A PRELIMINARY REPORT ON TIBETAN TONES

Olle Kjellin

In this paper I shall make two different attempts at describing phonologically the distinctive tones of modern spoken Lhasa Tibetan and how they change in certain circumstances.

First a few words about the syllable and word structure. Short syllables are of the common CV type. Long syllables have an extra vowel appended (CVV), or a nasal (CVN). These types constitute the majority of syllables. Then there are a number of CVC syllables, where the final consonant is a stop ([b] or [g]) or a liquid ([l] or [r]). But even most of these seem to be no more than alternative pronunciations of CVV syllables (which fact, of course, has its diachronic implications, which are beyond the scope of the present article). These, too, will be referred to as the long type.

Words may consist of only a single syllable, but very often there are two syllables in one word (which I shall call a compound) and sometimes three or even four.

The tones of the syllables in compounds will be called combination tones; those of monosyllabic words will be called isolation tones (without implying that such a word is uttered in isolation, but not excluding such instances, either).

The tones are described somewhat differently by different investigators, presumably as a consequence of the wide diversity of dialects (including, so it is said, some dialects without tonal distinctions) and the absence of a defined 'standard language'. Another drawback with the tonal descriptions in the literature so far is the peculiar fact that the writing has been taken as a basis for explaining the tones and tone changes. True, the writing can give hints and clues as concerns the historical development of the tones, but this paper is intended to treat them synchronically. In this situation and so far I have relied more on my one informant [1] than on any amount of printed information; and for the fundamental analysis I thank my first Tibetan teachers at Tokyo University of Foreign Languages, the Institute for Asian and African Studies.

According to them there are four tones to be found on isolated syllables (the pitch values are theirs, based on auditory impressions only):

Tone I [˥ , ˧˥ (55)] is high pitched and level.

Tone II [˧˦ (13)] rises from low to mid. Personally, I suspect "to high", but more extensive instrumental measurements will have to confirm or refute that. In short syllables, the tone seems to be just low, at most with a tendency to rise.

Tone III [˥˩ (53)] was said to be high-to-mid falling; perhaps one should call it high-to-low [˥˩ (51)]. A variant pronunciation cuts off the end of the fall with a glottal stop. Tone III occurs only in long syllables.

Tone IV [↗↘ (131)] is low pitched, begins with a slight rise, then turns to a fall; it could thus be called a 'convex' tone. This tone, too, occurs only in long syllables and has a variety with a glottal stop instead of the entire fall. It then resembles Tone II, which, however, has no glottal-stop variety.

This is only an approximate phonetic (auditory) description of how the tones are realized. Sound spectrograms have been consulted for a rough comparison with the impressionistic images, but a closer scrutiny with a pitch extractor still remains to be performed, as do intensity measurements.

The same has to be said about the combination tones: Auditorily, spectographically, and according to my teachers, the left combination tone (i. e. that which is in the first syllable of a compound as defined on the preceding page) is always level whether high or low, whether long or short, whereas the right combination tone may be either level or falling but has to begin on high pitch. In other words, to the left one may find ˥, ˧, ˩, or ˩; to the right ˥, ˧, or ↘. The right combination tone is found in the second syllable of a compound; all syllables after that (if there are any) have a kind of 'neutral' tone not yet analyzed by me, but apparently mid or low pitched. In certain morphological (paradigmatic) environments the second syllable is 'neutral'.

The high left combination tone stems from tone I or tone III.
The low left combination tone stems from tone II or tone IV.
The level right combination tone stems from tone I or tone II.
The falling right combination tone stems from tone III or IV.

