Ann. Bull. RILP, No. 9, 151-166 (1975)

HOW TO EXPLAIN THE "TONES" IN TIBETAN *

Olle Kjellin

0. It is generally believed that the development of tones in Southeast Asian languages is connected with a restructuring of consonantal systems in such a way that low tones are associated with originally voiced consonants and high tones with voiceless consonants (Haudricourt 1961; Maran 1971, 1973; Matisoff 1973). The same association of voicing and pitch (fundamental frequency) has also been observed in other languages not ordinarily held to be 'tone languages', such as e. g. English (Lehiste and Peterson 1961; Mohr 1971; Lea 1973).

The tones of an idiolect of Lhasa Tibetan (Kjellin 1974) were reanalysed in accordance with this theory and were found to correspond exactly with the given pattern: historically voiceless consonants have a high fundamental frequency (i. e. the following vowel begins high), and historically voiced consonants have a lower fundamental frequency (which rises in the following vowel) although most of these are voiceless in the modern language. It might seem puzzling, however, that not all voiced consonants (obstruents) have become devoiced. And further, the fact that the sonorants have both high and low pitched cognates does not seem to be immediately explained by the voicing-to-pitch hypothesis.

It will be shown, however, that with a slight elaboration of the hypothesis, together with a still closer look at the phonology of Tibetan, all of its pitch (fundamental frequency) contours are found to be the result of synchronic phonetic processes, and thus are predictable.

1. The 'tones' we are going to account for in this paper are given in (1), where capitals represent "high pitch target", and lower-case letters "low pitch target" (fundamental frequencies of about 140 Hz and 120 Hz, respectively, as extracted by digital analysis (Kjellin, 1975)); see also Fig. 1 and Appendix 2.

(1)	MA	MAA	MAa			
	mA	maA	mAa			
	MAMA	MAAMA	MAMAA	MAAMAA	MAMAa	MAAMAa
	maMA	maaMA	maMAA	maaMAA	maMAa	maaMAa

Most of the words in (1) are meaningful; at least, each tonal type is documented with some segmental constituents or other. The "other" consonant types are given in (2), for simplicity represented by the labial series only.

(2) P PH M W b ph m w

* To be presented to the 8th International Conference on Sino-Tibetan Languages and Linguistics, University of California, Berkeley, Oct. 24-26, 1975. An almost identical version appeared in CAAAL (Computational Analyses of Asian & African Languages, Project on Computational Analysis, National Inter-University Research Institute of Asian & African Languages & Cultures, Tokyo) No. 2, 1975.

Frequently in 'phonemic' analysis, these are paired together as tonally determined 'allophones', viz. P/b PH/ph M/m W/w (Simon 1970; Goldstein and Nornang 1970; Yukawa 1971, ch. 5), but as Hyman and Schuh have shown, this is implausible ("Consonants affect tone, but tone does not affect consonants" – Hyman and Schuh 1972, p. 40), and we shall show that in fact all of these 'allophones' have different underlying origins.

1.1. We posit the underlying initial consonants (C_i) as in (3), again represented by the labial series only.

(3) p ph b m w

with no specifications for tone or pitch or fundamental frequency (actually, tone etc. will be irrelevant to the issue until the very last part of this paper!); they are only specified for sonorancy, voicing, aspiration, and place-of-articulation (as in the Chomsky and Halle system). As final consonants (C_f) only /b m/ occur, thus contrasting just sonorancy, not voicing or aspiration.

Syllables will be posited with initial and final clusters:

(4) C_p C_i C_g V C_f C_s

where C_i is the initial consonant, C_p is a pre-initial "clustermaker" ("prefix" in the traditional Tibetological terminology, for sometimes C_p is a morphological marker), V is the syllabic nucleus, C_f is a final, and C_s is a second final ("suffix"). C_g is the glide y or the liquid r, which may enter and thus cause palatalisation or retroflexion, respectively, of C_i. Like C_f, C_p is picked from the voiced series only.

The smallest syllable is C_iV.

The existence of clusters has been noted by several scholars, e.g. P. M. Miller (1951), R. A. Miller (1954), Shefts and Chang (1967), but for some reason the synchronic implications have eluded them, and there is even talk of "allomorphs".

What troubled these investigators is the fact that C_p is never pronounced as such word-initially, but only medially in disyllabics as C_f of the first syllable – if there is space for it, i.e. if the first syllable is open. It is not inappropriate, however, to believe that it is just natural for a "lost" cluster-initial to be revived when the syllabification is altered; cf. English gnostic vs. agnostic, etc.

This process, which will be called 'transmigration' here, necessitates the positing of a word-internal syllable boundary (&) and is formulated as in (P1) (cf. Hooper 1972).

(P1) CV & CCV $\longrightarrow$ CVC & CV

Examples of transmigration are abundant, but here only numerals are given, being the most revealing.

