Speech Prosody 2016
31 May - 3 Jun 2106, Boston, USA

Tones are not abstract autosegmentals

Fang Hu

Institute of Linguistics, Chinese Academy of Social Sciences
hufang@cass.org.cn

Abstract

This paper renewed the representation problem of tone in phonology theory. Accumulating evidence from phonetic researches shows that tones are not abstract autosegmentals. In contrast, stable temporal structures are found regarding tonal alignments to segmental productions. That is, tone gesture is coupled with consonant and vowel gestures in syllable production. This paper described new evidence from Lhasa Tibetan, which is a well-known language of emergent tones or tonogenesis. Discussion covers such tonal languages as Wu Chinese dialects, known for its tonal spreading phenomenon that motivates the autosegmental nature of tones.

Index Terms: tone representation, autosegmental, gestural coordination, Lhasa tones.

1. Introduction

Since [1], phonological theories adopted an autosegmental view on the relation between segmental and suprasegmental productions. Segments are vowels and consonants, which differ from each other in sound quality (timbre); suprasegmentals are basically prosodic features, which differ in pitch (tones), loudness (stress), or duration (length). And in phonological processes such as nasalization and vowel harmony, superimposed modifications to primary articulations or extracted features from segments were also treated as autosegmental. On one hand, the autosegmental view of tones emphasizes the independency of tonal behaviors. Physically, fundamental frequencies (F0) are laryngeal productions and are an independently controlled modification to supralaryngeal articulations, i.e., segments. Phonologically, tonal phenomena, such as floating tones, tonal spreading etc., are not conditioned by their segmental bases. In short, tones and segments are in parallel tiers and thus autonomous to each other. On the other hand, autosegmental phonology does not explain how exactly tones and segments are aligned together. Rather, they are associated by lines under Well-Formedness Condition in an abstract fashion. (See [2] pp. 65-104 for a review and description of late development.)

In the research paradigm of speech prosodies in general, F0 behaviors in languages have also been treated separately to segments in both modeling and empirical studies (e.g. [3], [4], [5]). However, the relationship between segments and tones is not abstract. In stead, phonetic researches reveal that temporal alignment between tones and segments demonstrates stable, concrete patterns both in tone languages ([6], [7], [8]) and in non-tonal languages ([9], [10]). The results corroborate the traditional view that tones are an integral part of syllable ([11]).

More recently, the inquiry into how tones are organized in syllables is renewed in the framework of Articulatory Phonology. According to Articulatory Phonology ([12], [13], [14]), each individual articulation, or 'gesture', is an action unit, which involves a formation and release of a particular constriction in vocal tract. Unlike traditional phonological concepts, which is claimed to be autonomous or linguistic internal, gestures in articulatory phonology follow general principles of kinematics and dynamics, and can thus be defined as a task dynamic system with temporal and spatial properties ([15], [16], [17]). Specifically, it is observed that speech gestures exhibit a stable timing relations if they are in-phase coupled, i.e. synchronized, or they are anti-phase coupled, i.e. sequential with each other ([13], [18], [19]). As proposed in [13], syllable production follows two general principles: (1) the prevocalic consonants are in-phase coupled, i.e. synchronized with the following vowel; (2) individual prevocalic consonants are anti-phase coupled, i.e. sequentially coordinated with each other. Thus, in a C_1C_2V syllable, for instance, the production of C_1 will be shifted earlier relative to the production of V and the production of C_2 be shifted later relative to the production of V due to the competing phasing specification between C-C coupling and C-V coupling ([18]). As schematized in Figure 1, the onset of the V gesture coincides roughly with the mid point of the two C gestures. The "C-center effect", therefore, serves as an important measure in characterizing syllable structure ([13], [18]).

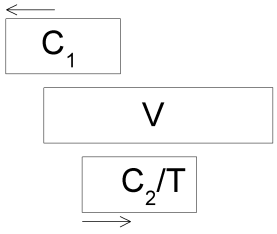


Figure 1: *The C-center organization of syllable production.*

Tone is not an exception; tone is a laryngeal gesture. Gao ([20], [21]) examined the gestural coordination in syllable production in Mandarin Chinese, and found that Mandarin CV syllables demonstrate a C-center-like organization among the C, V, and tone (T) gestures. That is, in Mandarin syllables, tone behaves as an additional consonantal gesture as C2 in Figure 1. In this connection, toned Mandarin CV syllables can be represented as CTV syllables, which are comparable to C_1C_2V syllables in languages with consonant clusters. The findings renewed the representation problem of tone. Tone is not an abstract autosegmental; it is a coupled gesture in syllable organization.

