Contributed to *DougSchrift: A Collection of Squibs and Puzzles presented to Doug Pulleyblank*, 2025. https://arts-pulleyblank-2024.sites.olt.ubc.ca/

Tone patterns of Xhosa nonce verbs

Wm. G. Bennett & Jeremy Perkins

1. Introduction

This squib seeks to drill down on a small question at the intersection of two themes woven across Doug Pulleyblank's work: tone, and emergence.

Some previous work has found compelling evidence that prosodic structure plays a critical role in determining linguistic structure and its expression. For example, Kitagawa & Fodor (2003, 2005) find that syntactic judgments in Japanese are sensitive to the prosodic structure. This implies that even when prosody is not overtly indicated (e.g. in writing), there is nonetheless a prosodic structure inferred. The obligatory presence of suprasegmental structure is also a recurrent theme in phonological analyses, including work involving principles of autosegmental representations and optimality theoretic imperatives (e.g. SPECIFY-T, HAVETONE, etc.). If prosody is an obligatory facet of mental representations of linguistic structure, and if speakers of tone languages are consequently impelled by grammar to devise appropriate tonal representations, then we naturally expect a nonce word production task to bear out whatever aspects of tonal well-formedness are phonologically productive.

The question we seek to shed light on here is: how do speakers decide what tonal pattern to assign to novel nonce morphemes? The approach we take here is a post hoc analysis of data already recorded for a production experiment on morphophonological alternations in isiXhosa. The particulars of Xhosa orthography and tonology are such that we think a closer examination of the pitch contours speakers produce in nonce verbs could shed further light on productivity and generality of other phonological patterns in the language. In particular, Xhosa's tone system exhibits some characteristics typical of pitch-accent systems, with high morphotactic predictability, and uneven but generally low contrastive functional load. This invites speculation along the lines already suggested by Roux (1998): is it **really** a *tone* system? If so, shouldn't speakers generalize tone contrasts to novel words?

Xhosa orthography does not indicate tone.[1] Our impression is that the hypothetical average member of our local speech community has some vague impression that Xhosa is a tonal language, but deeper awareness of tone and its role in grammar is not generally taught as part of basic language education. There are contrasting tonal minimal pairs in

[1] With two odd exceptions: class 10 nouns with the prefix *ii-* are differentiated in pronunciation from class 9 *i-* by a falling tone; this isn't marked as tone, but the vowel is spelled differently to indicate a distinction. Plausibly the class 10 prefix is historically *izi(N)-, with the medial [z] lost except where it contributes to bisyllabic minimality in short nouns. Class 2a is also spelled with a double vowel, as *oo-*, and is also pronounced with falling tone.

Xhosa, but their distribution throughout the lexicon is such that they carry very low functional load. Complicating things further, tonal minimal pairs often feature words that are related morphologically and/or etymologically, e.g. [ábàfúndì] 'students' and [àbàfúndì] 'they do not study' (Claughton 1992:10), which makes them especially easy to distinguish in most contexts without tone information. In sum: participants were provided with very little overt indication of what tone sequence to apply to the production stimuli.

We take up the following questions to attempt to answer here:

(1) Do speakers project underlying tone melodies onto nonce roots? If so, which ones?
(2) Is there inter-speaker variation in the tone assigned to novel wug items?
(3) Are consonant-tone interactions productive in nonce items?
    a. Do speakers show phonetic evidence of depressor consonant effects?
    b. Do tone melody assumptions change to accommodate depressor Cs?

The approach we take, coarsely, is one of brute force, using a statistical cluster detection algorithm to find structure within the normalized pitch track data. Further details are found in sections 4 and 5. Before delving into these specifics, we lay out the general background picture of the tone system of Xhosa (the slice of it we think we understand, at any rate), and what expectations the system sets up for our nonce words.

2. Tone basics of Xhosa

The tone system of Xhosa is one characterized by several very common tropes of Bantu tonology, and simultaneously by many conflicting descriptions of the finer details. Detailed analyses in various frameworks can be found in previous literature (e.g. Goldsmith et al. 1989, Downing 1990, Cassimjee 1998, etc.). The gist of what is relevant for verbs is as follows. To tip our hand in advance: this description accentuates the similarities between Xhosa's tonology and pitch systems normally regarded as accent more than tone.

Lanham (1958) sorts bisyllabic verb roots into three (surface) tonal patterns: HL, FL, and LL. Claughton (1992) notes, building on earlier suggestions by Meeussen, that the HL ~ FL contrast is a vestige of the historical loss of a vowel length distinction. That is, historically long vowels are FL, while historically short vowels are HL. Neither of these dovetails especially well with the consensus view in the Bantuist literature that the primary distinction at the tonal level is H vs Ø, and that L tones are not specified or represented underlyingly. We proceed in that vein anyway, under the assumption that the 'F' toneme of Lanham and Claughton is melodically an H or perhaps HL – i.e. that it implies an underlying /H/ autosegment is in the mix.

Many morpheme classes have straightforwardly predictable tones. For example: all third-person agreement markers bear /H/ underlyingly, and all first- and second-person agreement markers have low tones. Tense and aspectual inflections come with their own

pre-determined tone melodies, which tend to manifest at the cost of overwriting underlying contrasts on roots. These generalizations intersect to leave surprisingly little functional load for tone melodies on verb roots. The one extremely common contrast is between subject markers for 2nd.singular /u-/~>[ù-] and 3rd.singular.class1 /ú-/~>[û-].

H tones generally spread rightward toward the (ante)penult[2], which is generally assumed to bear a metrical accent, and which is very saliently lengthened at the ends of phrases. Strings of consecutive /H/ are generally ***not*** realized phonetically as a high pitch plateau, though, for two reasons. The first is that two underlying /H/ tones from different morphemes are normally realized with a downstep between them. The second is a pattern that Downing (1990) terms 'left branch delinking' and analyzes in crucially derivationally-ordered terms: underlying /H/ spreads rightward, and may subsequently de-link one or more of its left branches. Thus, according to Claughton (1992:24), the toneless (≈ Low) stem /-shukumisa/[3] 'shake', preceded by the augment-bearing infinitival prefix /ú-ku-/, has the following variant options (with the two Claughton identifies as most common marked with a check.)

(4) Rightward spreading and delinking lead to amorphous H realization in verbs:

Input:	/ú-ku-shukumisa/	/HLLLLL/
H spreading:	→ ú-kú-shúkúmìsà	ˇHHHHLL
Left delinking:	ù-kù-shùkúmìsà	ˇLLLHLL
(partial)	ù-kú-shúkúmìsà	LHHHLL
	ù-kù-shúkúmìsà	LLHHLL
		*HLHHLL

The essential generalization is that only the rightmost syllable in a contiguous H span is actually destined to be realized as phonetically H. Phonetically, such sequences are usually produced with a gradual rise up to a pitch peak on the rightmost TBU associated with the multiply-linked /H/. The rampant variation can, we suspect, be understood largely as flexibility in where one chooses to make the cut off between surface "L" and surface "H" in a sequence of syllables with gradually rising pitch. Consequently, we deviate from Claughton's notation and mark these variable tones as rising (with the admission that this representation obscures potentially important phonetic nuances).

Syllables preceding the rightmost H-linked syllable, and linked to the same H at the melodic level, could be claimed to set their pitch targets based on interpolation up to this final pitch peak. As such, the one constraint on the realization of such spans is that the Hs must all be contiguous: *HLH... is not acceptable as the start of such a pitch span. By

[2] H originating on the antepenult spreads to the penult. H originating further left spreads to the antepenult.

[3] Except where specifically noted as IPA, all examples are given in Xhosa orthography, adorned with morphological divisions and tone markings.

coincidence, that HLH structure is the normal result of adding an /H/-bearing object prefix, like class 6 /wá/. This second H tone spreads, and simultaneously seems to inhibit spreading of the H that precedes it. Counter-intuitively, this means that an underlying form with more /H/-bearing morphemes might well be realized with *fewer* surface H syllables.

(5) Routine non-transparency of tonal exponence

Input:	/ú-ku-shukumisa/	/ú-ku-wá-shukumisa/
	aug-inf-shake	aug-inf-OM.c6-shake
	'to shake'	'to shake them$_6$'
Surface:	ǔ-kǔ-shǔkúmìsà	ú-kù-wà-shǔkúmìsà
	HHHHLL or	HLLHHLL or
	LLLHLL	HLLLHLL

Functionally, then, the tone system effects a consolidation of the underlying tone distinctions into a few key locations. Regardless of which prefix contributes an /H/, it will land on the antepenult in /shukumisa/. Similarly, regardless of the underlying tone on the root, it will surface with an antepenult H as long as a prefix with an /H/ is present.

Cutting across this Gordian knot is a known effect of depressor consonants, which have a Janus-like pair of effects. First, depressor consonants depress (duh) the pitch of a H tone. (Implicitly, they do so by a noticeably greater degree than any phonetic pitch lowering effect they have on L tones.) Second, Lanham notes that depressor Cs prevent a following L from being raised after a preceding H. These depressor consonants are *not* the cause of the HL ~ FL contrast, though, and the FL roots do not necessarily contain depressors in the final syllable.

3. Expectations and predictions

Given the tone system of Xhosa, and given the set of data available for this post hoc analysis, what should we expect speakers to produce? We see a few intuitive hypotheses, which we lay out here, starting from the most general end.

