THE MANY USES OF $F_0$. *

Most phonetic parameters that we posit for the analysis of language are elusive to measurement because of the variability of their physical realization in different contexts. The pitch of the voice, however, has a rather straightforward physical correlate, namely the fundamental frequency of the glottal tone, or $F_0$. The challenge to the phonologist here is not so much how $F_0$ can be extracted from the speech wave and displayed, but how the many functions that $F_0$ performs in language can be systematically extracted from and related to each other. To meet this challenge we must establish an explicit relation between the abstract impressionistic units of linguistic description and the physical units of the laboratory.

The simple time-function of $F_0$ is the result of the interaction of two complex sets of factors: (1) instructions that realize certain types of linguistic formative contained in the surface phrase-marker of the utterance, and (2) operating characteristics (hereafter OC) of the speech mechanism. The set of linguistic instructions carries cognitive content in ways that vary according to the particular language which is involved. The formatives which determine these instructions are discrete and are usually organized hierarchically. The OC, on the other hand, are largely independent of the language in question, though they may vary from individual to individual, depending on such factors as age, sex, mood, and speech rate. There is a more gradient element with many OC that makes them less determinate and more subject to free variation than the linguistic instructions. The common core of OC, which do not vary significantly according to speakers (since they reflect the physiological and acoustical constraints on the speech mechanism in general), have been called “intrinsic”, (Wang and Fillmore, 1961).

When a man speaks, he is only concerned with the linguistic instructions: the OC are of course built in. But when we program a synthesizer to speak, we must be concerned with the OC in addition. When the synthesizer is speaking English, Japanese or Hawaiian, it must be remembered that it is speaking Homo sapiens as well. This

---

* This is a revised version of a paper presented at the 1968 Kyoto Speech Communication Symposium. This research is supported in part by National Science Foundation Grant GB-1430 and by the Pacific and Asian Linguistics Institute of the University of Hawaii.
obvious point has two aspects to it. When we program in the proper OC we are
rewarded with a more natural or more 'human sounding' output from the synthesizer.
Conversely, when we effectively factor out these OC from the speech wave we are in
a better position to understand the nature of the instructions to the speech mechanism
which are determined by the linguistic structure that underlies the utterance.

The above-mentioned considerations are already incorporated in the synthesis
strategy of many workers in the field, though with varying degrees of explicitness.
A topic that has not been in the literature as much, however, is the role that intrinsic
OC play in shaping particular phonological codes, especially from a diachronic view-
point. This topic promises to be an important and fertile area of research in which
the tools of the phonetician are crucial in pushing us beyond merely describing
phonetic systems into the explanatory realms of why sounds pattern the way they do
in language. Efforts to provide a phonetic basis for sound change (and therefore,
indirectly, for the genesis of phonetic systems) have a long tradition in modern
linguistics, cf. Whitney, 1887. It is only within recent years, however, that conceptual
and physical tools have been developed which promise to be equal to the task.

The linguistic instructions are contained in a set of surface structure phrase-markers,
that may be thought of as hierarchical arrangements of selected formatives, each of
which is, directly or indirectly, associated with particular instructions to the speech
mechanism. The formatives pertinent for $F_s$ are accents at three levels, depending on
the nature of the objects they mark: syntactic accent (or intonation), morphological
accent, and lexical accent. These accentual formatives, through the application of
phonological rules, determine the exact shape of $F_s$ in conjunction with the OC of
the speech mechanism.

Fig. 1
Examples of Cantonese tones

1 For a wide ranging survey of diverse intonation systems as well as some perceptive remarks regarding
their universality, see Bolinger, 1964.
accent of noun derivation has the effect of changing the rising \( F_r \) of the verb to the falling \( F_f \) of the noun.

