F1 and Center of Gravity Interplay in the Maintenance of Phonological Height within a Statistical Model of a Communal Grammar: The Case of Foodo [ATR] Acoustics

Coleen G. Anderson

Doctoral Candidate, Ph.D. Linguistics
The University of Texas at Arlington

Abstract: This paper presents the notion of an idealized or communal grammar as a statistical model of mode, mean or median as emerging from a representative number of a population, rather than in individual speech, per se. Like other languages with a 9-vowel Cross Height Vowel Harmony (CHVH) system, the most reliable correlate of ATR is F1; [+ATR] vowels have lower F1 values than their [-ATR] counterparts, while F2 differences show considerable variation across speakers. F1, however, fails to maintain phonological height differences as the [+ATR] mid vowels of level 3 overlap in acoustic space with the [-ATR] high vowels of level 2. Center of Gravity (CoG) mean differences prove to be a reliable means of maintaining phonological height in some cases, emerging also as the model for the whole sample.

1.0 Introduction

The mechanics and acoustics of the feature ATR have engendered much study in laboratory phonology during the past three decades. Although the vowel harmony systems of African languages such as Akan (Kwa) have been documented for well over 100 hundred years (viz. Christaller 1875), as Fulop et al. (1998) point out, “previous work … uncovered the operation of vowel harmony in many of these languages … [but] the precise nature of this parameter … eluded investigators” (p. 81). Work in recent decades has added much to our understanding of how this feature is realized phonetically; if the literature has revealed any consistency, it is that the value of the first formant (F1) is the most reliable measure of ATR (Lindau 1978; Hess 1992; Jacobson 1978; Fulop et al. 1998; Guion et al. 2004); [+ATR] vowels have consistently lower F1 values than their [-ATR] counterparts.

In addition, some studies (Lindau 1975, 1989; Hess 1992; Anderson 1999; Casali 2002) have noted acoustic overlap of cross-height pairs /e/ and /o/ in terms of F1 values and
questions arise as to whether these vowels have merged. Hess (1992) points out, however, that for Akan, these pairs differ in articulation, both in tongue height and tongue root position, in such a way that similar F1 values are produced. Furthermore, Akan speakers in Hess’s study are able to for the most part distinguish between these vowels uttered randomly in a CV context. Indeed, native speakers have been known to express surprise when expatriate researchers cannot hear the difference between the pairs (Casali 1997).

The ability of native speakers to perceive differences between acoustically overlapping vowels suggests that they may be listening for something other than F1 differences to distinguish the vowels. One possibility is that speakers are listening for voice quality differences such as those that have been reported for some languages, namely that [+ATR] vowels have a “breathy” or “hollow” sound associated with them, while [-ATR] vowels sound “choked” or “bright” (Berry 1957: Akan; Painter 1971: Anum; Casali 1997: Birifor; Anderson 1999: Ikposo). While such voice quality or timbral differences can be caused by differences in phonation type, this is not the case for Akan. Rather, they are more likely associated with differences in spectral slope between [+ATR] and [-ATR].

Previous studies have shown that [+ATR] and [-ATR] vowels present systematic differences in spectral slope. Two of these studies, Fulop et al. (1998) and Guion et al. (2004), applied a procedure which normalizes the effect of vocal tract resonances on the spectral slope, while Hess (1992) compared the bandwidths of vowels with similar formant structures (namely, [e] and [ʊ ʊ]). The results are similar in these cases. For spectral normalization [-ATR] vowels have higher frequency energy. In the Hess study, [-ATR] vowels have lower F1 amplitudes than the [+ATR] vowels.
In presenting formant and spectral analyses of the 9-vowel system of Foodo (Kwa; Benin), this study confirms that F1 faithfully differentiates vowel harmony sets. ANOVA modeling of the vowels in Foodo’s 5-height system shows heights 1 and 2 and heights 3 and 4 to be statistically significant ($p < 0.05$). F1 fails, however, to differentiate between heights 2 and 3 in three of the four speakers. Center of Gravity (CoG), a measurement of spectral mean (inspired by Kingston et al. 1997), provides the means not only to tease out heights 2 and 3 for two of the speakers, but leads to a statistical model in which [+ATR] vowels are shown to have consistently lower spectral means than [-ATR] vowels for three of four speakers. In the case of the remaining speaker for whom neither the means of F1 nor CoG means for heights 2 and 3 were significant, F1 bandwidth measurements differentiate between the vowel heights: the [+ATR] vowels /e o/ have significantly narrower bandwidths than [-ATR] /i u/.

The results of this study further indicate that while there is variation among individual speakers, these differences are leveled when the entire sample of four speakers is considered together. The sample as a whole produces the predicted model: the mean values of heights 2 and 3 are leveled for F1, but the means for all heights are significant for CoG, with heights 1 and 3 (the [+ATR] vowels) having lower spectral means than heights 2 and 4, (the [-ATR] vowels). Such leveling of variation to fit the predicted model suggests the possibility of an idealized or communal grammar, one which emerges from a representative number of the population rather than in individual speech, necessarily. This paper examines the notion of a communal grammar as representing a statistical mode, mean, or median. As will be seen, the results of this study suggest that either a mean or a median grammar exists for Foodo vowel production.