(6) Hypotheses and expected possible outcomes

a. H1: speakers might replicate the same set of tone melodies in the lexicon
b. H2: speakers might default to a single unmarked tone
c. H3: speakers might assign tones at random
   (and/or in probabilistic way(s) beyond our understanding; same thing)
d. H4: Depressor consonants might drive tone melody choice

Naively, we might expect that when speakers produce a set of nonce roots, they will ascribe tones to them in a way that roughly mirrors the distribution of tone melodies in the lexicon. If this is the case, we may then find a difference between three clusters of f0 contours: one cluster for each of Lanham's tone classes. Alternatively, we might find just two f0 contours: one cluster for roots with an underlying /H/, and one cluster for roots without (which might be analyzed either as underlyingly toneless, or as having underlying /L/ instead).

If the contrast between HL and FL roots is a historical vestige of the falling contour on long vowels, we should expect (i) that speakers will not generalize it to novel words they encounter, and/or (ii) to the extent that the contrast *is* generalized to nonce items, it should be analogous to its segmental distribution in real words.

On the other hand, it would also be unsurprising if the tones assigned by speakers show an emergence of the unmarked effect. Classically, H is regarded as more marked than L. This disparity is also echoed by the Bantuist tradition of analyzing two tone systems as H vs Ø, with the latter literally not being marked. If this is the main factor driving the choice, then speakers' responses might skew towards being predominantly low tone.

There is also, of course, the possibility that our expectations are entirely wrong and off-base. If so, we ought to then find that speakers are consistently doing something that is unlike either of the two scenarios above. For example, they might assign tone melodies at random. Or speakers might assign tones by analogizing from real words in the phonological neighbourhood of the nonce items, or from words semantically related to an ideophone that the wug item reminds them of. It would be impossible for us to distinguish between such possibilities, but we can still expect any of them to lead to a much less clustered set of data.

Our final prediction concerns depressor consonants. Given the lack of any requirements for any given trial to have any particular tone, we might expect that the presence or absence of depressor consonants will cause participants to assign different tones. This makes transparent sense in a phonetically-driven phonology framework: if depressor Cs are somehow (e.g. violably) incompatible with H tones, we might expect them to have a repulsive effect. That is, speakers should be more likely to assign a /H/ melody to forms that have no depressors. From a casual OT-adjacent perspective, we might expect this effect to emerge clearest where there is no underlying tone melody to be faithful to.

Some of our nonce roots have initial depressors; this was not controlled for. The medial Cs we picked for the nonce items are (IPA) [m ɲ mb ndʒ] (orth.) *m ny mb nj*. Lanham's list of depressor consonants includes /mb/ and /nj/, but not /m/ or /ny/. Thus, we might expect to see these roots split into different clusters if the depressors affect tone melody or pitch contour.

4. Methodology

4.1. Procedures & materials

Fifteen mother-tongue isiXhosa speakers were recruited by word of mouth and convenience/snowball sampling. All were adults residing in Makhanda (formerly Grahamstown; iRhini in Xhosa), in the Eastern Cape province of South Africa, where Xhosa is the dominant local language.

Participants were recruited for a suite of three production studies, all designed to test productivity of morpho-phonological alternations not relevant here, using a wug test style task. Our aim in this paper is a post hoc analysis of the production stimuli that were recorded for purposes of coding the data.

Recordings were made in the Sound Laboratory at Rhodes University, with a zoom h4 digital recorder and a Nady HM-10 head-mounted microphone (variably placed to accommodate participants' various headgear and hairstyles). For our analysis, each speaker's recorded production tokens were segmented into prefix and stem domains. Misreadings and other errors were excluded from our data set. Where participants made false starts or hesitated, these were excluded; if they subsequently corrected themselves and produced the test item correctly, these tokens were included (the notion being to get the most normal production token from each trial).

The task used to collect the data is described in Bennett & Braver (2020), but we summarize the basic details here. Speakers were presented with an inflected nonce verb form, and a blank to fill in a passive form for that verb. For example: *iyafamba* → *iya*___*wa*. Nonce items were presented with a frame of inflectional morphology for a present continuous verb with no following object, with a third person subject of noun class 9. Thus, '*iyafamba*' could be translated roughly as 'it is *famba*-ing'. Participants were asked to say the nonce word aloud, then generate a passive form for it. Our analysis in this paper is limited to the prompt that the participants read aloud.

The morphology aligns with the points of variation expected if different speakers assume different underlying tone patterns on the nonce items. Every word in every trial began with the prefix sequence *i-ya-*, 'it$_9$ is….'. Since these are the same prefixes every trial, we would naturally expect them to receive the same tone specification in each production token. If different speakers are projecting different tone melodies, the differences ought to manifest in the stem rather than the prefixes – modulo any kind of opaque interaction where the prefixes shift to accommodate a phonological co-occurrence restriction (such as

the OCP). We expect the prefix sequence /í-ya-/ to introduce an underlying H tone, which ought to spread rightward (unless the following stem contains another H).[4]

All of the wug stems were CVCV sequences constructed according to the following schema.

(7) Wug anatomy chart

{C}	a	m	a
	o	mb	
		ny	
		nj	

The canonical prosodic shape for verbs in Xhosa is CVCV, with final -a in citation form and present tense inflected forms. The set of test items was selected to balance the medial vowel and the consonant of the final syllable. Each item was presented once, and therefore participant-specific errors make the sample uneven by individual. Our test items were mixed with an equal number of real word distractor items, which were not controlled for length or tone.

4.2. Statistical analysis

The statistical methodology used here utilizes a time-series clustering analysis via the dtwclust (Sarda-Espinosa, 2024) package in R (R Core team, 2025). This methodology seeks to discover optimal groupings of f0 contours for a given token that minimize differences between contours assigned to different groupings and maximize similarity of contours assigned to the same groupings. Thus, in a study where the status of tone or f0 contour is unknown, it is ideal for diagnosing both the number of distinctive f0 contours in a data set and also the nature of those distinctive f0 contours.

The analysis was performed using the tsclust() function, with partitional clustering, using the "pam" centroid method and the "dtw_basic" distance measure. The first step in the analysis was to determine the optimal number of f0-contour clusters. Following that, the cluster assignments for each recorded token was assigned, with an analysis performed of how the factors Speaker, Word and Depressor consonant status (both medial and initial) correlated with the cluster assignments. This method allows us to optimally determine how many different intonational patterns exist, what they are, and finally, what factors correlate with each pattern. Depending on which factors found in the lexicon are

[4] We are somewhat unclear on what the expected outcome of this spreading ought to be, in part because it depends on finer points of analysis than we are have space to explore in depth here. Extrapolating from the paradigms provided by Claughton (1992), the wug stem *famba* ought to yield [ǐ-yá-fàmbà] (H spreading to the antepenult) if the stem is analyzed as toneless (=LL for Lanham), or should surface as [í-yà-fámbà] if the stem is analyzed as having /H/ (whether it be HL or FL).

generalized to nonce words, we expect different numbers of clusters. For example, there should be three clusters of f0 contours if speakers are producing the same 3 tone classes Lanham describes. If, on the other hand, speakers mentally represent these as a simpler H~Ø contrast, then we might expect to see just two clusters. Analogously, if depressor consonants have a strongly significant effect on the f0 contour, we might find them multiplying the basic tonal distinctions by two.

5. Results

The clustering analysis was tested by comparing diagnostic metrics for models using different numbers of clusters, to find which model fits our data set best. We tested this using 9 different clustering models, setting k, the number of clusters to 2, 3, 4, 5, 6, 7, 8, 9 and 10 for each model. We compared model fit by comparing three diagnostic metrics: Within-cluster Sum-of-squares (WSS), Silhouette score and Davies-Bouldin Index (DBI). First, we considered WSS, which measures how compact clusters are. WSS always decreases as the number of clusters increases and so we used the "elbow method" to identify an optimal number of clusters. The "elbow" refers to a point in a plot of WSS against k where the rate of WSS decrease slows dramatically. The k corresponding to this elbow suggests an optimal fit. Figure 1 shows how WSS decreased with increasing k for the nine models we tested.

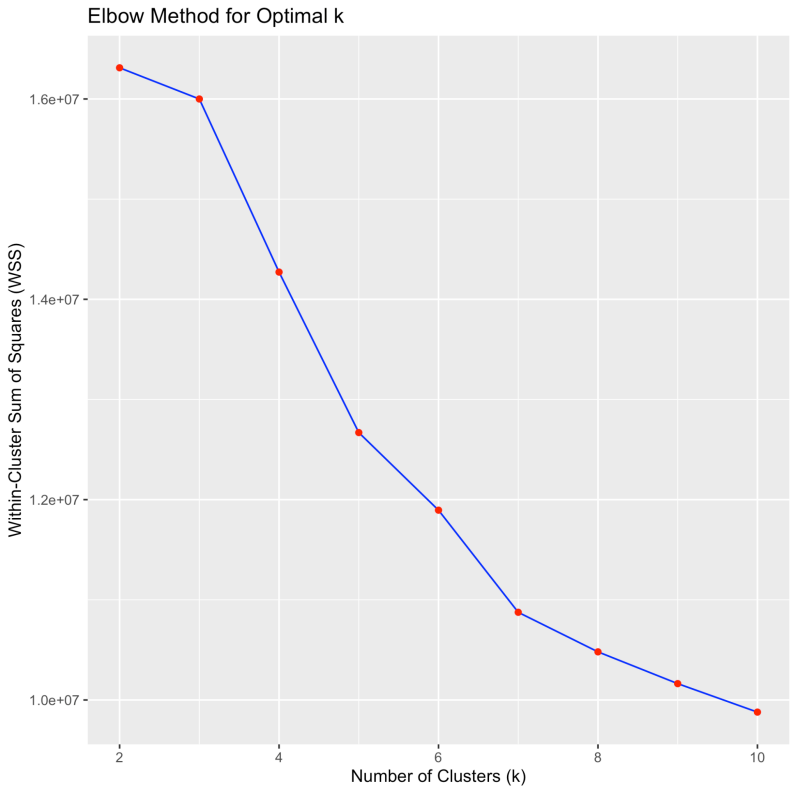


Figure 1: WSS against number of clusters; decreasing slope indicates optimal k

Figure 1 shows an "elbow" between k = 2 and 3 clusters, indicating that our data are consistent with 2 clusters of different f0 contours. Another elbow where the slope decreases noticeably occurs between k = 7 and 8 though.

