Acoustic correlates of creaky voice in Yoruba

Nate Koser
nate.koser@rutgers.edu
Rutgers University
2019

1 Introduction

This paper examines the acoustic correlates of creaky voice on the low tone in Yoruba. This is carried out through a pair of experiments. The first confirms previous experimental results and reports in the literature that attest to the presence of creaky voice in the language (Welmers 1974; Hayward et al. 2004; Yu 2010). The second experiment expands these results to disyllabic words, showing that consistent differences in creakiness between tone levels are present there as well.

Looking at disyllabic words, not just monosyllables as in Hayward et al. (2004), enhances our understanding of the phenomenon by addressing the following research questions. First, what correlate of creaky voice is the most consistent? What changes, if any, occur in the correlates from syllable to syllable? Second, how do the measured correlates change over the course of the vowel? Are they relatively stable or is there some portion of the vowel that is more or less creaky? What is the effect of the tone sequence on level of creak? And finally, is there a gradual shift towards creakiness when moving progressively lower in the speaker's register, or is the difference categorical?

Linear mixed effects models are built to address these questions for each acoustic measurement taken. The main result is that Harmonic-to-Noise Ratio (HNR) is the only consistent correlate of creaky voice in Yoruba. It marks low tones as categorically different from mid and high tones, and is consistent across and within syllables and speakers.

2 Background

2.1 Creaky Voice

Creaky voice is a mode of phonation in which the glottal folds are drawn closely together – but not completely together – allowing for voicing to occur. This produces vocal pulses at irregular intervals, reflected in waveforms as irregular pitch periods and lower intensity when compared to modal phonation (Ladefoged 1971; Laver 1980). Ladefoged (1971) suggests that creaky voice is one side of a linguistic continuum, with "most closed" glottal states (creaky voice; full glottal closure) on one end and "most open" (breathy voice; voiceless phonation) glottal states on the other. While Gordon and Ladefoged (2001) caution that this may be an oversimplification, it is still useful to think of non-modal phonation in these terms.

Laver (1980); Klatt and Klatt (1990), among others, have identified many acoustic properties of creaky phonation. This includes low F0, irregular F0, low spectral tilt, and level of glottal constriction, expressed as the difference between the first and second harmonics (H1-H2), to name a few. In a survey of different kinds of creaky voice, Keating et al. (2015) found H1-H2 to be the most common indicator of creak.

While there is no apparent *a priori* reason that non-modal phonation and tone should be linked in languages in which both are present, examples where both operate independently are rare (Silverman 1997), though Jalapa Mazatec is an example (Silverman et al. 1995). There are languages where pitch is perceptually primary, but there are consistent differences in phonation (Mandarin (Davidson 1991), Cantonese (Yu and Lam 2014), Cham (Brunelle 2012)). There is also the opposite case, with association of perceptually primary non-modal phonation types to certain pitch levels being an established property of tone languages in East and South East Asia such as Burmese and Hmong (Bradley 1982; Huffman 1987).

2.2 Yoruba

Yoruba is a Niger-Congo language spoken by approximately 28 million people. The highest concentrations of Yoruba speakers are in Nigeria, Benin, and Togo, where in all three countries it is an official language.

Yoruba has three tone levels - high (H), mid (M), and low (L). How the three tones relate to one another has been the subject of much discussion.

There is evidence that the "mid tone" is actually not a tonal unit at all, and is instead the absence of tonal features, as it is not affected by tonal processes in the same way as the high or low tone and thus should be considered a "default" (Akinlabi 1985; Pulleyblank 1986). Stahlke (1974) posits that the low and mid tone are the result of a historical split based on their distribution.

Low tone has been observed to fall in pitch as the utterance continues, while the high and mid tone show a relatively stable F0 value throughout their course (Connell and Ladd 1990). This effect is so salient that it plays an important role in perception of low tones, to the extent that Harrison (1996) found that none of his participants perceived any of his synthetic stimuli with flat F0 contours as low. Bakare (1995)'s results suggest a hierarchy in terms of which tones are most distinctive for Yoruba listeners, where high tone is most distinctive, the mid tone is the least distinctive, and the low tone is somewhere in the middle.

Other sources mention creaky voice in Yoruba in passing, but provide no acoustic analysis of the phenomenon (Welmers 1974; Yu 2010). Hayward et al. (2004) found that the low tone patterned differently from the mid and high tone with regards to phonation type based on several acoustic measures. Figure 1 shows an example of their findings.

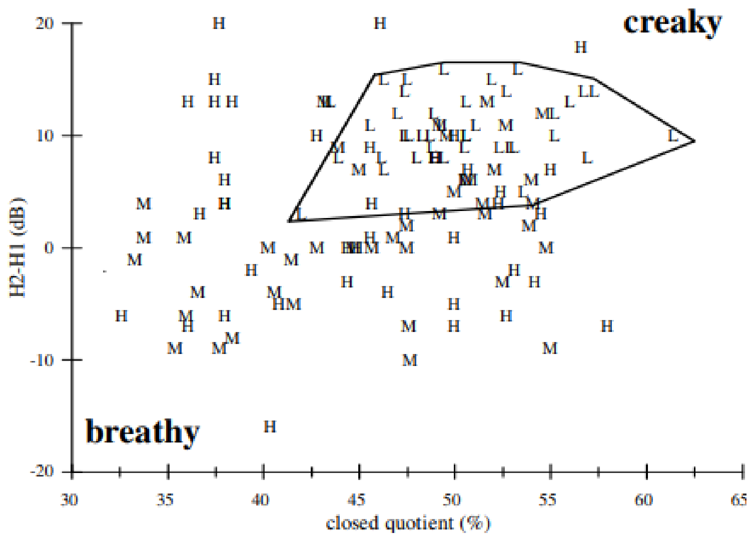

Figure 1: L tones cluster toward creaky end of spectrum (Hayward et al. 2004)

The graph plots the value of H2-H1 versus the measured closed quotient (CQ), which is the ratio of the duration of glottal closure to the entire period of glottal fold vibration. A higher CQ and H2-H1 are generally indicative of creaky voice quality. The graph shows that while high and mid tones are more freely distributed, the low tone data points cluster in the upper right corner – the area most associated with creaky voice based on the measures used. As the current study uses similar methodology to Hayward et al. (2004), a similar result is expected.

3 Procedure 1

3.1 Methodology

The first experiment confirms the presence of creaky voice on the Yoruba low tone. Given the success of Hayward et al. (2004) in pinpointing creaky voice in Yoruba, similar methods are employed here. The target words come from a list of 63 CV words representing all possible combinations of the seven Yoruba vowels (/i e ɛ a ɔ o u/) at the three tone levels (high, mid, low) with three initial consonants (/t n l/). This results in a mixture of actual and nonsense words. The tokens are then uttered in the following frame sentence:

(1) Sọ _____ lẹ kan sí i
/sɔ _____ lɛ kã sí i/
Say _____ once more

Participants are given a practice period to familiarize themselves with the task and the frame sentence, as it is not visible during the experiment. Tokens are randomly ordered and presented as single CV words using PsychoPy v3.0 (Peirce 2007). Each token is repeated five times for a total of 315 data points.

3.2 Participants

One recording has been made to this date. The participant was a 31 year old male who lived in Nigeria until the age of 26 and has since moved to the United States for school. He grew up in a bilingual Yoruba-English household, acquiring both simultaneously from birth. The speaker indicated that he spoke Yoruba "all the time" as a child, and that he still uses it frequently. He reported no difficulties in speaking or listening. A colleague, also fluent in Yoruba, was present during the recording and attested to the quality of the speech produced by the participant. He also engaged the participant in conversation in Yoruba before the task began.

The recording session took place in a sound-attenuated booth at the Phonology Laboratory at the Rutgers Center for Cognitive Science using a Logitech H390 USB microphone headset attached to the researcher's laptop running Audacity audio recording software version 2.3.0 recording in mono at a project rate of 44100Hz.

3.3 Data analysis

Statistical analysis is carried out using R (R Core Team 2017). The influence of tone (independent variable) on the various acoustic measures (dependent variables) is analyzed using linear mixed effects models as implemented in the `lme4` package (Bates et al. 2015) with *block* as a slope for a random *word* intercept. The goal in modeling this way is to understand what significant differences exist with regard to the measurements when moving from one tone level to another. The expectation is that the low tone corresponds to acoustic properties that are characteristic of creak, while the other two tone levels do not. An additional result is a conception of exactly how creaky voice is implemented in Yoruba.

Initial impressions are consistent with the findings of Hayward et al. (2004) – low tone is distinct from high and mid tone with regard to phonation type. Figure 2 shows waveforms with a representative sample of a high/low contrast in the recorded speaker for the word *tá* (gloss: *feel for*; left) and *tà* (gloss: *sell*; right):

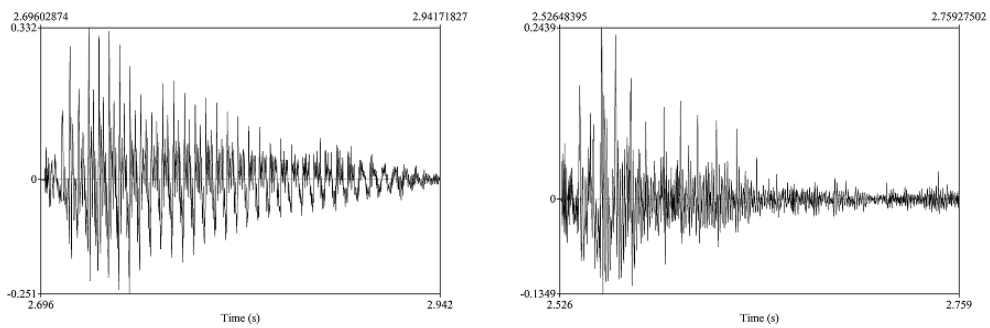

Figure 2: waveforms for *tá* (left) and *tà* (right)

Comparing the two, the low tone waveform exhibits hallmarks of creaky phonation: aperiodic pitch periods and decreased intensity. There are also pitch spikes further along in the signal than is usually observed in modal voicing. There are fewer pitch periods observed in the low tone waveform, and they occur at irregular intervals when compared to the high tone waveform, which displays more frequent, regularly-spaced pitched periods. This result is indicative of the creaky-non creaky dichotomy reported in Hayward et al. (2004) and others.

3.4 Measurements

Four measurements were taken to evaluate acoustic data. These included F0, vowel duration, spectral tilt – a measure in dB of glottal constriction measured as the difference in amplitude between the first and second harmonics (H1-H2), and Harmonic to Noise Ratio (HNR), a measure in dB of aperiodicity and turbulence of airflow at the glottis for which lower values indicate creakier phonation (Keating et al. 2015). For both spectral tilt and HNR, lower values are indicative of creaky phonation.

All segmentation was done in Praat (Boersma 2001). For /t/-initial tokens, the boundary between the stop and the vowel was marked at the zero-crossing of the first non-deformed pitch period. For /n/-initial tokens, the boundary was marked at the point where amplitude increased, seen as a clear darkening in the spectrogram for F2 and F3. For /l/-initial tokens, the boundary was marked in the same way. The end of the vowel was marked where intensity died off, as determined by Praat's automatic intensity detection algorithm. A script divided the vowel into four even slices. F0 was measured by finding the mean hertz value in each of the slices. Spectral tilt was measured by searching a range around the mean F0 for the peak hertz to get H1, searching a range around double the mean F0 for the peak hertz to get H2 and subtracting the latter from the former. HNR was measured by periodicty detection as described in Boersma (1993), measuring the mean amplitude of 4.5 periods per 0.01s frame above a silence threshold of 0.1dB. Eleven tokens were excluded from the analysis due to speaker error. Means and standard deviations for these measurements are shown in the table below.

	F0	HNR	spec tilt	duration
H	149.39	17.84	7.91	0.26
	6.44	*4.15*	*3.43*	*0.04*
M	126.97	13.56	4.84	0.27
	6.37	*7.22*	*2.30*	*0.05*
L	98.48	0.71	-0.19	0.17
	5.08	*3.06*	*4.73*	*0.04*

Table 1: Means (above) and standard deviations (below) for acoustic measurements

The acoustic measures taken from the data are suggestive of a distinctive creaky quality for the low tone in Yoruba. HNR and spectral tilt are much lower in the low tone. High standard deviations can be in part attributed to the various vowel qualities and initial consonants. In particular, the participant was observed to aspirate /t/-initial tokens before high vowels only, resulting in breathier phonation for those tokens as compared to others.

Among the measurements in Fig. 1, spectral tilt and HNR appear to show the clearest bifurcation of the tonal space in Yoruba. The plot in Fig. 3 makes this result clearer. A regression line with 75% confidence interval (CI) is shown for each tone.

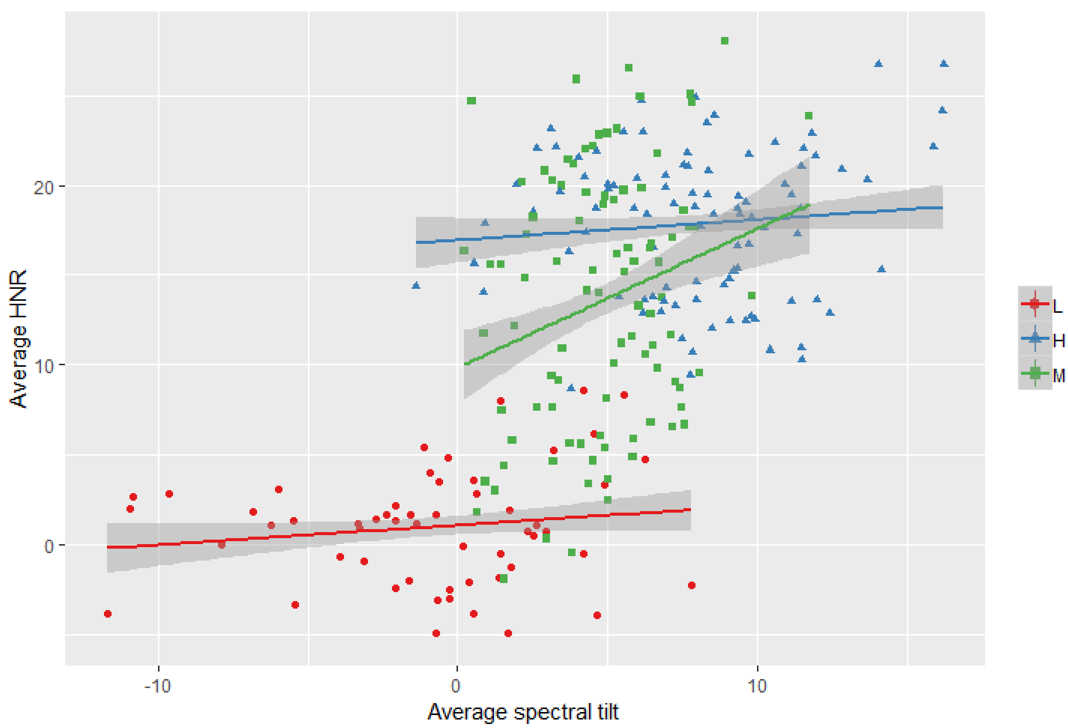

Figure 3: Average HNR and Average spectral tilt by tone

The low tones cluster in the region of lowest HNR and spectral tilt. Mid and high tones cluster in the opposite region. While the distribution of mid and high tones shows a great deal of overlap, there is very little overlap between low tones and either of the other two tonal categories. This suggests that low tones in Yoruba are distinguished from mid and high tones in a way that mid and high tones are not distinguished from each other. Low tones carry acoustic properties typical of creaky phonation, and mid and high tones generally do not. The following subsections present a linear mixed effects model-based assessment of the effect of tone level on each acoustic measurement individually.

