The Complex Tones of East/Southeast Asian Languages: Current Challenges for Typology and Modelling

Alexis Michaud 1 2

1 Centre National de la Recherche Scientifique, CNRS – LACITO, France
2 Centre National de la Recherche Scientifique, CNRS – CEFC, Hong Kong/Taipei
alexis.michaud@vjf.cnrs.fr

Abstract

In some of the tone systems of East and Southeast Asian languages, linguistic tone cannot simply be equated with pitch; some tones have phonation-type characteristics as part of their phonological definition; and there is no compelling evidence for analyzing tonal contours into sequences of levels. Salient findings are reviewed, first from a synchronic perspective, then from a diachronic one, to bring out facts that are relevant for tonal typology and for evolutionary approaches to phonology.

Index Terms: tonal typology, Asian tones, phonation-type characteristics, tonogenesis, registogenesis, diachrony of tone systems, level tones, complex tones

1. Introduction

Since the early 20th century, considerable progress has been made in the study of tonal systems of East and Southeast Asia. An increasing number of languages have come under academic scrutiny; together with synchronic descriptions, historical studies have attained an increasing degree of precision, and the diachronic origin of tones in many of the languages of the area is now well-understood. As far as tonal typology and the phonological modelling of tones are concerned, however, there remain some incomprehensions between researchers. The present paper reviews evidence on the complex tone systems of Asia, first from a synchronic perspective, then from a diachronic one, to bring out facts that are relevant for tonal typology and for evolutionary approaches to phonology.

2. A brief review of synchronic studies of complex tones

Chao Yuen-ren’s work on Mandarin Chinese in the early 20th century [1], [2] brought to the attention of linguists the complexities of its tone system [see also 3]. It is clear to linguists grappling with the tone systems of Chinese dialects (“Sinitic languages”) that “Chinese tones are complex suprasegmental events distinguishing (otherwise) homophonous syllables” [4]. Following sustained exchanges between Chao Yuen-ren and Kenneth Pike, Chao’s findings were taken up in Pike’s typological divide between registertone* and contour-tone*, which recognized that “the glides of a contour system must be treated as unitary tonemes and cannot be broken down into end points which constitute lexically significant contrastive pitches” [5]. Later studies [6–10] have brought out the importance, in some languages of Asia, of phonation-type characteristics that are part and parcel of the definition of tones. In these systems, phonation types are not a distinctive feature orthogonal to tone, unlike in the Oto-Manguean language Trique, for instance [11]. Pike’s two-way typology, while it emphasizes important properties of the languages which he was able to take into consideration, has some limitations in this respect: characterization as “contour tone” may not be an adequate label for tones such as those of Vietnamese, which contrast with the other tones of the system through a set of characteristics that include specific phonation types* in addition to the time course of fundamental frequency (‘height modulation’, ‘melodic contour’).

Languages with phonetically complex tone systems constitute a challenge for tonal typology and phonological modelling. There clearly remains some progress to make in this area. Given the complexity of linguistic tone, tonal typology is a formidable task, which requires taking into account considerable amounts of data. Specialists typically focus primarily on one area or language group, with the unwanted result that they sometimes tend to grant universal status to the characteristics that they repeatedly observe. Eugénie Henderson makes her choice clear: “My preference, derived both from professional training and experience, would be to present only material of which I have first-hand personal knowledge, since, though this may be fallible, one may at least suppose the same bias to run through the whole of it” [12]. This can lead one to consider with some suspicion the work of colleagues who, starting out from different languages, come to different conclusions. Eugénie Henderson, who is familiar with complex tones, warns that “… ‘tone’ is seldom, if ever, a matter of pitch alone” [12]; this “if ever” amounts to casting doubt on the validity of tonal descriptions that do not mention phonation type* and other potential phonetic correlates of tone [see also 13]. One of the conclusions of Martine Mazaudon’s study of the complex tones of the Tamang-Gurung-Thakali-Manangke group (Sino-Tibetan, Nepal) is that “The pitch-only tones which are reported especially from African languages might be a special case rather than the prototype, as they are commonly presented” [14]. Conversely, generative phonology adopted the “autosegmental” representations initially developed for the level tones* of African languages [15–17] and raised it to the status of universal representation of tone. Adherents of generative phonology therefore tend to assume that analyses into level tones* apply to all tone systems, in East Asia as elsewhere, even in the absence of compelling language-internal evidence.