Next, we considered silhouette score, which is a measure of how well each data point fits its assigned cluster. It is evaluated for each data point, with the overall average yielding a number between -1 and 1, with 1 indicating that it is well clustered and is clearly separated from other clusters, -1 indicating that it is misclassified perhaps, and 0 indicating that the point is not well clustered and lies in between two clusters. Figure 2 shows that k = 2 provided the highest silhouette score, with k = 3 nearly as high and a spike at k = 5.

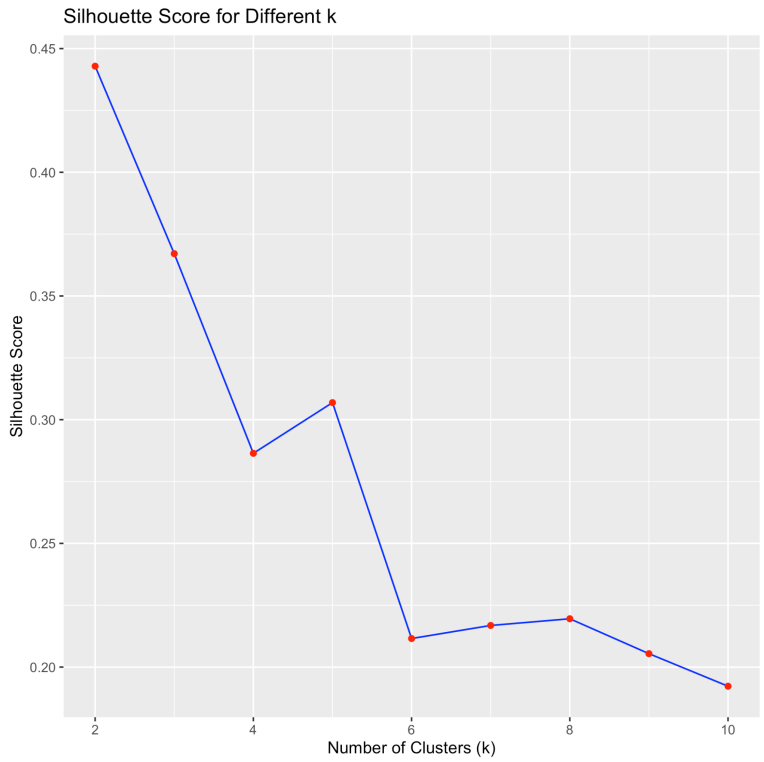


Figure 2: Silhouette score against number of clusters; higher values indicate optimal k

The third diagnostic was Davies-Bouldin Index (DBI). This involves calculating the ratio of within-cluster scatter to between-cluster distance for each cluster. Then for each cluster pair, the worst ratios are taken, and are averaged across all cluster pairs. Lower values of DBI indicate that clusters are tightly packed and well-separated from each other, indicating better model fit. Figure 3 shows that DBI is lowest for k = 2 with k = 4 close behind.

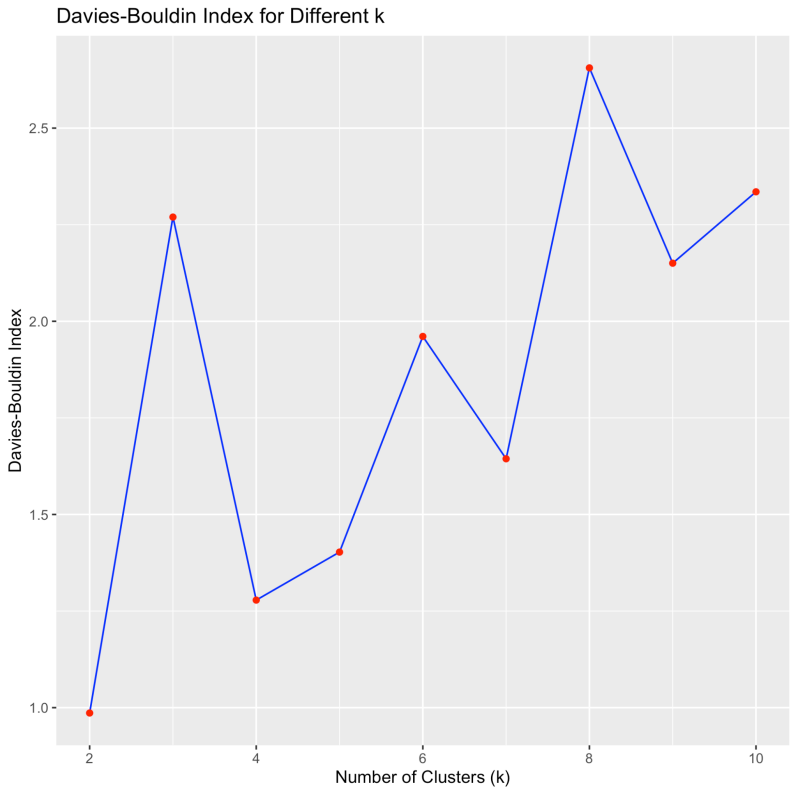


Figure 3: DBI across *k*

Taken together, our results show there are 2 clusters of differing f0 contours. Figure 4 shows smooths by cluster that illustrate one rising and one falling f0 contour.

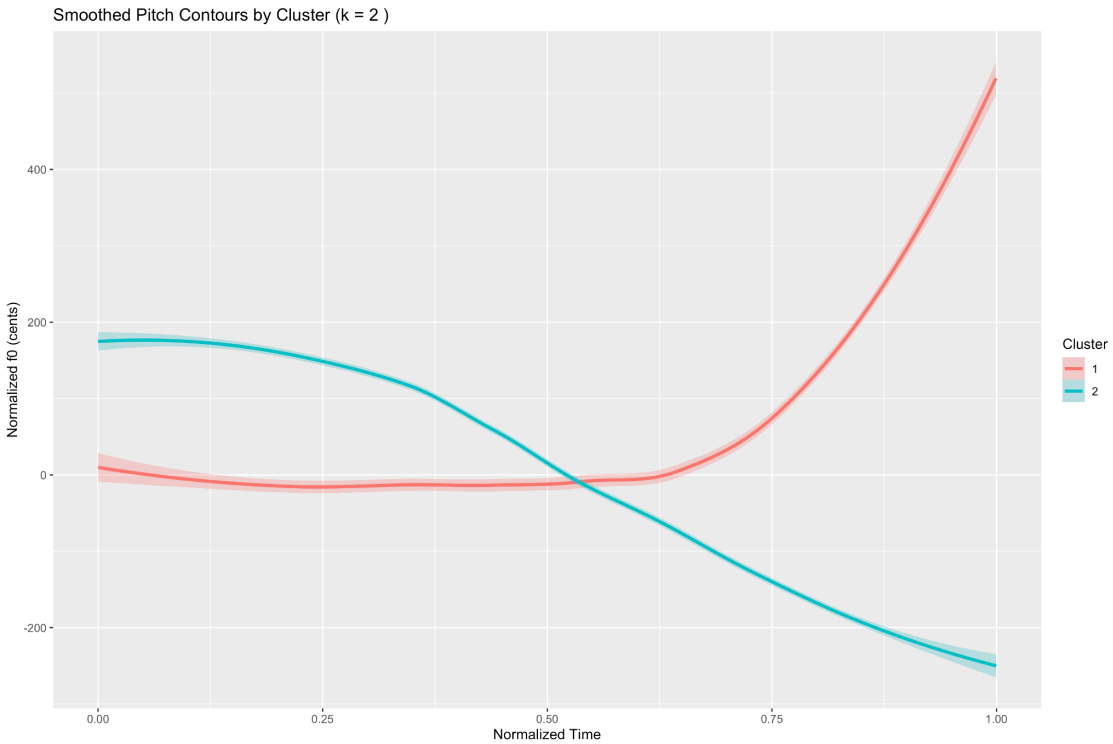


Figure 4: Smooths of normalized f0 by normalized time for the two clusters

Statistical tests were performed to test whether and to what extent the optimal clustering correlated with different factors among the nonce words, including Speaker (to test whether f0 clusters were mostly just related to differences between speakers), Initial Consonant Status (to test if the clusters were sensitive to initial consonant depressor status), and Medial Consonant Status (to test the same for medial consonants). A chi-squared test of independence was calculated to determine whether a significant association existed between the cluster grouping and levels of each factor. Cramér's V was also calculated to determine the strength of the association in each case. We found that the clustering was significantly correlated with Speaker ($\chi^2 = 15{,}398$, $p < .001$), medial consonant ($\chi^2 = 997$, $p < .001$) and initial consonant status ($\chi^2 = 718$, $p < .001$), but that the strength of this association with Speaker was large ($V = .573$), whereas the consonant effects were moderate at best (Initial C: $V = .124$; Medial C: $V = .146$).