2.0 Brief (Non-exhaustive) History of the Articulatory and Acoustic Studies of ATR
As noted above, the exact phonetic parameters of the vowels of the African languages with cross-height vowel harmony eluded early researchers. Nonetheless, as early as 1947, Pike (ctd. in Stewart 1967), suggested that one set of the vowels was produced by a combination of lowered larynx, tongue root advancement, and spreading of the faucal pillars. Stewart (1967) captured this observation proposing a new feature that has generally been accepted as [ATR], engendering a series of studies which have shed light on the articulation and acoustics of the vowels of languages with ATR harmony.

2.1 Articulatory Studies

In the decades following Pike’s initial account, articulatory studies, notably X-ray (Lindau 1975, 1978, 1979) and MRI studies (Tiede 1996), substantiated early observations for West African vowel harmony systems, while other investigations, notably Jacobson (1978), suggested that there may be more than one way of articulating the differences between the harmony sets across the continent.

Lindau’s analysis of three West African languages (Akan, Ijo and Igbo) suggested that within the oral cavity there is no apparent difference in tongue height for the harmonic pairs such as [i] ~ [ɪ]. Rather, the greater pharyngeal volume of the [+ATR] vowels is accompanied with downward vertical displacement of the larynx. The MRI conducted by Tiede on one speaker of Akan corroborates Lindau’s earlier (cine)radiography study by showing pharyngeal volume differences between [+ATR] and [-ATR] vowels of the same height: [+ATR] vowels have larger pharyngeal volumes than their [-ATR] counterparts. Tiede’s findings indicate that the pharyngeal volume increase for [+ATR] vowels is produced in a transverse direction, accompanied by expansion in the sagittal direction (from the advancement of the tongue root) and by lowering of the larynx. The converse was found to be true for the [-ATR] vowels:
pharyngeal constriction in both sagittal and transverse directions, coupled with a raising of the larynx.

However, unlike Lindau and Tiede, Jackobson (1978) was unable to substantiate any consistent downward vertical displacement of the larynx for the [+ATR] vowels of DhoLuo (Western Nilotic); radiographic tracings reveal that either tongue height or pharyngeal expansion may be used to distinguish the vowel sets.

Most recently, laryngoscopic studies conducted by Edmondson and Esling on Akan and Kabiye (Gur) have shed further light on the mechanism involved in the production of the ATR harmony sets in at least some West African languages. Their work supports Tiede’s findings for Akan, while revealing that a laryngeal stricture involving the aryepiglottic folds (valve 3 in Figure 1 below) most likely affects the expansion or constriction of pharyngeal volume (Edmondson & Esling 2006).

*Figure 1: Laryngoscopic View of the Glottis*

What the mechanics of this laryngeal stricture may be like in some languages with [ATR] harmony can be seen in Figures 2a,b, in which four slides from two harmonic pairs of the Esling et al. study of Kabiye are presented; [+ATR] vowels are on the left with their [-ATR]
F1 and Center of Gravity Interplay: The Case of Foodo [ATR] Acoustics

counterparts on the right. In the [-ATR] slides, valve 3, involving the aryepiglottic folds and the epiglottis, is clearly implicated.

Figure 2a: Laryngoscopic View of Kabiye [+ATR] /u/ and [-ATR] /ʊ/

![Figure 2a](image)

Figure 2b: Laryngoscopic View of Kabiye [+ATR] /e/ and [-ATR] /ɛ/

![Figure 2b](image)

2.2 Acoustic Studies

In addition to the study of the articulation of [ATR], certain impressionistic “voice qualities” associated with the vowel harmony sets have often been reported. As was noted earlier, [-ATR] vowels are perceived as “creaky” or “choked” while [+ATR] vowels appear “breathy” or “hollow”. Berry (1957) reports such impressions concerning Akan; these
differences have also been noted for Ikpősọ (Kwa) (Anderson 1999). Jacobson (1978: 2) also sites several works which refer to such voice quality differences for the Nilotic languages.

It was also noted earlier that coupled with the voice quality differences between the vowel harmony sets, there has also been a tendency for [-ATR] high vowels /i/ and /u/ and [+ATR] mid vowels /e/ and /o/ to overlap in acoustic space (Hess 1992; Casali 2002, among others).¹

In the search for an acoustic correlate of [ATR] that might account for the perceptual differences of acoustically overlapping vowels, researchers have offered several accounts. Hess (1992), for example, argues that the relevant acoustic correlate is the bandwidth of F1: [+ATR] mid vowels [e o] are shown to have statistically narrower bandwidths than [-ATR] high vowels [i u] for one speaker of Akan. Still, others such as Kingston et al. (1997) suggest that a measurement of spectral “flatness,” the difference (in dB) between the first and fourth (H1-H4) harmonics, might tease out the differences between the harmonic sets, which bandwidth measurements are unable to do. Similarly, Fulup et al. (1998) compared the advanced and retracted vowels of Degema (Edoid) with some success (reliable in only two of the five vowel pairs) using a normalized measure of relative formant intensity. Guion et al. (2004) show similar results for Maa (Nilo-Saharan). [+ATR] vowels display relatively less energy in the higher frequency regions than their [-ATR] counterparts. Finally, the recent research of Edmondson et al. (draft) pursues an acoustic measurement of spectral mean, “Center of Gravity,” typically used to measure the concentration of the frequencies produced by the constriction of fricatives, with promising results for Akan and Kabiye.

