Acoustic Correlates of Harmony Classes in Somali

Wendell Kimper // William Bennett // Christopher Green // Kristine Yu

Draft: 30 September 2016

1 Introduction

The vowel inventory of Somali (East Cushtic) is commonly described as containing five major vowel categories {i,e,a,o,u}, each of which is contrastive for length and for an additional feature that has been variously described as FRONT/BACK (Andrzejewski, 1955), ±ATR (Saeed, 1993), TENSE/LAX (Green et al., 2015), and (aryepiglotallically) SPHINCTERED/EXPANDED (Edmondson et al., 2004).

This last feature is of particular interest, since it is implicated in a phonological process of vowel harmony that Andrzejewski (1955) describes as extending iteratively beyond word boundaries. If this description is accurate, Somali may constitute the sole putative case of truly iterative harmony beyond word boundaries.

However, investigating this harmony process in Somali presents a number of interesting analytical challenges. The relevant feature contrast is neither represented orthographically nor noted in dictionaries of the language, a relatively small number of lexical items have been described as belonging to one class or the other, and there are few minimal pairs. Furthermore, Andrzejewski (1955) describes inter-speaker and dialect variation with respect to lexical classification. Finally, the articulatory dimensions ascribed to the relevant feature contrasts is acoustically diffuse, making clear identification of feature values difficult without articulatory data.

In this paper, we present acoustic data from four native speakers of Somali, with the aim of describing the acoustic correlates of harmony classes and developing a method for classifying tokens of vowels whose feature values have not been described. We find statistically significant differnces between harmony classes along several dimensions relevant to

	Series A	**Series B**
dhis	'build' (Imper. Sg.)	'he built'
hel	'find' (Imper. Sg.)	'he found'
kab	'a sandal'	'he set' (e.g. a fractured bone)
qod	'dig' (Imper. Sg.)	'he dug'
tus	'show' (Imper. Sg.)	'he showed'
diiday	'I fainted'	'I refused'
hees	'song'	'he sang'
laab	'chest (thorax)'	'he folded'
duushay	'she flew'	'she attacked'

Table 1: Minimal pairs (Andrzejewski, 1955).

tongue root and/or voice quality features, but no clear evidence to support a categorical phonological distinction.

2 Background

Andrzejewski (1955) describes the difference between harmony classes as fronting or tongue advancement:

> The difference between vowels of Series A and B is that the vowels of Series B are more 'front', i.e. articulated with the mid part of the tongue more advanced towards the hard palate and teeth-ridge than the corresponding vowels of Series A.

Throughout this paper, we follow Andrzejewski in adopting Series A and Series B as labels for the two harmony classes; minimal pairs can be seen in Table 1.

Edmondson et al. (2004) provide a careful articulatory description of the difference between Series A and Series B vowels, using laryngoscopic data from a single native speaker of Somali. They argue that the main difference between Series A and Series B vowels is constriction or expansion of aryepiglottalic folds, describing the differences as in (1). They also provide some acoustic data suggesting differences in F_1 and F_2 consistent with advancement or retraction of the tongue root, and oral airflow data showing that articulation of Series A vowels results in substantially lower airflow than Series B vowels.

(1) *Properties of Harmony Sets (Edmondson et al., 2004)*
Set 1 (Series A)

a. Sphincteric compacting of the arytenoid-epiglottal aperture in the posterior-anterior dimension.

b. Vowel quality that is more retracted.

c. Voice quality that is tense.

Set 2 (Series B)

a. Expansion of the arytenoid-epiglottal aperture in the anterior-posterior dimension.

b. Vowel quality that is more fronted and/or raised.

c. Voice quality that is lax.

Edmondson et al. (2004) note that these findings and previous descriptions are consistent with contrasts that are often referred to as *register* features in other languages. This characterization informs the acoustic dimensions under consideration in our study — duration, F_0, F_1, F_1 bandwidth, F_2, spectral slope, and center of gravity have all been implicated as potential correlates of register.

Duration and F_0 have been found to be relevant for contrasts involving voice quality (Edmondson and Li, 1994; Halle and Stevens, 1969). Spectral slope is another important correlate of voice quality Kingston et al. (1997), since lax voice quality results in a relative increase in the energy of the first harmonic. In addition, Edmondson et al. (2007) note that constriction in the aryepiglottic sphincter (as as found for Series A vowels) should result in a higher center of gravity.

F_1 and F_2 are the most likely correlates of a process involving advancement or retraction of the tongue root (Starwalt, 2008). F1 Bandwidth has also been shown to be relevant to timbre differences in tongue root contrasts in Akan (Hess, 1992) and other languages (Starwalt, 2008). We have also included F_3 in the set of measurements, as it is involved in tongue root retraction in Arabic pharyngealization Ghazeli (1977).

3 Methods

3.1 Subjects and Elicitation

The present data come from four native speakers of Somali. Speaker 1 (male) and Speaker 2 (female) are originally from regions in Northern

	Speaker 1			Speaker 2			Speaker 3			Speaker 4		
	A	**B**	**U**	**A**	**B**	**U**	**A**	**B**	**U**	**A**	**B**	**U**
[u]	24	12	89	23	9	43	0	0	0	0	0	70
[i]	50	72	116	32	37	61	30	88	172	0	0	30
[a]	80	86	239	89	52	90	86	78	246	0	0	104
[o]	41	44	88	38	18	23	62	36	82	0	0	22
[e]	30	55	36	18	33	13	46	30	54	0	0	0
	225	269	568	200	149	230	224	232	554	0	0	226

Table 2: Token counts for Series A, Series B, and unclassified vowels.