With these facts at hand, it seems as if there are certain characteristics of some of the tones that do not need to be included in the phonological representations, and carefully keeping the phonetic output problem out of the treatment the phonologist can allow himself to do a certain amount of 'tailoring' of the data in order to be able to find the simplest possible rules to account for the phonological changes.[2]

Thus, for instance, seeing that the rising character of tone II and of the beginning of tone IV does not appear to influence their respective combination tones - left or right - whatsoever, and that it is not even essential for the discrimination between any of the tones, thus having no linguistic value (other than redundancy), one might speculate that this dialect could do as well without this rise. Indeed, a simple preliminary experiment - which will be repeated under better controlled conditions[3] - seems to justify this view inasmuch a long low level tone was interpreted by the informant as tone II, and a low, further falling tone as tone IV. (The rise actually found in the phonetic output, then, will have to be hypothesized to be the result of some physiological constraint of the larynx, or just to be a redundancy element.)

Thus, by extracting the relevant parts of the tones, it becomes clear that the tones are either level or falling, on either a high or a low pitch. Consequently, the features needed are [High Tone] and [Falling Tone], yielding the matrix shown in Table 1.

Tone I	Tone II	Tone III	Tone IV
˥	˩	˧˥	˥˩
[+High, -Falling]	[-High, -Falling]	[+High, +Falling]	[-High, +Falling]

Table 1. The phonological representations of the isolation tones. Redundant features like [Contour], etc., are omitted.

Table 2 is a listing of all possible tone combinations and their resultant tone configurations:

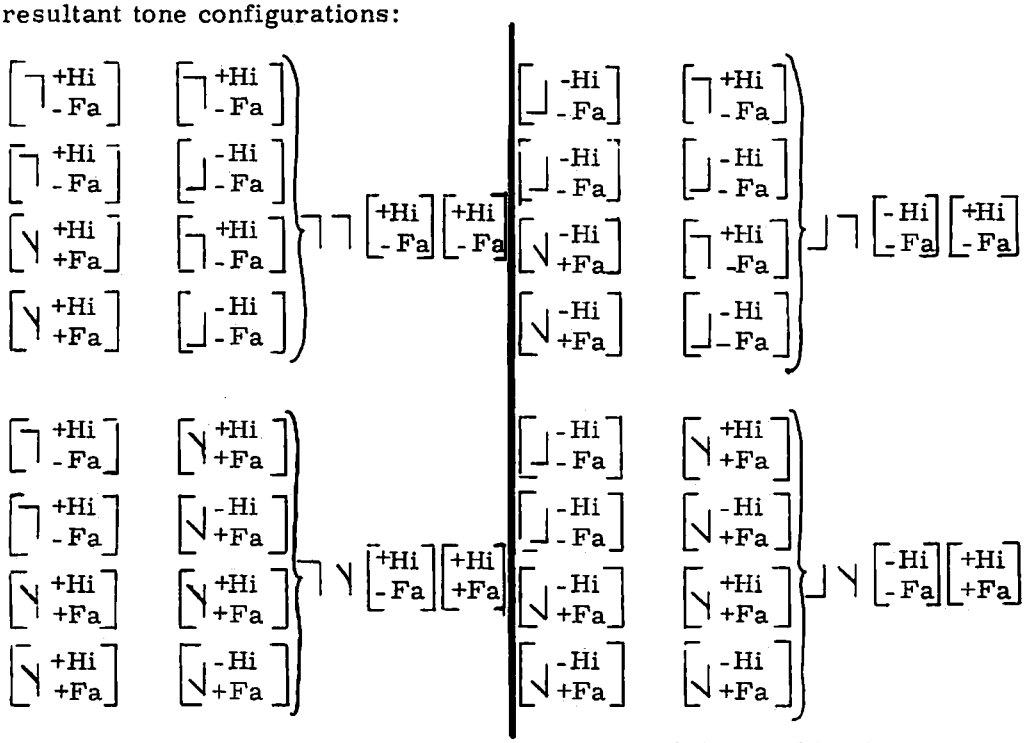

Table 2. The tone combinations and the combination tones.

