Proceedings of the Language Technologies for All (LT4All) , pages 43–46
Paris, UNESCO Headquarters, 5-6 December, 2019.

Handling Prosody and Tone Languages

Aijun LI, Wei WANG
Institute of Linguistics, Chinese Academy of Social Sciences, Beijing, China
No.5 Jianguomennei Dajie, Beijing, 100732, P.R.C.
liaj@cass.org.cn, waywon@qq.com

Abstract

Technology has played a vital role in advancing our understanding of prosody in human languages. In many languages, tones distinguish lexical meanings and variations of other prosodic features carry semantic or pragmatic information. Although deep learning and end-to-end technologies have been increasingly used in speech applications, challenges in technology for handling prosody and tone languages are still significant, especially in less familiar languages and dialects. In this squib, we will present as a case in point the study of tone, intonation and other prosodic features in Chinese and its relevance to speech technology.

Keywords: tone, intonation, prosody, prosodic modelling

Résumé

技术在我们对了解人类不同语言中起了非常关键的作用，语言学家采用科学仪器研究声音的声学、发音和感知特性，包括韵律和声调特性。声调可以区分词义，韵律特征的变化可以传递言语的语义和语用信息。对言语技术来说，尽管已经很容易采用深度学习和端到端技术来构建语音应用系统，面对智能语言系统来说，特别是对一些低资源的语言或者方言的研究，仍存在很多挑战，需要加强对韵律和声调语言的研究。本文将重点以声调语言汉语为例，围绕言语技术面临的这些挑战，介绍声调、语调和韵律的研究成果。

关键词: 声调，语调，韵律，韵律建模

1. Introduction

Tones can be defined as pitch variations that change either the lexical or grammatical meaning of a word. A language in which the meaning of a word depends on its tone is known as a tone language. The aspects of speech that extend beyond individual vowels and consonants are known as suprasegmentals or prosody. Narrowly speaking, prosody is sometimes used as a synonym only for intonation, which refers to the use of suprasegmental features to convey post-lexical or sentence-level pragmatic meanings in a linguistically structured way (Ladd, 1996).

Tone languages can have either register tones or contour tones. The vast majority of languages spoken in Africa are register tone languages, such as Igbo, Shona, Yoruba, and Zulu, to name just a few. In a contour tone language, pitch movement, instead of pitch level, serves to distinguish word meaning. Many of the languages spoken in Southeast Asia – including Mandarin Chinese, Chinese dialects, Thai, and Vietnamese – are contour tone languages.

Speech prosody is a systemic structure of various linguistic components in an utterance or multiple connected utterances, which conveys linguistic, paralinguistic and even nonlinguistic information via suprasegmental features (Fujusaki, 1997). As a result, prosody is not only related to tone, intonation, stress and rhythm, but also to discourse-information, as well as interlocutor's emotions and attitudes.

Deep learning and end-to-end technology are widely used in speech application systems. However, challenges are plentiful for technology in handling prosody and tone languages in real-life intellengent systems, to wit, how to describe complex tonal systems phonetically and phonologically; how to model tone and intonation from a typological point of view; how to model contextual prosody in interaction; and how to apply technology in prosodic annotations for speech corpora with different accents automatically or semi-manually.

To this end, in this squib I will focus on presenting as a case in point the study of tone, intonation and other prosodic features in Chinese, including phonetic and phonological descriptions of tones, tone sandhi and tonal coarticulation, typological study on intonation, intonation modeling and prosodic annotations.

2. Phonetic and Phonological Despcriptions of Tones

In the first comprehensive explication of the sound system in Chinese found in a Chinese rhyme dictionary called "Qieyun", created in 601 during the Sui Dynasty (581–618), a total of 12,000 character entries were arranged corresponding to the four tonal categories, referred to as "Ping", "Shang", "Qu", and "Ru" in the philological tradition. A new phase was initiated in the experimental study of tone when the kymograph was used in the measurement of tones in Chinese dialects by Liu Fu, Wang Li and Yuen-Ren Chao. Chao was also the inventor of "tone letters" used in International Phonetic Alphabets (IPA) for the transcription of tones, in which the total pitch range is divided into four equal parts, marked by five pitch levels, numbered 1, 2, 3, 4, 5, with 1 being the lowest pitch and 5 the highest (Chao, 1930).

Presently, there are ten major dialect groups of Chinese spoken in China (Maps of the languages in China, 2012), which exhibit great variations in phonology, syntax and other aspects of grammar.