Somalia; Speaker 3 (female) is originally from Central/Southern Somalia, and Speaker 4 (female) is originally from Central Somalia. Speakers 1, 2, and 4 currently reside in US diaspora communities, while Speaker 4 resides in South Africa.

Elicitation sessions for Speakers 1–3 consisted primarily of establishing familiarity with lexical items (and grammaticality of sentences) from Andrzejewski (1955). Clear repetitions were elicited for familiar lexical items, and additional items that the speakers volunteered were included for analysis. Elicitation for Speaker 4 consisted of a list of monosyllabic words, with CVC structure and flat tones; all items were previously unclassified.

3.2 Data Preparation

Measurements for F_1 bandwidth, spectral slope (band energy difference) and center of gravity were taken at vowel midpoints using Praat (Boersma and Weenink, 2008). Duration was measured from vowel onset to vowel offset, and mean measurements for F_{0-3} were taken across the middle 80% of the vowel's duration.

Only monophthongs were included in the analysis. The number of tokens of Series A, Series B, and unclasified vowels for each vowel category for each speaker is given in Table 2. To reduce collinearity and improve comparability, data were centered within each vowel category for each speaker.

4 Results

4.1 Acoustic Correlates

The first question to address is whether Series A and Series B vowels show significant differences along the predicted dimensions (and in the predicted directions). Speakers have been analysed separately, since there is reason to expect inter-speaker variation (Andrzejewski, 1955).

Because the relevant acoustic dimensions are collinear, linear models[1] (with series and vowel category as predictors) were fitted separately for each acoustic dimension, excluding extreme outliers ($|z| > 3$). Bonferroni correction was applied ($\alpha/8$) was applied to adjust for familywise error (corrected p-values are reported). For those dimensions which showed a statistically significant difference between Series A and Series B, Hartigan's Dip Test for Unimodality was applied. Data from Speaker 4 was excluded from this stage of the analysis, as it contained only unclassified tokens.

Distributions and means for Speakers 1–3 can be seen in Figures 1. Series A and Series B vowels differed in F_1 and F_1 bandwidth for all speakers ($p < 0.001$), as well as spectral slope ($p < 0.05$ for Speaker 1; $p < 0.001$ for Speakers 2–3). F_2 showed significant differences for Speakers 1–2 ($p < 0.001$) but not for Speaker 3, F_3 was significant only for Speaker 2 ($p < 0.01$), and center of gravity was significant only for Speaker 3 ($p < 0.05$). Neither duration nor F_0 showed significant differences for any speaker, however it is worth noting that Somali has tonal and prosodic processes (Green et al., 2015) that were not controlled for in elicitations, potentially resulting in noise that could obscure relevant differences.

Of the acoustic dimensions that showed significant differences, the only one to show any statistically detectable departure from unimodality was F_1 bandwidth, and only for Speaker 3. Furthermore, the source of this multimodality may not be directly related to vowel series — as can be seen in Figure 3, while the lower mode appears to consist primarily of Series A observations, the higher mode shows substantial overlap between Series A and Series B.

[1] Linear mixed effects models with random intercepts for either 'word' or 'sentence' were attempted, but rarely converged.

