INTERSPEECH 2021
30 August – 3 September, 2021, Brno, Czechia

Taiwan Min Nan (Taiwanese) Checked Tones Sound Change

Ho-hsien Pan[1], *Shao-ren Lyu*[1,2]

[1]National Yang Ming Chiao Tung University, Taiwan
[2]McGill University, Canada
hhpan@faculty.nctu.edu.tw, shaorenlyu@gmail.com

Abstract

The multifaced changes of Taiwan Min Nan (TMN) checked sandhi tones, S3 and S5 were investigated as well as the checked base tones, B3 and B5. Simultaneous EGG data, CQ_H and acoustic data, including duration, f0 offset at 80% vowel interval, and spectral tilt H1*-A3* from forty male and female speakers above 40 and under 30 years of age were analyzed. Though different measures progress at different paces, in general, as the coda stops [p, t, k, ʔ] from full stop closure, to energy damping and finally to complete deletion, vowel duration lengthening, f0 offset lowering, and more modal phonation were observed. Gender effects were found on f0 offset and CQ_H offset. The pace of progress is more advanced for base tone B5 with glottal coda stops. After coda deletion, the contexts conditioning the anticipatory co-articulation were removed and vowel and tone characteristics were modified to be similar to those found in open syllables.

Index Terms: Electroglottography, duration shortening, f0 lowering, non-modal phonation, H1*-A3*, contact quotient

1. Introduction

Chinese languages underwent two stages of tonal development: the first tonogenesis stage and the second tone split stage [1]. At the end of the two stages of tone developments, there were eight tonal categories, including two checked and six unchecked tones. Compared with unchecked tones, generally speaking, checked tones in CV[stop] syllables, are short in duration and often produced with non-modal phonation. The tone changes continued and, in language such as Mandarin, both checked tones were eventually lost. Even among languages that retained checked tones, such as Shanghai Chinese, Cantonese and Taiwan Min Nan (TMN), the idiosyncratic characteristics associated with checked tones are gradually disappearing. Currently, Shanghai Chinese low checked tone is gradually losing its breathy non-modal phonation [2]. Ohala [3] suggested that tone change could result from the misparsing of coarticulatory pitch variation from coda stops to preceding vowels. As coda stops weaken and disappear, the pitch variation on the vowel was perceptually misparsed as a feature of vowel thus retained even after the deletion of the coarticulatory source – coda stops. Ohala [4] further proposed that though sound sounds may have physical or physiological factors, language or culture-specific factors can still determine the implementation of sound changes. However, here in this study we explore TMN sound changes of which the coarticulatory effects from coda stops to preceding vowels were removed after deletion of coda stops.

This study shows that coda stop weakening contribute to duration lengthening, disappearance of non-modal phonation, and f0 contour changes of preceding vowels. Before discussing the on-going checked tone sound change, we will first explain basic TMN tonal properties. There is a high checked tone T5 and a low checked tone T3 in TMN. The tone numbers stand for pitch height. The larger the number is, the higher the pitch becomes. When several TMN tones come together, the surface tonal values may change. This is called tone sandhi. According to TMN tone sandhi rules, base tones surface in tone sandhi group domain-final positions, whereas sandhi tones surface in non-final positions of tone sandhi groups. There are two cyclic tone sandhi rules for T3 and T5, that is /3/ → [5] and /5/ → [3] in the non-final position of tone sandhi group. Sandhi checked tone /3/ [5] and /5/ [3] will be denoted as S5 and S3 respectively, whereas base tone /3/ [3] and /5/ [5] will be denoted as B3 and B5 respectively in the following texts [5, 6].

TMN checked tones S3, S5, B3 and B5 in syllables closed by voiceless unreleased coda stops [p, t, k, ʔ] are short and produced with glottalized phonation according to results of fiber optical studies [7]. Furthermore, according to results of inverse-filtered oral airflow study, some speakers produced checked tones with small amounts of airflow and a longer closed phase [8]. Pan [13] found that the vowels of high checked sandhi tones S5 was produced with a longer CQ_H, lower H1∗-A3∗ and a higher Cepstral Peak Prominence (CPP), suggesting a longer close phase, a more abrupt glottal closure and more periodic voicing than low checked sandhi tone S3. Table 1 provides a description of the measures above.

