

THE ROLE OF DURATION IN THE PERCEPTION OF CHECKED VERSUS UNCHECKED TONES IN TAIWANESE SOUTHERN MIN

Wei Zhang¹, Yu-An Lu²

¹Department of Linguistics, McGill University, Canada.

²Department of Foreign Languages and Literatures, National Yang Ming Chiao Tung University, Taiwan
wei.zhang16@mail.mcgill.ca; yuanlu@nycu.edu.tw

ABSTRACT

Taiwanese Southern Min has checked versus unchecked tonal contrast the perception of which has received limited investigation. This study examined the role of duration in perceiving this contrast in TSM, for both the mid and high registers. Identification experiments were carried out on stimuli sampled from ‘All-cue’ Continuum where all cues were proportionally manipulated and from ‘Duration’ Continuum where only the duration cue was manipulated. For the mid register, similar categorization patterns were found for Duration and All-cue Continua, indicating that duration is indeed used reliably to categorize checked vs. unchecked tones. Additionally, we observed a suppression effect of unchecked tone identification on tokens with glottalization, an accompanying cue to checked tones. For the high register, however, no categorical response was found on the Duration Continuum, suggesting that the reliance on duration cue is subtle and that glottalization cues guided the perception of checked vs. unchecked tones for the high register.

Keywords: Taiwanese Southern Min, checked tone perception, duration cue, glottalization cue

1. INTRODUCTION

Taiwanese Southern Min (TSM), also known as Taiwanese Hokkien, is the second popular language spoken in Taiwan [1]. TSM has a rich tonal system including seven lexical tones, two of which are checked tones occurring in CV[p, t, k, ?] syllables where the final stops are unreleased [2]. Unchecked tones, on the other hand, occur in CV(N) syllables. An example of the tonal system [3] is shown in Fig. 1. As can be seen, the checked tones (T5, T3, the numbers in the tone names indicate the pitch height [4], same below), are shorter than the unchecked tones (T55, T13, T51, T31, T33).

In recent years, there has been growing attention to the sandhi schemes of TSM [3], [5], [6], and the acoustic features of checked and unchecked tones [7]–[9]. However, to the best of our knowledge, investigations on how listeners perceive the checked vs. unchecked tone pairs in TSM has been limited.

There are two pairs of checked vs. unchecked tones for mid and high registers, respectively: T3 vs. T33 and T5 vs. T55. Phonologically, the distinction between checked and unchecked tones are the final unreleased stop. However, as shown in Fig. 1, acoustically, the duration difference is also very salient: unchecked tones are about twice as long as checked tones. Furthermore, previous studies using electroglottographic and acoustic data have shown that, while final stop closures may be absent, energy dipping and irregular glottal vibration can still be observed on the preceding vowels [8]. [3] also reported cases of coda deletion, in which over 80% of /ʔ/ codas were deleted, followed by /t/, /k/ and then by /p/, suggesting a possible increasing role of duration in checked tone perception. Similar to how voiceless stop codas in English which are often cued by a shorter preceding vowel and the presence of glottalization [10], the checked tones in TSM, which end with voiceless stops, are also shorter and often accompanied by glottalization. This study is aimed to investigate to what extent listeners make use of duration and spectral cues informed by the final stop in perceiving the TSM checked vs. unchecked contrast for both registers.

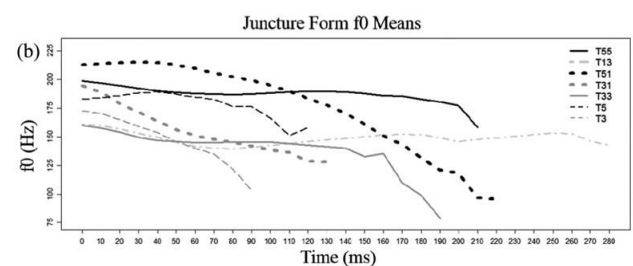


Figure 1: Tonal system of TSM (From Pan (2017: 38), the speaker was a female speaker over 40 years old. The numbers in the tone names indicate pitch height.)

