WORKING PAPERS

IN

LINGUISTICS

The notes and articles in this series are progress reports on work being carried on by students and faculty in the Department. Because these papers are not finished products, readers are asked not to cite from them without noting their preliminary nature. The authors welcome any comments and suggestions that readers might offer.

Volume 45(3)
May
2014

DEPARTMENT OF LINGUISTICS
UNIVERSITY OF HAWAI‘I AT MĀNOA
HONOLULU 96822

DEPARTMENT OF LINGUISTICS FACULTY
2014

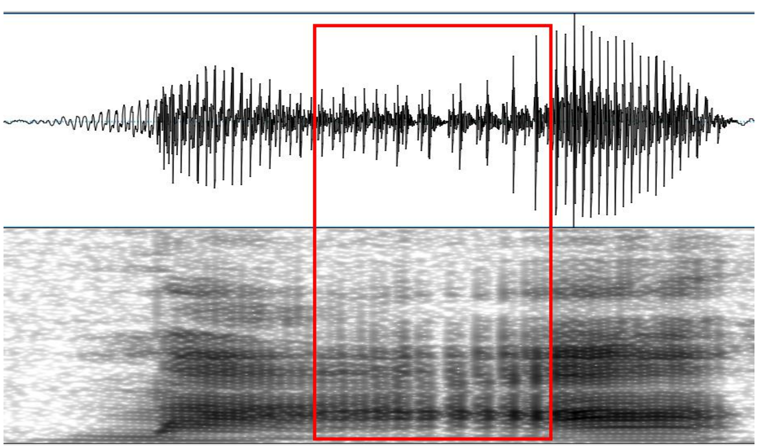
EFFECT OF CREAKY VOICE SIMULATION ON THIRD-TONE PERCEPTION IN MANDARIN CHINESE

JOHN VAN WAY

Mandarin tones have traditionally been described by differences in pitch. However, the second tone (rising) and third tone (low/dipping) are perceptually close and have been shown to be a problematic pair for L1 and L2 Mandarin learners. Creaky voice has been observed in third-tone production, and has been shown to aid accurate identification. This study finds that creaky voice resynthesized in second-tone tokens can cause third tone identification, which leads to a reconsideration of the role of phonation, in comparison with pitch, when listeners distinguish these two tones.

1. INTRODUCTION. The object of this study is an examination of how voice quality, creaky voice in particular, interacts with the perception of tone in Mandarin. Mandarin Chinese is widely-known as a textbook example of a "tone language." Standard Mandarin (Putonghua) has four contrastive tones, which are characterized by relative pitch (fundamental frequency, F0) levels or trajectories: T1 - high (55), T2 - rising (35), T3 - low/dipping (21(4)) and T4 - falling (51).[1] T3, described as 'low' and/or 'dipping,' is subject to phonological and morphophonological rules that determine its surface pitch realization. The rising component of the tone (the "4" in the "214") only surfaces when the word is utterance-final or in citation. Additionally, when T3 is followed by another T3, it changes to T2. Many studies have shown that T3 presents the most difficulty for both L1 and L2 learners of Mandarin, perhaps due to this allotony. Alongside the pitch differences, creaky voice is observed in T3 productions as well. See figure 1 for an illustration.

FIGURE 1. The (T3) word 吗 *ma*214 ('horse') uttered in isolation by a female native speaker of Mandarin.



Creaky voice is produced by a tightening of the vocal folds as compared to modal voice. The arytenoid cartilages hold the vocal folds tightly at one end so that they no longer vibrate as a whole (Ladefoged 2006, Ladefoged and Maddieson 1996). The result is a slower, more irregular vibration of the vocal folds. Acoustically, these irregular glottal pulses characterize creaky voice and can be identified clearly in a waveform, as in figure 1. Specifically, creaky voice manifests acoustically as "jitter" (cycle-to-cycle variations in pitch period or F0), "shimmer" (cycle-to-cycle variations in pulse amplitude) and a

[1] These numbers (5 = highest; 1 = lowest) are standard denotations for relative pitch levels observed in East Asian tone languages.

lower "spectral tilt" (ratio between the first and second harmonics of the vocal source) when compared to modal voice (Johnson 2012).

Languages employ creaky voice in different ways—some use it phonemically and others for non-phonemic uses. Many languages of East and Southeast Asia, where tones are widespread, exhibit phonemic uses of voice quality, such as breathiness or creakiness, alongside other laryngeal features like pitch. Languages such as Khmer use voice quality completely independent of any tone system (Huffman 1967), while others (e.g., Cantonese, Bai, Burmese) use voice quality along with pitch to contrast different tones (Yu and Lam 2011, Edmondson and Shaoni 1994, Bradley 1982). Other languages, in which voice quality is non-phonemic, employ non-modal phonation for various purposes. Particular attention has been given to the social and pragmatic uses of creaky voice in English. Cross-linguistically, it has been found that females use creaky voice more than men (Klatt and Klatt 1990, Mendoza et al. 1996).

In Mandarin, creaky voice has long been observed to co-occur with T3, as mentioned in many early works (e.g., Chao 1968). Special attention has also been given to the use of creaky voice among females (e.g., Hockett 1947). Since creaky phonation creates the percept of lowered F0, some have argued that the use of creaky voice in Mandarin is a by-product of articulation, which is often especially prevalent among females who have higher overall pitch, because the average fall in F0 during the production of T3 is greater than for males.