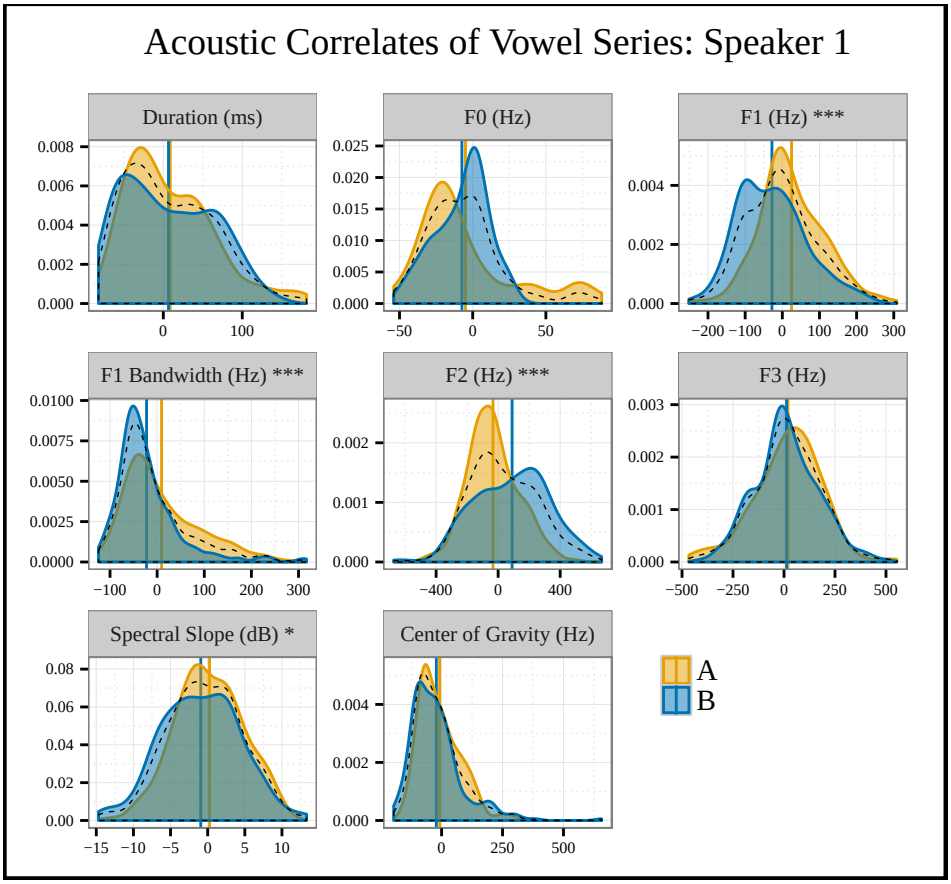

Figure 1: Density plots of Series A and Series B vowels for Speaker 1 (centered measurements, extreme outliers removed). Dashed lines represent combined distributions; vertical lines represent series means; asterisks indicate statistically significant differences (after Bonferroni correction).

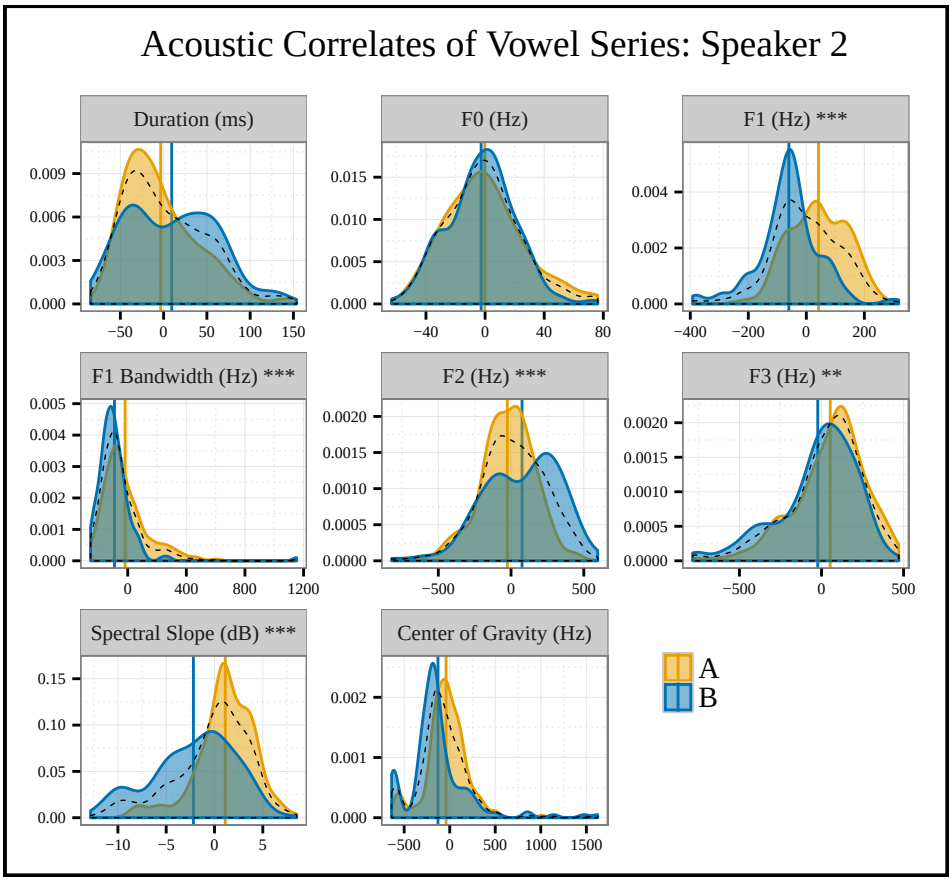

Figure 2: Density plots of Series A and Series B vowels for Speaker 2 (centered measurements, extreme outliers removed). Dashed lines represent combined distributions; vertical lines represent series means; asterisks indicate statistically significant differences (after Bonferroni correction).