There are relatively fewer investigations of the perception (than production) of checked tones or syllables (see [11] for a review). From the few investigations of checked tone perception in other languages/Chinese dialects, *F0*, *duration* and *voice quality* (such as *glottalization*) are the most important cues reported. Relative importance of the three cues varies across languages. In Burmese, voice quality and duration might be the distinguishing cues for the checked tone from the rest of the tones [12]. For both

Southern and Northern Vietnamese, [13] found that final-glottalization was a more important cue than F0 to the checked tone perception. For White Hmong, [14] found that listeners rely more on F0 and duration than glottalization for identifying checked tones. Additionally, for Xiapu Min, it was found that glottalization is less important than duration and F0 [11], and duration seems to be the most important cue [15] in the checked tone perception.

In terms of the checked vs. unchecked tonal contrast in TSM, as shown in Fig. 1, both the checked and unchecked tones have a falling contour acoustically. Additionally, [16] showed that when the durations are normalized, checked tones and unchecked tones exhibit very similar F0 contours, suggesting that F0 may not be an important cue to the checked vs. unchecked tonal contrast. With a special interest in duration, this study investigates the importance of the duration cue in perceiving checked vs. unchecked tones in TSM. We selected the CV[?] structure for the checked tones based on the aforementioned finding that [?] was more frequently deleted than [p, t, k] as the final coda of TSM checked tones. This may suggest that duration plays a relatively greater role in checked tones with a [?] coda than in those with the other three codas.

2. METHOD

2.1. Material

The segment was controlled for both checked and unchecked tones as well as for both registers. Specifically, the syllable /la/ was used for T33 (‘拈’, meaning ‘stir’) and T55 (‘拉’, meaning ‘pull’) while /laʔ/ was used for T3 (‘垃’, meaning ‘garbage’) and T5 (‘蠟’, meaning ‘wax’). A female speaker of TSM recorded each of the four tones/syllables naturally, in a sound-attenuated booth. From these recordings, one sample of each tone was selected based on less perturbation and a smoother amplitude envelope. These selected samples were then used to generate checked-unchecked tonal continua.

To evaluate the importance of duration in checked tone perception, we created two types of tonal continua. Type 1 is created by manipulating only the duration of the syllable between the checked and unchecked tones, named Duration Continuum hereafter. The Duration Continuum was created in two ways using the Pitch Synchronous Overlap and Add (PSOLA) algorithm in Praat [17]): (1) by increasing the duration based on the checked tone sample, and (2) by decreasing the duration based on the unchecked tone sample. The main difference between the two methods was whether vowel glottalization naturally contained in the checked

stimuli was preserved or not. Type 2 is created by using the checked and unchecked tones as endpoints, hence every contributing cue should be proportionally manipulated. This is called All-cue Continuum hereafter. Tandem-Straight [18], a speech analysis, modification and resynthesis framework, was used to create the All-cue Continuum. The spectrograms of the endpoint samples are shown in Fig. 2. To assess the extent to which TSM listeners rely on *duration* in categorizing checked and unchecked tones, we compared these listeners’ categorization of the same tonal pair in the two types of continua, one only differs in duration while the other in all relevant acoustic cues.

Altogether there were six continua used in this study: Duration Continuum based on T3, Duration Continuum based on T33, All-cue Continuum of register 3, and the same three continua for register 5.

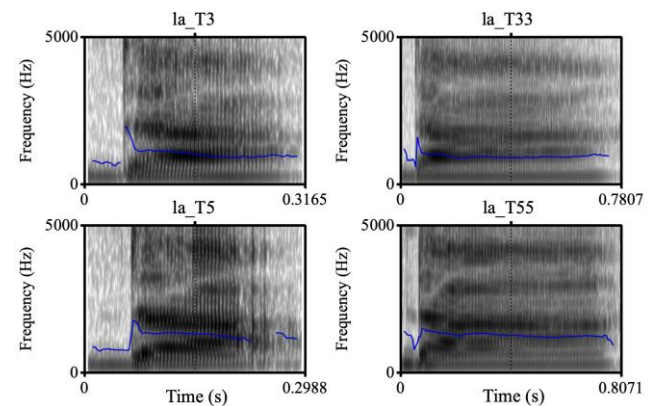


Figure 2: Spectrograms of base samples (top left: T3; top right: T33; bottom left: T5; bottom right: T55).