3.4.1 Duration

The mean duration values in Fig. 1 shows what appears to be a difference between low tones and high and mid tones. The model testing tone category as a predictor of duration did reveal a significant difference for low tones between both mid ($\beta = 0.10, SE = 0.01, df = 53.87, t - value = 13.90, p < 0.001$) and high tones ($\beta = 0.09, SE = 0.01, df = 54.03, p < 0.001$), with model $R^2 = 0.72$. Pairwise comparison finds no significant difference between high and mid tones. During segmentation of the raw data the end of the vowel was marked where intensity died off. If low tone vowels are creaky, then they should have lower/faster dying intensity, and so would have been systematically marked as shorter given consistent segmentation.

3.4.2 F0

Based on mean F0, there is a clear three-way partition of the register space in Yoruba. The mean low tone value of just under 100Hz was drawn from the entire duration of the vowel. This obscures the effect of the contour, which was quite noticeable. Consider the diagram in Fig. 4.

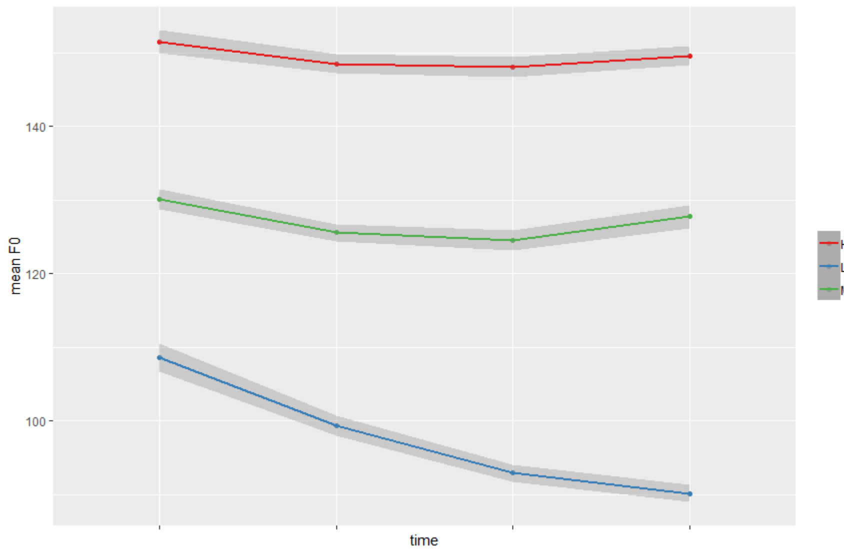

Figure 4: mean F0 over course of vowel

This plot shows the mean value of F0 as the vowel progresses (95% CI shown). The data indicate a ~20Hz drop in F0 over the entire duration of the vowel for the low tone. A linear mixed effects model with average F0 as the dependent variable and vowel portion as independent variable with the first slice of the vowel as the reference level indicates a significant difference between the start of the vowel and all proceeding portions (first vs second: $\beta = -9.23, SE = .82, df = 305.35, t-value = -11.32, p < 0.001$; first vs. third $\beta = -15.67, SE = .85, df = 307.96, t-value = -18.50, p < 0.001$; first vs. fouth $\beta = -18.55, SE = .91, df = 309.92, t-value = -20.50, p < 0.001$, $R^2 = 0.68$). A natural question then is whether other acoustic measures vary significantly within the same vowel. It is known that non-modal phonation can occur over a certain portion of the vowel, rather than its entire duration (Gordon and Ladefoged 2001). I address this question for each measure individually.

3.4.3 HNR

The Harmonic to Noise Ratio (HNR) is a measure of the irregularity of F0 and turbulent airflow at the glottis during production measured in dB. Lower values are indicative of creaky voice. Based on the tables and figures above, it appears that low tones in Yoruba do pattern differently than mid or high tones with regards to this acoustic property. The output of a linear mixed effects model with average HNR as the dependent variable and tone category as the independent variable is shown here in Table 2.

	β	SE	df	t-value	p
(intercept)	0.92	1.06	60.10	0.86	0.39
mid	13.93	1.43	58.13	9.77	< 0.001*
high	16.63	1.43	58.55	11.63	< 0.001*
random effects					
group	name	variance	σ	corr.	
word	(intercept)	20.67	4.58		
	block	1.20	1.10	-0.43	
residual		7.11	2.67		
fixed effects R^2: .64		model R^2: .91			

Table 2: Linear mixed effects model of tone on mean HNR

The low tone shows significantly lower HNR than the mid tone or the high tone. Post-hoc pairwise comparison finds a non-significant trend between the mid and high tones, such that high tones have slightly higher HNR ($\beta = 2.69, SE = 1.35, df = 0.58, t - value = 2.00, p < 0.05$). It is noteworthy that the difference between the mean HNR of high and mid tones is ~3dB, while the difference between low and mid tones is much higher, at ~13dB. That the distance between the means of the low and the mid is much greater than the distance between the means of the mid and high suggests that HNR does mark low tones as creaky in a way that is categorically different from mid or high tones, though a smaller effect of lowering in the register may exist, explaining the slight difference between mid and high tones.

Fig. 5 shows how HNR changes as the vowel progresses (95% CI).

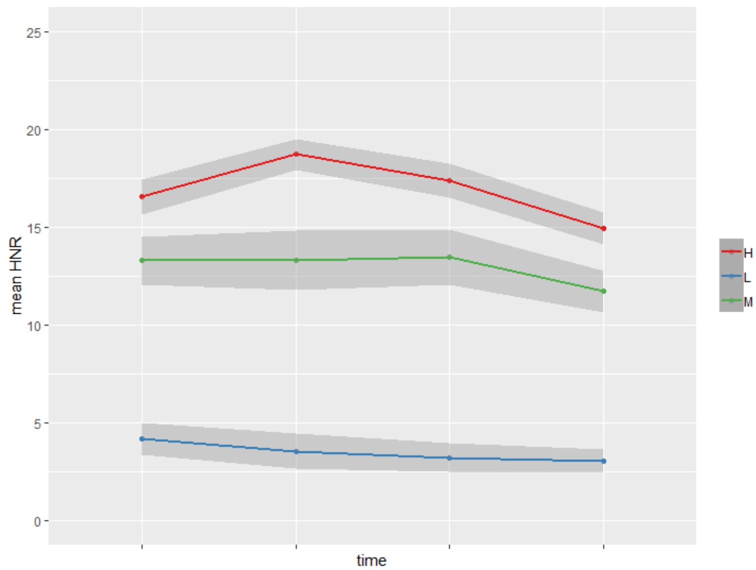

Figure 5: mean HNR over course of vowel by tone

It appears that, while the low tone is stable in terms of HNR throughout the vowel, there is movement in the high and mid tones. In high and mid tones, there is an increase in HNR from the first to the second slice, followed by a drop in HNR going through the remaining duration of the vowel. To formally assess how HNR changes over the course of the vowel, linear mixed effects models with HNR value as dependent variable and vowel slice as the independent variable with block as an intercept for word in the random effects structure were constructed for each tone. The results are shown in Table 3.

low										
	β	SE	df	t-value	p	random effects				
(intercept)	0.90	0.64	27.67	1.41	0.17	group	name	variance	σ	corr.
slice 2	-0.12	0.49	334.48	-0.26	0.80	word	(intercept)	10.50	3.24	
slice 3	0.37	0.49	335.71	0.80	0.43		block	0.19	0.44	-0.77
slice 4	0.68	0.50	337.46	1.37	0.17	residual		11.10	3.33	
						fixed effects R^2: .01		model R^2: .35		
mid										
	β	SE	df	t-value	p	random effects				
(intercept)	13.66	1.24	27.78	11.02	< 0.001*	group	name	variance	σ	corr.
slice 2	0.51	0.79	353.56	0.64	0.52	word	(intercept)	108.54	10.42	
slice 3	-1.77	0.79	353.56	-2.25	0.03		block	8.08	2.84	-0.87
slice 4	-3.63	0.80	353.77	-4.53	< 0.001*	residual		31.30	5.59	
						fixed effects R^2: .03		model R^2: .59		
high										
	β	SE	df	t-value	p	random effects				
(intercept)	17.08	0.86	23.26	19.97	< 0.001*	group	name	variance	σ	corr.
slice 2	2.83	0.38	360.25	7.40	< 0.001*	word	(intercept)	12.76	3.57	
slice 3	-0.15	0.38	360.25	-0.39	0.70		block	0.30	0.55	0.00
slice 4	-2.90	0.39	360.99	-7.42	< 0.001*	residual		7.55	2.15	
						fixed effects R^2: .15		model R^2: .73		

Table 3: Linear mixed effects model of vowel portion on HNR

For low tones, there is no significant slice-to-slice difference for any pairwise comparison of slices. Mid tones appear to drop in HNR over the course of the vowel after a slight increase between the first and second slice. Post-hoc pairwise comparison indicates that the second slice is significantly higher in HNR than the third ($\beta = 2.28, SE = 0.79, df = 353, t-value = 2.90, p = 0.004$) and fourth slice ($\beta = 4.13, SE = 0.80, df = 354, t-value = 5.18, p < 0.001$). The same pattern holds in the high tone. One interpretation of this result is that – in addition to different raw HNR values – the behavior of HNR over time also sets high and mid tone vowels apart from low tone vowels, supporting the idea of a categorical partition of the register space in Yoruba.

3.4.4 Spectral tilt

Spectral tilt is a measure of the degree to which intensity increases as frequency decreases, quantified by subtracting the amplitude value of the second harmonic peak, H2, from the first harmonic peak, H1. As it concerns phonation type, lower values are indicative of creaky voice. Looking back at Table 1, low tones do appear to have generally lower spectral tilt values than mid or high tones. The results of a linear mixed effects model with mean spectral tilt as the dependent variable and tone category as the independent variable with block as an intercept for word in the random effects structure is shown in Table 4.

	β	SE	df	t-value	p
(intercept)	-0.95	0.53	57.76	-1.80	0.08
mid	5.53	0.68	53.20	13.83	< 0.001*
high	9.37	0.68	53.08	8.18	< 0.001*
random effects					
group	name	variance	σ	corr.	
word	(intercept)	19.03	4.36		
	block	1.18	1.09	-0.95	
residual		7.18	2.68		
fixed effects R^2: .51		model R^2: .71			

Table 4: Linear mixed effects model of tone on mean spectral tilt

The mid and high tone show significantly higher spectral tilt than the low tone. Post-hoc comparison indicate that the mid and high tones differ significantly as well ($\beta = 3.84, SE = 0.60, df = 46.80, t - value = 6.44, p < 0.001$). For HNR, the mean value for low and mid tones is much farther apart than the mean value for mid and high tones. It is not clear that the same holds for spectral tilt – the difference between the low and mid is $\sim$5dB, while the difference between the mid and high is $\sim$4dB. This suggests that while HNR may mark Yoruba low tones as creaky categorically, differences in spectral tilt between tone levels is gradient, and perhaps emerge simply because the speaker is reaching a lower point in their register.

Fig. 6 shows how spectral tilt changes over the course of the vowel (95% CI).

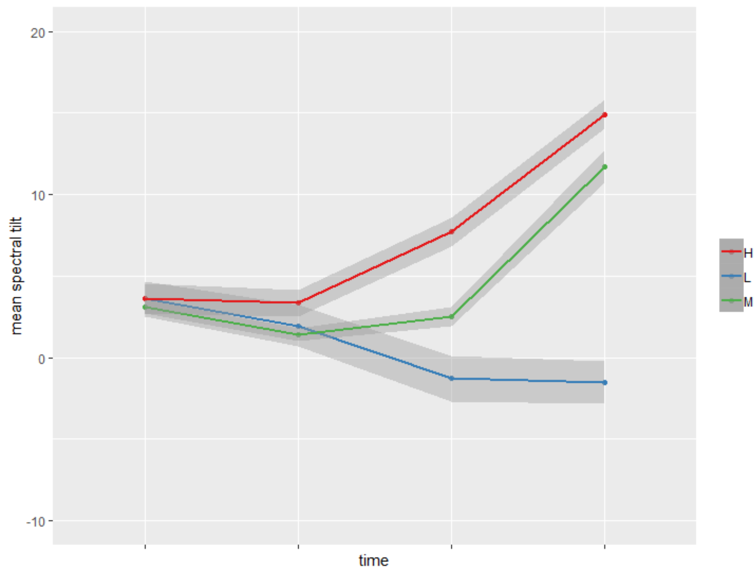

Figure 6: mean spectral tilt over course of vowel by tone

Some striking differences between the low tone and mid and high tones appear. In the first two slices, spectral tilt is quite similar between all three tone levels. As the vowel progresses, however, low tone spectral tilt lowers, while high and mid tone spectral tilt rises sharply. This suggests that, even if raw average spectral tilt values do not categorically separate low tones from the other tone levels, spectral tilt behavior over the course of the vowel does. Linear mixed effects models with spectral tilt value as dependent variable, vowel slice as independent variable, and block as an intercept for word in the random effects structure were constructed for each tone. The results are given in Table 5.

low										
	β	SE	df	t-value	p	random effects				
(intercept)	2.23	0.79	47.33	2.80	0.007*	group	name	variance	σ	corr.
slice 2	-1.98	0.85	302.95	-2.32	0.80	word	(intercept)	67.94	8.24	
slice 3	-5.94	0.89	306.56	-6.71	< 0.001*		block	5.44	2.33	-0.90
slice 4	-6.45	0.94	309.13	-6.82	< 0.001*	residual		34.308	5.56	
						fixed effects R^2: .13		model R^2: .40		
mid										
	β	SE	df	t-value	p	random effects				
(intercept)	2.81	0.36	102.38	7.89	< 0.001*	group	name	variance	σ	corr.
slice 2	-1.74	0.44	367.15	-3.94	< 0.001*	word	(intercept)	8.74	2.96	
slice 3	-0.64	0.44	367.15	-1.45	0.14		block	0.55	0.74	-0.97
slice 4	9.25	0.44	367.15	20.94	< 0.001*	residual		10.05	3.17	
						fixed effects R^2: .62		model R^2: .68		
high										
	β	SE	df	t-value	p	random effects				
(intercept)	3.95	0.61	39.33	6.44	< 0.001*	group	name	variance	σ	corr.
slice 2	-0.27	0.54	367.40	-0.51	0.61	word	(intercept)	16.21	4.03	
slice 3	4.25	0.54	367.40	7.92	< 0.001*		block	0.97	0.99	-0.84
slice 4	13.22	0.54	367.40	24.63	< 0.001*	residual		14.83	3.85	
						fixed effects R^2: .58		model R^2: .71		

Table 5: Linear mixed effects model of vowel portion on spectral tilt

For low tones, there is a general negative trend as the vowel progresses. Pairwise comparison shows a significant difference between slice two and slice three ($\beta = 3.96, SE = 0.88, df = 306, t - value = 4.51, p < 0.001$) but no significant difference between slice three and four ($\beta = 0.51, SE = 0.96, df = 305, t - value = 0.53, p = 0.59$), suggesting a "bottoming out" of spectral tilt about halfway through low tone vowels. In mid tones, after an initial significant decrease in spectral tilt between slice one and slice two, the value increases significantly between slice three and four. Pairwise comparison indicates that all previous slices are significantly lower than slice four. For high tones, the pattern is similar. Starting at the third slice, spectral tilt raises significantly. Pairwise comparison indicates that, other than slice one and two, there are significant differences between all slice comparisons.