¹ See Przedziecki (2005) for an example of a language variant with 9 surface vowels (7 phonemic) in which [+ATR] mid vowels and [-ATR] allophones of high vowels do not overlap in acoustic space.
As noted above, laryngoscopic video recording of the speech of one Akan and one Kabiye speaker revealed a constriction at the aryepiglottic folds, what Edmondson and Esling (2006) call “valve 3.” One of the acoustic (and thus perceptual) consequences of such a stricture at the larynx is that valve 3 may likely be responsible for the “choked” sound that has been reported for the [-ATR] vowels. Edmondson (personal correspondence) uses two analogies to describe the effect of this valve setting. One is that of a tube. For the [-ATR] vowels, it is helpful to imagine the end of the tube having been pinched. Air passing through the pinched end of the tube would likely create some degree of turbulence. Such turbulence may be perceptually interpreted as “choked” sounding. The other analogy he uses is that of a brass instrument. The vocal folds, the source of sound, would compare to a player’s vibrating lips and the vocal tract (the acoustic filter) to the body of the instrument. The laryngeal setting, again valve 3, acts as a mouthpiece. Change the size of the mouth piece of the instrument and a different timbre results.

Edmondson suggests that this constriction is measurable. For such a measurement we may consider “Center of Gravity,” a measure of spectral mean that has been used successfully in the study of fricatives (viz. Gordon et al. 2002 and studies cited therein). The relative intensity of the constriction should affect the strength of the frequencies of higher harmonics, [-ATR] vowels predicted to have overall higher mean frequencies than their [+ATR] counterparts.

3.0 Foodo Vowel System

In Foodo, a Guang language of the Kwa family spoken in Northern Benin, the vowel system consists of nine vowels /i ei a o u/, which participate in CHVH (see below). Plunkett (1991), in his thesis on Foodo tone, briefly notes that all vowels within the stem must agree in ATR value and that prefixes agree with the stem in terms of ATR. In the following examples,
the stem vowels are bolded while the prefix vowel is italicized and bolded. Only vowels within
the phonological word agree in harmony.

(1) a. o-sóóli haI1
   3s.C.A.2-gather Cl.M.
   ‘he gathered’

b. ṣ-sóólI haI1
   3s.C.A.-stretch Cl.M.
   ‘he stretched’

As noted above, Foodo is a 9-vowel language with Cross Height Vowel Harmony
(CHVH). Its harmony is called “cross height” because the process involves crossing height
barriers, as seen in Figure 3 below. In the 5-height CHVH system, vowels within the domain of
harmony agree in [ATR] regardless of their phonological height.3

\begin{figure}
\centering
\begin{tabular}{|c|c|c|}
\hline
Height & Vowel & ATR \\
\hline
1 & i & + \\
\hline
2 & u & - \\
\hline
3 & e & + \\
\hline
4 & o & - \\
\hline
5 & a & - \\
\hline
\end{tabular}
\caption{5-Height CHVH System}
\end{figure}

The 5-height system contrasts with 4-height (or 7-vowel) harmony systems in which
either the mid-vowels or the high-vowels must harmonize, depending on which height is
phonologically missing from the system. In the 4-height (M) System, where mid vowels must
harmonize for ATR, phonological height 2 vowels are absent from the system, though they may

\footnote{2 The following grammatical abbreviations are used: C.A. – Completed Aspect, Cl.M. – Clause Marker}

\footnote{3 H,M,L are abbreviations for High, Mid and Low vowels, respectively.}
occur allophonically in some languages, such as certain dialects of Yoruba (see Przezdzieki 2005).

In contrast, in the 4-height (H) System, high vowels harmonize for ATR and phonological height 3 vowels are missing from the system. Again, in some languages, allophonic height 3 vowels, [e] and [o] specifically, may occur. Kinande (Bantu, Democratic Republic of Congo) (viz. Gick et al. 2006 and authors cited therein) is one such language. Casali (2003) also cites Talinga-Bwisi (Bantu, Uganda) as another.

That Foodo is considered a 5-height system, and not a 4-height system, will become pertinent later when the acoustic facts are presented. As will be seen, the mean values of F1 alone are not sufficient to maintain the integrity of five phonological heights.

4.0 Methodology

The data analyzed in this study are a part of a larger corpus recorded in November 2003 in the village of Sèmèrè in Northern Benin. The recordings were made with a Shure dynamic microphone and Sony laptop computer, and digitized with CoolEdit2000. Five speakers participated in the recordings; data from four are considered here, two males and two females. Speakers range in age from 30 to 52. The data set consists of 18 words: two different words for each of the nine vowels, one with phonemic High Tone and the other with phonemic Low Tone. Eleven words have monosyllabic roots, the remainder di-syllabic ones. Each word was repeated 10 times in the carrier phrase “o/-ɔ- _____ hal” ‘he/she ______-ed’ for a total of 20 tokens for each target vowel. All words chosen for analysis have phonemically long vowels in order to avoid the centralization known to occur with short front vowels in most other North Guang languages (Casali 1997). The target vowel is always in the first syllable of the di-syllabic roots.
With the aid of the PRAAT software program, five measurements were taken: fundamental frequency (F0), frequency values for both F1 and F2, Center of Gravity (or spectral mean), and F1 bandwidth. For formant analysis, approximately 10 pulses were selected for measurement from the steady state of the vowel. For Center of Gravity measurements, approximately the center 2/3rd of the vowel was chosen to eliminate transitional effects seen in the harmonics of their narrow band spectrograms. The bandwidth and fundamental frequency measurements were taken from the center of the steady state of the vowel.