There were seven steps in each continuum. The duration of the checked and unchecked tone endpoints are 160 ms and 340 ms (~30 ms difference each step), respectively. The duration values were selected according to the authors’ previous production results of TSM tones and adjustment was made for balancing the categorical boundary to the middle steps.

2.2. Procedure and Participants

We recruited 17 TSM speakers (10M, 7F, $M = 23$ yrs, $sd = 1.6$) at Experimental Phonology Lab, National Yang Ming Chiao Tung University for the experiment. Two-alternative forced choice paradigm was used. Since there were twice as many stimuli on the Duration Continua as on the All-cue Continua, we doubled the number of stimuli on the All-cue Continua. Hence there were 8 continua in each block and the block repeated four times. Participants were allowed to take a break after each block. Altogether there were 224 (8[continua]*7[steps]*4[blocks]) trials in the experiment. A short instruction with audio

examples and 6 trials were provided before the experiment for practice.

Before the main task, the participants were asked to complete a questionnaire on their TSM, such as age of acquisition, frequency of use, listening level and speaking level (assessed on a 1-7 scale).

2.3. Statistical analysis

We excluded participants who self-reported their TSM listening level as 3 or lower (4 participants). For the rest of the data, statistical analysis was carried out with Bayesian mixed effects logistic regression, implemented using *brms* [19] in R [20]. The dependent variable was the response of the tone type (checked T3 and T5 were mapped to 0 and unchecked T33 and T55 were mapped to 1). Three fixed predictors were used: *step* index of the continuum (scaled and centered), manipulation *method* (treatment coded: All-cue Continuum was taken as the reference level), and *register* (register 3 as the reference level). The interactions between any two of them were included. By-participant random effects were specified as random intercept with the three fixed effects as random slopes.

3. RESULTS

The categorization results for each continuum are visualized in Fig. 3. Panel A and B of Fig. 3 show the patterns of responses for register 3 and register 5, respectively. For register 3, the results of the Duration Continuum show that increasing the duration between the checked and unchecked tones (moving from left to right on the x-axis) resulted in a gradual shift towards the unchecked tone, similar to the pattern observed on the All-cue Continuum, in which all cues were proportionally manipulated. In terms of register 5, however, no categorical response pattern was observed in the two Duration Continua. This suggests that, for the T3-T33 contrast but not for the T5-T55 contrast, duration plays a relatively larger role in distinguishing between checked and unchecked tones.

Results of the mixed effect logistic Bayesian model are presented in Table 1. *Rhat* values of all effects are 1, indicating the model does not show convergence issues. The probability of direction (*pd*) is computed using the R package *bayestestR* [21]. A *pd* value larger than 95% indicates a credible effect.

In Table 1, larger step index credibly increases the unchecked tone responses ($\beta = 4.25$, $pd = 100\%$), consistent with the expectation that unchecked tones are longer, when other predictors hold constant. Compared to the All-cue Continua, the Duration Continua based on checked tones credibly decrease the unchecked tone response ($\beta = -0.74$, $pd = 95\%$), whereas the Duration Continua based on unchecked

tones do not show credible effect ($\beta = 0.53$, $pd = 91\%$). On average, register 5 has credibly less unchecked tone responses ($\beta = -0.68$, $pd = 95\%$).

The interaction between step and method, and the interaction between method and register are all credible. Comparison results of ‘*step:method*’, using the *emmeans* package [22], revealed that step contributes a credible larger effect to the unchecked tone response in All-cue Continua than in Duration Continua based on checked tones ($\beta = 1.3$; *hpd*: [0.4, 2]), and a credible smaller effect to the unchecked tone response in All-cue Continua than in Duration Continua based on unchecked tones ($\beta = -2.6$; *hpd*: [-3.4, -1.8]).¹

