Checked Tone Merger in the Nanjing Dialect: An Acoustic Analysis

Christopher Oakden
Rutgers, The State University of New Jersey

This study provides new evidence relating to a reported tonal merger for younger speakers of the Nanjing Dialect, a Lower Yangzi Mandarin dialect. Recent impressionistic accounts (Song 2009, Gu 2015) report an in-progress merger of checked tone syllables containing a glottal stop coda, first by coda deletion and then syllable lengthening. In the current study, an acoustic analysis of checked tone syllable realizations was performed on a younger speaker of the dialect. Rime duration and vowel creak data extracted from recordings conflict with earlier characterizations of the merger. Lack of vowel creak (glottal coda residue) on canonically checked syllables indicates outright elimination of glottal stops, while a significant duration difference between checked tone syllables and other lexical tones is retained.

0. Introduction

A topic of enduring interest in discussions of Chinese phonology is the status of checked tone syllables, generally distinguished by the presence of an occlusive coda (Chen 2000:5) rather than a specific tonal melody. Understood as a relic of Middle Chinese phonology, the checked or entering tone (*rusheng*) category is attested across the majority of modern Chinese dialect families, but is mostly absent from the Mandarin dialects. Certain Mandarin subdialect classes do retain it, however, such as Lower Yangzi Mandarin or *jianghuai guanhua.*

One such example is the Nanjing dialect (Fei & Sun 1993, Liu 1995; henceforth NJD); earlier accounts describe it as a five-tone system, with one lexical tone category populated by checked tone syllables containing a glottal stop coda. Recent studies of the dialect (Song 2009; Gu 2015) suggest, though, that checked tone syllables have begun merging into other lexical tone categories, first by coda deletion and subsequent syllabic lengthening. Gu (2015) reports that the merger is not uniform across generations of speakers, and is reported to be more advanced in younger speakers (those between the ages of 16 and 25 years old).

The current study investigates the reported merger in NJD with the aid of a phonetic experiment. Specifically, it quantifies the extent of the merger by narrowing in on two acoustic parameters that distinguish checked tone from other tone categories: rime duration and vowel creak. Two main results of the experiment disagree with earlier impressionistic accounts of the merger; while a significant duration contrast remains

between checked tone rimes and those of other lexical tones, there is no evidence of creak on vowels to suggest residue from a glottal coda.

The remainder of the paper is organized as follows: Section 1 introduces the tonal phonology of the Nanjing dialect and reviews earlier reports of merger, focusing on the predictions these accounts make in regard to rime duration and vowel creak. Section 2 outlines the design of the acoustic experiment and discusses the necessary controls implemented. Data analysis and results from the experiment are presented in Section 3, while their implications and directions for further research are explored in Section 4. Section 5 concludes.

1. Background
1.1 Nanjing Dialect Tone

This study focuses on the Nanjing dialect, a Lower Yangzi Mandarin dialect spoken in Nanjing, the capital of Jiangsu province, and surrounding districts. Its tonal inventory consists of five lexical tones. One is a checked tone, distinguished by high, level pitch, short duration, and the presence of a glottal stop coda. Middle Chinese (MC) categories are given a numeral designation, which will be used throughout the remainder of the paper.

MC Category	Description	Melodic Representation	Chao Tone Letter
Yinping (1)	Falling contour	[HL]	31
Yangping (2)	Rising contour	[LH]	24
Shangsheng (3)	Low level	[LL]	11
Qusheng (4)	High level	[HH]	44
Rusheng (5)	High level checked	[H]	5

Table 1: Tonal Inventory of Nanjing Dialect

Earlier descriptive accounts of NJD make no mention of the instability of the checked tone category (Fei & Sun 1993, Liu 1995); however, recent accounts (Song 2009, Gu 2015) report that checked tone syllables have begun to merge into other lexical tone categories for NJD speakers.

1.2 Checked Tone Merger

What sets checked tone merger apart from other tonal merger processes is that it entails more than a simple change in melodic realization of syllables. Instead, checked tone merger is a complex of three primary contrast neutralizations, described schematically below.

(1) a. syllable shape (presence of occlusive coda vs. absence)
     b. duration (short *cusheng* vs. long *shusheng*)
     c. melodic realization (5 vs. 31, 24, 11, 44)

As a result, a degree of possible variability exists in terms of how the merger could proceed.

