Production and perception of tones by Dutch learners of Mandarin
Zou, T.; Zou T.

Citation

Version:	Not Applicable (or Unknown)
License:	[Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)
Downloaded from:	[https://hdl.handle.net/1887/52977](#)

Cover Page

The handle http://hdl.handle.net/1887/52977 holds various files of this Leiden University dissertation.

Author: Zou, T.
Title: Production and perception of tones by Dutch learners of Mandarin
Issue Date: 2017-9-28

Chapter one

Background

1.1 Introduction

Pitch movements (cued mainly via fundamental frequency or f0 acoustically) have different functions across languages. For non-tone language speakers, pitch information is mainly used at the post-lexical level to convey linguistic and paralinguistic meanings. Linguistically, it can be used to indicate prominence at the word and sentence levels (Birch & Clifton, 2002; Welby, 2003; Xu & Xu, 2005), delimit prosodic constituents (such as intonation phrase, utterance, paragraph and so on) (Cole, 2015; Cutler, Dahan, & Van Donselaar, 1997; Snedeker & Trueswell, 2002), and mark sentence types, such as statements and interrogatives (Pierrehumbert & Steele, 1989; Van Heuven & Haan, 2002). Some paralinguistic information, such as emotion (anger, surprise, joy, fear, and so on) and attitude (politeness, uncertainty, irony, dejection, and so on) can also be (partly) encoded with pitch movements (Chen, Gussenhoven, & Rietveld, 2004; Chen, 2005; Luthy, 1983; also see Shattuck-Hufnagel & Turk, 1996 for a detailed review).

Tone-language speakers, on the other hand, primarily employ pitch information to convey lexical meanings, while at the same time, in a much more complex and sometimes subtle way, signal various post-lexical information comparable to that in non-tone languages (e.g., Chen, 2000, 2012; Chen & Gussenhoven, 2008; Cole, 2015; Gussenhoven, 2004; Xu, 2001; Yip, 2002).[1]

Mandarin is a tone language. Fundamental frequency (f0) has been demonstrated as the primary acoustic correlate of tones (Howie, 1976; Xu & Wang, 2001; Yip, 2002), although other acoustic parameters (e.g., intensity and temporal properties such as duration and position of pitch turning point in some contour tones) can also be used to mark tonal contrast (Hallé, 1994; Moore & Jongman, 1997; Xu, 2009).[2] In Mandarin Chinese, tonal information is an integral part of a word and the meaning of the segments is associated with the pitch contour superimposed on them. Mandarin Chinese has four main tones, in addition to a neutral tone. Tone 1 is a high-level tone; Tone 2 is a mid-rising tone; Tone 3 is a low tone; Tone 4 is a high-falling tone. When produced in prepausal position or in isolation, Tone 3 is realized with a dipping contour.

---

[1] For some African and Asian languages, tone can be used to signal grammatical information. In this thesis, "tone language" refers to languages in which tones are solely used to convey lexical meaning.

[2] Mandarin is used here to refer to Standard Chinese, the official language spoken in Mainland China, which is based on Beijing Mandarin. Only when referring to other Mandarin dialects, will we identify the specific dialect within the Mandarin dialect family. "Mandarin speakers", without mentioning a specific Mandarin dialect, then, is used to refer to speakers of Standard Chinese who may or may not speak a Mandarin dialect other than Standard Chinese.

