The role of creaky voice in perception of spoken Yoruba

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

1 Introduction

This paper investigates the relationship between creaky voice and perception of 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 creak accompanying the low tone in spoken Yoruba (Welmers 1974; Hayward et al. 2004; Yu 2010). Data from this experiment is used to synthesize tokens for a perception experiment. The second experiment explores the importance of creaky voice in perception of Yoruba via cue weighting and confusion matrices. The main research question is as follows: if the creaky quality of a low-tone word is altered synthetically and played back to native-speaker listeners, do they still perceive the word as bearing a low tone? While previous studies have established the importance of tonal contour in perception of the low tone in Yoruba (Harrison 1996), the role of non-modal phonation has yet to be considered.

That creaky voice may affect perception of tone in Yoruba is interesting in that it would indicate that Yoruba speakers are using information in the acoustic signal other than F0 to make judgments about tonal contrast, even though Yoruba is thought of as a pure tone language where pitch is the only cue for tonal contrast. Yu and Lam (2014) find exactly this result for "tone 4" of Cantonese, where a battery of experiments demonstrated a clear perceptual bias towards tone 4 with increased creaky phonation. A similarly positive result for Yoruba would place it 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).

This line of research also helps delineate the importance of different

possible perceptual cues for low tone in Yoruba. Harrison (1996) shows the importance of the falling contour for the low tone. Is creaky voice also an important cue for listeners? The research to be carried out here will help disambiguate the role of contour versus creak in perception.

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). The opposite case is better known, with association of 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 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.

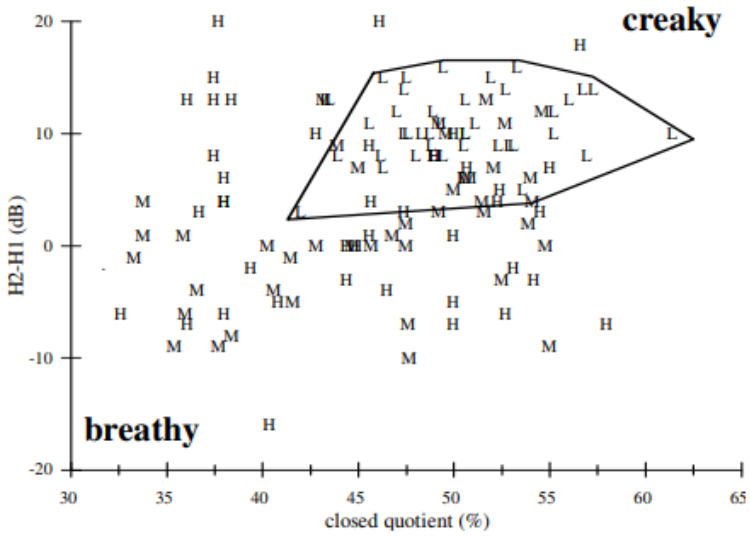

Figure 1: L tones cluster toward creaky end of spectrum (Hayward et al. 2004)

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 first experiment of the current study follows the methodology employed in 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) So _____ le kan sí i
/sɔ _____ le 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 vowel as the random effects. 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):

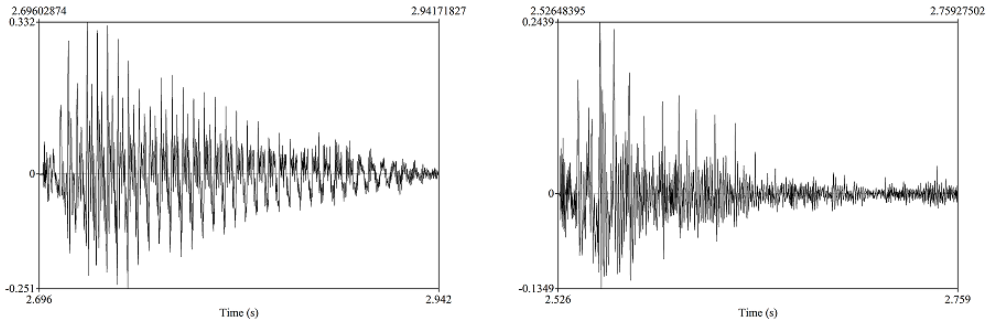

Figure 2: waveforms for *tá* (left) and *tà* (right)

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

Five 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), F1-F0 – measured as the

difference in amplitude between the first harmonic and the most energetic harmonic in the first formant peak (Kirk et al. 1984), and Harmonic to Noise Ratio (HNR), a measure of F0 irregularity for which lower values indicate creakier phonation (Keating et al. 2015). For F1-F0, higher values are indicative of creaky phonation. Lower spectral tilt values suggest increased glottal constriction, and so 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 anMeans and standard deviations for these measurements are shown in the table below.