Southern dialects generally have more complex tonal systems than northern dialects in terms of the number of tones and contour shapes. While most southern dialects have between 6 and 13 tones, northern dialects have about 2 to 5 tones. Mandarin dialects are not known for having too many tones, except those spoken in the lower reaches

of Yangtze River. Most of them have 5 to 7 tones, due to partly influences from the neighboring Wu dialects. On the other end of the spectrum, the Lanyin variety of Mandarin, which is spoken in the western part of China, has 2 to 4 tones, fewer than most other Mandarin dialects due to its direct contact with non-tonal minority languages over a long period of time. According to Ran (2018), the dialect with the most tones is a Gan dialect called the Jinxian dialect, which has 13 tones; and the dialect with the fewest tones is the Honggu dialect, with only 2 tones.

One area focused on the analysis of acoustic features such as f0, duration and amplitude of tones in isolation and in a sequence (e.g. Lin and Yan, 1992; Wu, 1982). An interesting development in recent years has been the study of phonation types such as modal voicing, breathy voice or creaky voice. Evidence was presented to show that lexical meanings can be distinguished on the basis of both pitch and phonation types, not just pitch alone as previously assumed for tone languages like Chinese (Kong, 2001; Zhu, 2012).

Phonological systems to represent tones have been proposed for Mandarin and other Chinese dialects. Two earliest studies (Wang, 1967; Wu, 1979) both distinguished register and contour features of tones. The earliest study concerning the representation of tones is Wang (1967). He distinguished three level features and a redundant contour feature. For example, a tone can be in the high or low register of pitch and it can take the shape of a level or contour pitch movement such as rising or falling. The dichotomy of register and contour is widely adopted in more formal treatment of representation of tones (Bao, 1990; Chen, 2000; Duanmu, 1990; Li, 2003; Yip, 1989, 2000). In a comprehensive study of tonal systems, Liu (2004) surveyed 1,186 Chinese dialects from published documentations. Tones in her study are transcribed with the tone letters system of Chao. She proposed register and contour feature to classify tone systems of Chinese dialects and accounted for contour variations using downtrend of pitch.

Zhu (2012) departed from previous systems of tonal features and proposed a new notational system, based on his own field work. According to him, in Chinese dialects some tones are not only cued by pitch movement, but also by phonation types. Therefore, phonation types and pitch movement are both relevant in defining three pitch registers and six pitch levels. He proposed a framework of tonal typology in which each tone can be defined by four features which are register, duration, level, and contour.

3. Tones in Context: Tone Sandhi and Tonal Coarticulation

Changes can happen as a result of two tones in juxtaposition. Tone sandhi is generally known as the process of tone change due to the influence of the other. The best-known tone sandhi in Mandarin is the third tone sandhi: in the combination of two syllables in Tone 3, the first syllable changes to Tone 2. In other words, Tone 3 + Tone 3 become Tone 2 + Tone 3. It is a macro change of tonal identity and is easily noticeable by fluent speakers. Tonal coarticulation, on the other hand, refers to much more subtle changes, caused by interactions of different articulatory gestures. It does not involve changes of tonal categories.

Major issues in the study of tone sandhi include the relationship between tone in isolation and in tone sandhi, the domain in which tone sandhi happens and the types of tone sandhi. Tone sandhi patterns can be classified in terms of direction, manifestation and domain. For example, the Shanghai dialect and other Northern Wu dialects exhibit the left-dominant tone sandhi pattern in which the left-most syllable in the tone sandhi domain – lexical compounds in this case – keeps its underlying tone and spreads its tone to the other syllables in the domain. Phrasal structures follow different patterns, though. The right-dominant pattern is illustrated in the third tone sandhi in Mandarin above: The Tone 3 in the right-most syllable triggers the tone sandhi in the previous syllable. A third pattern is called tonal substitution. An often-cited example is the southern Min dialect spoken in Xiamen (also known as Amoy). The right-most syllable keeps its tone and the other syllables in the same domain are replaced with other tones in a circular way, forming what is known as the Min circle (Chen, 1987). Chen (2000) provides in-depth analyses of tone sandhi patterns across Chinese dialects. It has been so far the most comprehensive study of tone sandhi in Chinese dialects.

Determination of tone sandhi domain has been challenging as it involves syntax, rhythm and tempo. Take the third tone sandhi in Mandarin as an example. The tone sandhi patterns can be quite complex when three or more third tones are present in the same domain and across two domains. As mentioned above in passing, compounds and phrasal structures in Wu dialects follow different set of tone sandhi rules.

