Chapter 11

Acoustic correlates of harmony classes in Somali

Wendell Kimper
University of Manchester

Wm. G. Bennett
University of Calgary and Rhodes University

Christopher R. Green
Syracuse University

Kristine Yu
University of Massachusetts at Amherst

In this paper, we present 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 find statistically detectable differences along the predicted acoustic dimensions (on the basis of previous articulatory descriptions), and use Linear Discriminant Analysis (LDA) to extend classifications to previously-uncategorized items. However, we find no clear evidence that these differences are categorical or phonological.

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 (purportedly) for an additional feature that has been variously described as FRONT/BACK (Andrzejewski 1955), ±ATR (Saeed 1993), TENSE/LAX (Green et al. forthcoming), 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 contrast are 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. While we do find statistically significant differences between harmony classes along several acoustic dimensions relevant to tongue root and/or voice quality features, we find no clear evidence to support a categorical phonological feature contrast, and instead suggest the possibility of a near merger between previously-distinct vowel categories.

2 Background

The first necessary step towards categorising vowels along the relevant feature dimension is to identify its likely articulatory and acoustic correlates. Andrzejewski 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. (Andrzejewski 1955)

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.

There is overlap between the retracted or backed tongue position characteristic of the Series A vowels and the coarticulatory effects of uvular and pharyn-

Table 1: Minimal pairs (Andrzejewski 1955).

	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'

geal consonants in the language (i.e. [q] and [χ]). Indeed, of the items for which Andrzejewski provides a classification, only Series A items contain uvulars or pharyngeals. For further discussion, see §4.4.

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 exhibits substantially lower airflow than Series B vowels.

(1)   *Properties of Harmony Sets (Edmondson et al. 2004)*
      Set 1 (Series A)
      1. Sphincteric compacting of the arytenoid-epiglottal aperture in the posterior-anterior dimension.
      2. Vowel quality that is more retracted.
      3. Voice quality that is tense.
      Set 2 (Series B)
      1. Expansion of the arytenoid-epiglottal aperture in the anterior-posterior dimension.
      2. Vowel quality that is more fronted and/or raised.
      3. Voice quality that is lax.

Edmondson et al. (2004) note that these findings and previous descriptions are consistent with *register* features, based primarily in voice quality rather than supra-laryngeal articulation. See e.g. Trigo (1991) for further discussion of the relationship between tongue root and register features.

Based on these previous descriptions, the acoustic dimensions under consideration in our study reflect the likely correlates of both register and tongue root features.

Duration and F_0 have been found to be relevant for contrasts involving voice quality (Edmondson & Li 1994; Halle & Stevens 1969), as has spectral slope (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 was 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). F_1 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 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 3 resides in South Africa; all speak some English.

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 & 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 unclassified 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.

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

	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

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 ($\alpha/8$) to adjust for familywise error (corrected p-values are reported). For

---

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

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 1–3. 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. forthcoming) 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.

4.2 Classification

Acoustic analysis of the previously-classified items shows that Series A and Series B items differ detectably along a number of the expected acoustic dimensions (F_1, F_1 bandwidth, F_2, F_3, center of gravity, and spectral slope). But do these differences pattern in a way that might allow listeners (or learners) to map acoustic realizations onto discrete phonological categories? The small effect sizes and lack of detectable departure from unimodality found above provides cause for doubt. In this section, we submit both classified and unclassified forms to cluster analysis, to determine the extent to which observations pattern into discoverable categories.

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

---

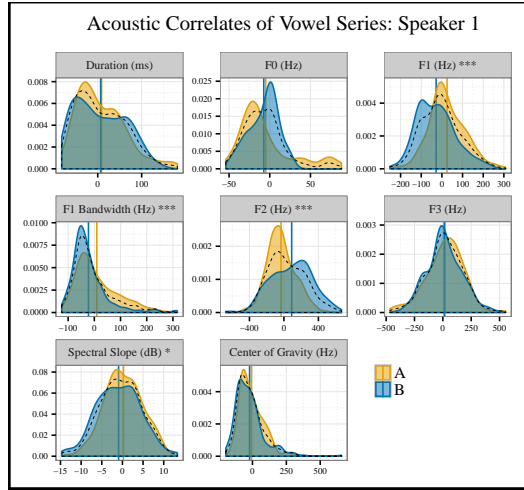[2] For Speaker 2, it was necessary to remove outliers prior to cluster analysis.