In Gao's work, contour tones are coupled elements of level ones. For instance, Mandarin tone 1 is composed of a H(igh) tone gesture; tone 2, the rising tone, a L(ow) tone gesture and a H tone gesture; tone 3 a L tone gesture; tone 4, the falling tone, a H tone gesture and a L tone gesture. And interestingly, Gao demonstrated that the LH gestures of the rising tone are in-phase coupled, that is, synchronized; whereas the HL gestures of the falling tone are anti-phase coupled, that is, sequentially produced. Following Gao's proposal, Hsieh pointed out the Mandarin tone 3 should be represented by anti-phase coupled LH tone gestures, which accounts for the

 doi: 10.21437/SpeechProsody.2016-62

dipping tone realization in citation form and the rising tone realization in tone sandhi ([22]). Yi and Tilsen examined the difference between the production of the Mandarin sandhi rising tone (i.e. the sandhi form of tone 3) and the citation rising tone (tone 2) and further elaborated the representation of tones in the framework of Articulatory Phonology ([23]).

This paper reports tones and syllable production in Lhasa Tibetan, and argues that the temporal organization of syllable organization in Lhasa supports the analysis of tones as an integral part of syllables. The gestural account of syllable production is of particular importance to Tibetan languages because it establishes the connection between non-tonal and tonal languages, and thus provides an explanation to the issue of the emergence of tones, or tonogenesis ([24]). Tibetan languages exhibit a variegated scenario of tonal developments: with the simplification of consonant clusters, Tibetan languages constitute a tonality continuum from completely non-tonal to highly tonal such that there is no clear dichotomy between a tonal and non-tonal language ([25], [26]). Lhasa Tibetan is a highly developed tone language; and its tonology is typical in the tonal development of Tibetan languages. It is generally agreed that Lhasa tones have a high vs. low contrast, whereas it remains controversial how many contour contrasts Lhasa Tibetan has (see [27] for a brief review). This is mainly attributable to the fact that tone contours are highly constrained by syllable types in Lhasa Tibetan.

2. Methodology

This paper examines the acoustics and articulation of Lhasa monosyllabic citation syllables and tones. Lhasa Tibetan has four syllable types; each can be associated with a high or a low tone. In order to have a better observation on the inter-gestural coordination in a natural conversational condition, meaningful monosyllabic words or morphemes with a labial initial consonant [p m] and a low or mid-low vowel [a ɛ] in all the eight possible syllable types and tone combinations were used as test syllables.

The test syllables, written in Tibetan script, were presented to the speaker in a random order on an LCD monitor. Each target syllable was embedded in a carrier frame:

X, ji ke? $_{\text{Tibetan Character}}$ ti_{this} X sa_{is} 'X, this is X'.

10 to 15 repetitions were recorded using the Carstens AG500 EMA system with a synchronized audio recording. Three native Lhasa female speakers were recorded. They were first or second year undergraduate students, 20 to 21 years old, in the Minzu University of China in Beijing.

The sensors were attached on the speaker's articulators along the midsagittal plane: two on the tongue (tip and body), two on the lips (lower and upper lips respectively), and one on the gum ridge at the lower incisors (jaw). Additional three sensors on the bridge of nose and behind the left and right ears served as references. The articulatory data were sampled at 200 Hz and smoothed with a 12 Hz low-pass filter. The acquired data were further corrected for head movements, and then rotated and translated to the speaker's occlusal plane.

The consonant gesture in the target syllable was characterized by lip aperture, i.e. the calculated Euclidean distance between the lower and upper lip sensors. The vowel gesture was characterized by the kinematics of the tongue body sensor. The tone gesture was based on the fundamental frequency (F0). Following Gao ([20], [21]), the preceding F0 minimum was taken as the onset of a high tone, and the preceding F0 maximum as the onset of a low tone.

Figure 2 illustrates the acoustic and articulatory data labeling procedures for the high toned syllable [mar]. The annotations consist of two acoustic levels, syllable and tone, and two articulatory levels, lip aperture (LA) and tongue body (TB). The label of syllable delimits the entire syllable, i.e. initial and rime. The label of tone delimits the rime part in a syllable. Thus, the interval between the onset of syllable and tone defines the consonant duration.