I am fortunate to be familiar with a few strikingly different tone systems of Asia: Yongning Na (Sino-Tibetan), which has many properties that will be familiar to Bantuists: tonal spreading, tonal reassociation, and tonal morphology [18], [19]; the closely related language Naxi, which likewise has H(igh), M(id) and L(ow) levels as basic units, but has

1 Terms in italics followed by an asterisk are defined in the Appendix (§5).
much fewer tonal alternations [20–22]; Vietnamese; and Mandarin Chinese. To me, it is equally clear (i) that Yongning Na is to be described as having a *level-tone* system, and (ii) that “there are no objective reasons to decompose Vietnamese tone contours into level tones or to reify phonetic properties like high and low pitch into phonological units such as H and L. In Vietnamese, tone features would only duplicate perceptual properties that have to exist independently in the phonetic grammar of listeners” [23; see also 24, 25].

I hope that this diversified experience will to some extent alleviate suspicions of a bias towards a given theoretical approach. *Level-tone* representations have proved useful way beyond the Subsaharan domain, for which they were initially developed: in the Americas [26–29], Asia [30–33, 20, 34, 35], New Guinea [36, 37] and elsewhere. They do not meaningfully apply to all systems, however. It is of course possible to propose *ad hoc* amendments to autosegmental representations to describe unitary contours. One such proposal [38] consists in describing the contours of Mandarin by sequences of level tones (e.g. Low+High for a rising tone) all of which are linked to a single tonal node, as in (1a) below, instead of one tonal node for each level in ‘true’ *level-tone* systems, as in (1b) below:

![Diagram](image-url)

(a) contour tone  b. tone cluster

V      V                  (tone-bearing unit)
\|\|\
\o\o\o                  (tonal node)
L H L H                  (tones)

Nick Clements (p.c. 2008) suggested instead that contours should be viewed as units unless there is positive evidence to the contrary. This burden of proof should not appear as too heavy: morpho-phonological tonal alternations can provide decisive evidence that a tone system is based on *level tones*, and in many languages such evidence is easy to come by, e.g. (in Asia) Prinmi/Pumi [31] and Hakha Lai [32].

Conversely, demonstrating that complex tones cannot be decomposed is no simple task, since the failure of attempts at decomposition could in theory result from shortcomings in the analysis, not from a real impossibility to decompose these tones. Relevant evidence can nonetheless be found, for instance from speech errors. If a tone consists in a sequence of *level tones*, one would expect to find speech errors in which only one of these levels is omitted or modified: e.g. a L or H realization of a L+H tone. This does happen in *level-tone* systems; some examples are found in a corpus of Yongning Na which is freely available online [19]. On the other hand, observations on *complex tones* suggest that, in tonal speech errors, one of the tones of the system is substituted holus-bolus for another, e.g. in Mandarin [39]. (Larry Hyman [p.c.] notes that tonal rules in Sinitic languages are likewise replacive, rather than assimilatory.) The authors of the study of Mandarin speech errors conclude that “phonological theories which require that all contour tones in every language must be represented as a sequence of level tones underlingly may be missing an insight into the possible underlying differences among tone languages” [39]. This echoes the observations of practitioners of tonal analysis in Asia [e.g. 40].

One can reasonably hope that these arguments will gradually lead to a general recognition of the existence of more than just one possible structure for lexical tones. “[In a number of languages] contour tones were clearly shown to be composed of a sequence of level tones. The phenomena that gave rise to the analyses of contour tones as such are not found in Chinese languages; however, in the interest of theoretical parsimony, the same analysis in which contour tones are treated as a sequence of level tones was carried over to Chinese tonal systems, in the absence of evidence to the contrary. However, we have seen enough evidence to doubt that a sequence of level tones is an appropriate representation for the contour tones under discussion and that, in fact, contour tones in Chinese tonal languages should be seen as unitary” [41].