5.0 Results

In this section, the results of the formant, Center of Gravity and bandwidth analyses are presented. The results of the fundamental frequency analysis are pertinent for only speaker 3.

5.1 Formant Analysis

This section presents the results of the statistical tests (ANOVA) ran to determine whether ATR vowel pairs differ in formant values for F1 and F2.

Figures 4 – 7 below show the F1 vs. F2 formant plots for the four Foodo speakers. As might be expected, visual inspection of the formant plots reveals no apparent overlap for F1 for any of the ATR harmonic pairs. On the contrary, for each speaker there is at least one harmonic pair, namely [o]–[ɛ], which appear to overlap for F2. All four harmonic pairs seem to overlap for speaker 4, while back vowel pairs overlap for speakers 1 and 3.

Contrary to the apparent separation of ATR harmonic pairs by F1, for each of the speakers, the formant charts reveal acoustic overlap or neutralization of heights 2 and 3 vowels, [-ATR] high vowels [i u] and [+ATR] mid vowels [e o].

4 Ellipses are for illustrative purposes only and not to indicate standard deviation.
The chart in Table 1, below, displays the range of F1 averages for the entire sample (males and females). The lower of the two values in any row is the lowest average of one of the males while the higher value is that of one of the females. Shaded here in yellow diagonal slashes, the same apparent overlap of the mean values of height 2 [-ATR] high vowels and height 3 [+ATR] mid vowels seen in the individual formant plots is also evident for the whole sample. A summary chart of the averages of the formants for each vowel for all speakers may be found in the Appendix.

Figure 4: F1 vs. F2 Plot for Speaker 1, Male

---

5 Such differences are generally due to the difference of size of the vocal tract between males and females. Males tend to have larger (i.e. longer) vocal tracts, leading to lower F1 values (among other things).
Figure 5: F1 vs. F2 Plot for Speaker 2, Male

Figure 6: F1 vs. F2 Plot for Speaker 3, Female
The F1 values of all the vowel pairs were submitted to a one-way ANOVA with F1 as the dependent variable and vowel quality as the independent factor. Tukey’s post hoc and descriptive statistics were run simultaneously. The level of significance was set at 0.05 for this and subsequent sections. Results confirm that for every speaker, [+ATR] vowels have lower F1 mean values than their [-ATR] counterparts ($p < 0.001$). The results also indicate that for three
of the speakers (1-3), there is no significant difference in F1 mean averages for non-harmonic vowel pairs, [+ATR] mid vowels [e o] and [-ATR] high vowels [i u]. For speaker 4, however, the apparent overlap of these vowels is superficial only. The $p$ values for the means of these vowel pairs for each speaker are found in Table 2.

**Table 2: $p$ Values for F1 Mean Averages of [+ATR] Mid Vowels and [-ATR] High Vowels**

<table>
<thead>
<tr>
<th>Vowel Pair</th>
<th>Speaker 1 (M)</th>
<th>Speaker 2 (M)</th>
<th>Speaker 3 (F)</th>
<th>Speaker 4 (F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[i]/[e]</td>
<td>1</td>
<td>0.928</td>
<td>0.596</td>
<td>0.001</td>
</tr>
<tr>
<td>[u]/[o]</td>
<td>0.274</td>
<td>1</td>
<td>0.981</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

The F2 values for all vowel pairs were also submitted to a one-way ANOVA with F2 as the dependent variable and vowel quality as the independent factor. Table 3 summarizes the $p$ values for the ATR harmonic vowel pairs. The results indicate that differences of F2 mean between ATR harmonic vowel pairs are significant in most cases for speakers 1-3 and insignificant in all cases for speaker 4. With the exception of the high back vowels of speaker 3, all of the statistical results confirm the apparent overlap seen in the formant plots above.

**Table 3: $p$ Values for F2 Mean Averages of ATR Harmonic Vowel Pairs**

<table>
<thead>
<tr>
<th>Vowel Pair</th>
<th>Speaker 1 (M)</th>
<th>Speaker 2 (M)</th>
<th>Speaker 3 (F)</th>
<th>Speaker 4 (F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[i]/[ɪ]</td>
<td>0.004</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>0.529</td>
</tr>
<tr>
<td>[e]/[ɛ]</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>1</td>
</tr>
<tr>
<td>[o]/[ɔ]</td>
<td>0.997</td>
<td>0.944</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>[u]/[u]</td>
<td>0.396</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>0.982</td>
</tr>
</tbody>
</table>

In addition, the average mean values of F2 are also significant in distinguishing non-harmonic pairs, [ɪ e] and [ʊ o] ($p < 0.001$).

**5.2 Center of Gravity (CoG) Analysis**

In order to test the integrity of Foodo as a 5-height system, the F1 mean values were submitted to another one-way ANOVA with Vowel Height as the factor. This ANOVA also
included Center of Gravity (CoG) as a dependent variable. In keeping with the results of the ANOVA ran for F1 and vowel quality, it was anticipated that the mean values of F1 for vowel heights 2 and 3 would not be significant. On the contrary, it was expected that those for CoG would be.

