INTRODUCTION

Mandarin Chinese is a tone language that contrasts four lexical tones, with Tone 1 having high-level pitch, Tone 2 high-rising pitch, Tone 3 low-dipping pitch, and Tone 4 high-falling pitch. A neutral tone, or zero tone, that is phonetically unstressed and much weaker in intensity and shorter in duration also exists in Mandarin. The phonetic features of tones are physically realized by the F0 (fundamental frequency) values and contours. For native Mandarin speakers, the primary cue for tone contrasts is F0 contour (Xu, 1997; Liu & Samuel, 2004). Therefore, native Mandarin listeners attach more importance to “contour” than “height” dimensions of tones (Wang, Jongman, & Sereno, 2003). Non-native speakers often have difficulties with the perception and production of Mandarin tones regardless of their first language (L1) prosodic background. Previous research has shown that both tone language speakers (such as Cantonese and Hmong) and non-tone language speakers (such as English and Japanese) have difficulties in differentiating Mandarin tones (So, 2005; So & Best, 2010; Wang, 2006). Despite these difficulties, beginning level Mandarin learners showed significant improvement in identification and production of Mandarin tones in isolation at a syllable level after taking perception and production training with both auditory and visual input.
visual input, or after taking perception training with auditory input only (Wang, 2008; Wang et al., 1999, 2003). However, perception and production of Mandarin tones in larger linguistic units such as phrases and sentences still poses significant challenges to learners who have made progress in isolated syllables. This study explored the effects of production training using auditory and visual input for learning Mandarin tones in larger linguistic units at phrase and sentence levels.

**TRAINING AND L2 SPEECH LEARNING**

Previous research has shown that intensive laboratory based training is effective in modifying adult speakers’ L2 speech perception and production in both segmental and suprasegmental aspects of spoken language (Bradlow et al., 1997; Flege et al., 2004; Hardison, 2003, 2004; Jamieson & Morosan, 1989, 1992; Kingston, 2003; Lively et al., 1993, 1994; Logan et al., 1991, 1993; Logan & Pruitt, 1995; Wang & Munro, 2004). Training studies on L2 segments have shown that after weeks of perception training, trainees improved significantly in perceptual accuracy of target L2 consonant contrasts (Jamieson & Morosan, 1986, 1989; Lively et al., 1993, 1994) and L2 vowel contrasts (Wang & Munro, 1999, 2004). Perceptual learning also transferred to production (better production of the target contrast) without undergoing any training in production (Bradlow et al., 1997). Though still very limited, perceptual training for learning L2 lexical tones has also been reported (Wang, 2008; Wang et al., 1999, 2003; Wayland & Guion, 2003, 2004). For example, Wang et al. (1999) used natural stimuli produced by multiple Mandarin speakers to train native English listeners on Mandarin lexical tones. The trainees’ identification scores increased significantly after two weeks of training. Perceptual learning also spread to the production mode without training in production (Wang et al., 2003). Recent studies on tones also compared the influence of the learners’ L1 prosodic background (So, 2005, So & Best, 2010), or the effectiveness of different training procedures (such as auditory input only versus auditory and visual input), on learning lexical tones (Wang, 2008).

Previous studies have also developed and advanced different training methodologies. In terms of training stimuli, some studies evaluated the impact of training using synthesized stimuli (Strange & Dittmann, 1984; Wang & Munro, 1999), and tested whether learning with synthesized tokens transferred to natural speech (Jamieson & Morosan, 1986, 1989; Rochet, 1995). Other studies used naturally produced stimuli by multiple speakers, the High Variability Paradigm (Logan et al., 1991, 1993), and a combination of synthesized and natural stimuli (Wang, 2002; Wang & Munro, 2004).

With regard to training procedures, some studies tested the use of discrimination training tasks such as requiring the trainees to tell whether the first sounds are the same or different when hearing two stimuli such as “right” and “light” (Strange & Dittmann, 1984; Wayland & Guion, 2003, 2004). Other studies used forced choice identification tasks such as requiring listeners to choose between /l/ and /r/ when hearing a single stimulus such as “right” (Jamieson & Rvachew, 1992; Logan & Pruitt, 1995). Still other studies evaluated a “fading technique” in which the key phonetic/acoustic features that distinguish the target contrasts such as the English /ʌ/-/o/ contrast (Jamieson & Morosan 1986, 1989) or English tense and lax vowel pairs /i/-/I/, /u/-/℧/, and /ε/-/æ/ (Wang, 2002) were expanded and contracted along a synthetic continuum to direct learners’ attention to the most relevant acoustic cues of the target phonetic contrasts for category formation.

Regarding types of input, most training studies involved perceptual training with auditory input only. By comparison, production training studies on L2 speech learning are still limited. One production training study on lexical tones (Leather, 1990, 1997) tested whether visual input of pitch contour change only without auditory input was sufficient to learn nonnative lexical tones. Though effective,