Electroglottograph and Acoustic Cues for Phonation Contrasts in Taiwan Min Falling Tones

Ho-hsien Pan\textsuperscript{1}, Mao-Hsu Chen\textsuperscript{1}, Shao-Ren Lyu\textsuperscript{1}

\textsuperscript{1}Department of Foreign Languages and Literatures, National Chiao Tung University, Hsinchu Taiwan

hhpan@faculty.nctu.edu.tw, r97725008@ntu.edu.tw, thousandshine@hotmail.com

Abstract

This study explored the effective articulatory and acoustic parameters for distinguishing Taiwan Min falling unchecked tones 53 and 31 and checked tones 5 and 3. Data were collected from Zhangzhou, Quanzhou, and mixed accents in northern, central, and southern Taiwan. Results showed that EGG parameters, Contact Quotients (CQ) and Peak Increase in Contact (PIC) were not effective in distinguishing checked from unchecked tones across speakers. In contrast, f0 contour and Cepstral Peak Prominence (CPP) consistently distinguished checked tones from unchecked tones across speakers. The f0 onset was highest for tone 53, followed in order by tone 3, 5 and 31. The f0 contours of tone 5 were the highest in the latter half of the vowels. CPP measures of checked tones were higher than those of unchecked tones in the latter portion of vowels.

Index Terms: Electroglottoraph, glottal wave form, f0, Cepstral peak prominence, Taiwan Min, phonation.

1. Introduction

In languages with mixed pitch and phonation tone systems, the effective acoustic parameters for characterizing phonation types vary across languages. As the production of pitch for lexical tones and voice qualities both involve larynx, the variations in pitch may contribute to the variation in voice qualities. In Burmese, inverse filter acoustic waveforms were used to characterize lexical tones produced with modal or creaky voice qualities accompanying lexical tones. It was found that peak flow and the asymmetry between rate of rise and fall in glottal pulses distinguished creaky and modal phonation [1]. Most previous studies on phonation contrasts use articulatory data to supplement acoustic data. For example, both glottal airflow and acoustic data were studied to characterize breathy, creaky, and modal phonation types accompanying three lexical tones in Green Mong. The closed phase duration of glottal airflow distinguished phonation types as did amplitude difference between first harmonic (H1) and second harmonic (H2), f0, vowel duration, vowel quality and voice onset time [2, 3, 4, 5]. The harmonic amplitude difference is referred to as spectral tilt (H1-H2). Additionally, both airflow and acoustic data were used to study languages including Jingpho, Hani, Yi (Nasu) and Wa, with tense / lax phonation contrasts accompanying lexical tonal contrasts. Lax vowels were found to have a greater flow to pressure ratio and a greater ratio of f0 to H2 for lax vowels [6]. An Electroglottograph (EGG) study of Hanoi Vietnamese found an increase in the open quotient for obstruent-final rhyme syllables and glottalization for nasal-final rhyme syllables [7]. In an articulatory and acoustic study of two languages with phonation and lexical tone contrasts, namely Hmong and Yi, it was found that two EGG parameters, including closed (contact) quotients obtained using a hybrid method (CQ\_H) and Peak Increase in Contac (PIC), distinguished modal, breathy, and creaky phonations in Hmong and distinguished tense and lax phonations in Yi [6, 8].

Acoustic parameters including spectral tilts which measure the amplitude differences between the first harmonic (H1) and the strongest harmonic in the first formant (A1), second formant (A2) and third formant (A3), i.e. H1*-H2* (corrected), H1*-A1* (corrected), H1*-A2* (corrected), H1*-A3* (corrected), and Cepstral Peak Prominence (CPP) distinguished phonation contrasts in Hmong and Yi [6, 8].

A previous fiberoptic study of Taiwan Min checked syllables with final unreleased stops [p t k ?] produced with tones 5 and 3, showed glottalization with adduction of ventricular folds for checked tone syllables in citation form. However, glottalization sometimes disappeared when checked tones were placed in a sentence or phrase [9]. An inverse filtered oral airflow study found low airflow and a long closed phase in the glottal waveform of checked tones produced by some speakers [10]. A more recent laryngoscopic study found consistent adduction of glottal folds for Taiwan Min checked tones, but less or no adduction of ventricular folds or aryepiglottic folds for younger speakers [11].

Although airflow, fiberoptic, and laryngoscopic studies found glottalization for checked tones, spectral tilt as measured by H1-A2 were not effective for distinguishing checked from unchecked tones [12]. Since acoustic data is more readily available during speech processing, it is essential to identify the effective acoustic parameters for documenting phonation contrasts in Taiwan Min.

