Journal of Phonetics 51 (2015) 6–35

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Effects of talker-to-listener distance on tone

Chilin Shih [a,*], Hsin-Yi Dora Lu [b]

[a] University of Illinois at Urbana-Champaign, United States
[b] National Taipei University of Education, Taiwan

ARTICLE INFO

Article history:
Received 5 July 2014
Received in revised form
16 January 2015
Accepted 6 February 2015
Available online 21 March 2015

Keywords:
Talker-to-listener distance
Tone production
L2 tone perception
Quadratic decomposition of tone
Tone difficulty
Tone learning
Invariant tone features

ABSTRACT

This paper explores the effects of talker-to-listener distance (TLD) on tone and examines its impact on tone perception by both native listeners and second language learners.

Speakers naturally adjust vocal effort to talk to people at different distances, which leads to changes not only in intensity, duration and formant frequencies, but also in fundamental frequency (*f0*). Fundamental frequency is the primary acoustic cue that differentiates Mandarin lexical tones from one another. This study aims to answer the question of whether changes in *f0* as a function of TLD affect tone perception by native (L1) and second language (L2) listeners. If so, what are the specific changes that have an impact on tone perception?

The production study investigates the acoustic correlates of the effects of TLD on tone, using 7959 monosyllabic Mandarin speech files recorded by three speakers under 11 levels of TLD. The perception study explores the effects of TLD on L1 and L2 tone perception by 2 native listeners and 52 L2 listeners.

The effects of TLD on speech production are systematic, and we present models and analyses with sufficient detail to simulate these effects. Intensity, duration, and initial and maximum *f0* increase along with TLD, while time-normalized toneshapes remain invariant. The results of the perception study show that native listeners' performance is robust under changes in TLD, while L2 listeners' perception of tone interacts with TLD. One significant finding is that tone 3 recognition improves with TLD.

This work investigates the relationship between speech production and perception when acoustic attributes change naturally in response to the demands of speech communication. The results have potential applications in speech synthesis, pronunciation training and second language testing.

Published by Elsevier Ltd.

1. Introduction

This paper explores the effects of talker-to-listener distance (TLD) on tone and examines how speech production under demanding conditions impacts tone perception by both native listeners and second language learners.

The distance between a talker and a listener has a strong impact on face-to-face speech communication because acoustic power decreases quickly along with distance. In an open field, acoustic power drops by a factor of 4, or 6 dB, for every doubling of the distance following the inverse square law. In anticipation of the rapid decrease in sound level that occurs when distance increases, talkers intuitively adjust their vocal output to account for that distance, and listeners have matching expectations. Adjustment in speech to account for TLD is a listener-centered adaptation whose goal is successful communication. In principle, everything else being equal, TLD-influenced changes in speech should enhance speech perception, given that this is presumably the intention of the speaker. Whether TLD effects actually help speech perception depends on whether the adjustments made in response to TLD can compensate for the demanding or suboptimal communicative environment created by such long distances. A further complication is the possibility that these adjustments could alter acoustic cues that are crucial for the identification of sound categories, and thus listeners may not be able to successfully decode all such changes.

Previous research has identified the adjustment of vocal effort as the primary talker response to TLD, and changes in vocal effort lead to changes in the acoustic attributes of speech (Cheyne, Kalgaonkar, Clements, & Zurek, 2009; Pickett, 1956; Pickett & Pollack, 1958). Studies of TLD effects and vocal effort have been investigated across segmental and suprasegmental levels, including isolated vowels (Eriksson & Traünmuller, 2002; Liénard & Benedetto, 1999), single words, word pairs or short phrases (Brungart &

* Corresponding author.
E-mail addresses: cls@illinois.edu (C. Shih), doralu0508@gmail.com (H.-Y.D. Lu).

0095-4470/$ - see front matter Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.wocn.2015.02.002

Scott, 2001; Cheyne et al., 2009; Traünmuller & Eriksson, 2000), real conversations (Pelegrín-García, Smits, Brunskog, & Jeong, 2011), different ranges of TLD, and different languages including French, Swedish and English (Cheyne et al., 2009; Liénard & Benedetto, 1999; Traünmuller & Eriksson, 2000). Salient acoustic parameters that are associated with increased vocal effort include increases in fundamental frequency (*f0*), intensity, duration, and changes in formant frequencies (Eriksson & Traünmuller, 2002; Fux, Feng, & Zimpfer, 2011; Liénard & Benedetto, 1999; Pelegrín-García et al., 2011; Thompson & Webster, 1964; Traünmuller & Eriksson, 2000). The effects of TLD are consistent across talkers, listeners and utterances.

The effect of TLD on *f0* is of particular interest when studying a tone language. If *f0* changes as a function of TLD, will that affect tone perception for native (L1) and second language (L2) listeners? If so, what are the changes that have an impact on tone perception?