Accurate tonal perception in Mandarin is generally attributed to the relative F0 values/trajectories of the four tones. However, T3, which contains a final rising portion when utterance-final or in citation, but otherwise remains low, is often confused with T2. This is usually attributed to T3's rising portion (which occurs in only some positions) and because of a tone-sandhi rule that changes T3 to T2 when it immediately precedes another T3. Confusion matrices (e.g., Huang 2001, Lee et al. 2008) show the closeness of T2 and T3 for Mandarin listeners. Maps of the perceptual tone space reveal that T2 and T3 are closer for Mandarin listeners than for English listeners. In L1 acquisition, Mandarin-learning children acquire adult-like production of T3 last, just after T2 (Wong 2012). L2 learners of Mandarin exhibit difficulty with the production of T2 and T3, relative to the other tones, regardless of their language backgrounds (Hao 2012).

Because differences in pitch have generally been ascribed as the primary means of distinguishing Mandarin tones, there is an expectation that other acoustic cues, such as voice quality, have little effect. In part because it seems that the use of creaky voice is just an articulatory coincidence of producing this low tone, some expect that creaky voice has a limited role, if any, to play in T3 identification. Gårding et al. (1986), in a study of T3 and T4 in Mandarin, argue that creaky voice does not aid listeners in identifying these two tones correctly. They did not, however, keep voice quality independent from F0 information associated with each tone and they only used stimuli from synthesized male speech. Davison (1991:1) reports that in the Tianjin dialect of Mandarin, which has two low tones (only one of which is produced with creaky voice), the slightly different pitch shape is a "more salient" feature in distinguishing the two tones than the different voice qualities. Others have investigated the role that creaky voice plays in Mandarin discourse (e.g., Belotel-Grenié and Grenié 2004).

On the other hand, many recent studies have shown that voice quality cues do aid listeners in perceiving certain tones. Some have looked more directly at how creaky voice influences the perception of T3 in Mandarin. Grenié and Belotel-Grenié (1994), using a gating experiment, showed that with creaky voice present, T3 identification was significantly faster than without it. Similarly, Yu and Lam (2011) tested the role of creaky voice in the perception of tone in Cantonese, which has two low tones, pitch patterns 21 and 33—the former is observed with creaky voice, the latter is not. They took stimuli from a corpus of natural speech, some tokens creaky and some not, and created a word-identification task, the results of which showed that listeners are much better at identifying the 21 tone when it is accompanied by creaky voice than when it is not. These studies suggest that creaky voice is a secondary cue for tone identification in related Chinese languages.

This study examines the role of voice quality in the identification of T2/T3 when F0 information, taken from T2, is held constant. No other studies to my knowledge have examined whether creaky voice can have a direct effect on the identification of T3 with the F0 information associated with a different tone. Kreiman and Gerratt (2010) found that Mandarin speakers are more sensitive to changes in the harmonics of the voice source (where one can find indicators of phonation type) than English speakers, who have no phonemic use of voice quality. Similar results were found when they compared English speakers with Gujarati speakers, who use voice quality distinctively (Kreiman et al. 2010). If native Mandarin listeners are more sensitive to cues of voice quality than speakers of languages who have no phonemic use for voice quality, this leads to the question of how significant voice quality is Mandarin. Can creaky voice alone, without T3's F0 cues, lead to the perception of T3? This study attempts to answer this question.

2. Methodology.

2.1 Participants. Twenty-five native Mandarin speakers participated in this study, for which each was compensated five U.S. dollars. All were living in Hawai'i at the time of the experiment, and were recruited from the University of Hawai'i with flyers advertising the study. Before the experiment, participants completed a questionnaire about where they were born and raised, how long they had lived in the U.S., and how much they used Mandarin and other languages on a daily basis. A summary of the participant information is presented in table 1.

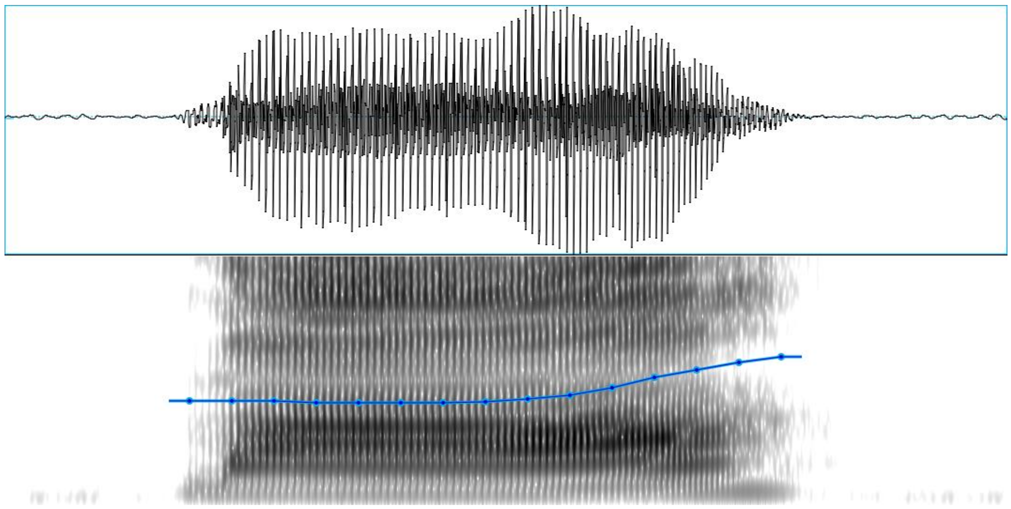
The demographics/groups were slightly unbalanced, as seen in table 1, because recruitment was not conducted with these factors in mind. Factors such as the amount of time lived in USA, and the percentage of Mandarin daily usage were each split into two groups that balanced the participant population as evenly as possible. All factors were tested post-hoc in the statistical models to see if they contributed an effect or interaction to the response patterns.