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).
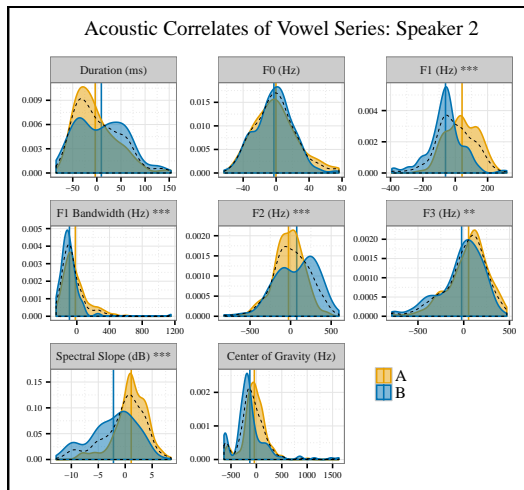

Figure 2: Density plots of Series A and Series B vowels for Speaker 2 (centered measurements, extreme outliers removed). Dashed lines represent combined distributions; vertical lines represent series means; asterisks indicate statistically significant differences (after Bonferroni correction).

Speaker 1, 62% for Speaker 2, and only 54% for Speaker 3.[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.

The acoustic correlates of classes differed considerably between speakers – the only acoustic dimension whose correlation with the discriminant was consistently medium-sized or larger was spectral 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 failed to detect any departure from unimodality.

4.3 Lexical status

The match between the cluster analysis and previous classifications, while fairly poor, was nevertheless above chance for Speakers 1 and 2 (and marginal for Speaker 3, from whom there were fewer observations). This suggests, as with the acoustic analysis, that there is some difference between Series A and Series B vowels that the cluster analysis is sensitive to. However, as before, the unimodality of the linear discriminant casts doubt on the presence of clear categories.

If the distinction between Series A and Series B vowels has *contrastive* status as a phonological feature, it should be lexically specified – we would therefore expect the realization of this feature to be consistent across tokens of an individual lexical item, and those tokens should be assigned to the same category in the classification procedure more often than expected by chance.

Classifications for individual segments were compared across multiple tokens of each lexical item, and all items which appeared more than once were categorized as either *invariant* or *variant* – for example, all 6 instances of the [i] in *biyo*

---

[3]95% Confidence Intervals: 62-70% for Speaker 1, 54-64% for Speaker 2, 49-58% for Speaker 3.

from Speaker 3 were classified as B, so this was categorized 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 and B for the remainder, so it was categorized as variant. Baseline frequencies of A and B classes (combined with the number of tokens for each item) were used to calculate the chance probability of invariance. As can be seen in Figure 4a, segments 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 categorized 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 categorized as variant. Figure 4b shows that vowels within the same word token were classified consistently more frequently than would be expected by chance ($p < 0.001$ for Speakers 1 and 3, $p < 0.01$ for Speaker 2).[4]

Turning to the purported minimal pairs, Figure 5 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.

4.4 Uvular and pharyngeal consonants

Recall from Section 2 that, for lexical items given classifications in Andrzejewski (1955), only Series A words contain uvular or pharyngeal consonants. Could this be a possible source of the effects presented above? If vowels in these words undergo (gradient) coarticulation, we would expect their presence in Series A (but not series B) to result in the kind of small but detectable differences in the acoustic correlates examined. Additionally, because flanking consonants would be held constant among tokens of a single lexical item, we would expect this to result in increased consistency of classification.