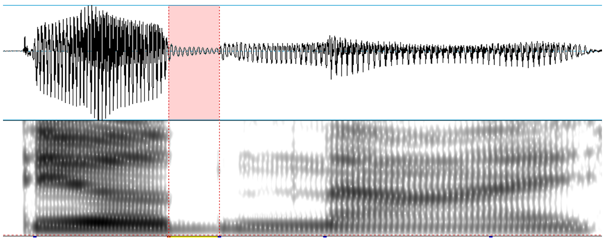
Table 1: *Summary of measures*

Measures	Descriptions
Duration	Duration of vowel measured in milliseconds
Fundamental Frequency (F0) Offset	Frequency measured in Hertz (Hz) at 80% (offset) vowel point time.
H1*-A3*	Amplitude differences between the corrected first harmonic and the strongest harmonic of the corrected first formant. Measured in dB. Correlates with voice quality (higher = breathier/more lax)
CQ_H	Contact quotient of glottal cycle taken with hybrid method. Taken from glottal wave form of Electroglottograph (EGG) signals. Measured in percentages. (higher percentage = longer contact quotient = creakier / tenser voice quality).
CPP	The amplitude of harmonics above the noise floor. Measure of periodicity of vocal fold vibration. (higher amplitude = more periodic)

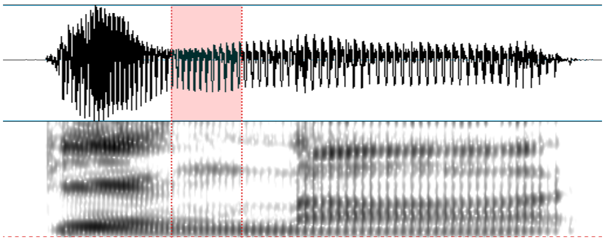
http://dx.doi.org/10.21437/Interspeech.2021-672

TMN checked tone sound changes were first reported by field workers for high checked base tones T5 from which the glottalization gradually disappeared [9], and [10]. Pan [13] found that the CQ_H, H1∗-A3∗ and CPP values for the non-modal phonation distinction were lost among checked base tones.

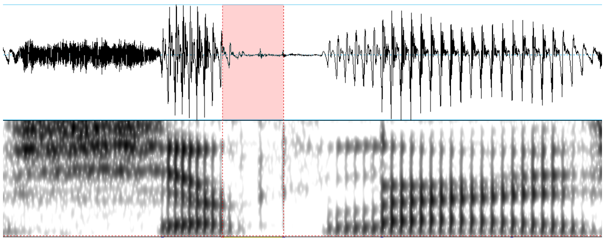
In addition to the disappearance of non-modal phonation, field work revealed the deletion of glottal stops [ʔ] among high checked base tone B5, but not among low checked base tone B3 [10], [11], and [12]. Pan [13] documented the gradual disappearance of coda stops. Four different weakening forms appear as the final voiceless stops decline from a full oral/glottal stop closure into an energy dip at the vowel's end, or an aperiodic voicing, and finally into a complete coda deletion (Figure 1). Currently, over 80% of /ʔ/ codas and less than 20% of /p t k/ codas were deleted. The undeleted /p, t, k/ codas were more likely to be produced with a full stop closure among low checked tone [13].



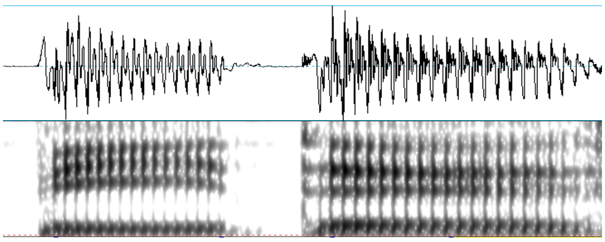
(a) Stop: [tek5 nã13] "bamboo forest" Female, < 30



(b) Damping: [tek5 nã13] "bamboo forest" Male, > 40



(c) Aperiodicity [sek 5 lai33] "indoor" Male, < 30



(d) Coda deletion [pit5 te33] "pencil case" Male, < 30

Figure 1: *Disyllabic word produced with (a) stop, (b) energy dip, (c) irregular glottal vibration during vowel, (d) with coda deletion.*

In addition to the disappearance of glottalized phonation and coda weakening, field workers also reported vowel lengthening and f0 lowering for high checked tone [10, 11 and 12]. The f0 lowering is dialect dependent and results in tones that sound similar to unchecked mid tone [10 and 12], mid-falling tone, high falling tone, or high-level tone [9, 10 and 11].