	F0	HNR	spectral tilt	F1-F0	duration
H	149.39	17.84	7.91	371.69	0.26
	6.44	*4.15*	*3.43*	*234.48*	*0.04*
M	126.97	13.45	4.79	401.96	0.27
	6.37	*7.22*	*2.30*	*192.02*	*0.05*
L	98.48	0.71	-0.19	442.99	0.17
	5.08	*3.06*	*4.73*	*211.75*	*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. F1-F0 seems to track higher, though this is less clear. 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.

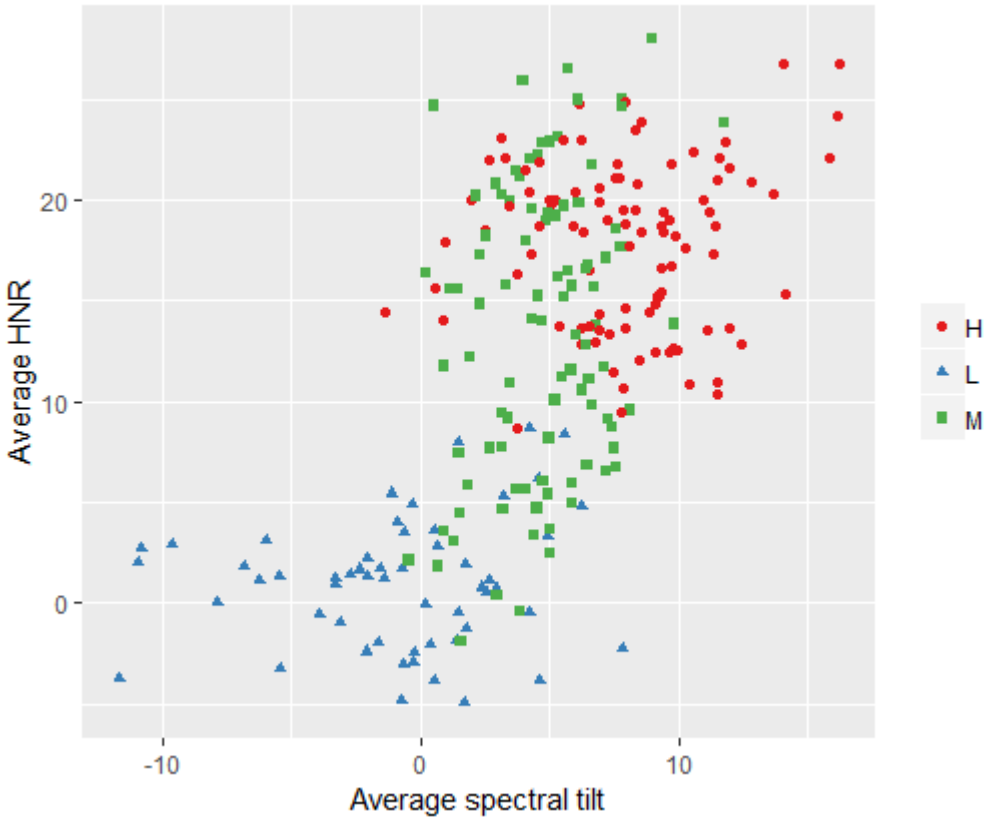

Figure 4: Average HNR and Average spectral tilt by tone

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 phona-

tion, 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 $R^2 = 0.68$. If it is the case that duration is also a cue for tone level, it is important to consider this in experiment two. A set duration should be used for all tokens in order to avoid possible participant reliance on duration as a cue for tone level. 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.