Even though it has a very different phonological structure from English, Chinese also has an accentual formative of noun derivation. It derives nouns from other grammatical categories by changing the tone of the morpheme in question, whichever tone it is, into the high falling tone. Thus we have ʂən 'fan' from ʂən 'to fan', ɬɪm 'chain' from ɬɪm 'to connect', ʂə 'number' from ʂə 'to count', and many others.

The English and Chinese forms above illustrate derivational accents. Many languages can be cited where accentual formatives function inflectionally, i.e., to signal such information as tense, number, gender, aspect, etc. In many native languages of America and Africa, elaborate paradigms of words are distinguished from each other just by extremely fine differences in \( F_r \). For instance in Nkuba, a language of the Belgian Congo, the forms ʂəʔ, ɬɪ and ɬɪŋ make up a paradigm of the verb meaning 'to put', but in four different tenses. Researchers suggestively refer to these morphological types as being 'vertically' patterned, to distinguish them from the more familiar 'horizontal' patterns of languages like German, Russian or Japanese, where the affixes are strung out along the dimension of time.

Morphological and syntactic accents occur in the phrase markers as independent formatives. Lexical accents, on the other hand, are part of the phonological make-up of individual morphemes. Instead of merely giving prominence to \( F_r \) at given positions, in the manner of syntactic accents, lexical accents may be quite complex in their composition of phonological features. A sound system with lexical accents may be classified by these three questions: (1) Does it have noncontour features? (2) Does it have external accents? (3) Is the accent recursive?

In a recent survey of several hundred tone systems of Asia by the author, it was found that very few languages have more than (a) four noncontour tones, (b) two rising tones, (c) two falling tones, or (d) two convex tones. (Wang, 1968:15). External accent is usually manifested on noncontour features. When the external accent is recursive, the pitch contours assume a remarkable step-function appearance that is visibly different from pitch contours of English or Chinese. The appended bibliography of studies on tone languages should give some idea of the scope of the scholarship in this area.

There is a time-honored tradition in linguistics for accents to be divided into pitch and stress accents. Some 100 years ago, for example, in formulating the famous sound law that bears his name, Karl Verner (1875) argued that the change from a chromatic accent (i.e., pitch accent or tonal accent) into an expiratory accent (i.e., stress accent) was an important phonetic factor that supported his explanation of intervocalic voicing of obstruents.

Thus it is said that Japanese has pitch accent, while English has stress accent, mean-

---

Footnotes:
1 Data taken from Nida 1949. The tone letters in these examples mark high, mid, low and dipping pitch contours respectively.
2 For discussion of these features and their inter-relations see Wang, 1967.
illustrates the two variants of the same tone category: the contour 53 (A2) changes to 55 (A1) in certain conditions. Notice that in the lower spectrogram of Figure 1, the tone on /jən/, that is usually transcribed phonetically as 23, remains relatively invariant in the two positions.

The tenth harmonics of the major pitch contours of Cantonese tones are shown in Figure 2, together with their usual phonetic transcriptions. Each A1 contour is shown directly above its corresponding A2 contour. Thus the first 53 in row 1 corresponds to the first 53 in row 2; the latter is the same morpheme occurring in utterance-final position. The 55 in row 1 and the 53 below it are traced from the upper spectrogram of Figure 1. These four contours represent a single historical category of tone, the YIN-PING, that appears to be undergoing a phonetic split.

The remainder of Figure 2 is to be interpreted the same way. There are altogether eleven morphemes illustrated in eleven frames, as shown in Table 1 below. As noted above, the morpheme in item b is /te'ən/. The six morphemes in a and c-g are all pronounced /jən/, differing only in \( F_0 \). The utterance-final pronunciation is considerably longer in duration for these morphemes. The morpheme in item h is /jət/ and has a relatively long vowel. The remaining morphemes, in i-k, have short vowels and are all pronounced /sk/. In Cantonese we have an instance where \( F_0 \) is carrying an unusually heavy functional load in the phonological system, especially at the lexical level.