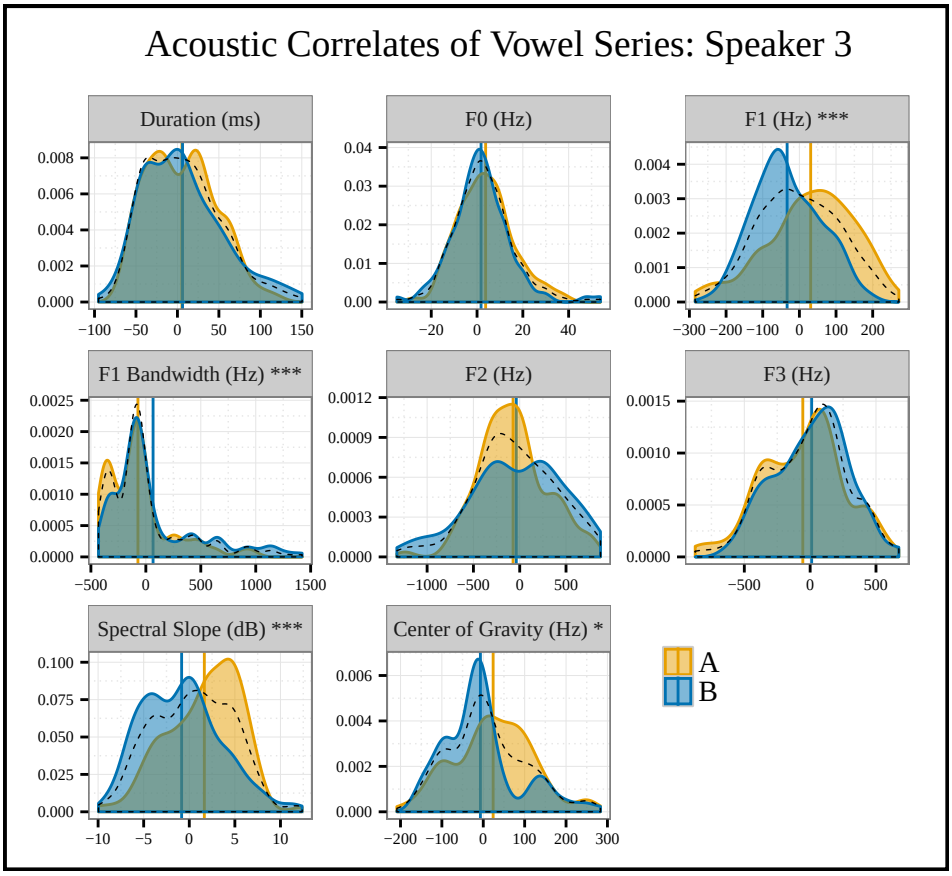

Figure 3: Density plots of Series A and Series B vowels for Speaker 3 (centered measurements, extreme outliers removed). Dashed lines represent combined distributions; vertical lines represent series means; asterisks indicate statistically significant differences (after Bonferroni correction).

4.2 Classification

For Speakers 1–3, data for both classified and unclassified tokens were subjected to k-means cluster analysis, using data from only those acoustic dimensions that had shown significant differences for any speaker in the previous stage of analysis. Series A and Series B means were used as initial centers for the clusters, and the analysis was done separately for each speaker.[2]. The results of cluster analysis matched prior classifications somewhat poorly — 66% of tokens for Speaker 1, 62% for Speaker 2, and only 54% for Speaker 3. The sets of matched tokens for each speaker (all acoustic dimensions) served as training data for a linear discriminant analysis (LDA), which was then used to predict classification values for the full set of tokens for that speaker.

For Speaker 4, Series A and Series B grand means from Speakers 1–3 served as the initial centers for k-means cluster analysis. Additionally, an initial LDA was trained on pooled classification-matched data from Speakers 1–3 and used to predict classification values for data from Speaker 4. Classifications from the cluster analysis and the initial LDA matched on 84% of tokens; the set of matched tokens served as training data for a second LDA, which was then used to predict classification values for the full set of tokens from Speaker 4.

Correlations between the resulting linear discriminant values and the measured acoustic demensions (along with the posterior probabilities resulting from LDA classification) can be seen in Figures 4–7. There is considerable variation between speakers — the only acoustic dimension whose correlation with the discriminant was consistently medium-sized or larger was spectrial slope (medium for Speaker 1, large for Speakers 2–4). All other acoustic dimensions showed medium-sized or larger correlations for at least one speaker, and all except F_1 bandwidth showed medium or larger correlations for three out of the four speakers.

As with the individual acoustic dimensions, the linear discriminant itself does not appear to show a bimodal distribution — for all three speakers, Hartigan's Dip Test on the linear discriminant fails to detect any departure from unimodality ($p > 0.05$).

Because it is not possible to verify the previously-unclassified items, alternative means are needed to assess the success of the classification procedure. Classifications for individual segments were compared across multiple tokens of each lexical item, and all items which appeared more

[2]For Speaker 2, it was necessary to remove outliers prior to cluster analysis.