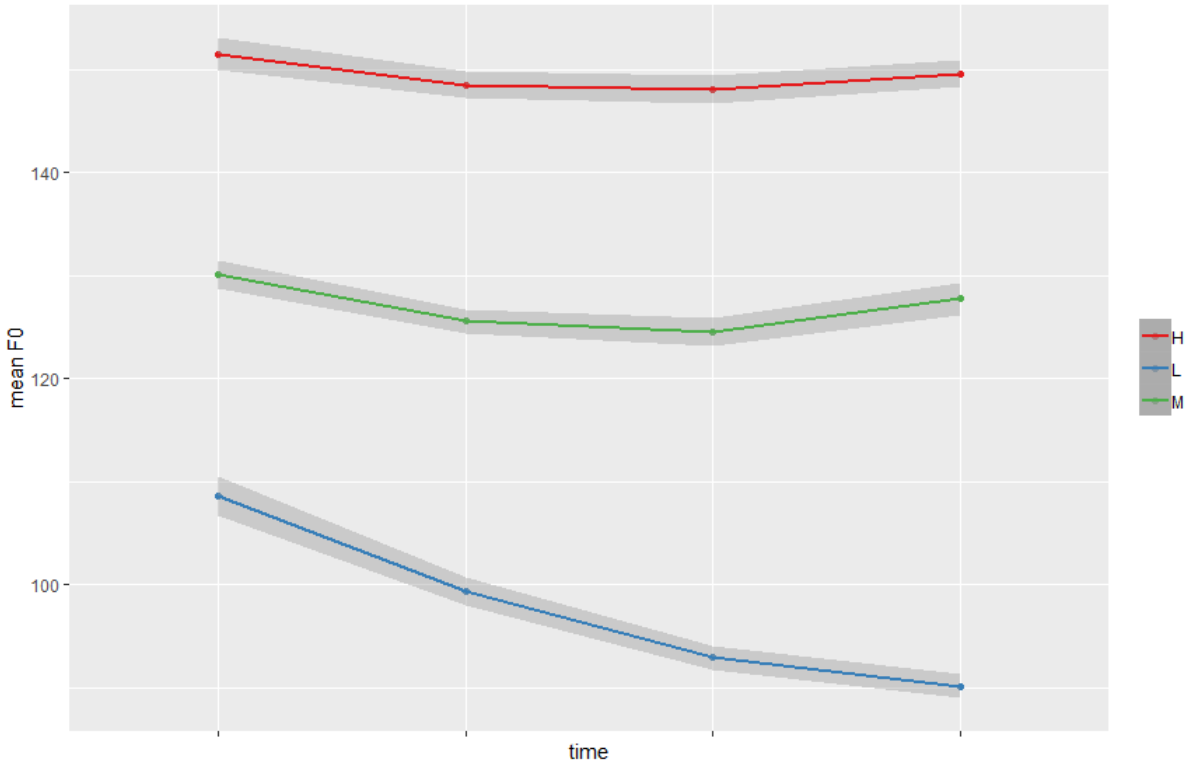

Figure 5: mean F0 over course of vowel

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 variables 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 as the predictor indicates a significant difference between both low and mid tones ($\beta = 11.76, p < 0.001$) and low and high tones ($\beta = 15.92, p < 0.001$) with model $R^2 = 0.80$. This result indicates that low HNR reliably marks low tones in Yoruba, and suggests that HNR is one acoustic implementation of creak in the language.

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 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 significant difference is found for HNR between the first slice of the vowel and the fourth slice ($\beta = -3.78, p < 0.001$). 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 ($\beta = -2.70, p < 0.001$). 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 binary 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 categorically lower spectral tilt values than mid or high tones. A linear mixed effects model with mean spectral tilt as the dependent variable and tone cate-

gory as the independent variable reveals a statistically significant effect. Spectral tilt increases both when moving from a low tone to a mid tone ($\beta = 5.02, p < 0.001$) and when moving from a low tone to a high tone ($\beta = 8.12, p < 0.001$) with model $R^2 = 0.51$. This result indicates that spectral tilt value marks low tones in Yoruba in a way that is different from mid or high tones, and suggests that spectral tilt is one acoustic implementation of creak in the language.

There is some variation in spectral tilt throughout the course of the vowel. A linear mixed effects model with low tone vowel portion as the independent variable and spectral tilt as the dependent variable shows a non-significant trend between the first and second slice ($\beta = -1.97, p = 0.03$) 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 ($\beta = -1.76, p < 0.001$) and first and fourth portions ($\beta = 9.16, p < 0.001$) 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 may not "kick in" until the latter portion of the vowel.

3.4.5 F1-F0

For F1-F0, higher values are indicative of creaky phonation. The means presented in Fig. 3 are suggestive, but unclear - it appears that low tones may pattern differently from the mid or high tones, but the standard deviations are quite high compared to the means. A linear mixed effects model with average F1-F0 as the dependent variable and tone level as the independent variable indicates a statistically significant difference between both low and mid tones ($\beta = -64.00, p = 0.01$) and low high tones ($\beta = -94.96, p =< 0.001$) with model $R^2 = 0.57$. This result indicates that F1-F0 value is distinctive among Yoruba low tones versus mid or high tones, and suggests that F1-F0 is an acoustic implementation of creaky

voice in the language.

