior and specific design, reaches a point of congruence with the phonetic process. We are suggesting that such situations in fact lead to changes in the grammar. We commonly assume that the government of the linguistic process lies in the grammar and the government of surface phonetics in the universal patterns and the internal constraints. Perhaps the interesting note of this model is the suggestion that there are moments when the premeditated process of grammar (as attested by the paradigm of finals in JP and NB) and the unpremeditated consequences of surface phonetics interdigitate in such a way that basic changes to the language are rendered transactable. The relevance of the full redundancy stage is in having underlined this point. For it seems obvious that without such a stage, cognitivization will most likely be arbitrary, costly and difficult to motivate internally. With the full redundancy stage the mechanisms for the transfer of cognitive functions are readily provided.

2.6. The Behavior of Tone Bearing Vowels

In this section we shall try to develop in a cursory fashion a typology of TB tonal categories and systems. Each system is provided a general and abstract model and each category is treated as an abstract specification. The objective is to try to show that there is a unifying thread underpinning the various tonal characteristics within TB, that even the more advanced and complex forms are derivable, in a developmental sense, from the more primitive versions.

Initially, we will need a set of conventions for the task and the following will be used. (1) The canonical syllable shall have the structure CVC. (2) The phonetic difference between the consonants (C) shall be indicated by numeral subscripting; thus C₁ ≠ C₂. (3) The tonal feature difference on the nucleus (V) shall be
indicated by the subscripted small-case letters x, y, and z; thus \( V_x \neq V_y \neq V_z \). By these conventions we can define the following elementary concepts.

(3) a. The pretonal forms shall have this type of contrast:

i. \( C V C_1 \neq C V C_2 \) or,

ii. \( C_1 V C \neq C_2 V C \)

etc., for nonprosodic basic contrast of items.

b. Derivability and Redundancy. A phenomenon is derivable if the following assimilatory relationship holds:

i. \( V \rightarrow V_x / \_ \_ C_1 \) or,

\( V \rightarrow V_x / C_1 \_ \_ \) and,

\( V \rightarrow V_y / C_2 \_ \_ \) etc.

Alternatively, derivability may be defined as where

ii. \( C \rightarrow C_1 / \_ \_ V_x \) or,

\( C \rightarrow C_2 / \_ \_ V_y \) etc.

Full redundancy as defined earlier may be stated as follows:

iii. \( V \rightarrow V_x / \_ \_ C_1 \) implies alternatively that

\( C \rightarrow C_1 / V_x \_ \_ \)

Also,

\( CV_x \neq CV_y \)

With the above notations for the basic concepts we shall begin to outline the general and common features of TB tonal systems. As we operate on the assumption that tonal features grew out of atonal earlier forms, derivability defined as (3b.ii) will be inapplicable, this being a case where tonal features are the underlying and given features of contrast. However, in order to be consistent with the theory of the development of tones with time we must allow for one thing. This concerns the fact that certain finals are re-analyzed and lost earlier than the others, and hence, for the for-
mer set of items in the lexicon, the property of providing contrast will belong to prosodic features. In other words, we expect the manner in which a tonal systems takes over to be gradual but pervasive. Examples to support this are not hard to find.

In the dialect of Tibetan described in Chang and Sheft (1964), the apical finals have already been lost and tones have replaced the finals d, n, l, which are retained in the writing system. In the case of standard JP the final h, associated with the high-open tone in Gauri, may be already lost. In the CB and SB dialects of Burmese there seems to be little concrete evidence to posit final h to derive the high-open tone, although in the more conservative dialects the opposite will appear to be the case. We shall assume that in those cases where a final cannot be motivated on independent grounds, the lexical item indeed is distinctive in terms of tonal features. Put differently, when we say that a family of languages becomes tonal historically we imply that redundancy is not quite absolutely extensive to all items in the language. Furthermore, it would be expected that the set of forms to which full redundancy applies will gradually decrease in membership whereas the set of forms where the basic contrast is tonal will increase in membership.

2.7. Nonvocalic Context and the Nucleus: the Influence

Where there is a redundancy or derivability as defined in (3b), we expect to find the following types of transfer mechanisms. Transfer mechanisms are rule-like procedures which explicate the connection between segmental arrangements and prosodic consequences.

(4) a. The final-based mechanism is a mapping of the following type:

i. \( V \rightarrow V_x / \_ \_ C_1 \) or,

ii. \( V \rightarrow V_y / \_ \_ C_2 \)

In NB and JP, the finals t, k, and ʔ (and p in JP) are associated exclusively with the high-(checked) tone. The tone rule will have the following form:

\( [+\text{tone}] \rightarrow [+\text{high tone}] / \_ \_ [+\text{obst}, -\text{voice}] \)
b. The initial-based mechanism is a transfer mechanism of the following type:

i. \( V \to V_x / C_1 \quad (C) \) or,

ii. \( V \to V_y / C_2 \quad (C) \)

The role of the initial segment of the syllable in influencing tonal development is reported in a number of sources (Haudricourt 1954, Matisoff 1970, Lehman, in preparation). Haudricourt's observation was based on Vietnamese, Matisoff's on Lahu and Lehman's on a dialect of Tibetan. However, in each of these, the initial-based transfer mechanism does not seem to apply without the final-based mechanism. As Lehman's well-documented case from Tibetan shows, the final-based mechanisms apply originally and then the initial-based mechanisms enter the derivation of tones. In other words, we may not have a tonal system in TB which depended entirely on the initial-based mechanism to develop and acquire its tones.

A situation where the development of tones relies both upon the initial and final environments shall be called a biconditional mechanism. However, as we know of no TB language where the biconditional mechanism starts with the initial, the sequence of transfer mechanisms will be an ordered one. The ordering of these mechanisms states that the initial step must be performed by the final-based mechanism. The following model is suggested for this type of process.

(4) c. The biconditional mechanism. Given \( C_2VC_1 \):

i. \( V \to V_x / \quad C_1 \)

ii. \( V \to V_y / C_2 \)

iii. Thus, \( C_2VC_1 \to C_2V_{xy}C_1 \) (as a result of i. and ii.)

Or, given \( C_2VC_2 \), we will expect an output having the shape \( C_2V_{yy}C_2 \), and if contraction is allowed, \( C_2V_yC_2 \). In other words, if the initial and the final differ crucially, such as in voicing, we may expect a kinetic tonal phenomenon such as High-Low or Mid-High, Low-Mid, etc. This kinetic phenomenon may occur on the single, original nucleus, or the nucleus may be reduplicated;
thus $C_2 V_{xy} C_1$ may in due course lead to $C V_x V_y C$.

To illustrate this biconditional transfer we shall return to the Tibetan dialect from Chang and Sheft (1964); here Maran (1968) shows that the lost final apicals $d$, $l$ are replaced by kinetic tones occurring over the nucleus; that is, falling-rising and rising-falling tones here quite likely owe their origins to the following kinds of transfer:

(5) a. $C_1 V C_2 \rightarrow C_1 V_{yx} C_2$ after which the final dropped
b. $C_2 V C_1 \rightarrow C_2 V_{xy} C_1$ after which the final dropped

The tonal features transferred to the nucleus will thus differ in the sequencing of the rising and the falling features.

In the case of the Written Tibetan final $r$, which was apparently also reanalyzed and lost in this dialect, geminate tones replace the final. We can conceive of this transfer as follows:

(5) c. $C_2 V C_2 \rightarrow C_2 V_{yy} C_2 \rightarrow C_2 V_y V_y$

In keeping with Lehman's (in preparation) finding that the initial-based mechanisms also play a part, we shall offer the above biconditional transfer models as an account of Tibetan tone development.

When a nonvocalic final is reanalyzed and lost, the closure segment is removed, and hence, some form of vowel-lengthening will be expected. The Tibetan phenomenon seen in (5c) will serve as an example of such a change.

We have completed illustrating the major tonal systems of TB by the use of two transfer mechanism models, the final-based transfer and the biconditional transfer. A survey of TB systems (Maran 1971a:91-96) reveals that there is no system there which could not be accounted for by the two transfer mechanisms demonstrated above.

3. CONCLUDING REMARKS

The objective of this brief paper has been to outline a general model of the development of tones in Tibeto-Burman. The model has been patterned after certain comparatively and internally sig-
nificant stages of change; the principle of the model asserts that given a continuing flow of changes of the type discussed, there will come a point or a moment when certain transfer mechanisms are conveniently set loose. Although one may get the impression that the model is rather streamlined and mechanical, we stress again that this is an abstract model.

NOTE

*I wish to thank F. K. Lehman and Robert H. Morse for providing me information and to Eric Svaan for technical assistance. Furthermore, I must thank Larry Hyman for exhorting me to finish this brief paper, a task I had considered insurmountable due to unfortunate and immediately pressing problems. I alone, however, am responsible for any weakness of the contention.

REFERENCES


Morse, R. H.  1963. The phonology of Rawang. Anthropological
Linguistics 5.5:17-41.
THE ROLE OF TONE IN SEGMENTAL PHONOLOGY

BY

WILLIAM R. LEBEN

COMMITTEE ON LINGUISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

Published by the Linguistics Program
University of Southern California
Los Angeles, California 90007
0. INTRODUCTION

0.1. Perhaps the most striking aspect of the description of tone is the number of distinct kinds of underlying representation that have been posited for it in various languages. For example, Edmondson and Bendor-Samuel (1966) claim that the tonal patterns on words in Etung, a language of Nigeria, are such that tone cannot reasonably be regarded as other than a feature of the phonological word. Welmers (1962) makes an observation about stress assignment in Kpelle which he regards as formulate only if tone is expressed as a feature on morphemes. Similarly, McCawley (1970) proposes that in Bantu verbs and in the Southern and Western Kyûshû dialects of Japanese, tone is phonologically represented as a morpheme feature assigning pitch to the pitch-bearing unit; this proposal draws on the observation in McCawley (1964) that in Southern and Western Kyûshû, the amount of tonal information necessary in an underlying representation is independent of the length of the morpheme. At the same time, Pike (1948) defines the syllable as the tone-bearing unit in a tone language; McCawley (1964, 1970) also adopts the syllable as the domain of tonal features in tone languages; Wang (1967) provides some arguments for viewing tone as a feature on syllables in some languages. Finally, Woo (1969), Schachter and Fromkin (1968), and Maddieson (1971) express tone phonologically as a feature on segments.

0.2. Is tone such a special phenomenon that it must be viewed as a feature on morphemes or larger units in some languages, as a feature on syllables in others, and as a feature on segments in still others? If so, then there is something left to explain: namely, why tone, unlike any other linguistic entity we know anything about, is capable of this many different types of representation. Phonological features like [continuant] or [strident] are almost universally maintained to be features only on segments. Features like [+Past] or [+First Conjugation] are universally maintained as features on morphemes or larger units, although Chomsky & Halle (1968), hereafter SPE, propose a convention assigning these morpheme features to individual segments.
in order to account for certain phonological facts; however, even in this case, features like [+Past] or [+First Conjugation] do not undergo the sorts of phonological changes that tone is seen to undergo. As for features on syllables, granting for the moment that some types of features can be expressed on syllables, we lack clear evidence which would distinguish the behavior of such features from the behavior of features on other linguistic units, such as the morpheme or the segment; this makes it hard to come by justification for positing tone as a feature on syllables.

If, on the other hand, we abandon the claim that tone is capable of all of these different kinds of representation, then an equally difficult question arises: what has caused all of these different kinds of representation to be proposed, and how could we reformulate these proposals in accordance with a theory that permits only one of these kinds of representation?

This paper is devoted, first of all, to making these questions appear answerable, and secondly, to at least beginning to formulate actual answers for them. The basic thesis that will be defended is that tone is not really all that different from other phonological phenomena that have been studied, that these other phenomena are also expressible as features on a number of different kinds of linguistic units, and that, given a proper analysis of the data, there is no contradiction or paradox involved in maintaining that these phenomena behave sometimes as segmental features and sometimes as suprasegmental features, where "suprasegmental" is taken to characterize features on linguistic units larger than the segment.

To demonstrate these points, we need criteria for differentiating segmental from suprasegmental behavior, and these criteria will have to be specified formally. The framework of generative phonology is an attractive vehicle for this purpose, since by its very nature it affords a concrete evaluation of claims about segmental or suprasegmental behavior. For example, if one maintains that tone is a segmental phenomenon, the formalism clarifies the nature of the task involved in demonstrating, first, that this claim has some content, and second, that the claim is true; one
would aim to show that there are non-segmental phonological phenomena in language and that tone behaves not like these but like other clear cases of segmental phenomena. Similarly, if one advocates representing tone as a feature on morphemes, one's evidence will exhibit similarities between the behavior of tone and the behavior of undisputed morpheme features.