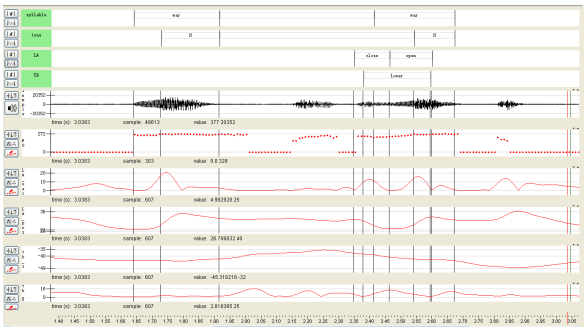


Figure 2. *Acoustic and articulatory labeling for [mar]. Levels of annotation: syllable, tone, lip aperture (LA), tongue body (TB); signal windows: audio, F0, LA velocity, LA position, TB position, TB velocity.*

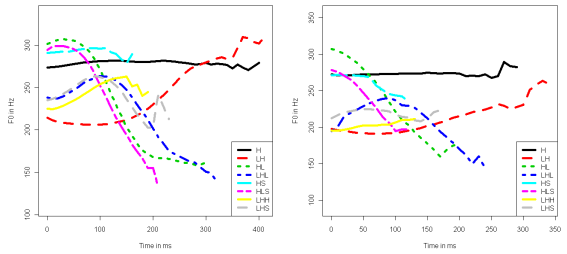
The target syllable in 'citation' position was labeled acoustically only and the discussion on the acoustic properties of Lhasa tones are based on these annotated 'tone' segments in 'citation' positions. Articulatory annotations apply to the target syllable in the sentence-mid position. The consonant gesture for the bilabial [p] or [m] was defined by lip aperture (LA), which is composed of a gesture of lip closing and lip opening. The production of the vowel [a] or [ɛ] was characterized by a lowering gesture of tongue body (TB). The annotations were based on the positional data with reference to the criterion of tangential velocity minimum. As shown in Figure 2, from the LA positional peak to its first valley was labeled as the gesture of lip closing (close), and accordingly from the valley to the following peak was labeled as the gesture of lip opening (open). And as shown in the figure, peaks and valleys occur where there are tangential velocity minima. Similarly, the lingual lowering gesture (lower) was labeled from a stable higher TB position to a stable lower TB position where there are tangential velocity minima. And the F0 minimum that precedes the target high tone, which is located around the offset of the preceding syllable [ti], was defined as the onset of the target high tone gesture.

3. Results

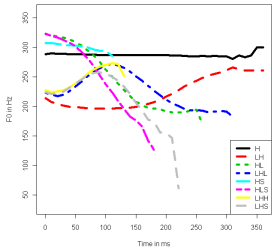
3.1. Acoustic results

Figure 3 gives the mean F0 contours associated with the eight different syllable type and tone combinations in Lhasa Tibetan from the three female speakers. After the examination of their variabilities, the F0 contours were averaged for each combination in the 'citation' position across all the repetitions of all the tested syllables (n = 2 test syllables ×10 to 15 repetitions). The F0 contour patterns were summarized in Table 1. The data is consistent across the three speakers. First, there is a clear high vs. low tonal contrast. Acoustically this feature is manifested on the onset part of the tone. The high

tones have a high F0 onset at around 270-320 Hz and the low tones a low F0 onset at around 190-240 Hz. Second, tonal contours are highly correlated with syllable types. It has been debated in the literature whether Lhasa has two, four, or six tones. It's quite clear from the acoustic data presented here that the complementary distribution of F0 contours leaves room for different phonological interpretations. A two-tone analysis emphasizes that the high-low contrast, or the so-called "register contrast", is the only phonological distinction for Lhasa tones ([28], [29]). A four-tone analysis further treats the difference in quantity or glottal stop as phonological contrast ([30]); a six-tone analysis treats both quantity and glottal stop differences as phonological ([31]).