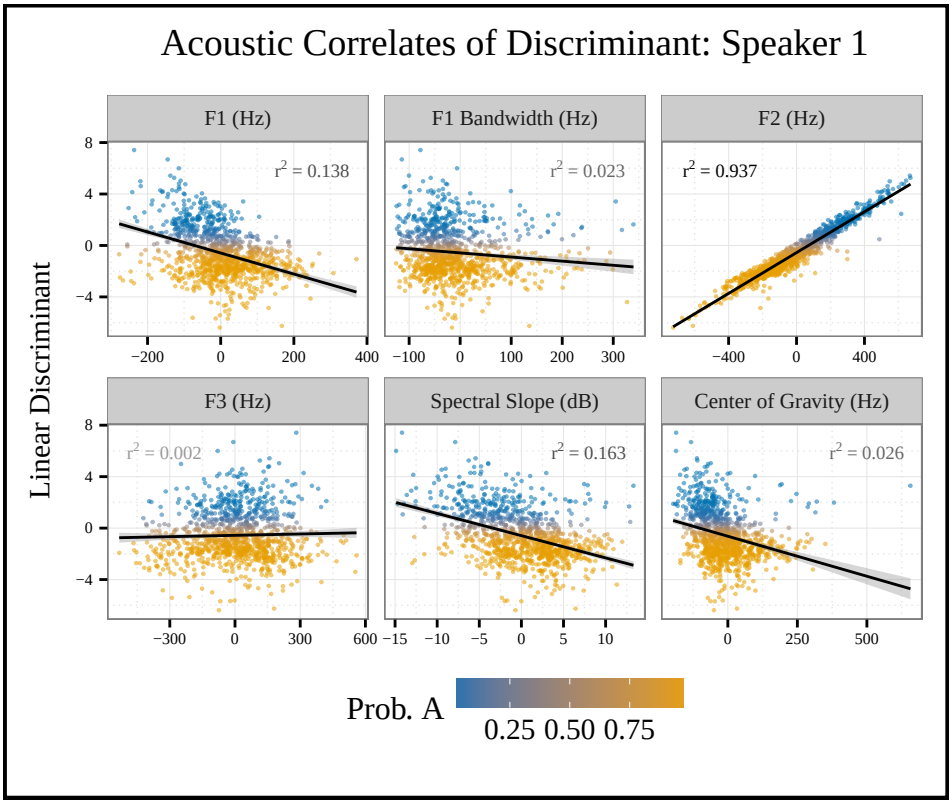

Figure 4: Relationships between acoustic measures and the LDA discriminant for Speaker 1 (centered measurements, extreme outliers removed). Straight lines represent best fit linear regression slopes; shading around regression lines indicates 95% confidence intervals.

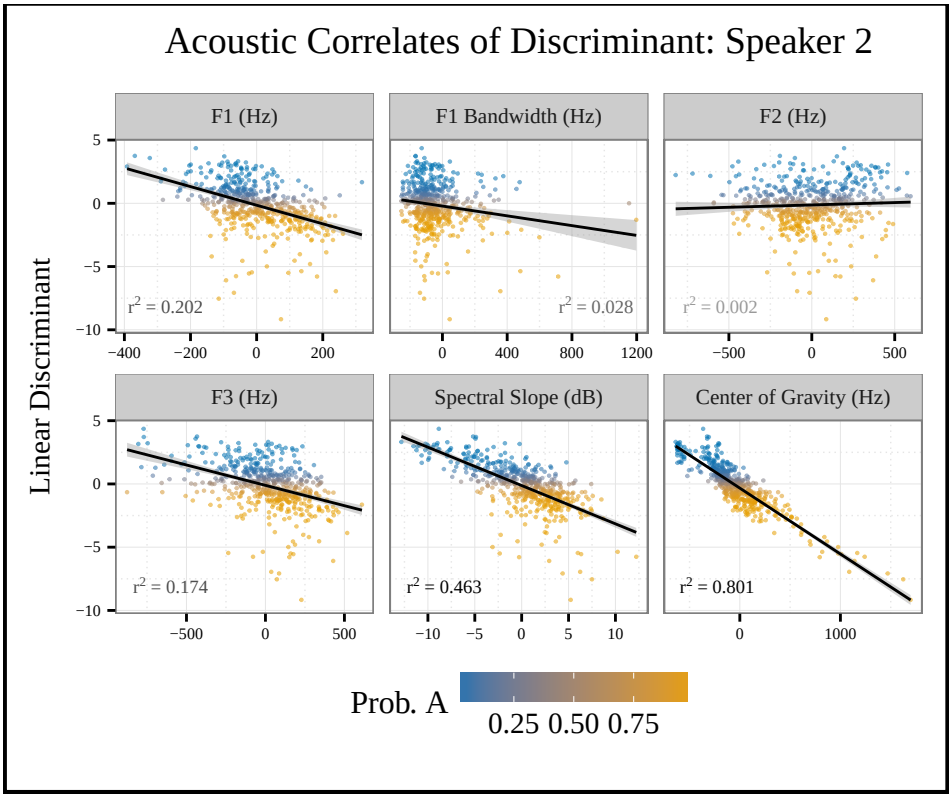

Figure 5: Relationships between acoustic measures and the LDA d iscriminant for Speaker 2 (centered measurements, extreme outliers removed). Straight lines represent best fit linear regression slopes; shading around regression lines indicates 95% confidence intervals.

