Creaky voice in Yoruba

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

1 Introduction

This paper investigates the presence 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 creakiness is present there as well.

Looking at disyllabic words, not just monosyllables as in Hayward et al. (2004), enhances our understanding of creaky voice in Yoruba by addressing the following research questions. First, are low tones more creaky in the second syllable versus the first? Second, is there any effect of the sequence of tones on the level of low-tone creakiness? For example, does the first L in an LL sequence exhibit stronger creak than the L in an LM sequence? There is also the question of whether differences in creaky voice between tone levels is a gradient or categorical one. Is there a gradual shift towards creakiness when moving from a high to a mid and then to a low tone, or is the low tone different from the mid and high tones categorically?

That creaky voice may mark the low tone 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).

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 in terms of contrast 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). The line of research laid out in this paper would add Yoruba to this constellation of languages where there is an interaction between tone and phonation type.

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 is said to have 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.

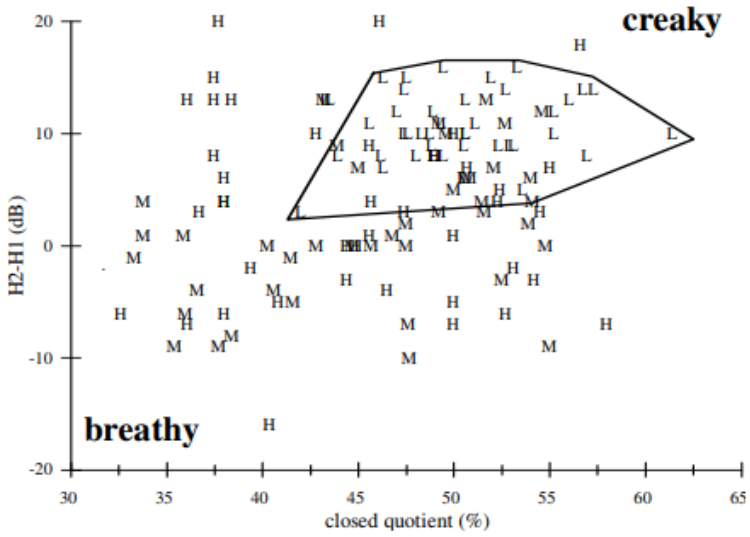

Figure 1: L tones cluster toward creaky end of spectrum (Hayward et al. 2004)

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 kã̀ 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 subject, repetition number, and word as random intercepts. 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):

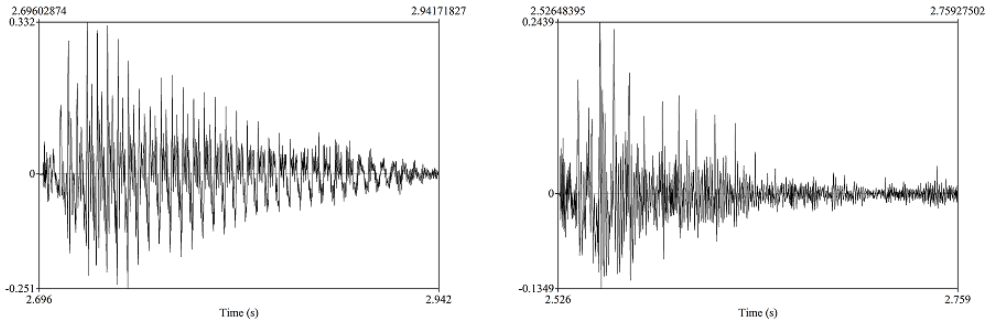

Figure 2: waveforms for *tá* (left) and *tà* (right)

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 – measured as the difference in amplitude between the first and second harmonics (H1-H2), and Harmonic to Noise Ratio (HNR), a measure of F0 irregularity 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. Measurements were taken at four evenly-spaced points over the course of the vowel using a script. Ten 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*

Figure 3: 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. 3, spectral tilt and HNR appear to show the clearest bifurcation of the tonal space in Yoruba. The plot in Fig. 4 makes this result clearer.