Speaker 1 Speaker 2



Speaker 3

Figure 3. *Lhasa tones. High CVS: H; low CVS: LH; high CVh: HS; low CVh: LHH; high CV?: HLS; low CV?: LHS; high CVN?: HL; low CVN?: LHL.*

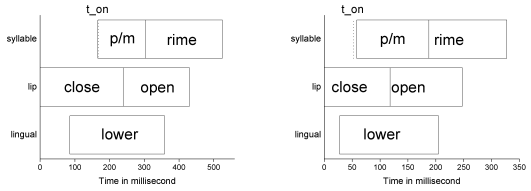
Table 1. *Syllable types and tonal melodies in Lhasa.*

tonal category	syllable type	F0 contour	label
high	CVS	long level	H
low		long rising	LH
high	CVh	short level	HS
low		short rising	LHH
high	CV?	short falling	HLS
low		short rising-falling	LHS
high	CVN?	short(?) falling	HL
low		short(?) rising-falling	LHL

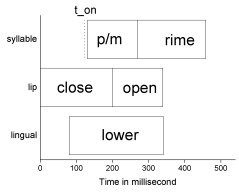
The results suggest that Lhasa has two phonological tones, each having four varieties of contour that is associated with syllable types. That is, the results support an eight-tone analysis. The emerging new tones on CVN? and CV? syllables differ significantly in length, although they share similar F0 contour patterns (see [32] for details).

3.2. Articulatory results

It has been shown so far that contrastive tones have emerged in Lhasa Tibetan. Meanwhile, the emergent tonal melodies are highly constrained by syllable structures, and are thus still under development from a historical phonological point of view. The articulatory study examines the temporal structure of intergestural coordination for the syllable production in Lhasa. Figure 4 summarizes the results from three speakers. Bars in the figures denote mean durations of gestural segmentations averaged across the repetitions and target syllables. The syllable tier denotes the acoustic landmarks of Lhasa syllable production: acoustically segmented [p]/[m] and rime, and the acoustically defined tone onset (t_on), which is represented by a dotted line in the figures. Lip gesture denotes the consonant gesture: a closing and an opening phase for the production of the bilabial consonant [p] or [m]. The lowering gesture of lingual articulation is taken as a measure of the vowel gesture for the low vowel [a] or mid-low vowel [ɛ].



Speaker 1 Speaker 2



Speaker 3

Figure 4. *Intergestural timing of Lhasa syllable production.*

As can be seen from the figure, the three speakers demonstrate a consistent pattern regarding the intergestural timing of syllable production. First, both lip gesture and lingual gesture begin before the p or m segment. That is, the articulation of both consonant and vowel begins before the acoustic onset of syllable. Second, the tone gesture begins around or shortly before the acoustic onset of syllable, too. Third, the lip opening gesture begins around the mid part of the p or m segment, and is accomplished around the mid part of the rime. Fourth, the tongue is lowered to its lowest position, i.e. the articulatory target for the vowel, during the first half part of the rime in general.

It can also be seen from the figure that the consonant gesture, vowel gesture, and tone gesture in Lhasa syllable production are generally consistent with the concept of C-center-like organization. That is, the vowel gesture begins around the midpoint between the consonant gesture and the tone gesture. To elaborate further if syllable types (CVS, CVN?, CV?, and CVh), initial consonants ([p m]) and tones (high and low) have an effect on the C-center-like alignment, the consonant-to-vowel (CV) lags and vowel-to-tone (VT) lags were further calculated for each speaker, and a three-way ANOVA with repeated measures was conducted on the difference between CV lag and VT lag. Results from Speaker 1 indicate no significant effect of syllable type ($F(3,320) = 0.6463$, $p = 0.5858$), initial consonant ($F(1,320) = 0.0272$, $p = 0.8692$), and tone ($F(1,320) = 0.0519$, $p = 0.8200$). And there is no significant effect of interactions between syllable type and initial consonant ($F(3,320) = 0.5314$, $p = 0.6611$), between syllable type and tone ($F(3,320) = 1.0751$, $p = 0.3598$),