This may seem complex at first, but it is evident that the left combination tone is always [-Fa] and thus this feature can be left unmarked whereas the value of the feature [Hi] has to be decided case by case; and that the reverse is true about the right combination tone, as shown in the following matrix:

[Left	[Right
mHi	uHi
uFa]	mFa]

Marking convention: $u \rightarrow \left\{ \begin{matrix} - \ / \ \# \ \underline{\quad} \ Z \\ + \ / \ \# \ Z \ \underline{\quad} \end{matrix} \right\}$ (where Z = 'syllable')

The marked features of course get their respective values directly from the 'original', underlying, tones, yielding the following output:

$$\# \begin{bmatrix} \alpha Hi \\ -Fa \end{bmatrix} \quad \begin{bmatrix} +Hi \\ \beta Fa \end{bmatrix}$$

Thus the phonological rules required to obtain this result become very simple, namely in effect identical to the marking convention:

(1) $Z \rightarrow [-Fa] \; / \; \#\_\_\_Z$

(2) $Z \rightarrow [+Hi] \; / \; \#Z\_\_\_$

(where Z = 'syllable').

Rule (1) makes the left combination tone non-falling;

Rule (2) makes the right combination tone high.

These rules (barely) cover the data but do not enable us to draw any conclusions or make any predictions about tones in general or in particular.

The feature system used above is the one most commonly used for Asian tone languages (see especially Wang (1967)).

In her dissertation, Nancy Woo (1969) argues convincingly against the idea of having features like [Rising], [Falling], [Convex], etc. (so-called dynamic features) on a whole syllable as such; because a matrix including dynamic features "does not represent a set of physiological instructions which are to be realized simultaneously, but sequentially, as are the instructions for series of matrices within boundaries" (p. 36 f.). Thus, for instance, in the case of the dipping third tone in Mandarin Chinese both the features [Rising] and [Falling] must be present (with plus marks), together with a 'traffic' feature [-Convex] to ensure that the fall comes before the rise.

However, should we really have such a feature with no articulatory correlate but only with an ordering function; that is, a feature that orders features which already, in themselves, order sequences of pitch?

Instead Woo suggests that tone features not be suprasegmental but segmental. Among all its other specifications each segment is also to have a pitch specification in its feature matrix. While the pitch of a vowel or a nasal can be varied by the speaker, certain segments seem to have intrinsic pitch: Voiced and voiceless obstruents naturally have low and high pitch, respectively (Mohr (1971) presents very interesting data on this point). Woo's investigations underlying her dissertation (op. cit.) have led her to conclude that the pitch of the initial segment (whether intrinsic or deliberate) of a tone-language syllable does not contribute significantly to the tone of the syllable (unless, of course, the initial segment is the syllabic segment).

Following the example set by Woo (1969), I shall now proceed to reformulate the Tibetan tone representations and rules.

The syllable-structure formula is CV(X), where C is the initial consonant, whose intrinsic pitch need not concern us unless we are interested in the origin of the tones; and where V is the vowel constituting the

syllabic nucleus of the syllable; and where X is a nonsyllabic segment which may be one of those that can end a syllable, namely a vowel, a nasal, and, rarely, a stop or a liquid. The nucleus V carries the significant pitch; and X, when present, completes it, resulting in either a level or a dynamic, long tone across the syllable. Even when X is a stop, its intrinsic pitch seems to be relevant to the shape of the tone.

In CV syllables there are only two possible tones, high level and low level, since there can be no dynamic tone curve without the X segment. (Thus this fact about Tibetan is correctly predicted by Woo's theory, which is based on work on other languages!)

In CVX syllables, consequently, there may be rising or falling dynamic tones, if both of the relevant pitch specifications are different (a case which may be considered parallel to that of tongue heights in diphthongs, as Woo also points out), and long level tones, as well, if the two specifications are equal.