The challenges for phonological modelling are obvious. It is difficult to pinpoint the exact nature of tones that are phonetically complex and do not behave phonologically as levels (or sequences of levels). Shorthand notations developed for Chinese dialects (Sinitic languages), such as Chao’s five-point scale [42] and the expanded version currently developed by Sean Zhu [43], aim to capture the relevant aspects of fundamental frequency contours. This stylization provides a relatively precise description of the tones’ contours. It does not offer any means to transcribe phonation-type distinctions, however, and specialists commonly consider that a thorough description requires experimental data: audio recordings, and supplementary phonetic data for phonation types where applicable [e.g. 10].

Diachrony can shed some light on this difficult topic. Section 3 reviews facts and hypotheses.

3. Diachronic insights into Asian tones

The present section briefly recapitulates salient facts about the evolutions leading up to the creation of complex tones. Some hypotheses are put forward about a possible relationship between the historical phasing of the various stages of tonogenesis and the properties of the resulting lexical tones.

3.1. Tonogenesis and registrogenesis

Tonogenesis can result from the loss of various phonemic oppositions, through a mechanism of compensation (*transphonologization*): lexical contrasts are preserved (at least in part) by means of a new opposition. These processes are now well understood [see 44 for a worldwide survey]. Taking the textbook case of Vietnamese [45, 46], Table 1 recapitulates the evolution from a stage when the language did not have tone (Table 1a) up to the present-day system (1c) via a stage where there were three tonal categories (with stop-final syllables as a distinct, fourth set).
Table 1. Vietnamese tones in diachronic perspective

Table 1a. Rhymes in Late Proto-Viet-Muong: open syllables without glottalization; final glottal constriction; final /h/; final /p/, /t/ or /k/. After [47].

<table>
<thead>
<tr>
<th>tone</th>
<th>da</th>
<th>da’</th>
<th>tap, tat, tak</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
</tbody>
</table>

Table 1b. First stage of tonogenesis in Vietnamese: three tones; no contrasts on stop-final syllables.

<table>
<thead>
<tr>
<th>tone</th>
<th>da</th>
<th>da’</th>
<th>tap, tat, tak</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
</tbody>
</table>

Table 1c. The tone system of contemporary Hanoi Vietnamese: tonal categories in etymological notation, and name given in the orthography. Tones A1 to C2 only appear on open or nasal-final syllables, and tones D1 and D2 on stop-final syllables.

<table>
<thead>
<tr>
<th>tone</th>
<th>da</th>
<th>da’</th>
<th>tap, tat, tak</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>A2</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
</tbody>
</table>

The details of the process whereby these transphonologizations take place are increasingly well understood. Pulleyblank was the first to suggest that phonation-type oppositions were an intermediate stage in the two-way split (second stage of tonogenesis) in Chinese: "...the split into upper and lower registers was conditioned primarily by the voiced aspiration, giving rise to breathy vowels, rather than simple voice" [48]. The study of Khmer by Henderson had already provided key insights on how a voicing opposition on initial consonants could become a phonation-type register* opposition [49; for a synthesis, see 50, 51]. Henderson’s article was of central importance to explain the evolution of voicing oppositions among initial consonants, shedding light on the diachronic links between consonants, registers and tones.

The following paragraph sets out a hypothesis concerning the link between the diachronic origin of tones and their synchronic properties.

3.2. The origin of complex tones: a hypothesis

A hypothesis proposed by Michel Ferlus (p.c.) is that tones comprising phonation-type characteristics are created in cases where the tones resulting from the first stage of tonogenesis (see Table 1b above) are still in a stage where they preserve phonation-type characteristics associated with earlier final consonants, such as glottalization or breathy voice, at the time when the second stage of tonogenesis begins (Table 1c).