This can be seen visually in Figure 5, where more variation is seen among clustering across speakers than across consonant status.

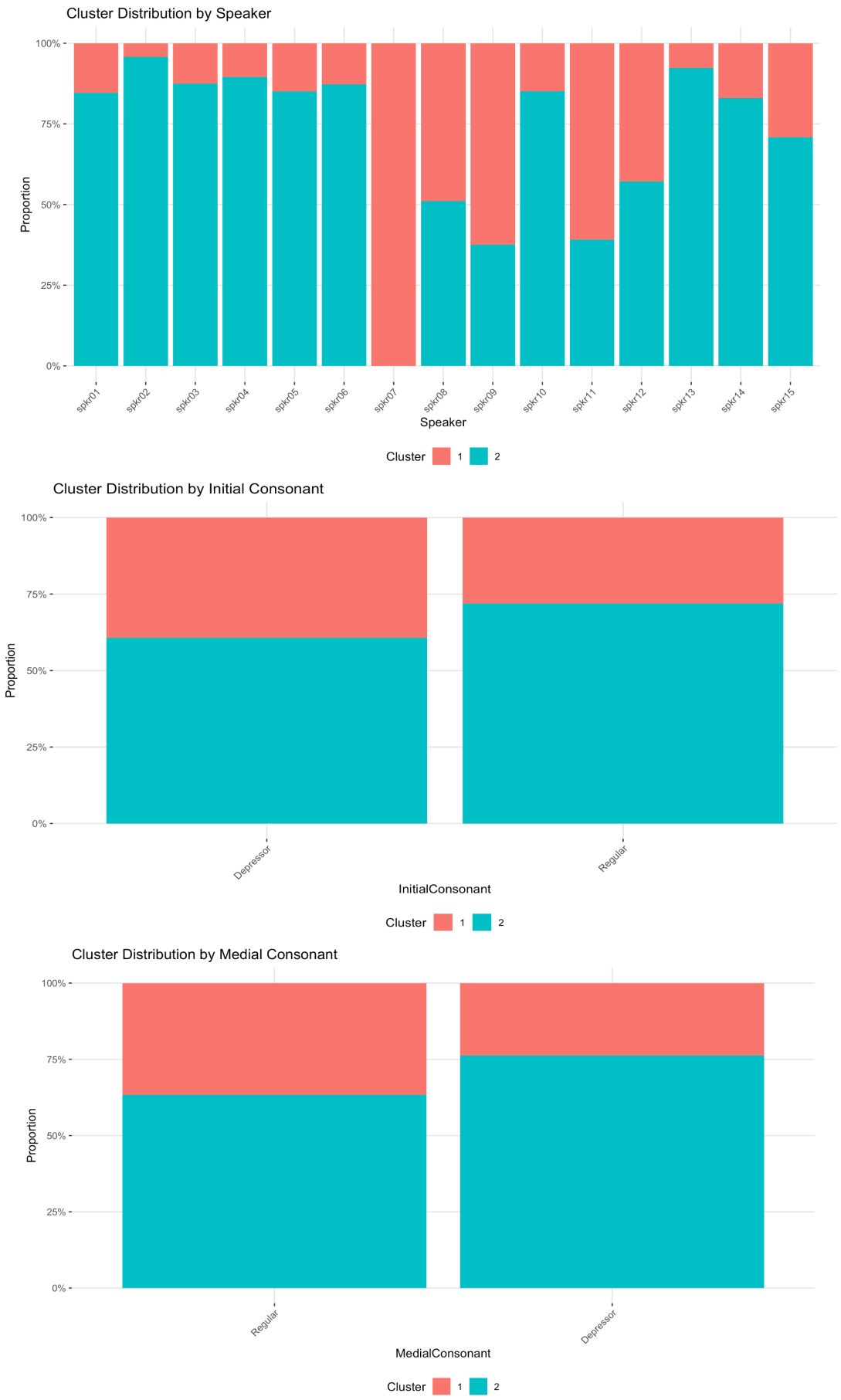


Figure 5: Cluster distribution by speaker (top), initial consonant (middle) and medial consonant (bottom).

Still, the results show a tendency for cluster 2 to correlate with depressor consonants more often, although the effect is muted in comparison with the kind of variation seen among speakers, where certain speakers gravitate completely to one or another cluster (see speaker 7 who produces the rising contour of cluster 1 almost exclusively). Basically, this suggests that most of the clustering seen is dependent on speaker differences, rather than any tonal differences. This supports the hypothesis that a single default tone, possibly equivalent to cluster 2, is being applied across the board.

In order to check correlation between these clusters and Xhosa tonal melodies, we collected recordings for real words from all 15 speakers to allow comparison with the three attested tones. The triplet zálà ('be full' or 'give birth', both high tone), lûmà ('bite', falling tone),and vùlà ('open', low tone) offer a simple illustration of these three tones. Smooths for each speaker for each of the three words are shown in Figure 6.

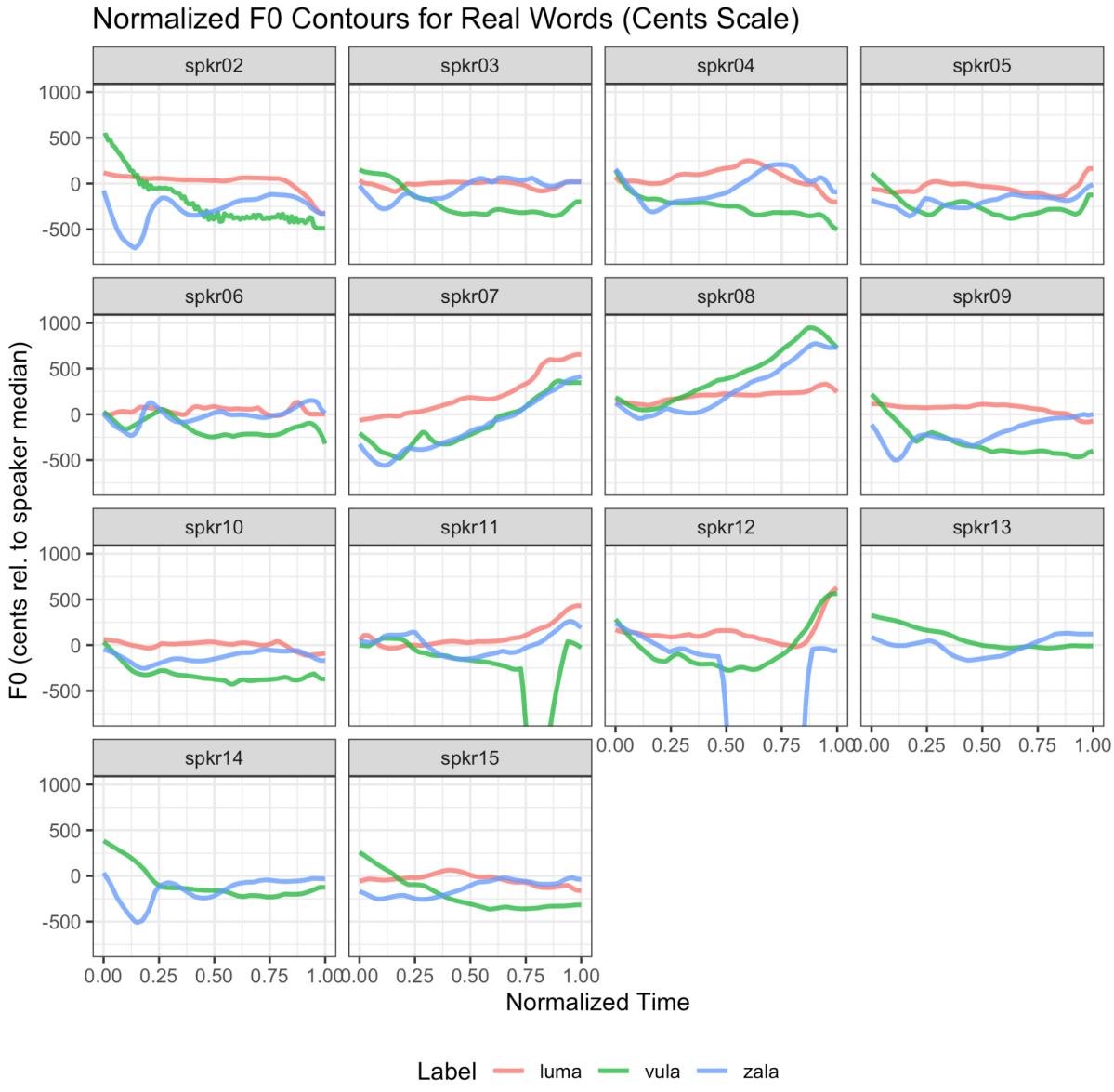


Figure 6: f0 Smooths of real words by Tone by Speaker for lûmà, vùlà, zálà

The latter two words include likely initial depressor consonants, explaining why zálà looks to have an LH tonal melody, and not an HL melody in many cases. However, the status of [v] as a depressor in vùlà seems unlikely, given the relatively high initial f0 for most speakers. Still, for its latter half, vùlà exhibits a relatively low f0, consistent with L tone. Finally, the falling tone word lûmà has a high f0 for most speakers, but there is no evidence of an f0 fall. However, it at least illustrates what an initial H tone looks like, without any depressor effect incurred by an initial consonant (cf. zálà). In fact, we may wonder if lûmà is perhaps being produced with high tone, rather than falling tone here.