So far, no acoustic data that illustrates the TSM checked tones sound change exists. This study provides acoustic and Electroglottography (EGG) data from syllables with various degrees of coda stops weakening to reveal the duration shortening, f0 lowering, and disappearance of non-modal phonation before full stops, stops realized as energy damping, aperiodic voicing, or completed deleted stops.

2. Method

Simultaneously recorded EGG and acoustic data of sandhi and base checked tones 3 and 5 with four coda types— stop, energy damping, aperiodic voicing and complete coda deletion—were compared.

2.1. Speakers

As changes in F0 contours differ across different dialect regions, the speakers were selected from the five major dialect regions on the west coast of Taiwan: Northern Zhangzhou (NZ), Northern Quanzhou (NQ), Central Zhangzhou (CZ), Central Quanzhou (CQ) and Southern Mixed (SM) [9]. The study recruited eight native speakers from each of the five dialect regions. These included two females and two males in their 20s, and two females and two males older than 40 years. In all, there were 40 speakers [(2 males + 2 females) × 2 age groups × 5 dialect regions]. These participants were selected because they speak TMN as their native tongue at home. Other than one young male speaker from central Quanzhou, all the speakers had lived in their native dialect region until they were 16 years of age. All participants also speak Mandarin.

2.2. Material

The materials included 236 disyllabic words embedded in the carrier sentence "Please listen to the tone of ____ first," which consisted of all the syllables associated with surface tone 33, as indicated in (1). The carrier sentences asked speakers to pay attention to tones. The words were familiar to native TMN speakers. Each of the 236 carrier sentences was repeated twice; therefore, each speaker produced 472 sentences in total.

"Please listen to the tone of ____ first," (1)

[siŋ^{33} tĩã^{33} ____ e^{33} sĩã^{33} tiau^{33}]

"First listen ____ (TSG) POSS tone."

Words were familiar lexical items to native TMN speakers. A tone sandhi group, as shown in (1) followed each target disyllabic word. The first syllable of target disyllabic lexical items carries a sandhi tone, whereas the second syllable carries a base tone at the final position of a tone group.

2.3. Instruments and Software

The Glottal Enterprise EGG system EG2-PCS picked up the EGG signals from two electrodes placed on each side of each subject's neck around the thyroid cartilage. A TEV Tm-728 II microphone placed 15 cm from each speaker's mouth picked up the acoustic sound waves. The recorded EGG and acoustic signals were simultaneously transferred to a laptop using Audacity software. Next, the EGG and acoustic data were preprocessed using EggWorks [14]. Praat was used for the manual transcription of the sound waves. The F0 data were analyzed using STRAIGHT algorithm [15]. Then, both the EGG and sound waves were analyzed together with VoiceSauce [16] to extract the acoustic and EGG measures.

2.4. Experiment Procedure

The speakers first read the order number for each sentence and then paused before reading the sentence. The recording of each speaker lasted approximately 35 minutes. Speakers were asked to pause after every hundredth sentence to take a drink of water before continuing. A technician worker was with the speaker in the recording room to monitor the EGG and acoustic signals and to indicate errors to the speakers.

2.5. Data Analysis

The tokens with miss-identified corrected F1* values were excluded from further analysis. The tokens with a mistracked F0 were not excluded from further analysis, since irregular glottal vibration often characteristizes glottalization.

Spectral tilt H1*-A3*, the amplitude difference between the first and strongest harmonic of the third formant, is also taken from the corrected F1. Mid-frequency spectral tilt, H1*-A3*, the amplitude difference between first harmonic (H1*) and the strongest harmonic of the third (A3*) formant, reflects the abruptness with which the airflow is cut off when the glottis closes [17]. The larger the H1*-A3* values, the less abrupt the glottis closes, corresponding to more lax voice quality.

CQ_H is the contact quotient during glottal vibration cycle, obtained by using the maximum value of the first derivative of the EGG signal at the onset of contact and the 25% threshold of peak-to-peak amplitudes of EGG signals at the offset of contact. CPP and CQ_H were found to distinguish tense and lax phonations in Yi [18].

Logistic mixed effect regression model (stop closure, energy damping, coda deletion) with speakers and lexical items as random factors were used to analyze the effect of coda categorization on duration, F0 offset at 80% vowel intervals and voice quality measures, namely H1*-A3* and CQ_H at 80% vowel intervals.