between initial consonant and tone ($F(1,320) = 0.1604$, $p = 0.6890$), and between syllable type, initial consonant and tone ($F(2,320) = 1.2462$, $p = 0.2890$). Results from Speaker 2 indicate no significant effect of syllable type ($F(3,208) = 0.7506$, $p = 0.5231$), initial consonant ($F(1,208) = 1.0917$, $p = 0.2973$), and tone ($F(1,208) = 0.0855$, $p = 0.7703$). And there is no significant effect of interactions between syllable type and initial consonant ($F(3,208) = 1.6034$, $p = 0.1897$), between syllable type and tone ($F(3,208) = 0.2495$, $p = 0.8616$), and between initial consonant and tone ($F(1,208) = 0.1123$, $p = 0.7379$), but there is significant effect of interaction between syllable type, initial consonant and tone ($F(2,208) = 3.3392$, $p = 0.0374$). Results from Speaker 3 indicate significant effect of syllable type ($F(3,214) = 3.7871$, $p = 0.0112$) and initial consonant ($F(1,214) = 8.9015$, $p = 0.0032$), but no effect of tone ($F(1,214) = 1.4429$, $p = 0.2310$). And there is no significant effect of interactions between syllable type and initial consonant ($F(3,214) = 2.0283$, $p = 0.1109$), between syllable type and tone ($F(3,214) = 0.9170$, $p = 0.4335$), between initial consonant and tone ($F(1,214) = 0.0282$, $p = 0.8668$), and between syllable type, initial consonant and tone ($F(2,214) = 0.0339$, $p = 0.9667$). Therefore, suffice it to conclude that the tone doesn't affect intergestural timing, and it is generally held that the C-center-like alignment is consistent across different syllable types and initial consonants.

The Lhasa case corroborates the results from Mandarin Chinese ([20], [21]). That is, the tone gesture behaves like an additional consonant gesture in terms of its intergestural temporal relation to consonant and vowel gestures, as is schematically illustrated by the coupling structure of Lhasa syllable production in Figure 1. The fact that the tone is temporally aligned like an additional consonant in syllable production establishes a connection between tonal and non-tonal languages. Diachronically, the gestural account provides a direct observation into phonological evolution ([33]); synchronically, the gestural account makes the comparison of syllable complexity possible among tonally different languages.

4. Discussion and conclusions

The acoustic data confirmed the high-low contrast of tones in Lhasa on the one hand and a high correlation between tonal contours and syllable types on the other. The intergestural timing revealed a C-center organization for the Lhasa syllable production, namely the vowel gesture begins approximately at the midpoint between the consonant gesture and tone gesture. That is, the tone gesture is coordinated like an additional consonant to the CV production. The Lhasa case corroborates the results from Mandarin Chinese ([20], [21]), a canonical syllable tone language. The results suggest that tones are not abstract autosegmental, which are conceptually associated with segmental bases or other tonal bearing units (TBU). Rather, tones are coupled laryngeal gesture in syllable production.

In tone languages such as Mandarin Chinese, tones are lexical representations and are thus integrated gestures in the coupling relation of syllable production. However, sentential pitch accents in non-tonal languages such as Catalan and German demonstrate different F0 alignment events. Sentential pitch accents occur as a post-lexical event in non-tonal languages, and the alignment of the tone gesture doesn't affect the coordination structure of the consonant and vowel gestures ([34]). In Tibetan languages, the emergence of tone is a process of how global F0 events are localized as an integrated gesture in syllable production ([24]).

Tone spreading is a well-known tonal phenomenon in Wu Chinese dialects (e.g., [35]). Tone spreading demonstrates the stability of tones and their irrelevance to the corresponding segments, and thus strongly motivates the autosegmental treatment of tones. But alternatively, tone spreading could be viewed as an indication that tones are coupled gestures in prosodic words or other larger prosodic units than syllables. Duanmu reports that syllable length in Shanghai Wu is significant shorter than that in Mandarin in normal speech ([36], [37]). In other words, in tone spreading languages such as Shanghai Wu, syllables within a prosodic word are more temporally related to each other and thus serves as an integrated unit for gestural coordination. Very recently research from Suzhou, another northern Wu dialect with right spreading phenomenon of tones, shows that the spreading patterns are quite variable regarding a number of disyllabic lexical items among young adult speakers ([38]). For instance, the spreading tonal pattern [55-21] (in Chao's tonal letters [39]) is observed to have new disyllabic words that are originally belonged to other spreading patterns. Interestingly, they are not random phenomenon of lexical diffusion; rather, they are conditioned by citation tones on second syllables. More strikingly, the same tonal pattern [55-21] demonstrates different temporal structures on different disyllabic items. The syllable ratio of the original [55-21] disyllabic words is 1.07, whereas that of the new items is 0.84 ([38]). That is, the first syllable is longer in well-established disyllabic words, as the second syllable is toneless and thus acquires a spreading tone from left according to the autosegmental account. But the second syllable is longer in new items, since the citation tones on the second syllable have an impact on gestural organization in the production of disyllabic prosodic words, which is beyond the autosegmental theory of tones but can be well explained in gestural theory.