This paper will point up a number of empirical predictions latent in a formalism; this situation, which is not at all uncommon, provides perhaps the most potent justification for concern with formalism in linguistic description. Insofar as the present paper is devoted to justifying a formalism for tone, it may appear subject to recent attacks on the "empty formalism which has characterized generative phonology and its derivatives to date" (Ohala 1972). Ohala (1973) has condemned "the sterile, faddish formalism and unchecked speculation that characterizes most of generative phonology and its offshoots". More specifically, there have been many criticisms of the important role assigned to formal simplicity in the evaluation of competing analyses, and there does appear to be a good deal of agreement on this. For example, Kiparsky (1971, 1972) has proposed a number of substantive constraints on phonological systems; these constraints are intended to be taken into account along with the formal constraint of simplicity. At the same time, however, Kiparsky (1972) also provides what is perhaps the clearest example to date of the benefits achieved from the current "fad" of formulating rigorous, explicit descriptions. SPE stresses the empirical content of the abbreviatory devices proposed therein; Kiparsky's example from two Finno-Ugric languages illustrates this point beautifully. Eastern Cheremis has a stress rule which, expressed informally, has two parts:

(1) **STRESS** (Eastern Cheremis)

   a. Accent the last unreduced vowel of the word
   b. If there is no unreduced vowel, accent the first syllable

In Komi, on the other hand, stress is assigned in the following way:
(2) STRESS (Komi)

a. Accent the first heavy vowel in the word
b. If there is no heavy vowel, accent the last vowel

Hence, if we symbolize reduced or non-heavy vowels with \( \checkmark \), we may schematize facts of these languages in the following way, using \( C_o \) to designate any number of consonants:

(3) a. Eastern Cheremis
   i. \( C_o \checkmark C_o V C_o \checkmark C_o \)
   ii. \( C_o \checkmark C_o V C_o \checkmark C_o \)

b. Komi
   i. \( C_o \checkmark C_o V C_o V C_o \)
   ii. \( C_o \checkmark C_o V C_o \checkmark C_o \)

Now, Kiparsky observes, if we attempt to formulate part (a) of the Eastern Cheremis stress rule in the simplest way (i.e. using the "simplistic" notion employed in SPE), we arrive at the following result, which has been modified from Kiparsky's published version to reflect some recent unpublished work by M. Halle and J.-R. Vergnaud:

(4) STRESS (Eastern Cheremis)

\[
V \rightarrow [+Acc] / \_\_\_\_ Q \#
\]

\( Q \neq \ldots [-\text{reduced}] \ldots \)

In this rule, \( Q \) is a variable which can stand for any string of segments that does not contain a word boundary; the condition imposed on \( Q \), which prohibits it from containing unreduced vowels, guarantees that the final unreduced vowel of the word will be accented. But in addition to expressing the first part of the informal stress rule in (1), (4) also automatically yields the second part; there is a general convention on the interpretation of variables like \( Q \) which requires that the first expansion of \( Q \) must be the maximal one (in this case, the expansion which makes the \( V \) to be accented the first vowel in the word); furthermore, once we establish that the rule applies in this environment, subsequent
applications of the rule to non-initial vowels are blocked by the convention on disjunctive ordering.

For similar reasons, the simplest statement of part (a) of the Komi stress rule automatically yields part (b) as a special case:

(5) \[ \text{STRESS (Komi)} \]

\[ V \rightarrow [+\text{Acc}] / \# Q \]

\[ Q \neq \ldots [+\text{heavy}] \ldots \]

Note that the point here is not simply that our notational conventions make possible a few elegant solutions to stress problems; rather, these notational conventions, which were arrived at on completely independent grounds, will, when coupled with an evaluation measure, permit us to explain why these two languages differ not just in part (a) or in part (b) of the stress rules (1) and (2), but in both parts of these rules. Since the analysis in (4) and (5) says that only one rule is involved in both of these cases, it follows from this analysis that a language which differs from Eastern Cheremis or Komi in one part of the stress rule in (1) or (2) will also optimally differ in the other part of the stress rule.

In addition, J. Greenberg and C. Ferguson have privately observed that the sort of situation depicted in Eastern Cheremis and Komi is quite common in languages. This is exactly what our notational conventions would predict, given that languages tend toward optimal representations for their rules.

These notational conventions, since they constitute empirical hypotheses about how language works, are, of course, subject to disconfirmation. Consider, for example, the complicated account of Hausa stress in Abraham (1934). From the data that Abraham gives, we may judge that stress is assigned by the following ordered informal rules:

(6) \[ \text{STRESS (Hausa)} \]

a. Assign [1 stress] to all H (for high) tones
b. Reduce [1 stress] to [2 stress] on all vowels when
either LH or HL precedes in the same word or when HL follows in a separate word

c. Disyllabic nouns with the pattern LH have their stress shifted back to the L vowel when a word beginning with H follows; the shifted stress is reduced to [2 stress]
d. In a reduplicative compound of the form LH-LH, stress on the first H is shifted back onto the L vowel and is reduced to [2 stress]

It is impossible to express this account in an elegant set of rules using the SPE convention. Hence, assuming Abraham's characterization of the Hausa stress facts to be correct, we might look for notational conventions to supplant those of SPE which would permit a more reasonable statement of (6) while at the same time providing for rules in English, Eastern Cheremis, Komi, and other languages in which the current notational conventions have proved adequate. On the other side of the coin, however, we see that the current notational conventions, since they appear to be justified by the facts of a number of languages, cast doubt upon the correctness of (6) as a representation of linguistic data, judging from the relatively awkward form of (6) when expressed under these conventions. Since Abraham's observations appear never to have been taken seriously by any scholar, the second side of the coin is most likely the appropriate one in this case.

This points up the function of a theory in scientific fact-gathering; insofar as a theory appears to be justified, it should be permitted to shape our expectations about how the facts will turn out; and insofar as a theory induces empirically incorrect expectations, it needs to be modified or abandoned. In the case of Hausa stress, of course, it may be the case that notational conventions are essentially correct and that the optimal representation of the facts of Hausa happens to appear quite awkward. This position could be disconfirmed by producing facts from a closely related language, as with Eastern Cheremis and Komi, which demonstrate that the subparts of (6) are more closely related than the present notational conventions would predict.

In the study of tone, as in any other study of linguistic structure, we have every reason to expect that concern with the formal aspects of representation, far from resulting in what some
have called a "pseudo-algebra" or a "sterile game", will produce
a real algebra which is subject to empirical disconfirmation when
applied to language and which will, in the end, shed light on the
behavior of tone.

1. SEGMENTAL REPRESENTATION OF TONE

1.0. As noted above, tone is treated as a segmental phenomenon in
Woo (1969), Schachter and Fromkin (1968), and Maddieson (1971),
among others. However, in all of the cases that I know of, justifi-
cation for the segmental representation of tone is either ex-
tremely complex or else totally unmentioned. For our purposes, it
will suffice to summarize an analysis of tone changes in Standard
Thai compounds, and then to review some facts from languages in
which tone is subject to influences from surrounding segments.

1.1. The analysis of tonal changes in Thai compounds, taken from
Leben (1971b), provides a quite succinct demonstration of one sort
of argument that may be made in favor of segmental representation
for tone. Data is from Henderson (1949).

Thai is regarded as having two styles of speech: the isolative
style, in which syllables are articulated relatively slowly,
and the combinative style, which is more rapid. In the combina-
tive style, distinctions in tone and vowel length are neutralized
to a considerable extent. In particular, in compounds, a long
vowel in the first element will be shortened in the combinative
style, and its tone, if it is a contour tone, will be simplified to
a mid tone.

(7)  Isolative                          Combinative
     a. thi:  nai                         thi'  nai        'where?'
       HL    LH                          M     LH
     b. thi:  ni:                         thi'  ni:       'there'
       HL    H                           M     H
     c. si:  kha:u                       si'  kha:u      'white'
       LH    LH                          M     LH
     d. sa:u  sa:u                       sau  sa:u        'young girls'
       LH    LH                          M     LH
     e. wa:ŋ  wa:ŋ                       waŋ  wa:ŋ        'at your leisure'
       HL    HL                          M     HL
The single dot after the vowel of the first element in the combina
tive style in (7a-c) indicates that some length is retained on
vowels not followed by a consonant or glide in the same syllable,
even though shortening applies. In cases where the first element
has an underlying level tone, shortening applies, but the level
tone is maintained as before; it is not changed to a mid tone.

(8)  Isolative     Combinative
     a. na:m  tɔha:
       H    M     nam  tɔha:  'tea'
          H    M
     b. na:m  ta:n
       H    M     nam  ta:n  'sugar'
          H    M

At this point, we may summarize the facts informally in the fol-
lowing way: in the combinative style, the vowel of the first ele-
ment of a compound is shortened, and if it has a complex tone,
this is simplified to a mid tone. The following two sets of forms
establish that there is a close connection between the shortening
of the vowel and the simplification of the tone:

(9)  Isolative     Combinative
     a. tɔŋ  ka:n
       HL  M     tɔŋ  ka:n  'want'
           HL  L
     b. thau  rai
       HL  M     thau  rai  'how much'
           HL  L

Here we see that if shortening does not apply (that is, if there is
no long vowel to be shortened), then tone simplification also does
not apply. We can account for the connection between what appeared
at first glance to be two separate operations by expressing long
vowels as geminates, VV, and by expressing contour tones as se-
quences of level tones, with each level tone expressed as a fea-
ture on a vowel. Vowels with falling and rising tones will then
be represented as in (10a,b), respectively:

(10) a.  V  V
       [+H]  [+L]

b.  V  V
    [L]  [H]

This permits us to express vowel shortening and tone simplification
in a single rule, making the connectedness of these operations an
automatic consequence:

(11) VOWEL SHORTENING (Thai)

\[ \text{VV} \rightarrow \text{V} \]

As has been suggested by Sapir and also other linguists, the result of collapsing two segments into one is in some sense a "compromise" between the two original segments; consider for example, a rule collapsing a vowel and a following nasal—the normal result is a nasalized vowel. Establishing this as a convention on rules of collapsing will cause (11) to merge the sequences HL and LH on vowels into a mid tone, while leaving a sequence HH as H. This, then, demonstrates that in some cases, tone does appear to be a feature on segments.

1.2. Another type of instance in which tone behaves like a segmental phenomenon appears in cases where it is subject to influences from the surrounding segmental environment. Greenberg (1968) discusses a number of languages in which tone exhibits, or historically has exhibited, this type of behavior. Synchronically it appears that in Xhosa a high tone is realized as a rising tone when preceded by a "depressor" consonant, such as bh, mbh, mh, v, h, but not when preceded by consonants like p, ph, t, ñ, m, f, h. As we might expect, a falling tone, which is really the sequence HL, is realized as a rising-falling, LHL, when preceded by a member of this same class of depressor consonants.

(12) a. into 'thing'    c. impuku 'rat'
    H H      H HL H
    [↑ -]     [- -]

   b. inja 'dog'    d. imbhila 'rabbit'
    H H      H HL H
    [↑ -]     [- -]

A number of proposals have been advanced to account phonetically for the influences of a preceding consonant on a tone; the reader is referred to Ladefoged (1971), Hale and Stevens (1971, 1972), and some of the other papers presented at this Symposium for discussion and further examples of this interaction. For the case at hand, it is sufficient to note that, while part of the consonantal influence on tone is a completely automatic consequence of the
physiology of speech, languages like Xhosa must also incorporate a phonological rule which prefixes a low-toned component to a H preceded by a depressor consonant, since the physiology of speech is not sufficient to account for the degree of lowering evidenced; this point is discussed at greater length in Fromkin (1972) and Hyman and Schuh (1972).

Given this, we may say that in the case of Xhosa, tone behaves like a segmental feature, since it is influenced by surrounding segments; later we will see that non-segmental features are not subject to such influences.

2. SOME APPARENT IDIOSYNCRASIES IN THE BEHAVIOR OF TONE

2.0. Evidence for segmental features of tone has led some researchers to reduce all instances of the behavior of tone to a representation in which tone is characterized exclusively as a feature on segments. This position is defended by Woo (1969). Schachter and Fromkin (1968) express tone in Akan as a feature on vowels in underlying representations; in addition, they permit phonological derivations to give rise to a tone feature represented on an entity which has no segmental features other than [+seg]. This proposal is reformulated in Fromkin (1972), where these segmentally unspecified entities are characterized as [-seg]. A somewhat similar proposal is made for Nigerian and other languages in Maddieson (1971), but with [-syllabic] substituted for [-seg].

In this section I will attempt to demonstrate that tone does not uniformly behave like a segmental feature in languages, that sometimes it exhibits behavior that must be characterized as suprasegmental. To make this point, I will discuss cases in which segmental tone features cannot explain and can only awkwardly characterize the operation of tone, while suprasegmental tone features make possible a quite satisfying explanation. In addition, I will show that if the use of segmental tone features is extended to characterize these cases, such an extension would make latent predictions about the possible behavior of other segmental features, and these predictions turn out to be false.

2.1. McCawley (1970) has observed that in Southern and Western
Kyūshū, only two types of tonal pattern are possible on lexical entries, a falling pattern and a non-falling pattern; he emphasizes that the possible tonal patterns are thus completely independent of the number of vowels or syllables in the lexical entry. Polysyllables can occur with only two possible patterns, and these patterns are the same as for monosyllables. This fact is captured by specifying each lexical entry for a value of the feature [Falling]; a phonological rule then maps the pattern onto the segments or syllables of the lexical entry.

The same point is demonstrated in Tiv by McCawley (1971): the tonal pattern of a given verbal form is arrived at by a phonological rule which combines the basic tone of the verb with the basic tone of the tense or aspect morpheme that is attached to the verb. This rule operates without regard for segmental information about the verb. Subsequently, a phonological rule maps the tonal pattern arrived at in this way onto the segments or syllables of the verbal form.

The fact that the operations which precede tonal mapping occur without reference to segmental information is captured on an analysis in which the tonal information is specified suprasegmentally; suprasegmental tones are by definition independent of any segments—rather than being expressed as features on segments, they are features on larger linguistic units. This makes it impossible for suprasegmental tone rules to refer to individual segments over which they may range, and hence predicts that suprasegmental tone rules will not need to refer to segmental information.

If tone is expressed as an underlying segmental feature in these languages, on the other hand, the question of why rules prior to tonal mapping operate independently of segmental information is left unanswered. The latter analysis would be consistent with a situation in which such rules did require segmental information. Insofar as we can show that this type of situation is unexpected in languages, we are justified in embracing the suprasegmental proposal over the segmental one, since the suprasegmental proposal makes an apparently correct prediction, while the segmental proposal seems to make no prediction at all. Later on, we will see that
there are indeed predictions latent in the segmental proposal, and that these predictions are quite certainly false.

2.2. Another type of tonal relationship that is captured without reference to segmental information occurs in Mende, a Mande language of Sierra Leone. Mende lexical entries can have the tonal contour LHL; this contour can occur on words of one, two, or three syllables. Hence, in Mende we have \( \text{nmba} \, LHL \) 'companion', \( \text{nyaha} \, \frac{L}{H} \, HL \) 'woman', and \( \text{nikili} \, \frac{L}{H} \, H \, \frac{L}{L} \) 'groundnut'. However, no Mende lexical item can have the tonal pattern HLH, and this restriction applies to words of arbitrarily many syllables; that is, the following are impossible in Mende: \( \text{*nmba} \, \frac{HLH}{H} \, \frac{L}{H} \, \frac{L}{L} \), \( \text{*nyaha} \, \frac{HLH}{H} \, \frac{H}{H} \), \( \text{*nyaha} \, \frac{H}{L} \, \frac{H}{L} \). In Leben (forthcoming), the representation of Mende tone as a feature on lexical items is justified in some detail. The fact that the constraint on morpheme structure described here is independent of the number of vowels or syllables in the lexical entry is a direct consequence of the proposed suprasegmental underlying representation. The restriction in question can only be described awkwardly, if we are forced to regard Mende tone as a property of segments or syllables.