Mandarin Chinese is a lexical tone language, and the *f0* height and contour over the domain of a syllable is the primary acoustic cue that differentiates among lexical tones. There are four lexical tones in Mandarin, traditionally referred to as tone 1 to tone 4. For example, the syllable *ma1*, with tone 1, a high-level tone, means “mother”; *ma2* with tone 2, a rising tone, means “hemp”; *ma3* with tone 3, a low falling or low falling-rising tone, means “horse”; and *ma4* with tone 4, a falling tone, means “to scold.” In general, changing the tone of a monosyllabic word alters the meaning of the word, though there are nuances behind this statement worth pointing out. First, there exist many syllable-tone combinations with multiple homophones for which meaning disambiguation is based on writing rather than sound. Second, context in continuous speech helps with the recovery of meaning even if the tone is unclear or wrong (Levow, 2005, 2006). That said, the functional load of tone in explaining phonological contrast in Mandarin is still as high as the functional load of vowels (Surendran & Levow, 2004); hence the role of *f0* in Mandarin speech communication cannot be ignored.

While it is generally agreed that relative *f0* height and *f0* shapes are the primary acoustic cues for Mandarin tones (Chao, 1968; Liu, 1924), many acoustic cues are present simultaneously that may affect the perception of tonal categories; these include contrasts in duration, intensity, creakiness, *f0* height, and *f0* turning point (Blicher, Diehl, & Cohen, 1990; Fu, Zeng, Shannon, & Soli, 1998; Gandour, 1983; Garding, Kratochvil, Svantesson, & Zhang, 1986; Hallé, Chang, & Best, 2004; Lee, Tao, & Bond, 2008; Luo & Fu, 2004; Moore, 1995; Shen & Lin, 1991; Shih, 1988; Shih & Ao, 1996; Wang & Li, 1967; Whalen & Xu, 1992). There is also considerable variation in the *f0* dimensions of natural speech. The *f0* contours of tones produced with contrastive focus and in emphatic speech may look distorted when compared to tones produced in normal speech modes, but they tend to be more separated in tone space, which actually helps with tone recognition (Surendran, 2007; Surendran, Levow, & Xu, 2005; Xu, 1999). Tones produced in prosodically weak positions in connected speech can be reduced to the extent that they lose their distinctive tone shapes (Kochanski & Shih, 2003; Kochanski, Shih, & Jing, 2003; Surendran et al., 2005). The acoustic correlates of tones are further influenced by factors including but not limited to speaker, context, sentence prosody, declination, speaking rate and speaker effort (Kochanski et al., 2003; Shih, 1988, 2000; Xu, 1997, 1999).

Tone learning is an important yet challenging task for second language (L2) learners, who need to recognize the characteristic acoustic attributes that distinguish one tone from another. This task is challenging for speakers of non-tone languages in which the typical function of *f0* is to convey speech act, discourse and paralinguistic information. The functional divergence of *f0* usage is revealed in the different patterns of pitch-shift responses observed in speakers with or without tone language backgrounds (Ning, 2014), as well as the observation that people with musical training are better at tone identification even without prior tone training (Lee & Hung, 2008; Ning, 2014). Making the switch to using *f0* to encode lexical information requires the recruitment of different cortical areas (Gandour et al., 2000; Kaan, Barkley, Bao, & Wayland, 2008; Klein, Zatorre, Milner, & Zhao, 2001; Wang, Spence, Jongman, & Sereno, 1999; Wang, Sereno, Jongman, & Hirsch, 2003). Native speakers of a tone language such as Mandarin are sensitive to tonal contours, while native speakers of American English without tone training pay attention to pitch levels (Huang & Johnson, 2010; Stagray & Downs, 1993), beginning and end points, and the similarity between tone shapes (Gottfried & Suiter, 1997; Kiriloff, 1969; So & Best, 2010; Yang, 2010). Tone 2 and tone 3 have similar shapes and are thus difficult to differentiate. Tone 3 is difficult even for Mandarin-speaking young children to identify (Wong, Schwartz, & Jenkins, 2005).

Having access to multiple examples, especially speech materials rich in variation, allows learners to evaluate the relative contributions multiple cues make to a sound category and to identify the primary cues, thus facilitating learning (Bradlow, Pisoni, Yamada, & Tohkura, 1995; Francis, Ciocca, Ma, & Fenn, 2008; Lively, Logan, & Pisoni, 1993; Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994; Logan, Lively, & Pisoni, 1991; McCandliss, Fiez, Protopapas, Conway, & McClelland, 2002; Wang et al., 1999). Manipulating TLD is an effective method for collecting natural variations in tone. It has been shown that TLD correlates with changes in *f0* minima, *f0* maxima and the initial and final slopes of *f0* contours (Fux et al., 2011). Speech variations induced by TLD have been employed successfully in the tone training of L2 learners (Shih, Lu, Sun, Huang, & Packard, 2010; Shih & Lu, 2010).

The current study builds on previous research on TLD and investigates the effects of TLD on Mandarin Chinese tones through both production and perception experiments. Studying TLD effects on tone will potentially provide a dynamic view of how acoustic attributes combine to account for robust tone classification extending from normal to demanding speech and to identify areas of weakness that arise if tone perception is impacted negatively by TLD. This line of research on the effects of TLD is similar to research on many other modes of speech that are produced under demanding or suboptimal conditions that may impair speech communication and prompt speakers to make adjustments to enhance speech, including the Lombard effect, in which speakers produce clear speech in a noisy environment (Bradlow & Alexander, 2007; Picheny, Durlach, & Braida, 1986, 1989). Clear speech is found to be more intelligible, suggesting that the enhancement measures adopted by speakers are usually effective (Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988). Nonetheless, L2 learners may have difficulty decoding the altered acoustic cues and thus may not get the full benefits of clear speech (Bradlow & Bent, 2002).