This tone also has two variants in connected speech: it becomes a low falling tone preceding Tone 1, Tone 2, Tone 4 and neutral tone, and is realized with a rising contour similar to Tone 2 preceding another Tone 3. The four full tones are demonstrated in Figure 1.1. The neutral tone always comes at the end of a word or phrase, associated with a weak syllable. It has a static and mid target, but the target is realized with more pitch variation compared with lexical full tones: the pitch of a syllable with neutral tone is substantially influenced by the tone in the preceding syllable (Chen & Xu, 2006).
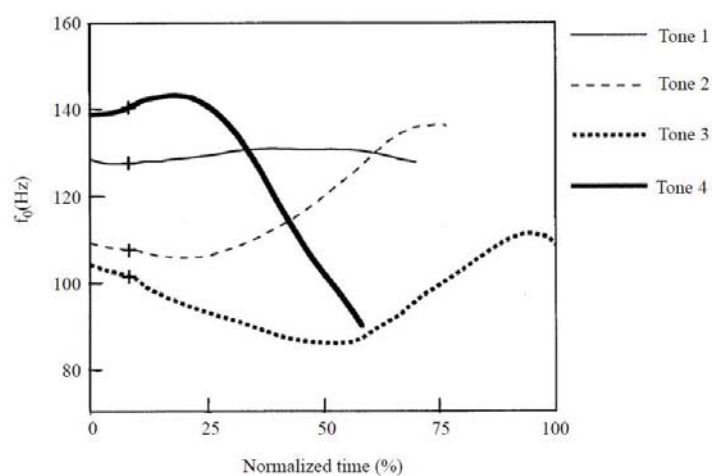

Figure 1.1. *Mean f0 contours of the four Mandarin tones in the monosyllable /ma/ produced in isolation (Xu, 1997). Averaged over 48 tokens produced by eight male native Mandarin speakers.*

When produced in isolation, different tones are realized with stable and distinctive pitch contours. However, when produced in context, tones can be influenced by adjacent tones and undergo substantial acoustic variation, leading to coarticulated tonal realization which are different from the canonical contours. As for the perception of tones, prior studies show a high level of interdependency in the processing of segmental and tonal dimensions by native Mandarin speakers (Choi, Tong, Gu, Tong, & Wong, 2017; Lin & Francis, 2014; Repp & Lin, 1990; Tong, Francis, & Gandour, 2008). That is, the segmental and tonal dimensions are integral and processed simultaneously by Mandarin native speakers.

With regard to perceptual identification and discrimination of lexical tones, previous research showed that tones can be perceived in a categorical fashion by native speakers (Francis & Ciocca, 2003; Hallé, Chang, & Best, 2004). In terms of the role of tone in word recognition, some previous studies suggest that tone serves as a weaker cue compared to segmental information. However, recent studies using online measures such as eye-tracking and event-related potentials (ERP) show parallel processing of segments and tones in word recognition, arguing that the role of tonal information is comparable to that of segmental information. For example, Schirmer,

Tang, Penney, Gunter, and Chen (2005) used ERPs to investigate the role of tone and segmental information in Cantonese word processing. Comparing the ERPs elicited by semantically congruous words and by tonally and segmentally induced semantic violations, they found that both segments and tones were accessed at a similar point in time and elicited an N400-like negativity. Malins and Joanisse (2012) offered further support for the comparable roles of segments and tones, showing that both segmental and tonal information could be accessed and used as soon as they become available during word processing. Taken together, the existing literature suggests that tonal information is exploited in spoken word recognition. It plays an early constraining role in lexical activation, and words with non-matching tone would not be activated as candidates. This effect can be captured and revealed more readily in online measurements with tasks more similar to real communication situations.

Speakers of tone and non-tone languages have been reported to tune their auditory systems to the same acoustic stimuli differentially due to their first language (L1) experience. Both behavioral and neuroscientific studies have suggested that speakers of tone and non-tone languages process pitch information differently.

Using multidimensional scaling, Gandour (1983) investigated the influence of different language backgrounds on tonal perception. Speakers of tone languages (Cantonese, Mandarin, Taiwanese, Thai) and a non-tone language (English) were asked to judge the dissimilarity of paired tones. The result showed that listeners from a non-tone language (English) attached more importance to pitch height and gave less weight to pitch direction than did listeners from most of the tone language group. It is suggested that the absence of lexical contrastive tones in monosyllabic words can account for the low saliency of the direction dimension in English listeners' dissimilarity judgments. Bent, Bradlow, and Wright (2006) investigated the influence of long-term linguistic experience on identifying non-speech rising, falling and flat pitches. The results showed that the rising and falling stimuli were treated in the same way by English listeners, but Mandarin listeners more often misidentified flat and falling pitch contours than the English listeners, in a manner that could be related to specific features of pitch contours of Mandarin lexical tones. The authors argued that, in Mandarin, the pitch range of the falling tone (Tone 4) is larger than that of the rising tone (Tone 2), so that Mandarin listeners might have a different criterion for the distinction between falling and rising contours and were more reluctant to label stimuli as falling than to label them as rising. Thus, it appears that listeners' perception of pitch movements can be shaped by the way pitch information is used in their native language. Neurophysiological studies also showed differences in the hemispheric specialization of pitch processing by tone and non-tone language speakers: tonal contrasts are processed mainly in the left hemisphere by tone-language speakers, but in the right hemisphere or bilaterally by non-tone language speakers (Gandour, Dzemidzic, Wong, Lowe, Tong, & Hsieh, 2003; Gandour, Tong, Wong, Talavage, Dzemidzic, & Xu, 2004; Krishnan, Xu, Gandour, & Cariani, 2005; Zatorre & Gandour, 2008).