To conclude, accumulating evidence shows that tones are not abstract autosegmentals, and gestural approaches have been revealing more details regarding tone production in particular and the production of syllable or other larger speech unit in general (c.f. [40]). It should be admitted, however, that more articulatory studies on different languages with diversities of tonal behaviors are required to further explore the nature of the production of tones.

5. Acknowledgements

This work is supported by CASS Innovation Project "Language universals and diversities: an experimental approach to the phonetics and grammar in Chinese dialects" (P.I.: Dr. HU Fang) and Chinese National Social Science Foundation (Project number: 15BYY073; P.I.: Dr. HU Fang).

6. References

[1] Goldsmith, J. A. *Autosegmental phonology*, Ph.D. dissertation, MIT, 1976.

[2] Yip, M. *Tone*, Cambridge, U.K.: Cambridge University Press, 2002.

[3] Fujisaki, H., Wang, C., Ohno, S. and Gu, W. "Analysis and synthesis of fundamental frequency contours of Standard Chinese using the command–response model", *Speech Communication*, 47, 59-70, 2005.

[4] Tao, J., Kang, Y. and Li, A. "Prosody conversion from neutral speech to emotional speech", *IEEE Transactions on Audio, Speech, and Language Processing*, 14(4), 1145-1154, 2006.

[5] Tseng, C., Pin, S., Lee, Y., Wang, H. and Chen, Y. "Fluent speech prosody: framework and modeling", *Speech Communication*, 46, 284-309, 2005.

[6] Xu, Y. "Consistency of tone-syllable alignment across different syllable structures and speaking rates", *Phonetica*, 55, 179-203, 1998.

[7] Xu, Y. "Effects of tone and focus on the formation and alignment of f0 contours", *Journal of Phonetics*, 27, 55–105, 1999.

[8] Xu, Y. "Speech melody as articulatorily implemented communicative functions", *Speech Communication*, 46, 220-251, 2005.

[9] D'Imperio, M., Loevenbruck, H., Menezes, C., Nguyen, N. & Welby, P. "Are tones aligned to articulatory events? Evidence from Italian and French", In Cole, Jennifer and Hualde, José Ignacio (eds.), *Laboratory Phonology*, 9, 577-608. Berlin, New York: Mouton de Gruyter, 2007.

[10] Mücke, D., Grice, M. Becker, J. & Hermes, A. "Sources of variation in tonal alignment: evidence from acoustic and kinematic data", *Journal of Phonetics*, 37, 321–338, 2009.

[11] Chao, Y.-R. *A Grammar of Spoken Chinese*, Berkeley, CA: University of California Press, 1968.

[12] Browman, C. P. and Goldstein, L. "Towards an articulatory phonology", *Phonology Yearbook*, 3, 219-252, 1986.

[13] Browman, C. P. and Goldstein, L. "Some notes on syllable structure in Articulatory Phonology", *Phonetica*, 45, 140-155, 1988.

[14] Browman, C. P. and Goldstein, L. "Articulatory Phonology: An overview", *Phonetica*, 49, 155-180, 1992.

[15] Saltzman, E. "Task dynamic coordination of the speech articulators: a preliminary model", in H. Heuer and C. Fromm (eds.) *Generation and Modulation of Action Patterns*, Berlin: Springer-Verlag, pp. 129–144, 1986.

[16] Saltzman, E., & Kelso, J. A. S. "Skilled actions: A task dynamic approach", *Psychological Review*, 94, 84-106, 1987.

[17] Saltzman, E. L. & Munhall, K. G. "A dynamical approach to gestural patterning in speech production". *Ecological Psychology*, 1(4), 333-382, 1989.

[18] Browman, C. P. and Goldstein, L. "Competing constraints on intergestural coordination and self-organization of phonological structures". *Bulletin de la Communication Parle*, 5, 25-34, 2000.