(5)	/btyu/	⟶	ju	'ten'
	/gtyig/	⟶	jig	'one'
	/btyu & gtyig/	⟶	jugjī	'eleven'
	/bzyi/	⟶	shi	'four'
	/btyu & bzyi/	⟶	jubshi	'fourteen'
	/bzyi & btyu/	⟶	shibju	'forty'
	/rgu/	⟶	gu	'nine'
	/btyu & rgu/	⟶	jurgu	'nineteen'
	/rgu & btyu/	⟶	gubju	'ninety'
	/lŋa/	⟶	ŋa	'five'
	/lŋa & btyu/	⟶	ŋabju	'fifty'
	/btyu & lŋa/	⟶	jȫŋa	'fifteen'

(The transcription of pronunciation forms is essentially that of Simon (1970), although it is rather inadequate. This usage is in order not to distract attention from the issue here, viz. to show that cluster consonants are pronounced word internally but lost initially. The last item also shows the effects of two other processes, namely Umlaut in front of a dental final — which is deleted and the umlauted vowel lengthened — and vowel lowering (ü ⟶ ö) before /a/. The fact that we can see the effect of the former in this example is a further piece of evidence for the existence of (P1), 'transmigration'.)

2. The next process to be explained will be called 'tenseness distribution'. But first the feature of tense/nontense must be defined, although our definition is very similar to Chomsky and Halle's.

We regard non-tenseness as a redundancy (originally) concomitant with voicing in obstruents: Voicing is a state of vibration of the vocal folds caused by the airflow through the closed glottis, as in vowels and other sonorants (van den Berg 1958). If the airflow is impeded, as in obstruents, voicing cannot easily be maintained. Therefore, in "voiced obstruents" the supraglottal cavity must be allowed to distend, or it must be expanded actively, all in order to prevent too fast a build-up of the supraglottal pressure, in other words, to make room for a sufficient transglottal airflow.

Such expansion of the supraglottal tract has been observed through experiments using X-ray, and it is probably implemented by complex muscular gestures, such as relaxing the pharyngeal walls, moving the tongue root forward, lowering the larynx, etc. (Perkell 1965, 1969; Kent and Moll 1969; Bell-Berti and Hirose 1972; Lindqvist 1972). A certain amount of air may also be permitted to escape through a "nasal leak" behind the velum (Rothenberg 1968; T. Ushijima, personal comm.). A further means of facilitating voicing at a decreasing transglottal pressure drop is by adjustment of the vocal folds directly, which involves partial relaxing of certain laryngeal muscles, resulting in "soft phonation at low pitch" (Hirano 1974).

All of these gestures originating from the need to maintain voicing in certain consonants will be comprised in the feature specification of [-tense], although (or because!) there is evidence that there are considerable individual variations in the organisation or balance of the various gestures (Bell-Berti and Hirose 1972).

Thus [-tense] will be redundantly assigned to voiced obstruents, in Tibetan as well as in many other languages. Sonorants, including the vowels, and voiceless obstruents will be redundantly [+tense] (cf. Chomsky and Halle 1968). However, absolute word-initial sonorants in Tibetan function in the same way as initial obstruents; this phenomenon will be discussed in more detail below.

In (P2), the collection of processes is formulated; sample derivations will be found in (6).

(P2)

(a)	[ ]	⟶	[- sonorant]	/ # ___
(b)	[-sonorant]	⟶	[α tense]	/ [- $\overline{\alpha}$ voiced]
(c)	[+sonorant]	⟶	[+ tense]	

(6)	'butter'	'wound'	'down'	'one'	'five'	'nine'
[±tense]	/mar/	/rma/	/mag/	/gtyig/	/lŋa/	/rgu/
after (P2b)	/- /	/- /	/- -/	/-+ -/	/- /	/-- /
- " - (c)	/- ++/	/- ++/	/- +-/	/-++-/	/-++/	/--+/
output:	/mAR/	/rMA/	/mAg/	/gTYIg/	/lŊA/	/rgU/

(where upper-case = [+tense], and lower-case = [- tense]).

The (P2)-distribution of tenseness is in good agreement with the general word phonology of Tibetan where, as was mentioned above, voicing is contrastive word-initially only, while sonorancy is contrastive in other positions. Thus the initial 'sonorants', being voiced, become nontense whereas all other sonorants become tense.

Incidentally, it may be this same phenomenon that explains why there a are no word-initial vowels in Tibetan: These are instead preceded by laryngeal 'obstruentalised' glides which, moreover, are of three kinds, viz. /ʔ h ɦ/, exactly matching the triplets of 'ordinary' obstruents, /p ph b/ (that is, voiceless unaspirated, voiceless aspirated, and voiced. We take the liberty of considering these glides to be $\begin{bmatrix}\text{-sonorant}\\ \text{-obstruent}\end{bmatrix}$, as they do not fit the plus-value definition of either of these features).

2. 1. The fact that absolute-initial sonorants seem to be closely related to obstruents also in other languages was noted in Woo (1969), and — at least as far as glides are concerned — is a cornerstone in Hooper (1972).

That noninitial sonorants are not like obstruents, of course, was also noted in Woo, and this may be further evidenced from the presumably unrelated language Japanese. While nasals and other consonants have the same initial distribution in Japanese, an (undifferentiated) nasal is the only consonant which can occur in final position. Therefore, loanwords with a final obstruent are reanalysed and augmented by one vowel so that such an obstruent instead becomes syllable initial, e.g. English /tayp/ > Japanese/taipu/ 'type', while final nasals (at least non-labial nasals) remain final to yield long ("two-mora") syllables paralleling those with geminate vowels, e.g. English /pen/> Japanese /peN/ 'pen'.

From such considerations we feel justified in believing that the environments in (P2) are perfectly natural; and that the processes themselves are natural is due to the fact that they are dictated by the characteristics of the human vocal tract plus aerodynamic requirements.