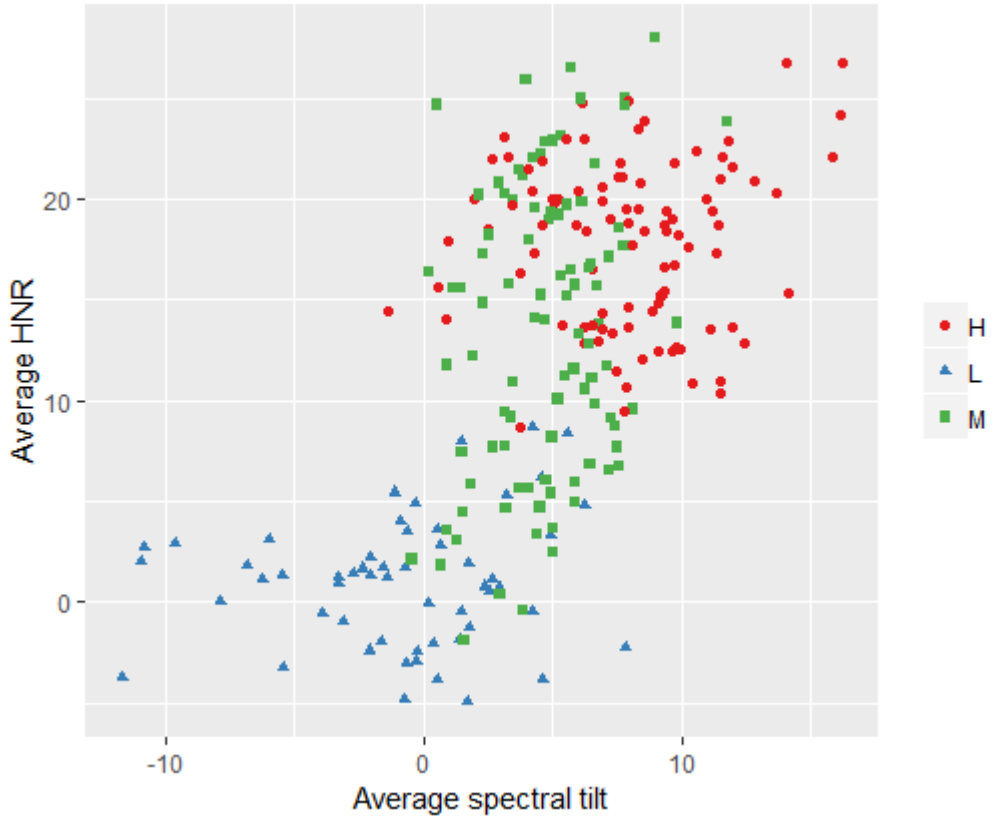

Figure 4: Average HNR and Average spectral tilt by tone

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 tone 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. 3 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 ($\beta = 0.10, p < 0.001$) and high tones ($\beta = 0.090, p < 0.001$), with model $R^2 = 0.68$. As mentioned above, 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. 5.