2.3. Bambara and Maninka, two closely related Mande languages, have the same sort of limitation on underlying tone patterns that was pointed out above for Southern and Western Kyushū and for Tiv; there are only two patterns possible, regardless of how many vowels or syllables a given word has. To express this fact, Welsmers (1949) posits the underlying forms in the left hand column in (13a,b,c); pronounced before a low tone, these would have the pitches represented in brackets below each form. For our purposes, these underlying forms will be re-expressed in the manner shown on the right.

(13) a. \( \text{jiri} \, [-] \quad \text{Hjiri} \, \text{tree'} \)
   b. \( \text{muso} \, [-] \quad \text{LH}_{\text{muso}} \, \text{'woman'} \)
   c. \( \text{sa} \, [+\text{/}] \quad \text{LH}_{\text{sa}} \, \text{'snake'} \)

The motivation for expressing these tones suprasegmentally in underlying representations can, as in the cases mentioned above, be based simply on capturing the limitation on the distribution of tone
patterns in these languages; a phonological rule of tone mapping eventually assigns the appropriate tones to individual vowels. What is particularly interesting about Bambara and Maninka is that it is possible to establish the existence of a rule in these languages which performs an operation on these tones that treats the tones as suprasegmental entities; as will be shown, if this rule had to treat the tones as segmental features, it would produce incorrect results.

The data for Bambara and Maninka is taken from Bird (1966, 1968) and from Spears (1968). The analysis departs from that proposed in these sources in a number of ways; an explicit comparison of these analyses is made in Leben (forthcoming). In both languages, there is a rule which lowers a H when it is preceded by L and followed by H. For example, in the Maninka expressions in (14b,c), the underlined vowel, which exhibited a H or LH in (13), is converted to L.

\[(14) a. \text{Hjiri Hte Lyan} \quad \text{\textquoteleft\textquoteright\ no tree is here} \quad [\_ \_ \_ \_] \]
\[b. \text{LH muso Hte Lyan} \quad \text{\textquoteleft\textquoteright\ no woman is here} \quad [\_ \_ \_ \_] \]
\[c. \text{LH sa Hte Lyan} \quad \text{\textquoteleft\textquoteright\ no snake is here} \quad [\_ \_ \_ \_] \]

Similarly, in Bambara, the underlined vowel is realized as L instead of H, because it is preceded by L and followed by H. (The immediately following H is a "floating" indefinite marker, proposed by Bird (1968), with no segmental realization. The question of "floating tones" will be returned to later.)

\[(15) \text{La Hma LH muso H(Indef.) Ye} \quad \text{\textquoteleft\textquoteright\ he didn't see a woman} \quad [\_ \_ \_ \_ \_] \]

This alternation between H and L in Maninka and Bambara is expressed by (16), TONE SPREADING, a process noted by Hyman and Schuh (1972) to occur in many languages of West Africa.

\[(16) \quad \text{TONE SPREADING} \]
\[H \rightarrow L / L \_ \_ \_ H \]

We must now ask whether TONE SPREADING applies to suprasegmental
units or to segmental units. Either interpretation of (16) would produce the correct result in the examples considered thus far; if the H to be lowered on \textit{muso} and \textit{sa} in (14) and (15) is a feature on the final vowel, the rule will convert it to L; and if the H is not yet associated with any vowel but is rather a feature on the lexical entries, it will still be operated on by TONE SPREADING, producing the outputs $L_{\text{muso}}$ and $L_{\text{sa}}$ in (14) and (15), which would later be mapped into $\frac{\text{muso}}{L_L}$ and $\frac{\text{sa}}{L}$. A crucial case which decides between these two interpretations of the rule is provided by tone assignment in Bambara and Maninka compounds.

The following observational statement of the Bambara compound rule is adapted from Woo (1969); the data in Spears (1966, 1968) shows that the Maninka rule is identical.³

(17) COMPOUND RULE (Bambara, Maninka)

a. If the first word is H, then every vowel of the compound is H

b. If the first word is L [i.e. LH in the reanalysis], then all vowels preceding the last word in the compound are L; the last word is H

Hence, the tonal shape of the compound is entirely dependent upon the tonal shape of the first member. The contrast between compounds with a high-toned first member and those with a non-high-toned first member is illustrated in (18):

(18) a. $H_{\text{jiri}}$ \textit{jiri-finman-nyiman} $HHHHHH$ 'good black tree'

b. $L_{\text{H}}_{\text{muso}}$ \textit{muso-finman-nyiman} $L_LLLHH$ 'good black woman'

The observational account in (17) gives us no clue as to why high-toned and non-high-toned nouns behave as they do in compounds; in effect, it says that there is no explanation for their behavior and that if these languages instead contained a compound rule such as the following, they would be no more complicated:

(19) COMPOUND RULE (artificial)

a. If the first word is H, then every vowel of the compound is H

b. If the first word is L, then all vowels preceding the last word are H; the last word is L
Now, there is nothing in itself wrong with an analysis which does not explain why certain facts are the way they are, since certainly there are many facts in languages for which no explanation will suffice. However, from what we know about the rest of the grammar of Bambara and Maninka, we can see that there is an explanation for why the compound rule is as it is expressed in (17); the explanation rests on the operation of TONE SPREADING, which causes a transition from L to H to be moved rightward in a word. That is, TONE SPREADING permits us to re-express (17) as (20):

(20) REVISED COMPOUND RULE (Bambara, Maninka)

a. Copy the last tone of the first word [i.e. a high tone, in all instances] onto all non-initial words in the compound

b. (=16) TONE SPREADING

If we interpret this rule as applying to segmental tone features, its output is not quite correct, since it causes the vowel of the penultimate syllable in (21b) to be L and it should be H. (For the moment, I will assume that TONE SPREADING applies iteratively.)

(21) a. jiri-finman-nyiman
    H H
    H H H H H H
    LH H H H H H
    L L H H H H
    L L L H H H
    L L L L H H
    L L L L L *L H

If the rule can be stopped from applying to the vowel of the penultimate syllable, the correct output will be obtained. This result is exactly what is achieved by interpreting this rule as applying to suprasegmental tones, as illustrated in (22):

(22) a. Hjiri-finman-nyiman
    H jiri H finman H nyiman
    L H muso H finman H nyiman
    L H muso H finman H nyiman
    L H muso H finman H nyiman

Subsequently, the tones will be mapped onto individual vowels by the same mapping rule alluded to above; this will yield the forms
given in (18).

There is a question as to whether phonological rules ever apply iteratively. In a version of phonological theory which does prohibit iterative rules, the most promising way of restating the compound rule would be to have it apply cyclically to forms which are bracketed as in the first line of (23). Derivations would then proceed in the following way:

(23) a. \[\text{[[H} \text{jiri}] \text{finman} \text{H} \text{nyiman}]\]  
\[\text{H} \text{jiri H} \text{finman}\]  
\[\text{H} \text{jiri H} \text{finman H} \text{nyiman}\]  

b. \[\text{[[L} \text{H} \text{muso}] \text{finman} \text{H} \text{nyiman}]\]  
\[\text{L} \text{muso H} \text{finman}\]  
\[\text{L} \text{muso L} \text{finman H} \text{nyiman}\]  

If this latter analysis were maintained, one would, of course, seek independent justification for this bracketing; for example, one might see whether the adjective finman restricts only jiri while the adjective nyiman restricts the full constituent jiri-finman. Since I have no data to support this view, and since the need for this alternative to the first suprasegmental formulation has not been conclusively demonstrated anyway, I will simply end this discussion with the observation that the suprasegmental analysis can explain the tonal behavior of compounds on the basis of principles arrived at on completely independent grounds from the behavior of non-compounds; the segmental analysis, on the other hand, can only observe this behavior.

The existence of a phonological rule which operates on suprasegmental tones, along with the fact noted in preceding sections that certain restrictions on tone patterns can be most adequately stated suprasegmentally, establishes strong support for the notion that tone cannot be solely regarded as a segmental feature.

2.4. A demonstration similar to the one just given for Bambara and Maninka can also be made for Hausa. The analysis of Hausa on which this demonstration relies is justified in Leben (1971a) and Leben (forthcoming).
The following forms illustrate an alternation between H and L in Hausa nouns and adjectives. The vowels on which the alternation occurs are underlined; the tones are those which appear prior to the operation of LOW TONE RAISING, a rule discussed in Leben (1971a).

(24) a. jinjir+ii (m.)  jinjir+nii+aa (f.)  'baby'
    \H L L \H H L L

b. ma+bi+ii (m.)  ma+bi+ii+aa (f.)  'follower'
    \H L L \H H L L

The L of the underlined vowel in the masculine form corresponds to a H in the feminine form. In (24b), the morpheme bi 'to follow' has an underlying H tone; but in the agentive construction exemplified here, the underlying tones of the morpheme in this position are disregarded and a morphological tone assignment rule assigns a level L to the morpheme in the masculine form.

The rule which expresses the correspondence between the masculine and feminine forms is as follows:

(25) TONE SPREADING (Hausa)

\[ L \rightarrow H / H \quad L + \]

This rule is formally similar to Bambara and Maninka TONE SPREADING. The plus sign in the environment is needed to prevent jinjir+ii \H L L \H H L from becoming *jinjir+ii \H H L. Thus, the rule does not apply to the masculine forms; in the feminine, it converts the underlined vowels in (24) from L (which corresponds to the L in the masculine) to H. Does this rule apply to tone features on vowels, or does it apply to suprasegmental tone features? To answer this question, it suffices to examine a masculine-feminine pair in the same agentive construction exemplified by (24b), but this time containing a verb with more than one syllable.

(26) ma+aikat+ii (m.)  ma+aikat+ii+aa (f.)  'worker'
    \H L L L \H H H L L

In the masculine, the verb root aikat is assigned a level low-toned contour, and in the feminine the entire root has its tone raised to H. If we apply TONE SPREADING to segmental tones, the following incorrect derivations will result:
If we assume that TONE SPREADING does not apply iteratively, the incorrect outputs *ma+aikat+ii* and *ma+aikat+ii+aa* will be obtained. Even if we permit iterative application, the first of these incorrect outputs will still not be avoided. What TONE SPREADING needs to capture is that the level low tone on aikat acts like a single unit; if aikat fits into the environment of TONE SPREADING, the tone pattern of the entire morpheme must be raised to H. This is precisely what the suprasegmental analysis would yield.

This constitutes a second instance in which a phonological rule treats tones as suprasegmental units.

2.5. The discovery of facts which appear to pose a problem for the segmental representation of tone is by no means unprecedented. In the past, some researchers have responded to this by extending the segmental descriptive apparatus in various ways in order to meet the demands of the problematical tonal facts.

Consider, for example, the issue of "floating tones". Bird (1966) has posited a "floating" L for the Bambara definite marker. This tone has no segmental realization; the L simply becomes attached to the immediately preceding vowel, which, if it is H prior to attachment, will then be realized as HL; less convincingly, perhaps, Bird (1968) has posited a "floating" H to mark the Bambara indefinite, as seen in example (15). Similarly, Hyman (1972a,b) proposes a "floating tone" for the Igbo and Fẹ́ẹ́ọ̀ associative marker.

In the works cited, the existence of such tones hardly appears to constitute a serious problem; this is due largely to the fact that the authors did not concern themselves sufficiently with a formalization of their results. However, in works like Schachter and Fromkin (1968), Fromkin (1972), and Maddieson (1971), where
formalism was a concern, the problem immediately surfaces. This
points up one of the values of formalism in grammar. The formalism
in some cases brings up anomalies and in other cases suggests ex-
planations which would otherwise go unnoticed.

Schachter and Fromkin (1968) represent "floating" L in Akan
with the matrix [+segment, +L], with no other segmental specifica-
tions. This proposal is clearly ad hoc, since no other segmental
feature has ever been shown to be capable of this sort of representa-
tion. For example, it has not been proved that languages need
matrices like [+segment, +continuant] with no other feature speci-
fications filled in. This constitutes an anomaly on the segmental
view. Why should one type of segmental feature be able to "float",
when no other segmental feature appears to have this property? If
we permit this question to go unanswered, the statement that tone
is a segmental feature becomes undisconfirmable and therefore un-
interesting; the claim that a given entity is a segmental feature,
if accompanied by an explanation for why this entity is capable of
behavior different from that of other segmental features, opens the
door to an endless number of possible claims of segmental status
for entities that clearly are not segmental. Hence, this proposal
for a "segmental" representation for "floating" tones owes us an
explanation before we can be expected to accept this view.

The suprasegmental theory, on the other hand, practically
leads us to expect the existence of "floating" tones. This theory
says, roughly, that two separate sequences of feature matrices can
be given for a given lexical entry in a tonal language; the first
represents a sequence of segments, and the second represents a se-
quency of suprasegmentals. We know that in the languages argued
above to have suprasegmental tones lexically, there exist certain
morphemes which are nonetheless inherently toneless, their tones
being in all cases assigned by phonological rule. Such morphemes
are thus specified underlyingly with only one of the two permissible
sequences of feature matrices, the segmental one. In view of this,
we may also expect that in such languages other morphemes will be
specified only with the sequence of suprasegmental feature matrices;
this is exactly what a "floating tone" is.
Other proposals which have attempted to escape the suprasegmental nature of "floating tones" are subject to pretty much the same criticisms as the unspecified [+segment] proposal.
Maddieson's (1971) assignment of tone features to "segments" which have the feature [-syllabic] but no other segmental features is obviously as ad hoc as the proposal which has [+segment] where Maddieson has [-syllabic]. Under a proposal in Fromkin (1972), a "floating" L would be represented by the matrix [-segment, +L]; the assignment of a "segmental" feature to [-segment] is, again, totally ad hoc, leaving unanswered the question of why languages do not appear to contain matrices like [-segment, +strident].