3. The 'second final' (C_s) has a very restricted optional occurrence, viz. only after the nasal finals /m n ŋ/. Since all clusters will be simplified and since C_s cannot transmigrate (there will be no space for it, all syllables beginning with a consonant), this C_s will never appear on the surface. Therefore it will not be marked for place of articulation, etc., but only as [-sonorant, +consonantal, -vocalic]. However, it cannot disappear without a trace so its derived nontenseness is mapped onto (or assimilated into) the preceding nasal, as shown in (P3).

(P3)

$$[\text{-vocalic}] \longrightarrow [\text{-tense}] \;/\; \_\_\_\_ [\text{-tense}] \left\{ \begin{matrix} \& \\ \# \end{matrix} \right\}$$

Alternatively, it is very possible that it is the nonsonorancy of C_s that is regressively assimilated; in that case this happens before the tenseness distribution process. Whatever the proper solution is, the example in (7) shows how the process operates (the C_s is symbolised with /z/; Tibetan orthography has -s or -d, probably reflecting historical allophones of /z/).

(7)

/khamz/ 'Kham district' $\xrightarrow{(P2)}$ KHÀMz $\xrightarrow{(P3)}$ KHAmz

Obviously this process resembles those of the so-called floating tones reported from several tonal languages of Africa and America. As Eunice Pike puts it, "The vowels may be lost, the consonants may be lost, but the tone is usually retained" (Pike 1974). Note, however that in our Tibetan case it is not tone but nontenseness (alternatively nonsonorancy) that is "floating". A similar case is found in some dialects of American English, where /panting/ is pronounced with a "nasal flap", as [pæ̃Nɪŋ]. Here we see the nonsonorancy of /t/ being retained in the nasal. As another example, in Swedish the derived form litna is further transformed into lilla in a process where the sonorancy of n is retained in the nonnasal. In the English and Swedish cases as well as in the Tibetan ones the mapping process is regressive. (If we had the formal means, we should perhaps rather write it as a 'simultaneous reciprocal assimilation'.)

With this this argumentation we consider (P3) sufficiently justified and natural.

4. The next process to be stated, 'devoicing of initial obstruents', is language specific but not at all unusual among Southeast Asian languages. It is also observable as a tendency in certain German dialects where children[1] tend to say "kuck mal" instead of "guck mal" 'look!', and Danish is well-known to have devoiced initial obstruents. Such observations suggest that this is a universal innate phonetic process.

In Tibetan the process only applies to single-initial obstruents, and not to those of clusters, i. e. with a 'pre-initial' (C_p); see (P4). (Retroflexed and palatalised consonants, i. e. those with underlying C_g, do not count as clusters; those processes are irrelevant here.)

(P4)

$$\begin{bmatrix} \text{-sonorant} \\ \text{-tense} \end{bmatrix} \longrightarrow [\text{-voiced}] \;/\; \# \underline{\quad\quad} V$$

(Actually, this formulation needs some refinement, for fricatives are always devoiced regardless of environment. Further, it only applies to nonderived obstruents.)

The explanation for (P4) is not self-evident, but it might be speculated that one of the adjustments mentioned above for the maintaining of voicing in obstruents, namely the vocal fold relaxation (Hirano 1974), is exaggerated with the result that the glottis is opened too much, and as a consequence the vibrations stop. (This is why [-tense] was included to the left of the arrow although it is not needed there explicitly as a distinctive feature).

At an intermediate degree of 'nontenseness adjustment' of the vocal folds, the result would be breathiness, which might spread to the following vowel (together with nontenseness as such; cf. (P7), below). True breathy vowels have not been observed in our informant's speech, but Sprigg (1954, 1955), describing another idiolect of Lhasa Tibetan, mentions breathiness, and Gordon (1969) on Sherpa, and Strahm and Maibaum (1971) on Jirel — both relatively closely related to Lhasa Tibetan — also report breathy, or "lax", vowels. They are further reported from other languages in Southeast Asia, e. g. in Haudricourt (1965); cf. Matisoff's "lax-larynx syndrome" (Matisoff 1973).

The relation of nontenseness with breathiness and the devoicing process is mere speculation, but it remains a fact that at the output of (P4), stops sound somewhat aspirated, and what is there to distinguish devoiced breathiness from 'ordinary' aspiration? (There is a difference, and the answer will be given below; Richter (1964) calls it "secondary aspiration"!).

The fact that (P4) does not apply to cluster consonants may possibly be related to the well-known redundancy in English and many other languages, where p is aspirated in pot but not in spot.

5. An important process in the development of many related and unrelated languages in Southeast Asia is that of cluster simplification, which obviously operates synchronically in the Lhasa Tibetan described here (in several, especially Western, dialects it does not operate at all or only more restrictedly; cf. for example Sprigg 1972). The process (P5) simply deletes that consonant which is farthest away from the syllabic nucleus (the vowel):

(P5) $[+\text{segment}] \longrightarrow \emptyset \;/\; * \underline{\quad\quad} CV$

[1] I am not sure about the grown-ups.

where the environment of course must not contain a boundary. The asterisk has the ordinary "mirror image" interpretation, so that the process applies both to initial and to final clusters:

(8)

/rma/ 'wound' $\xrightarrow{(P2,\ P3)}$ rMA $\xrightarrow{(P5)}$ MA

/khamz/ 'Kham' $\longrightarrow$ KHAmz $\longrightarrow$ KHAm

Of course, (P5) also applies to syllables in compounds. (Again, and hereafter also, capitals represent [+tense] segments, and lower-case [-tense]).

6. Somewhere at this stage in the derivations there are a number of rules of no direct relevance to the issue, but since they are necessary for the correct outputs they will be mentioned briefly though informally, collected in (P6):

(P6)

(i) Vowels are nasalised before nasals;
(ii) Vowels are umlauted before dental finals (/d n l/);
(iii) Final dentals are obligatorily converted into 'alpha-vowels', i. e., the same as the preceding vowel, yielding geminate vowels, but with the original tenseness specification retained (e. g. /nad/ 'illness' ⟶ nAd ⟶ nÄd ⟶ nÄa);
(iv) Other final consonants are optionally converted into 'alpha-vowels', labials less frequently than velars, and /r/ almost always (e. g., /mar/ 'butter' ⟶ mAR ⟶ mAA).

7. An interesting process — or perhaps rule — in Tibetan is that which assimilates the tenseness value from the initial segment to the subsequent segment(s) in certain environments, as formalised in (P7). The explanation for this process is not clear, but may at least partly be found in the inertia of the articulators, as with other assimilations. The significance of the environment will be discussed presently.

(P7)

$$[+\text{sonorant}] \longrightarrow [\alpha\,\text{tense}] \;/\; \#[\alpha\,\text{tense}] \_\_\_ \begin{Bmatrix} [+\text{tense}] & \text{(a)} \\ \& & \text{(b)} \end{Bmatrix}$$

In the sample derivations in (9), the effect of (a) feeding (b) is seen in the third example; that and the second are the only ones affected by (P7).

(9)

	'fire'	'butter'	'lamp'	'down'	'wound'
	/me/	/mar/	/mar&me/	/mag/	/rma/
by (P2)	mE	mAR	mAR ME	mAg	rMA
by (P5, P6)	"	mAA	mAA ME	mAa	MA
by (P7a)	"	maA	maA ME	"	"
by (P7b)	"	"	maa ME	"	"
output:	mE	maA	maaME	mAg	MA

8. Now, as far as these simple examples are concerned, those who are acquainted with Tibetan have surely already realised that we can assign 'low pitch' to the nontense segments, and 'high pitch' to the tense segments (the picture is somewhat more complicated with clitics and trisyllabic words). But why is this so?

We believe the following to be the answer: Some of the gestures mentioned above in the definition of [-tense] as being involved in maintaining voicing in phonologically voiced obstruents, are also connected with the production of low fundamental frequency. These include the lowering of the larynx and the adjustment of the vocal folds directly. Thus, although it is impossible (and perhaps unimportant!) to tell which is the cause and which is the effect, we have the factum: each instance of [-tense] intrinsically has a low fundamental frequency associated with it. The sonorant [+tense] instances carry the ordinary utterance intonation, including emphatic intonation, etc., and the obstruent instances of [+tense] do not seem to perturb this.

8.1. Implicit in this claim is also the claim that pitch is dependent on the degree of stress, which, of course, is not at all surprising. Whatever "stress" is, at this level we regard it as having been projected onto each segment within its range. And since we are dealing with 'degrees', the pitch (fundamental-frequency) specifications must be n-nary, rather than binary. With [4] as the highest pitch, (P8) shows how to reinforce tenseness or nontenseness with pitch:

(P8)

$$
\begin{array}{ll}
\text{(a)} & \\
\text{(b)}\ [+\text{tense}] \longrightarrow & \left\{ \begin{array}{l} [4\ \text{pitch}]\ /\ [\overline{\text{emphatic}}] \\ [3\ \text{pitch}]\ /\ [\overline{+\text{stress}}] \\ [2\ \text{pitch}]\ /\ [\overline{-\text{stress}}] \end{array} \right\} \\
\text{(c)} & \\
\text{(d)}\ [-\text{tense}] \longrightarrow & [1\ \text{pitch}]
\end{array}
$$

In principle there is stress on each syllable. Particles, etc., are unstressed, but the two auxiliaries /na/ and /da/, corresponding to English 'if' and 'of course... but', respectively, are lexically (semantically?) [emphatic]. There are no true words with more than two syllables, though there may be reduplicated forms and phrases; see the examples in (12) and (13) below. But first a few simpler examples:

(10)

	'fire'	'butter'	'lamp'	'wound'	'pus'
	/me/	/mar/	/mar&me/	/rma/	/rnag/
by (P2-P7)	/mE	maA	maaME	MA	NAa
by (P8)	1 3	1 1 3	1 1 1 3 3	3 3	3 3 1

	'meat pie'	'curtain'	'Tibet's'
	/mog&mog/	/zyal&yol/	/bod <u>gi</u>/$_{\text{particle}}$
by (P2-P7)	/mooMOo	syääYÖÖ	phÖö gI
by (P8)	1 1 1 3 3 1	1 1 1 3 3 3	1 3 1 1 2

	'Kham' /khamz/	'if... not buy' /ma&nyo&na/emph.	'eight' /bgyad/	'fifty-five' /ŋa&lŋa/
by (P1)				ŋal&ŋa
by (P2-P7)	KHAm	maNYO NA	gyÄä	ŋääŊA
by (P8)	3 3 1	1 1 3 3 4 4	1 3 1	1 1 1 3 3

As can be seen, it is in most cases possible to write a broad phonetic transcription using capital letters for the higher pitch and lower-case letters for the lower pitch. For words with internal obstruents, however, I have not yet obtained the exact data regarding their pitch. As to the other features, in slow and careful pronunciation of compounds the monosyllabic lexical contrasts of the initial obstruent of the second syllable may be optionally retained (which, incidentally, is a further piece of evidence for the "performance", or surface-phonetic, character of these processes) as shown in (10) (without pitch indexes).

(11) /bzyi&btyu/ 'forty' ⟶ syibdyU or syibtyU or syibTYU

In fact, the second syllable is always perceived as beginning on the higher pitch even when the obstruent is voiced. This may be explained by Lea's finding that "The central values of F_0 in a voiced consonant... immediately before the second vowel... were also considerably higher when the second syllable was stressed" (Lea 1973, p. 58).

Probably the intrinsic pitches of the obstruents simply are irrelevant if there are two or more syllables in the word, so in the examples in (12) and (13) only the sonorants are given pitch indexes (and to avoid transcriptional trouble all second-syllable initials are written with capitals, even when denoting voiced obstruents).

(12)

	'everywhere' /ga&sa ga&la/NP	'in India' /rgya&gar la/pt.	'of course I want to go, but...' /ŋdro&ŋdod&da/emph.
by(P1-P7)	khaSA kaLA	gyaGAA lA	droōDÖÖ DA
by (P8)	1 3 1 22	1 3 3 1 2	11 3 3 4

(13)

	'Darjeeling' /rdo&rdye gliŋ/NP	'library' /pe&mdzod khaŋ/NP
by (P1-P7)	dooDYE LIŊ	PENDZÖö KHAŊ
by (P8)	11 3 222	3 3 3 1 2 2

To our knowledge, the pitch contours given in (12) and (13) have been carefully avoided in the literature on Tibetan tones so far, or at least they have not been adequately accounted for. This is only to be expected, as all investigators have assumed Lhasa Tibetan to be a lexical "tone language" with two distinctive pitch levels combined in 4-6 distinctive tones, plus a 'neutral tone'. Of course we do not agree with such a theory.

8. 2. Returning to the problem of consonantal pitch targets, the examples in (14) show that in contrast with polysyllables, the pitch targets of the consonants of monosyllables are highly relevant for the distinction between tense and nontense consonants (cf. also Fig. 1).

(14)

	'wound'	'mother'	'battle'	'down'	'place'	'eat'
	/rma/	/ma/	/dmag/	/mag/	/sa/	/za/
by (P2-P7)	MA	mA	MAa	mAa	SA	sA
by (P8)	3 3	1 3	3 3 1	1 3 1	3 3	13

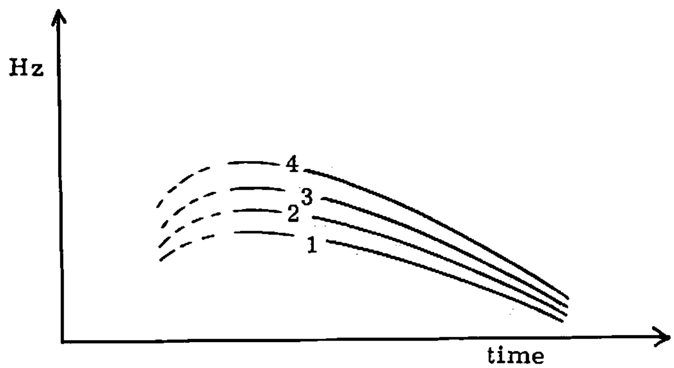
	'accumulate'	'fall'	'end'	'now'
	/sag/	/zag/	/tha/	/da/
by (P2-P7)	SAa	sAa	THA	thA
by (P8)	3 3 1	1 3 1	3 3	1 3

If Tibetan is regarded as a "tone language", the examples in (14) would refute Woo's suggestions that initial sonorants cannot have contrasting pitch and that contour tones cannot be contrastive in short vowels (short syllables) (Woo 1969). However, since the Tibetan described here seems rather to be a nontonal language, Woo's observations may well be correct.

9. The last two examples in (14) provide a clue to the answer to the question above as to what distinguishes aspiration from devoiced breathiness: namely the pitch transition in the following vowel. In whispering, where (P8) is inhibited by a later 'all-devoicing-rule', the tenseness contrasts must be reinforced with something other than fundamental frequency; probably with the 'strength' of the aspiration, or, in the case of the fricatives, with the degree of stridency.

10. The absolute frequencies, we speculate, will be determined by the position of a particular word on the overall utterance-intonation contour, where the pitch indexes 'ride', as it were, on the curves of a frequency scale, as in (15):

(15)



Thus all contours are intrinsically falling even when the specification is e. g. [333]. As if to compensate for this and to increase redundancy, a sequence of [331] at the end of an utterance is converted into [341], as is evident from Fig. 1. Therefore the following is added to (P8):

(P8e) [3 pitch] ⟶ [4 pitch] / [3 pitch] ___ [1 pitch]

By this process /dmag/ 'battle' goes from $\begin{bmatrix} \text{MAa} \\ 3\,3\,1 \end{bmatrix}$ to $\begin{bmatrix} \text{MAa} \\ 3\,4\,1 \end{bmatrix}$, and so it becomes more distinct from $\begin{bmatrix} \text{MAA} \\ 3\,3\,3 \end{bmatrix}$ 'to the wound'.