Next, in order to illustrate correlation between tones and clusters, the f0 profiles of the three real words, along with five other real words were included in a subsequent clustering analysis:

(8) Real word practice items available as tone comparators

a. bôphà ‘tie’
b. thándà ‘love’
c. hlâmbà ‘wash’
d. thêngà ‘buy’
e. fúndìsà ‘teach’

This allowed us to see which cluster each real word would be assigned to, allowing any tonal patterns to emerge. Figure 7 shows individual f0 contours of each real word in colour (faint lines) with the corresponding average by-speaker smooths by tone. Black smooths by cluster for each speaker across all words (nonce and real) are overlaid for comparison to show which tones (if any) align with which cluster for each speaker.

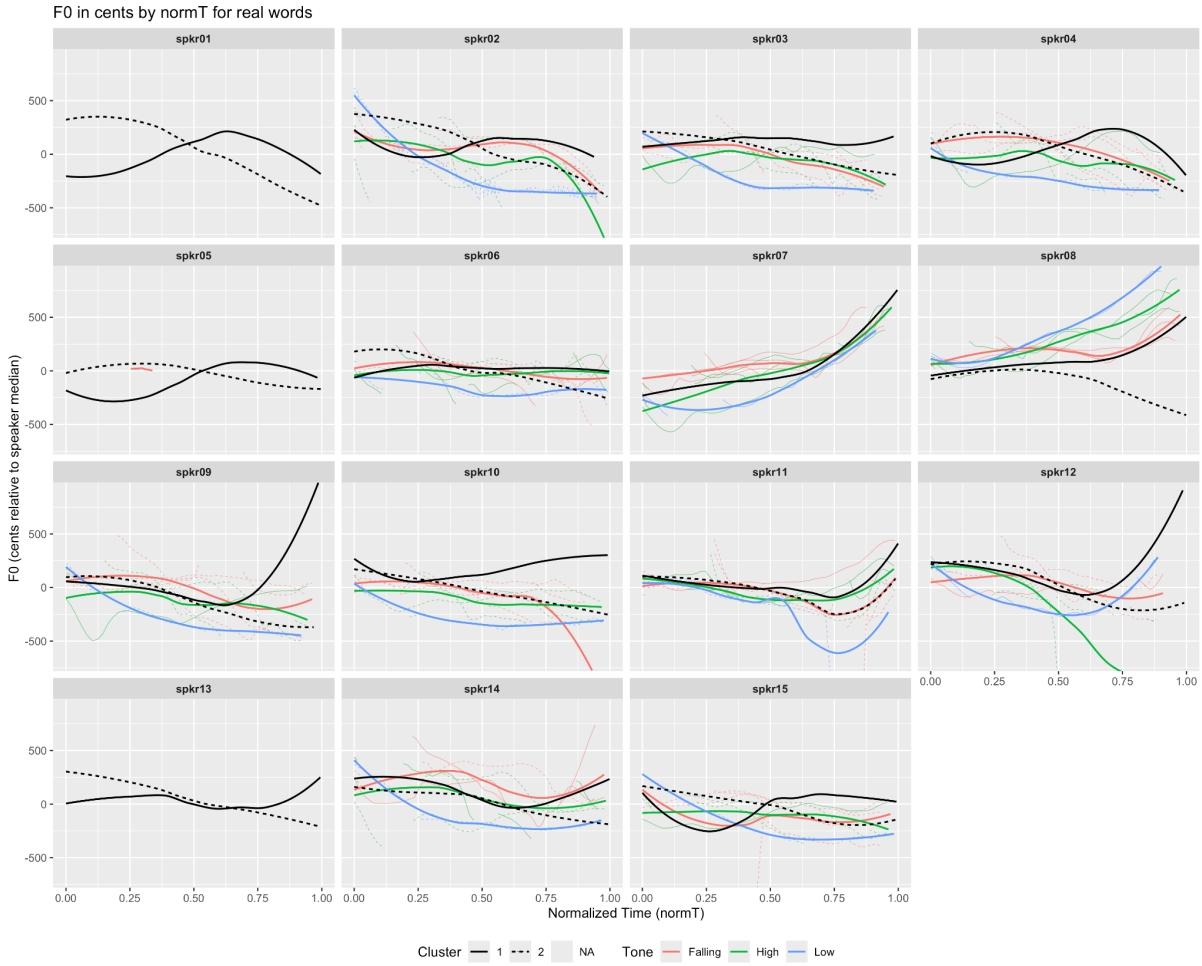


Figure 7: Smooths of real words by tone vs. smooths by cluster overlaid

Inspection of Figure 7 shows that cluster 2 tends to align with falling tone most often and sometimes high tone, but rarely with low tone. This would seem to indicate that speakers tend to produce falling or high tone in nonce words more often. However, there is one caveat: We had only one L tone real word that happened to come with an initial depressor consonant, and so more real word data should be collected in order to make definite conclusions regarding this generalization.

Next, we examine whether depressor consonants affect tonal contours in nonce words. A significant depressor effect, while moderate (see above), was seen in the expected locations among nonce words. In words with initial depressor consonants, f0 was lowered at the start of the word, and in words with medial depressor consonants, f0 was lowered midway through the word, as expected. Furthermore, the f0-lowering effects extended far into the word as shown in Figure 8, suggesting a phonological effect rather than a phonetic effect. The plot below shows the entire nonce CVCV stem, and thus the 50% mark approximates the location of the medial consonant.

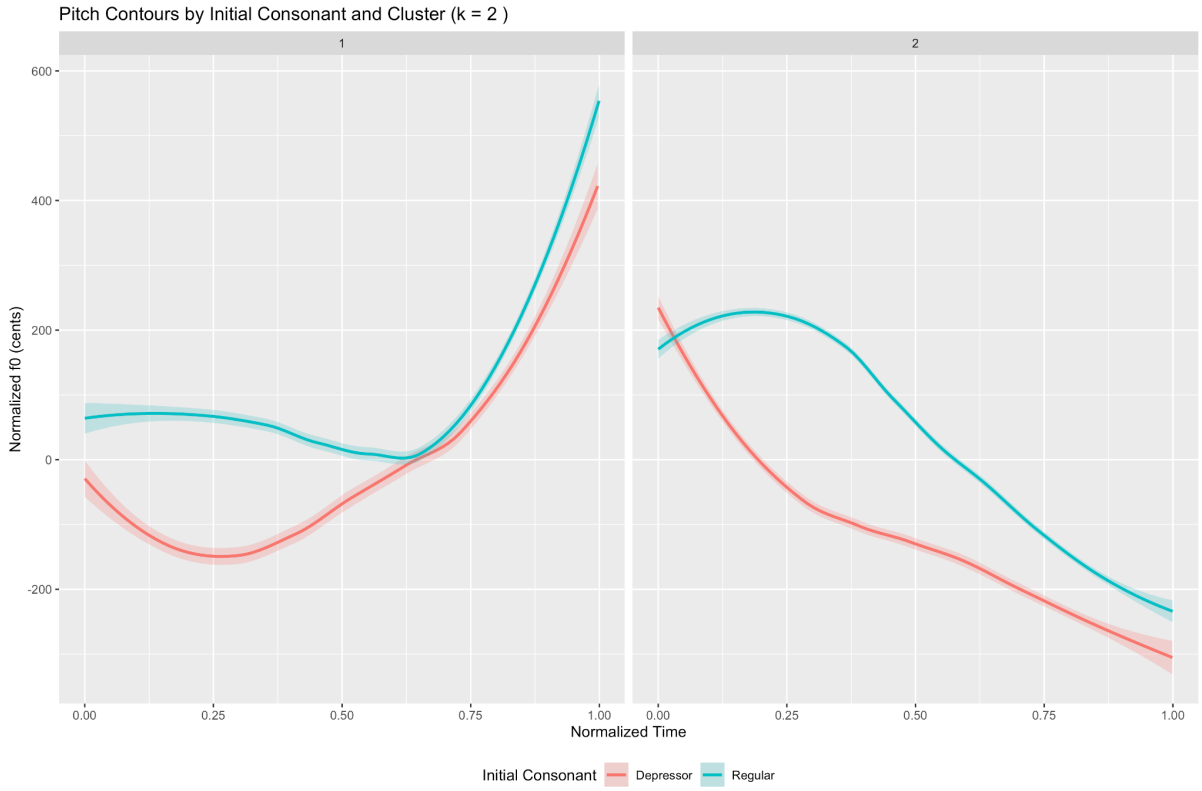


Figure 8: f0 contours following depressor and non-depressor initial consonants by cluster

The same effect can be seen for medial depressor consonants as shown in Figure 9, but this time the effect starts around the midpoint of the word, as expected.