Two previous studies have identified an in-progress merger for NJD checked tone syllables, each decomposing the trajectory of the merger into distinct stages. These stages make predictions about the phonetic realization of checked tones as the merger progresses, which allows for testable hypotheses regarding the correlation of certain acoustic cues.

Song (2009) identifies three stages of merger. The first stage comprises glottal coda weakening accompanied by gradual lengthening of the rime. By the second stage, the glottal stop coda has neutralized completely, and duration of checked syllables becomes indistinguishable from those of non-checked syllables. However, these syllables retain a unique melodic realization. The final stage predicts complete merger into other tonal categories producing a four-tone system. Song argues that the current state of the dialect is at an intermediary phase between stages one and two.

A more recent impressionistic study by Gu (2015) also creates a three-way partition of the merger process, but narrows its focus on both generational differences among speakers and the effects of Standard Mandarin on diachronic change. In the analysis, the first stage represents a pre-merger state characterized by a robust glottal coda and a clear durational contrast between checked and non-checked tones. This is followed by coda weakening/neutralization and rime lengthening. No distinct melodic realization remains at this point of the merger; formerly checked syllables assume the tonal melodies of other lexical tones in the dialect. The final stage proposed in (Gu 2015) involves a total supplanting of NJD melodies on checked tone syllables with Standard Mandarin melodies. In that study, survey participants between the ages of 16 and 25 showed signs of being at a more advanced stage of the merger than older speakers.

Despite differences in the projection of merger, both studies converge on the same crucial generalization regarding the glottal coda and rime duration: lengthening accompanies coda neutralization. Phonetically, the correlation of these two phenomena is expected; lack of full glottal closure results in concomitant creak on the preceding vowel and compensatory lengthening (Garellek 2013).

These generalizations give rise to testable hypotheses. If earlier accounts are on the right track, there should be a clear correlation between glottalization and duration such that checked syllables with no evidence of a glottal coda (or even residue) should be indistinguishable from non-checked syllables with respect to rime duration. In addition, this effect should be particularly evident in younger speakers of the dialect, who are reported to be at a relatively advanced stage of merger.

2. Experiment

An acoustic experiment was designed and implemented to test the hypotheses regarding NJD checked tone merger generalized from the results of Song (2009) and Gu

(2015). The primary goal was to explore the interaction of duration and glottalization on checked syllables in the speech of younger NJD speakers. To achieve this goal, a native speaker was recorded producing NJD's five lexical tones. Rime duration was compared across lexical tone categories, and canonically checked syllables were examined for evidence of glottalization.

To ensure maximally accurate duration and glottalization results, a number of controls were applied in the process of selecting target words. Targets were limited to [CV] syllable shape to aid in segmentation and minimize potential durational effects (Turk et al. 2006, van Santen 1992). Syllables with onglides [CGV], nasal codas [CVN], or a combination of the two [CGVN] were excluded from consideration due to the difficulty of identifying precise [GV] and [VN] transitions. Additionally, only targets with the low, back vowel [a] as a rime were included to control for inherent durational differences between vowels (Toivonen et al. 2015 and numerous sources cited therein). Table 2 includes the number of tokens for each tonal category and total number of checked/non-checked target words (see Appendix A for a complete list of targets).

Tone Category	Number of Tokens		
Tone 1	8		
Tone 2	3		
Tone 3	5		
Tone 4	8	*Non-Checked Tokens*	*Checked Tokens*
Tone 5	16	**24**	**16**

Table 2: Number/Type of Tokens

In addition to the controlled targets, sixteen fillers (eight with checked tones and eight with non-checked tones) of varying syllabic structure and segmental content were added to the list of tokens.

Two frame sentences were used to illicit vernacular speech production. The sentences described an informal classroom setting in which the subject was asked to imagine a casual conversation among students during a break between class periods. When a Chinese character is written on the blackboard, the participants in the conversation observe that the character is written in an ugly manner. Subjects read the target word in both frame sentences:

(2) a. 黑板　上　　写 了 一个 ___ 字。

   heiban shang xie le yige ___ zi
   'On the blackboard is written a ___ character.'

  b. 这个　___　字 写 滴 有　点儿　丑。

   zhege ___ zi xie di you dianr chou
   'This ___ character is written a bit ugly.'