2.6. For similar reasons, any theory which requires that tone be expressed exclusively as a segmental feature is hard put to explain how restrictions like the ones noted in sections 2.1 and 2.2 could exist. Taking the Mende prohibition against the sequence HLH as an example, the segmental theory would lead us to expect similar prohibitions to apply to clear cases of segmental features. We might expect that it would be possible for a language to prohibit the sequence [+nas] [-nas] [+nas] in words of three segments, while permitting the sequence [-nas] [+nas] [-nas]. In this language, the restriction would rule out words like mam, ata, and iln while permitting words like amb, tät, and ini. It is quite certain that restrictions of this type are nonexistent in natural languages.

Another type of case that we might look for is a language in which an undisputed segmental feature, such as [round], was subject to a prohibition of the following form:

\[
\begin{align*}
(29) \quad &a. \quad C^o \quad +F \quad C^o \quad +F \quad C^o \quad +F \quad C^o \\
&b. \quad C^o \quad +F \quad C^o \quad +F \quad C^o \\
&c. \quad C^o \quad +F \quad C^o \quad +F \quad C^o
\end{align*}
\]

where the tie \(\sim\) is taken to indicate a sequence of features on a single segment. There are two problems here. The first involves the expression of sequences of features on single segments; this will be examined below in the discussion of contour tone features.
The second problem involves the expression of "melodies" containing transitions in segmental feature specifications ranging over lexical entries or over the vowels of lexical entries. One type of phenomenon that may appear related to this is vowel harmony, whereby all of the vowels up to a given point in a word are specified as agreeing in some feature, such as [tense], [back] or [round]. But this is the most trivial sort of segmental "melody", involving not a transition in values of a feature but rather agreement in the value of a feature. One might also mention a restriction in a language which would permit words of the form VCV but rule out words of the form CVC. There are many languages that appear to exhibit precisely this kind of restriction. Here, however, it is necessary to distinguish between prohibitions or rules applying to segments and prohibitions or rules applying to segmental matrices.

Consider, for example, phonological metathesis rules. The examples that I have run across, such as those in Greek, Arabic, and Sundanese, involve a rule which interchanges two segments in a given environment. Examples involving the interchanging of a segmental feature, such as a putative operation on [+nasal] [-nasal] in na that would reverse these features, yielding tā, are unattested. This suggests as a first approximation that phonological metathesis rules, which many researchers have found to be an overly powerful device anyway, be restricted to applying only to segments. The same sort of restriction can also be suggested for morpheme structure conditions like the prohibition of CVC mentioned in the preceding paragraph. This would rule out the use of segmental features to specify possible and impossible segmental "melodies".

What effect does this have on the representation of tone? Above, it was hypothesized that lexical entries can be specified for two separate sequences of phonological matrices, one of them segmental and the other suprasegmental. In this view, a given tone has the same status with respect to a suprasegmental sequence as a given segment has with respect to a segmental sequence. Since the constraint on metathesis rules was proposed solely to rule out the possibility of phonological rules which interchanged single segmental features, we might expect that just as rules can metathesize
full segmental matrices, so can they metathesize full suprasegmental matrices. And, in fact, rules which metathesize sequences like HL have been proposed.

This in turn leads us to expect that, since constraints on sequences of segments are expressible in phonology, constraints on sequences of tones should also be expressible, and this is exactly what we have in the case of the Mende prohibition against HLH. The segmental view, on the other hand, would be forced to rule out the impermissible operations discussed here except when they apply to tone features. The suprasegmental theory is not simply defining away the problem here by calling tone features non-segmental; rather, for reasons quite independent of the present problem (cf. sections 2.3 and 2.4), it has established that two separate sequences of matrices are associable with a given lexical entry in a tonal language; from this, it follows that a tone has the same status with respect to a suprasegmental sequence as a segment has with respect to a segmental sequence. Hence, all other things being equal, the fact that a segment undergoes certain operations should lead us to expect that a suprasegmental can also undergo these operations. This expectation is confirmed by the facts. It is perhaps belaboring the point to observe that the segmental theory leaves us with no expectations in this regard. Since this theory is forced to say that tone features behave differently from other segmental features, it makes no predictions about how tone features actually do behave.

2.7. Another phenomenon which the segmental theory of tone appears unequipped to deal with, involving syllable inversion games in languages, was first brought to my attention by Larry Hyman. Haas (1969) points up a language game in Thai and Burmese in which the syllable finals of adjacent words in a constituent are interchanged; for example, the bracketed expressions in the Thai example (30a) are interchanged, producing the output in (30b).

(30) a. k[ôn] j[àj] 'big bottom' b. käjôn

Note that the tones move along with the rest of the syllable finals. Haas compares this game with English spoonerisms, such as the conversion of half-formed wish into half-warmed fish. Jean-Marie
Hombert in some as yet unpublished work has established that a similar language game is played in Bakwiri, a Bantu language of Cameroon. In this game, the first syllable is moved to the end of a word. For example, the word in (31a) is converted into (31b):

(31) a. lùùngá 'stomach'
    b. ngaâlú

Note that in this language, the tone (as well as vowel length) pattern of the input is preserved in the output—the tone does not move along with the shifted syllable. If we can establish independently that tone is a suprasegmental phenomenon in this language, a likelihood in a Bantu language, then the suprasegmental theory has an explanation for this difference between the language games of Bakwiri and Thai, namely, that in Bakwiri the tone pattern is a property of the word, while in Thai, as can be gathered from the discussion of Thai in section 1.1, tone is a property of individual segments and hence moves along with the segments. The theory that tone is only a property of segments provides no explanation for this difference.

3. CONTOUR TONE FEATURES

3.0. One of the most difficult questions to answer in the study of tone concerns the possibility of expressing contour features, such as [Rising], [Falling], and [Rising-Falling]. If such features are disallowed, the suprasegmental theory would receive another boost, since it alone appears capable of adequately expressing the underlying tone of Mende words like \(
\text{mbu}_\text{HL} \text{ 'owl', mba}_\text{LH} \text{ 'rice', and mba}_\text{LHL} \text{ 'companion'},
\)
all of which have been reported to contain a short underlying vowel. The alternative of representing these contours by adding entities labelled [+segment], [-segment], or [-syllabic] capable of bearing tone features was discarded in section 2; other conceivable alternatives to the suprasegmental theory will be treated in this section.

It would perhaps be reckless to suggest that underlying contour features do not exist; for example, Wang (1967) has presented interesting cases of phonological rules which employ these features in Chinese, and in recent unpublished work Margaret Sung has demon-
strated that rules employing Wang’s contour features have wide application in Amoy Chinese; to reduce [Rising] to LH, and so on, in these cases, would complicate the rules immensely and needlessly.

However, it is possible to demonstrate that in certain languages which exhibit surface contour tones on single syllabic nuclei along with complex tonal melodies, such as those in Mende, in lexical representations, these contour tones do not have the properties of known phonological features; specifically, these contours do not behave as indivisible units, as the feature representations [Rising], [Falling], and [Rising-Falling] would imply.

3.1. An instance in which this point is easy to see is the phenomenon of downdrift in African languages. The phenomenon involves the lowering of successive high tones when they are preceded by low tones. (Successive non-contiguous low tones may also be lowered.) For example, the phonological sequence HLHLH would, if this contour remained unchanged in the course of a derivation, be assigned the phonetic pitch representation in brackets below:

\[
\begin{array}{c}
\text{H L H L H} \\
- - - \\
- - 
\end{array}
\]

The way that pitch assignment is normally expressed (e.g. in Schachter and Fromkin (1968) and Voorhoeve, Meeussen and de Blois (1969)) is by having a preceding low tone cause a lower pitch to be assigned to a high tone than would otherwise be the case. Now, in any language with downdrift, it so happens that the initial part of a falling tone is subject to downdrift just as is any high tone; furthermore, the final part of a falling tone occasions downdrift, just as any low tone would. The same point can be made with reference to rising tones and rising-falling tones. This fact is an immediate consequence of regarding a falling tone as a sequence of HL, a rising tone as a sequence of LH, and so forth. One apparent alternative to treating contour tones as sequences of level tones would involve adding the redundant specifications [+H] and [+L] to contour features, as illustrated below:
(33) a. Falling b. Rising c. Rising-Falling

Not only is this use of distinctive features entirely ad hoc—contradictory specifications such as [+H, +L] are unheard of in phonology—this solution does not even work! The reason is that, although we know that the initial part of a falling tone behaves like a low tone, there is nothing in this solution which prevents us from incorrectly taking the feature [+H] in the matrix (33c) as causing the initial part of this tone to undergo lowering after a low tone (as the feature [+H] in the matrix (33a) does), or as permitting a high tone following this tone to remain on the same pitch as the final part of this tone (as the feature [+H] in the matrix (33b) does). To avoid this undesirable situation, it is necessary to add to the grammar a convention for interpreting the feature complex [+Convex, +H, +L] as a sequence L-H-L, and so on. Needless to say, this extra convention is unnecessary in the suprasegmental theory, which represents the tone in question directly as the sequence L-H-L. The theory of segmental representation, on the other hand, is here forced to recognize the feature complexes in (33) as sequences also, by adding a convention which states this. But this contradicts what could otherwise be a basic claim of the theory of segmental representation—namely, that the features [Rising], [Falling], and [Rising-Falling] are indivisible features. Even if this contradiction can be avoided, the segmental theory incorporates a great deal of ad hoc machinery in trying to express a point elegantly stated in the suprasegmental theory: contour tones are sequences of level tones in languages with a level of suprasegmental representation.

3.2. Another instance in which contour tones do not behave like indivisible units involves the rule of TONE COPYING, reported in Leben (1971c, forthcoming) for Hausa and Mende. This rule copies the immediately preceding tone onto an inherently toneless element. When the immediately preceding element has what might be described as a contour tone, such as Mende \( \text{mbu}_{\text{HL}} \) or \( \text{mba}_{\text{LH}} \), the tone copied is not a falling or a rising tone, but rather the last level tone of
the sequences HL or LH. The suprasegmental theory, which disallows contour features in languages with a suprasegmental level of representation, again makes the correct prediction.

4. TONE MAPPING RULES

4.0. If we accept the proposal that a suprasegmental level of representation is possible in some languages, there are some questions to be answered concerning the nature of the extension of phonological theory that such a level would require and concerning the apparent uniqueness of tone in appearing in both suprasegmental and segmental representations.

4.1. The evidence presented in section 3 is sufficient to warrant any extension of phonological theory that would be required in order to permit the expression of a suprasegmental level of representation for lexical items whose tones exhibit the non-segmental behavior pointed out. In fact, the required modification of phonological theory can hardly be termed an extension at all. There are already many clear cases of features attached to lexical entries: lexical category features, features marking a given lexical entry as an exception to a general phonological rule or as undergoing a minor rule, and the like. Among these features is at least one "sequence" feature, the feature which specifies the order of segments in the lexical item. Suprasegmental representation requires simply that another sequence feature be expressed on a lexical item, a feature specifying the order in which suprasegmentals occur on the lexical item. In addition, a phonological mapping rule is required, to map the suprasegmentals onto the segmentals at some point in the derivation. This type of rule can be shown to exist independently of tone; the demonstration will also serve to point out that tone is not completely idiosyncratic in its behavior.

4.2. Bendor-Samuel (1970) points out that the first person marker in Terena is a nasalization prosody which is mapped onto the unmarked form of a noun or a verb by the following rule:

(34) FIRST PERSON (Terena).

a. Nasalize all vowels and semivowels in the word up to the first stop or fricative
b. Nasalize the first stop or fricative in the word as follows: mb replaces p, nd replaces t, ñg replaces k, nz replaces both s and h, and ng replaces both f and hy

Examples of the operation of this rule are given below; the unmarked noun to which the rule applies is in fact the third person form.

(35) a. emo'ú 'his word'   ēmō'ū 'my word'
    b. ayo 'his brother'      āyō 'my brother'
    c. owoku 'his house'      ōwōngu 'my house'
    d. piho 'he went'          mbiho 'I went'
    e. ahya'afo 'he desires'   ānya'afo 'I desire'

Bendor-Samuel also reports a formally similar phonomenon involving a palatalization prosody for the second person marker; however, its description appears less straightforward than the above nasalization rule, which will in any case, suffice to make the point. Bendor-Samuel's arguments show that in this case, to treat the feature [nasal] as a segmental feature in underlying representations would complicate the description and would completely distort the plausible observations in (34). In this case, the transition from [+nasal] to [-nasal] on first person forms must rather result from some rule triggered by the presence of the first person morpheme. The fact that in the course of a derivation, the feature [nasal] is incorporated into segments and that the incorporation process is subject to segmental information (namely, it must find the first stop or fricative in the word) provides some independent basis for determining what kind of behavior we are to expect from a prosodic tone feature; if the tone feature does not behave in a similar fashion, this will constitute counter-evidence for the claim that tone is a prosodic or suprasegmental feature.

4.3. An instance in which the operation of a tone mapping rule appears to have some of the interesting properties of the Terena nasalization rules is provided by Ngizim, a Chadic language of Nigeria, described in Hyman and Schuh (1972). Ngizim has a rule of TONE SPREADING, which causes certain low tones to become falling tones after a high tone. This is illustrated by the following examples:
If we view tone as a suprasegmental in Ngizim, then we might say that what is happening in these cases is that the H of the first element is combining with the LH of the second element to form a contour HLH, and that subsequently this contour is mapped onto the segments of the words involved by a rule which has the effect of placing the tones where they are on the right in (36). Hyman and Schuh note, however, that Ngizim TONE SPREADING must be exempted from applying in those cases where the spreading high tone would have to cross over a voiced obstruent, such as the underlined b in (37a). That is, we do not get (37b).