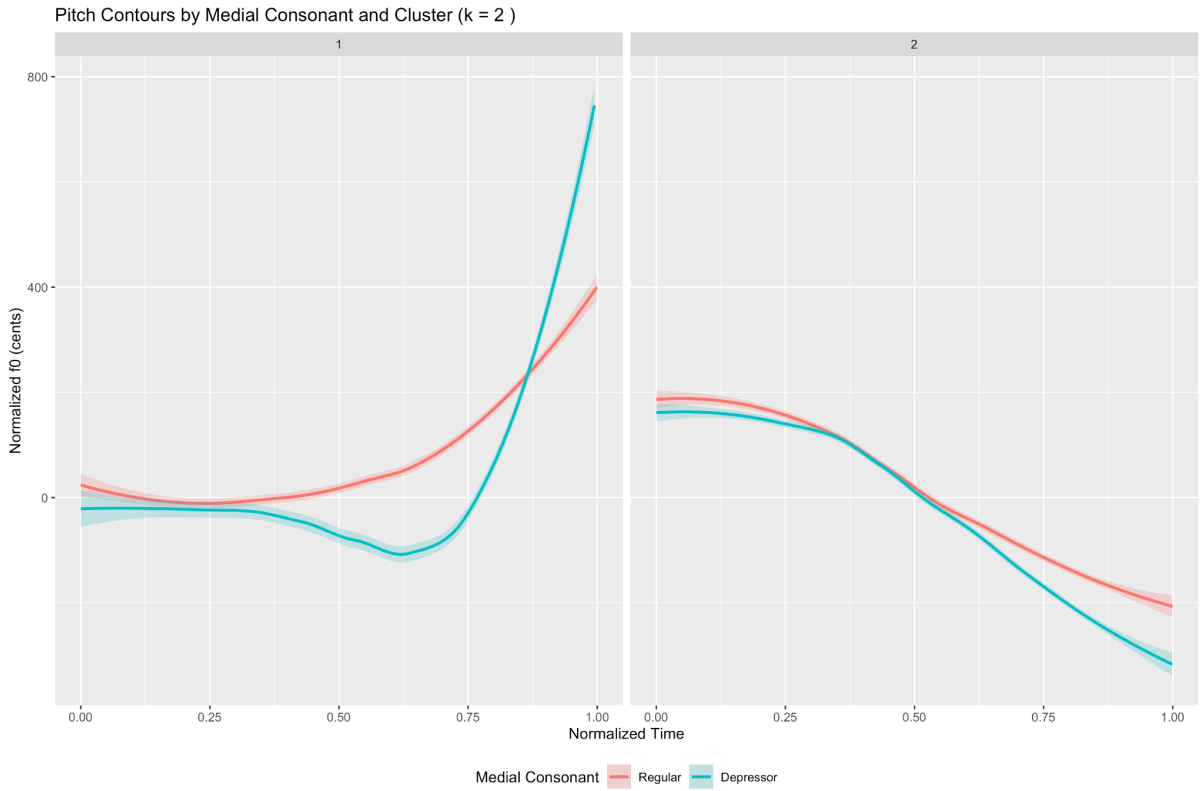


Figure 9: f0 contours following depressor and non-depressor medial consonants by cluster

Notably, this time, f0 is roughly the same for the first half of the word, but is lowered to some degree in the 2nd vowel (0.50 onward). This effect is less pronounced visually than that in the initial vowel. This can be explained by the way in which time was normalized: The anchoring points were the first and last f0 measurement of the word. This is ideal for localizing initial consonant effects, but the medial consonant effect will vary across the normalization, depending on the actual timing of the release of the second consonant, which is not an anchoring point for the normalization.

6. Discussion and conclusions

6.1. Uptalk as a task effect

Our first and least significant finding is that some tokens were produced with a pronounced final pitch rise. This is pretty obvious at a glance at the f0 contour for cluster 1 in Figure 4.

We analyze this as a task effect, though there are two plausible mechanisms that could lie behind it. The first is a difference in how speakers performed the task. Some participants read each word on its own, as though they were in isolation. This was the intended response anticipated by the task design. However, a few participants parsed the two words into a shared phrase. This means that the first word – the one we are interested in measuring – is not subject to H nonfinality and does not have the penultimate lengthening and phrase-final pitch fall that normally characterizes pre-pausal words. Intuitively, we would expect that the unfamiliarity of a laboratory context might have led some participants to rush in their responses, à la 'White Coat Syndrome' and the Lab Coat Effect. This could lead to a bias towards parsing the two words of each trial into a single prosodic phrase. Conceivably, that difference might eliminate the nonfinality effect that prevents H from spreading onto final syllables, by rendering the first word non-phrase-final.

Alternatively, and probably more likely, it could be that these participants were producing the phenomenon of "uptalk" rather than uniformly converging on a different prosodic structure than the rest of the group. When presented with unfamiliar pseudowords, participants might have reasonably applied a rising intonation. While this phrase-final rise is not the typical intonational pattern for wh-questions in isiXhosa, all participants were familiar with English, which does use this intonation pattern in questions – particularly echo questions. Speaker 7 in particular produced quite a lot of uptalk, with the overwhelming majority of their tokens being assigned to cluster 1 by our statistical analysis.

We cannot distinguish these factors with our data, we can only observe that some speakers have a systematic final high pitch for reasons that are probably unrelated to their assumptions about underlying tone melodies of each word. There is also the possibility of

synergistic interactions between both factors: the 'fill in the blank' nature of the task lends itself to treating the trial as a question-answer pair, where the stimulus is a question and its passive form the answer.

Whatever the reason, some tokens had an unexpected rising pitch. This puts them outside the scope of the questions we are primarily after, namely how speakers apply tones to unfamiliar roots. Future work can sidestep this issue by re-running the clustering analysis, while excluding all data assigned to cluster 1, thus removing the "uptalk" portion from the data and focusing only on the data that lacks the uptalk effect. The clustering analysis would undoubtedly be affected by this and so it may more accurately identify real tone differences.

6.2. Do speakers project tone contrasts onto nonce roots? No.

An admirably simple result would be that speakers assign nonce verb roots to one of the three tonal classes. This is not what we find. A similarly simple finding would be that speakers assign each stem either an /H/ or an /L/. This is also not what we found.

Speakers do not project tone contrasts onto nonce roots. Instead, our findings indicate that speakers vary considerably from each other, but that there is a tendency to produce falling or high tones rather than low tones in nonce words. A more thorough phonetic analysis of the actual tonal melodies in real words would be needed to make a definite conclusion here. Still, this finding where falling or high tone is more common in nonce words is unexpected as it is the opposite of that predicted by markedness theory, where an unmarked tone, low in this case, is expected.

6.3. Depressor consonants

Depressor consonants were found to affect f0 in nonce words in the ways expected: f0 was lowered in vowels following depressor consonants. The effect extended throughout the vowel, and was large enough to be a phonological effect rather than a phonetic effect.

Interestingly, despite the noted difference, the clustering analysis did not yield separate clusters for f0 contours following depressors vs. non-depressors. This may indicate that the size of the uptalk effect was large enough to overwrite any other effects. Future work would seek to remove this effect; in fact, a follow-up analysis of this work could simply exclude the cluster 1 data and re-analyze the remaining cluster 2 data. This may allow finer distinctions in f0 contours to emerge, such as the depressor vs. non-depressor distinction, as clusters that were previously hidden. It is indeed even possible that some tonal differences may emerge in that case as well. We leave this to future work.

Finally, our data allowed us to measure the size of the effect of depressor consonants. In order to accurately measure this, we focused on our cluster 2 at the 25% time point. The reason for this is that cluster 1 seems to involve a marked intonational

pattern corresponding to a possible task effect, whereas cluster 2 seems to represent the default f0 pattern. In addition, our time normalization was aligned to the left edge of each nonce stem, allowing the initial consonant effect to be measured more accurately than the medial consonant effect. As such, we focused at 25%, roughly midway through the first vowel. By inspecting Figure 8, it's apparent that depressor consonants involve lowering of f0 by 350 cents (from 220 cents to −50 cents or 3.5 semitones). This establishes a normalized baseline for the size of the depressor effect in Xhosa.

References

Bennett, Wm. G., & Braver, A. (2020). Different speakers, different grammars: Productivity and representation of Xhosa labial palatalization. *Phonological Data and Analysis*, *2*(6), 1-29

Cassimjee, Farida. (1998). Isixhosa tonology: an Optimal Domains Theory analysis. Munich: Lincom Europa.

Claughton, John Sellick. 1992. The Tonology of Xhosa. Makhanda (Grahamstown), ZA: Rhodes University PhD dissertation. http://hdl.handle.net/10962/d1002171

Downing, Laura J. (1990). Local and metrical tone shift in Nguni. *Studies in African Linguistics*, *21*(3), 261-318.

Goldsmith, John, Peterson, K., & Drogo, J. (1989). Tone and accent in the Xhosa verbal system. *Current approaches to African linguistics*, *5*, 157-178.

Kitagawa, Yoshihisa and Janet Dean Fodor. (2003). Default Prosody Explains Neglected Syntactic Analyses of Japanese, McClure, William, ed., In *Japanese/Korean Linguistics* 12, 267-279. CSLI Publication.