11. Now the environment of the tenseness assimilation process (P7) may be explained. The process is repeated here for the reader's convenience:

(P7)

$$[+\text{sonorant}] \longrightarrow [\alpha\,\text{tense}] \;/\; \#\,[\alpha\,\text{tense}] \;\_\_\_\; \begin{Bmatrix} [+\text{tense}] \\ \& \end{Bmatrix}$$

An assimilation is not only 'facilitated' in a certain environment, it must also be blocked or limited at some point (otherwise the whole utterance would be only a single sound!) and in usual formalism this is done by the 'lack of environment'. In (P7), however, the limits are explicit, and this suggests that there is a conspiracy going on, a conspiracy to have at least <u>one</u> [+tense] segment in each word. Thus there are only six possible monosyllabic forms in Lhasa Tibetan, as given in (16):

(16)

MA	mA	MAA	maA	MAa	mAa	(*ma	*maa	*mAA)
a	b	c	d	e	f	g	h	i

Looking at it the other way around, one might ask why assimilation is at all possible in (d) <u>maA</u> but not in the other monosyllables. Unless the need for having at least one tense segment per word is accepted as a 'physiological constraint' or perhaps an 'acoustical constraint' which blocks the assimilation in (b) <u>mA</u> and (f) <u>mAa</u>, there must be another explanation. This explanation, we believe, is <u>redundancy</u> (a 'redundancy constraint'); in other words, the need for maximal distinctiveness facilitates the assimilation in (d): <u>maA</u> rather than <u>*mAA</u> is maximally distinct from all the other monosyllables, as shown in (17). Compare also (16).

(17)

maA, not *mAA versus mA
↑_↑ ↑_↑ ↑_↑

maA, not *mAA versus MAA
↑_↑ ↑_↑ ↑↑

etc.

11.1 Implicit in this is the slight feeling that the tenseness distribution (reinforced with pitch if not whispered) may indeed have a linguistic function, although it is completely predictable with 'performance-related' characteristics. This is even more evident in two-syllable words with a nontense initial consonant. Here the whole first syllable will become nontense and get a level low pitch e.g. maaME 'lamp'. As a result such a syllable will function as a 'flag', and it will signal that the listener should expect another syllable before the next word boundary. This is of course related to the conspiracy demanding at least one tense segment per word, so if a long syllable passes by without such a segment there must be another syllable coming along. All this makes Tibetan resemble a pitch accent language, though it is but a tendency. Is perhaps Tibetan on its way to becoming a pitch accent language?

If the first syllable is short, however, as in maMA 'wet-nurse', it is no more of a flag than the ma- of maA 'butter'; similarly with all-tense syllables, as in MAAMI 'soldier'. Therefore we will not in fact admit Lhasa Tibetan into the number of pitch accent languages.

11.2. However, it is a curious coincidence, though, (or is it not a coincidence... ?) that the so-called pitch accent in Swedish has exactly the same flag function as in Tibetan, signalling that it is on the first syllable (if it is stressed) of a word with more than one syllable, or on the first stressed syllable in a compound (which will have two stressed syllables, as in Tibetan). Furthermore, in my idiolect, representative of Central Standard Swedish, it is even manifested with a pitch very similar to that of the Tibetan flag, namely a low and slightly falling pitch contour, whereby the 'ordinary' utterance intonation is delayed, or pushed, as it were, to the following syllable(s), which thus become(s) high — exactly as in Tibetan!

Interestingly, the Swedish case is different from the Tibetan in one very important respect. Although the place of the flag of either language is morphologically predictable, its pitch contour in Swedish has no apparent relationship at all to the segmental constituents of the syllable, whereas in Tibetan, as we have seen, all pitch contours are phonetically predictable. As nobody would claim that Swedish is a tone language just because Tibetan is not, this calls for yet another look at the definition of "tone language" (but we shall not deal with that question here).

12. Another problem that needs further examination is that of the historical development of tone. At least for Southeast Asian "tone languages" it seems to be quite widely assumed that tonogenesis is due mainly to the loss of final consonants (or the loss of the voiced-voiceless distinction in them), and that the tones have split further with the loss of the voicing distinction in the initial consonants. Now, if Tibetan is accepted as a budding (not only 'buddist'!) tone language, it must be noted that initial and final consonants are equally important in causing the transitions between intrinsic pitches to sound like tones, and to the extent that there can be talk of any "split" at all, it rather seems to be the final consonants that one has to watch. Sinologists and others are requested to reconsider their presuppositions on this matter.

13. In conclusion, we should like to point out that the account of the Tibetan tone-like phenomena as outlined in this paper has no bearing whatsoever on the question of whether tone is segmental or suprasegmental (cf. e.g. Leben 1973). The answer to this question depends rather on what we want to do: describe tones or explain them. Suprasegmental tone representations are very convenient as abbreviations for processes such as those formulated in

this paper, and were successfully and economically used for that purpose in Kjellin (1974), describing the very same idiolect of Lhasa Tibetan as here.

But if we want to explain anything, we must obviously look at the segmental level, as in this study. In less transparent cases we might have to make do with lexically tone-marked segments, as in Woo (1969), but even that method seems no more than an abbreviation of the potential explanation, which in such a case can probably be found by diachronic reconstruction.

Summary

We have tried to show that Lhasa Tibetan is still today a non-tonal language though perhaps well on its way to becoming tonal, and that it therefore may provide valuable clues as to tonal development in general. It was also shown that Lhasa Tibetan still today preserves underlying initial and final consonant clusters. Due to the different behaviour of cluster and non-cluster consonants, such phenomena as devoicing and tensing, it was argued, have paved the way for tone-like development, and all the processes are believed to be synchronic. As a by-product, a tentative explanation for why Tibetan does not possess word-initial vowels was offered.

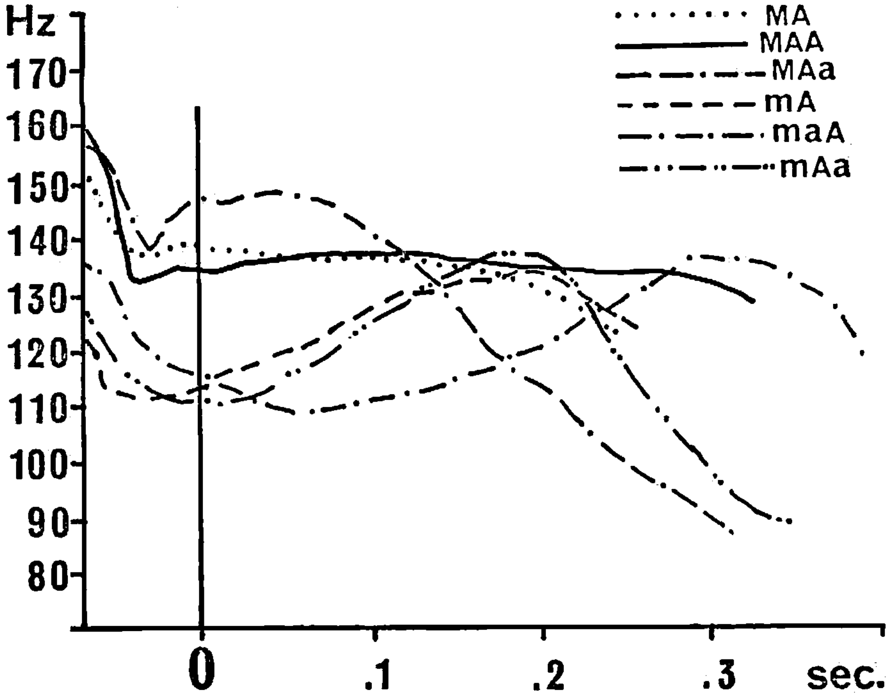


Fig. 1: The six fundamental-frequency contours of monosyllables. (The curves are manually smoothed averages of ten tokens of each type uttered in isolation and are aligned at the moment of oral release as determined by comparison of spectrograms, oscillograms, and intensity curves. The F_0 extraction was made on computer (FACOM 230-25) with the programs of Fujisaki and Mitsui (1973), by which, in essence, each fundamental glottal period (T_0) is measured and then the value for every 0.0128 sec. is interpolated and printed out. For further details, see Kjellin, 1975.)

APPENDIX 1.
Summary of the processes.

(P1) Transmigration;
(P2) Tenseness distribution
(P3) -Mz ⟶ -mz (tenseness) assimilation;
(P4) Devoicing of initial obstruents - if single;
(P5) Simplification of clusters;
(P6) Umlaut, nasalisation, alpha-vocalisation;
(P7) Tenseness assimilation;
(P8) Pitch assignment.

APPENDIX 2.
What happens to what syllable if compounded?

1st syll. \ 2nd syll.	MA	MAA	MAa	mA	maA	mAa
MA	1. MAMA	2. MAMAA	3. MAMAa	1. MAMA	2. MAMAA	3. MAMAa
MAA	4. MAAMA	5. MAAMAA	6. MAAMAa	4. MAAMA	5. MAAMAA	6. MAAMAa
MAa	4. MAAMA	5. MAAMAA	6. MAAMAa	4. MAAMA	5. MAAMAA	6. MAAMAa
mA	7. maMA	8. maMAA	9. maMAa	7. maMA	8. maMAA	9. maMAa
maA	10. maaMA	11. maaMAA	12. maaMAa	10. maaMA	11. maaMAA	12. maaMAa
mAa	10. maaMA	11. maaMAA	12. maaMAa	10. maaMA	11. maaMAA	12. maaMAa

= only 12 different types out of 36 theoretically possible

References

Bell-Berti, Fredericka, and Hajime Hirose (1972) Stop Consonant Voicing and Pharyngeal Cavity Size. Haskins Laboratories, Status Report on Speech Research, SR-31/32, pp. 207-211.

van den Berg, Jw. (1958) Myoelastic-aerodynamic theory of voice production. J. Speech and Hearing Research 1, pp. 227-244.

Chomsky, Noam, and Morris Halle (1968) The Sound Pattern of English. Harper & Row, New York, Evanston, and London.

Fujisaki, Hiroya and Yasuyoshi Mitsui (1973) An Improved Method for Pitch Extraction of Speech in the Time Domain. Reports of the 1973 Autumn Meeting of the Acoustical Society of Japan, pp. 129-130 (In Japanese).

Fujisaki, Hiroya, and Yoshihisa Tanabe (1972) A Time-Domain Technique for Pitch Extraction of Speech. Annual Report of the Engineering Research Institute (Faculty of Engineering, University of Tokyo), Vol. 31, pp. 259-266. (A shorter version appeared in Journal of the Acoustical Society of Japan 29:7, 1973, pp. 418-419.

Goldstein, Melvyn C., and Nawang Nornang (1970) Modern Spoken Tibetan. Univ. Washington Press, Seattle.

Gordon, Kent (1969) Sherpa phonemic summary. Summer Institute of Linguistics and Tribhuvan University, Kathmāndu.