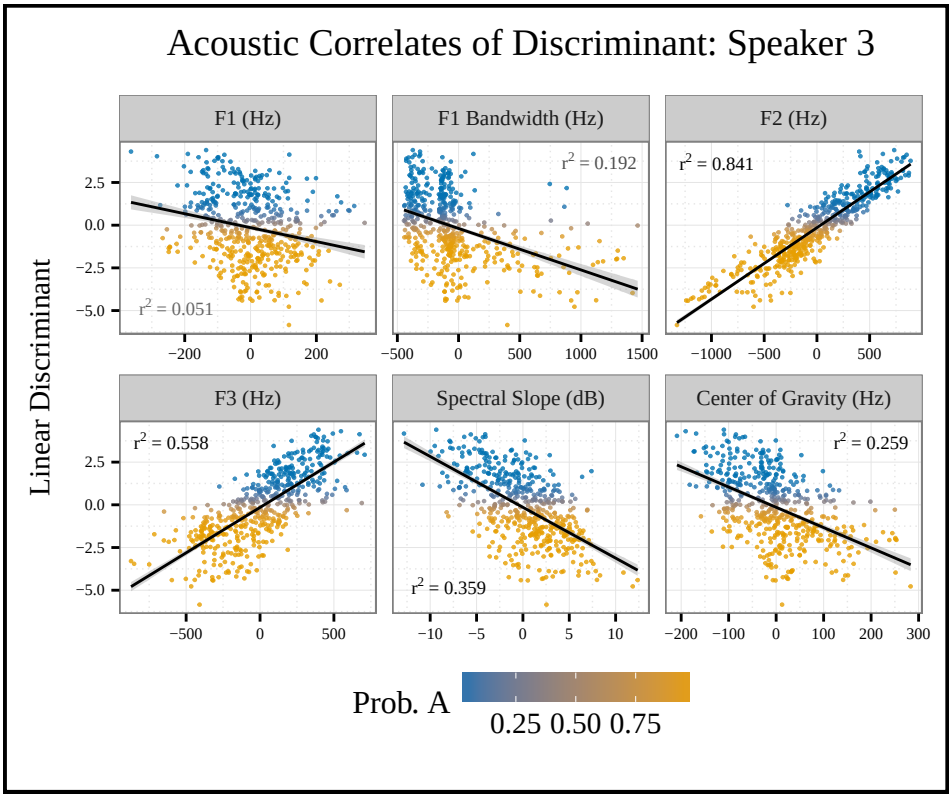

Figure 6: Relationships between acoustic measures and the LDA discriminant for Speaker 3 (centered measurements, extreme outliers removed). Straight lines represent best fit linear regression slopes; shading around regression lines indicates 95% confidence intervals.

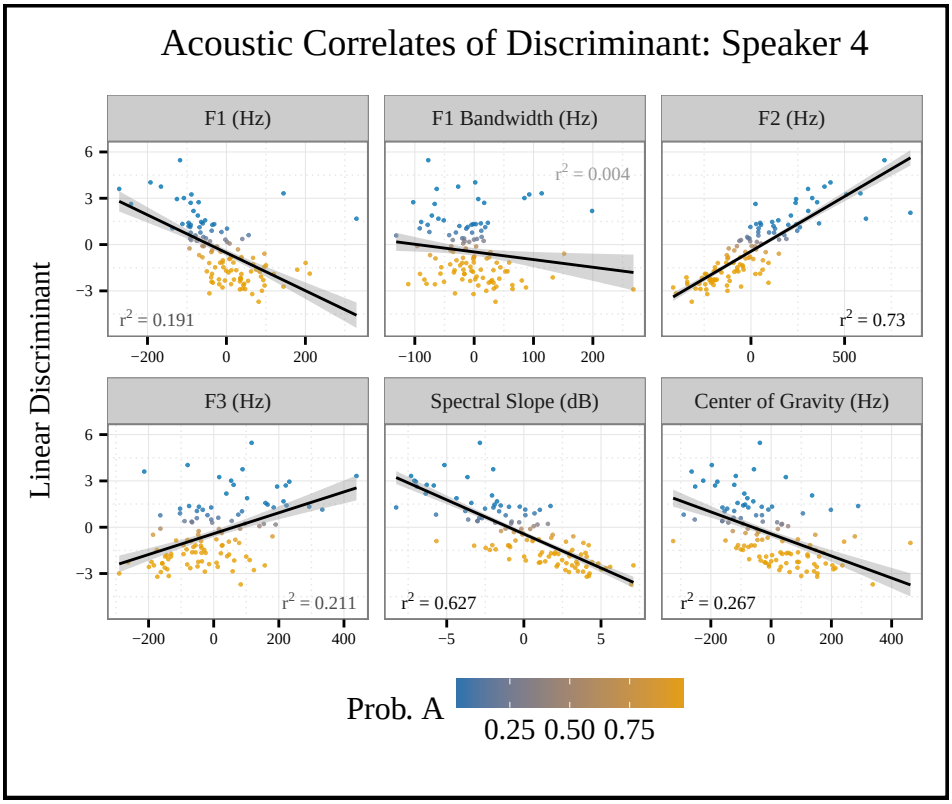

Figure 7: Relationships between acoustic measures and the LDA discriminant for Speaker 4 (centered measurements, extreme outliers removed). Straight lines represent best fit linear regression slopes; shading around regression lines indicates 95% confidence intervals.

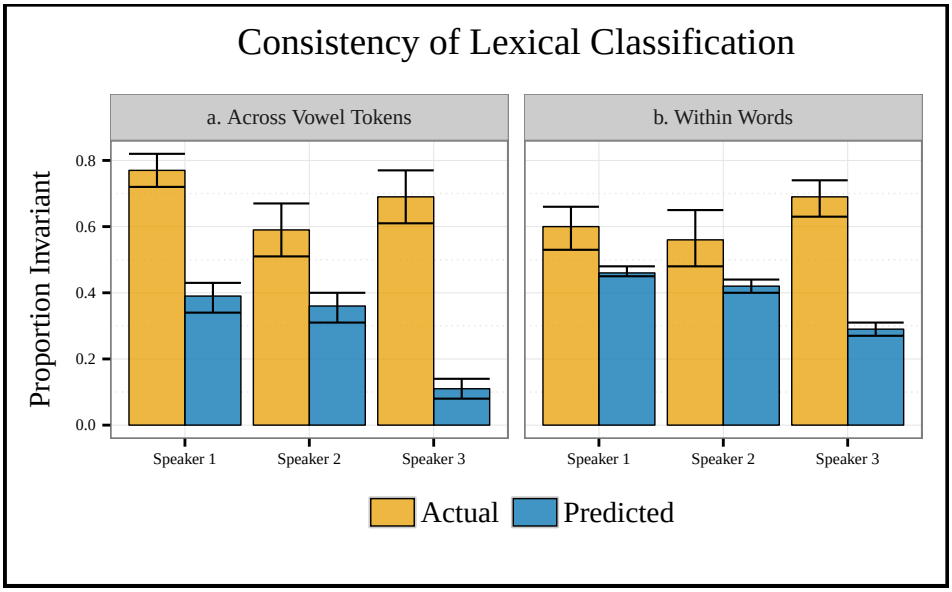

Figure 8: Invariance of classification (a) among vowel tokens for each position of each word, (b) within individual word tokens, and (c) consistency of invariance across tokens of the same word. Error bars represent 95% Confidence Intervals; predicted values represent means of the chance probabilities for each item.