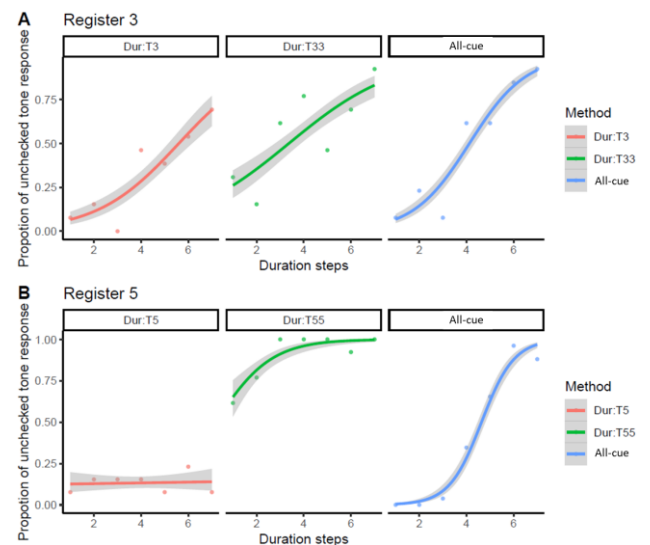


Figure 3: Tonal categorization results, grouped by register (top: Register 3; bottom: Register 5) and colored by manipulation method (left: Duration Continua based on checked tones; middle: Duration Continua based on unchecked tones; right: All-cue Continua). The abbreviation, for example, ‘Dur:T3’ indicates the Duration continuum based on T3 sample.

Comparison results of ‘*method:register*’ showed firstly that there is no credible difference between Duration Continuum based on T3 and All-cue Continuum of register 3, or between Duration Continuum based on T33 and All-cue Continuum of register 3, indicating that duration manipulation elicits comparable response as manipulating all cues in perception for the mid register. However, the effect of Duration Continuum based on T3 was credibly smaller than that of Duration Continuum based on T33 ($\beta = -1.3$; *hpd*: [-2.6, 0]), suggesting that the glottalization, which exists in the T3 base but not in the T33 base, plays a role in perception and reduces the unchecked tone response. Secondly, for register 5, there is a credible difference between any two of the three continua. Specifically, All-cue Continuum of register 5 shows a credible larger effect on the unchecked tone responses than the Duration Continuum based on T5 ($\beta = 1.7$; *hpd*: [0.8, 2.5]), and

a credible smaller effect than the Duration Continuum based on T55 ($\beta = -4.6$; hpd : [-5.6, -3.7]). The effects indicate that, in contrast to the mid register, duration manipulation does not elicit any comparable response as manipulating all cues for the high register. Thirdly, when comparing across registers, the Duration Continuum based on T3 has a credible larger effect than that based on T5 ($\beta = 1.7$; hpd : [0.7, 2.6]). Specifically, the Duration Continuum based on T33 has a credible smaller effect than that based on T55 on the response variable ($\beta = -3.4$; hpd : [-4.5, -2.4]). These results suggest again that glottalization plays a predominant role in checked tone perception for the high register.

	β	l-95% CI	u-95% CI	pd
Intercept	-0.13	-1.04	0.78	62 %
step	4.25	2.2	5.89	100%
methodChecked	-0.74	-1.52	0.18	95%
methodUnchecked	0.53	-0.28	1.29	91%
register	-0.68	-1.54	0.18	95%
step:methodChecked	-3.33	-3.98	-2.68	100%
step:methodUnchecked	-2.38	-3.08	-1.67	100%
step:register	0.17	-0.41	0.76	72%
methodChecked:register	-0.98	-1.53	-0.43	100%
methodUnchecked:register	4.1	3.42	4.82	100%

Table 1: Results of the mixed-effect logistic Bayesian model. Probabilities of direction (pds) of credible effects are bolded.

4. DISCUSSION

This study investigated the perception of the checked vs. unchecked tonal contrast in TSM. As the spectral cue, glottalization, is disappearing with the deletion of syllable coda in checked tones, this study examined whether duration shows an increased cue weight. Through identification experiments on All-cue and Duration Continua for both registers, we found that for the mid tones, duration is indeed used reliably to identify checked vs. unchecked tones based on the linear function of their responses as the duration step increased. The tokens without glottalization were categorized similarly as the all-cue tokens. We did see a suppression effect of unchecked tone identification on tokens with glottalization cues. The reliance on duration cue, however, was subtle for the high tones. It appears that glottalization cues guided the perception of checkedness for the high tones.