This process in Ngizim appears to be formally quite similar to the mapping of a nasal melody defined for Terena in (34); recall that in the Terena rule, what determined the transition point between [+nas] and [-nas] in the mapping rule was the location of the first obstruent. In Ngizim, the spreading of the initial H of a HL or HLH melody is stopped by the first voiced obstruent. In both Terena and Ngizim, the selection of the point at which [+nasal] and [+H] are to stop spreading is phonetically motivated; but the phonetics does not make it necessary that the spreading stop at these precise points, since it is not phonetically impossible for languages to extend nasализation across obstruents and rules which extend high tone across voiced obstruents. Thus, in Terena and Ngizim, a condition containing segmental information must be placed on the mapping rules; the rules themselves appear to be quite similar, involving the association of nasализation or tonal "melodies" with certain strings of segments, and the conditions on the rules are obviously similar.

We should expect that before the mapping rules convert these prosodic patterns into features on segments, the patterns will not be capable of undergoing operations that depend on the nature of the
segments over which they range. This is captured by the suprasegmental theory as an automatic consequence of a constraint that is already built into present phonological theory. That is, for the same reason that the present formalism will not permit the operation described in (38a), it will prohibit (38b):

(38) a. Lexical items with the feature [+First Conjugation] acquire the feature [-Rule 1001] when their final segment is [-consonantal]

b. Lexical items whose suprasegmental matrix has the positive value of the feature [[+H][−H]] acquire the positive value of the feature [[+H][−H][+H]] when their final segment is [-consonantal]

Notice that this does not prohibit all operations on suprasegmentals; in particular, it permits the suprasegmental TONE SPREADING rules motivated in section 2, which do not refer to segmental information. The question of whether these rules are in fact types of mapping rules is tackled in Leben (forthcoming), where it is shown that mapping is a distinct process.

As a consequence, this analysis provides an ordering principle for the operations described in other papers at this Symposium; it says that any rule in which tone is affected by surrounding segments must be ordered after any rule in which tone is represented as a suprasegmental. Similarly, in a language with a nasalization phenomenon like Terena's, we would obviously expect the interaction of nasals and surrounding segments to await the mapping of the nasalization melody onto segments. This perfectly obvious principle goes unstated in a purely segmental analysis.

NOTES

*This work was supported in part by the NDEA African Language and Area Center at Stanford University. This is a modified version of papers presented at the U.S.C. Symposium on Consonant Types and Tone, the Fourth Annual Conference on African Linguistics at Queens College, C.U.N.Y., and a linguistics colloquium at the University of Massachusetts, Amherst. I am grateful to the participants at these meetings for their thoughtful comments and am particularly indebted to Larry M. Hyman and Morris Halle for their detailed criticisms.

1The change on the final syllable from M to L, which does not appear to be related to the other phenomena under discussion, may be the result of a common tendency in tonal languages to lower a tone
before pause when its level is intermediate between H and L if it is preceded by HL. This same phenomenon occurs in Hausa, Bambara, and Maninka. In Hausa, for example, there is no Mid tone, but any H following a sequence HL is lowered to a pitch intermediate between the pitch of the preceding H and L; thus, relative to the preceding H and L, this lowered H is a Mid tone. But before a pause, the pitch is only slightly higher than that of the preceding L. This observation has been verified by pitch recordings; some data which demonstrates this phenomenon is published in Wängler (1967). This accounts for the fact that beginning students mistranscribe Hausa HLH before pause as HLL. Essentially the same phenomenon appears to occur in Bambara and Maninka, as reported in Bird (1966, 1968) and Spears (1968). In all of these languages, the pre-pausal lowered H is then lowered to L only when preceded by HL—not, for example, when preceded only by L.

It should be noted that the accuracy of this data has been contested. D. Dwyer (personal communication) has observed that mba (LHL), although it is given in Spears (1967a,b) with a short vowel, may actually have a long vowel and may have originated from a bisyllabic word. If the vowel is long, this still does not affect my argument insofar as the argument seeks to establish that the permissible and impermissible tone patterns for monosyllables are the same as for other words. If mba originates synchronically from a bisyllabic word, then another instance of a monosyllable with the tonal pattern LHL would have to be sought. In addition, Dwyer has suggested that nikili (LHL) and all other monomorphemic three-syllable words in Mende may be fairly recent borrowings. In this event we would have no native monomorphemic three-syllable words on which to observe the behavior of the pattern LHL. However, there is nothing to suggest that the behavior of nikili (LHL) is irregular in any other respect; furthermore, it is demonstrated in Leben (forthcoming) that bi-morphemic words whose last morpheme is inherently toneless behave tonally in the same way as monomorphemic words. Assuming this, a word like nyaha-ma (LH-L) 'on a woman', could be used to make the same point as nikili (LHL) in the text.

I am grateful to Karen Courtenay for correcting my Bambara facts on several occasions.

A very interesting alternative for expressing the downdrift rule, involving a rule of LOW RAISING, is brought up in Hyman and Schuh (1972). This alternative is still subject to the point made in the text.
REFERENCES


McCawley, J. D. 1964. What is a tone language? Presented at the summer meeting of the Linguistics Society of America.


THE ROLE OF CONSONANT TYPES IN NATURAL TONAL ASSIMILATIONS

BY

LARRY M. HYMAN

LINGUISTICS PROGRAM
UNIVERSITY OF SOUTHERN CALIFORNIA
LOS ANGELES, CALIFORNIA 90007
O. INTRODUCTION

The relevance of consonant types to tone has been known for quite some time. In Oriental languages, tone contrasts have often been seen as deriving from earlier consonant contrasts (see Matisoff 1973 for a general statement). In African languages, the phenomenon of "depressor" consonants, particularly in Southern Bantu (e.g. Xhosa and Zulu), has received considerable attention (cf. Beach 1923). The aim of the present paper is to view this relevance of consonant types to tone within the framework of natural tone rules. Two separate sections will therefore be necessary. First, I shall discuss the nature of tonal assimilations; and second, I shall discuss the ways in which consonant types take part in these natural tonal assimilations. In both cases I shall draw my examples from various African languages. These languages are for the most part distantly related in the Niger-Congo language phylum, though some (in particular, Hausa and Ngizim) are unrelated and belong to the Chadic group of Afro-Asiatic.

It is important to note that the generalizations I shall make in both discussions pertain only to "syllable-tone" languages, i.e. to languages where tone has been analyzed as a property of a phonological unit (i.e. the syllable, or perhaps the segment). In other words, what I shall have to say does not apply to languages (sometimes termed "pitch-accent") which, like stress languages, isolate one phonological unit per word/phrase to receive prominence. Thus, languages such as the European "tone languages" (Scandinavian and Slavic), as well as certain East African Bantu languages (e.g. Safwa (Voorhoeve 1973) and Kinga (Schadeberg 1973)) are not expected to be covered by the generalizations I shall make in the following sections. This should not be surprising. While a language which assigns tone to syllables may change an underlying H-L-L (high-low-low) sequence to phonetic H-H-L as a natural assimilation, such a process would be self-defeating in a language where the organizational principle is the placement of one mark of prominence per word.

The above assimilation has in this case the effect of "spreading prominence over two syllables. Since these European and East
African Bantu languages do not have such "spreading" rules, it follows that only languages organized on the basis of tone per phonological unit display the kind of phonological interference from consonant types presented in section 2 below. While phonological information such as "syllable weight" (Newman 1972) is an important factor, I should like to hypothesize that languages which have stress or tone "placement" rules do not ever restrict these rules according to consonant types. That is, no language has a rule that says: place stress (high tone, rising/falling tone, etc.) on the penultimate syllable unless this syllable begins with a voiced consonant—in such a case, place stress (tone) on the ante-penultimate syllable.

1.0. THE NATURE OF TONAL ASSIMILATIONS

In this section I shall discuss two dichotomies which prove useful in the study of tonal assimilations: vertical vs. horizontal assimilation and anticipatory vs. perseverative assimilation. It will be argued that the role of consonant types in tone languages cannot properly be understood except in the context of a theory of natural tonal assimilations.

1.1. Vertical vs. Horizontal Assimilation

This first dichotomy distinguishes between assimilations in an up and down movement and assimilations from one syllable to another.

1.1.1. In vertical assimilations, tones are raised or lowered in the environment of a higher or lower tone. Thus, in (1),

(1) \[ H \rightarrow M / L \]

a \( H \) (high tone) becomes a \( M \) (mid tone) after a \( L \) (low tone). The most well known case of such a vertical assimilation involving lowering is downdrift. In many African languages a sequence \( H-L-H \) is not realized as on the left of (2), but rather as on the right:

(2) \[ / - - / \rightarrow [ - - ] \]

That is, the two high tones are not pronounced on the same pitch (though they are phonologically identical), but rather the second
high tone is lower in pitch than the first high tone. Stated differently, the interval from H to L is greater than from L to H. Thus, downdrift is frequently viewed as a case of a low tone pulling down a high tone, a vertical lowering assimilation. An Igbo example is given in (3):"  

\[(3) \quad \text{Ọ nà àmpà ūnyà īgwè} \quad \text{'he is trying to ride a bicycle'}\]  

\[
\begin{array}{ccccccc}
\text{H} & \text{L} & \text{H} & \text{L} & \text{H} & \text{L} & \text{L} \\
- & - & - & - & - & - & - \\
\end{array}
\]

As can be seen in the above example, the downdrift effect can extend over several HLH sequences. In this sentence there are no fewer than five distinct phonetic pitch levels. There is some disagreement over whether low tones are also subject to downdrift. Generally, if there is a difference between the pitch values of low tones in such sequences, it is not nearly as perceptible as the discrepancies between high tone pitch levels. In Hausa, however, both H and L are subject to quite marked downdrift (Kraft and Abubakar 1965).

Vertical assimilation involving the raising of a L in the environment of a neighboring H is termed low-raising (Hyman and Schuh 1972). Though not as frequently mentioned in the literature as downdrift, it is characteristic of a number of languages where a low tone is raised before a high tone (and, if present, before a mid tone). In many languages the tendency is present, though not of any particular phonological significance. Thus, while the [o] of [ôké] 'rat' is slightly higher in pitch than the [ô] of [ôpì] 'horns' in Igbo, this difference is barely perceptible. In Mbuiga-Bamileke, on the other hand, the tendency is so marked that the low tone is raised to the level of a mid tone before a high tone (Hyman 1972). Thus, the rule in (4)

\[(4) \quad L \rightarrow M / \_ \_ \_ H\]

is responsible for the realization of /nibáu/ 'breast' as [nibáu]. A similar case is discussed from Ewe in section 2.2, though involving consonant information.
A third vertical pitch adjustment which should be mentioned in conjunction with the above is final-lowering. Many languages have a rule by which a final low (especially the last of a series of L's) falls to a rest position. Thus, Dschang-Bamileke /àpà/'bag' is pronounced [ _ _ ]. It is interesting to note that the Dschang word [lèsɔŋ] 'tooth', on the other hand, is pronounced as a level [ _ _ ], i.e. without the final fall, since it carried a historical final high tone, as in the reconstruction *lèsɔŋ*. This H prevented the preceding L from falling, though it is not pronounced in the synchronic citation form.

1.1.2. Horizontal assimilations result from a non-synchrony between the tones and the segments (syllables) over which they have domain. In a partial horizontal assimilation, a contour tone, either rising or falling, results, as shown in the rules in (5):

(5) a. L H → L L̂ H (i.e. /àwà/ → [àwà])
   b. H L → H Ĥ L (i.e. /áwà/ → [áwà])

A L-H sequence becomes L followed by a rising tone (here indicated as a L̂ H in the same syllable), and a H-L sequence becomes H followed by a falling tone (here indicated as a Ĥ L in the same syllable). Examples from Gwari illustrating the two assimilations are seen in (6):

(6) a. /dʟ̃p̃̃/ → [d₇kp₇] 'length'
   b. /s₈u₇k₇N₇u/ → [s₇U₇k₇] 'bone'

In these examples, the tone of the first syllable has "spread" into the second syllable, creating a rising tone from L to H in 'length' and a falling tone from H to L in 'bone'. This kind of assimilation of the second tone to the first tone is quite different from the vertical tonal assimilations discussed in the previous section. A vertical assimilation would result if the L of /dʟ̃p̃̃/ rose to M, i.e. [ɒkpɒ], or if the H of /dʟ̃p̃̃/ lowered to M, i.e. [d₇kp₇]. In [d₇kp₇] (L-L̂ H), however, we do not have a simple up and down adjustment, but rather the spreading of a L into a H tone syllable.

The need for this distinction between vertical and horizontal tonal assimilations is seen in cases of complete horizontal assimilation. Specialists of tone have frequently pointed out the assym-
metry between L and H in tonal assimilations (Schachter 1969, Stevick 1969). Namely, L seems to assimilate to H more readily than H assimilates to L. Thus, while some languages modify H-L-H to H-H-H (e.g. Mbuï), no language would ever assimilate L-H-L to L-L-L. That is, a L surrounded by H's is infinitely more vulnerable to assimilation than a H surrounded by two L's. While this is certainly true, there are well-documented cases of H assimilating to L in a systematic fashion, and these are particularly frequent in African languages (especially Bantu, but also Maninka (Leben 1973) and Ngizim (see below)). The form of the rule is given in (7a); this should be compared with (7b), which does not occur.

(7) a. L H H → L L H
   b. L H L → *L L L

A H can assimilate to a preceding L only if this H is followed by a second H. Thus, in Kikuyu (Pratt 1972), /gɔr/ 'to buy' + /ɪrɛ/ 'immediate past tense marker' are realized as indicated in (8):

(8) /gɔr/ + /ɪrɛ/ → [gɔrɪrɛ] 'bought' (imm. past)

The question is: why can L-H-H become L-L-H, but L-H-L cannot become L-L-L?