3. Results

3.1. Socio-phonetic effects: age, gender and dialect

There were no dialect or age effects on duration, f0 offset, H1*-A3*, or CQ_H. However, there were significant effects of gender on the f0 offset and CQ_H at 80% vowel intervals. Female speakers' f0 offsets and CQ_H 80% were significantly higher ($p< .001$***) and longer ($p<.001$***) than those of male speakers respectively. The higher f0 of female speaker were physically determined. The longer CQ_H suggested more glottalized phonation by female speakers. Thus, the f0 offset and CQ_H data were further divided into two sets according to gender.

3.2. Duration for sandhi and base tones 3 and 5

Figure 2 demonstrates that vowel durations of sandhi and base tones Base tone 3 (B3) and Base tone 5 (B5) with final coda [ʔ] were longer than those before base tones B3 and B5 closed off with [p, t, k] codas, which in turn were longer than sandhi tones sandhi tone 3 (S3) and sandhi tone 5 (S5) closed off with [p, t, k] codas. Aside from base tone B3 with coda [ʔ], vowel duration lengthening can be observed as coda stops weaken from energy damping to complete deletion. Though previously existing field work reported lengthening only among B5 [10 and 12], in this study vowel lengthening was also recorded among sandhi tones S3 and S5 after the deletion. Vowel duration of base tone B3 was not lengthened during coda weakening. Results of mixed effect regression model showed that vowel duration in syllables with coda deletion were significant longer than those with coda produced as either as full stop (β=-37.424, $p< .001$) or energy damping (β=-36.128, $p< .001$).

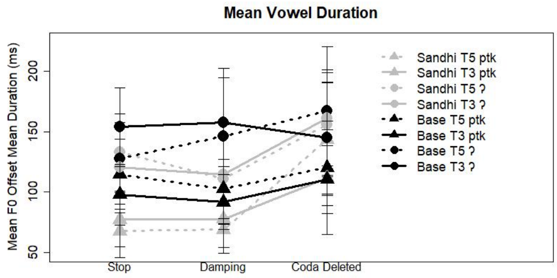


Figure 2: *Duration of syllables carrying sandhi and base tones T3 and T5 with glottal and oral coda stops.*

3.3. F0 offsets for sandhi and base tones T3 and T5

As shown in Figure 3, the effect of coda weakening on f0 offset were most pronounced for base tone B5 with coda [ʔ]. F0 offset lowering can also be found among sandhi tone S5 with coda [p, t, k, ʔ]. Though existing field works reported the F0 onset being lowered only among base tone B5, the current study found f0 lowering among sandhi tone S5 as well [9, 11].

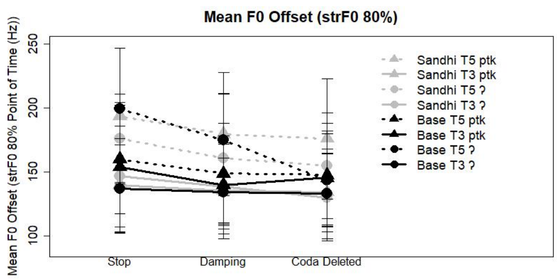


Figure 3: *F0 offset (F0 at 80%) of sandhi and base tones 3 and 5 with glottal and oral coda stops.*

Table 2: *Results of linear mixed effect regression model (coda type) on f0 offset at 80% vowel intervals. 3: low checked tones, 5: high checked tones.*

		Estimate β	Std. Error	Pr(>|z|)	Contrast
Female	5	3.330	1.295	<.05*	Delete < damping
		4.099	1.389	<.001***	Delete < stop
Male	3	6.930	1.479	<.001***	Delete < damping
		13.092	1.670	<.001***	Delete < stop
		6.161	1.455	<.001***	Damping < stop
	5	8.836	1.444	<.001***	Delete < damping
		12.897	1.759	<.001***	Delete < stop
		4.061	1.404	<.05*	Damping < stop

Results of linear mixed effect regression models (Table 2) showed that, generally, the f0 offsets in syllables with deleted codas were lower than those with coda produced as either full stop or energy damping for both male and female speakers.

3.4. Spectral tilt measure, H1*-A3*

With reference to H1*-A3* measured at 80% within the vowel, H1*-A3* of sandhi tone S5 with codas [p, t, k] decreased as coda stops weakened from full stop to energy damping and finally to complete deletion (Figure 4). Similarly, H1*-A3* of base tones B5 and B3 with coda [ʔ] decreased as final stops weakened from full stop to energy damping. The decrease of H1*-A3* suggests that the phonation become less breathy / lax and more modal-like as coda stops weakened.