Since non-tone language speakers are not familiar with pitch information which conveys lexical meaning, tones can present a great difficulty to them. Such difficulty in tone production and perception experienced by beginning learners of Mandarin has been tested in a number of studies. In terms of tone production, Wang, Jongman, and Sereno (2003) tested tone production in monosyllables by beginning learners of Mandarin and showed that only 57 percent of the learners' tone productions was correctly identified by native Mandarin listeners. After a short-term training, 78

percent of their tone production can be correctly identified. This shows that beginning learners can be trained to improve their production accuracy in monosyllables. Other studies also showed that tones in monosyllabic words can be produced correctly by second language (L2) learners of Mandarin (e.g., Hao, 2012).

As for the perception of tones, Wang, Spence, Jongman and Sereno (1999) showed that, in a tone identification task, beginning English learners of Mandarin showed an identification accuracy rate of 69 percent with a prevailing Tone 2-Tone 3 confusion. This study also showed that learners could be trained to significantly improve their tone production and identification accuracy. After a two-week perceptual training, English learners of Mandarin improved their Mandarin tone identification accuracy by 18 percentage points.

Taken together, lexical tone plays an important role in Mandarin. Tones and segments are processed in an integral manner by native speakers and the role of tones in spoken word recognition is comparable to that of segments. For non-tonal L2 learners, establishing new tonal categories can be challenging since they do not make extensive use of suprasegmental features in lexically contrastive way in their L1. Previous studies demonstrated that beginning L2 learners of Mandarin can correctly produce tones in monosyllables and perceive tones with high accuracy in simple identification and discrimination tasks. These findings have presented a promising start, but the process of tone acquisition is more demanding. To produce natural and native-like words, learners need to learn coarticulation patterns of tones in addition to the canonical tonal contours. To process tonal information effectively, learners are also required to attach more perceptual weight to the previously ignored suprasegmental dimension. Furthermore, they need to learn to attune to the most reliable phonological tonal information for word form identification in connected speech. Last but not least, effective exploitation of tones is of great importance in spoken word recognition. Therefore, the crucial research questions are: (1) to what extent can advanced learners achieve these goals; and (2) what mechanisms underlie the development of their L2 tone acquisition? Few studies in the existing literature have addressed these issues. To fill this research gap, this dissertation examines the production and perception of Mandarin tones by both beginning and advanced Dutch learners of Mandarin, with native Mandarin speakers as a control group. Specifically, four experimental studies are reported in this dissertation, viz. on L2 tonal coarticulation patterns in disyllabic tone production (Chapter 2), attention redistribution in L2 tone processing (Chapter 3), L2 tone processing at the phonological level (Chapter 4), and the role of tones in lexical access for L2 learners (Chapter 5). The following sections provide the backgrounds of these topics.

1.2   Tonal coarticulation by native Mandarin speakers and L2 learners

Coarticulation refers to the influence of one sound on a neighboring sound in speech production, since speech is produced as "a sequence of sounds flow to articulatory movements" and "there is 'blurring of the edges' of segmental articulations as the vocal tract moves from one articulation configuration to the next" (Bell-Berti, Krakow, Gelfer, & Boyce, 1995). A speech sound is influenced by both the preceding sound (i.e., carryover effect) and the subsequent sound (i.e., anticipatory effect). Bidirectional coarticulatory effects have been reported in vowels and consonants, e.g., carryover

effect as shown in Recasens (1984) and Beddor, Harnsberger, & Lindemann (2002); and anticipatory effects as shown in Martin & Bunnell (1981) and Grosvald (2009). Whalen's (1990) findings demonstrate that the carryover effect is a result of physiological constraints in realizing some motor program, since this effect remains robust when cognitive planning is constrained. The anticipatory effect, on the other hand, would be a reflection of speech planning, in that it decreases when the participants' planning mechanism is inhibited.