It is noteworthy that spectral tilt values are quite similar across all tone levels in the first – and even the second – slice of the vowel. It is not until the latter half of the vowel that low tones diverge from mid and high tones.

Following the target word, there is a slight pause, and so it is possible that this is positional, allophonic variation. Just as with HNR, this grouped behavior where mid and high tones behave similarly and low tones pattern another way suggest a categorical phonetic difference for the low tone in Yoruba.

3.5 Discussion

The results of the first experiment suggest a creaky quality for the low tone in Yoruba that is generally absent from the mid or the high tone. This is in line with both previous impressionistic descriptions, as in Welmers (1974), and acoustic experiments, as in Hayward et al. (2004). The most consistent acoustic correlate of creaky voice is HNR, which marks low tones as different from mid or high tones categorically, while changes in spectral tilt from tone level to tone level may be more gradient and only emerge at the end of the word. Behavior over the duration of the vowel with regards to both HNR and spectral tilt also divides the register space in Yoruba, paritioning the low tone and higher tones. The next section details the results of the second experiment.

4 Procedure 2

The second experiment expands the analysis of creaky voice in Yoruba to disyllabic words. This results in a better understanding of creakiness in Yoruba by addressing the research questions stated in §1. What differences appear between and within syllables with regards to the measured correlates of creak? What is the effect of tone sequence? Do the results suggest a categorical division of the register space, as in the first experiment?

4.1 Methodology

The methodology of the second experiment is largely the same as that of the first experiment. The target words come from a list of 81 CVCV words representing all possible combinations of the three Yoruba vowels /i u a/ (targeting the corners of the vowel space) at all three tone levels with the consonant /n/. /n/ was chosen over /t/ and /l/ because the aspiration of /t/ can interfere with measures of F0 and creaky voice, and /l/ is more vowel-like and thus more difficult to process post-experiment. The words were presented in the frame sentence from experiment one, randomly ordered and repeated four times in PsychoPy v3.0 (Peirce 2007) for a total of 324 tokens per speaker.

For this experiment two speakers were recorded – the speaker from the first experiment, and a second speaker. The second speaker is a 49 year old male who lived in Nigeria until moving to the US at the age of 30. He is bilingual in Yoruba and English, reporting that he spoke Yoruba "most of the time" at home as a child and still uses it every day. He reported no difficulties in speaking or listening. The recording session took place in a sound-attenuated booth at the Phonology Laboratory at the Rutgers Center for Cognitive Science using a Shure SM10A head-worn unidirectional microphone and a Marantz PMD660 recording device recording in mono at a project rate of 44100Hz.

All segmentation and extraction of measurements was done in Praat (Boersma 2001) using the same methods described in §3.4 and all post-experiment analysis was done in R (R Core Team 2017).

4.2 Measurements

Three measurements were used to evaluate acoustic data. These included F0, HNR, and spectral tilt. Means and standard deviations for these values, separated by speaker and syllable are given in the table below:

	spectral tilt				HNR				F0			
	spkr 1		spkr 2		spkr 1		spkr 2		spkr 1		spkr 2	
	Syll1	Syll2	Syll1	Syll2	Syll1	Syll2	Syll1	Syll2	Syll1	Syll2	Syll1	Syll2
H	-1.99	-0.58	1.17	6.78	17.19	14.61	18.92	13.17	157.25	150.64	184.34	177.43
	4.12	*4.71*	*4.45*	*4.47*	*2.49*	*5.53*	*4.77*	*7.22*	*7.14*	*11.31*	*16.10*	*21.47*
M	-10.38	-9.55	3.41	8.82	18.73	17.64	19.87	17.40	137.05	131.69	162.24	158.47
	2.80	*3.45*	*3.45*	*3.69*	*4.00*	*4.19*	*3.65*	3.46	*7.72*	*5.34*	*12.18*	*12.59*
L	-9.55	-6.84	2.41	3.56	4.27	1.46	1.14	1.37	115.29	117.25	108.29	120.34
	2.09	*2.51*	*5.31*	*6.85*	*5.92*	*3.22*	*6.34*	*4.40*	*8.20*	*14.86*	*10.60*	*24.15*

Table 6: Means (above) and standard deviations (below) by speaker

There is some inter-speaker variation in terms of spectral tilt and F0. In Fig. 7, I give a plot of average HNR against average spectral tilt separated by tone, with a regression line and 95% confidence interval shown for each tone.

Figure 7: Average HNR and Average spectral tilt by tone, CVCV

It appears that – while HNR marks low tones generally as creakier than mid or high tones – there is no clear division of the tonal space with regards to spectral tilt. It should be noted that, because of the presence of falling and rising contours in the second syllable of LH and HL tone sequence words, the raw mean value is not as informative as might be hoped. For each measurement, I first comment on the overall pattern, followed by analysis of inter and intra-syllable differences, concluding with a discussion of the effect of tone sequence.

4.3 F0

Fig. 8 shows average word F0 divided by tone category.

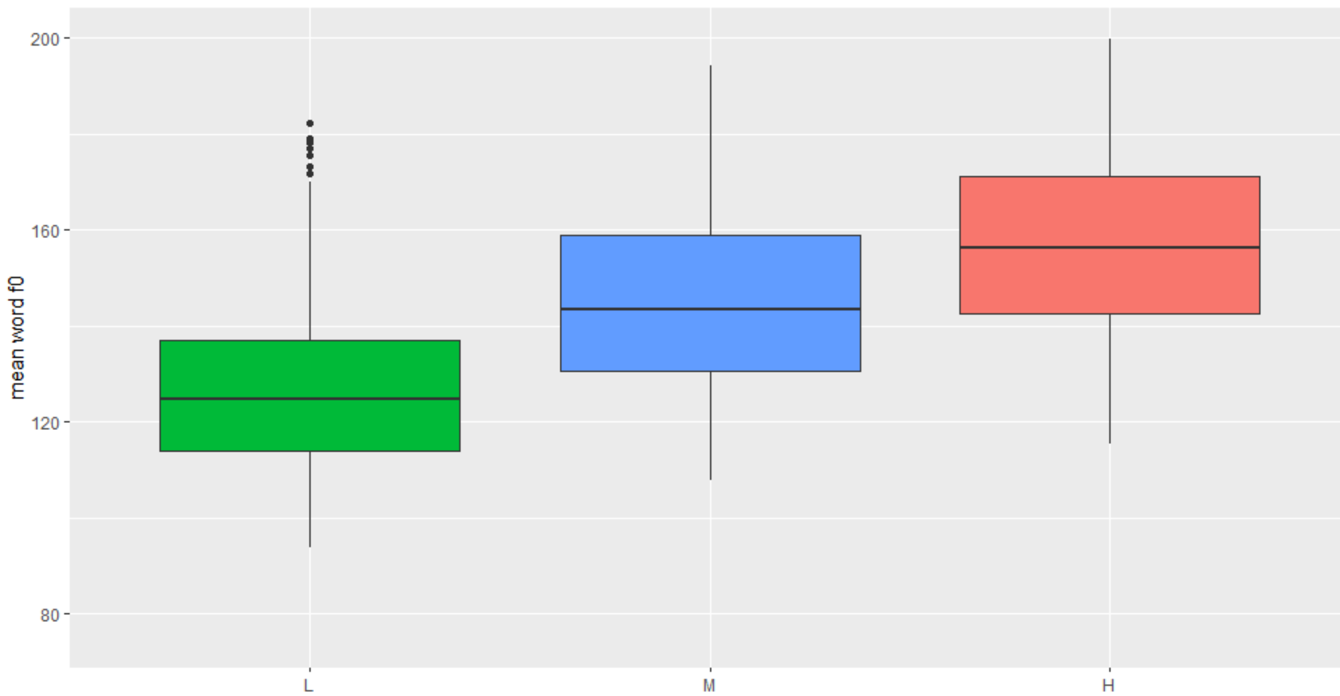

Figure 8: Average word F0

Though the difference in tone level is still visible, there is a great amount of overlap due to the inter-speaker variation seen in Fig. 6. Speaker two's mid tones are generally higher in pitch than speaker one's high tones. If speaker one has a lower pitch range in general, this may partly explain the lower spectral tilt values for speaker one as opposed to speaker two. Taking the mean F0 of the entire word blurs the picture, as an LL sequence will have a higher average F0 than an LH sequence, for example. To this end, Fig. 9 shows average F0 in each syllable separately.

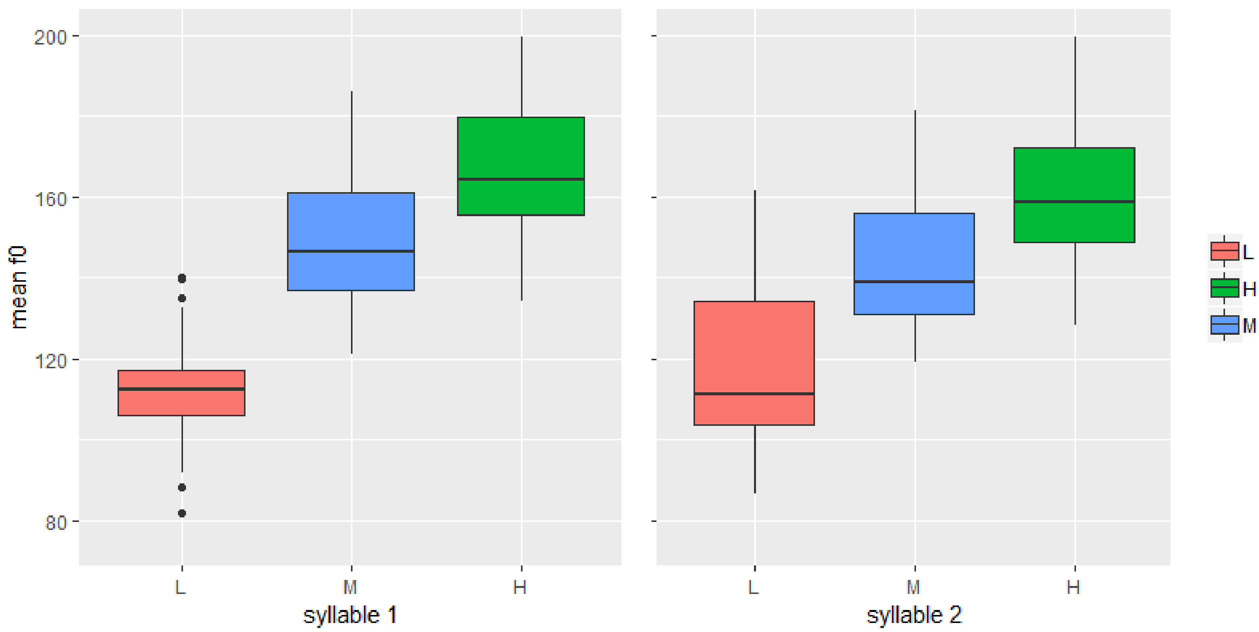

Figure 9: Average syllable F0

However, looking only at the mean values obscures the effect of tonal contour. As reported in Akinlabi and Liberman (1995), the most extreme contours occur in the second syllable of LH and HL words, such that the final high tone is a rising tone, and the final low tone is a falling tone. This can be seen in the following F0 tracks, separated by tone sequence and syllable:

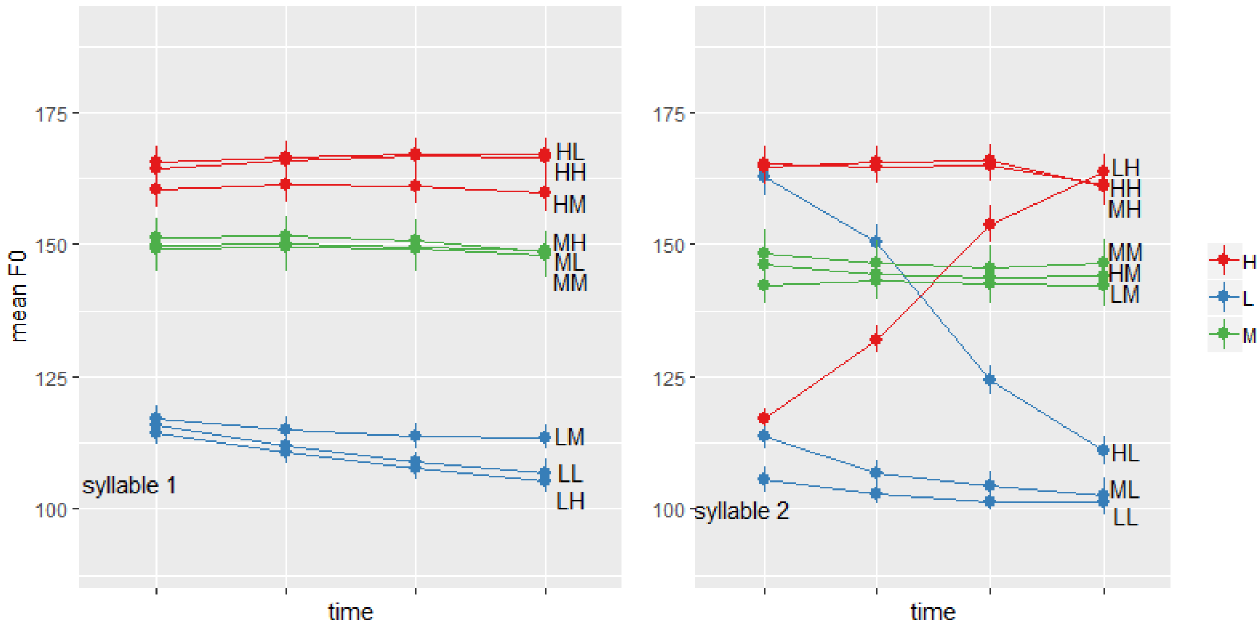

Figure 10: F0 over time by tone sequence

First comparing sequences with no contours, each tonal category behaves similarly. Low tones have a falling character in both syllables similar to what was seen in experiment one, no matter what tones precede/proceed them. Mid tones are the most stable of all three, showing almost no movement. Second syllable high tones in MH and HH sequences are stable until a slight drop at the end of the vowel.First-syllable high tones are stable and, with some variation based on the following tone.

The low tone in an HL sequence shows a drop in F0 of around 50Hz, and the high tone in an LH sequence shows a rise of a similar value in F0 over the course of its duration. The presence of these contours is one reason for analysing dependent variables in each vowel slice separately, rather than using the average value of the entire vowel. It is reasonable to guess that the falling low tone in an HL sequence may carry less creaky voice than other low tones, as it starts in a register space that is unusually high for a low tone. Conversely, it is possible that the rising high tone in a LH sequence is creakier than other high tones, as it starts at a much lower point in the register than the ultimate high target is located. I address this by examining the effect of contour on spectral tilt and HNR in the following sections.

4.4 Spectral tilt

Fig. 11 shows the average word spectral tilt values divided by tone and speaker.