The results of linear mixed effect regression models showed significant increase of H1*-A3* as coda of high checked tone weakened from full stop to energy damping and finally to complete deletion (Table 3). Similarly, H1*-A3* of low checked tone increased significantly as coda weaken form energy damping to complete deletion.

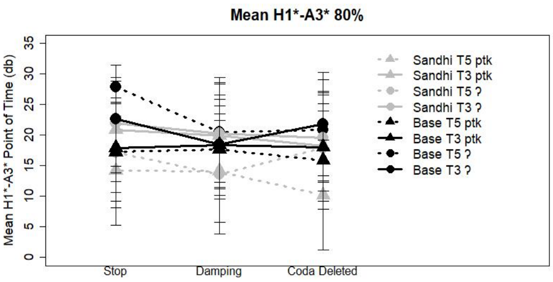


Figure 4: *H1*-A3* of sandhi and base tones 3 and 5 with glottal and oral coda stops.*

Table 3: *Results of linear mixed effect regression model (coda type) on H1*-A3* at 80% vowel intervals. 3: low checked tones, 5: high checked tones*

	Estimate β	Std. Error	Pr(>|z|)	Contrast
3	.824	.214	<.001***	Damping < delete
5	.915	.244	<.001***	Damping < delete
	-.845	.333	<.05*	Stop < damping

3.5. Glottal contact quotient

Regarding the contact quotient, during glottal waves obtained through EGG, generally speaking, the CQ_H at 80% vowel interval before B5 and S5 with coda [ʔ] gradually decrease as coda [ʔ] weakened from full stop to energy damping and finally to complete deletion (Figure 5). However, the contact quotient increased as coda stop [ʔ] weaken in syllables carrying base tone B3. Due to the contrasting directions of phonation changes, further detailed study across different ages and dialect regions are necessary.

The results of linear mixed effect regression model revealed significant increases in contact quotient as final stop of checked tones weakened from full stop to either energy damping or complete deletion for both male and female speakers (Table 4).

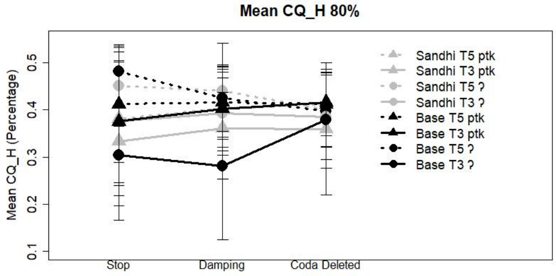


Figure 5: *CQ_H of sandhi and base tones 3 and 5.*

Table 4: *Results of linear mixed effect regression model (coda type) on CQ_H at 80% vowel intervals. 3: low checked tones, 5: high checked tones.*

		Estimate β	Std. Error	Pr(>|z|)	Contrast
Female	3	-.012	.003	<.001***	Stop < delete
		-.012	.002	<.001***	Stop < damping
	5	.006	.002	<.05*	Delete < damping
		-.009	.003	<.01**	Stop < delete
		-.015	.002	<.001***	Stop < damping
Male	3	-.014	.004	<.001***	Stop < delete
		-.016	.003	<.001***	Stop < damping
	5	-.013	.003	<.001***	Stop < delete
		-.012	.003	<.001***	Stop < damping

In sum, this study found duration lengthening and f0 offset lowering in checked tones. The direction of phonation changes is more variable between high and low checked tones. The pace of change varies among the different measures. In sum, the checked tone change was more progressive among syllables ending in [ʔ] than among syllables ending in [p, t, k].

4. Discussion

This study addresses the sound change of TMN checked sandhi and base tones by examining the contributions of post-vocalic voiceless consonant deletion to vowel duration, F0 and phonation. In general, after coda deletion the contexts conditioning the anticipatory co-articulation are removed and vowel and tone characteristics are modified to reflect those found in open syllables. These include longer vowel duration, lower F0, and a less constricted voicing.

Though sound changes can result from timeless physiological factor such as coarticulation, however, language specific factor can determine the direction of sound change. Instead of retaining the short duration, non-modal phonation and f0 contours caused by coarticulation from coda stops to preceding vowels. Here TMN choose to eliminate these coarticulatory effect after coda deletion. TMN listeners did not mis-parse the coarticulatory cues. Though mis-parsing of coarticulatory cues caused tonogenesis in the past, the current TMN checked tone changes do not involve mis-parsing.