Following the study on Yi and Hmong, the current study explores the effectiveness of EGG parameters including CQ and PIC and acoustic parameters including spectral tilt measures and CPP in order to characterize reliable cues for Taiwan Min checked tones. As Taiwan Min tone 5 is currently undergoing sound change during which the f0 onset of tone 5 was lowered from high to mid f0 range, the identification of the effective EGG and acoustical parameters can help explore how voice quality vary along with the f0 change [13]. It is hypothesized that voice quality contrast in terms of spectral tilt and Cepstral Peak Prominence will be maintained to help distinguishing checked tones from unchecked tones while the contrast in f0 onset is undergoing changes. However, there should be inter-speaker variations as sound change spread around different regions and ages.

2. Method

2.1. Subjects

Following the isoglosses defined by Ang [14], we recruited eight native speakers from Taiwan who spoke Zhangzhou (潮}

Copyright © 2011 ISCA 649

28 – 31 August 2011, Florence, Italy
Subject backgrounds. F: Female, M: Male, N: Northern, C: Central, S: Southern

<table>
<thead>
<tr>
<th>Subject</th>
<th>Age</th>
<th>Gender</th>
<th>Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>N. Zhangzhou</td>
<td>&gt;40 years old</td>
<td>M, F</td>
<td>C</td>
</tr>
<tr>
<td>N. Quanzhou</td>
<td>1 F</td>
<td>M</td>
<td>C</td>
</tr>
<tr>
<td>C. Zhangzhou</td>
<td>1 F</td>
<td></td>
<td>C</td>
</tr>
<tr>
<td>C. Quanzhou</td>
<td>1 M, 1 F</td>
<td></td>
<td>C</td>
</tr>
<tr>
<td>S. Mixed</td>
<td>2 F</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.2. Corpus

There were 390 randomized monosyllabic items in the reading list, including target items with checked tones 5 and 3, control items with tones 53 and 31, and filler items with tone 55, 13 and 33.

2.3. Instruments

Glottal Enterprise EGG system EG2-PCS and TEV Tm-728 II microphones were used to simultaneously pick up acoustic and EGG data which were then recorded with Audacity onto a laptop. The EGG and acoustic signals were then separated into different channels, inverted and analyzed with EggWorks [15]. The acoustic signals were tagged using Praat to mark vowel boundaries, and then analyzed with VoiceSauce to obtain spectral tilt and CPP measurements [16].

2.4. Procedure

During the recording, two electrodes were placed on speakers’ necks while the microphone was placed within 15 cm in front of all speakers’ mouths. Speakers first read the number of each token and then paused before reading each target token on the reading list. Speakers paused after every 100 tokens to take a rest. The recording lasted for about 40 minutes.

2.5. Data analysis

As shown in Figure 1, for EGG signals, two parameters, namely the CQ of glottal waveform and instantaneous peak rate of change in contact (Peak Increase in Contact, PIC), were calculated at nine equal intervals (10% through 90% during vowel phonation) using EggWorks. CQ was calculated using a hybrid method (CQ_H). In this method, the edges of the contacting phase of glottal pulses are defined using two methods. The beginning of the contact phase is the positive peak in the first derivative of the EGG signal. The end of the contact phase is arbitrarily determined using a 25% threshold [16]. The PIC is the peak positive value in the derivative of the EGG signal, equivalent to DECPA measure [17].

VoiceSauce was used to query acoustic parameters including F0, spectral tilt (corrected H1*-H2*, H1*-A1*, H1*-A2*, H1*-A3*) and CPP [16, 18]. F0 was estimated using the STRAIGHT algorithm at 1 ms intervals [19]. Since vowel height and tongue advancement were not controlled for, the corrected amplitude of the first three formant, bandwidth, and corresponding spectral tilts H1*-H2*, H1*-A1*, H1*-A2*, H1*-A3* was calculated using Snack Sound Toolkit [20]. CPP is a measure of cepstral peak amplitude normalized over all amplitudes [21]. CPP measures represent how prominent the cepstral peak is against “background noise,” and indicates how periodic the signals are.

A one-way ANOVA (checked vs. unchecked tones) was used to analyze the EGG and acoustic parameters at 9 vowel time points for each speaker.

Figure 1: Acoustic waveform (top panel) and EGG waveform (lower panel).

3. Result

If Taiwan Min checked tones are produced with tense/creaky voice, then the checked tones should have longer CQ and higher PIC values. Additionally, the spectral tilt of checked tones should show a less steep spectral slope, thus having smaller values of H1*-H2*, H1*-A1*, H1*-A2* and H1*-A3*.

The CPP values tend to be less for phonation with breathiness.