than once was categorised as either *invariant* or *variant* — for example, all 6 instances of the [i] in *biyo* from Speaker 3 were classified as B, so this was categorised as invariant. On the other hand, the initial-syllable [a] in *dabqaad* from Speaker 2 was classified as A for 2 out of 4 tokens, so it was categorised as variant. Baseline frequencies of A and B classes (combined with the number of tokens for each item) was used to calculate the chance probability of invariance. As can be seen in Figure 8a, segements were invariant considerably more frequently than would be expected by chance ($p < 0.001$ for all speakers).

For each word with more than one monopthong, consistency was examined between the vowels in each token. For example, in one token of *aha* from Speaker 1, both vowels were assigned to class A, so it was categorised as invariant. On the other hand, in one token of *culus* from Speaker 2, the first [u] was classified as B while the second was classified as A, so it was categorised as variant. Figure 8b shows that vowels within the same word token were classified consistently more frequently

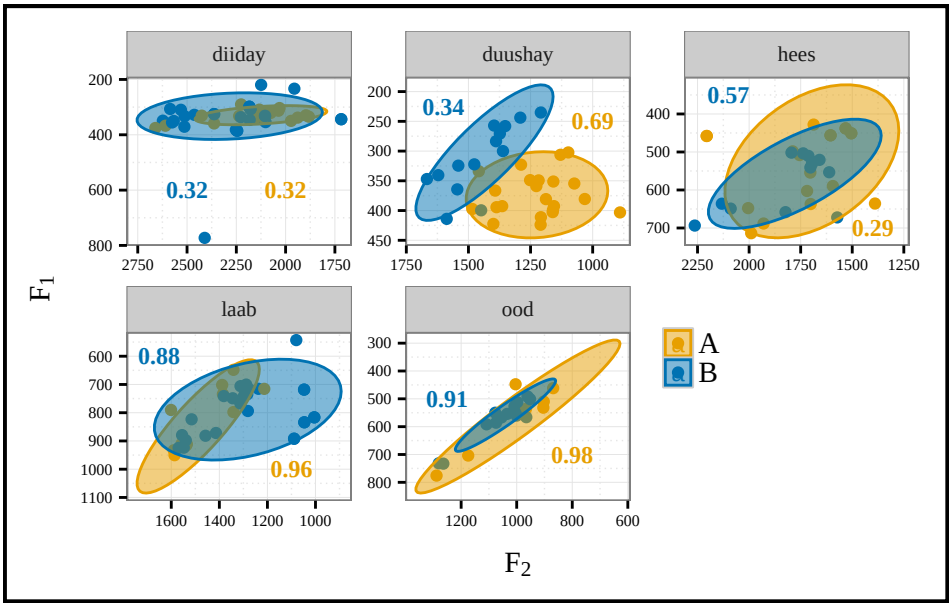

Figure 9: Formant plots for minimal pairs, pooled data for all speakers. Ellipses represent 90% confidence; overlaid numbers represent the proportion tokens for each member of the pair that were classified as A.

than would be expected by chance ($p < 0.001$ for Speakers 1 and 3, $p < 0.01$ for Speaker 2).[3]

Turning to the purported minimal pairs, Figure 9 shows the high degree of acoustic variability of tokens belonging to each member (compared with the differences between members). There was also considerable variation in classification between tokens — none were consistent across all speakers, and no speaker produced any minimal pairs where both members were consistently classified distinctly.

[3]Calculations of chance probability were done under the assumption of independence, which does not entirely hold in this case — vowel-to-vowel coarticulation influences the acoustic dimensions on which classification was based, and would be expected to slightly increase the likelihood of vowels in the same word token sharing the same classification. As such, this result should be viewed with appropriate caution.

	F_1		F_1 Band.		F_2		F_3		Sp. Slope		C. Grav.	
Sp. 1	✔	**M**	✔	S	✔	**L**	✗	XS	✔	**M**	✗	S
Sp. 2	✔	**M**	✔	S	✔	XS	✔	**M**	✔	**L**	✗	**L**
Sp. 3	✔	S	✔	**M**	✗	**L**	✗	**L**	✔	**L**	✔	**L**
Sp. 4	N/A	**M**	N/A	S	N/A	**L**	N/A	**M**	N/A	**L**	N/A	**L**

Table 3: Summary of results of acoustic analysis and classification. Checkmarks represent statistically significant effects, and effect sizes of correlation coefficients from classification are listed alongside.

5 Discussion

The aim of this study was to provide a detailed acoustic description of the feature distinguishing harmony sets in Somali, and to develop a method of classification that can be applied to vowels whose feature specification have not been described. The data presented in the previous section show that there is considerable gradience and variability, but some clear patterns do emerge; a summary of results is presented in Table 3.