Formal analyses of tone sandhi are often given in tonal features such as H (high) and L (low). Different theories have been proposed for tone sandhi patterns in different dialects, but there are still remaining issues that have yet to be fully accounted for.

Tonal coarticulation happens between adjacent tones, but it does not cause changes of tonal categories. Effects of tone sandhi in Mandarin have received intensive attention since the 1990s (Peng, 1997; Shen, 1990; Shih, 1986; Xu 1994, 1997, 1999). Both carryover and anticipatory effects have been identified in the production of connected tones. For example, in Mandarin when the Tone 3, a low tone, is preceded by the other three tones and itself, the pitch trajectory of the low tone is greatly affected by the preceding tones. Specifically, its high point is the highest at the syllable boundary when preceded by a high level tone. The carryover effect in Mandarin is primarily assimilatory. The anticipatory effect, much smaller in scale than the carryover effect, is seen in the raising effect of the low tone on the preceding tone. For example, the high f0 (fundamental frequency) of the falling tone in Tone 4 is realized higher after Tone 3 than the other three tones, ergo the effect is mainly dissimilatory.

In addition to four tones in Mandarin, there is a special category called the neutral tone. A syllable is said to be in neutral tone when it does not carry one of the four lexical tones in Mandarin. Neutral tone has received ample attention in research, but its status in the Mandarin phonological system has been controversial as to whether the neutral tone is a tonal category, or it is related to stress. According to Chao (1979), neutral tone is related to weak stress. When a syllable is in weak stress, its tonal range is almost reduced to zero and its duration significantly

shortened. Lin (1957, 1962) used the term "weak stress" to capture the metrical property of syllables in neutral tone. Lu and Wang (2005) distinguished "neutral tone" from "weak stress", reserving the former as neutralized tone in the tonal system and the latter as unstressed syllable in the metrical system. Their view resonates with Chao's. Li (2017) approached the neutral tone by considering the effects of prosodic boundary and information structure in the acoustic analysis, and examined its pitch and durational properties in different prosodic contexts.

4. Typological Study on Intonation

After the emergence of the prosodic or intonational annotation system such as INTSINT (Hirst, 1998) and ToBI (Silverman et al. 1992), a number of studies have been conducted within these frameworks, in which the view of local typology is widely assumed. Intonational typology does not merely compare intonation structures of different languages, but is also relevant to theoretical issues, among which question intonation and the interplay between intonation and lexical prosody such as tone are most discussed. The difficult development of intonational typology is due to the complexity of intonation as well as the lack of a widely accepted annotation system. Moreover, the researchers still do not reach a consensus as to the parameters of cross-linguistic comparisons of intonation. Jun (2005, 2014) present analyses of intonation in 27 languages, including Mandarin, and describe how intonation is constructed out of tones or accents that are tied to stressed syllables and syllables at the end of a prosodic domain, following the autosegmental-metrical (AM) approach to intonation (Ladd, 1996, 2008).

The interplay of tone and intonation in Chinese has been most intriguing in intonational phonology. Since the seminal work of Yuen-Ren Chao (Chao, 1932, 1933), a rich body of research, both descriptive and experimental, has been produced to advance our understanding of the linguistic functions and physical properties of tone and intonation in Mandarin Chinese, especially regarding the interaction of tone and intonation. Chao came up with two metaphors that have been widely known in characterizing how tone and intonation interact: the "rubber band effect" and the "small wave and big wave" theory. According to the latter theory, tone and intonation are related in the form of superimposition – either successive or simultaneous – just like small waves sitting on top of big waves.

The study of intonation in Chinese dialects is still in its infancy, but has shown great potential in making meantinful contributiongs to the typological study of intonation. More comprehensive studies other than Mandarin can only be found in Cantonese (Fox et al., 2008). The study of Cantonese intonation started early in 1970s, when Vance investigated the tone and intonation in Cantonese (Vance, 1973). Other studies in the early days include those of the Chengdu dialect (Chang, 1958), Toishan (Lee, 1986), Cantonese (Johnson, 1986) and Changsha (Shen, 1991) dialects.

The study of Cantonese treated different aspects of intonation: the focus structure (Man, 2002), the effect of question intonation on lexical tone (Lee, 2004; Ge, 2018), the interaction between sentence final particles and intonation (Wu, 2008), and the modeling of Cantonese intonation using the command-response model (Gu, 2004). The most notable property of Cantonese intonation is the rising question intonation, as it has been demonstrated in several studies on acoustics (Lee, 2004; Ge, 2018) and perception (Mai, 2000). Another noteworthy phenomenon is its rich inventory of sentence final particles (SFPs). As is shown by Wu (2008), the pitch manifestations of SFPs combine the effects of intonation and lexical tones.