Previous research focusing on coarticulation patterns in Asian tone languages shows that tonal coarticulation is also bidirectional, which is in parallel with the cases of vowel and consonant coarticulation. The carryover effect is generally strong in terms of the magnitude and temporal domain, while the anticipatory effect is generally weaker (e.g., Thai: Abramson, 1979; Gandour, Potisuk, & Dechongkit, 1994; Vietnamese: Brunelle, 2009; Han & Kim, 1974). The patterns of Mandarin tonal coarticulation generally agree with the findings from other tone languages. Xu (1997) examined tonal coarticulatory patterns in disyllabic non-words /mama/ with all possible tonal combinations in Standard Chinese produced by native Beijing Mandarin speakers. The results showed that the carryover effect exhibits an assimilatory nature; a high offset of the first tone can raise the onset of the following tone, while a low offset lowers the onset of the following tone. The anticipatory effect, however, is largely dissimilatory. Xu (1997) showed that, in Standard Chinese, the anticipatory effect is mainly on the maximum f0 of the preceding tone.

In terms of L2 production of Mandarin tones, considerable evidence indicates that L2 learners are able to correctly produce lexical tones in isolation (e.g., Hao, 2012; Wang, Jongman, & Sereno, 2003). Producing tones in connected speech, however, does present a great challenge, evident in the higher error rates and decreased intelligibility of L2 speech (Hao, 2012; Shen, 1989; Sun, 1998; Yang, 2011, 2016). However, the acquisition of fine-grained tonal coarticulation patterns has received less research attention. Recently, Brengelmann, Cangemi and Grice (2015) tested tonal coarticulation in disyllabic sequences by German learners of Mandarin and found much f0 variation in the last 20 percent of the tone contours on the first syllable, which suggested a strong but non-native-like anticipatory effect. The extent to which the pattern is general among learners of non-tonal languages is an interesting issue to investigate.

As reviewed earlier, tonal coarticulation has been mainly investigated in three aspects: the directionality (carryover or anticipatory), the nature (assimilatory or dissimilatory), as well as the magnitude and temporal extent of the effects. Thus far, the underlying mechanism and source of the tonal coarticulatory effect for native speakers, as well as L2 acquisition of tonal coarticulation have been under-investigated. Therefore, Chapter 2 sets out to investigate these issues using a disyllabic tone production task with a high cognitive load. By testing native speakers, we aim to shed further light on the mechanisms underlying tonal coarticulation. By recruiting both beginners and advanced Dutch learners of Mandarin, we will investigate the developmental trajectory and mechanisms of tonal coarticulation that underlie the ultimate attainment of acquisition in a tone language by non-tonal L2 learners.

1.3   Attention redistribution and segment-tone integration in L2 acquisition

When learning a foreign language, learners are often confronted with difficulties in both low-level auditory processing and in the phonological processing of non-native segmental and suprasegmental contrasts. Different theoretical models have been proposed to account for such difficulties. The Speech Learning Model (SLM) holds that L2 learners perceive non-native sounds by referring to the phonetic categories of their L1 sound system (Flege, 1995). The mechanisms involved in L1 acquisition, such as category formation, remain intact through one's life and can also be used in L2 learning, although this ability tends to decrease as the learner's age of learning increases. PAM-L2 (Best & Tyler, 2007), based on the Perceptual Assimilation Model (PAM) (Best, 1994), assumes that a listener's perceptual system will automatically assimilate non-native speech sounds to the nearest categories in the L1 sound system, and the discrimination of non-native contrasts can be predicted from the way in which they are assimilated. For the case of L2 acquisition of Mandarin tones, both SLM and PAM-L2 suggest that a novel L2 speech contrast can potentially be acquired by learners.