The rationale behind selecting a written character is that it allows for isolation of a single syllable in a natural context, and facilitates construction of a frame sentence that equally accommodates any syllable, regardless of part of speech. Targets were placed sentence-medially in each frame sentence to control for phrase-final lengthening.

Three repetitions of the targets and fillers within the frame sentences in (2) were recorded for the experiment, totaling 144 non-checked tokens and 96 checked tokens. The order of targets was randomized and counterbalanced in each repetition, with fillers evenly spaced among stimuli. Additionally, five fillers were placed at the beginning of each repetition to accommodate any initial hyper- or mis-articulation in case the subject felt anxious about the task.

The participant in this experiment was a native NJD speaker born in Nanjing, Jiangsu, and grew up speaking the language at home. At the time of recording, the participant—a first-year undergraduate student at Rutgers University—had been living in the United States for approximately two months, but still spoke NJD daily with classmates and family members. The participant had no training in linguistics and no history of speech impairment.

Recording took place inside a sound-attenuated booth at the Phonology and Field Research Laboratory at Rutgers University. The subject was fitted with an AKG C420 head-worn microphone with behind-the-neck headband to maintain a constant distance from the mouth. The recording was made using *GoldWave* v6.10 software at a 44.1k Hz sampling rate and 16-bit quantizing rate in mono. A visual aide to the prompts was projected on a laptop screen that the subject scrolled through manually as the experiment progressed. The prompt contained an ugly hand-written character on an image of a blackboard, with the character in standard font to the left, and the frame sentences at the bottom of the screen for reference, as in (3).

(3)
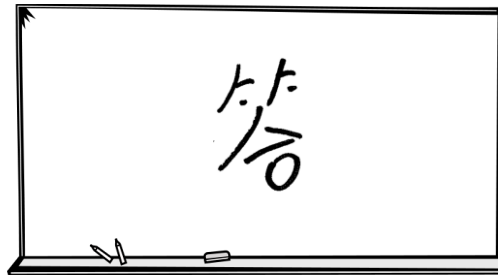
黑板上写了一个__字。这个__字写滴有点儿丑。

All three repetitions were completed during a single recording session with a 10-15 minute break between each repetition. The next section outlines processing and analysis procedures for the data collected, as well as the results from statistical analysis.

3. Data Analysis and Results

After data collection concluded, the sound files were annotated using TextGrids in Praat (Boersma & Weekik 2015). Each file was labeled with four intervals: the vowel of the target syllable in both frame sentences, as well as a fixed interval in each frame sentence for the purpose of duration normalization. The left edge of the target vowel was identified by placing a boundary at the zero crossing of the first periodic, non-deformed waveform of the vowel (Francis et al. 2002); the right edge was determined by inserting a boundary at the end of the second formant on the vowel (Turk et al. 2006: 7). Duration and jitter (as a measure of vowel glottalization; see Koike 1973, Koike & Hirano 1973, Surana & Slifka 2006) figures were extracted from these labeled sound files with the use of customized Praat scripts. Analysis of both parameters and results are discussed separately below.

3.1 Duration Results

Duration measurements were normalized using a fixed point in each frame sentence to control for variation in speech rate. Table 3 outlines mean rime duration and standard deviation post-normalization.

Tone Category	Frame 1	Frame 2
Tone 1	212.7 (33.7)	161.8 (22.1)
Tone 2	214.1 (33.9)	156.8 (24.7)
Tone 3	212.8 (56.4)	167.4 (25.6)
Tone 4	292.1 (64.8)	182.5 (20.6)
Tone 5	101.8 (23.7)	105.6 (28.5)

Table 3: Mean Normalized Target Rime Durations (ms) and Standard Deviation

Tone 5 (checked) rimes were shorter than those of the other lexical tones; this is particularly stark in the first frame sentence, where checked rimes' duration was on average less than half that of other tones. The distribution of durational data across both frame sentences is represented graphically in (4).