By external accent I mean that the effect of the accent is realized elsewhere than the syllable marked by the accent; such as in Japanese. The words /hanə/ ‘nose’ and /hanə/ ‘flower’, for example, are completely homophonous in Tokyo speech when said in isolation, even though one of them has a lexical accent of down on the second syllable. DOWN is an instruction to lower the \( F_0 \) of the following syllable, and it can be carried out, of course, only when there is a following syllable. When we add a copula, say, da, then the two utterances become obviously distinct.

Since /hanə/ and /hanə/ are written with different characters in Japanese and have different morphophonemic representations, native speakers sometimes need to be convinced that they are indeed homophonous in utterance-final position. When the utterances /konə hanə da/ ‘it is this nose’ and /konə hanə da/ ‘it is this flower’ are examined spectrographically, as in Figure 3, the drop in \( F_0 \) between /hanə/ and /da/ is seen to be considerably greater than that between /hanə/ and /da/. Native speakers have no difficulty in distinguishing between these two utterances. On the other hand, when the /da/ is removed by taping, the two utterances cannot be easily identified.

In a Japanese utterance, DOWN cannot further lower the \( F_0 \) that has already been lowered by an earlier DOWN. This is to say, the lexical accent here is not recursive. Down is recursive, however, in some West African languages, where a succession of this accent produces what has been called a “terrace level” effect. (Welmers, 1949)

Several years ago Unice Pikes was kind enough to present me with some data on a Mexican Indian language in which there is a recursive t.p., where successive occurrence-

---

### TABLE 1

<table>
<thead>
<tr>
<th>row</th>
<th>Cantonese tones</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
</tr>
<tr>
<td>a.</td>
<td>因果</td>
</tr>
<tr>
<td>b.</td>
<td>春天</td>
</tr>
<tr>
<td>c.</td>
<td>隠居</td>
</tr>
<tr>
<td>d.</td>
<td>佛山</td>
</tr>
</tbody>
</table>

For languages such as French, which do not have lexical accents, the questions enumerated above are of course not applicable. The language with the richest inventory of lexical accents that I know of is Cantonese (see Table 1). The \( F_0 \) patterns of its tones are shown in Figure 1 and Figure 2.

The narrow-band spectrograms of Figure 1 and Figure 2 are made from a tape recording of Mrs. Teresa Cheng, who is a native speaker of Hong Kong Cantonese. Each utterance is of the form ‘\( A_1 X \) ke \( A_2 \)’, where \( A_1 \) and \( A_2 \) are instances of the same morpheme in utterance-initial and utterance-final positions respectively. Thus the upper spectrogram in Figure 1 illustrates the utterance /te'ən t'in ke te'ən/, which means “spring as in ‘springtime’”. This utterance is of special interest because it

---

4 This is not to say that it is impossible to vary \( F_0 \) by either mechanism above. Thus people who speak from tank respirators or with an artificial larynx vary \( F_0 \), largely independently of subglottic activity. In Mandarin, the continuous amplitude contours of the four tones are very similar whereas the syllables are VOICED or WHISPERED, which suggests that the subglottal pressure plays an important role in producing the \( F_0 \) contours. It is also interesting to note that Mandarin speakers identify weak stresses in English with low tones in Mandarin; tone sandhi rules which are activated by low tones are also activated by weakly stressed English syllables embedded in Chinese sentences; see C. C. Cheng (1968).

4 These spectrograms were prepared by Teresa Cheng. A detailed study of Cantonese phonology can be found in T. Cheng, 1968. There are constraints on the occurrence of these pitch patterns according to the segmental phonemes in the syllable.
ces of this accent take the $F_c$ all the way to the top of the $UP$-staircase. In Figure 4 we see the $F_c$, in a set of sentences from Acatlan Mixtec which were designed to illustrate $UP$. The transcription below is by Kent Wistrand of the Summer Institute of Linguistics, where the numbers increase with pitch height.