This hypothesis would explain in part the limited geographic extent of these lexical tones. Their appearance requires the conjunction of two specific structural properties: the inception of a split of the tone system at a stage when a previous tonogenetic process (typically: tonogenesis by final laryngeal consonants) is still in progress, i.e. before the tones have reached a 'pure-pitch' stage. This conjunction took place in a number of East and Southeast Asian languages because of similarities in their syllable structure and in the evolution of their syllable canon [see 51 for a synthesis].

As for level-tone* systems, their wide geographical distribution makes sense in light of the fact that they can have various diachronic origins: they can result from the transphonologization of contrasts on initial consonants, as in Oceanic languages [52–54], or on final consonants, as in Athabaskan [26].

After they come into existence, tones involving phonation-type characteristics may change, and their specific phonation type disappear: for instance, Hanoi Vietnamese still has glottalization in two of its tones, whereas Southern Vietnamese does not retain any phonation-type characteristics [23, 25]. It seems that no hard-and-fast boundary can be drawn between tones with and without phonation-type characteristics. In the vast domain of Sinitic languages (Chinese dialects), the tones of some varieties clearly have specific phonation [e.g. 7, 9] whereas for others (e.g. Mandarin dialects) it can be debated to what extent the nonmodal phonation which is occasionally present for some tones is part of their phonological definition.

The final section of this paper sets out reflections on possible differences between tone systems in terms of diachronic evolution.

3.3. Differences in evolutionary potential between level tones and complex tones

There appear to be salient differences in evolutionary potential between different tone systems. While Pike’s two-way distinction between “register tones” (level tones*) and “contour tones” is certainly not the last word in prosodic typology, the differences that it aims to capture are reflected in diachronic evolution. Non-decomposable tones such as those of Vietnamese, Thai and Mandarin undergo a gradual phonetic evolution – apart from tone mergers, which are categorical and irreversible. The evolution of level tones*, on the other hand, is punctuated by categorical changes. Under given circumstances, noncontrastive details in the realization of tone (i.e. conditioned allotonic variation) can be reinterpreted as differences between tonal categories; as a result, the phonological system is modified. For instance, a slight phonetic raising of a H tone preceding a L tone has been noted in various tone languages: taking the Gulmancema data in Table 2 as an example, the syllable /kan/ will be realised phonetically higher in the sequence /kan/ than in /kan/. This phonetic phenomenon does not affect the phonological nature of the tones. In the closely related language Moba, on the other hand, the super-high phonetic variant of the high tone has gained contrastive status – i.e. a lexical extra-high tone has emerged – following the loss of word-final vowels.

Table 2. A comparison showing the origin of the extra-high tone of Moba. Data and analysis from [55].

<table>
<thead>
<tr>
<th>meaning</th>
<th>Gulmancema</th>
<th>Moba</th>
</tr>
</thead>
<tbody>
<tr>
<td>he stepped over</td>
<td>13Ŭ 15ƙan 4di</td>
<td>13Ŭ 15ƙant</td>
</tr>
<tr>
<td></td>
<td>( ő kândi )</td>
<td>( ū kánt )</td>
</tr>
<tr>
<td>he steps over</td>
<td>13Ŭ 15ƙan 4di</td>
<td>13Ŭ 15ƙant</td>
</tr>
<tr>
<td></td>
<td>( ő kândi )</td>
<td>( ū kánt )</td>
</tr>
</tbody>
</table>

