Available online at www.sciencedirect.com

ScienceDirect

Speech Communication 78 (2016) 34–41

www.elsevier.com/locate/specom

Generating tonal distinctions in Mandarin Chinese using an electrolarynx with preprogrammed tone patterns

Liana Guo[a,b], Kathleen F. Nagle[b], James T. Heaton[a,b,*]

[a] *Massachusetts General Hospital Institute of Health Professions, 36 1st Ave, Boston, MA 02129, United States*

[b] *Massachusetts General Hospital Center for Laryngeal Surgery and Voice Rehabilitation, 1 Bowdoin Sq, 11th Floor, Boston, MA 02114, United States*

Received 1 July 2015; received in revised form 7 January 2016; accepted 13 January 2016
Available online 21 January 2016

Abstract

An electrolarynx (EL) is a valuable rehabilitative option for individuals who have undergone laryngectomy, but current monotone ELs do not support controlled variations in fundamental frequency for producing tonal languages. The present study examined the production and perception of Mandarin Chinese using a customized hand-held EL driven by computer software to generate tonal distinctions (tonal EL). Four native Mandarin speakers were trained to articulate their speech coincidentally with preprogrammed tonal patterns in order to produce mono- and di-syllabic words with a monotone EL and tonal EL. Three native Mandarin speakers later transcribed and rated the speech samples for intelligibility and acceptability. Results indicated that words produced using the tonal EL were significantly more intelligible and acceptable than those produced using the monotone EL.

Keywords: Electrolarynx; Mandarin Chinese; Tonal control; Speech intelligibility.

1. Introduction

Each year, thousands of individuals undergo a total laryngectomy, a standard surgical treatment for advanced laryngeal cancer that results in complete removal of the larynx and leaves them without the ability to phonate normally. Fortunately, several voice prostheses, including the commonly used electrolarynx (EL) provide an alternative source of verbal communication for these individuals. To produce speech, the EL transmits electromechanical vibrations, which can then be shaped by movements of the articulators, through the neck tissue. Due to its portability, ease of use, and readiness to serve as a backup when experiencing difficulty with other modes of alaryngeal speech, the EL is chosen by more than half of Laryngectomees as their primary mode of communication (Hillman et al., 1998).

Although the EL requires little training for users to achieve voicing, most ELs generate little, if any, pitch variation during phonation, contributing to a robotic, unnatural speech quality. Conventional ELs that vibrate on a single fundamental frequency (F_0) have been shown to create a particular deficit in the speech intelligibility of tonal languages such as Thai, Mandarin, and Cantonese (Gandour et al., 1988; Liu et al., 2006; Ng et al., 1998). For words to be correctly perceived in a tonal language when context clues are lacking, a listener needs to hear not only the speech sounds, but also the underlying pitch changes within each syllable (tones). In Mandarin, each syllable contains one of four basic tones (plus a fifth, neutral tone) that make use of F_0 to differentiate the meaning of words with the same sound pattern. Tone 1 has high-level (HL) pitch, Tone 2 middle-rising (MR) pitch, Tone 3 falling-rising (FR) pitch, and Tone 4 high-falling (HF) pitch. For example, the syllable "ma," produced with the four tones, means "mother," "numb," "horse," and "scold," respectively. Liu et al. (2006) found that speakers using a conventional, monotone EL produced F_0 contours that were invariably level. Consequently, listener identification of tones

* Corresponding author at: Massachusetts General Hospital Center for Laryngeal Surgery and Voice Rehabilitation, 1 Bowdoin Sq, 11th Floor, Boston, MA 02114, United States.
E-mail address: james.heaton@mgh.harvard.edu (J.T. Heaton).

http://dx.doi.org/10.1016/j.specom.2016.01.002
0167-6393/

was significantly poorer than those produced by normal laryngeal (NL) speakers.