Kitagawa, Yoshihisa and Janet Dean Fodor. (2005). Prosodic influence on syntactic judgments. *IULC Working Papers*, *5*(2).

Lanham, L. W. (1958). The tonemes of Xhosa. *African Studies* 17(2):65-81.

R Core Team (2025). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

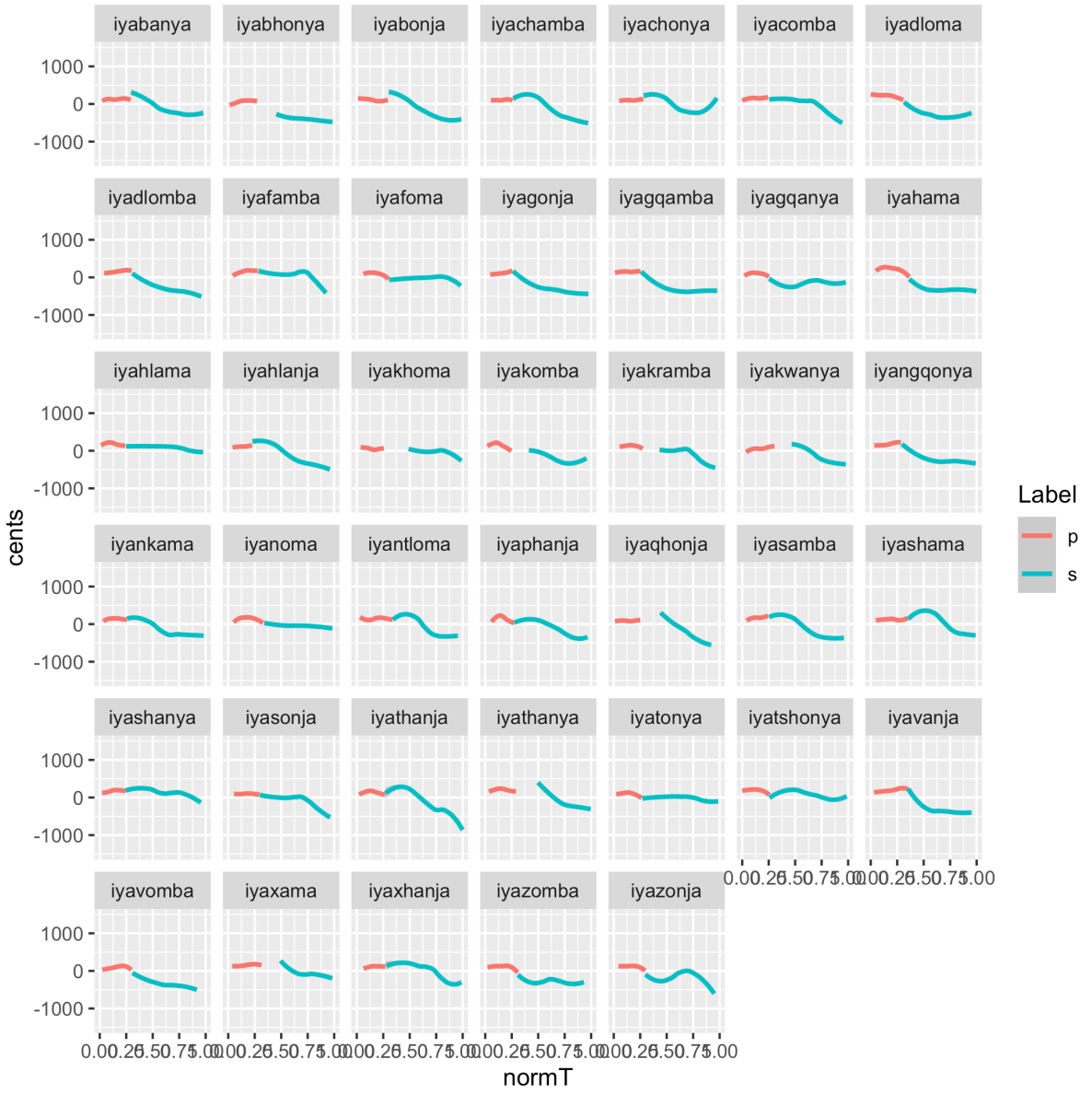
Roux, Justus C. (1998). Xhosa: A tone or pitch–accent language? *South African Journal of Linguistics*, *16*(sup36), 33–50.

Sarda-Espinosa, A. (2024). *dtwclust: Time Series Clustering Along with Optimizations for the Dynamic Time Warping Distance*. R package version 6.0.0, <https://CRAN.R-project.org/package=dtwclust>.

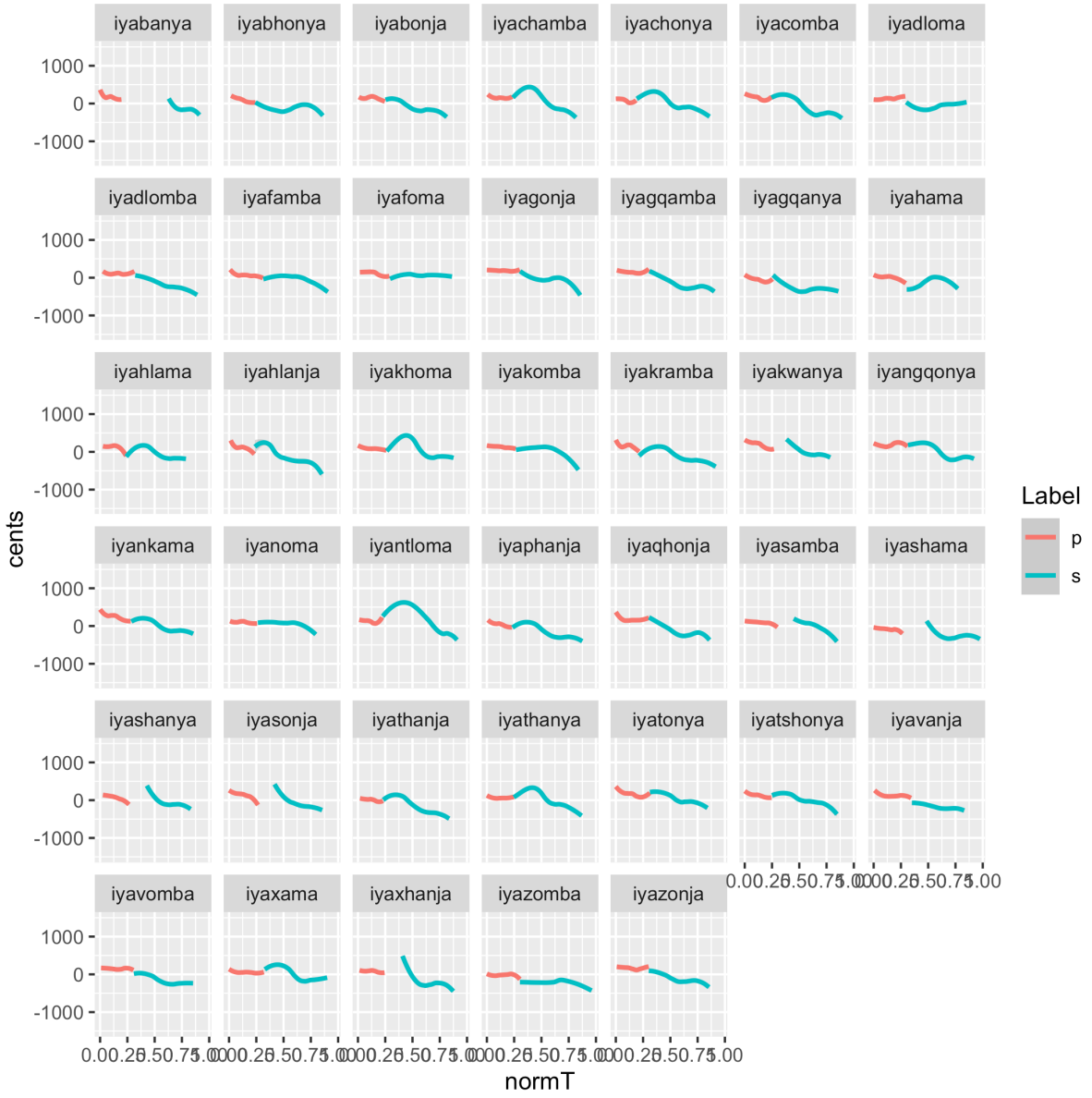
Appendix

The following figures show sets of pitch tracks for a couple of speakers, organized per token, to give an impression of degree of within-individual variation. Intervals are marked as ‘p’ for prefixal domain, or ‘s’ for the stem. The juncture between them coincides with the start of each wug root.