(4)

146
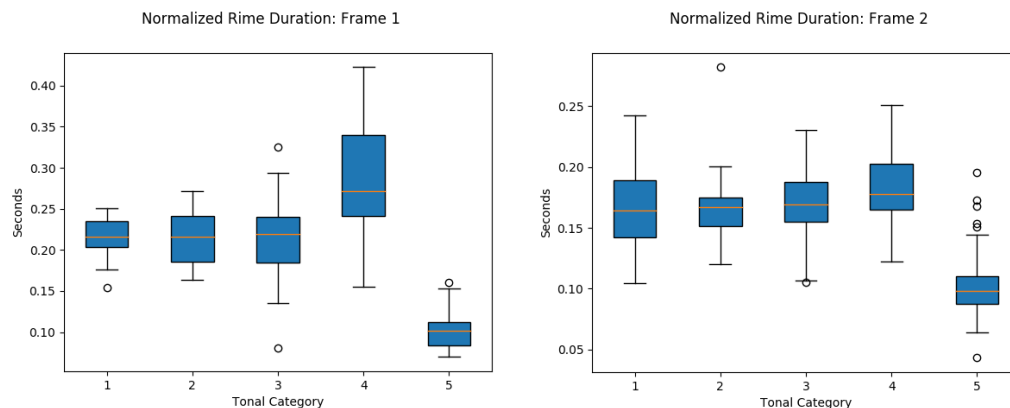
Two-sample Student's t-tests between checked tone rime duration measurements and those of other tonal categories confirmed impressionistic judgments about durational differences, showing significant effects ($p < 0.01$) for each test between checked tone and the non-checked tones. Since duration values were normalized by frame sentence, cross-sentential comparisons were not conducted. Within each frame sentence, rimes on checked tone syllables were shown to be significantly shorter than those of each of the other four lexical tones.

3.2 Jitter Results

To determine the degree of glottalization on the vowel in target syllables, Praat's Jitter(ddp) function values were extracted. This function measures frequency perturbations on the vowel (vocal creak) as differences in the durations of consecutive glottal periods. To calculate the degree of perturbation, Jitter(ddp) averages the absolute durational differences between consecutive intervals, then divides by the average interval, yielding a numerical value between 1 and 6. A higher value corresponds to a higher degree of vocal creak.

Table 4 below presents the mean jitter results and standard deviations for each NJD lexical tone across both frame sentences.

Tone Category	Frame 1	Frame 2
Tone 1	3.198 (1.47)	2.181 (1.06)
Tone 2	3.346 (1.35)	2.947 (1.22)
Tone 3	3.217 (1.84)	2.175 (1.33)
Tone 4	1.932 (1.23)	2.551 (2.23)
Tone 5	1.988 (1.39)	2.246 (1.77)

Table 4: Mean Jitter(ddp) Values and Standard Deviation

Jitter values for checked tone syllables (as well as Tone 4) are noticeably lower than Tones 1, 2, and 3 in the first frame sentence. Student's t-tests performed individually between checked tone and lexical tones 1, 2, and 3 each reveal differences that are statistically significant ($p < 0.01$). The same results were obtained between Tone 4 and lexical tones 1, 2, and 3. In the second frame sentence, no significant effects were observed; degree of vocal creak was comparable on vowels of all five lexical tones.

This result indicates that vowels on checked tone syllables do not show evidence of glottalization; they exhibit less creak than uncontroversially open CV syllables. If residue of glottal stop codas was still present on these vowels, a higher degree of vocal creak compared to non-checked syllables would be expected. Rather, the inverse was observed in the current study. The jitter values corroborate an advanced stage of glottal coda weakening in checked tone syllables, and are compatible with a total neutralization analysis.

The generalization that emerges from combining duration and jitter results is problematic for earlier hypotheses regarding the trajectory of NJD checked tone merger. Instead of a clear correlation between durational and glottalization effects, there is a durational contrast but no evidence of a robust glottal coda (or even residue of a glottal coda). The next section will discuss these results in greater detail as they relate to previous hypotheses, as well as suggest possible explanations for the results.

4. Discussion

The acoustic profile of checked tone syllable rimes established in this study does not confirm the hypothesis that predicts a correlation between glottal weakening and rime lengthening as a component of NJD checked tone merger. No evidence of glottal coda residue was found on checked tone syllables, suggesting outright glottal deletion; however, a clear durational contrast between checked and non-checked syllables was observed.