Reexamining the phonetic descriptions on page 2, it becomes obvious that in this new approach to the analysis of tone it is possible to follow the phonetic reality more closely; there is no longer the urgent need to 'disregard' certain phenomena for the convenience of the phonologist. The new feature matrices of the tones are presented in Table 3:

	Tone I	Tone II	Tone III	Tone IV
	˥, ˦	˩ ˧	˨˦	˧˥˧
	CV(X)	(CV(X)	CVX	CVXX
High pitch	[+(+)]	[−(+)]	[+ −]	[− − −]
Low pitch	[− (−)]	[+(−)]	[− +]	[+ − +]

Table 3. The refined phonological representation of the isolation tones.

It should be noted, here, that Tone II and Tone III are 'dynamic' to the extremes rather than to the mid (i. e. ˧˥ instead of ˧˦, and ˨˦ instead of ˨˧ or ˨˩, respectively). Although the reason for this is partly in the anticipation of the rules to come (any of the alternatives would complicate the rules considerably[4]), it by no means amounts to 'tailoring': my own perception of these tones suggests these forms, and the simple spectrograms taken so far do not deny them.

Tone IV will need an 'intrusive' segment because of its circumflex shape requiring three different pitch specifications, but the presence of this extra nonsyllabic segment to carry the ridge of the convex tone is not so strange as it first might seem. Woo uses the same device to explain the dipping third tone of Mandarin in isolation (see for instance op. cit. pp. 45, 71, or 76f.), and thereby also predicts it would be longer than any of the other tones, which her acoustic measurements actually confirm (see tables in chapter II). This longer duration is also easily discriminated by the ear, but I see no reason to claim that it has to be longer: After all,

the intrusive segment is supposed to be non-syllabic, and at most it can be expected to cause a shortening of the other segments in the same syllable in order to keep the entire syllable duration within a (conceivably) pre-set norm (for studies on this matter, see Rapp (1971) and Lindblom & Rapp (1973)).

Thus, in the case of the Tibetan convex tone the same prediction, or claim, could have been made a la Woo, but my simple data so far do not suggest that it is true, although there is a very slight tendency of prolongation. Anyway, (to extend the parallel on top of the previous page:) triphthongs are not always longer than diphthongs, and yet they must be thought of as having two nonsyllabic segments in connection with the syllabic nucleus.

It is likely, however, that the intrusive segment is absent from the underlying representation of the tone. Of course it is desirable to have underlying forms that resemble the surface forms as much as possible, but they do not necessarily have to be identical, because then (in this case, at least) the isolation and corresponding combination tones would each need different underlying forms, and so the situation would even be worse[5].

The underlying representations I have found adequate for tones I, II, and III, happen to conform with the isolation tones shown in Table 3. That for tone IV, however, is posited without the intrusive X; see Table 4.

	Tone I	Tone II	Tone III	Tone IV
	˥, ˥	˩. ˧	˥˩	˩˥
	CV(X)	CV(X)	CVX	CVX
High pitch	[+(+)]	[−(+)]	[+−]	[−−]
Low pitch	[−(−)]	[+(−)]	[−+]	[++]

Table 4. The refined phonological representation of the underlying tones.

It is worth noting that under the present analysis the underlying form of tone IV has the same contour as had that of tone II in the previous one, i. e., low level. This may be more than a coincidence, since in fact the difference between II and IV cancels out in the left combination position.

Similarly, the difference between tones I and III also cancels out in the left combination position, while those that become equalized in the right combination position are II with I, and IV with III. These processes are summarized in Table 5, and the feature matrices of the combination tones are shown in Table 6.