5.2.1 F1 and Center of Gravity Mean ANOVAs

Summaries of the mean values (in hertz rounded to the nearest hertz) of F1 and CoG for each of the five phonological heights in Foodo are found in Tables 4 and 5. Shaded areas represent heights for which the differences in mean values for F1 or CoG are not significant. In Table 4, for example, we see that the mean differences of F1 for speakers 1-3 as well as the pooled results are not significant and thus fail to distinguish between heights 2 and 3. For speaker 4, however, the mean differences of F1 between vowel heights are significant.

Table 4: ANOVA of F1 Means with Height Factor

<table>
<thead>
<tr>
<th>Height</th>
<th>ATR</th>
<th>Speaker 1 (M)</th>
<th>Speaker 2 (M)</th>
<th>Speaker 3 (F)</th>
<th>Speaker 4 (F)</th>
<th>Pooled</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>[+ATR]</td>
<td>352</td>
<td>335</td>
<td>326</td>
<td>363</td>
<td>344</td>
</tr>
<tr>
<td>2</td>
<td>[-ATR]</td>
<td>480</td>
<td>430</td>
<td>511</td>
<td>555</td>
<td>494</td>
</tr>
<tr>
<td>3</td>
<td>[+ATR]</td>
<td>460</td>
<td>416</td>
<td>518</td>
<td>606</td>
<td>495</td>
</tr>
<tr>
<td>4</td>
<td>[-ATR]</td>
<td>683</td>
<td>592</td>
<td>682</td>
<td>777</td>
<td>684</td>
</tr>
<tr>
<td>5</td>
<td>[-ATR]</td>
<td>897</td>
<td>825</td>
<td>904</td>
<td>949</td>
<td>894</td>
</tr>
</tbody>
</table>

Table 5: ANOVA of Center of Gravity Means with Height Factor

<table>
<thead>
<tr>
<th>Height</th>
<th>ATR</th>
<th>Speaker 1 (M)</th>
<th>Speaker 2 (M)</th>
<th>Speaker 3 (F)</th>
<th>Speaker 4 (F)</th>
<th>Pooled</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>[+ATR]</td>
<td>448</td>
<td>430</td>
<td>512</td>
<td>488</td>
<td>469</td>
</tr>
<tr>
<td>2</td>
<td>[-ATR]</td>
<td>652</td>
<td>555</td>
<td>693</td>
<td>770</td>
<td>667</td>
</tr>
<tr>
<td>3</td>
<td>[+ATR]</td>
<td>556</td>
<td>448</td>
<td>610</td>
<td>673</td>
<td>572</td>
</tr>
<tr>
<td>4</td>
<td>[-ATR]</td>
<td>843</td>
<td>681</td>
<td>834</td>
<td>943</td>
<td>824</td>
</tr>
<tr>
<td>5</td>
<td>[-ATR]</td>
<td>1015</td>
<td>960</td>
<td>1200</td>
<td>1094</td>
<td>1067</td>
</tr>
</tbody>
</table>

The harmonic mean sample size for these data and the other individual results presented below is 34.386. The group sizes are unequal and therefore the harmonic mean of the group sizes is used. The statistical tool does not guarantee type 1 error levels.
Table 5 indicates that the mean values of CoG for heights 1-5 are significant for speakers 1 and 4 and the pooled data. In the case of speaker 2, the differences in mean value for height 1 and 3 are not significant but those for heights 2, 4, and 5 are, whereas for speaker 3 the difference in mean value for height 3, highlighted in green, is neither significant from that of height 1 (yellow) nor height 2 (blue). The mean values of heights 1, 2, 4, and 5, however, are significant from each other and levels 1, 4 and 5 from that of height 3, as well. The difference in mean values for the CoG of the height 2 and 3 vowels is significant for two of the three speakers (speakers 1 & 2) whose F1 means for these heights were not significant. This would indicate that CoG plays a role in disambiguating acoustically overlapping vowels for these speakers. This is also so for the pooled data. For speakers 2 and 3, the mean averages of CoG for heights 1 and 3 are not significant. In one sense, since these heights are comprised of vowels of the same ATR value, the lack of significance may not be important. This is difficult to evaluate, however, without some sort of normalizing procedure with which to compare these values. This is most true in the case of speaker 2 whose mean values for these two heights differ by only 18 Hz. Nonetheless, see discussion of Table 6 below.

Of more importance is the insignificance of the CoG mean values for speaker 3’s heights 2 and 3, as these two heights are also neutralized for F1. It had been noted in casual observation that CoG values for this speaker’s low tone vowels tended to be lower than for high tone vowels. While it is unlikely that fundamental frequency differences in and of themselves would affect spectral means, unless those differences are accompanied by phonation type differences, something that cannot as yet be attested for Foodo, it was deemed prudent to separate the tokens by tone class to see how the results differed from the combined classes.
The tokens split by tone class produced differing results. The ANOVA for high tone tokens indicated the same insignificance of F1 mean values for height 2 and 3 vowels as in the combined tone model. But the ANOVA evaluating the CoG mean values showed those differences to be significant. The low tone tokens, on the other hand, generated different results. Like speaker 4, the mean values of F1 for each height are significant. But the CoG mean values were not significant for heights 2 and 3. Given the significance of CoG mean values for heights 2 and 3 for high tone tokens, it may be concluded that there is a tendency for the means of CoG to be significant for height 2 and 3. As will be seen below, differences in bandwidth shed further light on this particular speaker’s behavior.