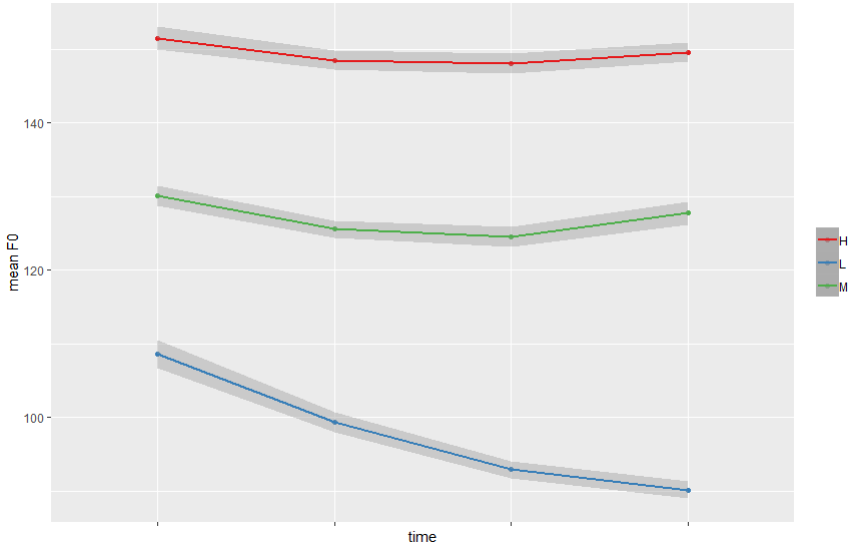

Figure 5: mean F0 over course of vowel

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: $\beta = -9.10$; first vs. third $\beta = -15.49$; first vs. fouth $\beta = -18.23$; $p < .001$ for all, $R^2 = 0.64$). 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. A linear mixed effects model with average HNR as the dependent variable and tone category and vowel as the predictors indicates a significant difference between both low and mid tones ($\beta = 11.75, p < 0.001$) and low and high tones ($\beta = 15.76, p < 0.001$) with model $R^2 = 0.87$. This result indicates that low HNR marks low tones in Yoruba, and suggests that HNR is one acoustic implementation of creak in the language.

Post-hoc comparisons using the Tukey test indicate a significant difference between the high and mid tones as well ($p < 0.001, 95\%$ CL). Interpreting this as a gradual decline in creakiness from high to mid to low tone seems to support the hypothesis that differences in creak between tone levels are gradient, rather than categorical. It is noteworthy, however, that the difference between the mean HNR of high and mid tones is ∼4dB, while the difference between low and mid tones is ∼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 leaves open the possibility that HNR does mark low tones as creaky in a way that is categorically different from mid or high tones. More data may shed further light on this issue.

HNR does not appear to vary widely throughout the course of low tone vowels. The lowest average HNR is in the second slice of the vowel, at 0.34dB, while the highest is in the fourth slice, at 1.22dB. A linear mixed effects model with HNR as the dependent variable and vowel portion and vowel as the independent variable does not indicate any significant differences. For mid and high tones, however, variation in HNR over the course of the vowel is wider, with a range of ∼4dB for mid tones and ∼6dB for high tones. For mid tone, a non-significant trend is found for HNR between the first slice of the vowel and the third slice ($\beta = -1.83, p < 0.09$) and the first slice and fourth slice ($\beta = -3.48, p = 0.001$.) For high tone, a significant difference is found between the first slice and the second slice ($\beta = 2.83, p < 0.001$) and the first and fourth slice ($\beta = -2.69, p < 0.001$). One possible interpretation of this result is that a change in HNR over time sets high and mid tone vowels apart from low tone vowels, supporting the idea of a categorical partition of the register space in Yoruba with regards to correlates of creaky voice.

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 Figures 3 and 4, low tones do appear to have generally lower spectral tilt values than mid or high tones. A linear mixed effects model with mean spectral tilt as the dependent variable and tone category, vowel, and onset as the independent variables reveals a statistically significant effect. Spectral tilt increases both when moving from a low tone to a mid tone ($\beta = 5.22, p < 0.001$) and when moving from a low tone to a high tone ($\beta = 8.30, p < 0.001$) with model $R^2 = 0.55$. This result indicates that spectral tilt value marks low tones in Yoruba, and suggests that spectral tilt is one acoustic implementation of creak in the language.

Post-hoc comparisons using the Tukey test indicate a significant difference between the high and mid tones as well ($p < 0.001, 95\%$ CL). With HNR, interpretation of the same result was tempered by the observation that 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 ∼5dB, while

the difference between the mid and high is ∼3dB. 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. Again, more data will hopefully offer a clearer generalization.