We provide some possible sources of why duration plays different roles in perceiving the checked tone for the two registers. First, literature on TSM revealed that coda [ʔ] was deleted more often in the high checked tone than in the mid checked tone [23-24, cited after 25], which could result in different weightings of glottalization and duration for the two

registers in general. Second, [25] found that in TSM when the deletion of [ʔ] happened, the preceding vowel was lengthened for the high tones, causing a possible disconnection between the high checked tone and shorter duration. For the mid tones, however, the duration remained short after the deletion of the coda, strengthening the connection between the mid checked tone and shorter duration. That is, the duration cue may not be reliable in categorizing high checked vs. unchecked tones but remains reliable for the mid register. Third, when creating the stimuli, we used separate base samples for the two registers, hence the degree of glottalization was not strictly controlled across the registers. As shown in Fig. 2, there seems to be more glottalization for ‘la_T5’ than for ‘la_T3’, which might bring about larger glottalization effect in the Duration Continuum based on T5 than on T3. However, this difference could not be the only reason accounting for the asymmetrical results between Duration continua based on T33 and T55, since both continua did not contain glottalization. What might be a reason instead is the psycho-acoustic effect: a sound with higher pitch is perceived longer than with lower pitch when their physical durations are equal [26-27]. Hence, steps in Duration Continuum based on T55 can be perceived longer (than steps in the continuum based on T33), which may increase the unchecked tone responses. As shown in Fig. 3, as duration decreases from step 4 to 1, there are increasing checked tone responses. If even shorter durations were tested, the categorical pattern might emerge, and this direction awaits further investigation.

The results of this study may have implications for the tone merger in TSM. The two checked tones are merging and the final codas are undergoing loss [1, 3]. The confirmed role of duration in distinguishing between the checked and unchecked tones, along with the ongoing trend of stop coda deletion [3], suggests a possible outcome of tone merging in TSM: a long vs. short tonal contrast. The deletion of stop codas suggests a decreasing phonological awareness of checkedness, while the confirmed role of duration in this study indicates awareness of quantity. It is possible that the contrast between checked and unchecked tones could be phonologized into a long vs. short contrast in TSM.

One limitation of this study is the relatively small sample size and the variability in the degree of glottalization across the different registers. Additionally, further investigation is needed to explore the perception of checkedness in the high register under shorter duration ranges, and the individual-level perception-production link on checkedness.

5. ACKNOWLEDGEMENTS

We would like to thank Yin-Ching Chang for her help on collecting data. This work was supported by Ministry of Science and Technology of Taiwan (MOST110-2628-H-A49-001-MY2) to Yu-An Lu.