Among the properties which distinguish NJD checked tone syllables from non-checked tone syllables, the durational contrast was robust in the speech of the experiment's participant. Checked rimes in the first frame sentence were on average less than 50% of the duration of those from other lexical tone categories. Though this ratio was less extreme in the second frame sentence, the differences evident in both frames were determined to be statistically significant.

Jitter measures provided no evidence of vocal creak on checked tone rimes relative to non-checked tones. In fact, jitter values on tones 1, 2, and 3 were found to be significantly higher (indicating a higher degree of vocal creak) than both tones 4 and 5 in the first frame sentence. One possible explanation for this result is a correlation between vocal creak and [L] tone; precisely what distinguishes these tones is the presence or absence of [L] tone in the melodic representation. Examination of the spectrograms of tones of each type corroborates this hypothesis, with typical examples given in (5).
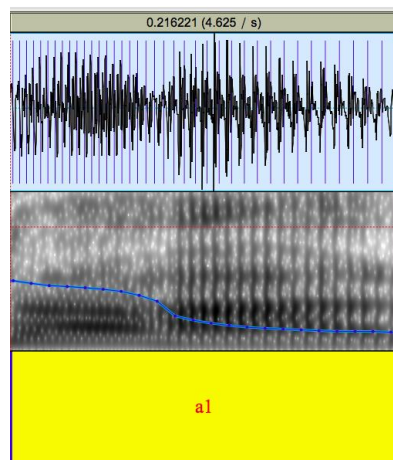
(5)   Tone 1 [HL]   Tone 5 [H]
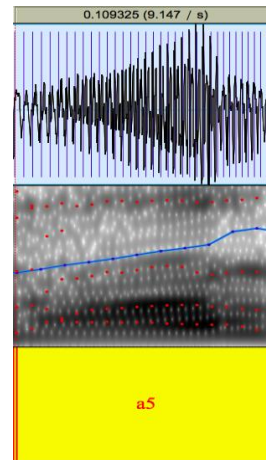
Striations in the spectrogram indicating vocal creak are apparent in the [L]-tone portion of tone 1 in (5); no such striations appear in the spectrograms of checked tone syllables. In terms of creak, then, checked tone syllables pattern with tone 4 [HH], an uncontroversially non-checked tone category.

A number of open questions remain regarding mid-merger NJD checked tone syllables and the nature of tone 5 as a lexical tone category in general. Chief among them is their melodic realization; earlier studies differ in their predictions about the pitch realization of NJD checked tone at different stages of merger. Song (2009) argues for initial retention of a unique melody after rime lengthening followed by merger into other NJD tonal categories. Gu (2015) predicts merger terminating in full adoption of Standard Mandarin melodies. The current study suggests a stage in merger—neutralization of syllabic shape contrast only—that does not fit either model. This obscures the possible predictions each hypothesis might make about the melodic realization of checked tone syllables, thus complicating any evaluation of the earlier predictions. I leave analysis of melodic realization, then, to future work.

5. Conclusion

This study has provided new evidence regarding checked tone merger in the Nanjing Dialect. An acoustic analysis performed on the recorded speech of a younger NJD native speaker found a robust durational difference between checked and non-checked tone syllables, but no evidence of glottal residue in checked tone syllables. These results conflict with previous accounts of the merger process, which predict a correlation between glottal coda weakening and rime lengthening in checked tone syllables. Thus, among the three contrast neutralizations inherent in checked tone merger, there is only evidence for the neutralization of contrast in syllable structure between checked and non-checked tone syllables.

REFERENCES

BOERSMA, PAUL and DAVID WEENINK. 2016. *Praat: doing phonetics by computer* (version 6.0.25). http://praat.org.

CHEN, MATTHEW C. 2000. *Tone sandhi: Patterns across Chinese dialects*. Cambridge, UK: Cambridge University Press.

FEI, JIA and SUN LI (eds.) 1993. *Nanjing fangyan zhi* [*Survey of the Nanjing dialect*]. Nanjing: Nanjing Publishing House.

FRANCIS, ALEXANDER; VALTER CIOCCA; and JOJO MAN CHING YU. 2002. Accuracy and variability of acoustic measures of voicing onset. *The Journal of the Acoustical Society of America* 113.2: 1025-1032.

GARELLEK, MARC. 2013. Production and perception of glottal stops. PhD Dissertation, University of California Los Angeles.