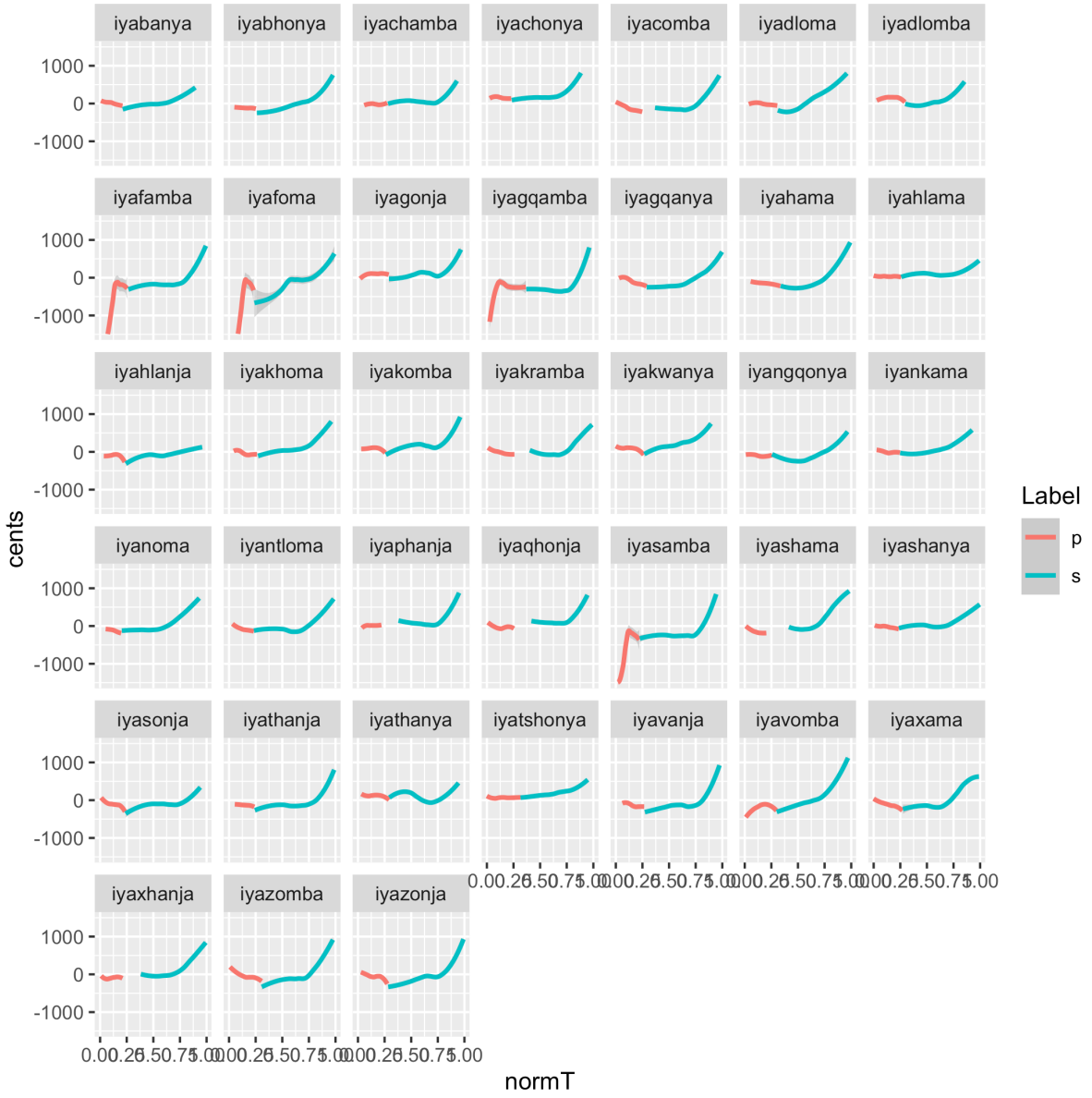
Speaker 3:



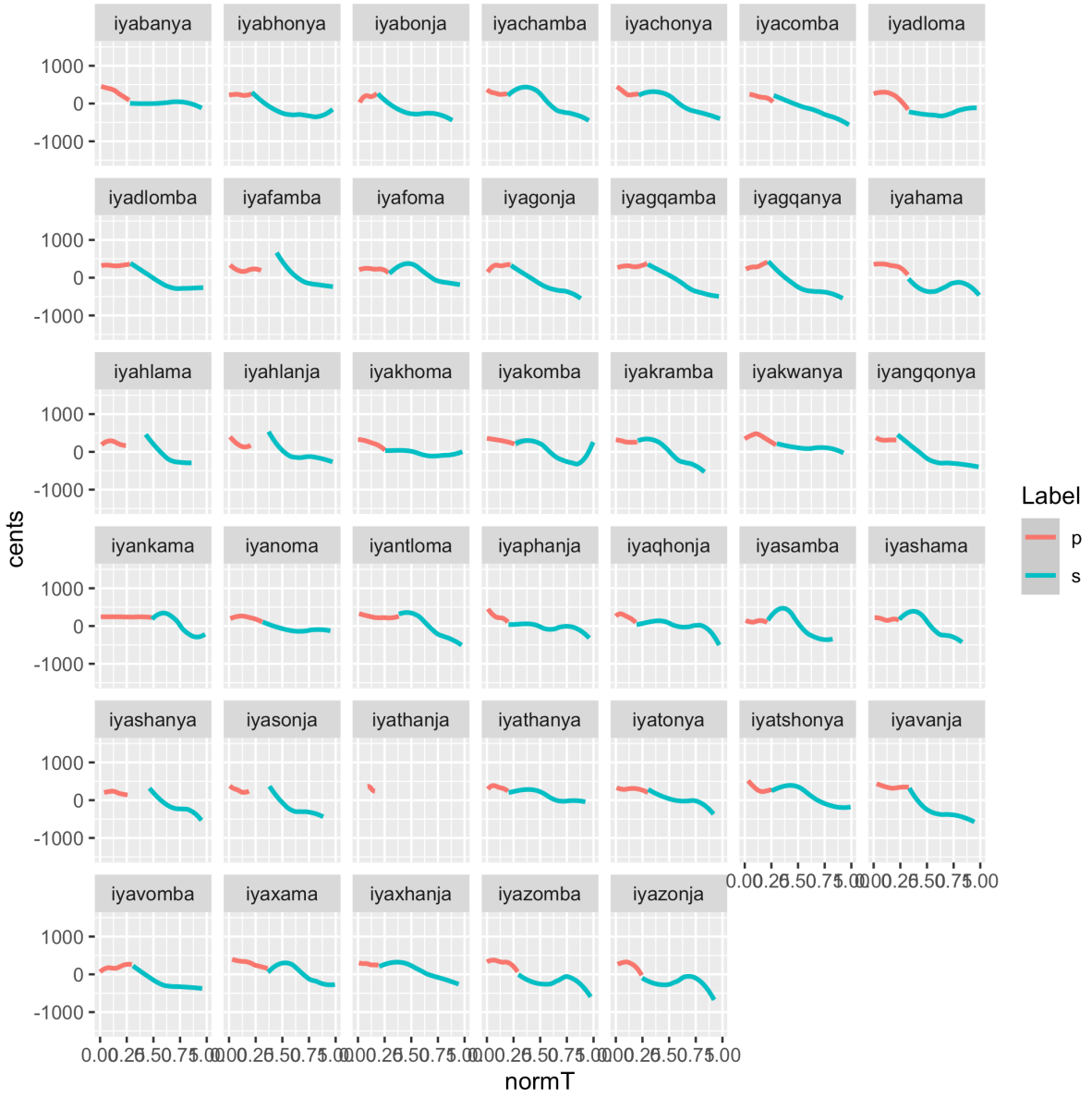
Speaker 6:


iyabanya
iyabhonya
iyabonja
iyachamba
iyachonya
iyacomba
iyadloma
iyadlomba
iyafamba
iyafoma
iyagonja
iyagqamba
iyagqanya
iyahama
iyahlama
iyahlanja
iyakhoma
iyakomba
iyakramba
iyakwanya
iyangqonya
iyankama
iyanoma
iyantloma
iyaphanja
iyaqhonja
iyasamba
iyashama
iyashanya
iyasonja
iyathanja
iyathanya
iyatonya
iyatshonya
iyavanja
iyavomba
iyaxama
iyaxhanja
iyazomba
iyazonja
1000
0
-1000
0.00
0.25
0.50
0.75
1.00
cents
normT
Label
p
s

Speaker 7:


iyabanya
iyabhonya
iyachamba
iyachonya
iyacomba
iyadloma
iyadlomba
iyafamba
iyafoma
iyagonja
iyagqamba
iyagqanya
iyahama
iyahlama
iyahlanja
iyakhoma
iyakomba
iyakramba
iyakwanya
iyangqonya
iyankama
iyanoma
iyantloma
iyaphanja
iyaqhonja
iyasamba
iyashama
iyashanya
iyasonja
iyathanja
iyathanya
iyatshonya
iyavanja
iyavomba
iyaxama
iyaxhanja
iyazomba
iyazonja
1000
0
-1000
0.00
0.25
0.50
0.75
1.00
cents
normT
Label
p
s

Speaker 13:


iyabanya
iyabhonya
iyabonja
iyachamba
iyachonya
iyacomba
iyadloma
iyadlomba
iyafamba
iyafoma
iyagonja
iyagqamba
iyagqanya
iyahama
iyahlama
iyahlanja
iyakhoma
iyakomba
iyakramba
iyakwanya
iyangqonya
iyankama
iyanoma
iyantloma
iyaphanja
iyaqhonja
iyasamba
iyashama
iyashanya
iyasonja
iyathanja
iyathanya
iyatonya
iyatshonya
iyavanja
iyavomba
iyaxama
iyaxhanja
iyazomba
iyazonja
1000
0
-1000
0.00
0.25
0.50
0.75
1.00
cents
normT
Label
p
s