Table 1. Demographic and language-use information about participants

Characteristics	Factors		Min.	Mean	Median	Max.
Gender	Female 16	Male 9	NA	NA	NA	NA
Age (in years)	Under 30 15	30 or Over 10	18	29.9	28	47
Time spent in USA	< 1 year 12	≥ 1 year 13	2 weeks	2.1 years	1 year	14 years
Mandarin Usage	≤ 50% 14	> 50% 11	10%	56%	50%	95%
Country of Origin	Taiwan 9	P.R.C. 16	NA	NA	NA	NA

2.2 Design. In order to test whether a T2 token could be identified as T3 based solely on the amount of creaky voice simulation, an experiment was designed to test participants' reaction to a stimulus continuum of creaky voice manipulation. F0 information in each of the target stimuli was left constant, with a rising pitch characteristic of T2. An example of an unmanipulated T2 token, *ma*35 (麻), in which the pitch is relatively high and rises over time, is shown in figure 2. The effect of the creaky voice manipulation at each level of the continuum is considered to be present if listeners are significantly more likely to identify tokens as T3 than T2 (i.e., level 0), since pitch information from T2 is constant on every token. Because this experiment was meant to explore not just *whether* creaky voice could have an effect on T3 identification, but also *how much* creaky voice simulation was needed for such an effect, it was not known whether there would be a gradual increase in T3 as the level of creakiness went up, or if there would be a categorical jump from T2 to T3 at a certain level.

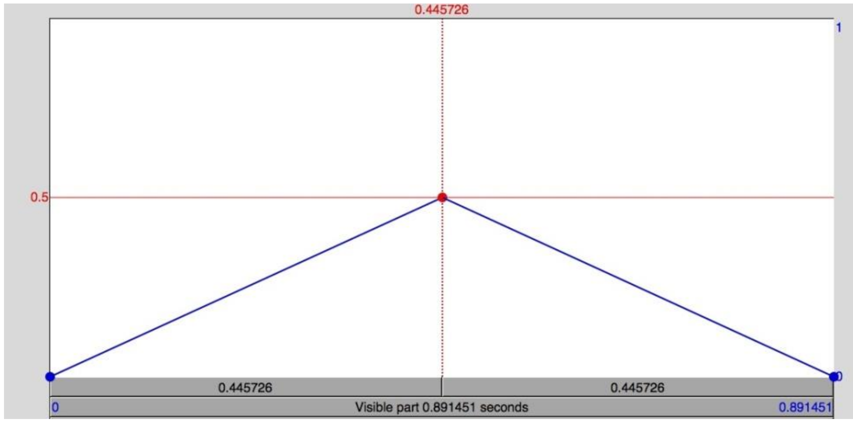
FIGURE 2. Example waveform and spectrogram of a T2 token, *ma*35 (麻), spoken by a female native Mandarin speaker. The line in the spectrogram shows the results of the automatic pitch tracker in Praat (Boersma and Weenink 2013).



2.3 STIMULI. Participants were presented with stimuli recorded by native Mandarin speakers, which were then manipulated using Praat's KlattGrid (Boersma and Weenink 2013). Two speakers, one female and one male, were recorded saying *ma*55 (妈), *ma*35 (麻), *ma*214 (马) and *ma*51 (吗), which are a four-way minimal set contrasting only in tone. The speakers said the minimal set in isolation as many times as they wished until they were satisfied with a final version. The T2 tokens from these recordings were then manipulated by attenuating every other glottal pulse[2] by a given percentage in order to create the percept of creaky voice. This method simulates shimmer (the cycle-to-cycle variation in the amplitude of glottal pulses) observed during creaky voice production. The amount of glottal pulse attenuation ranged from level 0 (0%) to level 10 (100%) in single-level steps of 10% resulting in a stimulus continuum from modal voice to maximum creaky voice (see figure 4). Rather than maintaining this creaky voice simulation throughout the entire token, the glottal pulse attenuation peaked at the midpoint of the token, rising linearly from zero at the beginning and falling to zero at the end (see figure 3 for illustration). Using the midpoint of the token for maximum double-pulsing approximates the creaky voice that is prevalent during the "dipping" portion (i.e., lowest pitches) of a T3 word in isolation (see figure 1 for a naturalistic example).

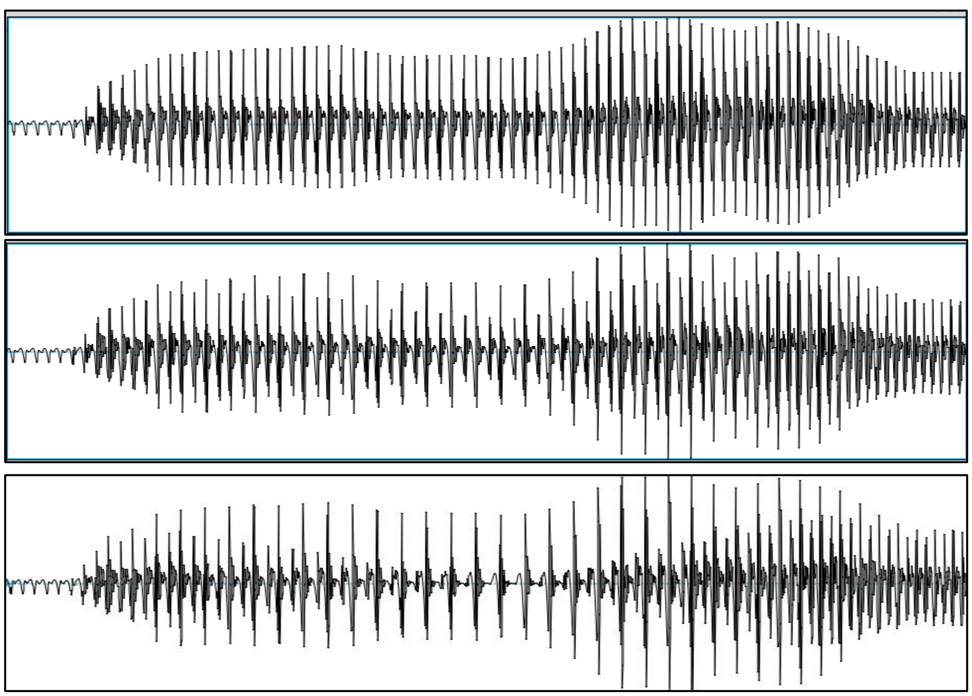
[2] This function is called "double pulsing" in Praat's KlattGrid.