The acoustic analysis from Section 4.1 was repeated for all subjects with items containing either uvular or pharyngeal segments removed. The results were by

---

[4]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.
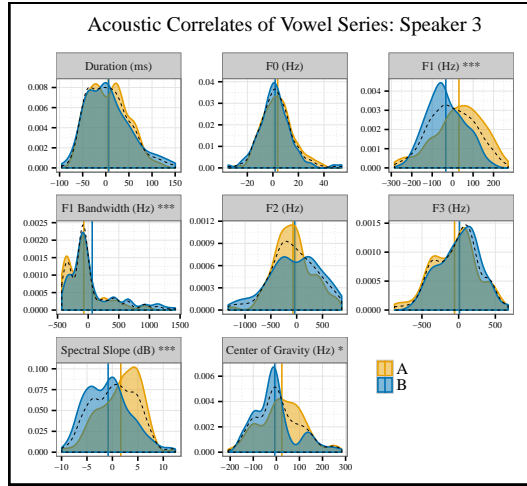
Figure 3: Density plots of Series A and Series B vowels for Speaker 3 (centered measurements, extreme outliers removed). Dashed lines represent combined distributions; vertical lines represent series means; asterisks indicate statistically significant differences (after Bonferroni correction).
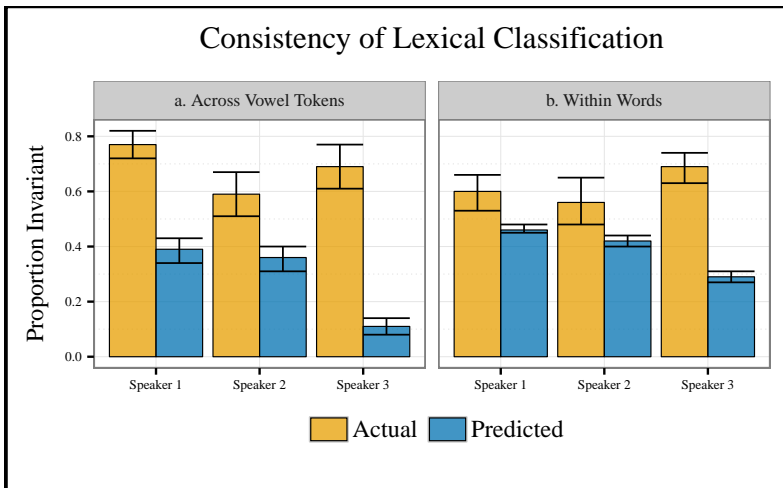

Figure 4: Invariance of classification (a) among vowel tokens for each position of each word, (b) within individual word tokens, and (c) consistency of invariance across tokens of the same word. Error bars represent 95% confidence intervals; predicted values represent means of the chance probabilities for each item.
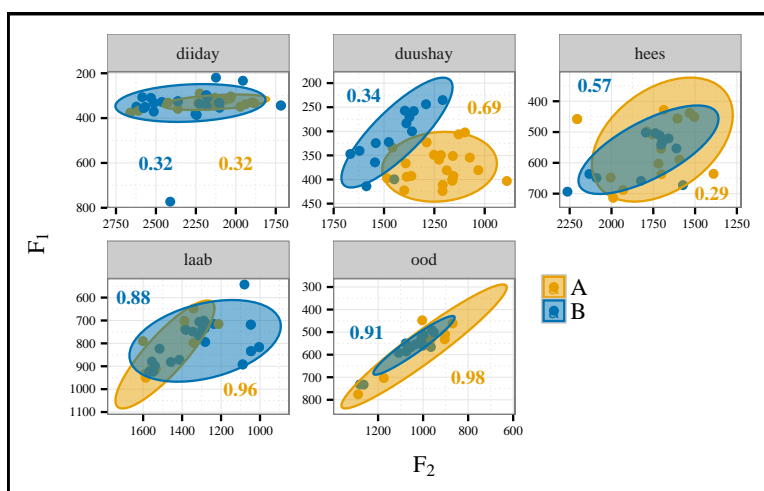
Figure 5: Formant plots for minimal pairs, pooled data for all speakers. Ellipses represent 90% confidence; overlaid numbers represent the proportion tokens for each member of the pair that were classified as A.