a. $t^i \text{mba}^8 \text{le}^4$ ‘small compadre’

b. $ma^1 \text{ti}^3 \text{mba}^4 \text{le}^6$ ‘not small compadre’

c. $ma^1 \text{ti}^3 \text{yu}^4 \text{a}^4 \text{mba}^3 \text{le}^8$ ‘not small father compadre’

d. $ma^1 \text{ti}^3 \text{yu}^4 \text{a}^4 \text{di}^2 \text{y}^2 \text{mba}^8 \text{le}^4$ ‘not small father mother compadre’

In the bottom sentence the informant has approached something of a screech at the end, even though he anticipated the pitch rise by starting very low. This is not to say, of course, that he has any explicit understanding of the phonological structure of these sentences — native speakers typically do not have the vaguest idea of the physical basis of the prosodic phenomena in their language.

Languages with recursive accents seem to have a strikingly unique appearance on narrow-band spectrograms. Whereas so-called tone languages like Cantonese, Japanese, and Swedish look very much like the average European languages in $F_c$, these languages with recursive accents are distinctive in the extent to which $F_c$ remains largely parallel to the base line, as can be seen in Figure 4.

Regardless of the accentual system a language has, or does not have, $F_c$ is consistently influenced by intrinsic factors, all of which have either an acoustic or physiological motivation. The point of interest here is that sometimes a small effect conditioned by intrinsic factors has grown in time into significant differences that participate in the morphophonemic alternations in the language. For example, high vowels are known to raise the $F_c$ by a small increment. Since the raising of the tongue body for their articulation indirectly pulls on the thyroid cartilage, this raising contributes to the tensing of the vocal folds. In a language like Foochow Chinese, this relation between $F_c$ and vowel height has grown into morphophonemic alternations.

In Foochow, as in almost all Chinese dialects, tone sandhi is typically regressive. This means that in a disyllabic sequence $S_1 \text{ $s_2$}$, the tone of $S_2$ stays constant whereas the tone of $S_1$ changes. Sometimes the nature of the change is determined by the tone on $S_2$, as in Foochow; sometimes it is not, as in Amoy. The special interest that Foochow tone sandhi has for us is that it is accompanied by changes in vowel height. This can be seen in the following examples from Yuan (1960: 295-97).

a. $t^i \text{yu}^4 \text{but $t^i \text{yu}^4 \text{u}^4$}

b. $k^i \text{ei}^4 \text{but $k^i \text{ei}^4 \text{a}^4$}

These five examples illustrate the vowel alternations

$c \rightarrow o \rightarrow y$

$\text{a} \rightarrow \text{e} \rightarrow \text{i}$

Each time a tone with a lower $F_c$ changes to one with a higher $F_c$, the vowel also changes to a higher vowel. Since the tone sandhi is a more general phonological rule (i.e., there are vowel qualities such as [a] which are not affected by the sandhi), it is descriptively more economical to let the sandhi environment condition the vowel raising rule. From the viewpoint of the physiology of speech production, however, it is perhaps more likely that the vowel raising brought about the tone sandhi, which was then generalized even to those vowels that do not get raised.

Another respect in which small intrinsic effects may assume phonemic status has to do with various phenomena of glottalization. These phenomena frequently co-occur with tones which are either at the extremes of the pitch range of the speaking voice or which are very rapidly changing in value, such as converge, convex or very short tones. What may have started sporadically as weak breaks in the voice when the vocal folds make adjustments to implement these tones, frequently stabilize into a consistent aspect of the tones, and may remain with these tones even though the tones themselves change their $F_c$ contours.

In CV syllables the manner of the articulation of C can be seen to exert a strong influence on the $F_c$ contour. The situation found in Yabem of New Guinea is fairly typical, where “voiced plosives are followed by a low-toned vowel, and voiceless plosives by a high-toned vowel.” Such correlations have been reported again and again in the literature for languages ranging from southern Africa to China and native America. This is of course not surprising since they are motivated by the intrinsic OC of the speech mechanism. Yabem is of somewhat additional interest in this respect because “consonants other than plosives are neutral, i.e., they may be followed by either tone, and no rule can be given for them.”