6. REFERENCES

- [1] U. Ang, "The distribution and regionalization of varieties in Taiwan," *Language and Linguistics*, vol. 14, no. 2, p. 315, 2013.
- [2] J. Norman, *Chinese*. Cambridge University Press, 1988.
- [3] H. Pan, "Glottalization of Taiwan Min checked tones," *Journal of the International Phonetic Association*, vol. 47, no. 1, pp. 37–63, Apr. 2017, doi: 10.1017/S0025100316000281.
- [4] Y.-R. Chao, "A system of tone letters," *Le maître phonétique*, 1930.
- [5] C.-H. Kuo, "Perception and acoustic correlates of the Taiwanese tone sandhi group," PhD Thesis, UCLA, 2013.
- [6] Y.-F. Chien and A. Jongman, "Tonal Neutralization of Taiwanese Checked and Smooth Syllables: An Acoustic Study," *Lang Speech*, vol. 62, no. 3, pp. 452–474, Sep. 2019, doi: 10.1177/0023830918785663.
- [7] H. Pan, M.-H. Chen, and S.-R. Lyu, "Electroglottograph and acoustic cues for phonation contrasts in Taiwan Min falling tones," in *Twelfth Annual Conference of the International Speech Communication Association*, 2011.
- [8] H. Pan, H. Huang, and S. Lyu, "Coda Stop and Taiwan Min Checked Tone Sound Changes.," in *INTERSPEECH*, 2016, pp. 1011–1015.
- [9] H. Pan, "Voice quality of falling tones in Taiwan Min," in *Ninth European Conference on Speech Communication and Technology*, 2005.
- [10] J. Penney, F. Cox, and A. Szakay, "Weighting of Coda Voicing Cues: Glottalisation and Vowel Duration.," in *INTERSPEECH*, 2018, pp. 1422–1426.
- [11] Y. Chai, "Phonetics and Phonology of Checked Phonation, Syllables, and Tones," Ph.D., University of California, San Diego, United States -- California, 2022. Accessed: Sep. 30, 2022. [Online]. Available: <https://www.proquest.com/docview/2714115642/abstract/26EACB3B47A6499FPQ/1>
- [12] J. F. Gruber, *An articulatory, acoustic, and auditory study of Burmese tone*. Georgetown University, 2011.
- [13] M. Brunelle, "Tone perception in northern and southern Vietnamese," *Journal of Phonetics*, vol. 37, no. 1, pp. 79–96, 2009.
- [14] M. Garellek, P. Keating, C. M. Esposito, and J. Kreiman, "Voice quality and tone identification in White Hmong," *The Journal of the Acoustical Society of America*, vol. 133, no. 2, pp. 1078–1089, 2013.
- [15] Y. Chai, "Perception of checked tones in Xiapu Min," *The Journal of the Acoustical Society of America*, vol. 150, no. 4, pp. A309–A310, Oct. 2021, doi: 10.1121/10.0008390.
- [16] M.-H. Lin, "Pitch and duration reflexes in Taiwanese Southern Min tones across generations.," M.A., National Yang Ming Chiao Tung University, Hsinchu, 2022.
- [17] P. Boersma, "Praat: doing phonetics by computer [Computer program]. Version 6.1.16," <http://www.praat.org/>, 2020.
- [18] H. Kawahara, "STRAIGHT, exploitation of the other aspect of VOCODER: Perceptually isomorphic decomposition of speech sounds," *Acoustical science and technology*, vol. 27, no. 6, pp. 349–353, 2006.
- [19] P.-C. Bürkner, "brms: An R package for Bayesian multilevel models using Stan," *Journal of statistical software*, vol. 80, no. 1, pp. 1–28, 2017.
- [20] R Core Team, *R: A language and environment for statistical computing*. Vienna, Austria, 2013.
- [21] D. Makowski, M. S. Ben-Shachar, and D. Lüdtke, "bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework," *Journal of Open Source Software*, vol. 4, no. 40, p. 1541, 2019.
- [22] H. Singmann, J. Love, P. Buerkner, and M. Herv, "Package 'emmeans,'" *CRAN R Project. Available online: https://cran.r-project.org/web/packages/emmeans/emmeans.pdf (accessed on 10 April 2020)*, 2019.
- [23] U. Ang, "The motivation and direction of sound change: On the competition of Minnan dialects Chang-chou and Chuan-chou, and the emergence of General Taiwanese," *Unpublished PhD dissertation, National Tsing Hua University, Taiwan*, 2003.
- [24] S. Chen, "New Sound Variation in Taiwan Southern Min: Vowel Systems and the Lower Register Entering Tone in Taipei, Changhua, and Tainan," *Language and Linguistics*, vol. 11, no. 2, pp. 425–468, 2010.
- [25] H. Pan and S. Lyu, "Taiwan Min Nan (Taiwanese) Checked Tones Sound Change," in *Interspeech 2021*, Aug. 2021, pp. 2641–2645. doi: 10.21437/Interspeech.2021-672.
- [26] C. Gussenhoven and W. Zhou, "Revisiting pitch slope and height effects on perceived duration," in *INTERSPEECH 2013: 14th Annual Conference of the International Speech Communication Association*, 2013, pp. 1365–1369.
- [27] Y.-A. Lu and S.-I. Lee-Kim, "The effect of linguistic experience on perceived vowel duration: Evidence from Taiwan Mandarin speakers," *Journal of Phonetics*, vol. 86, p. 101049, May 2021, doi: 10.1016/j.wocn.2021.101049.

¹ (HPD interval probability: 0.95. An HPD interval excluding 0 is assumed to indicate a credible effect.)