Finally, in the case of speaker 4, where both the mean values of F1 and CoG are significant for distinguishing all the heights, the spectral mean is higher for height 2 vowels than for height 3 vowels. This is true for both speakers 1 and 2 and the pooled data, and seen also in the high tone tokens for speaker 3. This is what would be expected if a constriction at valve 3 were present in the production of the [-ATR] vowels. It is interesting to note that the spectral mean for the [+ATR] mid vowels for speaker 4 is less than 70 Hz higher than the corresponding F1 averages. For the [-ATR] high vowels, on the other hand, which have lower F1 mean values than the [+ATR] mid vowels, the spectral mean values are well over 200 Hz higher than their corresponding F1 averages. Table 6 summarizes the distance, as it were, between F1 means and CoG means for the five heights for all the speakers.

---

7 Whether Foodo [-ATR] vowels are produced with a constricted valve 3 or not is hypothetical. Most likely only direct observation via laryngoscopy would confirm this.
Table 6: “Spectral Distance” in Hz of the Spectral Mean from F1 Mean

<table>
<thead>
<tr>
<th>Height</th>
<th>ATR</th>
<th>Speaker 1 (M)</th>
<th>Speaker 2 (M)</th>
<th>Speaker 3 (F)</th>
<th>Speaker 4 (F)</th>
<th>Pooled</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>[+ATR]</td>
<td>96</td>
<td>95</td>
<td>186</td>
<td>125</td>
<td>125</td>
</tr>
<tr>
<td>2</td>
<td>[-ATR]</td>
<td>172</td>
<td>125</td>
<td>182</td>
<td>215</td>
<td>173</td>
</tr>
<tr>
<td>3</td>
<td>[+ATR]</td>
<td>96</td>
<td>32</td>
<td>92</td>
<td>67</td>
<td>77</td>
</tr>
<tr>
<td>4</td>
<td>[-ATR]</td>
<td>160</td>
<td>89</td>
<td>152</td>
<td>166</td>
<td>140</td>
</tr>
<tr>
<td>5</td>
<td>[-ATR]</td>
<td>118</td>
<td>135</td>
<td>296</td>
<td>145</td>
<td>173</td>
</tr>
</tbody>
</table>

Note that while these numbers cannot be evaluated for their significance at this point, they can give us an overall sense of the spectral slope for each of the vowel heights. Lower numbers indicate a CoG relatively close to F1. In general, we see that [+ATR] vowels have average lower center of gravities than [-ATR] vowels. With the exception of speaker 3, the [+ATR] vowels always have average lower center of gravities than their [-ATR] counterparts. Speaker 3 has the lowest F1 averages and highest F2 averages for /i/; it is possible that contributes to the higher Center of Gravity for height 1.

5.3 Bandwidth Results

The results of F1 bandwidth analysis is found in Table 7. The mean values (in dB) are rounded to the nearest one-tenth of a dB. Shading again represents the lack of significance of the mean values in dB for those particular heights.

Table 7: ANOVA of F1 Bandwidth Means in dB with Height Factor

<table>
<thead>
<tr>
<th>Height</th>
<th>ATR</th>
<th>Speaker 1 (M)</th>
<th>Speaker 2 (M)</th>
<th>Speaker 3 (F)</th>
<th>Speaker 4 (F)</th>
<th>Pooled</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>[+ATR]</td>
<td>85.0</td>
<td>186.0</td>
<td>224.6</td>
<td>176.6</td>
<td>168.0</td>
</tr>
<tr>
<td>2</td>
<td>[-ATR]</td>
<td>85.9</td>
<td>83.3</td>
<td>134.3</td>
<td>122.1</td>
<td>106.4</td>
</tr>
<tr>
<td>3</td>
<td>[+ATR]</td>
<td>52.5</td>
<td>27.8</td>
<td>41.0</td>
<td>29.3</td>
<td>37.6</td>
</tr>
<tr>
<td>4</td>
<td>[-ATR]</td>
<td>76.8</td>
<td>58.7</td>
<td>64.0</td>
<td>157.1</td>
<td>89.1</td>
</tr>
<tr>
<td>5</td>
<td>[-ATR]</td>
<td>89.3</td>
<td>83.7</td>
<td>120.4</td>
<td>112.8</td>
<td>101.6</td>
</tr>
</tbody>
</table>

Even at a quick glance at this table, it is obvious that there is considerable overlap between the mean values of various heights. All speakers as well as the pooled data have heights for which the mean bandwidth values are not significant in comparison to one or more other
heights. However, for all speakers and the pooled data, the mean averages of the bandwidths of
heights 2 and 3 are significant. This includes especially speaker 3 for whom F1 means were not
significant between heights 2 and 3 and yet Center of Gravity only marginally distinguished
these heights. These results are quite in keeping with the bandwidth results in Hess (1992) for
one speaker of Akan in which [+ATR] mid vowels had significantly narrower bandwidths than [-
ATR] high vowels. In the case of the Akan speaker, as well as with speakers 1-3 and the pooled
data, [+ATR] mid vowels and [-ATR] high vowels overlap significantly for F1.