In general, voiced consonants tend to lower $F_c$, even when they do not involve full oral constriction. Other things being equal, this is what we may expect, because any oral constriction will impede the air flowing out of the mouth and reduce the pressure differential across the glottis. Since this pressure differential is the major mechanism whereby the subglottal air separates the vocal folds in the vibratory cycle, a reduction in its magnitude naturally causes a lowering in the $F_c$. When the vocal tract moves into a vowel configuration, the $F_c$ typically rises gradually as the pressure differential increases with the removal of the constriction.

7 Capell, 1949: 188.
The interaction between the laryngeal and supralaryngeal activities can be seen to be very complex when we examine X-ray moving pictures of the larynx during speech. The thyroid cartilage moves in intricate ways both vertically and in the anterior-posterior dimension, changing its geometric relation to the cricoid cartilage. A controlled study of the effects that these external movements have on \( F_v \) may be quite instructive.

Recently, Halle and Stevens have considered the possibility that there are "rather drastic adjustments in vocal-cord positioning and in the manner in which the vocal cords vibrate when voicing is to be maintained during certain consonants" (1967:269). Assuming that "the rate at which vocal-cord positioning can be achieved is relatively slow", and that the wide separation for unvoiced consonants "can be achieved more rapidly than the more finely adjusted smaller separation for a voiced consonant", the authors went on to explain the greater intrinsic duration of vowels before voiced consonants in terms of this difference in the rate of laryngeal adjustment. These vowels must be longer, they argue, to allow the larynx to adjust into a suitable configuration before a voiced consonant. As partial support for their explanation, they note that a vowel is shorter before \([m]\) than before \([b]\), presumably because, in the former, no special adjustment is required to sustain vocal fold vibrations.

The degree to which laryngeal adjustments are actually made can be determined eventually as techniques of electromyographic recordings of laryngeal activity get refined (Ohala 1970). There are two observations which may be mentioned in connection with the explanation advanced above. One is that when we compare vowel durations in \((a)CV+\), \((b)CV+\) voiced consonant and \((c)CV+\) unvoiced consonant, we find that the most conspicuous fact is that the vowel in \((c)\) is shorter than the other two. If \((b)\) requires special adjustments, then we should expect it to be longer than the other two. But actually, its duration is quite comparable to \((a)\). The other observation has to do with sequences of \(vowel+\) glide, \(vowel+\) liquid and \(vowel+\) nasal. Preliminary measurements show that the second member of such sequences shorten approximately proportionally before unvoiced consonants. If the durational difference were merely due to the adjustment time required by the following consonant, we might expect that the second member of such sequences would be influenced much more. Clearly, for a language to be able to signal those small differences in \( F_v \) exemplified by tone systems like Cantonese, we must be able to achieve very delicate control over the laryngeal musculature. The exact nature of this control awaits detailed investigation.

---

8 I am indebted to Peter MacNeilage who made available X-ray films which show these movements and also for discussing their significance with me.
9 This fact was noticed by Matthew Chen, who has made a systematic investigation of this phenomenon across several languages; c.f. (Chen, 1970).
10 "We may further observe that vowel duration is affected in the same way by the 'voicing' of the following consonant, even in whispered speech (Scharf, 1964). Since presumably voicing is not realized in the whispered consonant, it may be argued that the duration differential cannot be due to laryngeal adjustments. The situation is more complicated, however. We know that many primary phonetic features are habitually accomplished by secondary features during normal speech: as the primary consonantal feature here (be it [voiced], [tenseness], or whatever) is accomplished by the secondary vowel feature of [long], and as the primary tone feature is accomplished by the secondary amplitude feature (compare footnote 4). As we change from the normal mode into some other mode of production, e.g. whispering, some primary features are lost. The secondary features remain, even when their causes are no longer present, and replace the primary features as the major distinguishers. The situation is quite parallel to many historical instances of phonological change where a primary feature becomes lost and a secondary feature takes up its role."
SELECTED BIBLIOGRAPHY OF TONE STUDIES

Abbreviations:
ANPE: Archives Néerlandaises de Phonétique Experimentale
BEFEQ: Bulletin de l'école française d'Extrême-orient, Saigon
BIPA: Bulletin of the Institute of History and Philosophy
BSOAS: Bulletin of the School of Oriental and African Studies
DL: Dai Lu Zha Zhi
FUYPTJE: Fang Yun Yin Pa TING Hua Ji Kin
IJAAL: International Journal of American Linguistics
IAOS: Journal of the American Oriental Society of America
IAASA: Journal of the Acoustical Society of America
Lg: Language
LSA: Linguistic Society of America
MP: La Maitre Phonistique
POLA: Project on Linguistic Analysis
UCLP: University of California Publications in Linguistics
VTY: Yu Yin Yen Sia
UJYW: Zhong Guo Yu Wen
1938 The Phonetics of the Hottentot Language (Cambridge, England, W. Heffer & Sons Ltd.)
Benedict, Paul Kings, 1948 “Tonal Systems in Southeast Asia”, JAOS 68. 184-91 (July-Sept.).
Chang, T., 1953 “On the Tone System of the Miao-Yao Languages”, Lg. 29: 3. 374-78.
1956 “A Comparative Study of the Yao Tone System”, Lg. 42: 2. 303-10.
1928 Studies in the Modern Wu Dialects (Peking).
1933 “Tone and Intonation in Chinese”, BIPA 4: 2. 121-34.
Cowen, J Milton, 1936 “Pitch and Intensity Characteristics of Stage Speech”, Supplement to Archives of Speech 1.
De Angelus, Jaime, 1926 “Tone Patterns and Verb Forms in a Dialect of Zapotec”, Lg 2: 238-50.
1931 "A Comparative Study in Shona Phonetics" (Johannesburg: The University of Witwatersrand Press).
Egeland, Soren, 1956 "The Langu Dialect" (Copenhagen, E. Munksgaard).
Fairbanks, Grant, 1940 "Recent Experimental Investigations of Vocal Pitch in speech. JASA 11: 456-66.
Fü, Mao-Chhi, 1940 "A Study of the Moso Language (We-hei Dialect)", Part II, grammar (University of Western China).
Gedney, W., 1968 "Future Directions in Comparative Tai Linguistics", Lecture and notes delivered at University of California (Berkeley).
Guthrie, Malcolm, 1940-42 "Tone Ranges in a Two-Tone Language (Lingalal), BSOAS 10: 469-78.
Harris, Zelg S., 1945 "Navaho Phonology and Hojice's Analysis", IAAL 11 (Oct): 239-46.
THE MANY USES OF “P,”

Mark, Cornelia,
Malinberg, Bertil,
Malingqui, Goeran,
Meeussen, A.,
Merrifield, William R.,
Meyer, E.A.,
1937, 1954 Die Intonation im Schwedischen, Parts 1, 2 (Stockholm).
Meyer-Eppler, W.,
Migeot, André,
Miller, Charles Boardman,
Öhman, Sven E.G.,
Öhman, Sven E.G. and J. Linqvist,
Ohtsda, M.,
Pankratz, Leo and Eunice V. Pike,
Pike, Eunice,
Pike, Kenneth L.,
1948 “Tone languages (Ann Arbor).”
Pike, Kenneth L. and Eunice V. Pike,
Plattner, Oreste,
Robbins, Frank E.,
Roffe, G. Edward,
Root, Alfred R.,
The Many Uses of “Fe”

Zeng, Fu-Bing, 1964 “Concerning the Problem of Subdividing the Yin-ping Tone of Cantonese”, ZGYW 132 (May): 376-89.

ADDENDA