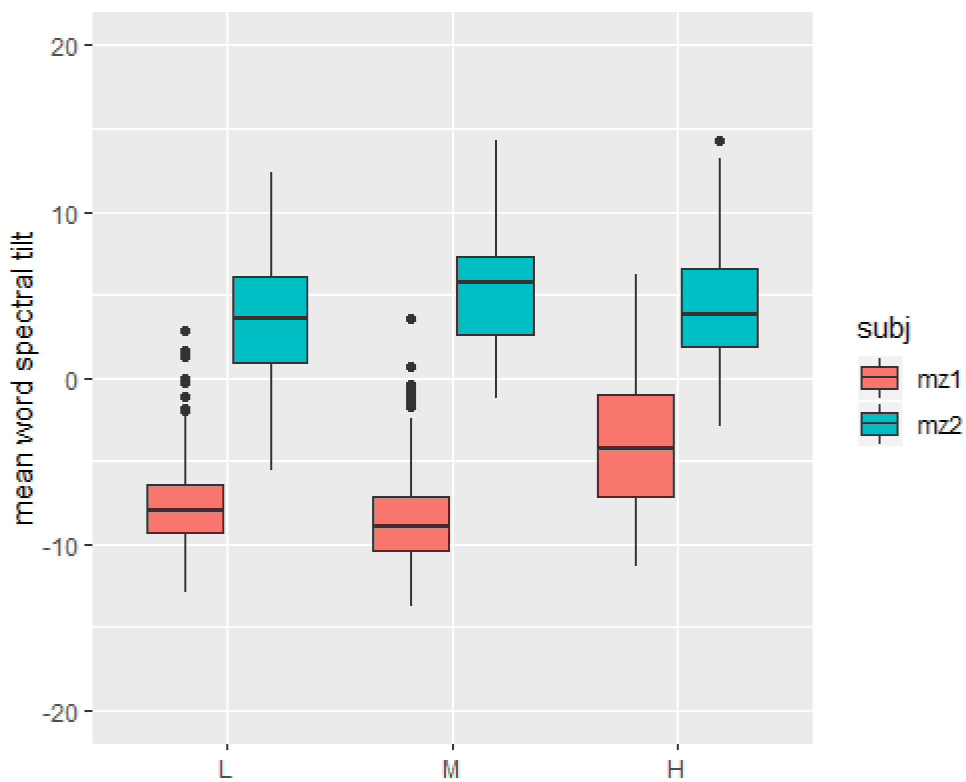

Figure 11: Average word spectral tilt

There is not an apparent division of the tonal space with regards to spectral tilt. There is also a great deal of inter-speaker variation. The mean values of the two speakers are quite different – speaker one has negative mean spectral tilt values in all tone levels, and speaker two has positive values. This can be interpreted as speaker one having a creakier voice in general, which is consistent with what the researcher observed. Interestingly, the mid and the low tone seem to pattern together in terms of spectral tilt for speaker one. It appears that the mid tone has slightly lower spectral tilt than even the low tone, and the high tone is marked with higher spectral tilt values. This is fairly consistent across syllables, as shown here in Fig. 12.

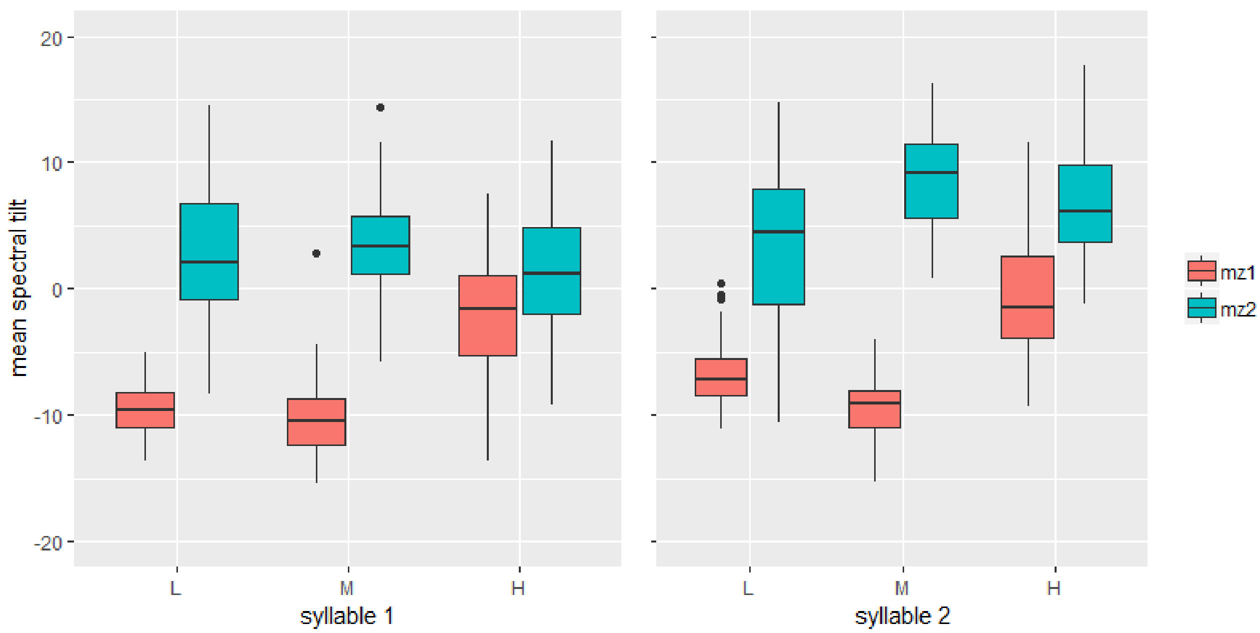

Figure 12: Average syllable spectral tilt by tone and speaker

For both speakers, spectral tilt values appear to raise from syllable one to syllable two. One possible interpretation of the data is that speakers use spectral tilt to mark word or phrase boundaries, which is consistent with the first experiment. Noting that speaker two here is the participant from the CV experiment, and that the CV words there and the second syllable of the words in this experiment are in the same environment (a word boundary where there is a slight pause), the behavior seems comparable. It appears that though there are consistent differences in spectral tilt, it does not mark low tones *in particular* as creaky. It will be shown that, in line with the first experiment, HNR does this more robustly.

I start by examining average spectral tilt level within each syllable for each tone. Different linear mixed effects models were built for each syllable and speaker with average spectral tilt value as dependent variable and tone as the independent variable, with *block/word* in the random effects structure. The models are reported in the following table.

speaker 1 syllable 1					
	β	SE	df	t-value	p
(intercept)	-9.64	0.33	78.78	-29.22	< 0.001*
mid	-0.33	0.47	79.11	-0.70	0.49
high	9.02	0.47	78.32	19.35	< 0.001*
random effects					
group	name	variance	σ	corr.	
word	(intercept)	20.99	4.58		
	block	1.43	1.20	-0.99	
residual		5.66	2.38		
fixed effects R^2: .65		model R^2: .81			

speaker 1 syllable 2					
	β	SE	df	t-value	p
(intercept)	-6.75	0.49	86.52	-13.89	< 0.001*
mid	-2.81	0.66	79.64	-4.24	< 0.001*
high	6.12	0.66	79.93	9.23	< 0.001*
random effects					
group	name	variance	σ	corr.	
word	(intercept)	19.20	4.38		
	block	2.35	1.53	-0.88	
residual		4.89	2.21		
fixed effects R^2: .56		model R^2: .82			

speaker 2 syllable 1					
	β	SE	df	t-value	p
(intercept)	2.41	0.46	76.83	5.22	< 0.001*
mid	1.00	0.63	69.29	1.57	0.06
high	-1.22	0.64	69.85	-1.92	0.12
random effects					
group	name	variance	σ	corr.	
word	(intercept)	0.00	0.00		
	block	0.01	0.11	NA	
residual		19.61	4.43		
fixed effects R^2: .04		model R^2: .05			

speaker 1 syllable 2					
	β	SE	df	t-value	p
(intercept)	3.17	0.80	79.22	3.95	< 0.001*
mid	5.64	1.02	69.87	5.51	< 0.001*
high	3.59	1.02	67.60	3.52	< 0.001*
random effects					
group	name	variance	σ	corr.	
word	(intercept)	5.39	2.32		
	block	0.002	0.05	1.00	
residual		17.71	4.21		
fixed effects R^2: .51		model R^2: .71			

Table 7: Linear mixed effects models of tone on mean syllable spec tilt

For speaker one in syllable one, there is a significant difference between low and high tones, but no significant difference between low and mid tones. Pairwise comparison indicates a significant difference between mid and high tones as well ($\beta = 9.34, SE = 0.47, df = 78.70, t - value = 20.05, p < 0.001$). In syllable two, the mid tone is lower than the low tone, and the high tone is significantly higher than both (M-H: $\beta = 8.93, SE = 0.64, df = 72.80, t - value = 14.01, p < 0.001$). For speaker two, a similar trend is observed. There are no significant differences between the low tone and other tone levels in syllable one. Post-hoc pairwise comparison indicates a significant difference between mid and high tones ($\beta = -2.22, SE = 0.62, df = 62.5, t - value = -3.59, p < 0.001$). In the second syllable, the low is significantly lower in spectral tilt than either the mid or high tones. There is a non-significant trend in which the mid shows higher spectral tilt than the high tone ($\beta = -2.05, SE = 0.89, df = 55.1, t-value = -2.30, p = 0.03$).

These results show that, while there are some differences between tone levels with regards to spectral tilt, it is not the case that low tones are categorically marked with low spectral tilt. In fact, the strongest differences in spectral tilt do not appear until the second syllable for both speakers, suggesting that spectral tilt is most influenced by the position of the vowel.

A more nuanced understanding of spectral tilt comes from considering changes in spectral tilt as the vowel progresses. Fig. 13 separates spectral tilt tracks by tone and speaker, showing some noteworthy trends.

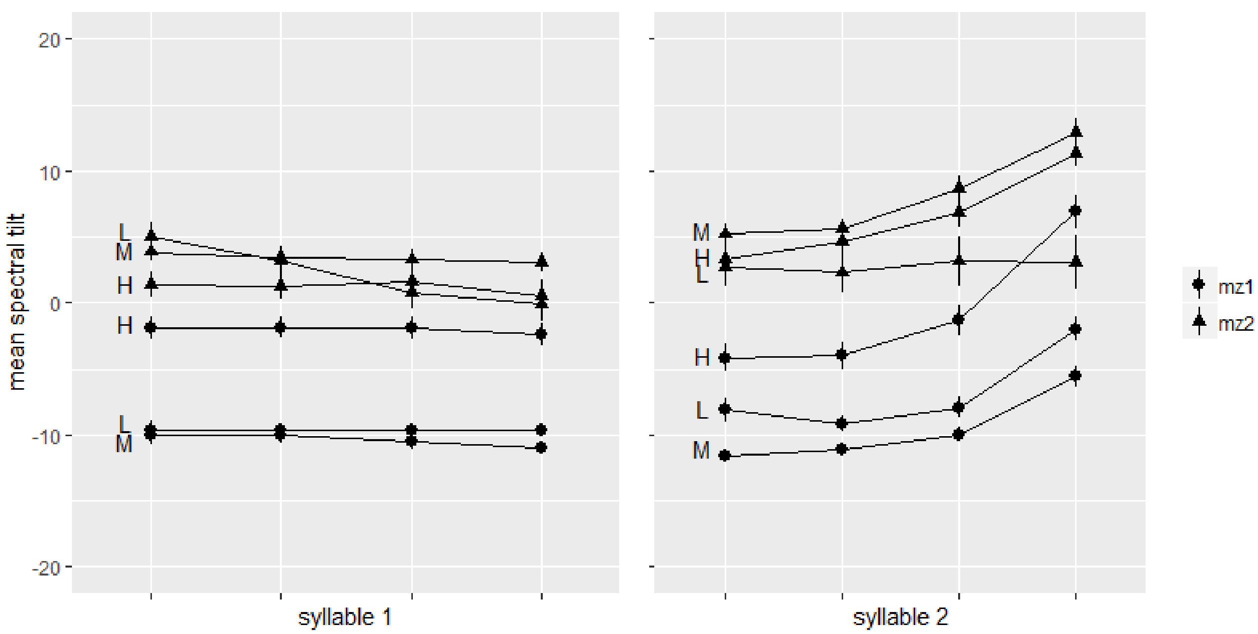

Figure 13: Average syllable spectral tilt over time by tone and speaker

There is a general upward trend in the second syllable that is absent in the first syllable for both speakers. The outputs of linear mixed effects models with spectral tilt value as dependent variable and vowel slice as independent variable with *block/word* as a random effect are shown below.[1]

[1]Here and in one other table, I omit information on random effects due to space concerns. Refer to the Appendix to see the full models.

speaker 1						speaker 1					
syllable 1;	low					syllable 2;	low				
	β	SE	df	t-value	p		β	SE	df	t-value	p
(intercept)	-9.58	0.34	49.83	-28.50	< 0.001*	(intercept)	-8.17	0.44	99.71	-18.48	< 0.001*
slice 2	-0.05	0.29	359.64	-0.18	0.86	slice 2	-1.06	0.52	331.99	-2.02	0.04
slice 3	-0.11	0.29	359.64	-0.37	0.72	slice 3	0.19	0.53	332.53	0.34	0.72
slice 4	-0.06	0.29	360.12	-0.19	0.85	slice 4	6.22	0.56	338.91	11.08	< 0.001*
fixed effects R^2: .0002		model R^2: .31				fixed effects R^2: .31		model R^2: .40			
speaker 1						speaker 1					
syllable 1;	mid					syllable 2;	mid				
	β	SE	df	t-value	p		β	SE	df	t-value	p
(intercept)	-9.70	0.49	86.52	-13.89	< 0.001*	(intercept)	-10.55	0.31	52.72	3.95	< 0.001*
slice 2	0.03	0.66	79.64	-4.24	0.92	slice 2	0.50	0.29	366.15	5.51	0.08
slice 3	-0.45	0.66	79.93	9.23	0.16	slice 3	1.58	0.29	366.15	3.52	< 0.001*
slice 4	-0.91	0.66	79.93	9.23	0.005*	slice 4	5.97	0.29	366.15	9.23	< 0.001*
fixed effects R^2: .02		model R^2: .43				fixed effects R^2: .38		model R^2: .71			
speaker 1						speaker 1					
syllable 1;	high					syllable 2;	high				
	β	SE	df	t-value	p		β	SE	df	t-value	p
(intercept)	-0.24	0.48	40.51	-0.50	0.62	(intercept)	-5.07	0.72	46.29	-7.05	< 0.001*
slice 2	-0.01	0.35	367.03	-0.04	0.97	slice 2	0.14	0.60	365.27	0.23	0.82
slice 3	-0.02	0.35	367.03	-0.07	0.94	slice 3	2.80	0.60	365.27	4.69	< 0.001*
slice 4	-0.53	0.35	367.03	-1.54	0.12	slice 4	11.33	0.60	365.27	18.98	< 0.001*
fixed effects R^2: .002		model R^2: .70				fixed effects R^2: .37		model R^2: .67			
speaker 2						speaker 2					
syllable 1;	low					syllable 2;	low				
	β	SE	df	t-value	p		β	SE	df	t-value	p
(intercept)	4.98	0.78	47.59	6.31	< 0.001*	(intercept)	2.67	1.02	43.84	2.61	0.01*
slice 2	-1.91	0.66	339.92	-2.87	0.004*	slice 2	-0.63	0.84	274.53	-0.75	0.45
slice 3	-4.23	0.67	340.45	-6.34	< 0.001*	slice 3	0.21	0.90	276.90	0.23	0.82
slice 4	-5.22	0.68	341.07	-7.72	< 0.001*	slice 4	-0.71	1.02	279.65	-0.70	0.49
fixed effects R^2: .10		model R^2: .44				fixed effects R^2: .005		model R^2: .07			
speaker 2						speaker 2					
syllable 1;	mid					syllable 2;	mid				
	β	SE	df	t-value	p		β	SE	df	t-value	p
(intercept)	3.54	0.36	138.74	9.79	< 0.001*	(intercept)	5.74	0.49	98.57	11.81	< 0.001*
slice 2	-0.40	0.46	362.76	-0.88	0.38	slice 2	0.38	0.57	345.75	0.66	0.51
slice 3	-0.60	0.46	362.76	-1.31	0.19	slice 3	3.61	0.57	345.75	6.32	< 0.001*
slice 4	-0.77	0.46	362.76	-1.68	0.09	slice 4	10.33	0.58	346.15	17.37	< 0.001*
fixed effects R^2: .006		model R^2: .21				fixed effects R^2: .40		model R^2: .59			
speaker 2						speaker 2					
syllable 1;	high					syllable 2;	high				
	β	SE	df	t-value	p		β	SE	df	t-value	p
(intercept)	1.59	0.54	167.13	2.93	< 0.001*	(intercept)	3.25	0.65	98.53	5.00	< 0.001*
slice 2	-0.15	0.71	379.25	-0.21	0.84	slice 2	1.31	0.74	394.18	1.77	0.82
slice 3	0.20	0.71	379.25	0.29	0.77	slice 3	3.97	0.74	394.18	5.33	< 0.001*
slice 4	-0.80	0.71	379.25	-1.13	0.26	slice 4	9.49	0.75	394.28	12.72	< 0.001*
fixed effects R^2: .005		model R^2: .07				fixed effects R^2: .28		model R^2: .38			

Table 8: Linear mixed effects models of vowel slice on spectral tilt by tone

In syllable one, speaker one shows no significant differences in spectral tilt between vowel slices for any tone level, except for a slight drop towards the end of mid tones. In syllable two low tones, the fourth slice is significantly higher than all previous tone levels. In mid tones, pairwise comparison indicates that the third slice is higher in spectral tilt than the preceding two slices, and the fourth slice is higher than all other slices. The high tones display the exact same pattern. So, while spectral tilt is relatively stable in syllable one, there is a significant increase in syllable two, regardless of tone level.