FIGURE 3. Example of KlattGrid's double-pulsing tier with 0.5 (level 5) value at midpoint.



Example waveforms of the target stimuli are found in figure 4. The result of the manipulation is visible in the level 5 and level 10 tokens in figure 4. The tokens with creaky voice synthesis at levels 5 and 10 approximate the jitter and shimmer of naturalistic creaky voice production. The only manipulation is the attenuation of alternating pulses; all other information, including duration and F0, was left unaltered.[3]

FIGURE 4. Waveforms of T2 token at level 0 (top), level 5 (middle) and level 10 (bottom) creaky voice manipulation.

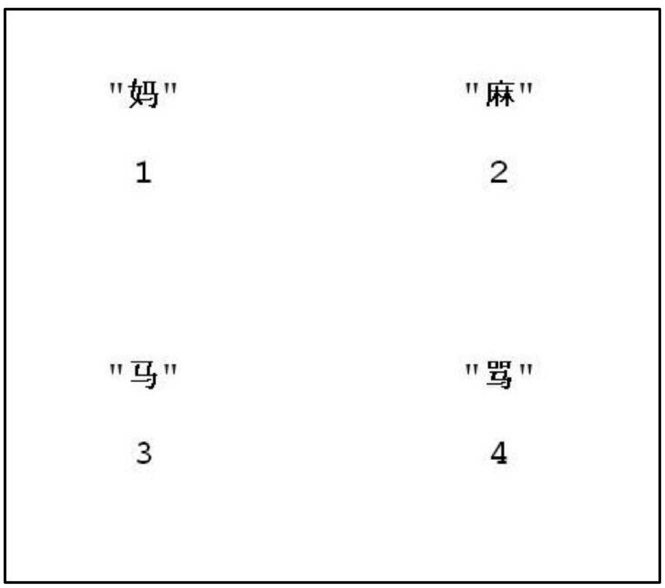


[3] The manipulated waveforms cause the automatic pitch tracker in Praat to report erroneous pitch values. It detected a longer pitch cycle during the periods of manipulation because it disregarded the attenuated pulses. Recording manual pitch measures based on cycles at every pulse, as described in Ladefoged 2003, reveals uniform pitch levels consistent with T2 throughout levels 1-9 tokens. An exception was the level 10 token, which had actual lower pitch values because every other pulse was attenuated by 100% at the midpoint, and thus the absent pulses created longer cycles and lower pitch.

First and fourth tone tokens were presented as fillers during the experiment. In order to make them comparable to the target stimuli, they were also resynthesized in the KlattGrid so that all tokens had a base-level of manipulation (and all sounded equally "artificial"). However, no double-pulsing was added to the filler tokens, nor to the T2 token at level 0; compare the waveform in figure 2 with the level 0 waveform in figure 4. No other manipulation was performed on the filler tokens; all distinctive information such as duration, pitch and intensity was left unchanged.

2.4 TASK. Participants were presented with a word-identification task in which they had four options based on the four tones of Mandarin. Participants first completed one trial of a practice session, which included both first- and second-tone stimuli from the experiment to become familiarized with the task and the stimuli. During the test phase, participants were asked to identify which of the four words they heard (see figure 5) and respond with the corresponding number. There were two blocks in the experiment, one containing all the stimuli from the first speaker (a female), and another containing all the stimuli from the second speaker (a male), with a short break between the blocks in which they had no discussion with the experimenter. In each block, the participant heard each target stimulus three times (the stimulus continuum contained eleven levels for a total of 33 target tokens in the stimulus continuum) and the filler tokens (T1 and T4) fifteen times each; no natural T3 tokens were included in the experiment because participant response to the creaky voice continuum was the object of interest. The order of the tokens was randomized within each block.

FIGURE 5. Screenshot of word identification task.



3. RESULTS

3.1 OVERALL. Participants responded to the task as expected. T1 and T4 filler responses were nearly uniformly correctly identified (only one of each was incorrectly identified over the entire dataset). Simulation of creaky voice showed an overall effect on listeners' identification of T3. As shown in figure 6, with all responses from both blocks combined, a trend is observed toward higher T3 responses (and fewer T2 responses) as the level creakiness increases. T1 responses to the target stimuli (only three in the entire dataset; no T4 responses) were excluded because they do not help answer the research question, which involves the role of creaky voice in the perception of T3 (not T1 or T4). This also allowed a simpler binomial statistical model to test responses (rather than a multinomial model) to be used.