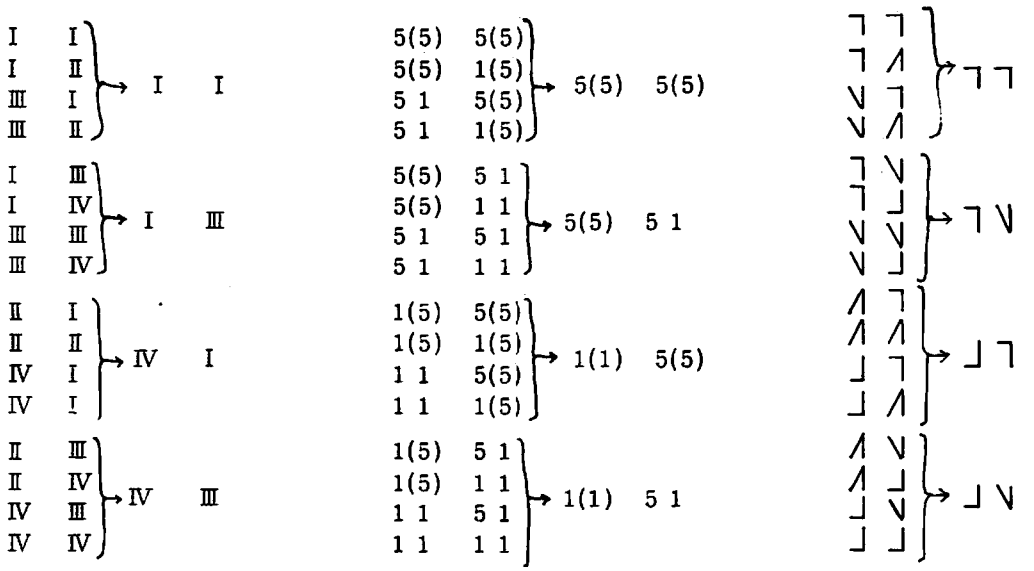

Table 5. The tone combinations: in tone 'names', pitch values, and in tone letters.

(a) The left combination tones:

	Tone I	Tone II = IV	Tone III = I	Tone IV
	˥, ˥	˩, ˩	˥	˩
	CV(X)	CV(X)	CVX	CVX
High pitch	[+(+)]	[−(−)]	[++]	[−−]
Low pitch	[−(−)]	[+(+)]	[−−]	[++]

(b) The right combination tones:

	Tone I	Tone II = I	Tone III	Tone IV = III
	˥, ˥	˥, ˥	˩	˩
	CV(X)	CV(X)	CVX	CVX
High pitch	[+(+)]	[+(+)]	[+ −]	[+ −]
Low pitch	[−(−)]	[−(−)]	[− +]	[− +]

Table 6. The feature matrices of the combination tones.

Now the point has been reached where it is possible to state the new (refined) rules, by which the underlying tones as given in Table 4 are transmuted into the combination tones (Table 6), and into the isolation tones (Table 3).

For the technical statement of the rules I have chosen to retain the 'CV(X)' shorthand, rather than state long lists of features such as

[Sonorant] or [Syllabic], in order to make the rules easier to read: but I have abandoned the Z for the whole syllable and spell it out as CV(X) instead. Thus, the reader will recall, C contains all the features for the initial consonant, V those for the vowel that constitutes the syllabic nucleus, and X those for the optional nonsyllabic segment that may make a syllable long. [Hi] and [Lo], of course, refers to pitch and not to tongue or vowel height.

(R 1) $$X \rightarrow \begin{bmatrix} \alpha \text{Hi} \\ \beta \text{Lo} \end{bmatrix} \; / \; \# \; C \; \begin{bmatrix} V \\ \alpha \text{Hi} \\ \beta \text{Lo} \end{bmatrix} \; \underline{\quad\quad} \; CV(X)$$

(R 2) $$V \longrightarrow \begin{bmatrix} +\text{Hi} \\ -\text{Lo} \end{bmatrix} \; / \; \# \; CV(X) \; C \underline{\quad\quad} (X)$$

Rule (R 1) will only affect long dynamic tones in the left part of the compound, namely tones II and III, and make them level by assigning to X the same pitch as the nucleus has. Rule (R 2) will affect low-pitched syllabic nuclei in the right part, namely of tones II and IV, and make them high-pitched.