Despite having disjoint spectral tilt ranges, speakers one and two behave in much the same way. The difference is in the behavior of speaker two's low tone, which lowers significantly throughout the course of first-syllable vowels, but does not change significantly in syllable two. The mid and high tones pattern like speaker one. The models and further pairwise comparison find no significant differences between any slice at either tone level in syllable one. In syllable two, starting at slice three in both mid and high tones, spectral tilt increases significantly. While the dichotomy in low tone behavior from syllable to syllable in speaker two is noteworthy, this result is interesting in that it indicates that both speakers behave similarly with regards to spectral tilt. The overall conclusion is that while tones do differ in spectral tilt in Yoruba, changes over time can mark word or phrase boundaries, rather than marking low tones in particular as creaky.

Knowing about the contours present with certain tone sequences, it is reasonable to guess that tone sequence may affect the spectral measures. To investigate this possibility, linear mixed effects models with tone sequence as dependent variable and average syllable spectral tilt as independent variable with *block/slope* as a random effect were created for each speaker and syllable. The results are shown in the following table, where an LL sequence is the intercept.[2]

[2] For full pairwise comparisons of tone sequence models, refer to the Appendix.

speaker 1 syllable 1					
	β	SE	df	t-value	p
(intercept)	-9.56	0.49	119.80	-19.40	< 0.001*
HH	9.96	0.68	115.71	14.60	< 0.001*
HL	9.72	0.69	118.18	14.06	< 0.001*
HM	6.22	0.68	115.71	9.12	< 0.001*
LH	-0.35	0.68	119.81	-0.51	0.61
LM	0.08	0.68	115.71	0.12	0.91
MH	-0.69	0.68	115.71	-1.02	0.31
ML	0.17	0.69	116.53	0.25	0.80
MM	-1.00	0.69	117.43	-1.44	0.15
random effects					
group	name	var.	σ	corr.	
word	(int.)	17.78	4.22		
	block	1.46	1.21	-1.00	
residual		5.56	2.36		
fixed effects R^2: .68		modelR^2: .80			

speaker 1 syllable 2					
	β	SE	df	t-value	p
(intercept)	-6.79	0.55	78.78	-12.39	< 0.001*
HH	9.80	0.73	69.72	13.39	< 0.001*
HL	1.69	0.76	74.98	2.23	0.03
HM	-2.17	0.73	69.72	-2.96	0.004*
LH	3.09	0.73	69.96	4.22	< 0.001*
LM	-2.13	0.73	69.72	-2.91	0.005*
MH	9.62	0.73	69.72	13.14	< 0.001*
ML	-0.54	0.85	96.28	-0.64	0.53
MM	-2.35	0.74	72.41	-3.17	0.002*
random effects					
group	name	var.	σ	corr.	
word	(int.)	28.63	5.35		
	block	2.46	1.57	-0.99	
residual		4.87	2.21		
fixed effects R^2: .69		modelR^2: .85			

speaker 2 syllable 1					
	β	SE	df	t-value	p
(intercept)	3.95	0.76	291.97	5.22	< 0.001*
HH	-3.58	1.06	291.97	-3.39	< 0.001*
HL	-2.25	1.07	291.97	-2.10	0.04
HM	-2.49	1.06	291.97	-2.32	0.02
LH	-4.00	1.11	291.98	-3.62	< 0.001*
LM	-0.88	1.09	291.98	-0.81	0.42
MH	0.22	1.05	291.97	0.21	0.83
ML	-0.54	1.06	291.97	-0.51	0.61
MM	-1.35	1.06	291.97	-1.29	0.21
random effects					
group	name	var.	σ	corr.	
word	(int.)	0.00	0.01		
	block	0.00	0.003	-1.00	
residual		0.00	4.45		
fixed effects R^2: .10		model R^2: .10			

speaker 2 syllable 2					
	β	SE	df	t-value	p
(intercept)	-3.58	1.27	104.91	-2.83	0.006*
HH	10.15	1.53	81.82	6.63	< 0.001*
HL	9.86	1.62	83.07	6.10	< 0.001*
HM	12.68	1.54	84.66	8.24	< 0.001*
LH	11.53	1.55	83.31	7.44	< 0.001*
LM	11.85	1.55	84.92	7.66	< 0.001*
MH	9.46	1.53	80.87	6.20	< 0.001*
ML	8.45	1.80	102.44	4.69	< 0.001*
MM	13.24	1.54	83.99	8.59	< 0.001*
random effects					
group	name	var.	σ	corr.	
word	(int.)	4.25	2.06		
	block	0.06	0.24	-1.00	
residual		17.60	4.20		
fixed effects R^2: .30		model R^2: .38			

Table 9: Linear mixed effects models of tone sequence on syllable spec tilt

For speaker one in syllable one, the intitial L in an LL sequence has significantly lower spectral tilt than any H-initial sequence. In fact, post-hoc pairwise comparison shows a significant difference for every L-initial to H-initial and M-initial to H-initial comparison, with no differences for any initial-L, initial-M, or L to M comparisons. This is in line with what is seen in Figures 12 + 13. Pairwise comparison also finds significant differences between initial H in an HM versus an HH or HL such that the H in an HM sequence has lower spectral tilt, though HH and HL do not themselves differ significantly (HM-HH: $\beta = 3.73, SE = 0.67, df = 111, t-value = 5.60, p < 0.001$; HM-HL: $\beta = 3.50, SE = 0.68, df = 114, t-value = 5.17, p < 0.001$).

In syllable two, the final L in an LL sequence has significantly lower spectral tilt than any final H, and significantly higher spectral tilt than any final M. Post-hoc comparison shows all final-M sequences to be significantly lower in spectral tilt than any final-H sequence. There is no significant difference between LL and ML sequences, or LM and ML sequences, but a non-significant positive trend between LL and HL. Comparing ML and HL directly, the same pattern is found. Given that HL-final L is actually a falling contour, it is not surprising that low tones in that environment have somewhat higher spectral tilt. The same difference appears when comparing LH to HH and MH sequences. Though HH and MH do not differ significantly from each other, final H in both show significantly higher spectral tilt than an LH-final H (HH: $\beta = 6.71, SE = 0.69, df = 60, t-value = 9.75, p < 0.001$; MH: $\beta = 6.53, SE = 0.69, df = 60, t-value = 9.49, p < 0.001$). This is again the effect of the contour. The effect is so strong that HL-final L and LH-final H do not differ significantly from each other in terms of spectral tilt ($\beta = -1.41, SE = 0.71, df = 65.7, t-value = -1.94, p < 0.57$). There are no significant differences between sequences with a final M.

For speaker two in syllable one, there is a significant difference or non-significant negative trend between the initial L in an LL sequence and every H-initial sequence. LL also differs significantly from LH, which shows a significant negative slope even further from LL than that of the H-initial sequences. Pairwise comparison shows that this holds for LH and LM as well ($\beta = -3.12, SE = 1.12, df = 292, t-value = -2.78, p = 0.006$), and that LH-initial L has lower spectral tilt than any M-initial sequence. This is also true of HH, which shows a significant difference or non-significant negative trend between any M-initial sequence. No initial-H only or initial-M only differences are found. Significant differences do appear between HH and MH sequences, where initial H is lower

($\beta = -3.79, SE = 1.03, df = 292, t - value = -3.68, p = 0.008$). LL and LM do not differ significantly from any M-initial sequence.

In syllable two, the final L tone of an LL is significantly lower in spectral tilt than any other sequence. Post-hoc pairwise compairson shows that ML and HL do not differ significantly, suggesting that the HL-final contour does not significantly alter spectral tilt for speaker two. The same holds for the LH contour, as there are no final-H sequence significant differences. No final-M sequences show significant differences with each other. MM-final M shows significantly higher spectral tilt than either its ML counterpart ($\beta = -4.79, SE = 1.55, df = 80.8, t - value = -3.09, p = 0.002$) or MH counterpart ($\beta = -3.77, SE = 1.22, df = 49.2, t - value = -3.10, p = 0.003$), which do not themselves differ. LM and ML-final tones show no significant differences.

These differences in behavior regarding spectral tilt lead to the conclusion that spectral tilt fluctuates a result of word or phrase position, or possibly changes gradiently as a result of moving from one tone level to another – but it is not a robust indicator of creaky voice in the low tone in Yoruba. The effect of contour varied between speakers, and was most apparent in the second syllable, where spectral tilt changes over the duration of the vowel are also most apparent. The positional variation may be allophonic – Blankenship (1997, 2002) find that contrastive non-modal phonation is more conspicuous and longer in duration than allophonic non-modal phonation. Given the positional nature of the variation and considering the more robust results for HNR described below, the results of this paper support those findings. Expanding the area of investigation beyond just words of one shape has provided a clearer conception of how creaky voice is implemented in the language.

4.5 HNR

For HNR, the generalization is clearer, both across syllables and across speakers. Higher HNR marks high and mid tones, while much lower HNR marks low tones. The difference has a categorical appearance, in that high and mid tones occupy a similar space that is set apart from the space occupied by the low tones. It is also noteworthy that although the speakers differed in spectral tilt, their behavior with regards to HNR is almost identical. I interpret this as an indication that HNR is the most robust acoustic implementation of creaky voice in Yoruba.

Figure 6 suggests that there is not much inter-speaker variation with regards to HNR, and initial models considering each speaker individually confirm this – the patterns are the same, both for HNR over the course of each syllable and for HNR differences between tone levels in each vowel slice. As such, data for each speaker is pooled for models assessing HNR for the sake of simplicity.

Fig. 14 shows the average word spectral HNR values divided by tone.

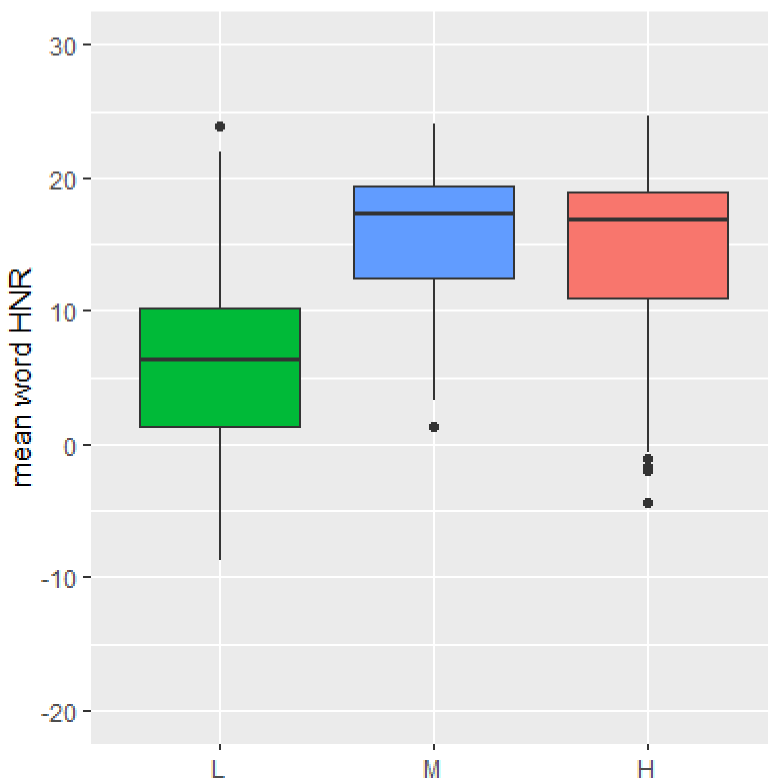

Figure 14: Average word HNR

Though there is a great deal of overlap due to taking the average of the entire word, the difference appears to have a categorical quality, such that mid and high tones pattern together with regards to HNR, while low tones generally show lower HNR. The pattern is clearer when looking at HNR in each syllable individually. This is shown in Fig. 15.