5. Acknowledgements

The study was supported by the Ministry of Science and Technology in Taiwan.

6. References

[1] A. G. Haudricourt, “The origin of tones in Vietnamese,” In Evolution of Languages and Techniques (Ed.) Martine Mazaudon, Boyd Michailovsky and Alexis Michaud. Berlin: De Gruyter Mouton. https://halshs.archives-ouvertes.fr/halshs-01678018/file/Haudricourt1954_OriginOfTonesInVietnamese.pdf

[2] J. Gao, “Sociolinguistic motivation in sound change: on-going loss of low tone breathy voice in Shanghai Chinese,” Papers in Historical Phonology, Volume 1, pp.166-186, 2016.

[3] John J. Ohala, “Sound change as nature's speech perception experiment, Speech Communication,” Volume 13, Issues 1–2, pp. 155-161, 1993.

[4] John, J. Ohala, “Sound change is drawn from a pool of synchronic variation,” In Language Change, (Eds.) Leiv Breivik nd Ernst H. Jahr, Berlin, New York: De Gruyter Mouton, pp. 173-198, 2011, https://doi.org/10.1515/9783110853063.173

[5] R., Cheng, “Tone sandhi in Taiwanese,” *Linguistics*, vol. 41, pp. 19-42, 1968.

[6] R., Cheng, “Some notes on tone sandhi in Taiwanese,” *Linguistics* vol. 100, pp. 5-25, 1973.

[7] R., Iwata, M. Sawashima, and H., Hirose “Laryngeal adjustments for syllable-final stops in Cantonese,” *Annual Bulletin Research Institute Logopedics and Phoniatrics*, vol. 15, pp. 45-54, 1981.

[8] H. -H, Pan, “Voice quality of Amoy falling tones,” *J. Acoust. Soc. Am.*, vol. 90, pp. 2345, 1991.

[9] S. -C., Chen, “台灣閩南語元音系統及陰陽入聲調的變異與變化─台灣閩南語的字表調查分析 (The vowel system changes and the Yin-/Yang-entering tonal variations in Taiwanese Hokkien),” *Journal of Taiwanese Language and Literature*, vol. 3, pp. 157-178, 2009.

[10] S. -C., Chen, “台灣閩南語新興的與音變異：台北市、彰化市及台南市元音系統與陽入原調的調查分析 (New sound variation in Taiwan Southern Min: Vowel System and the lower register entering tone in Taipei, Changhua and Tainan),” *Language and Linguistics*, vol. 11, no. 2, pp. 425–468, 2010.

[11] J. -C., Liao, “台語入聲調之現況分析 (Investigation on the Entering Tones in Taiwanese),” MA thesis, National Hsinchu University of Education, Taiwan, 2004.

[12] U. -J., Ang, “音變的動機與方向：漳泉競爭與台灣普通腔的形成 (The Motivation and Direction of Sound Change: On the Competition of Minnan Dialects Chang-chou and Chuan-chou, and the Emergence of General Taiwanese),” Ph.D. thesis, National Tsing Hua University, Taiwan, 2003.

[13] H. -H., Pan “Glottalization of Taiwan Min checked tones”, *Journal of International Phonetic Association*, vol. 47, no. 1, pp. 37-63, 2017.

[14] H., Tehrani, “EggWorks. This is available online at <http://www.linguistics.ucla.edu/faciliti/facilities/physiology/EGG.htm>, 2010. (date last viewed 18/03/2021).

[15] H., Kawahara, I., Masuda-Katsuse, and A., de Cheveigné, “Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds,” *Speech Communication*, vol. 27, pp. 187-207, 1999.

[16] Y. -L., Shue, P., Keating, C., Vicenik, and K., Yu, “VoiceSauce: A program for voice analysis,” *J. Acoust. Soc. Am.*, vol. 126, pp. 2221, 2009.

[17] K. N., Stevens, and H. M., Hanson, “Classification of glottal vibration from acoustic measures,” in *Vocal Fold Physiology: Voice Quality Control*, pp. 147-170, edited by O. Fujimura and M. Hirano (San Diego: Singular Publishing), 1995.

[18] J. -J., Kuang, and P., Keating, “Glottal articulation of phonation contrast and their acoustic and perceptual consequence,” *UCLA working paper in Phonetics*, vol. 111, pp. 123-161, 2012.