There was a significant positive correlation between T3 identification and creaky voice. Table 2 shows the output of a mixed-effects logistic regression model run over all the T2 and T3 responses in which each level of creaky voice and the trial order in which the participants heard the stimuli throughout the experiment were tested as predictors, and the intercepts and slopes for each level of creakiness for each participant were treated as random effects.[4]

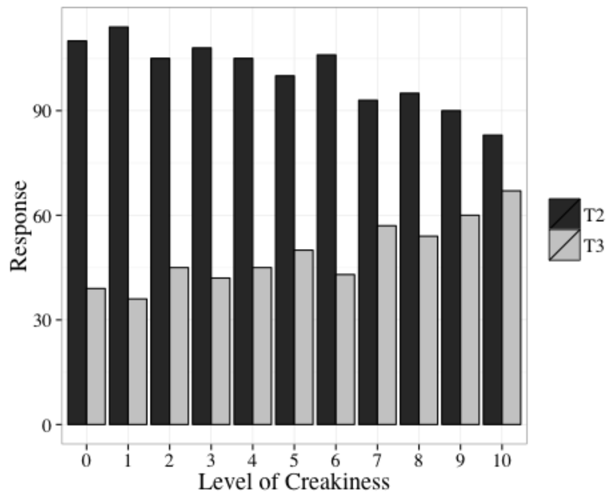


FIGURE 6. Number of overall responses to creaky voice stimuli per level.

TABLE 2. Output of mixed-effects logistic regression model of T3 perception.

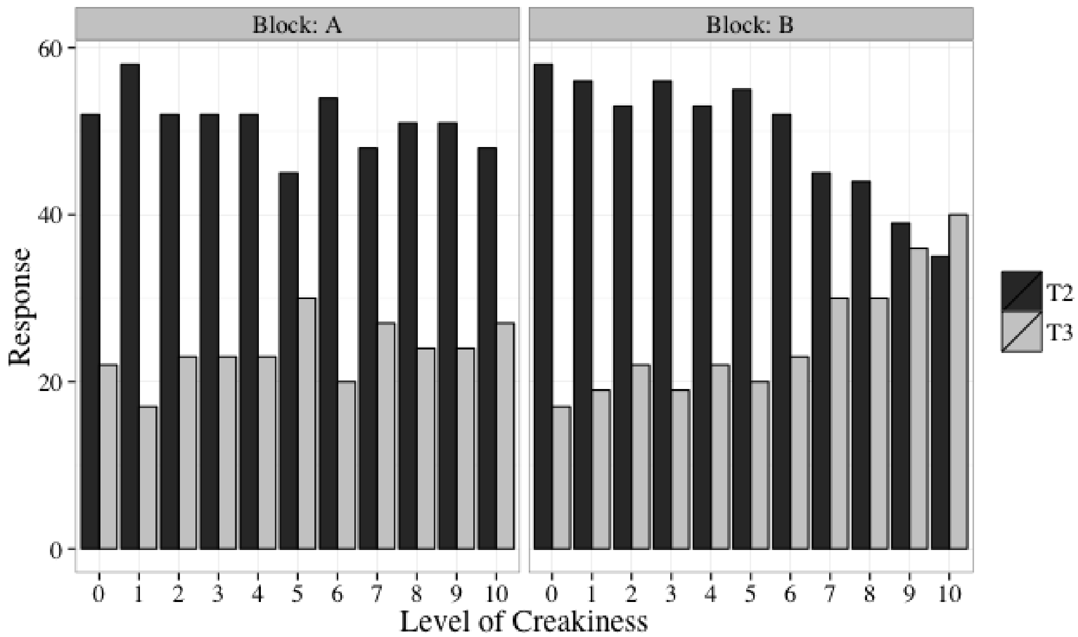
	Estimate	Std. error	z value	Pr(>|z|)
(Intercept)	-2.050	0.577	-3.556	< .001
Level 1	-0.214	0.346	-0.619	.536
Level 2	-0.248	0.400	-0.619	.536
Level 3	-0.314	0.388	-0.808	.419
Level 4	0.321	0.323	0.992	.321
Level 5	0.110	0.420	0.263	.792
Level 6	0.123	0.357	0.346	.729
Level 7	0.596	0.384	1.554	.120
Level 8	0.658	0.356	1.849	.064
Level 9	1.039	0.319	3.259	< .01
Level 10	1.452	0.337	4.308	< .001

Logistic regression models are able to predict a categorical outcome with respect to independent variables. In this case, the model predicts how likely a Mandarin listener is to respond T3 given a certain level of creakiness. In table 2, the estimated effect sizes are in log-odds, and are coefficients which are interpreted as the likelihood that a participant will respond T3 to a creaky T2 token; positive coefficients mean positive correlations and vice versa. The effect of level reaches significance at levels 9 and 10. All other levels of manipulation do not show a significant effect on a participant's response. Trial order failed to reach significance ($p > .1$), which shows that the effect observed in the responses is not due to a learning effect throughout the course of the experiment, but is actually an effect of the creaky voice.

3.2 SEPARATED BY BLOCK. The effect of creaky voice, however, is carried entirely by responses in Block B, as shown in figure 7.

[4] All statistical values reported in this paper were computed using R (R Core Development Team 2008).

FIGURE 7. Responses to creaky voice stimuli, separated by block.



Block A, which contained the voice of a female, shows no trend of higher T3 responses, or significant effects of the manipulation as the level of creaky voice increases. A mixed-effects logistic regression model was run over the responses in Block A, in which each level of creakiness was tested as a predictor and an interaction between level and trial order was tested; random intercepts and slopes for each level were included for each participant. No significant trend of creaky voice on T3 response was observed in the Block A model.

Block B, which contained a male voice, however, has a significant positive effect of creaky voice simulation on T2 tokens. A mixed-effects logistic regression model was fit to the T2 and T3 responses only in Block B, in which the trial order is included as an interaction, and intercepts and slopes for each participant were included as random effects. Each of the demographic/language use factors discussed in section 3.2 was tested in the model; time spent in USA was included as a main effect because it showed significance and was a model that better fit the data. The output of the model is shown in table 3.