[19] Goldstein, L., Byrd, D. and Saltzman, E. "The role of vocal tract gestural action units in understanding the evolution of phonology", In Michael A. Arbib (ed.) *Action to language via the mirror neuron system*, pp. 215-249, Cambridge: Cambridge University Press, 2006.

[20] Gao, M. *Mandarin Tones: An Articulatory Phonology Account*, Ph. D. Dissertation, Yale Uiversity, 2008.

[21] Gao, M. "Gestural coordination among vowel, consonant and tone gestures in Mandarin Chinese", *Journal of Chinese Phonetics*, 2, 43-50.

[22] Hsieh, F.-Y. "A gestural account of Mandarin tone 3 variation", In *Proceedings of ICPhS XVII, Hong Kong*, 2011.

[23] Yi, H. and Tilsen, S. "Gestural timing in Mandarin tone sandhi", *Proceedings of Meetings on Acoustics*, 22, 060003, 2015.

[24] Hu, F. Tonogenesis in Lhasa Tibetan – Towards a gestural account. In Hoole, P. et al. (eds.) *Consonant Clusters and Structural Complexity*, pp. 231-254. De Gruyter, 2012.

[25] Huang, B. "Conditions for tonogenesis and tone split in Tibetan dialects [in Chinese]", *Minzu Yuwen*, 3, 1-9, 1994. English translation by Jackson T.-S. Sun in *Linguistics of the Tibeto-Burman Area* 18: 43-62, 1995.

[26] Sun, Jackson T.-S. "Variegated tonal developments in Tibetan", In Bradley, David, Randy LaPolla, Boyd Michailovsky and Graham Thurgood (eds.) *Language variation: papers on variation and change in the Sinosphere and in the Indosphere in honour of James A. Matisoff*, pp. 35-51, 2003.

[27] Qu, A. "Tibetan tone and its historical development [in Chinese]", *Yuyan Yanjiu*, 1, 177-194, 1981.

[28] Sprigg, R. K. "Verbal phrases in Lhasa Tibetan", *Bulletin of the School of Oriental and African Studies*, 16, 1954.

[29] Sprigg, R. K. "Tone in Tamang and Tibetan, and the advantage of keeping register-based tone systems separate from contour-based systems", *Linguistics of the Tibeto-Burman Area*, 13.1: 33-56, 1990.

[30] Sun, Jackson T.-S. "The typology of tone in Tibetan", *Chinese Languages and Linguistics IV: Typological Studies of Languages in China* (Symposium Series of the Institute of History and Philology, Academia Sinica, Number 2), 485-521. Taipei: Academia Sinica, 1997.

[31] Hu, T., Qu, A. and Lin, L. "Experimental studies on Lhasa Tibetan tones [in Chinese]", *Yuyan Yanjiu*, 2, 18-38, 1982.

[32] Hu, F. and Xiong, Z. "Lhasa tones", In *Proceedings of Speech Prosody 2010*, 100163: 1-4, Chicago, 2010.

[33] Goldstein, L., Byrd, D. and Saltzman, E. "The role of vocal tract gestural action units in understanding the evolution of phonology", In Michael A. Arbib (ed.) *Action to language via the mirror neuron system*, pp. 215-249. Cambridge: Cambridge University Press, 2006.

[34] Mücke, D., Nam, H., Hermes, A., and Goldstein, L. "Coupling of tone and constriction gestures in pitch accents", In Hoole, P. et al. (eds.) *Consonant Clusters and Structural Complexity*, pp. 205-230. De Gruyter, 2012.

[35] Kennedy, G. "Two tone patterns in Tangsic", *Language*, 29, 367-373, 1953.

[36] Duanmu, S. "Rime length, stress, and association domains", *Journal of East Asian Linguistics*, 2, 1-44, 1993.

[37] Duanmu, S. "Syllabic weight and syllabic duration: a correlation between phonology and phonetics", *Phonology*, 11, 1-24, 1994.

[38] Wang, F. "New developments of disyllabic tone sandhi of unchecked yang-tones in Suzhou [in Chinese]", paper presented at *the Second Workshop on Experimental Dialectology*, Nanjing, 2015.

[39] Chao, Y.-R. "A system of tone-letters", *Le Maître Phonétique*, 45, 24–27, 1930.

[40] Karlin, R. "The articulatory TBU: gestural coordination of lexical tone in Thai", *Cornell Working Papers in Phonetics and Phonology*, pp. 124-149, 2014.