A linear mixed effects model with F1-F0 as the dependent variable and vowel portion as the independent variables finds no significant effect in the low tone category. A significant effect does appear in the mid and high tone categories. This effect occurs between the first and fourth vowel slices in both cases, with $\beta = 286.83, p < 0.001$ for mid tones and $\beta = 310.37, p < 0.001$ for high tones. By contrast, for low tones moving from the first portion of the vowel to the fourth brings a change of only $\beta = 67.58$ with $p = 0.15$. This supports the claim that high and mid tones behave differently from the low tone as concerns acoustic correlates of creaky voice.

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 data regarding change in the measurements over the course of the vowel suggest a possible confinement of creak to a certain portion of the vowel, or at least a clear division of the tonal space with low on one end and high and mid on the other in terms of measurements that reflect creaky voice. In the future, addition of data from more speakers will hopefully support the conclusions drawn here. The next section details the results of the perception experiment.

4 Procedure 2

The second task is to test what role creaky voice has to play in the perception of spoken Yoruba. This is achieved by way of an experimental task that presents Yoruba speakers with an audio token and asks them to classify the token based on tonal category. The tokens themselves come from the data of the production task described above that have been altered (or not) in a way that facilitates answering the stated research questions.

There are multiple hypotheses surrounding this experiment and further potential experiments that can be tested. The first, the null hypothesis, is that creaky voice has no effect on perception. Adding creaky quality

to non-creaky tokens or removing creaky quality from creaky tokens does not result in significant categorization errors on the part of participants.

A second hypothesis is that creaky voice effects perception in Yoruba such that low tones with creak removed are miscategorized as being of some other tonal category with significant frequency.

A third hypothesis is that creaky voice effects perception in Yoruba such that mid and high tones with artificially added creaky quality are miscategorized as being of some other tonal category with significant frequency.

A final hypothesis is that creaky voice is as important of a perceptual cue for low tone in Yoruba as the falling contour. Harrison (1996) found that none of his stimuli with flat F0 contours were perceived as low by Yoruba-speaking participants, suggesting that the fall in F0 is important for perception. Is creaky voice a similarly salient cue in the acoustic signal?

Participants are presented with stimuli from the first experiment. For low-tone tokens, half are chosen to be creaky and half non-creaky, following the acoustic measurement-based criteria for creaky voiced described above. Within these two categories of low tone, half of the tokens keep their natural falling F0 contour, while others are synthesized to have a constant F0 value over the duration of the vowel based on mean F0 for low tone for the recorded speaker. This is done using Pitch Synchronous Overlap and Add (PSOLA) in Praat. The result is four different kinds of low tone tokens – creaky with falling F0, creaky with steady F0, non-creaky with falling F0, and non-creaky with steady F0. This allows for the isolation of creak as a perceptual cue and for comparison with contour in terms of perceptual prominence. It also contributes to the difficulty of the task and will help ensure participants are not performing at ceiling. Duration and amplitude are also controlled for using PSOLA to assign each token a duration equal to the grand mean of token durations and scale average intensity of each token to a constant value.

When the participant hears a token, they are simultaneously given a decision task where they must match the audio data to a written word – one of three that will appear on screen, representing each possible tone level for that CV syllable. Figure 6 below gives an example:
This is what participants who have just heard *ta* at any tone level or phonation value will see. They then select which word they feel best matches the audio input. Participants are told that the token was extracted from the frame sentence in §3.1 and that they will be evaluated on speed and ac-

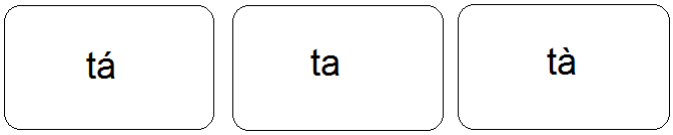

Figure 6: sample visual token for task two

curacy of response. The stimuli and possible responses are presented in a random order. As above, the task and recording of responses are managed via PsychoPy v3.0. The number of stimuli and repetitions is the same as in the experiment described above.

4.1 Measurements 2

The confusion matrix for the entire task will be calculated, and further detailed analysis of the low tone category will be conducted. Mixed effects logistic regression will be used to assess the accuracy of identification within low tone category including fixed effects *creaky* (yes, no), *contour* (flat, fall), *participant sex* (male, female), *consonant* (/l/, /n/, /t/), *nonce* (yes, no) and their interactions.