Haudricourt, A. G. (1961) Bipartition et tripartition des systemes de tons dans quelques langues d'Extrême-Orient. Bulletin de Société Linguistique de Paris 56:1, pp. 163-180.

Haudricourt, André G. (1965) Les mutations consonantiques des occlusives initiales en mon-khmer. Bulletin de Société Linguistique de Paris 60:1, pp. 160-172.

Hirano, Minoru (1974) Morphological Structure of the Vocal Cord as a Vibrator and its Variations. Folia phoniat. 26, pp. 89-94.

Hooper, Joan B. (1972) The Syllable in Phonological Theory. Language 48:3, pp. 525-540.

Hyman, Larry M. (Ed.) (1973) Consonant Types and Tone. Souther California Occasional Papers in Linguistics, No. 1. (Los Angeles: Linguistics Program, University of Southern California).

Hyman, Larry M., and Russell G. Schuh (1972) Universals of Tone Rules: Evidence from West Africa. Working Papers on Language Universals 10, pp. 1-50. (Revised version in Linguistic Inquiry 5:1, 81-115, 1974).

Kent, R. D., and K. L. Moll (1969) Vocal-Tract Characteristics of Stop Cognates. J. Acoustical Society of America 46:6/2, pp. 1549-1555.

Kjellin, Olle (1974) A Preliminary Report on Tibetan Tones. Ann. Bull, RILP 8, pp. 115-123.

__________ (1975) Progress Report on an Acoustical Study of Pitch in Tibetan. In this issue.

Lea, Wayne A. (1973) Segmental and Suprasegmental Influences on Fundamental Frequency Contours. In Hyman 1973, pp. 16-70.

Leben, William R. (1973) The Role of Tone in Segmental Phonology. In Hyman 1973, pp. 116-149.

Lehiste, Ilse, and Gordon E. Peterson (1961) Some Basic Considerations in the Analysis of Intonation. J. Acoustical Society of America 33:4.

Lindqvist, J. (1972) Laryngeal Articulation Studied on Swedish Subjects. STL-QPSR (Speech Transmission Laboratory, Quarterly Progress and Status Report) 2-3, pp. 10-27.

Maran, La Raw. (1971) Burmese and Jingpho: A Study of Tonal Linguistic Processes. Occasional Papers of the Wolfenden Society on Tibeto-Burman Linguistics, Vol. IV. Urbana, Illinois.

__________ (1973) On Becoming a Tone Language, A Tibeto-Burman Model of Tonogenesis. In Hyman 1973, pp. 98-114.

Matisoff, James A. (1973) Tonogenesis in Southeast Asia. In Hyman 1973, pp. 72-95.

Miller, P. M. (1951) The Phonemes of Tibetan (U-Tsang Dialect) with a Practical Orthography for Tibetan-speaking Readers. J. Asiatic Society, Letters 17:3, pp. 191-216.

Miller, Roy Andrew (1954) Morphologically Determined Allomorphs in Spoken Tibetan. Language 30:4, pp. 458-460.

Mohr, B. (1971) Intrinsic Variations in the Speech Signal. Phonetica 23, pp. 65-93.

Perkell, J. S. (1965) Cineradiographic Studies of Speech: Implications of a Detailed Analysis of Certain Articulatory Movements. Reports to the Fifth International Congress of Acoustics 1, A32, Universite de Liege.

__________ (1969) Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study. MIT Press.

Pike, Eunice V. (1974) A Multiple Stress System Versus a Tone System. International Journal of American Linguistics 40:3, pp. 169-175.

Richter, Eberhardt. (1964) Grundlagen der Phonetik des Lhasa-Dialektes. Schriften zur Phonetik, Sprachwissenschaft und Kommunikationsforschung Nr. 8. Akademie-Verlag, Berlin.

Rothenberg, M. (1968) The Breath-stream Dynamics of Simple-released-plosive Production. New York. S. Karger, Publishers.

Shefts, Betty, and Kun Chang (1967) Spoken Tibetan Morphophonemics: p. Language 43:2, pp. 512-525.

Simon, Walter (1970) Some Suggestions Towards a Romanisation of Modern Tibetan (Lhasa Dialect). In Roman Jakobson and Shigeo Kawamoto (eds.), Studies in General and Oriental Linguistics, pp. 535-539. TEC, Tokyo.

Sprigg, R. K. (1954) Verbal Phrases in Lhasa Tibetan - I. BSOAS (Bulletin of the School of Oriental and African Studies) 16:1, pp. 134-156.

__________ (1955) The Tonal System of Tibetan (Lhasa Dialect) and the Nominal Phrase. BSOAS 17:1, pp. 133-153.

__________ (1972) A Polysystemic Approach, in Proto-Tibetan Reconstruction, to Tone and Syllable-initial Consonant Clusters. BSOAS 35:3, pp. 546-587.

Strahm, Esther, and Anita Maibaum (1971) Jirel Phonemic Summary. Summer Institute of Linguistics, Tribhuvan University, Kathmandu.

Woo, Nancy (1969) Prosody and Phonology. MIT Dissertation. (Reproduced 1972 by Indiana University Linguistics Club).

Yukawa, Yasutoshi (1971) Some Problems in Tibetan Phonology. In Yukawa, Y., Basic Problems in Linguistics, Ch. 5, pp. 141-160. Taishuukan, Tokyo. (In Japanese).