Within Block B, we can see from table 3 that there is a significant effect of creaky voice on the likeliness of a T3 response, beginning at level 4. A positive trend in the direction of T3 response likelihood is observed at level 6 and above. Trial order has a significant effect only at level 6 ($p < .05$). Trial order is not significant at any other level, which indicates there was no overall learning effect in the task.

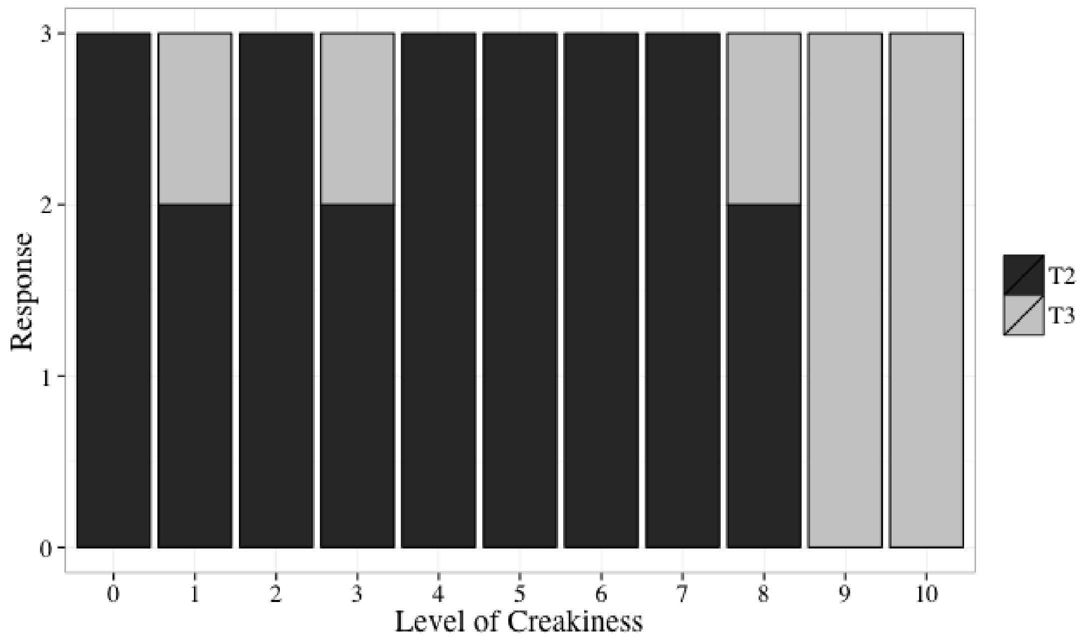
TABLE 3. Output of mixed-effects logistic regression model on Block B responses

	Estimate	Std. error	z value	Pr(>|z|)
(Intercept)	-2.436	1.291	-1.888	.059
Level 1	-2.359	1.443	-1.634	.102
Level 2	1.217	1.262	0.964	.335
Level 3	0.210	1.478	0.142	.887
Level 4	3.094	1.307	2.367	< .05
Level 5	-9.885	3.721	-2.657	< .01
Level 6	2.978	1.439	2.069	< .05
Level 7	3.689	1.256	2.936	< .01
Level 8	2.954	1.209	2.445	< .05
Level 9	4.036	1.196	3.374	< .001
Level 10	5.379	1.269	4.238	< .0001
Level 0:Trial	0.020	0.025	0.783	.433
Level 1:Trial	0.063	0.037	1.686	.092
Level 2:Trial	-0.003	0.033	-0.085	.932
Level 3:Trial	-0.021	0.040	-0.520	.603
Level 4:Trial	-0.057	0.036	-1.592	.111
Level 5:Trial	0.087	0.059	1.477	.140
Level 6:Trial	-0.083	0.037	-2.247	< .05
Level 7:Trial	-0.045	0.032	-1.402	.161
Level 8:Trial	-0.022	0.032	-0.666	.505
Level 9:Trial	-0.037	0.033	-1.107	.268
Level 10:Trial	-0.058	0.032	-1.812	.070
Time in USA (<1yr)	-2.941	0.713	-4.125	< .0001

Individual participants respond in different ways to the stimuli in perception experiments, and some of these perceptual strategies observed in this experiment are shown in section 3.2. An ordinary logistic regression model would treat two responses from the same person the same as two responses from different people. Effects of individual participants are each treated as random effects in the mixed-effects logistic regression models used here. Using random intercepts for each participant allows the model to expect a given amount of variation to the overall response patterns, while using random slopes for each participant includes random variation by participant for each level of creakiness. Both slopes and intercepts for each participant were found to have an effect on the response patterns in the data (as opposed to intercepts alone) because the direction of the random effects changed throughout the experiment. With random slopes and intercepts, all participants are assigned their own coefficients throughout the model. This makes a single participant's trends carry less weight than if they were all treated together and each response was given equal weight. This model is thus generalizable to a larger population of native Mandarin listeners, each of whom may exhibit different patterns of perception and tone identification.