If we attempt to answer this question by viewing the assimilation as vertical in nature, we immediately fall into an inexplicable situation. If the assimilation in (7a) and the non-occurring but logically possible assimilatory process in (7b) are viewed as vertical, then the H which is surrounded by L's should be more vulnerable to being pulled down (vertical assimilation) than a L which is "protected" by a following H. Instead, a following H makes the lowering of the H of a L-H sequence possible, just as a following L makes it impossible. Once we view this assimilation of L-H-H to L-L-H as the spreading of the preceding L the picture becomes clearer. The spreading rule given in (5a) above predicts that L-H-H will become L-LH-H, i.e. low-rising-high, a case of partial horizontal assimilation. At this point, another horizontal assimilation, termed absorption (Hyman and Schuh 1972), comes into play. The two possible cases are illustrated in (9), though I shall be concerned only with (9a):
(9) a. \( \text{L} \rightarrow \text{L} \) (i.e. [ǎwá] \( \rightarrow \) [àwá])
   b. \( \text{H} \rightarrow \text{H} \) (i.e. [àwà] \( \rightarrow \) [àwà])

A rising (\( \text{L} \)) tone followed by a \( \text{H} \) is realized as \( \text{L} \rightarrow \text{H} \), while a falling (\( \text{H} \)) tone followed by a \( \text{L} \) is realized as \( \text{H} \rightarrow \text{L} \). That is, the second component of a contour tone is "absorbed" into a like tone in the following syllable. The derivation of \( \text{L} \rightarrow \text{L} \) from \( \text{L} \rightarrow \text{H} \rightarrow \text{H} \) must be seen as involving, then, both (5a) and (9a):

\[
\begin{align*}
(10) & \quad \text{L} \rightarrow \text{L} \rightarrow \text{H} \rightarrow \text{L} \\
& \quad \text{(a)} \quad \text{(b)}
\end{align*}
\]

In (10a) \( \text{L} \rightarrow \text{H} \rightarrow \text{H} \) has become \( \text{L} \rightarrow \text{L} \rightarrow \text{H} \) through spreading (5a); in (10b) intermediate \( \text{L} \rightarrow \text{H} \rightarrow \text{H} \) has become \( \text{L} \rightarrow \text{L} \rightarrow \text{H} \) through absorption (9a). As claimed by Hyman and Schuh (1972), the conversion of \( \text{L} \rightarrow \text{H} \rightarrow \text{H} \) to \( \text{L} \rightarrow \text{L} \rightarrow \text{H} \) always involves an intermediate diachronic stage \( \text{L} \rightarrow \text{H} \rightarrow \text{H} \). While the synchronic facts of a language do not always provide proof for the prior existence of such a stage, frequent evidence is available in languages exhibiting such a horizontal assimilation. Thus, in Kikuyu (Pratt 1972), \( \text{L} \rightarrow \text{H} \rightarrow \text{H} \) is realized as \( \text{L} \rightarrow \text{L} \rightarrow \text{H} \), but \( \text{L} \rightarrow \text{H} \rightarrow \text{H} \) before pause, where absorption cannot convert the intermediate \( \text{H} \rightarrow \text{H} \) contour to \( \text{L} \). Also, in Ewe (Stahlke 1971a), \( \text{L} \rightarrow \text{H} \rightarrow \text{H} \) if the intervening consonant is voiced, as seen in (11a):

\[
\begin{align*}
(11) & \quad \text{(a)} \quad /\text{àvú}/ \rightarrow [\text{àvú}] \quad \text{'dog'} \\
& \quad \text{(b)} \quad /\text{àfí}/ \rightarrow [\text{àfí}] \quad \text{'ashes'}
\end{align*}
\]

The \( \text{L} \rightarrow \text{H} \) of 'dog' becomes \( \text{L} \rightarrow \text{H} \) (low-rising), since the intervening consonant [v] is voiced. The \( \text{L} \rightarrow \text{H} \) of 'ashes', on the other hand, remains \( \text{L} \rightarrow \text{H} \) phonetically, since the intervening consonant [f] is voiceless. As seen in (12a),

\[
\begin{align*}
(12) & \quad /\text{àvú}/ + /\text{lá}/ \rightarrow [\text{àvú lá}] \quad \text{'the dog'} \\
& \quad /\text{àfí}/ + /\text{lá}/ \rightarrow [\text{àfí lá}] \quad \text{'the ashes'}
\end{align*}
\]

the \( \text{L} \rightarrow \text{H} \) sequence of 'dog' resulting from (11a) is modified by absorption to \( \text{L} \rightarrow \text{L} \) before \( \text{H} \). On the other hand, the \( \text{L} \rightarrow \text{H} \) of 'ashes' is not affected when a \( \text{H} \) occurs in the following syllable. These facts are consistent with the view that \( \text{L} \rightarrow \text{H} \rightarrow \text{H} \) becomes \( \text{L} \rightarrow \text{L} \rightarrow \text{H} \) in two steps. If the input of (12a) is [àvú lá], the absorption process can be stated without reference to consonant types. On the other
hand, if the input to (12a) is left as L-H-H, then much of the environment of the rule needed for (11a) will be needlessly duplicated. Thus, at least in Kikuyu and Ewe, an intermediate synchronic stage of L-\( \hat{L} \hat{H} \)-H is motivated.

Returning to the case of L-H-L, we find that horizontal assimilations yield the following possibilities:

\[
\begin{align*}
(a) & \\
(b) & \\
(13) & L H L \rightarrow L \hat{L} H L \rightarrow \{L L \hat{H} L\} \{L H L\}
\end{align*}
\]

If spreading occurs to a L-H-L sequence, then a low-rising-low sequence is obtained, as in (13a). This sequence can be modified, again by spreading, to low-low-falling, as in the first line of (13b). Or, a more likely possibility, the language may convert L-\( \hat{L} \hat{H} \)-L back to L-H-L (or equivalently, not allow the change in (13a) to occur at all), since a rising tone followed by a L tends to be simplified to a H (Hyman and Schuh 1972). In fact, the process of absorption seen in (9) can be viewed to be in part motivated by the tendency of African languages to rid the tonal system of contours (especially rising tones). Thus, in terms of contour simplification, the first line of (13b) cannot be seen to be preferable to the second line.

The failure of L-H-L to become L-L-L is therefore dependent on the generalization that rising tones (which result from partial horizontal assimilation) tend to become H before L. An explanation for this fact is what is required. Namely, why doesn’t \( \hat{L} \hat{H} \) simplify to L before L? We can, however, conclude from the above discussion that H does not assimilate to L by vertical assimilation. Instead, the horizontal assimilations of spreading and absorption convert L-H-H to L-L-L. The reverse of this conclusion is not true, however. That is, L can assimilate to H by either vertical or horizontal assimilation.

That this is the case can be seen from the following dialectal realizations of 'three goats' in Igbo:

(14) a. /éwú/ + /átọ/ \( \rightarrow \) [éwú 'átọ] 'three goats'
    b. /éwú/ + /ètọ/ \( \rightarrow \) [éwú è'tọ]
Both Standard Igbo (14a) and Aboh Igbo (14b) start out with a H-H ('goat') followed by a L-H ('three'). In (14a), the L of 'three' assimilates vertically to the level of the (downdrifted) H of the following syllable. The result then is that the intermediate sequence [ -- ] (i.e. H-H-L-H, where the last H is lowered by the preceding L) becomes [ -- ]. In (14b), on the other hand, rather than saying that the L of 'three' assimilated directly to the preceding H of 'goat', it is argued in Hyman and Schuh (1972) that an intermediate H-H-H-L-H (i.e. with a falling tone on the L of 'three') derived from partial horizontal assimilation is motivated. Rather than repeating the arguments presented in Hyman and Schuh, let us simply note that many languages show a tendency to convert H-L-H to H-H, i.e. [ -- ] (a H followed by a "downstep"). The second H is of course lowered by the L component of the preceding falling tone. For example, where Ngizim allows H-L-H to develop into H-H-L-H (see section 2.1 below), the H-L-H sequence will be progressively more likely to develop into H-L-H as the original H-L syllable has less and less "weight" (Hyman and Schuh 1972). Thus, if the H-L occurs on a CV syllable, H-L-H will be extremely likely to become H-H. Thus, it is argued that Aboh Igbo H-H-L-H, as in (14b), becomes H-H-H-H, i.e. [ -- -- -- ], through an intermediate H-H-H-L-H.

1.2. Anticipatory vs. Perseverative Assimilation

The second dichotomy relevant to the study of tonal assimilations is that between anticipatory (progressive) and perseverative (regressive, lag) assimilations. In other words, given two tones, does the first assimilate to the second, or does the second assimilate to the first?

1.2.1. As already seen in section 1.1.1 above, both anticipatory and perseverative vertical assimilations occur. Thus, when L becomes M before H (cf. (4) above), this is a case of anticipatory assimilation. The L is raised in anticipation of the following H. On the other hand, when H becomes M after L (cf. (1) above), this is a case of perseverative assimilation. The H is lowered as a perseverative or lingering effect of the preceding L.

While vertical assimilations exhibit both kinds of effects,
it is important to note that there are restrictions on the naturalness of vertical anticipatory and perseverative assimilations. Thus, while the assimilations in (15a) have been said to be natural and frequently attested, those in (15b) are unnatural and almost non-attested:

(15) a. \( L H \rightarrow L M \) (i.e. \(/àwá/ \rightarrow [àwà])
    \( L H \rightarrow M H \) (i.e. \(/àwá/ \rightarrow [àwà])

b. \( H L \rightarrow *M L \) (i.e. \(/áwà/ \rightarrow [áwà])
    \( H L \rightarrow *H M \) (i.e. \(/áwà/ \rightarrow [áwà])

While \( L \) can pull down a following \( H \), and a \( H \) can pull up a preceding \( L \) (15a), a \( L \) does not generally pull down a preceding \( H \), nor does a \( H \) pull up a following \( L \) (15b). In other words, a \( L-H \) sequence is vulnerable to vertical assimilations (both anticipatory and perseverative), but a \( H-L \) sequence is not.

1.2.2. While vertical assimilations exhibit both anticipatory and perseverative rule types, horizontal assimilations are generally perseverative in nature. Thus, while the assimilations in (16a) have been said to be natural and frequently attested, those in (16b) are unnatural and almost non-attested:

(16) a. \( L H \rightarrow L Lâ \) (i.e. \(/àwá/ \rightarrow [àwà])
    \( H L \rightarrow H Hâ \) (i.e. \(/àwá/ \rightarrow [àwà])

b. \( L H \rightarrow *Lâ H \) (i.e. \(/áwà/ \rightarrow [áwà])
    \( H L \rightarrow *Hâ L \) (i.e. \(/áwà/ \rightarrow [áwà])

While \( L-H \) frequently develops into a low-rising sequence, and \( H-L \) into a high-falling sequence (16a), \( L-H \) generally does not become rising-high, nor does \( H-L \) become falling-low (16b). Thus, whenever there is a non-synchrony between the tones and their assigned syllables, tone-spreading occurs perseveratively into following syllables, and not anticipatorily into preceding syllables. That is, the earlier tone lasts too long, rather than the later tone starting too early. In so doing, an earlier tone enlarges its domain to encompass part or all of a following syllable.

One of the important generalizations that can be made about tone spreading is that in a given language, it always takes place first where the interval between the two differing tones is great-
est. Thus, in a three tone system with H, M and L, it is most likely to occur in L-H and H-L sequences first, and only then can it extend to L-M, M-L, etc. Thus, the following implicational universal can be stated: if in a language tone spreading takes place from a M into a following L (i.e. M-L becomes M-Ṃ\textsuperscript{L}), then tone spreading also takes place from a H into a following L (i.e. H-L becomes H-Ṃ\textsuperscript{L}). The reverse is not true, i.e. that H-L spreading implies M-L spreading, since Yoruba is characterized by the former, but not the latter. In Gwari, on the other hand, both H-L and M-L are subject to spreading, as seen in (17):

(17) a. /súkünü/ → [sūkû] 'bone' (cf. (6b))
    b. /ōz̄a/ → [ōzā] 'person'

Because of this implicational universal, tone spreading from M to L in Gwari must be a more recent innovation than tone spreading from H to L (which characterizes Yoruba, as well as Gwari). On the other hand, no language has been found to have tone spreading from M to L without also having tone spreading from H to L.

While Gwari has developed two falling tones through spreading, I know of no language which has developed two rising tones in a similar way. That is, no language to my knowledge converts L-H to L-Ṃ\textsuperscript{H} and also L-M to L-Ṃ\textsuperscript{M}. This presumably has to do with the fact that L-H and L-M are subject to the vertical assimilations treated in the previous section. The interval between them is then more reduced than that between H or M and L. Furthermore, when Ṃ\textsuperscript{H} or Ṃ\textsuperscript{M} is assigned to one syllable, it is much more likely that the interval will be narrowed than if the tones were on separate syllables. Thus, in Nupe, the rising tone resulting from spreading in the example in (18),

(18) /ādè/ → [ādə] 'cloth' [\underline{\_\_}]  

barely rises to the level of a mid tone. On the other hand, where spreading cannot occur (because of an intervening voiceless consonant—cf. section 2.1 below), the interval between L and H is much greater (there is no downdrift in Nupe):

(19) /ātā/ → [ātā] 'speed' [\underline{\_\_}]  

The tendency to reduce the interval between a L and a H in the same
syllable is so great that there would perhaps be no room for a second distinct rising tone to develop from L-M. Instead, the phenomenon of a "lowered-mid" is sometimes observed, as in the Gwari example in (20):

(20) /òdã/ → [òdã] 'father' [ _ _ ]

Notice, finally, that there does not appear to be any implica-
tional relationship between rising and falling tones resulting from spreading. Some languages exhibit spreading from L to H but not from H to L (Nupe), while some languages exhibit spreading from H to L but not from L to H (Aboh Igbo). However, since downdrifting and low-raising decrease the interval from a L to a following H, we might expect to find fewer instances of L-H developing into L-ampil than H-L developing into H-ampil in languages characterized by down-

2.0. CONSONANT TYPES AND NATURAL TONE PROCESSES

As can be inferred from the preceding section, the kinds of tonal assimilations exemplified above sometimes receive interference from consonant types. Let us consider first the case of spreading (horizontal assimilation):

2.1. Horizontal (Perseverative) Assimilation

It has already been seen in (16) that Nupe has a rule of L-H spreading. As seen in (21) (where /à/ is the present progressive marker), this spreading is restricted to cases where the intervening consonant is voiced (George 1970):

(21) /pà/ 'peel' : [èpà] 'is peeling'
    /bà/ 'be sour' : [èbà] 'is sour'
    /wà/ 'want' : [èwà] 'is wanting'

This rule can be written informally as in (22):

(22) L [+voice] H → L [+voice] L-ampil

Where the intervening consonant is [-voice], tone spreading does
not occur (e.g. [èpá]).