6.0 Summary and Discussion

The results of each of the statistics run for F1, F2, CoG and F1 bandwidth means reveal a
part of the story of the acoustics (that is phonetic realization) of Foodo vowel phonology. Within
this story we have seen two major plots: the main acoustic correlate distinguishing vowel
harmony pairs /i/ /e/, /ε/, /o/ o, and /u/ u/ and phonological vowel height. F1 is the main
character in differentiating harmonic pairs. As was seen consistent across speakers, F1 mean
values are significantly lower for the [+ATR] member of the pairs.8 F2, however, is a minor
character in distinguishing harmonic pairs. In all cases, F2 mean values were not significant for
back mid vowels, and for one of the speakers, they were insignificant for all four vowel pairs.
Speaker variation is a major theme for F2.

One of the consequences of the role of F1 in distinguishing the ATR harmonic pairs is its
impact on phonological height. While optimizing acoustic space for differentiating harmonic
pairs, the distinction between vowel heights 2 and 3 is leveled for 3 of the 4 speakers.

8 Such is how the literature has described the statistical significance between [+ATR] vowels and their [-ATR]
counterparts. It has been shown that [+ATR] vowels are often accompanied with a lowered larynx. Put another way,
F1 mean values are significantly higher for the [-ATR] member of the pairs, which is a predictable outcome of
aryepiglottal sphinctering, if the hypothesized constriction at valve 3 produces a shortened vocal tract.
Phonological height cannot be reliably maintained by F1 alone. While mean averages of F2 are significant for vowel qualities (i.e. for vowels such as /e/ and /o/, which overlap for F1), it cannot come to the aid of distinguishing phonological height in a statistical model. The average F2 values of front and back vowels all neutralize around 1400 – 1500 Hz along the height axis.

In sum, some other acoustic correlate other than F1 or F2 is necessary to maintain the integrity of a phonological 5-height vowel system such as that of Foodo. And overall, though there is variation among speakers, the results for Center of Gravity (CoG) are very promising in this regards. With the exception of speaker 3, CoG aids in the differentiation of heights 2 and 3, which otherwise are neutralized by the overlap in F1 mean values. As suggested in Kingston et al. (1997), speakers are most likely in tune with the extremes of spectral flatness where perceptual differences are more salient. The average means of the spectral slopes of the [+ATR] vowels of height 2 are lower (thus a steeper overall drop of energy from lower to higher harmonics) than the means of the slopes of the [-ATR] vowels of height 3. The lack of significance of these means for combined high and low tone tokens of speaker 3’s heights 2 and 3 vowels raises the question of whether the discipline accepted $\alpha = 0.05$ significance standard is adequate for measurements of spectral flatness. Presumably, if other speakers have no trouble perceiving the difference between speaker 3’s height 2 and height 3 vowels, then there must be something in the acoustic signal they perceive as significant. To what degree production variation among speakers is tolerated as accepted variation and not “speech error” feeds into the results of the sample as a whole and what these suggest in terms of the notion of a communal or idealized grammar.

---

9 A perceptual study would need to be designed to test this hypothesis.
The results of the F1 and Center of Gravity ANOVAs for the pooled data seen in Tables 4 and 5 of this sample of the population appear to support the notion of a communal or idealized grammar governing an entire population. In this case, the assumed idealized grammar for Foodo vowels is one in which F1 is sufficient for distinguishing between vowel harmony pairs but not necessary so for distinguishing vowel heights. Center of Gravity disambiguates vowel heights without compromising existing vowel height categories. Such a grammar begs the question: Just how large of a sample will produce the idealized grammar?

To consider this question, the communal grammar may be understood in one of three ways: as representing some sort of statistical mode, mean, or median. In a model of statistical mode, the grammar would be held by the largest number of people, regardless of how many that number may be. For a community of 10 with 9 grammars, for example, the one grammar held in common by the two who share it would become the grammar. The others would be variations from the “modal” grammar. A model of statistical mean, however, would be an “averaging” of behaviors. While no one person in the community actually possesses the mean grammar, the mean grammar is an idealized central tendency. The median grammar, however, would be one that is “most central” in the community. Being the furthest away from the extreme grammars, it would best approximate the largest number of grammars within the community. It bears similarity to the mean grammar, but with one difference. It is the actual grammar of at least one individual within the community. Finally, a communal grammar would need to tolerate a range of variation between individuals while maintaining perceptual integrity. It would allow for vocal

---

10 Credit for the notion of a communal grammar must be given to David Silva from whose observations (yet to be published) on the differences between individual and communal language behavior ideas emerge.
tract differences between individuals or sexes as well as other factors that may be contributing to
disambiguating failed fits.

The results of the sample of the Foodo population available thus far do not support the
possibility of a statistical mode. In order to have a mode grammar, more than one person in the
population must share the grammar. Our sample of 4 is most likely too small to have generated
all the possible variations of the idealized grammar. A larger sample could quite possibly
produce a statistical mode.

The Foodo data do, however, support either a statistical mean or median grammar. If the
model generated by the total population represents an averaging of behaviors, and the mean
grammar being an idealized central tendency, as stated above, then speaker 1, whose results
come the closest to this mean grammar, support such a notion. This sample of 4 is sufficient to
produce a mean grammar with one of the individuals approximating very closely the ideal
grammar. The results also support a median grammar, with the “most central” grammar being
the furthest away from extreme grammars. Speaker 1 falls into this category again as being
center set. Speakers 3 and 4 (the two women) fall on either side of the centralized grammar. In
the results for speaker 3, neither F1 nor spectral mean consistently distinguished heights 2 and 3,
while they both did so for speaker 4.