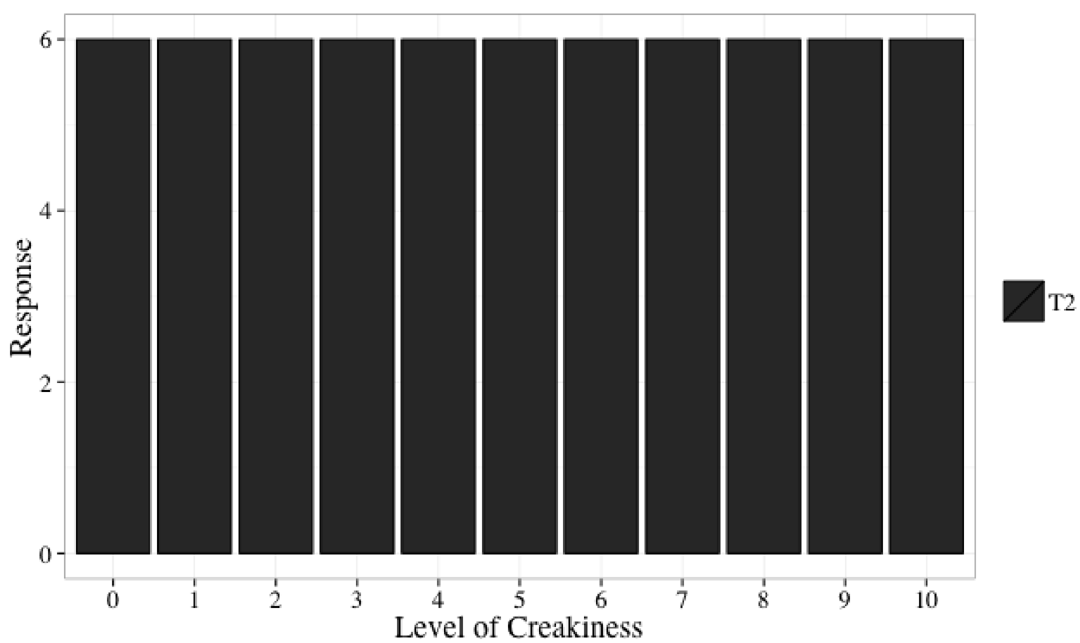
3.2 PERCEPTUAL STRATEGIES OBSERVED. Examining the responses of each subject, patterns emerge that suggest distinct perceptual strategies employed by different speakers. Some listeners use the simulated percept of creaky voice as a reliable indicator of T3. As the level of creakiness increases, the T3 responses rise as well. This contributes to the overall effect of creaky voice on T3 identification discussed in the previous section. An example of this response pattern is shown in figure 8.

FIGURE 8. Participant 12's response to the creaky voice manipulation in Block B.



Other listeners do not use the creaky voice simulation as a cue for T3. These listeners exhibit a "flat-line" response to the stimulus continuum. Clearly, for these speakers, F0 in T2 tokens trumps the creaky voice cues that have been added in, which others use for T3 identification. An example of this flat-line response pattern is seen in figure 9.

FIGURE 9. Participant 11's response to the creaky voice manipulation in both blocks combined.



4. Discussion.

4.1 Overall effect. The results demonstrate an effect of creaky voice on T3 identification. Some listeners use creaky voice alone to identify T3 even in the presence of T2 pitch information, overriding the T2 pitch information. This finding is important because it challenges the assumption that Mandarin tones are perceived and identified entirely or primarily by pitch cues; on the contrary, phonation is salient enough for some listeners to override pitch cues in identifying tones. This helps explain why Mandarin listeners are more attuned to phonation differences than are English listeners, who do not use voice quality phonemically (Kreiman and Gerratt 2010). Voice quality, therefore, may be a fruitful area for future studies in both L1 and L2 Mandarin tone acquisition.

4.2 Speaker gender. When speakers heard stimuli from a female voice (in Block A), the effect of creaky voice on T3 identification was not significant, but when they heard a male voice (in Block B), a significant trend toward T3 identification is observed. Further research should explore the effect of speaker gender by incorporating more than two individuals' voices. Since no significant effect of the trial order was observed on the overall model, it is unlikely to expect that this difference is due to a learning effect throughout the experiment. Because only voices from one female and one male were used in this study, the apparent effect of gender may in fact be due to idiosyncrasies between the two speakers used in the study, not their difference in gender.

Nevertheless, if the trend observed in this study were to hold—(highly) significant effect on male voice, no (or less) significant effect on female voice—an explanation could be that the threshold at which creaky voice becomes salient is higher for females than males. This could be because females are observed to use creaky voice more often both cross-linguistically (Klatt and Klatt 1990, Mendoza et al. 1996) and specifically in Mandarin to extend the pitch range to lower pitches.

An extension of this line of research should also investigate whether there is any effect or interaction between the gender of the listener and speaker gender. In the current study participant gender was found to have no significant effect on T3 response. A study with a balanced number of male and female participants and a higher number of speakers and participants (balanced for females and males) is needed to check for this.

4.3 Individual strategies of perception. Participants exhibited different response patterns to the stimuli in the experiment. Some participants showed a positive correlation between creaky voice and T3 identification. For other listeners, however, the voice quality information was not enough to override the T2 pitch cues, and they consistently identified T2 for all stimuli. It is possible that this is the result of interactions which were not observed in this experiment, or simply idiosyncratic differences among listeners regarding perceptual strategies.

A possible reason for the discrepancy in discrimination strategies observed here is the artificial nature of the creaky voice resynthesis. It is possible that creaky voice produced by a human would have an even more significant effect on T3 identification, in more speakers. Perhaps some listeners were less forgiving of the artificiality of the manipulation, and did not allow it to affect their identification decisions.

Exploratory analysis suggests that participants who had spent less time outside of their Mandarin-speaking home country at the time of the experiment were less likely to respond T3 to creaky T2 tokens. This effect deserves further study in an experiment carefully controlled to investigate it.

4.4 Further issues. Place of origin of participants and speakers must also be carefully controlled for. For example, studies show that Taiwan Mandarin speakers reliably produce creak in T3 and not in T2, though they both have low pitch, unlike "standard" Mandarin, which has rising T2. (For more on tone production in Taiwan Mandarin, see Fon et al. 2004.) Although this experiment used mainland Chinese speakers' voices to create the stimuli, a useful follow-up study would investigate the role of creaky voice in tone identification among Taiwan Mandarin speakers using stimuli that are based on Taiwan Mandarin production.