While Nupe exhibits a *partial* horizontal assimilation restricted by consonant type, Ngizim exhibits a corresponding *complete* horizontal assimilation (Schuh 1971). As seen in the examples in (23), L-H-H becomes L-L-H if the second syllable begins with a voiced consonant:

(23) /mùgbà bái/ → [mùgbà bái] 'it's not a monitor'
/màaràm tén/ → [màaràm tén] 'big nose'

If, on the other hand, the second syllable begins with a voiceless consonant (or an implosive—see below), no change occurs:

(24) /šlítá bái/ → [šlítá bái] 'it's not pepper'

That is, we do not obtain *[šlítá bái]. As argued above, Ngizim is claimed to have undergone an intermediate historical stage with only partial horizontal assimilation, as in (25):

(25) /mùgbà bái/ → mùgbà bái → [mùgbà bái]

The first change is comparable to the above Nupe data. L-H develops into L-L-H only if the intervening consonant is voiced. The second stage involves absorption and accounts for the change from L-H-H to L-H in the second and third syllables in the example. Ngizim does not permit rising tones to result from tone spreading.\(^{12}\)

It might be argued that tone spreading is possible in the specific Nupe and Ngizim environments because the intervening voiced consonant allows continuous voicing from the first syllable to the second.\(^{13}\) That this is not the case is first seen from the fact that implosives (which are *voiced*) have the same effect on tone as voiceless obstruents. Thus, in Ngizim, implosives do not allow the spreading of L to occur, as seen in (26):

(26) /kìidá bái/ → [kìidá bái] 'he didn't eat (it)'

That is, we do not obtain *[kìidá bái].

That continuous (vs. interrupted) voicing from one syllable to another is not responsible for the role of voiced consonants in tone spreading is also seen from the facts of H-spreading in Ngizim. As seen in (27),
(27) /ná kàsúw/  →  [ná kàsúw]  'I swept'
    /á rëpcí/  →  [á rëpcí]  'open!'

A H spreads into a following L if the intervening consonant is voiceless or a sonorant (e.g. /r/). On the other hand, as seen in (28),

(28) /ná bàkú-w/  →  [ná bàkú]  'I burned (it)'

H-spreading does not occur when the intervening consonant is a voiced obstruent. Thus, we do not obtain *[ná bákú]. This rule can be informally written as in (29), which should be compared with the rule given in (22) above for L-spreading:

(29) $H \left\{ \begin{array}{c} [-\text{voice}] \\ [+\text{son}] \end{array} \right\} L \rightarrow H \left\{ \begin{array}{c} [-\text{voice}] \\ [+\text{son}] \end{array} \right\} H_L$

However, this rule is not quite complete, since implosives also allow the spreading of H into a following L syllable:

(30) /ná dànkú-w/  →  [ná dànkú]  'I sewed'

Ignoring the implosives for the moment, the interaction of consonants and tone in Nupe and Ngizim can be summed up as in (31), where /p/ stands for the class of voiceless obstruents, /b/ the class of voiced obstruents, and /w/ the class of sonorants:

(31) /òpà/  →  [òpà] /òpà/  →  [òpà]
    /àbà/  →  [àbà] /àbà/  →  [àbà]
    /òwà/  →  [òwà] /òwà/  →  [òwà]

In Hyman and Schuh (1972) it was argued that voiceless obstruents block the spreading of L and voiced obstruents block the spreading of H. On the other hand, sonorants exert no blocking effect, but rather allow any tone to spread through them. In other words, /p/ is impermeable to low tone and /b/ is impermeable to high tone.¹⁴

Since L-H and H-L tend to become L-H and H-H as a natural horizontal assimilation, it can now be observed that the natural tendency of tones to assimilate sometimes encounters obstacles from intervening consonants. Voiceless obstruents are adverse to L-spreading, and voiced obstruents are adverse to H-spreading.
The inherent propensities of consonants and tones are thus often in conflict with one another. In some languages (e.g. Nupe, Ngi- zim, Ewe, Zulu), the consonants win out, and tone spreading occurs only where the consonants are favorably disposed to it. In other languages (e.g. Yoruba, Gwari), the tones win out, as tone spreading takes place regardless of the disposition of intervening consonants.

The important fact in the above examples is that the consonants should not be seen as motivating the tonal assimilations, but rather as permitting vs. blocking them. If the consonants are viewed as providing the primary motivation for the tonal assimilations in Nupe and Ngizim, then the following assimilations of tones to consonants in (32) should also be found in languages:

(32) a. /pə/ → *[pə]
    b. /bə/ → *[bə]

That is, we should expect to find that the "high toneness" of voiceless obstruents would cause a L to become a ĤL fall as in (32a). Similarly, we should expect to find that the "low toneness" of voiced obstruents would cause a H to become a L̂H rise, as in (32b). As far as I know, such "assimilations" are unattested in African languages.15

In addition, if the consonants provide the primary motivation for the above assimilations in Nupe and Ngizim, the assimilations in (33) should also be found:

(33) a. /əpə/ → *[əpə]
    b. /əbə/ → *[əbə]

That is, a L-L sequence with an intervening voiceless obstruent should develop into a L followed by a ĤL fall, and a H-H sequence with an intervening voiced obstruent should develop into a H followed by a L̂H rise. Neither of these is attested as a natural assimilation in tone languages. Rather, the more deeply one investigates languages with consonant interaction with tone, the more widespread the principles outlined above appear to be.

As a further illustration, consider the Xhosa example cited
by Leben (1973). As shown in (34),

(34) /ínjá/ → [ínjá] 'dog'

the (synchronic) derivation of H-LH from a sequence of H-H with an intervening voiced obstruent ("depressor consonant") appears to be an instance of rule (33b). While a language could conceivably have such a synchronic rule, it is only as a result of tonological restructuring. As is evident from the Swazi pronunciation [ínjá] (Euphrasia Kunene, personal communication), this word reconstructs as HL-H. The L of the HL noun prefix spreads into a following H syllable if the intervening consonant is permeable to low tone spreading. In fact, in Zulu (Cope 1970) /ínjá/ is pronounced [ínjá], where the final H is "displaced" by the incoming L of the HL prefix. Thus, the original spreading nature of the assimilation is revealed from this diachronic analysis.

Elsewhere in Zulu the same blocking effect of H-spreading by voiced obstruents seen in Ngizim is observed (Cope 1970):

(35) a. /a+bá+y₁+bónl/ → [ábáy₁'bóni] 'they do not see it'
   b. /a+bá+z₁+bónl/ → [ábáz₁bóni] 'they do not see them'

In (35a) the H of /bá/ 'they' spreads into the L of /y₁/ 'it'.
(The following syllable of /bóni/ 'see' is realized on a downstep, i.e. a H lowered by the preceding L, which is itself smothered by the spreading of the H of /bá/.) In (35b), on the other hand, spreading does not occur, because the /z/ of /z₁/ 'them' blocks it. Since Ngizim and Zulu are genetically unrelated and are separated by thousands of miles, the complete convergence of the two systems in regard to consonant types and tone spreading provides support for the view presented in this section.¹⁶

2.2. Vertical (Anticipatory) Assimilation

It has already been asserted that horizontal assimilations are generally perseverative in nature. Thus, in order to examine the nature of consonant interference in anticipatory tonal assimilations, it is necessary to turn to vertical tonal assimilations.¹⁷ A particularly revealing case of such interaction between consonant types and anticipatory vertical tonal assimilation is found in Ewe (Stahlke 1971a,b).
There are two related instances of low-raising rules which refer to consonant types in Ewe.\(^1\) The first involves low tone consonant-initial syllables which become mid as observed in (36):

\[
\begin{align*}
(36) & \text{a. } \text{dá 'snake': dà lá 'the snake'} \\
& \text{há 'pig': hà lá 'the pig'} \\
& \text{b. } \text{fù 'sea': fù lá 'the sea'} \\
& \text{kpol 'stick': kpô lá 'the stick'} \\
& \text{c. } \text{nyi 'cow': nyî lá 'the cow'} \\
& \text{ɣè 'sun': ñë lá 'the sun'}
\end{align*}
\]

In these examples (taken from Stahlke) a \text{L} becomes \text{M} before \text{H} unless the initial consonant of the \text{L} syllable is a voiced obstruent. (Stahlke identifies [h] as a voiced pharyngeal fricative and [ɣ] as a velar approximant = sonorant.) The rule can be informally written as in (37):

\[
\begin{align*}
(37) & \left\{ \begin{array}{c}
\text{[-voice]} \\
\text{[+son]} \\
\end{array} \right\} \text{L H} \rightarrow \left\{ \begin{array}{c}
\text{[-voice]} \\
\text{[+son]} \\
\end{array} \right\} \text{M H} \\
& \text{C} \quad \quad \text{C}
\end{align*}
\]

The tendency of \text{L-H} to become \text{M-H} has already been pointed out and illustrated. Thus, while rule (37) incorporates a consonantal specification of either \text{[-voice]} or \text{[+son]}, these feature specifications do not in themselves provide the primary motivation for the tonal assimilation. Rather, their complement—namely \text{[+voice, -son]}, the class of voiced obstruents—blocks the application of the natural low-raising tonal assimilation. This is not surprising, since we have observed the low tone nature of voiced obstruents in horizontal assimilations. What is of interest here, however, is that it is the initial consonant of the first syllable—and not the intervening consonant—which blocks the application of this natural tone assimilation. As we shall see, this is important, since although this is a case of vertical assimilation, both the Ewe low-raising (as illustrated so far) and the Nupe, Ngizim and Zulu spreading rules involve the effect of a consonant on the tone of the syllable it initiates (or, equivalently, on the tone of the following vowel).

A case of low-raising in Ewe involving the intervening con-
sonants concerns both the rule in (37), which applies to consonant-initial syllables, as well as the raising of the low tone /à/ noun prefix. The effect of consonant types on preceding syllables is summed up in (38), where the initial /à/ is the noun prefix:

(38) /àpà/ → [àpá]
    /àbà/ → [àbà]
    /àwà/ → [àwà]

The low tone /à/ prefix is raised to M before H only if the intervening consonant is a sonorant. Thus, /à/ stays low in both [àpá] and [àbà], but becomes mid in [àwà]. For the first time we see consonant types grouping in a different fashion. Instead of grouping voiceless obstruents and sonorants (which permit H-spreading and low-raising of a following vowel) and voiced obstruents and sonorants (which permit L-spreading), in (38) both voiceless and voiced obstruents form a class in opposition to the class of sonorants. Thus, both voiceless and voiced obstruents appear to exert a low tone influence on preceding vowels. This fact is particularly intriguing, since voiceless obstruents have been seen to exert a high tone influence on following vowels.

This observation is, however, consistent with the findings reported by Lea (1973) and Matisoff (1973). Lea's acoustic experiments show that voiceless obstruents have a raising effect and voiced obstruents a lowering effect on the fundamental frequency of following vowels, while sonorants ride along the fundamental frequency contour determined by stress and intonational factors. On the other hand, he reports a dip in fundamental frequency on a vowel at the onset of a following (voiceless or voiced) obstruent. Matisoff (1973) points out that in tonogenesis in various Southeast Asian languages the voicing contrast is relevant only with respect to the following vowel, while the glottal consonants /h/ and /ʔ/ can exert a respective lowering and raising effect on preceding vowels.

Returning to Ewe, the same restriction on the intervening consonant illustrated in (38) with respect to the /à/ noun prefix actually characterizes the low-raising situation illustrated
in (36) with respect to consonant-initial syllables. The examples in (36) all involve a L CV-syllable followed by /lá/ 'the', which begins with a sonorant and therefore permits low-raising. The two cases of low-raising are therefore clearly sub-parts of the same rule. This rule can be stated in prose as follows: a L becomes M before a H if 1) any intervening consonant is a sonorant, and 2) the L syllable does not begin with a voiced obstruent. If there is no intervening consonant, i.e. there is an underlying L±H (rising tone) sequence in the same syllable, this L±H is modified to a M±H rise unless the syllable which it is in begins with a voiced obstruent (this is covered by part (l) of the prose rule; cf. Stahlke 1971a).

Finally, it should be recalled from (11a) and (38) that Ewe too has a rule of L-spreading which is dependent on consonant type. Thus, /ávú/ 'dog' is realized as [ávú], but /áfí/ 'ashes' is realized as [áfí]. Again, the voiceless consonant blocks the spreading of low tone. As perceptively demonstrated by Stahlke (1971a), Ewe therefore has some rising tones (L±H) which come from an underlying bisyllabic L-H sequence. It also has underlying L±H rising tones on long vowels, and these become M±H after voiceless obstruents and sonorants. Thus, all phonetic L-L±H sequences have an intervening voiced obstruent, though some of these sequences derive from underlying L-H and others from underlying L-L±H.

The question is: why doesn't tone spreading occur when the intervening consonant of an underlying L-H sequence is a sonorant? As shown in (38) above, /áwá/ is realized as [áwá], i.e. M-H with no contour tone. The answer is provided by a generalization which was made in section 1.2.1 above. Namely, tone spreading always occurs first in sequences where the interval between the two tones is greatest. Since low-raising applies to a L-H sequence whose intervening consonant is a sonorant, the interval between the two tones is reduced. Thus, in Standard Ewe tone spreading applies only to a phonetic (i.e. post low-raising) L-H sequence. It has not yet reached phonetic M-H sequences.