There is some variation in spectral tilt throughout the course of the vowel. A linear mixed effects model with low tone vowel portion, onset, and vowel as the independent variables and spectral tilt as the dependent variable shows a non-significant trend between the first and second slice ($\beta = -1.97, p < 0.04$) and a significant difference between the first and third slice ($\beta = -5.79, p < 0.001$) and first and fourth slice ($\beta = -6.20, p < 0.001$) with model $R^2 = 0.29$. For the mid tone, a significant difference is found between the first and second ($\beta = -1.74, p < 0.001$) and first and fourth portions ($\beta = 9.25, p < 0.001$) with model $R^2 = 0.64$. For the high tone, the same is found for the first and third ($\beta = 4.25, p < 0.001$) and first and fourth portions ($\beta = 13.22, p < 0.001$) with model $R^2 = 0.63$. Note that while there is a general negative trend seen in the slopes of the low tone vowel portions, in the mid and high tone vowels there is a more pronounced positive slope increase as the vowel progresses. Additionally, spectral tilt values are quite similar across all tone levels in the first slice of the vowel. It is not until the latter half of the vowel that low tones diverge from mid and high tones. This suggests that creaky voice in Yoruba, as implemented by lower spectral tilt, may not "kick in" until the latter portion of the vowel.

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). A possible interpretation of the results is that low HNR marks low tones as different from mid or high tones categorically, while changes in spectral tilt from tone level to tone level are more gradient. Addition of data from more speakers will hopefully support the conclusions drawn here. 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 how robust creakiness is in Yoruba low tones. If both the first and second syllables have a similar creak profile, then the conclusion is that creak is not merely a reflex of word or utterance position. Through this experiment questions of creaky voice as it relates to the tone sequence are also answered. For example, it will be shown that the L in an HL sequence is generally less creaky than the second L in an LL sequence.

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 with a similar biographic profile. 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 was done in Praat (Boersma 2001) using the same methods described in §3 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*

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

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. There are some obvious trends that can be commented on, but for the purpose of statistical modeling, I investigate each of the four vowel slices individually and in relation to each other in order to clarify the generalizations regarding creaky voice in Yoruba.

4.3 Spectral tilt

I start with some observations on inter-speaker variation. For spectral tilt, 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 that even the low tone, and the high tone is marked with higher spectral tilt values. This is fairly consistent across syllables. The second speaker shows some variation in spectral tilt from syllable to syllable. There is no clear difference between tone levels in syllable one, but an increase in spectral tilt in syllable two in high and mid tones with a slight increase seen in the low tones as well. One possible interpretation of the data is that speaker one uses spectral tilt to mark the lower

part of their register in general, while speaker two manipulates spectral tilt to mark word or phrase boundaries. 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 is possible that spectral tilt does not mark low tones *in particular* as creaky as was concluded in §3.

Different models are built for each speaker. I start by examining spectral tilt level across each syllable for each tone. Linear mixed effects models with spectral tilt value as dependent variable and vowel slice as the independent variable with *block* and *word* in the random effects structure indicates no significant difference between slices for any tone level in the first syllable for speaker one. In syllable two, however, there is a significant difference between the first slice and fourth slice ($\beta = 6.27, p < 0.001$) with $R^2 = .40$. A post-hoc Tukey test indicates a significant difference between the second and fourth slice and third and fourth slice as well ($p < 0.001$, CL = 95%). The progression through the low tone in the second syllable shows a positive upward trend. Interestingly, this trend holds in second syllable mid and high tones as well. For mid tones, there is a significant difference between the first and third slice ($\beta = 1.65, p < 0.001$) and first and fourth slice ($\beta = 6.04, p < 0.001$), with a Tukey test indicating significant differences between the second slice and later slices and between the third slice and fourth slice. For high tones, a similar pattern is found (1st-3rd $\beta = 2.80, p < 0.001$; 1st-4th $\beta = 11.33, p < 0.001$), with further significant differences indicated by the Tukey test for all pairwise comparisons except the first and second slice. This result indicates that while spectral tilt is relatively stable in syllable one, there is a significant increase in general in syllable two, regardless of tone level. This supports the idea that spectral tilt level marks a word or phrase boundary rather than any tone in particular, and that any difference observed between tone levels may be a byproduct of the speaker approaching the bottom of their register – not an intentional application of creaky voice.