GU, QIAN. 2015. Nanjing fangyan dangjiang liangshe rushengyun de gongshi bianyi ji shili yanbian yanjiu [On synchronic variation and diachronic evolution of the entering tone in the dang and jiang classes in the Nanjing dialect]. *Linguistic Sciences* 14.4: 384-393.

KOIKE, YASUO. 1973. Application of some acoustic measures for the evaluation of laryngeal dysfunction. *Studia Phonologica* 7: 17-23.

KOIKE, YASUO and HIRANO MINORU. 1973. Glottal-area time function and subglottal pressure variation. *The Journal of the Acoustical Society of America* 54: 1618.

LIU, DANQING. 1995. *Nanjing fangyan cidian* [Dictionary of the Nanjing dialect]. Nanjing: Nanjing Publishing House.

SONG, YIDAN. 2009. Nanjing fangyanzhong de rusheng housaiwei shiyan yanjiu [Experimental research on entering tone in the Nanjing dialect]. *Journal of the School of Chinese Language and Culture Nanjing Normal University* 2009.2: 166-172.

TOIVONEN, IDA; LEV BLUMENFELD; ANDREA GORMLEY; LEAH HOITING; JOHN LOGAN; NALINI RAMLAKHAN; and ADAM STONE. Vowel height and duration. Proceedings of the 32nd West Coast Conference on Formal Linguistics, edited by Ulrike Steindl et al., 64-71. Somerville, MA: Cascadilla Proceedings Project.

TURK, ALICE; SATSUKI NAKAI; AND MARIKO SUGAHARA. 2006. Acoustic segment durations in prosodic research: a practical guide. *Methods in empirical prosody research*, edited by Stefan Sudhoff, Denisa Lenertova, Roland Meyer, Sandra Pappert, Petra Augurzky, Ina Mleinek, Nicole Richter and Johannes Schlieber. Berlin: Mouton de Gruyter.

SURANA, KUSHAN and JANET SLIFKA. 2006. Acoustic cues for the classification of regular and irregular phonation. Proceedings of INTERSPEECH 2006 and the 9th International Conference on Spoken Language Processing, 693-696. International Speech Communication Association.

VAN SANTEN, JAN P. 1992. Contextual effects on vowel duration. *Speech Communication* 11: 513-546.

Appendix A: Target List

Character	IPA	Tone	Melody	Gloss
八	[pa]	5	H	'eight'
拔	[pa]	5	H	'pluck'
罚	[fa]	5	H	'punish'
法	[fa]	5	H	'law'
发	[fa]	5	H	'launch'
乏	[fa]	5	H	'lack'
达	[ta]	5	H	'arrive'
搭	[ta]	5	H	'put up'
答	[ta]	5	H	'answer'
塌	[tʰa]	5	H	'collapse'
杀	[sa]	5	H	'kill'
杂	[tsa]	5	H	'mix'
撒	[sa]	5	H	'sprinkle'
擦	[tsʰa]	5	H	'scrub'
扎	[tsa]	5	H	'tie up'
踏	[tʰa]	5	H	'step on'
巴	[pa]	1	HL	*used in names*
他	[tʰa]	1	HL	'he'
渣	[tsa]	1	HL	'crumbs'
叉	[tsʰa]	1	HL	'fork'
沙	[sa]	1	HL	'sand'
纱	[sa]	1	HL	'yarn'
莎	[sa]	1	HL	*used in names*
砂	[sa]	1	HL	'grit'

爬	[pʰa]	2	LH	'climb'
茶	[tsʰa]	2	LH	'tea'
查	[tsʰa]	2	LH	'check'
打	[ta]	3	LL	'hit'
傻	[sa]	3	LL	'stupid'
把	[pa]	3	LL	'grasp'
塔	[tʰa]	3	LL	'pagoda'
卡	[kʰa]	3	LL	'get stuck'
爸	[pa]	4	HH	'father'
霸	[pa]	4	HH	'bully'
怕	[pʰa]	4	HH	'fear'
大	[ta]	4	HH	'big'
榨	[tsa]	4	HH	'press'
差	[tsʰa]	4	HH	'bad'
耙	[pa]	4	HH	'shovel'
坝	[pa]	4	HH	'dam'