That this is the correct explanation for the failure of /áwá/ to develop a contour in the second syllable is seen from the facts
of tone spreading in the Adangbe dialect of Ewe (Sprigge 1967). As seen in the comparison in (39),

(39) Standard Ewe  Adangbe Ewe

/ápá/ → [ápá]  /ápá/ → [ápá]
/ábá/ → [ábá]  /ábá/ → [ábá]
/áwá/ → [áwá]  /áwá/ → [áwá]

the Adangbe dialect differs from Standard Ewe in that low-raising is not operative in the case of L-H with an intervening sonorant. It is significant that when the L of underlying /áwá/ is not raised to M, tone spreading occurs exactly as in underlying /ábá/. This is precisely because the phonetic intervals are the same—and because both voiced obstruents and sonorants are permeable to L-spreading, as argued above.

3. CONCLUSION

From this cross-linguistic investigation of consonant types and tone rules, one clear generalization emerges: consonants affect tone, but tone does not affect consonants. Thus, I have given numerous examples of where consonants interfere with natural tonal assimilations. There are no examples I know of where tones interfere with natural consonantal assimilations. Thus, Fromkin (1972) points out that there is no language which restricts intervocalic voicing as in rule (40):

(40) $p \rightarrow b / L \_ L$

Although we have seen the low toneness of voiced obstruents (e.g. [b]), there seems to be no greater tendency to voice consonants between low tone vowels than between any other vowels.  

Kim (1971) points out a possible counter-example. Stating that "...a low tone seems to induce voicing, a high tone, voicelessness, and vice versa..." (p. 93), he cites voicing in Kpelle. As seen in the examples in (41),

(41) [pólu] 'back'  [bólu] 'his/her back'

the morpheme 'his/her' is sometimes realized by voicing a voiceless obstruent and by lowering the pitch of the following vowel
(here indicated as a change from H to M tone). Welmers (1962) analyzes the morpheme 'his/her' (when used with "relational" nouns) as a "floating" low tone prefix, and forms such as 'his/her back' as /'pölù/. The low tone is said to both voice the /p/ to a heavily voiced [b] and condition a lower allotone on the following vowel. Historically, however, as argued by a number of Africanists (see in particular, Bird 1971 and Hyman 1973a), it is necessary to reconstruct a low tone nasal prefix which constitutes the sole motivation for voicing the /p/ in 'his/her back'. In (42) it is seen that this same morpheme nasalizes a stem-initial non-nasal sonorant:

(42) [lēe] 'mother'  [nēe] 'his/her mother'

In addition, the morpheme 'my' is always realized as a high tone nasal prefix, and it has the effect ofvoicing /p/ to [b]:

(43) [pölù] 'back'  [nōlù] 'my back'

Thus, an internal reconstruction reveals that it was the nasality and voicing of the historical prefix in *N+pölù 'his/her back' which voiced the *p to [b], and not the low tone.

NOTES

1Much of the material presented in this paper was first discussed in Hyman and Schuh (1972). I am extremely indebted to my friend Russell Schuh for providing much of the information upon which this paper is based, and for discussing and developing these ideas with me. I would also like to thank Will Leben for his detailed comments on the original Hyman and Schuh paper and John Ohala for enlightening me on many of the phonetic issues involved in the study of consonant types and tone.

2What I shall have to say does characterize languages which map "melodies" (Leben 1973) onto grammatical units (e.g. the morpheme or word). Such languages including Kpelle (Welmers 1962) and Mende (Dwyer 1971, Leben 1971) require that a contour, e.g. H-L, be assigned independently of the number of syllables involved in the grammatical unit. These languages derive historically from the kind of syllable-tone languages I shall draw my data from in this study—precisely by application of the spreading and absorption rules to be discussed.

3This does not mean that the physiologically-based "intrinsic" effects of consonants, particularly on following vowels, are not felt. Quite to the contrary, the tendency of certain consonant types to raise or lower the fundamental frequency of a following
vowel has been shown for English (Lehiste 1970, Mohr 1971, Lea 1973), a stress language, and should be expected in all languages, stress or tonal. However, the claim I am making is that this information is not available for "phonologization" in stress languages, e.g. as a conditioning environment for stress placement rules. A syllable can reject a stress because of its inherent "weakness" or "lightness", but not because of its initial consonant. As pointed out by Newman (1972), the weight of a syllable is determined wholly by what follows the initial consonant, e.g. V, VC, VV, etc.

"The tone markings used in this study are as follows:

(') H(igh)   ('') L(h) (rising)
( ) M(id)   (') H(L)(falling)
(>) Lower-Mid
( )) L(ow)
( '"') Downstep

Tones which are assigned to different syllables are separated by -, e.g. L-L. Tones assigned to the same syllable (contours) are not separated by -, and are redundantly tied by a ligature, e.g. Lh.

5 Or, conceivably, if both occurred, producing [ôkpâ], i.e. M-M, a phenomenon termed "total downstep" by Meeussen (1970). That is, the L has been raised to the same level that the H has been lowered to. The notion of horizontal assimilation is indispensable for explaining some of the "tonal reversals" or "flip-flops" that have occurred in Bantu (cf. van Spaandonk 1971). Imagine a word which has a H-L prefix (two syllables) and a H-L stem, i.e. H-L=H-L. If each tone is to spread one to the right we obtain H-H=L=L (with the final L falling off at the end). Already the stem tones H-L have been reversed to L-H. If the initial H should, by virtue of its weak position, become L, a complete reversal is accomplished.

6 This notion of absorption is comparable to segmental cases of absorption such as the following rule which characterizes Kpelle (Welmers 1962):

\[ \eta^w \rightarrow \eta / u, o \]

The labialized velar [\eta^w] loses its rounding before the rounded vowels /u/ and /o/. Rather than viewing this as a "deletion" of a /w/, it is correct to view this as the absorption of rounding into a rounded V. Similarly, one speaks of the absorption of the second component of a contour into a following like tone, and not of the deletion of this component (cf. Hyman 1973b for further examples of absorption).

7 In fact, as reported by Kenneth Pike during this symposium, a L will, if anything, make a preceding H more prominent. An example supplied to me by William E. Welmers (personal communication) comes from Xhosa, where the final H of a H-H-H-L sequence will be realized higher in pitch than the two preceding H's.

8 The only counter-example I know of to this generalization is reported by Kotey (1973). Certain grammatical morphemes in Gâ
(which are perhaps encliticized or suffixed) are raised from L to a downstep (or M) after a H. This is not, however, a general process in the language, but is rather restricted to certain grammatical environments.

The tonological vulnerability of L-H (as opposed to H-L) is probably due to the observation of Ohala (1973) that it takes greater time to go up a certain pitch interval than to descend the same interval. Ohala hypothesizes that it therefore takes greater effort as well.

While I am not quite sure why this should be so physiologically, what this means is that it must take greater time to make the necessary pitch-regulating adjustments than to make the different segmental adjustments.

It is important to emphasize that I am speaking here only of purely phonetic assimilations—that is, where the only motivation for the assimilation is the juxtaposition of two differing tones. Where other factors are involved, there is sometimes a leftward (or anticipatory) shift of tone. Thus, when the final /ú/ of Ngizim /váwú/ 'he shot (it)' is apocopated, the high tone is assigned to the previous (remaining) syllable, i.e. [váw] (Hyman and Schuh 1972). This is not a counter-example, since the loss of a final syllable triggers the shift. The assignment of the H to the preceding L syllable (to yield a rising tone) is therefore not a purely phonetic assimilation. On the other hand, what would constitute a counter-example to the claim that tones spread perseveratively is if /váwú/ were to become [váwú], i.e. rising-high. While historical linguists sometimes posit such logical intermediate forms, to my knowledge no one has ever turned up such a counter-example to my claim.

It does allow rising tones to result from the "compression" of two syllables—cf. note 11.

This view has been expressed by Harms (1973) and privately to me by others not generally familiar with the effects of consonants on tone—it is therefore worth mentioning here.

The use of the terms "tonpermeable" and "tonimpermeable" was introduced by Lukas (1969). Leben (1973) illustrates the comparability of consonants which are permeable or impermeable with respect to nasalization. (See also Schourup (1972) for a general statement regarding the blocking of nasalization spreading by obstruent consonants.

As mentioned in note 3, the phonetic effect of different consonants influencing the fundamental frequency of the following vowel is expected to be present. In Asian languages where an older voicing contrast has given rise to a tonal contrast (Maran 1973, Matisoff 1973), I prefer to speak of the imminent loss of voicing as the motivation for the "phonologization" of the formerly intrinsic tonal effects of the different consonant types. Thus, if I am correct, this too is not a 'pure' phonetic assimilation.
Similarly, the motivation for the exaggeration of vowel lengthening before voiced consonants in English (Chen 1970) is the imminent loss of a voice contrast in final position in English (cf. Hyman, in preparation). It is interesting to repeat Matisoff's (1973) observation that voicing contrasts do not give rise to contour tones, but rather to what he terms "registers", i.e. to higher vs. lower allophones of pre-existent tones. However, as we know from Punjabi (Manjari Ohala, personal communication) and certain Tibeto-Burman languages (Glover 1970), "breathiness" (i.e. breathy voice) does appear to have the ability in itself to create contour tones.

Notice in (35) that the implosive [ɓ] also blocks the spreading of low tone, as it does in Ngizim.

The fourth possibility, consonantal interference in perseverative vertical assimilations, has not, to my knowledge, been demonstrated in any language. Thus, I know of no language where downdrift occurs only if the potentially downshifted syllable begins with a depressor consonant. That is, no language exhibits the following phonetic contrast:

/áwàpá/ → [−−] (no downdrift because /p/ blocks it)

/áwàbá/ → [−−] (downdrift permitted by /b/)

Both Smith (1968) and Stahlke (1971a,b) start with the mid tone as basic and provide rules of lowering (both before obstruents and finally, as will be seen below). This approach interprets the consonants as providing the basic motivation for the tonal alternations, and is therefore incompatible with the conceptualization of tone rules and consonantal interference being presented here.

Voiced obstruents might of course have been expected to have a tone lowering effect. Lea (1973) explains the identical lowering effect of voiceless obstruents by saying that "...the closure of the vocal tract for unvoiced [his emphasis] consonants usually occurs before voicing ceases..." (section 3.1). Thus, a voiceless obstruent in post-vocalic position starts off being voiced, and presumably has a lowering effect as a result.

The African data which I have accumulated suggests the following concerning h-like sounds. First, the voiced [ɓ] in Ewe has the same lowering effect on preceding and following vowels as other voiced fricatives. Thus, a hypothetical /áhá/ would be realized as [áhá], with the /h/ keeping the preceding vowel down, and allowing L-spreading. The voiceless [h] in Nupe, on the other hand, has the same blocking effect on L-spreading as other voiceless fricatives. Thus, /èhá/ 'is hanging' is realized as [èhá] and not as *[èhá]. These data suggest to me that /h/ is not a glide in any sense, but rather a fricative. There does not appear to be any phonetic h-like sound which like the sonorants (glides, liquids, nasals) is neutral with respect to both L- and H-spreading.
The only possible exception to this statement is found in Matisoff's (1973) reinterpretation of Maran (1971). According to Maran, modern Jinghpao has a co-occurrence between L and final voiced obstruents, and H and final voiceless obstruents. He suggests that the tone can be predicted on the basis of the final voice contrast. This of course violates the general principle that the voicing contrast is tonally relevant only on following vowels, and not on preceding vowels. Matisoff, on the other hand, claims that the voicing of the final consonants can be predicted from the tones. This interpretation, if correct, constitutes evidence that tones can affect the voicing of consonants.

APPENDIX

Identification of Languages Surveyed

<table>
<thead>
<tr>
<th>Language</th>
<th>Linguistic Classification</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bamileke</td>
<td>Benue-Congo (Bantu)</td>
<td>Cameroon</td>
</tr>
<tr>
<td>Ewe</td>
<td>Kwa</td>
<td>Ghana, Togo</td>
</tr>
<tr>
<td>Gâ</td>
<td>Kwa</td>
<td>Ghana</td>
</tr>
<tr>
<td>Gwari</td>
<td>Kwa</td>
<td>Nigeria</td>
</tr>
<tr>
<td>Hausa</td>
<td>Chadic</td>
<td>Nigeria, Niger</td>
</tr>
<tr>
<td>Igbo</td>
<td>Kwa</td>
<td>Nigeria</td>
</tr>
<tr>
<td>Kikuyu</td>
<td>Benue-Congo (Bantu)</td>
<td>Kenya</td>
</tr>
<tr>
<td>Kinga</td>
<td>Benue-Congo (Bantu)</td>
<td>Tanzania</td>
</tr>
<tr>
<td>Kpelle</td>
<td>Mande</td>
<td>Liberia</td>
</tr>
<tr>
<td>Maninka</td>
<td>Mande</td>
<td>Senegal, Mali+</td>
</tr>
<tr>
<td>Mende</td>
<td>Mande</td>
<td>Sierra Leone</td>
</tr>
<tr>
<td>Nupe</td>
<td>Kwa</td>
<td>Nigeria</td>
</tr>
<tr>
<td>Ngizim</td>
<td>Chadic</td>
<td>Nigeria</td>
</tr>
<tr>
<td>Safwa</td>
<td>Benue-Congo (Bantu)</td>
<td>Tanzania</td>
</tr>
<tr>
<td>Swazi</td>
<td>Benue-Congo (Bantu)</td>
<td>Swaziland</td>
</tr>
<tr>
<td>Xhosa</td>
<td>Benue-Congo (Bantu)</td>
<td>South Africa</td>
</tr>
<tr>
<td>Yoruba</td>
<td>Kwa</td>
<td>Nigeria, Dahomey</td>
</tr>
<tr>
<td>Zulu</td>
<td>Benue-Congo (Bantu)</td>
<td>South Africa</td>
</tr>
</tbody>
</table>
REFERENCES

Beach, D. M. 1923. The science of tonetics and its application to Bantu languages. Bantu Studies, Vol. II.


Hyman, Larry M. In preparation. Synchronic and diachronic naturalness.


can Linguistics, U.C.L.A. Studies in African Linguistics, 
Supplement 2. 183-200.
Smith, Neilson V. 1968. Tone in Ewe. Quarterly Progress Report No. 88. 293-305. Research Laboratory of Electronics, M.I.T.