The behavior of speaker two lends further support to this hypothesis. In the first syllable, there is a significant negative slope between the first and third ($\beta = -4.28, p < 0.001$) and first and fourth slice ($\beta = -5.17, p < 0.001$) for the low tone, but no significant difference between any vowel slice for the mid and high tones. Much like speaker one, there is relative stability across the syllable in terms of spectral tilt – at least for mid and

high tones. In syllable two, while there is no effect of vowel slice on spectral tilt for low tones, a similar effect observed in speaker one is found for both mid and high tones. For the mid tone, there is a significant difference between the first and third ($\beta = 3.61, p < 0.001$) and first and fourth slices ($\beta = 9.98, p < 0.001$) with a Tukey test indiciating significant differences between all slices except the first and second. The same pattern obtains in the high tones (1st-3rd $\beta = 3.94, p < 0.001$, 1st-4th $\beta = 9.46, p < 0.001$). 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, despite having disjoint spectral tilt ranges. The overall conclusion is that spectral tilt in Yoruba marks word or phrase boundaries, rather than marking low tones in particular as creaky.

Linear mixed effects models with *block* in the random effects structure investigating the effect of tone level on spectral tilt within a slice of the vowel further enforce this idea. For speaker one in syllable one, there is a significant difference between the low and high tone in all slices ($7.3 \leq \beta \leq 7.9, p < 0.001$ for all) but no significant difference between low and mid tones until the fourth slice ($\beta = -1.25, p = 0.004$). In the second syllable, however, there is a significant difference in spectral tilt between each tone level in all four slices of the vowel with a positive slope between low and high tones and a smaller, negative slope between low and mid tones and $p < 0.001$ for all comparisons. For speaker two in syllable one, there is a significant difference between low and high tones in slice one ($\beta = -3.61, p < 0.001$) and slice two ($\beta = -1.89, p = 0.005$), and a significant difference between low and mid tones in slice three ($\beta = 2.52, p < 0.001$) and slice four ($\beta = 3.23, p < 0.001$). In syllable two, there is a significant difference between the low and the mid in the first slice ($\beta = 2.22, p = 0.007$) and a significant difference between the low and mid and low and high in all proceeding slices with similarly positive slopes for both and $p < 0.001$ for all comparisons. Here, post-hoc Tukey tests indicate no significant difference between mid and high tones with regard to spectral tilt in any slice of second-syllable vowels. For speaker one, it appears that mid tones have lower spectral tilt values, while speaker two shows lower spectral tilt in high tones in the first syllable and higher values in both mid and high tones in the second syllable. 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, as the first experiment gave reason to believe. 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.

4.4 HNR

For HNR, the generalization appears 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.

Linear mixed effects models with HNR value as the dependent variable, vowel slice as the independent variable and *block* and *subject* in the random effects structure were built for each tone level to assess how HNR might change over the course of each syllable. In the first syllable, there is no significant difference in HNR between slices among low tones. For mid tones, there is a significant difference between the first slice and each other slice (1st-2nd $\beta = 2.15$; 1st-3rd $\beta = 3.25$, 1st-4th $\beta = 1.72, p < 0.001$ for all) and a further difference between the second and fourth slice and third and fourth slice as indicated by the Tukey test with CL = 95% and $p \leq 0.004$ for both. For high tones, there is also a significant difference between the first slice and each other slice (1st-2nd $\beta = 3.18$; 1st-3rd $\beta = 4.20$, 1st-4th $\beta = 3.18, p < 0.001$ for all) with no further significant differences between slices. These results show that while HNR is steady and low in the low tones, there is a small positive slope over the duration of the first vowel for mid and high tones.