While these models of L2 acquisition have focused on whether new L2 categories can be acquired, much less has been investigated on how they are acquired. As discussed in the previous section, past research has shown that the same pitch movements can be attended to differentially by tone and non-tone language speakers. Braun and Johnson (2011) showed that Mandarin speakers were attentive to the rising and falling pitch contours on both the initial and final syllables in a disyllabic non-word. These contours signal two different lexical tones in Mandarin. Dutch speakers, in contrast, were much more sensitive to pitch movements in the final position than in the initial position, possibly because a Dutch final pitch movement serves as a salient cue for non-lexical meanings, such as question vs. statement.

Moreover, prior studies suggest that the processing of segmental and tonal dimensions by native Mandarin speakers is more interdependent than by non-tone language speakers. The segment-tone integration has been revealed in some studies testing the so-called Garner interference. That is, there is an increase in reaction time due to the inclusion of irrelevant information during perceptual processing (Garner, 2014). For example, Tong, Francis and Gandour (2008) tested the interactions between segmental and suprasegmental dimensions of Mandarin Chinese by asking participants to attend to one dimension while ignoring the other. Their results suggested that variations in the segmental dimension interfered more with tone classification than *vice versa*. While in non-tone languages, like English, the two dimensions are much less integrated, and therefore listeners are able to tune their attention to only one dimension and suppress interference from the other (Lin & Francis, 2014). This integrality of tones and segments in tone languages such as Mandarin Chinese and Cantonese has also been found in recent neuroscientific studies (Choi et al., 2017; Gao, Hu, Gong, Chen, Kendrick, & Yao, 2012; Tong, McBride, Lee et al., 2014). For example, Choi et al. (2017) tested perceptual integration of vowels and tones in native Cantonese speakers using the passive oddball paradigm. Tone-MMN, vowel-MMN and double-MMN were elicited. The results showed that double-MMNs were significantly smaller in amplitude than the sum of single feature MMNs, suggesting the perceptual integration of tones and vowels at the phonological level.

Therefore, the issues we address here are: during the course of their acquiring a tonal system, can Dutch learners of Mandarin learn to redistribute their attention to segmental and tonal information like native speakers and can they develop a more integral processing of these two dimensions? Chapter 3 investigates these questions by

examining how beginners and advanced Dutch learners of Mandarin process tonal information in an ABX matching-to-sample task, compared to both native Mandarin speakers and native Dutch speakers without any tone language experience.

1.4 Phonological processing of tone contrasts by L2 learners

The automatic selective perception (ASP) model (Strange, 2011), which has been developed to characterize L1 and L2 speech perception, highlights the role of attention in language acquisition. It further differentiates between a phonological mode and a phonetic mode of perception. The phonological mode is employed by native listeners, in which automatic selective perception routines are used to detect phonologically contrastive information for identifying word forms. This automatic processing is shaped by language experience, and therefore costs little cognitive effort. The phonetic mode is employed by native speakers to detect fine-grained allophonic details, and requires more cognitive effort. It is hypothesized that at the beginning stage of L2 learning, the phonetic mode of perception has to be used when processing novel contrasts. The L2 learning process involves the development of new selective perception routines that optimize the attunement to information that is reliable for word-form recognition. The role of the task is also emphasized by the ASP model: in tasks with a high memory load and phonetic variability, L2 listeners are less likely to detect fine-grained phonetic details, and therefore have to use the phonological mode of processing; in less demanding tasks with simple stimuli, the phonetic mode can be used.

The problem of Japanese listeners' discrimination of the English /r/-/l/ contrast (Strange & Dittmann, 1984) is a good example of acquisition difficulty that can be accounted for by the ASP model. The L2 listeners showed a good performance in basic identification and discrimination tasks in which the phonetic mode of processing could be used. In a more demanding task with complex stimuli asking for the phonological mode of processing their performance was poor, since the selective perception routines of English had not been established yet. Likewise, for perception of a non-native /e/-/ɛ/ contrast as predicted by the ASP model, the level of difficulty is a function of task and stimulus factors (Pallier, Bosch, & Sebastián-Gallés, 1997; Sebastián-Gallés & Soto-Faraco, 1999; Sebastián-Gallés, Echeverria, & Bosch, 2005). These findings also suggest the important role of task demands and stimulus complexity in the assessment of participants' processing ability of non-native segmental and suprasegmental contrasts at the phonological level.