For non-compounds the following rule is needed:

(R 3) $$\emptyset \rightarrow \begin{bmatrix} X \\ \alpha F \\ -\text{Hi} \\ -\text{Lo} \end{bmatrix} \; / \; \# \; C \; \begin{bmatrix} V \\ \alpha F \\ +\text{Lo} \end{bmatrix} \; \underline{\quad\quad} \; \begin{bmatrix} X \\ +\text{Lo} \end{bmatrix} \; \#$$

which gives the convex contour to the long low-pitched tone IV in isolation ([α F] = [α Features] to make this X identical with V in all other respects)."

This method of representing (the) tones further enables one to describe naturally the process in Tibetan in which the falling pitch contour may be cut off with a glottal-stop-like abruptness as in tone III and IV, in isolation as well as in right combination:

(R 4) $$[\,+\text{Lo}\,] \longrightarrow \text{ʔ} \; / \; [\,-\text{Lo}\,] \; \underline{\quad\quad} \; \#$$ (optional)

The rule suggests that the laryngeal gesture required to produce a low pitch is simply exaggerated when it is to follow immediately upon a non-low pitch; but whether in fact it is so 'simple' has yet to be investigated more thoroughly.

I have presented the tones of modern spoken Lhasa Tibetan and suggested some tone rules stating them in two ways: in the spirit of Wang (1967) in the first instance, and in that of Woo (1969) in the second. Although neither should be considered the 'final solution', I am apt to regard the latter as more suitable for the purpose, perhaps more accurate and

inspiring as concerns side implications, and more consistent with phonological theory in general.

Acknowledgments

Apart from my informant and teachers already mentioned, I must especially thank Robert N. Romero for his comments and criticism on an earlier version of this paper.

Since the author is only in the beginning of his Tibetan research all kinds of comments and suggestions will be highly appreciated. Particularly I lack information on other, recent, investigations and references on Tibetan phonetics and phonology. (I am not, however, unaware of E. Richter's "Grundlagen der Phonetik des Lhasa-Dialektes', Berlin, 1964, or of R. K. Sprigg's articles in the Bulletin of the School of Oriental and African Studies, London, but for various reasons I have found it best simply to disregard these works.)

Notes

1) born in Shigatse, where the TSANG dialect is spoken; between 7 and 22 years of age residing and studying in Lhasa.

2) Many phonologists do so, at least; whether it is recommendable is a different matter.

3) Without revealing my intentions, I pronounced syllables (words) with a long low tone and with a low falling tone and asked what they meant, pretending I did my best to pronounce them "correctly. " Of course this experiment must be performed with synthetic speech; the present results are not reliable.

4) Moreover, as Romero (forthcoming) points out, there is probably no language in which a distinction is made by the end points in pairs of dynamic tones such as ˩ and ˩˧, or ˧˥ and ˥.

5) Admittedly, it is impossible to define what is legitimate in this respect, and what is unwarrantable.

References

Lindblom, B. , and K. Rapp: "Some Temporal Regularities of Spoken Swedish," in PILUS (Papers from the Institute of Linguistics, University of Stockholm) 21/Sept. 1973.

Mohr, B. : "Intrinsic Variations in the Speech Signal," Phonetica 23/1971.

Rapp, K. : "A Study of Syllable Timing." Speech Transmission Laboratory-Quarterly Progress and Status Report 1/1971.

Romero, R. N. : "The South Min Tone Circle," forthcoming.

Wang, W. S-Y. : "Phonological Features of Tone," International Journal of American Linguistics 33. 2/1967.

Woo, N. : "Prosody and Phonology," MIT dissertation, 1969; reproduced by the Indiana University Linguistics Club, Jan. 1972.

Annual Bulletin (Research Institute of Logopedics and Phoniatrics, University of Tokyo), No. 8 (1974)