Following a similar procedure, (Yu and Lam 2014) found an 80% identification accuracy for creaky Cantonese tone 4 but only 60% identification accuracy for non-creaky Cantonese tone 4 tokens. If creaky phonation does have some perceptual reality in Yoruba, a similar result can be expected. Yu and Lam chose Cantonese specifically because it has six tones, and – they suppose: "phonation differences are more likely to be used when a language has more tones, which thus are less reliably distinguished by pitch alone" (p.1321). Results from Yoruba, a language with only three tone levels, will provide evidence in support or against this claim. Overall levels of confusion between all tones may also shed light on how the tones of Yoruba relate to one another, as discussed above in §2.2.

5 Discussion

This paper proposes a two-step method for a beginning to an answer for the question: what role does creaky voice play in the perception of tone

in Yoruba? The first step is an acoustic experiment in the same vein as Hayward et al. (2004) to confirm the presence of creaky voice on Yoruba low tone and to gather a sample of CV words that can be manipulated and utilized in the perception task. In the perception task, Yoruba speakers are presented with a series of audio stimuli that they then must match to a tonal category.

The question of the importance of creaky voice in Yoruba perception is addressed through four hypotheses: the null hypothesis, that creak plays no role in perception; a hypothesis that creaky voice affects low-tone perception such that creakless low tones will be miscategorized; a hypothesis that introducing creak to high and mid tones will cause them to be miscategorized; and a hypothesis that creaky voice is as important for low tone perception as is the falling contour. The results of the experiment will allow for rejection or acceptance of these hypotheses, giving an answer to the stated research question. The first two, as well as the final hypothesis can be directly assessed via the perception experiment. The hypothesis regarding addition of creak to non-creaky tokens may be beyond the scope of the current project but may be addressed, at least in a pilot study, if time permits.

There are several theoretical points of interest as well. If it is true that creaky voice is important in perception of spoken Yoruba, then it indicates that non-phonological information in the acoustic signal is being used to make phonological classifications (tone level), placing it in the same realm as Cantonese (Yu and Lam 2014), providing further evidence for the claim that tone and register languages are part of a continuum with fuzzy boundaries (Abramson and Luangthongkum 2009). This research project could also pull apart the importance of the contour versus the importance of creak for perception of low tone in Yoruba.

Of great importance for future work in this research program is the efficient synthesization of creaky voice. Synthesis of creaky voice has received much attention in the literature. Drugman et al. (2012) and Raitio et al. (2013), in efforts to to improve naturalness of text-to-speech systems, propose methods for synthesization of creak that are useful here. They note the familiar hallmarks of creaky voice: lower F0, irregular periodicity, and "secondary laryngeal excitations", seen as pitch spikes further along in the waveform than are observed in modal voicing. Several such excitations are visible in the creaky voice waveform in Figure 2. Considering these properties of creaky phonation, they state that: "a non-creaky voice

is ... successfully transformed to use creak by modifying the F0 contour and excitation of the predicted creaky parts" (Raitio et al. 2013, p.2316).

The authors test four systems. One that is judged to be quite natural in subsequent evaluations by listeners replaces the (lack of) creaky laryngeal excitation properties of a non-creaky speaker with those of a prototypical creaky voice speaker. The result is speech that retains the original F0 level (and so preserves the tone) but now carries some marks of creaky phonation. A second system that is also evaluated to sound reasonably natural substitutes the F0 stream of a non-creaky speaker for that of a creaky speaker in addition to the aforementioned laryngeal excitations. This method may not be suitable for a tone language such as Yoruba, which relies on F0 to differentiate words.

Another option is to alter the *jitter* value of tokens artificially. Jitter is a measure that indicates the degree of aperiodicity present in the audio signal. Jitter can be manipulated via Praat (Boersma 2001) with a script that adds jitter to selected sound objects. Whichever method is ultimately employed, a native Yoruba speaker will evaluate the result in terms of naturalness before any audio tokens are used in the task.

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.

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.

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

Drugman, T., Kane, J., and Gobl, C. (2012). Modeling the creaky excitation for parametric speech synthesis. *Proceedings of Interspeech*.

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.

Kirk, P., Ladefoged, P., and Ladefoged, J. (1984). Using a spectrograph for measures of phonation types in natural language. *UCLA Working Papers in Phonetics*, 61:102–113.

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.

Raitio, T., Kane, J., Drugman, T., and Gobl, C. (2013). Hmm-based synthesis of creaky voice. *Proceedings of Interspeech*, pages 2316–2320.

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.