Studies have found that compared to other phonetic components, including duration and amplitude, F_0 contour provides the most important cue for tone perception for tonal languages (Whalen & Xu, 1992; Zhang, Qi, Song & Liu, 1981). Therefore, ELs that provide dynamic pitch variation can potentially better serve the communicative needs of Mandarin EL speakers. Current EL devices have the capacity to modulate F_0 through several methods: applying varying amounts of finger pressure on a single button (e.g. Western Electric #5, Western Electric; TruTone EL, Griffin Laboratories); controlling expiration pressure from the neck stoma (Uemi et al., 1994); filtering electromyographic (EMG) signals obtained from neck muscle contractions (Goldstein et al., 2004); and adjusting forearm tilt movement (Matsui et al., 2013). These devices have allowed speakers to convey natural intonation patterns in English and Japanese with varying degrees of success; however, effects on the vocal rehabilitation of tonal languages have not yet been shown.

Whereas F_0 contours of intonation can occur over the course of several seconds, tonal contours typically span milliseconds (Xu, 1997). Effective application of these EL devices to Mandarin requires the ability to modulate F_0 rapidly to successfully produce the four tones. To achieve optimal intelligibility, pitch contours generated by the ELs would also need to closely match the typical shape and frequency values for each tone. Considering these factors, the aforementioned EL devices with real-time pitch control have limitations that may fall short of normal F_0 control. Using expiration or finger pressure to provide precise pitch control is difficult to master and may reduce normal speaking rate (Liu and Ng, 2007), and the responsiveness of the EMG-EL's low-pass filter appears to be too slow to support tone production. Additionally, the fixed initial F_0 settings of the TruTone EL and EMG-EL make it difficult to generate appropriate starting tones and F_0 height for each tone.

Wan et al. (2012) presented a viable EL option for Mandarin using the movement of a trackball to control pitch (WT-EL). Users were required to manipulate the trackball with their thumb by moving the direction of the trackball to reflect the differences in tones during phonation. They found that, compared to the monotone EL, the WT-EL performed significantly better in measures of perceptual accuracy and acceptability. However, technical limitations of the device (i.e., 100 ms required to reset to initial F_0) made it difficult for users to consistently generate tonal contours similar to normal speech and produce continuous speech. Additionally, users may find the trackball difficult to master since they needed adequate hand control to precisely manipulate the trackball.

To further enhance Mandarin EL speech, ELs need to closely approximate natural tonal contours without sacrificing convenience. The present study explores the feasibility of achieving EL tonal control for Mandarin using an EL controlled by computer software to generate tonal distinctions. Speakers were trained to articulate their speech coincidentally with preprogrammed pitch patterns to produce words in Mandarin. Use of preset pitch patterns allowed for generated tones to resemble those produced by the NL voice while reducing variability caused by individual speaker differences. This study examined the production and perception of Mandarin using a tone-capable (tonal) EL compared to a monotone EL.

Fig. 1. A TruTone® EL modified to receive an input sine wave determining the instantaneous fundamental frequency (F_0) of the EL voice. Quarter provides size reference.

2. Methods

2.1. Tonal EL design

A TruTone® electrolarynx (EL) was modified by Griffin Laboratories (Temecula, CA) to receive an input signal that determined vocal fundamental frequency (F_0). The TruTone® normally detects the amount of pressure applied to its activation button and modulates F_0 proportionally. The modified TruTone® was physically altered to receive an external electrical signal through a custom input channel (see the black wire in Fig. 1) instead of using the pressure-sensitive activation button. The circuitry within the TruTone® was reprogrammed to detect the instantaneous frequency of a sine wave input and drive its mechanical transducer (e.g. sound source) at the same frequency. Sine wave patterns matching the four Mandarin tones or a monotone condition (see Section 2.2) were synthesized using audio editing software (Adobe Audition CS6, Adobe Systems), and inputted into the TruTone® through the audio output of a personal computer. To produce speech, users depressed the EL activation button to initiate a ready state, began playback of a preset tone on