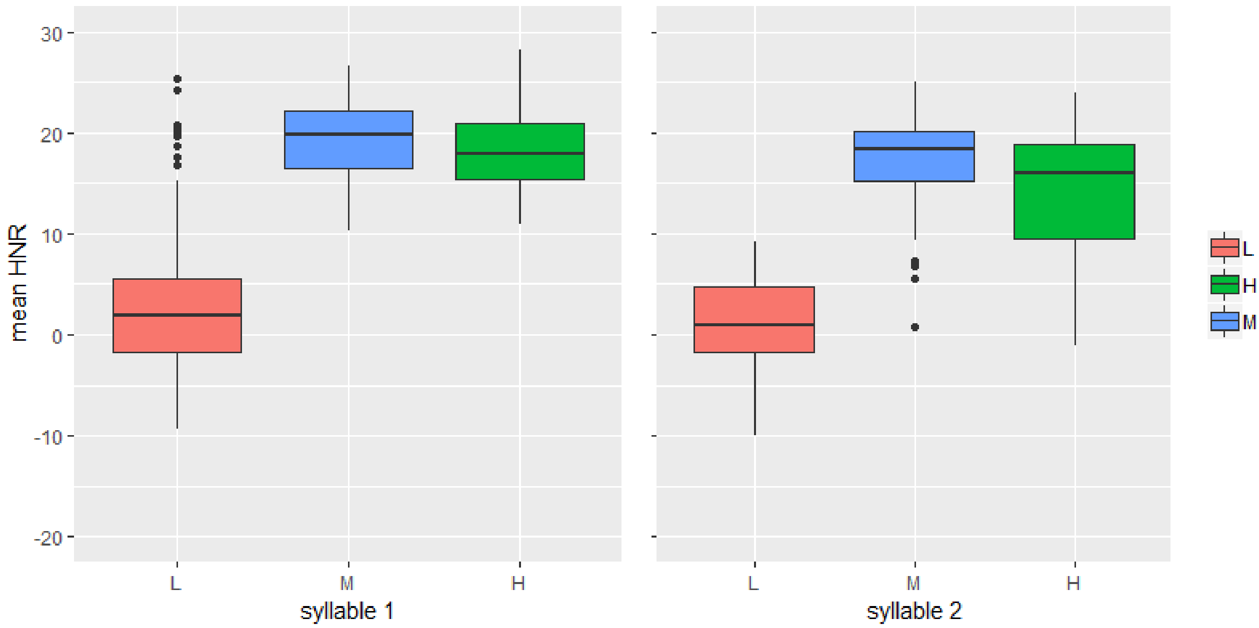

Figure 15: Average syllable HNR by tone

I first examine HNR level in each syllable for each tone. The output of linear mixed effects models with syllable HNR value as the dependent variable and tone as the independent variable with *subject* and *block/word* in the random effects structure are shown in Table 10:

syllable 1						syllable 2					
	β	SE	df	t-value	p		β	SE	df	t-value	p
(intercept)	2.15	0.56	79.00	3.85	< 0.001*	(intercept)	1.28	0.90	25.10	1.43	0.17
mid	17.00	0.79	78.30	21.64	< 0.001*	mid	16.12	1.16	78.61	13.99	< 0.001*
high	15.95	0.79	78.86	20.26	< 0.001*	high	12.65	1.16	78.72	10.91	< 0.001*
random effects						random effects					
group	name	var.	σ	corr.		group	name	var.	σ	corr.	
word	(int.)	4.36	2.09			word	(int.)	17.60	4.20		
	block	0.62	0.79	-0.10			block	0.0002	0.01	-1.00	
	subject	0.00	0.00				subject	0.26	0.51		
residual		13.74	3.71			residual		5.79	2.41		
fixed effects R^2: .73		model R^2: .83				fixed effects R^2: .67		model R^2: .92			

Table 10: Linear mixed effects models of tone on mean syllable HNR

In both syllables, the low tone differs significantly from the mid and high tones. Pairwise comparison of the mid and high tone in the first syllable does not find a significant difference between the two ($\beta = -1.05, SE = 0.79, df = 78.2, t - value = -1.34 p = 0.18$). In the second syllable, the difference between the mid and high tone is enough to be found statistically significant ($\beta = -3.56, SE = 1.16, df = 78.1, t - value = -2.08 p = 0.003$). Given the much greater difference between low tones and high or mid tones compared to the much smaller difference between high and mid tones, these findings support the idea that low HNR marks low tones in Yoruba as creaky categorically, rather than just being the result of speakers lowering in their register.

Fig. 16 plots the progression of HNR over the course of the vowel in each syllable.

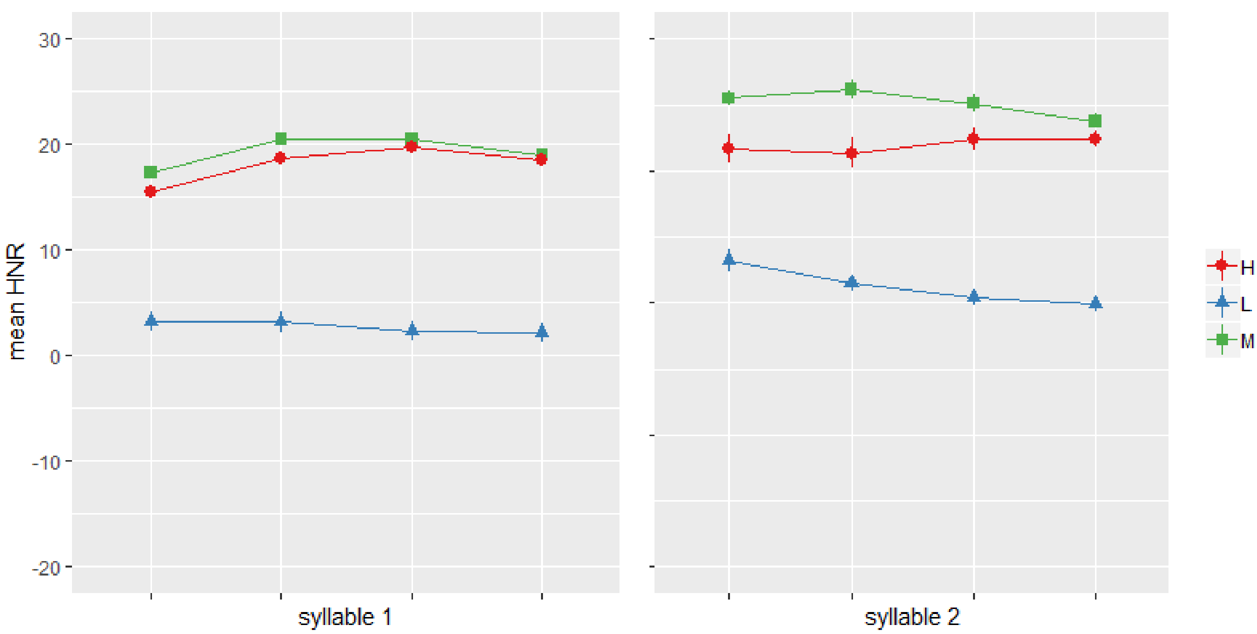

Figure 16: Average syllable HNR (dB) over time by tone

In syllable one, the most movement appears between the first and second slice for mid and high tones, though the low is quite stable. In the second syllable, there is more movement in general. The downward trend between the first and second slice of the low tone is most likely due to the falling contour in HL sequences, where the final L starts in a higher register space than usual and so is less creaky. The same applies to the high tone, which shows a slight upward trend that is likely due to the rising contour in an LH sequence. The differences between the three tone levels also support a categorical analysis of HNR differences. The mid and high tone are very close and far removed from the low tone, and the mid tone has even slightly higher HNR than the high tone. This can be seen in the mean values shown in Fig. 1, where the mean HNR of a mid tone is slightly higher than that of a high tone for both speakers. If HNR differences in Yoruba were merely the result of speakers lowering progressively in their register, this is not what we would expect to find.

To formally analyze HNR differences between vowel slices, linear mixed effects models with mean syllable HNR value as dependent variable and vowel slice as independent variable with *subject* and *block/word* in the random effects structure were created. The results are given in Table 11.

syllable 1;	low					syllable 2;	low				
	β	SE	df	t-value	p		β	SE	df	t-value	p
(intercept)	2.56	1.75	1.37	1.37	0.33	(intercept)	2.87	0.71	18.33	4.02	< 0.001*
slice 2	-0.00	0.52	764.89	-0.00	0.99	slice 2	-1.40	0.36	751.26	-3.88	< 0.001*
slice 3	-0.84	0.52	764.93	-1.62	0.11	slice 3	-2.53	0.36	751.30	-7.04	< 0.001*
slice 4	-1.09	0.52	764.81	-2.09	0.04	slice 4	-3.09	0.37	751.29	-8.27	< 0.001*
fixed effects R^2: .005		model R^2: .46				fixed effects R^2: .06		model R^2: .47			
syllable 1;	mid					syllable 2;	mid				
	β	SE	df	t-value	p		β	SE	df	t-value	p
(intercept)	17.50	0.74	5.41	23.79	< 0.001*	(intercept)	17.97	0.68	13.31	26.22	< 0.001*
slice 2	3.16	0.29	779.24	11.06	< 0.001*	slice 2	1.86	0.33	766.53	5.70	< 0.001*
slice 3	3.26	0.29	779.24	11.40	< 0.001*	slice 3	-0.36	0.33	766.53	-1.10	0.27
slice 4	1.72	0.29	779.24	6.03	< 0.001*	slice 4	-3.74	0.33	766.11	-11.38	< 0.001*
fixed effects R^2: .09		model R^2: .54				fixed effects R^2: .16		model R^2: .55			
syllable 1;	high					syllable 2;	high				
	β	SE	df	t-value	p		β	SE	df	t-value	p
(intercept)	15.48	1.10	1.60	14.04	0.01*	(intercept)	13.54	1.27	16.75	10.68	< 0.001*
slice 2	3.19	0.33	772.90	9.56	< 0.001*	slice 2	2.19	0.33	777.75	6.55	< 0.001*
slice 3	4.20	0.33	772.90	12.60	< 0.001*	slice 3	0.25	0.33	777.75	0.74	0.46
slice 4	3.21	0.33	772.60	9.58	< 0.001*	slice 4	-0.87	0.34	778.67	-2.54	0.01*
fixed effects R^2: .11		model R^2: .48				fixed effects R^2: .03		model R^2: .75			

Table 11: Linear mixed effects models of vowel slice on HNR by tone

In syllable one, there is a very slight downward trend in the low tones. Pairwise comparison finds no significant differences between slices for the low tone. For the mid tone, every slice has significantly higher HNR than the first slice. Pairwise comparison finds the fourth slice to have significantly lower HNR than the second slice ($\beta = -1.44, SE = 0.29, df = 779, t - value = 5.02p < 0.001$) and third slice ($\beta = -1.54, SE = 0.29, df = 779, t - value = 5.37p < 0.001$), indicating a slight drop off from maximum HNR level at the end of the first syllable. The profile of the high tone is similar, with significantly higher HNR in every slice proceeding the first, the third slice significantly higher than the second ($\beta = 1.02, SE = 0.33, df = 773, t - value = -3.06p = 0.002$), and the fourth slice significantly lower than the third ($\beta = -1.00, SE = 0.33, df = 773, t - value = 2.99, p = 0.003$).

In syllable two, HNR in the low tone trends downward as the vowel progresses. Pairwise comparison finds the second slice to be significantly higher in HNR than the third slice ($\beta = -1.13, SE = 0.35, df = 751, t - value = 3.22p = 0.001$) and fourth slice ($\beta = -1.70, SE = 0.37, df = 752, t - value = 4.61p < 0.001$). The mid tone shows a significant rise before dropping off at the end of the vowel. Pairwise comparison indicates that – other than the first and third slice – every slice is significantly different from every other slice. The pattern in the high tone is identical, with a significant increase in HNR between the first and second slices, a significant decrease at the end of the vowel, and a significant difference for all level comparisons except the first and third slice. HNR in the low tone decreases over the course of the vowel in both syllables. High and mid tone HNR increases before dropping off towards the end of the vowel in both syllables. This mirrors the results for spectral tilt, where the mid and high pattern together apart from the low.

As with spectral tilt, the effect of contour on HNR value was examined via linear mixed effects models with mean syllable HNR as the dependent variable, tone sequence as independent variable, and *subject* and *block/word* in the random effects structure. The results are shown here in Table 12, where an LL sequence is the reference level.

syllable 1						syllable 2					
	β	SE	df	t-value	p		β	SE	df	t-value	p
(intercept)	1.85	0.97	72.24	1.91	0.06	(intercept)	-0.33	0.95	34.27	-0.35	0.73
HH	16.66	1.37	71.81	12.17	< 0.001*	HH	18.03	1.27	71.49	14.24	< 0.001*
HL	15.88	1.37	73.17	11.55	< 0.001*	HL	4.80	1.27	73.12	3.77	< 0.001*
HM	16.29	1.37	71.78	11.91	< 0.001*	HM	17.61	1.26	70.86	13.94	< 0.001*
LH	-0.64	1.37	72.56	-0.46	0.64	LH	6.82	1.26	70.60	5.41	< 0.001*
LM	2.48	1.37	72.50	1.81	0.08	LM	16.90	1.26	70.61	13.41	< 0.001*
MH	17.56	1.37	71.22	12.86	< 0.001*	MH	18.13	1.27	71.42	14.33	< 0.001*
ML	16.93	1.37	71.87	12.37	< 0.001*	ML	-0.14	1.27	72.59	-0.11	0.91
MM	17.78	1.37	72.52	12.96	< 0.001*	MM	18.12	1.26	71.03	14.34	< 0.001*
random effects						random effects					
group	name	var.	σ	corr.		group	name	var.	σ	corr.	
word	(int.)	7.04	2.65			word	(int.)	5.03	2.43		
	block	0.62	0.79	-0.39			block	0.01	0.12	1.00	
subject	(int.)	0.00	0.00			subject	(int.)	0.22	0.47		
residual		13.75	3.71			residual		5.77	2.40		
fixed effects R^2: .73		modelR^2: .83				fixed effects R^2: .82		modelR^2: .92			

Table 12: Linear mixed effects models of tone sequence on syllable HNR

In the first syllable, initial low tone sequences show significantly lower HNR than initial high or mid tone sequences. An LM-intial L may be slightly higher than an LL-initial L, though the positive slope of 2.48 is not found statistically significant. The same is true of LM and LH, where pair-wise comparison finds a non-significant trend such that LH-initial L has lower HNR ($\beta = -3.11, SE = 1.37, df = 72.8, t - value = -2.27, p = 0.03$). No significant differences are found between any H-initial sequence and M-initial sequence, nor for any comparison of initial-H sequences with each other or initial-M sequences with each other.

In the second syllable, we see the effect of the contour. LL and ML-final L do not differ, but HL-final L is significantly higher than both. This is because of the falling contour in HL sequences. Conversely, while HH and MH-final H do not differ from each other and are similar to their initial H counterparts, LH-final H shows significantly lower HNR than both due to the rising contour (HH: $\beta = 11.52, SE = 0.59, df = 563, t - value = 19.43, p < 0.001$; HM: $\beta = 11.22, SE = 0.58, df = 562, t - value = 19.301, p < .001$). There are no HH or MH-final H differences with M-final sequences, nor any M-final only sequence differences.

4.6 Discussion

The results of the second experiment show that HNR is the most robust indicator of creaky voice in Yoruba, and that spectral tilt may mark word or phrase boundaries rather than marking low tones in particular. It was also shown that, despite having different spectral tilt ranges, the two speakers exhibit similar patterns of behavior that group the mid and high tone together, separate from the low tone. HNR values were somewhat higher in mid tones than high tones. If differences in HNR are gradient as speakers lower in the register, then the mid tone should carry slightly lower HNR values than the high tone – not the other way around. These results strongly support the hypothesis that Yoruba low tones are marked categorically as creaky with low HNR.

Examining tone sequences, the expected effects of the tonal contours were found. A falling L showed higher HNR and a rising H showed lower HNR than their flatter counterparts. Apart from cases with contour tones, there was not an apparent influence of preceding/proceeding tone on the spectral measures, such that non-contour tones showed similar values in both syllable positions, particularly for HNR.

5 Conclusion

What correlate most consistent
syllable to syllable
within syllable
sequence
gradient vs categorical

This paper described a pair of experiments designed to address the question of creaky voice in Yoruba. While there is sparse reference to creaky voice in the Yoruba low tone (Welmers 1974; Yu 2010), there is only one previous acoustic study examining non-modal phonation in the language, where it was found that the low tone does have a creaky character that the mid and high tone do not (Hayward et al. 2004).

The first experiment presented here sought to replicate the results of Heyward et al., measuring creaky voice in Yoruba CV words via spectral tilt and HNR. The results indicated that HNR is the most robust correlate of creaky voice, and that low tones were marked categorically in this way. Variation in spectral tilt only occurs as the vowel progresses – at the beginning of vowels levels of spectral tilt are similar between each tone level. Notably, for both HNR and spectral tilt, the vowel-course variation seems to group the mid and high together, while the low patterns differently alone. That the tones pattern together in this way is further evidence of a phonetic grouping of the mid and high tone that is distinct from the low tone.