The most consistent acoustic correlates of harmony series were F_1, F_1 bandwidth, and spectral slope, which were statistically detectable for all subjects from whom previously classified items were avialable. This is consistent with Edmondson et al. (2004)'s articulatory findings — constriction of the aryepiglottic fold should result in a lowered position of the tongue root, resulting in higher F_1, while the resulting effects on voice quality predict a steeper spectral slope. It is not clear at present whether differences in F_1 bandwith are an independent measure of voice quality or simply a reflection of the effects on F_1, since the two are highly correlated.

However, we find no clear evidence in this data for a categorical phonological distinction. First, there is no detectable departure from unimodality along the relevant acoustic dimensions[4] or along the abstract dimension of the linear discriminant. Additionally, the mean differences between previously-classified series A and series B vowels, while statistically detectable, is fairly small; for F_1 they range from 27.93 Hz for Speaker 1 — which is just barely above the just noticeable difference threshold for F_1 (Kewley-Port, 1995) — to 57.89 Hz for Speaker 2.

The purported minimal pairs fared even worse, with a mean difference of 6.32 Hz for Speaker 1 and 14.98 Hz for Speaker 2, both of which fall

[4]The one exception here is F_1 bandwith for Speaker 3, but as mentioned above this is not necessarily related to vowel series.

below the threshold of perceptibility.[5] There is therefore no evidence from this data that these actually are minimal pairs, at least for these speakers. We have found fewer than a dozen minimal pairs described in the literature; of these, many minimally-distinct roots take obligatory suffixing morphology, and others are uncommon words that were not known to all of our speakers. The remaining pairs show no differences that rise above the threshold of perceptibility.

One finding that does provide a suggestion that vowel series distinctions might possibly be phonologically relevant is the lexical consistency of classification — a given vowel exhibits similarities across different tokens of the lexical item it belongs to, resulting in consistent classification far higher than would be expected by chance. This suggests that there is some lexically-specified property which affects vowels along the relevant acoustic dimensions. This is consistent with a phonological feature specification, but it is also consistent with potential effects of surrounding consonants (in particular pharyngeal and uvular segments, which would have coarticulatory effects consistent with the acoustic differences seen in our data) — further investigation of consonant environment is warranted, but beyond the scope of this paper.

6 Conclusion

In this paper, we have presented pilot data from a small number of native speakers of Somali, investigating the acoustic correlates of the tongue root and/or voice quality feature relevant to vowel harmony in that language. We have found statistically detectable differences along the predicted acoustic dimensions (on the basis of previous articulatory descriptions) but no clear evidence that these differences are categorical or phonological.

It is difficult to draw any broad conclusions with a small number of speakers, particular with respect to a phenomenon that has been described as subject to dialect and individual variation. However, it does seem likely from our data that the categorical distinction between series A and series B vowels has been lost in at least some varieties of Somali. Further research is warranted, with higher numbers of speakers from a broader variety of dialect regions, more controlled and balanced word lists, and a variety of elicitation tasks.

[5] Speaker 3 did not produce a sufficient number of minimal pair tokens.

References

Andrzejewski, B.W. 1955. The problem of vowel representation in the Isaaq dialect of Somali. *Bulletin of the School of Oriental and African Studies* 17:568–580.

Boersma, Paul, and David Weenink. 2008. Praat: Doing phonetics by computer (Version 5.0.17) [Computer program]. Retrieved April 1, 2008 from http://www.praat.org/. Developed at the Institute of Phonetic Sciences, University of Amsterdam.

Edmondson, Jerold A., John H. Esling, and Jimmy G. Harris. 2004. Supraglottal cavity shape, linguistic register, and other phonetic features of Somali. Ms, University of Texas at Arlington and University of Victoria.

Edmondson, Jerold A., and Shaoni Li. 1994. Voice quality and voice quality change in the Bai language of yunnan province. *Linguistics of the Tibeto-Berman Area* 17:49–68.

Edmondson, Jerold A., Cécil M. Padayodi, Zeki Majeed Hassan, and John H. Esling. 2007. The laryngeal articulator: Source and resonator. In *Proceedings of ICPHS XVI*, 2065–2068.

Ghazeli, S. 1977. Back consonants and backing coarticulation in arabic. Doctoral Dissertation, University of Texas at Austin.

Green, Christopher R., Michelle E. Morrison, Nikki B. Adams, and Evan Jones. 2015. A grammar of common somali. Draft, University of Maryland — Center for Advanced Study of Language (CASL), September 2015.

Halle, Morris, and Kenneth Stevens. 1969. On the feature "Advanced Tongue Root". Technical Report 94, MIT. MIT Research Laboratory of Electronics Quarterly Progress Report.

Hess, Susan. 1992. Assimilatory effects in a vowel harmony system: an acoustic analysis of advanced tongue root in Akan. *Journal of Phonetics* 20:475–492.

Kewley-Port, Diane. 1995. Thresholds for formant-frequency discrimination of vowels in consonantal context. *Journal of the A* 97:3139–46.

Kingston, John, Neil A. Macmillan, Laura Walsh Dickey, Rachel Thorburn, and Christine Bartels. 1997. Integrality in the perception of tongue root position and voice quality in vowels. *Journal of the Acoustical Society of America* 101:1696–1709.

Saeed, John. 1993. *Somali reference grammar*. Dunwoody, Kensington, MD.

Starwalt, Coleen. 2008. The accoustic correlates of ATR harmony in seven- and nine-vowel african languages: A phonetic inquiry into phonological structure. Doctoral Dissertation, University of Texas at Arlington.