and large the same – the effects for spectral slope for Speaker 1 and center of gravity for Speaker 2 fell below the threshold for statistical significance, but the outcomes for all other measures for all three speakers were unchanged. Likewise, the lexical consistency analysis was also repeated with items containing uvular or pharyngeal consonants removed. For Speakers 1 and 2, the effect was retained – classification was invariant across tokens of a single lexical item more often than would be expected by chance. However, for Speaker 2, the lexical consistency effect was not found in the absence of uvulars and pharyngeals.

These results suggest that coarticulatory effects are unable to fully explain either the acoustic difference between Series A and Series B vowels or the consistency of classification across tokens of individual lexical items.

5 Discussion

The aim of this study was to provide a detailed acoustic description of the feature distinguishing harmony sets in Somali, to develop a method of classification that can be applied to vowels whose feature specification has not been described, and to begin to ascertain its phonological status in the language. 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.

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.

	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

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 available. 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[5]. Additionally, the mean differences between previously-classified Series A and Series B vowels, while statistically detectable, are fairly small; for F_1 they range from 27.93Hz for Speaker 1 – which is just barely above the just noticeable difference threshold for F_1 (Kewley-Port 1995) – to 57.89Hz for Speaker 2.

The purported minimal pairs fared even worse, with a mean difference of 6.32Hz for Speaker 1 and 14.98Hz for Speaker 2, both of which fall below the threshold of perceptibility.[6] 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.

---

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

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

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.

The distinction between Series A and Series B vowels in Somali seems, then, to have an intermediate status – neither fully contrastive nor entirely absent. This is consistent with a near merger (Labov et al. 1972), and suggests several avenues for further research. First, data from a larger number of speakers and representing a more carefully balanced sample of lexical items is needed to be certain that the lack of categoricity is not a symptom of noisy data. Additionally, perceptual data is needed to determine whether listeners are able to accurately distinguish minimal pairs.

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, suggesting the possibility of a near merger.

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 is in the process of being 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.

References

Andrzejewski, Bogumił W. 1955. The problem of vowel representation in the Isaaq dialect of Somali. *Journal of the School of Oriental and African Studies* 17. 567–580.

Boersma, Paul & David Weenink. 2008. Praat: Doing phonetics by computer (Version 5.0.17) [Computer program]. http://www.praat.org/, accessed 2008-4-1. Developed at the Institute of Phonetic Sciences, University of Amsterdam.

Edmondson, Jerold A., John H. Esling & 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. & Shaoni Li. 1994. Voice quality and voice quality change in the Bai language of Yunnan Province. *Linguistics of the Tibeto-Berman Area* 17(2). 49–68.

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

Ghazeli, Salem. 1977. *Back consonants and backing coarticulation in Arabic*. University of Texas at Austin. (Doctoral dissertation).

Green, Christopher R., Michelle E. Morrison & Nikki B. Adams. Forthcoming. *A grammar of Common Somali*. Berlin: Mouton de Gruyter.

Halle, Morris & Kenneth N. Stevens. 1969. On the feature "Advanced Tongue Root". *MIT Research Laboratory of Electronics Quarterly Progress Report* 94. 209–215.

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 Acoustical Society of America* 97(5 Pt. 1). 3139–46.

Kingston, John, Neil A. Macmillan, Laura Walsh Dickey, Rachel Thorburn & Christine Bartels. 1997. Integrality in the perception of tongue root position and voice quality in vowels. *Journal of the Acoustical Society of America* 101(3). 1696–1709.

Labov, William, Malcah Yaeger & Richard Steiner. 1972. *A quantitative study of sound change in progress*. Philadelphia: U.S. Regional Survey.

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

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

Trigo, Loren. 1991. On pharynx-larynx interactions. *Phonology* 8. 113–136.