There are a few of studies of intonation in other Chinese dialects lately, such as the Tianjin dialect (Zhang, 2018), and the Kaifeng dialect (Wang, 2018). Several other studies also looked at the focus structure in other dialects, such as Teochew (Hsu et al., 2018) and the Shanghai dialect (Ling and Liang, 2017). The focus structure of some Shandong dialects has also been investigated by Jia and colleagues (Jia, 2011; Duanmu et al.2013; Duanmu and Jia, 2015). As in Mandarin, PFC is found prevalently in many Chinese dialects, except Cantonese. Cantonese distinguishes nine lexical tones (including three checked tones), and it does not show the PFC effect. Another dialect that has been shown to lack this phenomenon is Taiwanese (Chen et al., 2009). More work is in need in this area.

5. Tone and Intonation: Modeling and Beyond

Following the footsteps of Chao, many were engaged in developing models of intonation in Mandarin to account for the interaction of tone and intonation. Wu (2004), who inherited and expanded Chao's theory, proposed the "transposition model" of intonation, which accounts for obligatory and optional tone sandhi patterns in Chinese. Shen (1992, 1994) characterized intonation in terms of the upper line and lower line of f0 that define a pitch register and argued that the two lines can be manipulated independently of each other in different intonation patterns. Xu (2004) proposed the Parallel Encoding and Target Approximation (PENTA) model of speech prosody, which is a framework for conceptually and computationally linking communicative meanings to fine-grained prosodic details, based on an articulatory-functional view of speech.

In a monograph on the experimental study of tone and intonation in Chinese, Lin (2012) took Chao's insights as a point of departure and explicitly adopted the autosegmental-metrical (AM) model of intonation. In his model, focal prominence and boundary tone are the two key elements in describing intonation in Chinese. For example, the difference between declarative intonation in statements and interrogative intonation in questions without sentence-final question particles resides in the boundary tone, which is realized acoustically as pitch register and slope of the contour. Shi (2013) looked at intonation from a broader perspective and proposed a systematic method to define an "intonation pattern" with three parameters – f0 contours, pause-lengthening ratio and sound intensity. In his study of declarative and interrogative intonations in Putonghua Cantonese and Korean, Shi was trying to figure out cross-linguistic patterns in intonation in terms of the quantifiable measurements of f0, duration and intensity.

Focus was probably given the most attention in the study of intonation. Xu (1999) looked into the effects of tone and focus on the alignment of F0 contours and found that focus exerts influence on pitch range in different ways: the pitch range of the syllables before the focal position remains

unmodulated, and that of the syllables is dramatically expanded in the focal position and compressed after focus, a phenomenon he termed "post-focus compression" (PFC). Other studies analyzed phonetic realizations of different types of focus and situations in which there are one, two or multiple foci in the utterance (Jia and Li, 2012). Wang and Xu (2011) reported an experimental investigation of the prosodic encoding of topic and focus in Mandarin by examining disyllabic subject nouns elicited in four discourse contexts. Their major findings were that focus causes post-focus f0 lowering while topic allows a gradual f0 drop afterwards; the effects of downstep, sentence length on initial f0 are independent of topic and focus, and the effects of topic, focus, downstep and sentence length are largely cumulative.

In addition to focus, prosodic structure has been another closely-examined area in the study of intonation. Tseng (1999) proposed a HPG model and modelling the tone and intonation under the frame of discourse prosody.

The link between intonation and emotion was explored as early as Bolinger (1989). Experimental studies have flourished on the influence of emotions on intonation patters in recent years (e.g. Bänziger and Scherer, 2005). Li (2015) undertook an extensive study on the role of intonation in conveying emotion in a tone language like Chinese, with focus on f0 levels and pitch contours in what she termed "successive addition boundary tone". She proposed that the boundary tone is composed of two components – the base tone of the syllable and an addition contour.

Intention understanding and generating in human-machine interactions calls for greater integration of discourse-level prosodic information in spoken dialogue systems. In a series of studies on the interface of prosody and discourse, Li A., Jia and their collaborators conducted detailed analyses on prosodic features in connection with discourse structure, information structure and dialogue acts (Jia, 2018; Li, 2018; Li et al., 2019).

Prosodic information is widely used for the detection of disfluencies and utterance boundaries, the segmentation of dialogue acts, the detection of sentence mood and modality, accent and so on. Prosodic annotation system, which provides a tool to highlight significant prosodic events and which is essential to statistical prosody modeling in these tasks. The prosody annotation systems based on ToBI framework and its modifications, such as C-ToBI (Li, 2002). and KToB, are most popular. Since manual prosodic annotation, is generally time-consuming and expensive to administer. It is important to develop automatic annotation of prosodic information. Many algorithms or models were proposed: CRF and HMM for annotation of Japanese accent types and phrase boundaries (Koriyamay et al., 2014), unsupervised joint prosody labeling and modeling by Chiang et al. (2009) for read speech and spontaneous Mandarin speech by Lin et al. (2016), and transfer learning and RNN-based model for L2 prosodic annotation and evaluation (Lee, 2019; Chen, 2019)

6. Concluding Remarks

Technology for handling Prosody and tone has now become more interdisciplinary and better integrated with other disciplines than ever before. Looking forward, we expect that availability of advanced research instruments is adopted to explore speech production and perception mechanisms at the neurological level, to direct more attention and resources to cross-linguistic and cross-dialectal typological and applicational research in order to better serve diverse linguistic and dialectal groups, and to conduct studies on contextual tonal and prosodic variations to meet the demands of speech and language technology. Finally, technology in speech and language technology will continue to serve as an invaluable tool in our joint efforts to document and preserve endangered languages and dialects in order to strengthen linguistic diversity.

7. Acknowledgements

This research is supported by the Key NSSFC Granting (No.15ZDB103), the National Key R&D Program of China (No. 2017YFE0111900).

8. Selected References

Bao, Z. (1990). *On the nature of tone*. Doctoral dissertation, MIT.
Chao, Y. R. (1933). Tone and Intonation in Chinese. In *Linguistic Essays by Yuenren Chao*, Beijing: The Commercial Press.
Chao, Y. R. (1934). The Non-uniqueness of Phonemic Solutions of Phonetic System, in *Linguistics Essay by Yuenren Chao*, Beijing: The Commercial Press.
Chen, M. Y. (2000). *Tone Sandhi: Patterns across Chinese Dialects*. Cambridge: Cambridge University Press.
Chinese Academy of Social Sciences. (2012). *Atlas of Languages in China*. Beijing: The Commercial Press.
Duanmu, S. (1990). *A formal study of syllable, tone, stress and domain in Chinese languages*. Doctoral dissertation, MIT.
Hirst, D., ed. (1998). *Intonation Systems: A Survey of Twenty Languages*. Cambridge: Cambridge University Press.
Jun, S. (2005). *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford, New York: Oxford University Press.
Jun, S. (2014). *Prosodic Typology II*. Oxford, New York: Oxford University Press.Li, A. (2002). Chinese prosody and prosodic labeling of spontaneous speech. In *Proceedings of Speech Prosody*.
Li, A. (2015). *Encoding and Decoding of Emotional Speech: A Cross-Cultural and Multimodal Study between Chinese and Japanese (Prosody, Phonology and Phonetics)* 1st ed. Springer.
Li, Z. (2003). *The phonetics and phonology of tone mapping in a constraint-based approach*. Doctoral dissertation, MIT.
Lin, M. (2012). *The experimental study of Chinese intonation*. Beijing: China Social Sciences Press.
Peng, S. H. (1997). Production and perception of Taiwanese tones in different tonal and prosodic contexts. *Journal of Phonetics*, 25(3): 371-400.
Potisuk, S., Gandour, J., and Harper, M. P. (1997). Contextual variations in trisyllabic sequences of Thai tones. *Phonetica*, 54(1): 22-42.
Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., ... and Hirschberg, J. (1992). ToBI: A standard for labelling English prosody. In *Proceedings of ICSLP*.
Shen, X. S. (1990). Tonal coarticulation in Mandarin. *Journal of Phonetics*, 18(2): 281-295.
Shih, C. L. (1986). *The Prosodic Domain of Tone Sandhi in Chinese*. Doctoral dissertation, University of California, San Diego.
Wu, Z. (2004). *Linguistics Essay*. Beijing: The Commercial Press.
Xu, Y. (1994). Production and perception of coarticulated tones. *Journal of the Acoustical Society of America*, 95(4): 2240-2253.
Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, 25(1): 61-83.
Xu, Y. (1999). Effects of tone and focus on the formation and alignment of F0 contours. Journal of Phonetics, 27(1): 55-105.
Xu, Y. (2004). Transmitting tone and intonation simultaneously - the parallel encoding and target approximation (PENTA) model. *Proceedings of International Symposium on Tonal Aspects of Languages*.
Yip, M. (1989). Contour tones. *Phonology*, 6(1):149-174.
Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.