In the second syllable, there is a significant difference among low tones between the first slice and each other slice (1st-2nd $\beta = -1.47$; 1st-3rd $\beta = -2.60$, 1st-4th $\beta = -3.01, p < 0.001$ for all) and a significant difference between the second and fourth slice (CL = 95%, $p = 0.008$). For mid tones, the pattern is similar, though it is not as immediately apparent – there is a significant positive slope between the first and second slice ($\beta = 1.84, p < 0.001$) but a significant negative slope between the first and fourth slice ($\beta = -3.54, p < 0.001$), with additional significant differences between all following slices with CL = 95% and $p < 0.001$. For high tones, there is also a significant positive slope between the first and second slices ($\beta = 2.16, p = 0.001$) and a significant negative slope between the second and fourth slice with CL = 95% and $p < 0.001$. This shows that HNR tends to drop off as the second syllable progresses – though the differences – measured in dB, are small.

Differences in HNR between tone levels were also assessed at each slice of the vowel with models taking in-slice HNR as the dependent variable and tone level as the dependent variable with *subject* as a random intercept. In syllable one there is a significant difference between low tones and mid tones in all slices ($12 \leq \beta \leq 18, p < 0.001$ for all) and low and high tones in all slices ($14 \leq \beta \leq 18, p < 0.001$ for all). In order to address the question of creaky voice appearing categorically on low tones, or gradiently as speakers descend in their register, it is important to also compare the mid and high tones. The post-hoc Tukey test indicates a small increase in HNR from high tones to mid tones with estimated slope around 2 and $p \leq 0.002$ for the first and second slices. Not only is the difference small – it is backwards from what we would expect if lowering in one's register led to lower HNR and thus gradiently more creaky phonation. Additionally, in the third and fourth slice, the significant difference between mid and high tones disappears.

In the second syllable, the pattern is similar. There is a significant difference between low tones and mid tones in all slices ($14 \leq \beta \leq 18, p < 0.001$ for all) and low and high tones in all slices ($10 \leq \beta \leq 14, p < 0.001$ for all). Here, the post-hoc Tukey test indicates a significant difference between high and mid tones in each slice of the second vowel, but it is still a much smaller difference than the difference between the low and other tones and is again backwards, with a positive slope of around 4dB and $p < 0.001$. 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.

4.5 F0

For F0, there is some inter-speaker variation. 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 seen for speaker one as opposed to speaker two. 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 for the second syllable in HL and LH sequences:

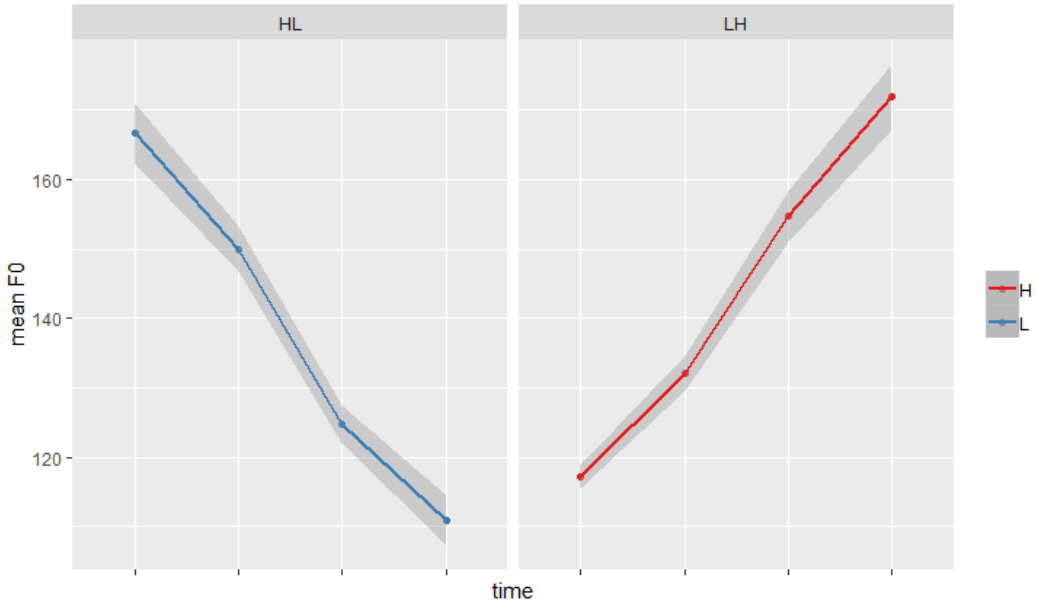

Figure 7: F0 over time of HL low tone and LH high tone

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 spectral tilt or HNR 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. The question, then is whether tone sequence has any effect of the level of creak seen in the tones. Is there any difference in creak between the initial low tone in an LL, LM, or LH sequence? Is there any difference in creak in the second tone of an LL, ML, or HL sequence? What other notable differences emerge?

To address these questions, linear mixed effects models for each syllable were built, using average HNR in the syllable as the dependent variable, tone sequence as the independent variable, and *subject* and *block* as random intercepts. For low tones in the first syllable, there is no significant difference between an LL and an LH ($\beta = -0.33, p = 0.63$). There is, however, a significant difference between an LL and an LM, such that initial low tones have slightly higher HNR in an LM versus an LL. ($\beta = 3.63, p < 0.001$) with $R^2 = 0.80$. This result is consistent with Akinlabi and Liberman (1995), which showed that F0 transitions in sequences involving a mid tone occur earlier than in HL or LH sequences. A post-hoc Tukey test finds no other significant creaky voice differences in the initial syllable within a tone level based on the following tone.

In the second syllable, the effects of the contour on creaky voice are seen – the low tone in an HL sequence has significantly higher HNR than its counterpart in an LL ($\beta = 4.56, p < 0.001$) or ML ($\beta = 4.61, p < 0.001$) with $R^2 = 0.91$. Similarly, the high tone in an LH sequence is significantly lower in HNR than in an HH ($\beta = -11.04, p < 0.001$) or MH ($\beta = -11.03, p < 0.001$). No other significant differences are observed within a tone level in the second syllable based on preceding tone. Based on these results, it can be concluded that the tone sequence present in a disyllabic word can affect the level of creaky voice found in each syllable.

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, contrary to the results of experiment one. It was also shown that, despite having quite different spectral tilt ranges, the two speakers exhibit similar patterns of behavior. Further conclusions regarding creaky voice and tone sequence were drawn, showing that the level of creakiness in a syllable may depend on the tone that precedes or proceeds it.

5 Conclusion

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 (H1-H2) and Harmonic-to-Noise Ratio (HNR). The results indicated that both spectral tilt and HNR are reliable cues for creaky voice, though it was not clear that low tones were marked categorically in this way, or if the difference was merely the gradient result of the speaker stepping down in their register from tone level to tone level.

The second experiment looked to expand previous results to words of different shapes by examining creaky voice charateristics in CVCV words. With the first experiment as a reference point, the hypothesis was that both HNR and spectral tilt would indicate creaky voice in the low tone, but this was not confirmed – it was found that spectral tilt may act as some sort of boundary marker, and that it does not mark low tones in particular. HNR, however, was much more robust, 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. Lastly, questions related to the sequence of tones were addressed, revealing that the level of creak in a high or low tone may vary depending

on the following or preceding tone.

To improve the results of this study, more speakers should be added. 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.

The most fruitful 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.

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.

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.