The second experiment looked to expand previous results to words of different shapes by examining creaky voice in CVCV words. Here as well, HNR emerged as the more consistent indicator of creak in the low tones, with consistent behavior across syllables and speakers. The difference also appeared to be categorical – high and mid tones have similar HNR profiles, while low tones are markedly different. Mid tones also show higher HNR than high tones, which is not what we would observe if the difference was the result of gradual lowering in one's register. For spectral tilt, there was no clear grouping of tones based on the mean values alone. Despite the distinct ranges of the two speakers, change in spectral tilt over time was similar between the two. There was almost no change throughout the first syllable, but a sharp rise in spectral tilt towards the end of the second syllable. This is in line with what was seen in experiment one, and suggests possible positional, allophonic variation in spectral tilt. Lastly, questions related to the sequence of tones were addressed, showing that measures of creak vary between tone levels in the cases where a contour is present. Other than sequences with contours, no cross-syllable tone sequence differences were apparent.

In the future, adding more speakers will improve the quality of this study. As both speakers recorded so far were male, it might be useful to record a female speaker to get a fuller range of data. There are also possible enhancements to data collection – having access to an electroglottograph (EGG) would give precise measures of the "closed quotient" (CQ) which is informative for non-modal phonation studies. Incorporating a measure of F1-F0 as in Hayward et al. (2004) may prove more useful than spectral tilt as well, as F1-F0 is more resilient in the presence of a variety of vowels and tones.

Another possible direction for future research is perception. It has been established that Yoruba low tones carry creaky voice. Are listeners then sensitive to the creak? For instance, if a low tone had its creaky quality removed, is there a chance a listener might mis-perceive it as a mid tone instead? Conversely, if creaky voice were added to a high or mid tone, might a listener perceive it as a lower-level tone? The answers to these questions are sure to be interesting, as they potentially show that Yoruba speakers use information other than just F0 to differentiate tones, but a full-scale investigation of perception is beyond the scope of this paper. Potential perceptual salience of creaky voice in Yoruba is also interesting in that it would place Yoruba in a growing body of evidence that the boundary between tone and register languages, where phonation type is contrastive, is "fuzzy" (Abramson and Luangthongkum 2009).

References

Abramson, A. and Luangthongkum, T. (2009). A fuzzy boundary between tone language and voice-register languages. In *Frontiers in Phonetics and Speech Science*, pages 149–155. The Commercial Press, Beijing, China.

Akinlabi, A. (1985). *Tonal underspecification and Yoruba tones*. PhD thesis, Nigeria: University of Ibadan.

Akinlabi, A. and Liberman, M. (1995). On the phonetic interpretaion of the yoruba tonal system.

Bakare, C. (1995). Discrimination and identification of yoruba tones: perception experiments and acoustic analysis. In *Language in Nigeria*, pages 435–450. Ibadan: Group Publishers.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.

Blankenship, B. (1997). The time course of breathiness and laryngealization in vowels.

Blankenship, B. (2002). The timing of non-modal phonation in vowels. *Phonetics*, 30:163–191.

Boersma, P. (1993). Accurate short term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences*, 17:97–110.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, 5:9/10:341–345.

Bradley, D. (1982). Register in burmese. In Bradley, D., editor, *Papers in South-East Asian Linguistics 8: Tonation*, pages 117–132. Canberra: The Australian National University.

Brunelle, M. (2012). Dialect experience and perceptual integrality in phonological registers: Fundamental frequency, voice quality and the first formant in cham. *Journal of the ASA*, 131:3088–3102.

Connell, B. and Ladd, D. R. (1990). Aspects of pitch realisation in yoruba. *Phonology*, 7(1):1–29.

Davidson, D. (1991). An acoustic study of so-called creaky voice in tianjin mandarin. *UCLA Working Papers in Phonetics*, 78:50–57.

Gordon, M. and Ladefoged, P. (2001). Phonation types: a cross-linguistic review. *Journal of Phonetics*, 29:383–406.

Harrison, P. (1996). An experiment with tone. *UCL Working Papers in Linguistics*, 8:575–593.

Hayward, K., Watkins, J., and Oyètádé, A. (2004). The phonetic interpretation of register: evidence from yorùbá. In Local, J., Ogden, R., and Temple, R., editors, *Phonetic Interpretation: Papers in Laboratory Phonology VI*, Papers in Laboratory Phonology, page 305–321. Cambridge University Press.

Huffman, M. (1987). Measures of phonation type in hmong. *Journal of the Acoustical Society of America*, 81:495–504.

Keating, P., Garellek, M., and Kreiman, J. (2015). Acoustic properties of different kinds of creaky voice.

Klatt, D. and Klatt, L. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J. Acoust. Soc. Am.*, 87:820–857.

Ladefoged, P. (1971). Preliminaries to linguistic phonetics.

Laver, P. (1980). The phonetic description of voice quality.

Peirce, J. (2007). Psychopy – pyschophysics software in python. *Journal of Neuroscience Methods*, 162 (1-2):8–13.

Pulleyblank, D. (1986). *Tone in Lexical Phonology*. PhD thesis, Reidel, Dordrecht.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Silverman, D. (1997). Laryngeal complexity in otomanguean vowels. *Phonology*, 14:235–261.

Silverman, D., Blankenship, B., Kirk, P., and Ladefoged, P. (1995). Phonetic structures in jalapa mazatec. *Anthropolog. Linguist.*, 37:70–88.

Stahlke, H. (1974). The development of three-way tonal contrast in yoruba. In Voeltz, E., editor, *Third Annual Conference on African Linguistics*, pages 138–145. Bloomington: Indiana University.

Welmers, W. (1974). *African Language Structures*. University of California Press.

Yu, K. and Lam, H. W. (2014). The role of creaky voice in cantonese tonal perception. *Journal of the Acoustical Society of America*, 136(3):1320–1333.

Yu, K. M. (2010). Laryngealization and features for chinese tonal recognition.

A Omitted random effects

speaker 1											
syllable 1;	low					syllable 2;	low				
	β	SE	df	t-value	p		β	SE	df	t-value	p
(intercept)	-9.58	0.34	49.83	-28.50	< 0.001*	(intercept)	-8.17	0.44	99.71	-18.48	< 0.001*
slice 2	-0.05	0.29	359.64	-0.18	0.86	slice 2	-1.06	0.52	331.99	-2.02	0.04
slice 3	-0.11	0.29	359.64	-0.37	0.72	slice 3	0.19	0.53	332.53	0.34	0.72
slice 4	-0.06	0.29	360.12	-0.19	0.85	slice 4	6.22	0.56	338.91	11.08	< 0.001*
random effects						random effects					
group	name	var.	σ	corr.		group	name	var.	σ	corr.	
word	(int.)	2.45	1.57			word	(int.)	4.66	2.16		
	block	0.64	0.25	-0.47			block	0.52	0.72	-0.81	
residual		4.01	2.10			residual		14.16	3.76		
fixed effects R^2: .0002		model R^2: .31				fixed effects R^2: .31 model R^2: .40					
syllable 1;	mid					syllable 2;	mid				
	β	SE	df	t-value	p		β	SE	df	t-value	p
(intercept)	-9.70	0.49	86.52	-13.89	< 0.001*	(intercept)	-10.55	0.31	52.72	3.95	< 0.001*
slice 2	0.03	0.66	79.64	-4.24	0.92	slice 2	0.50	0.29	366.15	5.51	0.08
slice 3	-0.45	0.66	79.93	9.23	0.16	slice 3	1.58	0.29	366.15	3.52	< 0.001*
slice 4	-0.91	0.66	79.93	9.23	0.005*	slice 4	5.97	0.29	366.15	9.23	< 0.001*
random effects						random effects					
group	name	var.	σ	corr.		group	name	var.	σ	corr.	
word	(int.)	19.22	4.38			word	(int.)	17.81	4.22		
	block	1.79	1.34	-0.98			block	1.07	1.03	-0.97	
residual		5.38	2.32			residual		4.31	2.08		
fixed effects R^2: .02		model R^2: .43				fixed effects R^2: .38 model R^2: .71					
syllable 1;	high					syllable 2;	high				
	β	SE	df	t-value	p		β	SE	df	t-value	p
(intercept)	-0.24	0.48	40.51	-0.50	0.62	(intercept)	-5.07	0.72	46.29	-7.05	< 0.001*
slice 2	-0.01	0.35	367.03	-0.04	0.97	slice 2	0.14	0.60	365.27	0.23	0.82
slice 3	-0.02	0.35	367.03	-0.07	0.94	slice 3	2.80	0.60	365.27	4.69	< 0.001*
slice 4	-0.53	0.35	367.03	-1.54	0.12	slice 4	11.33	0.60	365.27	18.98	< 0.001*
random effects						random effects					
group	name	var.	σ	corr.		group	name	var.	σ	corr.	
word	(int.)	65.14	8.07			word	(int.)	34.13	5.84		
	block	5.22	2.29	-0.97			block	5.86	2.42	-0.86	
residual		6.33	2.52			residual		18.89	4.35		
fixed effects R^2: .002		model R^2: .70				fixed effects R^2: .37 model R^2: .67					

Table 1: Linear mixed effects models of vowel slice on spectral tilt by tone, speaker 1

speaker 2

syllable 1;	low					syllable 2;	low				
	β	SE	df	t-value	p		β	SE	df	t-value	p
(intercept)	4.98	0.78	47.59	6.31	< 0.001*	(intercept)	2.67	1.02	43.84	2.61	0.01*
slice 2	-1.91	0.66	339.92	-2.87	0.004*	slice 2	-0.63	0.84	274.53	-0.75	0.45
slice 3	-4.23	0.67	340.45	-6.34	< 0.001*	slice 3	0.21	0.90	276.90	0.23	0.82
slice 4	-5.22	0.68	341.07	-7.72	< 0.001*	slice 4	-0.71	1.02	279.65	-0.70	0.49
random effects						random effects					
group	name	var.	σ	corr.		group	name	var.	σ	corr.	
word	(int.)	27.77	5.27			word	(int.)	42.39	6.51		
	block	2.01	1.42	-0.78			block	3.51	1.87	-0.75	
residual		22.27	4.72			residual		34.23	5.85		
fixed effects R^2: .10		model R^2: .44				fixed effects R^2: .003		model R^2: .41			

syllable 1;	mid					syllable 2;	mid				
	β	SE	df	t-value	p		β	SE	df	t-value	p
(intercept)	3.54	0.36	138.74	9.79	< 0.001*	(intercept)	5.74	0.49	98.57	11.81	< 0.001*
slice 2	-0.40	0.46	362.76	-0.88	0.38	slice 2	0.38	0.57	345.75	0.66	0.51
slice 3	-0.60	0.46	362.76	-1.31	0.19	slice 3	3.61	0.57	345.75	6.32	< 0.001*
slice 4	-0.77	0.46	362.76	-1.68	0.09	slice 4	10.33	0.58	346.15	17.37	< 0.001*
random effects						random effects					
group	name	var.	σ	corr.		group	name	var.	σ	corr.	
word	(int.)	5.00	2.34			word	(int.)	36.05	6.00		
	block	1.38	1.18	-0.95			block	4.09	2.02	-0.97	
residual		11.00	3.32			residual		16.65	4.08		
fixed effects R^2: .006		model R^2: .21				fixed effects R^2: .40		model R^2: .59			

syllable 1;	high					syllable 2;	high				
	β	SE	df	t-value	p		β	SE	df	t-value	p
(intercept)	1.59	0.54	167.13	2.93	< 0.001*	(intercept)	3.25	0.65	98.53	5.00	< 0.001*
slice 2	-0.15	0.71	379.25	-0.21	0.84	slice 2	1.31	0.74	394.18	1.77	0.82
slice 3	0.20	0.71	379.25	0.29	0.77	slice 3	3.97	0.74	394.18	5.33	< 0.001*
slice 4	-0.80	0.71	379.25	-1.13	0.26	slice 4	9.49	0.75	394.28	12.72	< 0.001*
random effects						random effects					
group	name	var.	σ	corr.		group	name	var.	σ	corr.	
word	(int.)	0.09	0.30			word	(int.)	0.80	0.89		
	block	0.35	0.59	-1.00			block	0.24	0.49	-1.00	
residual		25.83	5.08			residual		29.18	5.40		
fixed effects R^2: .005		model R^2: .07				fixed effects R^2: .28		model R^2: .38			

Table 2: Linear mixed effects models of vowel slice on spectral tilt by tone, speaker 2

syllable 1;	low				
	β	SE	df	t-value	p
(intercept)	2.56	1.75	1.37	1.37	0.33
slice 2	-0.00	0.52	764.89	-0.00	0.99
slice 3	-0.84	0.52	764.93	-1.62	0.11
slice 4	-1.09	0.52	764.81	-2.09	0.04
random effects					
group	name	var.	σ	corr.	
word	(int.)	25.38	5.04		
	block	5.15	2.27	-0.81	
subject	(int.)	5.14	2.27		
residual		27.84	5.28		
fixed effects R^2: .005	model R^2: .48				

syllable 2;	low				
	β	SE	df	t-value	p
(intercept)	2.87	0.71	18.33	4.02	< 0.001*
slice 2	-1.40	0.36	751.26	-3.88	< 0.001*
slice 3	-2.53	0.36	751.30	-7.04	< 0.001*
slice 4	-3.09	0.37	751.29	-8.27	< 0.001*
random effects					
group	name	var.	σ	corr.	
word	(int.)	9.14	3.02		
	block	0.003	0.06	-1.00	
subject	(int.)	0.15	0.39		
residual		12.79	3.58		
fixed effects R^2: .06 model R^2: .47					

syllable 1;	mid				
	β	SE	df	t-value	p
(intercept)	17.50	0.74	5.41	23.79	< 0.001*
slice 2	3.16	0.29	779.24	11.06	< 0.001*
slice 3	3.26	0.29	779.24	11.40	< 0.001*
slice 4	1.72	0.29	779.24	6.03	< 0.001*
random effects					
group	name	var.	σ	corr.	
word	(int.)	9.94	3.15		
	block	0.23	0.48	-0.48	
subject	(int.)	0.43	0.66		
residual		8.56	2.93		
fixed effects R^2: .09	model R^2: .54				

syllable 2;	mid				
	β	SE	df	t-value	p
(intercept)	17.97	0.68	13.31	26.22	< 0.001*
slice 2	1.86	0.33	766.53	5.70	< 0.001*
slice 3	-0.36	0.33	766.53	-1.10	0.27
slice 4	-3.74	0.33	766.11	-11.38	< 0.001*
random effects					
group	name	var.	σ	corr.	
word	(int.)	11.98	3.46		
	block	0.59	0.77	-0.53	
subject	(int.)	0.19	0.44		
residual		11.00	3.31		
fixed effects R^2: .16 model R^2: .55					

syllable 1;	high				
	β	SE	df	t-value	p
(intercept)	15.48	1.10	1.60	14.04	0.01*
slice 2	3.19	0.33	772.90	9.56	< 0.001*
slice 3	4.20	0.33	772.90	12.60	< 0.001*
slice 4	3.21	0.33	772.60	9.58	< 0.001*
random effects					
group	name	var.	σ	corr.	
word	(int.)	5.00	2.24		
	block	0.08	0.28	0.15	
subject	(int.)	1.89	1.38		
residual		11.55	3.40		
fixed effects R^2: .11	model R^2: .48				

syllable 2;	high				
	β	SE	df	t-value	p
(intercept)	13.54	1.27	16.75	10.68	< 0.001*
slice 2	2.19	0.33	777.75	6.55	< 0.001*
slice 3	0.25	0.33	777.75	0.74	0.46
slice 4	-0.87	0.34	778.67	-2.54	0.01*
random effects					
group	name	var.	σ	corr.	
word	(int.)	36.59	6.05		
	block	0.08	0.28	-0.32	
subject	(int.)	0.58	0.76		
residual		11.85	3.44		
fixed effects R^2: .03 model R^2: .75					

Table 3: Linear mixed effects models of vowel slice on HNR by tone

B Pairwise comparisons for tone sequence

speaker 1 syllable 1						syllable 2					
	β	SE	df	t-value	p		β	SE	df	t-value	p
LL-HH	9.96	0.68	115.71	14.60	< 0.001*	LL-HH	9.80	0.73	69.72	13.39	< 0.001*
LL-HL	9.72	0.69	118.18	14.06	< 0.001*	LL-HL	1.69	0.76	74.98	2.23	0.03
LL-HM	6.22	0.68	115.71	9.12	< 0.001*	LL-HM	-2.17	0.73	69.72	-2.96	0.004*
LL-LH	-0.35	0.68	119.81	-0.51	0.61	LL-LH	3.09	0.73	69.96	4.22	< 0.001*
LL-LM	0.08	0.68	115.71	0.12	0.91	LL-LM	-2.13	0.73	69.72	-2.91	0.005*
LL-MH	-0.69	0.68	115.71	-1.02	0.31	LL-MH	9.62	0.73	69.72	13.14	< 0.001*
LL-ML	0.17	0.69	116.53	0.25	0.80	LL-ML	-0.54	0.85	96.28	-0.64	0.53
LL-MM	-1.00	0.69	117.43	-1.44	0.15	LL-MM	-2.35	0.74	72.41	-3.17	0.002*
HH-HL	-0.23	0.68	114.00	0.34	0.73	HH-HL	-8.12	0.71	65.50	11.40	< 0.001*
HH-HM	-3.74	0.67	111.00	5.60	< 0.001*	HH-HM	-11.98	0.69	59.90	17.42	< 0.001*
HH-LH	-10.31	0.67	116.00	15.42	< 0.001*	HH-LH	-6.71	0.69	60.00	9.75	< 0.001*
HH-LM	-9.88	0.67	111.00	14.80	< 0.001*	HH-LM	-11.94	0.69	59.90	17.37	< 0.001*
HH-MH	-10.6	0.67	111.00	15.96	< 0.001*	HH-MH	-0.18	0.69	59.90	0.26	0.80
HH-ML	-9.78	0.67	112.00	14.55	< 0.001*	HH-ML	-10.34	0.81	88.70	12.78	< 0.001*
HH-MM	-10.96	0.68	113.00	16.17	< 0.001*	HH-MM	-12.16	0.70	62.70	17.45	< 0.001*
HL-HM	-3.50	0.68	114.00	5.17	< 0.001*	HL-HM	-3.86	0.71	65.50	5.14	< 0.001*
HL-LH	-10.07	0.68	118.00	14.84	< 0.001*	HL-LH	1.41	0.71	65.70	-1.97	0.05
HL-LM	-9.64	0.68	114.00	14.23	< 0.001*	HL-LM	-3.82	0.71	65.50	5.36	< 0.001*
HL-MH	-10.42	0.68	114.00	15.37	< 0.001*	HL-MH	7.94	0.71	65.50	-11.15	< 0.001*
HL-ML	-9.55	0.68	115.00	13.99	< 0.001*	HL-ML	-2.23	0.83	93.30	2.68	0.009*
HL-MM	-10.72	0.69	116.00	15.59	< 0.001*	HL-MM	-4.04	0.72	68.30	5.60	< 0.001*
HM-LH	-6.57	0.67	116.00	9.83	< 0.001*	HM-LH	5.26	0.69	60.00	-7.65	< 0.001*
HM-LM	-6.14	0.67	111.00	9.20	< 0.001*	HM-LM	0.04	0.69	59.90	-0.53	0.96
HM-MH	-6.91	0.67	111.00	10.36	< 0.001*	HM-MH	11.79	0.69	59.90	-17.16	< 0.001*
HM-ML	-6.05	0.67	112.00	8.99	< 0.001*	HM-ML	1.63	0.81	88.70	-2.01	0.05
HM-MM	-7.22	0.68	113.00	10.65	< 0.001*	HM-MM	-0.18	0.70	62.70	0.26	0.79
LH-LM	0.43	0.67	116.00	-0.64	0.53	LH-LM	-5.23	0.69	60.00	7.60	< 0.001*
LH-MH	-0.34	0.67	116.00	0.51	0.61	LH-MH	6.53	0.69	60.00	-9.50	< 0.001*
LH-ML	0.52	0.67	116.00	-0.78	0.44	LH-ML	-3.63	0.81	89.20	4.84	< 0.001*
LH-MM	-0.65	0.68	117.00	0.95	0.34	LH-MM	-5.45	0.70	62.90	7.81	< 0.001*
LM-MH	-0.78	0.67	111.00	1.16	0.25	LM-MH	11.76	0.69	59.90	-17.11	< 0.001*
LM-ML	0.09	0.67	112.00	-0.14	0.89	LM-ML	1.59	0.81	88.70	-1.97	0.05
LM-MM	-1.08	0.68	113.00	1.59	0.11	LM-MM	-0.22	0.70	62.70	0.31	0.75
MH-ML	0.87	0.67	112.00	-1.29	0.20	MH-ML	-10.16	0.81	88.70	12.56	0.53
MH-MM	-0.31	0.68	113.00	0.45	0.65	MH-MM	-11.98	0.70	62.70	17.19	0.002*
ML-MM	-1.17	0.68	114.00	1.72	0.09	ML-MM	-1.81	0.82	2.22	2.22	0.002*

Table 4: Spectral tilt tone sequence pairwise comparisons, speaker 1

speaker 2 syllable 1						syllable 2					
	β	SE	df	t-value	p		β	SE	df	t-value	p
LL-HH	-3.58	1.06	292.00	3.93	< 0.001*	LL-HH	10.15	1.53	81.80	-6.63	< 0.001*
LL-HL	-2.52	1.07	292.00	2.10	0.04	LL-HL	9.86	1.62	83.10	-6.10	< 0.001*
LL-HM	-2.45	1.06	292.00	2.32	0.02	LL-HM	12.68	1.54	84.70	-8.24	< 0.001*
LL-LH	-4.00	1.11	292.00	3.62	< 0.001*	LL-LH	11.53	1.55	83.30	-7.44	< 0.001*
LL-LM	-0.88	1.09	292.00	0.81	0.42	LL-LM	11.85	1.55	84.90	-7.66	< 0.001*
LL-MH	0.22	1.05	292.00	-0.21	0.83	LL-MH	9.46	1.53	80.90	-6.20	< 0.001*
LL-ML	-0.54	1.06	292.00	0.51	0.61	LL-ML	8.45	1.80	102.40	-4.69	< 0.001*
LL-MM	-1.35	1.06	292.00	1.27	0.21	LL-MM	13.14	1.54	84.00	-8.59	< 0.001*
HH-HL	1.33	1.06	292.00	-1.26	0.21	HH-HL	-0.29	1.32	52.90	0.22	0.82
HH-HM	1.32	1.04	292.00	-1.09	0.28	HH-HM	2.53	1.22	49.80	-2.07	0.04
HH-LH	0.43	1.09	292.00	0.39	0.70	HH-LH	1.38	1.24	51.10	-1.12	0.27
HH-LM	2.70	1.07	292.00	-2.52	0.01	HH-LM	1.70	1.23	51.20	-1.38	0.17
HH-MH	3.80	1.03	292.00	-3.68	< 0.001*	HH-MH	-0.69	1.21	47.30	0.57	0.57
HH-ML	3.04	1.04	292.00	-2.93	0.003*	HH-ML	-1.70	1.54	78.90	1.10	0.27
HH-MM	2.23	1.05	292.00	-2.13	0.03	HH-MM	3.09	1.22	50.00	-2.52	0.015
HL-HM	-0.20	1.06	292.00	0.19	0.85	HL-HM	2.83	1.33	54.60	-2.13	0.04
HL-LH	-1.75	1.11	292.00	1.58	0.11	HL-LH	1.67	1.34	55.60	-1.25	0.22
HL-LM	1.37	1.09	292.00	-1.26	0.21	HL-LM	1.99	1.34	55.70	-1.49	0.14
HL-MH	2.47	1.05	292.00	-2.36	0.02	HL-MH	-0.39	1.31	52.10	0.30	0.77
HL-ML	1.72	1.06	292.00	-1.63	0.11	HL-ML	-1.41	1.62	80.30	0.87	0.39
HL-MM	0.90	1.06	292.00	-0.85	0.40	HL-MM	3.38	1.33	54.70	-2.54	0.014
HM-LH	-1.56	1.09	292.00	1.43	0.16	HM-LH	-1.15	1.25	53.30	0.93	0.36
HM-LM	1.57	1.07	292.00	-1.46	0.14	HM-LM	-0.83	1.24	52.90	0.67	0.51
HM-MH	2.67	1.03	292.00	-2.58	0.01*	HM-MH	-3.22	1.22	49.00	2.65	0.011
HM-ML	-1.91	1.04	292.00	-1.84	0.07	HM-ML	-4.23	1.55	81.30	2.73	0.0008*
HM-MM	-1.10	1.05	292.00	-1.05	0.29	HM-MM	0.56	1.23	51.60	-0.45	0.65
LH-LM	3.12	1.12	292.00	-2.78	0.006*	LH-LM	0.32	1.25	54.30	-0.26	0.80
LH-MH	4.22	1.08	292.00	-3.89	< 0.001*	LH-MH	-2.06	1.23	50.20	1.68	0.10
LH-ML	3.47	1.09	292.00	-3.18	0.002	LH-ML	-3.08	1.56	80.70	1.98	0.05
LH-MM	2.66	1.10	292.00	-2.42	0.02	LH-MM	1.71	1.25	53.20	-1.37	0.18
LM-MH	1.10	1.07	292.00	-1.03	0.30	LM-MH	-2.39	1.23	50.30	1.95	0.06
LM-ML	0.35	1.07	292.00	-0.32	0.75	LM-ML	-3.40	1.56	81.80	2.19	0.03
LM-MM	-0.47	1.08	292.00	0.43	0.67	LM-MM	1.39	1.24	53.10	-1.12	0.27
MH-ML	-0.76	1.03	292.00	0.73	0.46	MH-ML	-1.02	1.54	78.00	0.66	0.51
MH-MM	-1.57	1.04	292.00	1.51	0.13	MH-MM	3.78	1.22	49.20	-3.10	0.003*
ML-MM	-0.81	1.05	292.00	0.78	0.44	ML-MM	4.79	1.55	80.80	-3.09	0.003*

Table 5: Spectral tilt tone sequence pairwise comparisons, speaker 2

syllable 1						syllable 2					
	β	SE	df	t-value	p		β	SE	df	t-value	p
LL-HH	16.66	1.37	71.80	-12.17	< 0.001*	LL-HH	18.03	1.27	71.50	-14.24	< 0.001*
LL-HL	15.88	1.38	73.20	-11.55	< 0.001*	LL-HL	4.80	1.27	73.10	-3.78	< 0.001*
LL-HM	16.29	1.37	71.80	-11.91	< 0.001*	LL-HM	17.61	1.26	70.90	-13.94	< 0.001*
LL-LH	-0.64	1.37	72.60	0.46	0.64	LL-LH	6.82	1.26	70.60	-5.41	< 0.001*
LL-LM	2.48	1.37	72.50	-1.81	0.08	LL-LM	16.90	1.26	70.30	-13.41	< 0.001*
LL-MH	17.56	1.37	71.20	-12.86	< 0.001*	LL-MH	18.13	1.27	71.40	-14.33	< 0.001*
LL-ML	16.93	1.37	71.90	-12.37	< 0.001*	LL-ML	-0.14	1.27	72.60	0.11	0.91
LL-MM	17.78	1.37	72.50	-12.96	< 0.001*	LL-MM	18.12	1.26	71.00	-14.34	< 0.001*
HH-HL	-0.78	1.37	72.70	0.57	0.57	HH-HL	-13.23	1.27	73.40	10.38	< 0.001*
HH-HM	-0.37	1.37	71.30	0.27	0.79	HH-HM	-0.42	1.26	71.10	0.33	0.74
HH-LH	-17.29	1.37	72.10	12.63	< 0.001*	HH-LH	-11.21	1.26	70.90	8.87	< 0.001*
HH-LM	-14.18	1.37	72.00	10.35	< 0.001*	HH-LM	-1.13	1.26	70.60	0.89	0.37
HH-MH	0.90	1.36	70.80	-0.66	0.52	HH-MH	0.10	1.27	71.70	-0.08	0.93
HH-ML	0.27	1.37	71.40	-0.20	0.84	HH-ML	-18.16	1.27	72.80	14.28	< 0.001*
HH-MM	1.12	1.37	72.10	-0.82	0.42	HH-MM	0.09	1.26	71.30	-0.07	0.94
HL-HM	0.41	1.37	72.60	-0.30	0.77	HL-HM	12.81	1.27	72.70	-10.07	< 0.001*
HL-LH	-16.52	1.38	73.40	12.00	< 0.001*	HL-LH	2.02	1.27	72.50	-1.59	0.12
HL-LM	-13.41	1.38	73.30	9.74	< 0.001*	HL-LM	12.10	1.27	72.30	-9.53	< 0.001*
HL-MH	1.67	1.37	72.10	-1.22	0.23	HL-MH	13.33	1.27	73.30	-10.46	< 0.001*
HL-ML	1.05	1.37	72.70	-0.76	0.45	HL-ML	-4.94	1.28	74.40	3.86	< 0.001*
HL-MM	1.90	1.38	73.40	-1.38	0.17	HL-MM	13.23	1.27	72.90	-10.47	< 0.001*
HM-LH	-16.93	1.37	72.10	12.36	< 0.001*	HM-LH	-10.79	1.26	70.20	8.56	< 0.001*
HM-LM	-13.81	1.37	72.00	10.09	< 0.001*	HM-LM	-0.71	1.26	70.00	0.56	0.56
HM-MH	1.27	1.36	70.70	-0.93	0.36	HM-MH	0.52	1.26	71.00	-0.42	0.68
HM-ML	0.64	1.37	71.40	-0.47	0.64	HM-ML	-17.74	1.27	72.20	13.98	< 0.001*
HM-MM	1.49	1.37	72.00	-1.09	0.22	HM-MM	0.52	1.26	70.70	-0.41	0.68
LH-LM	3.11	1.37	72.80	-2.27	0.03	LH-LM	10.08	1.26	69.70	-8.02	< 0.001*
LH-MH	18.19	1.37	71.50	-13.31	< 0.001*	LH-MH	11.31	1.26	70.80	-8.96	< 0.001*
LH-ML	17.56	1.37	72.20	-12.82	< 0.001*	LH-ML	-6.96	1.27	72.00	5.49	< 0.001*
LH-MM	18.41	1.37	72.80	-13.41	< 0.001*	LH-MM	11.30	1.26	70.40	-8.97	< 0.001*
LM-MH	15.08	1.37	71.50	-11.03	< 0.001*	LM-MH	1.23	1.26	70.50	-0.98	0.33
LM-ML	14.45	1.37	72.10	-10.55	< 0.001*	LM-ML	-17.04	1.27	71.70	13.45	< 0.001*
LM-MM	15.30	1.37	72.70	-11.42	< 0.001*	LM-MM	1.22	1.26	70.10	-0.97	0.34
MH-ML	-0.63	1.36	70.80	0.46	0.65	MH-ML	-18.27	1.27	72.80	14.37	< 0.001*
MH-MM	0.22	1.37	71.50	-0.16	0.87	MH-MM	-0.01	1.26	71.20	0.01	0.99
ML-MM	0.85	1.37	72.10	-0.62	0.54	ML-MM	18.26	1.27	72.40	-14.38	< 0.001*

Table 6: HNR tone sequence pairwise comparisons