When interpreting the results of this experiment, it is important to remember not to mistake the resynthesized creaky voice used in these stimuli for actual creaky voice produced by a human voice. It may be that some listeners would unambiguously show T3 responses to non-synthetic stimuli, where they had responded negatively to the synthesized stimuli.

Finally, this experiment dealt with words in isolation in a forced-choice identification task. The nature of the task may have led some participants to expect T3 tokens, causing a bias in their responses. Also, since T3 production varies across positions and dialects, it would be useful to investigate the role of creaky voice on T3 identification in connected speech as well as in isolation.

5. CONCLUSION**.** This study tested whether T3 identification could result from synthesis of creaky voice on T2 tokens, in order to investigate whether voice quality is a salient enough cue to override pitch information. It was found that, overall, creaky voice has a significant positive effect on T3 identification. Regarding the speaker, different patterns emerged when responses were separated by speaker voice (one female and one male); further studies are needed to determine whether the sex of the speaker is responsible for this difference. Regarding listeners, responses to the creaky voice manipulation varied with respect to perceptual strategies. In sum, creaky voice is shown to play a significant role in T3 identification, even overriding T2 pitch cues for some listeners. This finding suggests that the role of phonation in Mandarin tones deserves further examination.

REFERENCES

BELOTEL-GRENIÉ, AGNÈS, and MICHEL GRENIÉ. 2004. The creaky voice phonation and the organization of Chinese discourse. Paper presented at International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages, Beijing.

BOERSMA, PAUL and DAVID WEENINK. 2013. Praat: Doing phonetics by computer [Computer program]. Version 5.3.53, retrieved 9 July 2013 from http://www.praat.org/

BRADLEY, DAVID. 1982. Register in Burmese. In *Papers in South-East Asian Linguistics, No. 8 Tonation*, ed. by David Bradley, pp. 117–32. Canberra: Research School of Pacific Studies, Australian National University.

CHAO, YUEN REN. 1968. *A grammar of spoken Chinese*. Berkeley, CA: University of California Press.

DAVISON, DEBORAH S. 1991. An acoustic study of so-called creaky voice in Tianjin Mandarin. *UCLA Working Papers in Phonetics* 78:50-57.

EDMONDSON, JEROLD A., and LI SHAONI. 1994. Voice quality and voice quality change in the Bai language of Yunnan Province. *Linguistics of the Tibeto-Burman Area* 17(2): 49–68.

FON, JANICE, WEN-YU CHIANG, and HINTAT CHEUNG. 2004. Production and perception of the two dipping tones (Tone 2 and Tone 3) in Taiwan Mandarin. *Journal of Chinese Linguistics* 32:249–81.

GÅRDING, EVA; PAUL KRATOCHVIL; JAN-OLAF SVANTESSON; and JIALU ZHANG. 1986. Tone 4 and Tone 3 discrimination in modern standard Chinese. *Language and Speech* 29(3):281–93.

GRENIÉ, MICHEL, and AGNÈS BELOTEL-GRENIÉ. 1994. Phonation types analysis in standard Chinese. Paper presented at the Third International Conference on Spoken Language Processing, Yokohama.

HAO, YEN-CHEN. 2012. Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics* 40(2):269–79

HOCKETT, CHARLES F. 1947. Peiping phonology. *Journal of the American Oriental Society* 67(4):253–67.

HUANG, TSAN. 2001. The interplay of perception and phonology in tone 3 sandhi in Chinese Putonghua. *Ohio State University Working Papers in Linguistics* 55:23–42.

HUFFMAN, FRANKLIN EUGENE. 1967. *An outline of Cambodian grammar*. Cornell University PhD dissertation.

JOHNSON, KEITH. 2012. *Acoustic and auditory phonetics, 3rd edition.* U.S. and U.K.: Wiley/Blackwell.

KLATT, DENNIS H., and LAURA C. KLATT. 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America* 87:820–57.

KREIMAN, JODY and BRUCE R. GERRATT. 2010. Perceptual sensitivity to first harmonic amplitude in the voice source. *Journal of the Acoustical Society of America* 128(4):2085–89.

KREIMAN, JODY, BRUCE R. GERRATT and SAMEER UD DOWLA KHAN. 2010. Effect of native language on perception of voice quality. *Journal of Phonetics* 38(4):588–93.

LADEFOGED, PETER. 2003. *Phonetic data analysis: An introduction to fieldwork and instrumental techniques.* Massachusetts: Blackwell.

LADEFOGED, PETER. 2006. *A course in phonetics*, *5th edition*. Boston: Thomson/Wadsworth.

LADEFOGED, PETER and IAN MADDIESON. 1996. *The sounds of the world's languages*. Oxford: Blackwell.

LEE, CHAO-YANG, LIANG TAO, and Z. S. BOND. 2008. Identification of acoustically modified Mandarin tones by native listeners. *Journal of Phonetics* 36(4):537–63.

MENDOZA, ELVIRA; NIEVES VALENCIA; JUANA MUÑOZ; and HUMBERTO TRUJILLO. 1996. Differences in voice quality between men and women: Use of the long-term average spectrogram (LTAS). *Journal of Voice* 10:59–66.

R CORE DEVELOPMENT TEAM. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org

WONG, PUISAN. Acoustic characteristics of three-year-olds' correct and incorrect monosyllabic Mandarin lexical tone productions. *Journal of Phonetics* 40(1):141–51.

YU, KRISTINE M., and HIU WAI LAM. 2011. The role of creaky voice in Cantonese tonal perception. In *Proceedings of the 17th International Congress of Phonetic Sciences*, 2240–43.

jvanway@hawaii.edu