Data from more speakers would need to be analyzed to judge which of these statistical
models best represents the Foodo population at large. Nonetheless, it is interesting to note that
when the data were reanalyzed, separating the males from the females, the same grammar was
generated, adding perhaps more weight to a statistical mean grammar. The model thus appears
to allow for vocal tract differences between the sexes, even of a very small sample of the
population at large.
Finally, to what degree does the model accommodate failed fits? It is not readily apparent how such a question may be evaluated. However, in an attempt to see how differing the sample members might affect the communal grammar, the data were reanalyzed once more excluding speaker 4 for whom both F1 and CoG means were significant in distinguishing phonological height in Foodo. Once again, the same model was generated. What may be concluded is that for this particular sample, regardless of how the sample is viewed, this communal grammar is quite robust.

7.0 Conclusion

The findings of this paper contribute to a growing body of articulatory and acoustic studies of languages with the feature [ATR]. Like for other Niger-Congo languages that have been featured in the literature of the past 30 years, the most reliable acoustic correlate of [ATR] in Foodo is F1 frequency. [+ATR] vowels have consistently lower F1 mean values than their [-ATR] harmonic counterparts while differences in F2 between the vowel pairs are less consistent across speakers. Unlike other studies, this paper focuses on the F1 mean overlap of [+ATR] mid vowels and [-ATR] high vowels in terms of phonological height rather than vowel quality differences. Center of Gravity, a measure of spectral mean, is shown to be 1) a significantly reliable means of disambiguating the phonetic neutralization of phonological vowel height in two of the speakers, 2) an enhancement in the case of the one speaker for whom F1 mean averages were significant for all heights, and at the worst, 3) reliable for high tone tokens for the remaining speaker. Finally, this paper presents a novel way of considering phonetic variation among a sample of a population, namely by pooling the data in a model of statistical mode, mean or median. An idealized or communal grammar emerges that is more robust than the results of any one speaker thus being potentially capable of handling variation due to vocal
tract differences or to less robust results of a single acoustic correlate. The flexibility of such a
model is deemed necessary to accommodate a phonological feature, such as [ATR], which if it
involves a constriction somewhere in the pharyngeal cavity, is a complex or bundle of acoustic
features involving F1, spectral mean and bandwidth differences.

There are admittedly limitations to such claims until further research is conducted, both in Foodo and in other languages with overlapping phonological vowel height. One serious
limitation to this study is the lack of a normalization process of spectral mean as a baseline from
which to compare differences in Center of Gravity mean between harmonic vowel pairs. Such a
comparison is undoubtedly necessary for cases such as speaker 4 where F1 values were sufficient
to maintain phonological vowel height, yet Center of Gravity means also suggested much higher
values for levels comprised of [-ATR] vowels than for [+ATR] vowels. Also unexplainable at
this point of the research is why differences in F1 bandwidth values for the overlapping vowels
of speaker 3 are a more robust measurement than Center of Gravity values. In spite of such
limitations, others will hopefully build upon this modest study and test the notion of a statistical
model of a communal grammar in other areas of phonology where speaker variation is an issue.
Appendix: Summary Table of F1 and F2 Means (in Hertz) for all Speakers

<table>
<thead>
<tr>
<th>Vowel</th>
<th>1-M F1</th>
<th>1-M F2</th>
<th>2-M F1</th>
<th>2-M F2</th>
<th>3-F F1</th>
<th>3-F F2</th>
<th>4-F F1</th>
<th>4-F F2</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i/</td>
<td>330</td>
<td>2114</td>
<td>316</td>
<td>2092</td>
<td>312</td>
<td>2522</td>
<td>358</td>
<td>2264</td>
</tr>
<tr>
<td>/ɪ/</td>
<td>458</td>
<td>2056</td>
<td>426</td>
<td>2020</td>
<td>587</td>
<td>2260</td>
<td>539</td>
<td>2315</td>
</tr>
<tr>
<td>/ɛ/</td>
<td>457</td>
<td>2185</td>
<td>416</td>
<td>1943</td>
<td>514</td>
<td>2090</td>
<td>587</td>
<td>1949</td>
</tr>
<tr>
<td>/ɛ/</td>
<td>705</td>
<td>1925</td>
<td>572</td>
<td>1752</td>
<td>653</td>
<td>1954</td>
<td>753</td>
<td>1951</td>
</tr>
<tr>
<td>/a/</td>
<td>898</td>
<td>1494</td>
<td>825</td>
<td>1388</td>
<td>904</td>
<td>1555</td>
<td>949</td>
<td>1532</td>
</tr>
<tr>
<td>/ɔ/</td>
<td>662</td>
<td>1017</td>
<td>613</td>
<td>1003</td>
<td>712</td>
<td>1284</td>
<td>802</td>
<td>1253</td>
</tr>
<tr>
<td>/o/</td>
<td>485</td>
<td>1006</td>
<td>433</td>
<td>984</td>
<td>551</td>
<td>1282</td>
<td>625</td>
<td>1252</td>
</tr>
<tr>
<td>/ʊ/</td>
<td>503</td>
<td>875</td>
<td>435</td>
<td>832</td>
<td>536</td>
<td>1118</td>
<td>571</td>
<td>989</td>
</tr>
<tr>
<td>/u/</td>
<td>375</td>
<td>841</td>
<td>355</td>
<td>711</td>
<td>342</td>
<td>1021</td>
<td>369</td>
<td>1043</td>
</tr>
</tbody>
</table>
References:


