Tone and intonation processing: from ambiguous acoustic signal to linguistic representation

Liu, M.

Summary

Spoken language processing is a task that humans continuously perform from birth. It, however, is not always easy, as speech signals are inherently noisy and ambiguous. Prosody, which structures speech, is a determinant of the form of spoken language. Ambiguity in prosody therefore can be a very important source of speech ambiguity. Yet very little is known about how ambiguity in prosodic information affects spoken language processing. Not uncommonly, the same form of a speech signal can represent different prosodic information and cause speech ambiguity. This dissertation concerns how ambiguous acoustic signals representing different prosodic information affect spoken language processing.

The most prominent prosodic feature of tonal languages such as Standard Chinese is their use of pitch (i.e., tone) to distinguish lexical meanings. However, speech ambiguity arises in Standard Chinese because the same pitch contour can cue both tone and intonation. Previous studies have shown that the dual functions of pitch in Standard Chinese cause ambiguity in speech signals and result in pitch processing difficulties at the behavioral level. However, what are the underlying neural mechanisms leading to the eventual behavioral decisions of tone and intonation processing? How do native listeners resolve the pitch processing difficulties? These issues are less well-understood and further research on tone and intonation processing is needed.

The same or similar pitch contours can also cue the same linguistic representation (e.g., tone), but different categories of that representation in two linguistic systems (e.g., different tonal categories in two tonal systems) of the same speaker. Indeed, most Standard Chinese speakers also speak a local Chinese dialect. Across some Mandarin dialects and Standard Chinese, the same or similar pitch contour(s) is often used to characterize two different tonal categories, which in turn result in different lexical meanings. For example, the high-level pitch contour with a pitch value of 55 corresponds to a different tonal category in Xi'an Mandarin (a Mandarin dialect) than in Standard Chinese. In the former, the syllable *ma55* means *scold*. In the latter, it means *mother*. For bi-dialectal speakers of such Mandarin dialects and Standard Chinese, the

question arising is whether the same or similar pitch contours from the two tonal systems are taken as representations of the same tone in pitch processing. Furthermore, what role does tone play in the activation and processing of bi-dialectal lexical representation?

This dissertation therefore set out to investigate how pitch is processed within a linguistic system (i.e., Standard Chinese) and across two linguistic systems (i.e., Standard Chinese and Xi'an Mandarin) when the same pitch contour cues different linguistic functions (i.e., tone and intonation) or different categories of the same linguistic function (i.e., tone). The dissertation is composed of six chapters.

Chapter 1 introduced the research questions to be addressed and provided a brief overview of each chapter.

Chapter 2 tapped into the neural correlates of tone and intonation processing in Standard Chinese using the event-related potential (ERP) technique. Native Standard Chinese listeners were presented with semantically neutral Standard Chinese sentences, which contrast in final tones (rising T2 or falling T4) and intonations (Question or Statement). Their behavioral and electrophysiological responses were recorded. A clear P300 effect was observed for the question-statement contrast in sentences ending with T4, but no ERP effect was found for the question-statement contrast in sentences ending with T2. These results provide ERP evidence for the interaction of tone and intonation in Standard Chinese, confirming the findings from behavioral metalinguistic data that native Standard Chinese listeners can distinguish between question and statement intonation when the intonation is associated with a final T4, but fail to do so when the intonation is associated with a final T2. Chapter 2 extends our understanding of online processing of tone and intonation 1) from the pre-attentive stage to the attentive stage; and 2) within a larger domain (i.e., multi-word utterances) than a single word utterance.

Chapter 3 further investigated the role of semantic context in resolving the pitch processing difficulties in tone and intonation processing in Standard Chinese reported in Chapter 2. Tone and intonation identification experiments were conducted in both semantically neutral and constraining contexts with the same group of native speakers of Standard Chinese. Results showed that the

overall performance of tone identification was better than that of intonation. Tone identification was seldom affected by intonation information, irrespective of semantic contexts. Intonation identification, particularly question intonation, however, was susceptible to the final lexical tone identity and was greatly affected by the semantic context. Specifically, in the semantically neutral context, questions were difficult to identify, as evidenced in the lower response accuracy and longer reaction time, regardless of the lexical tone identity. In the semantically constraining context, both intonations took significantly less time to be identified than in the semantically neutral context. Moreover, questions ending with a falling tone were more accurately identified than questions ending with a rising tone. These results suggest that top-down information provided by the semantically constraining context can play a facilitating role for listeners to disentangle intonational information from tonal information, especially in sentences with a lexical falling tone in final position. Chapter 3 provides strong evidence for the role of semantic context in resolving pitch processing difficulties in Standard Chinese, particularly from the reaction time patterns, which have not been reported in earlier studies. The results reported in Chapter 3 also resolved the puzzle of the reversed patterns of question intonation identification in sentences ending with T2 and T4 in normal natural context versus in low-pass filtered context. It was found that the stronger and more informative the linguistic context is (semantically constraining context > semantically neutral context > low-pass filtered context), the better the identification of questions ending with T4. The opposite pattern holds for questions ending with T2.

One long-neglected fact in linguistic research on Standard Chinese is that most speakers of Standard Chinese also speak a local dialect, which may share phonological features with Standard Chinese. Among these dialects, Xi'an Mandarin is particularly interesting for the seemingly simple, yet intricate mappings between its lexical tones with those in Standard Chinese. **Chapter 4** empirically compared the tonal systems of Xi'an Mandarin and Standard Chinese. Tones with similar contours from Xi'an Mandarin and Standard Chinese were paired. Both tone production and perception experiments were carried out on highly proficient bi-dialectal speakers of Xi'an Mandarin and

Standard Chinese. The two experiments together showed that there are indeed systematic mappings of tones between Xi'an Mandarin and Standard Chinese. The degree of the similarity of the mapped tone pair in tone perception was largely dependent on the acoustic phonetic similarity between the tones in tone production, with a phonological rule playing a role in the tone pair of low contour. Chapter 4 compared the two systems in tone production with a more balanced design compared to the previous acoustic study. Moreover, it provides new empirical evidence for the mapping of the two tonal systems from a perceptual point of view. It also allows for an investigation of the relationship between tone production and perception in bi-dialectal tonal language speakers.

Chapter 5 further examined the effects of cross-dialect phonological similarity in segment and tone on bi-dialectal lexical access in spoken word recognition. Balanced bi-dialectal speakers of Xi'an Mandarin and Standard Chinese took part in an auditory-auditory priming experiment with a generalized lexical decision task. The primes were monosyllabic homophones from either Xi'an Mandarin or Standard Chinese, while the targets were disyllabic Xi'an Mandarin or Standard Chinese words. Primes and the first syllable of the target words had five configurations. They either overlapped in both segment and tone within a dialect (identical) or across two dialects (interdialectal homophones), or they overlapped in segment only within a dialect or across two dialects. The baseline condition was that they overlapped neither in tone nor segment within a dialect. Results showed that Standard Chinese primes did not yield significant priming effects for within- or cross-dialect segment-only overlap targets. Standard Chinese primes did not produce significant priming effects for within-dialect identical targets either. However, they did yield significant inhibitory priming effects for cross-dialect homophone targets. This overall pattern was reversed for Xi'an Mandarin primes because these primes were not treated differently from their interdialectal homophonous primes in the current mixed dialect setting. These results suggest that cross-dialect phonological similarity in segment alone did not affect lexical access in bi-dialectal auditory word recognition, while cross-dialect phonological similarity in both segment and tone posed a threat to the recognition system of bi-dialectal listeners. It is clear that tonal information

played a significant role in constraining word activation in bi-dialectal auditory word recognition. The results reported in Chapter 5 extends our understanding of the role of segment and tone in auditory word recognition in tonal languages from the monolingual context to the bi-dialectal context, and reveals a non-selective processing mechanism in bi-dialectal lexical access during auditory word recognition, as has been demonstrated for bilingual lexical access.

Chapter 6 recapped the research questions and summarized the main findings of this dissertation. The implications for future research were also discussed in this chapter.

In summary, this dissertation has demonstrated that pitch processing in Standard Chinese is subject to both within- and cross-linguistic influences. The ambiguous acoustic signals due to dual functions of the F0 channel in signalling tone and intonation in Standard Chinese cause pitch processing difficulty at the sentential level. This pitch processing difficulty has a neural correlate and can be resolved via top-down information provided by a constraining semantic context. Acoustic ambiguities in Standard Chinese can also arise from a closely related Chinese dialect that shares tonal similarities with Standard Chinese, here Xi'an Mandarin. The cross-dialect tonal similarities affect tone processing and further interfere in lexical access during spoken word recognition in bi-dialectal tonal language speakers. Together, this dissertation revealed two of the most prominent pitch processing difficulties tonal language speakers encounter from within and across languages, advancing our current understanding of pitch processing from various aspects.