Furthermore, learning to use tonal information in lexical access is another crucial issue in L2 tonal acquisition. However, to our knowledge, no systematic empirical research has been done to investigate tone processing in lexical access by L2 learners. To examine the developmental trajectory of the Dutch learners' phonological processing of tonal contrasts and the use of lexical tones in lexical activation, Chapter 4 adopts a cognitively demanding sequence recall task and a lexical decision task, testing both beginning and advanced Dutch learners of Mandarin and native Mandarin speakers as a control group.

1.5   Segmental versus tonal information in lexical access by L2 learners

As the studies investigating real-time spoken word recognition accumulate, it is becoming increasingly clear that as the input unfolds, lexical candidates are activated immediately with receipt of a minimal amount of acoustic information (McMurray, Clayards, Tanenhaus, & Aslin, 2008); the activation is updated incrementally (Dahan & Gaskell, 2007; Shen, Deutsch, & Rayner, 2013); multiple words are activated in parallel and compete with each other during the recognition process.

To capture these characteristics of spoken word recognition, current models (such as TRACE: McClelland & Elman, 1986 and Shortlist: Norris, 1994; Norris & McQueen, 2008) assume that, as speech input unfolds, the incoming sound can be mapped onto phonemic and lexical representations in the mental lexicon, and then a set of lexical candidates compete for recognition. Since these models were developed using non-tone languages, they do not encode lexical tones. As suggested by recent online studies on lexical tone processing (e.g., Malins & Joanisse, 2010), it is increasingly clear that tone plays a comparable role as segments in constraining lexical activation. Tones have been incorporated into the TRACE model in a recent simulation of monosyllabic spoken word recognition of Mandarin Chinese (Shuai & Malins, 2017), based on the finding and suggestions from previous studies (Malins & Joanisse, 2010; Ye & Connine, 1999; Zhao, Guo, Zhou, & Shu, 2011).

There has also been an abundance of studies that have tested the perception and production in beginning Mandarin L2 learners. For example, Wang et al. (1999) show that English learners of Mandarin improved their tone identification accuracy in monosyllabic words from 69% to 90% after a two-week training. The training-induced improvement also generalized to new words and speakers. In addition to tones in isolated syllables, the perception of longer stimuli also has been tested. Hao (2012) found that both English and Cantonese learners of Mandarin performed better in monosyllabic tonal identification than in disyllabic identification. Both learner groups showed better Mandarin tone mimicry than tone identification and reading. The former task only involved low-level auditory perception and articulation while the latter task required a more abstract representation of tones. This suggests that the main difficulty in tone learning is the establishment of robust associations between pitch contours and tone categories.

More recently, learning to use lexical tone information in word recognition by naive non-native speakers of Mandarin have been tested in several training studies. The sound-to-word learning paradigm, which trains participants to associate minimal tone pairs with different meanings, has been employed in these studies to examine the contribution of individual variability in cue weighting in tone learning (Chandrasekaran, Sampath, & Wong, 2010), the effect of individual musical experience (Wong & Perrachione, 2007), as well as the influence of tonal context in tone learning (Chang & Bowles, 2015). Some studies also found training-induced changes in the participants' neural system (Wong, Chandrasekaran, Garibaldi, & Wong, 2011; Wong, Perrachione, & Parrish, 2007). Although the focus varied across these studies, the convergent result is that naive non-native speakers of Mandarin can be trained to use pitch information lexically.

While much research effort has been devoted to learning lexical tones by naive non-native Mandarin speakers and beginning learners, the processing of tones and segments by advanced L2 learners of Mandarin and the developmental trajectory have

not been studied before. Moreover, L2 processing of tonal information has not been investigated using on-line methods. Therefore, Chapter 5 sets out to examine the role of tones and segments in auditory spoken word recognition using the Visual World Paradigm by monitoring the eye movements of both beginners and advanced Dutch learners of Mandarin. Native Mandarin speakers were also tested, as a control group.