This is a classical case of transphonologisation (transfer of distinctiveness): from the tone of the word-final vowel to the one that precedes. The allotonic variation paves the way for the diachronic change, but the change itself (the modification of the tone system) is triggered by the loss of final vowels.
I have not come across reports showing that, in a *level-tone* system, the tones could undergo a gradual phonetic evolution away from their original values, such as a H tone gradually becoming falling (evolving towards H+L). By contrast, unitary contours appear to undergo gradual change, due to the continuous effect of phonetic factors. Unlike in the case of the change attested by the comparison of Gulmanacema with Moba, the evolution of unitary contours can take place without any conspicuous phonological change. These tones are defined in terms of an overall contour (as well as phonation-type characteristics in some cases), which can vary somewhat so long as the contrast among the tones present in the language is preserved. The Tamangic group of Sino-Tibetan is an especially well-documented example, revealing various evolutionary stages reflected in the spatial diversity of dialects, as well as a remarkable amount of cross-speaker differences within the same village, and even for one and the same speaker. Risiangku Tamang illustrates an early stage in the gradual phonetic evolution of complex tones*: the four tones of this language, breaking off their last ties with the earlier voicing correlation on initial consonants, become free to evolve away from their original fundamental frequency range (namely: relatively lower tones after former voiced initials, higher after former unvoiced initials). The evolution is more advanced in Marphali and in Taglung Tamang, where tone 4, which etymologically belongs in the low series, is now phonetically high; same with tone 3 in Manangke [56–58]. “Once it is constituted, the tonal system evolves without remembering its origins” [59].

These arguments drawn from dialectology are confirmed by phonetic evidence, in the case of tone systems for which there is a sufficient time depth in experimental studies. A well-described example is Bangkok Thai, which has been documented experimentally at intervals for a hundred years. For instance, the tone which in 1908 was the highest, with a final fall, has now become rising [60, fig. 2]. The number of distinctive tones has remained the same; their phonetic evolution is gradual, and the evolution of one tone has consequences on that of the other tones with which the risks of confusion are greatest. In this process, one sees at play the familiar antagonist forces of (i) the tendency towards simplification, on the one hand, and (ii) the pressure towards the preservation of distinctive oppositions, on the other [61].

From a synchronic point of view, there is no difficulty in proposing a level-tone analysis for any system, for instance labelling the five tones of Bangkok Thai as H, L, H+L, L+H and zero on the basis of an approximative stylization of fundamental frequency tracings [for an example of such analyses: 62]. But the linguist is then at a loss to describe the diachronic change mentioned above: how come tone 4 changed from high-falling to rising? This clearly looks like a continuous process – evidenced also by the study of different dialects, as mentioned above for Tamang. Here as in many other cases, autosegmental approaches provide precious insights for phonological modelling. (See e.g. a detailed argument concerning the limitations of a flatly synchronic description of the Cantonese tone system: [63].)

In view of a rather substantial literature on the analysis of the tones of Mandarin or Thai into *level tones*, it might seem presumptuous to say that the usefulness of these analyses has not been convincingly demonstrated. However, a review suggests that the conclusion about Vietnamese cited in Section 2.2 can be extended to Thai, Tamang, Mandarin and some other languages of Asia: there is no conclusive evidence to support an analysis into level tones for these languages.¹

### 4. Conclusion

The above review aimed to recapitulate salient synchronic and diachronic properties of some of the tone systems of Asia. The considerable progress made in the study of the synchronic properties and diachronic evolution of these systems warrants a (cautiously) optimistic conclusion: it does not seem unreasonable to hope that refined models of tone will be developed in the next few years, taking into account the diversity of the fascinating complex-tone systems of Asia.

Beyond the typology and modelling of lexical tone systems, this issue has a bearing on intonation studies. “Autosegmental-metrical” models of intonation borrow extensively from studies of Subsaharan tone systems [66]. A significant number of researchers, some of whom initially argued against the modelling of intonation into discrete levels [67], now advocate models whose basic tenets are familiar concepts of autosegmental tonology, such as *level tones*, downstep and tone spreading [witness the following textbooks: 68, 69]. Supposing, as the evidence reviewed here suggests, that there are lexical tone systems for which autosegmental models of tone do not tell the full story, new developments in the study of these tones could provide models and concepts for non-autosegmental approaches to intonation. The latter, which include the Kiel Intonation Model and its developments [70–73], superpositional approaches [74–81], and other approaches that have pre-generative roots [82–86], are currently outside the mainstream of intonation studies, in the same way as non-autosegmental analyses of tone systems fall outside mainstream (generative) phonology [87, 88]. Exchanges between researchers working within different frameworks is made difficult by differences in research goals, reflected in a different understanding of basic concepts. The present paper has achieved its goal if it has successfully conveyed some of the reasons why the generalization of representations based on *level tones* is problematic.