3.1. EGG parameters

As shown in Table 2, only five of the eight speakers showed significant differences in the mean contact quotients (CQ_H) during the nine vowel time points. CQs at vowel midpoints were significantly different among checked 5 and 3 and unchecked tones 53 and 31 for these five speakers only.

As shown in Table 2, again significantly different PIC values between checked and unchecked tones can only be found among six of the eight speakers. Consequently, neither CQ_H nor PIC was effective measures in documenting phonation type across speakers.

Table 2. EGG, spectral tilt and CPP at nine vowel time points in checked tones 5 and 3 and unchecked tones 53 and 31 syllables. NZ: Northern Zhangzhou, NQ: Northern Quanzhou, CZ: Central Zhangzhou, CQ: Central Quanzhou, SM: Southern Mixed, F: Female, M: Male. ** p<.01
3.2. Acoustic parameters

3.2.1. F0

As shown in Figure 2, the f0 onset was the highest for tone 53, followed by tone 3, tone 5 and finally tone 31. After the vowel midpoint (50%) f0 contours were the highest for tone 5 and the lowest for tone 31.

3.2.2. Spectral Tilt

Table 2 reveals how effective it is to use spectral tilt values to distinguish checked tones from unchecked tones. H1*-A1* and H1*-A2* were the only two spectral tilt measures that effectively distinguished all speakers' data. However, there was no single vowel time point in which significant spectral tilt differences were observed among all speakers. At most were five speakers with significant differences in H1*-A1* and H1*-A2* at the 80% vowel time points.

Figure 2: F0 contours of Taiwan Min tones 53, 3, 5, and 3.

3.2.3. Cepstral Peak Prominence (CPP)

Figure 3: CPP of Taiwan Min unchecked tones 53 and 31 and checked tones 5, and 3.
As shown in Table 2, CPP significantly differentiated checked tones from unchecked tones produced by all speakers at 20% and 70% vowel time points (NZ1F: 20% F(I, 310)=21.41, p<.01, 70% F(I, 310)=13.79, p<.01; NZ2M: 20% F(I, 323)=7.91, p<.01, 70% F(I, 323)=49.82, p<.01; NQ1F: 20% F(I, 309)=20.52, p<.01, 70% F(I, 309)=19.29, p<.01; CZ1F: 20% F(I, 324)=35.45, p<.01, 70% F(I, 324)=55.8, p<.01; CQ1F: 20% F(I, 316)=21.39, p<.01, 70% F(I, 316)=11.38, p<.01; CQ1M: 20% F(I, 315)=30.85, p<.01, 70% F(I, 315)=11.63, p<.01; SM1F: 20% F(I, 325)=41.55, p<.01, 70% F(I, 325)=25.33, p<.01; SM2F: 20% F(I, 325)=42.22, p<.01, 70% F(I, 325)=31.82, p<.01).

Figure 3 shows detailed CPP values from 10% to 70% vowel time points in checked tones 5 and 3, and unchecked tones 53 and 31. It was found that the CPP values of checked tones were lower than unchecked tones at the 20% time point, but higher than unchecked tones at the 70% time point. In sum, EGG parameters, including Contact Quotient (CQ_H) and Peak Increase in Contact (PIC), did not reveal consistent pattern across speakers. Among the acoustic parameters, f0 contours and Cepstral Peak Prominence (CPP) showed consistent cross speaker patterns at 20% and 70% vowel point times. Inter-speaker variation in spectral tilt such as H1*-H2*, H1*-A1* and H1*-A2* may indicate dialectal variation.

4. Discussion

The lack of consistent results using EGG indicators of phonation type indicates that further articulatory studies are necessary for characterizing the phonation types of Taiwan Min checked tones. Acoustic measures are more consistent in identifying phonation types. CPP measures indicate periodic voicing is more consistent within the production of the latter half of syllables carrying a checked tone.

Tone 5 is undergoing a sound in change in Taiwan Min. Laryoscopic study showed no or little adduction of ventricular folds for checked tones produced by younger generation. The inter-speaker variation observed in spectral tilt, such as H1*-H2*, H1*-A1* and H1*-A2*, may reflect this sound change. While CPP consistently distinguish checked from unchecked tones, spectral tilt parameters may show dialectal variation due to sound change. For future study, CPP and spectral tilt measures could be used to document how the phonation of checked tones vary among young and middle aged speakers in different dialect regions.

5. Acknowledgements

This project was supported by National Science Council in Taiwan (NSC 98-2401-H-009-040-MY3). The software VoiceSauce and EggWorks were provided by the UCLA phonetics lab. We would like to thank Yi- chu Ke for help in designing the corpus.

6. References