### 5. Appendix. Definition of terms

This Appendix provides explanations about a few central concepts that receive widely different definitions in different frameworks. “After all, the goal is to typologize linguistic properties, not linguists” [89].

**Contour tone.** This term refers to two distinct things. (i) In the autosegmental framework, it refers to a *tone* composed of two *level tones* (or more). For instance, a L+H sequence realized over a single syllable is referred to as a contour. (ii) In Pike’s two-way distinction between “register tones” (=level tones) and “contour tones”, it refers to a *tone* defined by a certain time course of pitch over the syllable [5].

**Level tone.** A *tone* defined by a discrete level of relative pitch. In a language distinguishing two levels, the following

¹ Even though the analysis of tones into levels and their analysis into features are two distinct issues, it may be useful to point out that attempts at decomposition based on two tonal levels plus two pitch register features, such as [64], do not meet the criteria against which phonological analyses are usually evaluated. An in-depth review concludes that these analyses suffer from “difficulties which make arguments for a register feature less than fully convincing” [65].
tones may be observed: H(igh), L(ow), H=L, and L+H. Level-tone systems have two to five levels. Systems with more than three levels are uncommon [Bariba: 90; Bench, a.k.a. Gimira: 91, 92]. One single case of six-level system has been reported: Chori [93], for which a reanalysis is possible [94]. Five may seem a low figure, in view of the number of musical notes that the ear can distinguish. However, in speech, where attention is not exclusively focused on the recognition of successive pitches, four- or five-level systems present a high degree of complexity.

**Phonation type.** Mode of vibration of the vocal folds during phonation. This is commonly referred to as voice quality: see the entry on phonation-type registers.

**Phonation-type register.** See Register; phonation-type register.

**Register; phonation-type register.** In languages with a phonation-type register system, phonation type has a lexically distinctive role. Thus, the Mon language has a ‘clear’ voice (also called ‘modal’ voice) register contrasting with a breathy/whispery voice register; this was still the case of Khmer less than a century ago [49]. Even more than other linguistic features, registers tend to have multiple correlates: mode of vibration of the vocal folds, but also greater duration of rhymes carrying nonmodal phonation, differences in vowel articulation, and differences in fundamental frequency. Instrumental studies of register systems include [95–104]. The term ‘phonation-type register’ is preferred over ‘voice quality’, as the latter is more general: “In a forthcoming book on general phonetics Professor David Abercrombie of Edinburgh accepts ‘register’ as an appropriate phonological term but suggests that ‘phonation-type’ is a more suitable term for its phonetic realization, thus reserving the expression ‘voice quality’ for more general use” [12].

**Tone; toneme.** The terms ‘toneme’ and ‘tone’ as used here are strictly equivalent. The word ‘toneme’ brings out the parallel with phoneme: like phonemes, tones are units that contrast with others within a system; they are defined as part of a paradigm. In this respect, tone is of a different nature from lexical stress, which has a specified position, i.e. a syntagmatic specification. It may be useful to add, in order to distinguish lexical tone from lexical phonation type registers, that fundamental frequency has a great importance in the phonetic realization of tone, unlike in the realization of phonation-type registers. It is unlikely, however, that a hard-and-fast line can be drawn between phonation-type register systems and tone systems.

6. Acknowledgments

Many thanks to Larry Hyman, Martine Mazaudon, Boyd Michailovsky and Jacqueline Vaissière for sustained exchanges. Special thanks to Michel Fertl as for sharing his insights into the tone systems of Tai-Kadai and Austroasiatic languages. Needless to say, none of them is to be held responsible for the views expressed here.

7. References
