Acoustic Phonetic Aspects of a 7/9 Vowel Inventory: A Study of the [ATR] Harmony Language Zaghawa

Timothy K. Mathes

1. Introduction

Zaghawa is classified as a Saharan language of the Nilo-Saharan phylum, and is primarily spoken in Northern Darfur, Sudan and Eastern Chad, with a worldwide population total of approximately 169,000 (Lewis 2009). Zaghawa exhibits a phonological pattern of vowel harmony based on the feature Advanced Tongue Root [ATR], which is often used to label two sets of contrastive vowels: the [+ATR] vowels are articulated with a more advanced tongue root position in the vocal tract compared to their [-ATR] counterparts (Ladefoged 1964; Stewart 1967; Painter 1973). In the case of ATR harmony, the vowels in some domain must share the same [ATR] feature (Lindau et al. 1972; Ladefoged and Maddieson 1996; Casali 2003). Although ATR harmony has been studied in a number of languages, very little acoustic phonetic work has been conducted on ATR harmony systems in Nilo-Saharan languages (Guion et al. 2004).

The current study examines acoustic correlates of the feature [ATR] in the Kube dialect of Zaghawa. In addition, I will show that the [+ATR] mid vowels [e], [o] and [-ATR] high vowels [i], [u] may be distinguished phonetically despite their close proximity in an F1 versus F2 plane. The similar F1 and F2 values of [+ATR] mid vowels as compared to [-ATR] high vowels have been found in a number of studies of ATR harmony languages (e.g., Hess 1992; Fulop et al. 1998; Casali 2002; Guion et al. 2004; Local and Lodge 2004; Gick et al. 2006; Starwalt 2008; Kenstowicz 2009), and the issue of how these vowels might be disambiguated is a topic of continuing debate.

2. The Zaghawa vowel inventory

Jakobi and Crass (2004) and Wolfe (2001) claim the Kube dialect has 9 underlying vowels: [+ATR] /i, e, o, u/, and [-ATR] /ɪ, ɛ, ɔ, ʊ, a/, as outlined in (1).

(1) Zaghawa (Kube dialect) vowel inventory (Jakobi and Crass 2004:19; Wolfe 2001:36)

<table>
<thead>
<tr>
<th></th>
<th>front</th>
<th>central</th>
<th>back</th>
</tr>
</thead>
<tbody>
<tr>
<td>[+ATR] vowels</td>
<td>high</td>
<td>i</td>
<td>u</td>
</tr>
<tr>
<td></td>
<td>mid</td>
<td>e</td>
<td>o</td>
</tr>
<tr>
<td>[-ATR] vowels</td>
<td>high</td>
<td>ɪ</td>
<td>ʊ</td>
</tr>
<tr>
<td></td>
<td>mid</td>
<td>ɛ</td>
<td>ɔ</td>
</tr>
<tr>
<td></td>
<td>low</td>
<td>a</td>
<td></td>
</tr>
</tbody>
</table>

(2) shows examples of Zaghawa vowel harmony in which all of the stem vowels are either [+ATR] or [-ATR] (tones have been omitted for clarity).

* Timothy K. Mathes, New York University, tkm237@nyu.edu. Many thanks to Lisa Davidson, Gillian Gallagher and Greg Guy for comments and corrections on an earlier draft of this paper as well as the audience at the CUNY Conference on the Feature in Phonology and Phonetics for suggestions. A special thanks goes to the Zaghawa consultants for their participation. I'd also like to give a warm thanks to Dan Kaufman of the Endangered Language Alliance for his tireless help. As usual, any and all errors remain mine.
(2) a. [+ATR] stem vowels
- meri 'hot season'
- suru 'shell'
- sobu 'ashes'
- biri 'dog'
- bi: 'water'

b. [-ATR] stem vowels
- isti 'four'
- tibe 'field'
- msa 'pot'
- nditi 'heron'
- sura 'lion'

Phonemic 9-vowel ATR harmony languages are extremely common geographically and genetically within the Niger-Congo and Nilo-Saharan phyla, with Akan and Maasai as well-known examples (Casali 2008). The system described in (1) is unsurprising. However, in a review of the Jakobi and Crass grammar, Anonby (2007:219) writes: 'Jakobi and Crass are dealing with a fascinating expression of vowel harmony in Beria [Zaghawa], but their treatment of it (like that of earlier researchers on the language) obscures the most exceptional characteristic of the system: the particular shape of skewing between underlying contrast and surface realization...the [+ATR] mid vowels [e, o] are always found in the context of a [+ATR] high vowel i or u, and are never found alone...such a gap demonstrates that while nine vowels are phonetically present, only seven are underlingly contrastive; the [+ATR] mid vowels [e, o] are in fact realizations of ɛ, ɔ in a [+ATR] environment.' This type of 7/9 (underlying/surface) ATR system is uncommon (ibid; Casali 2008), although Kutsch Lojenga and Waag (2004) reports allophonic [e], [o] in the context of [+ATR] high vowels [i], [u] in the Sudanese language Fur. Thus, Zaghawa likely has the 7-vowel inventory seen in (3).

(3) Zaghawa (Kube dialect) vowel inventory (revised)

a. [+ATR] vowels
- front    central    back
  high    i         u

b. [-ATR] vowels
- front    central    back
  high    i         u
  mid     ɛ         ɔ
  low     a

The inventory shown in (3) also occurs in Kinande (Mutaka 1995; Gick et al. 2006). In contrast, (4) is a 7-vowel inventory that lacks the [-ATR] high vowels /ɪ/ and /ʊ/, which is found in Yoruba (Awobuluyi 1967; Archangeli and Pulleyblank 1989).

(4) Yoruba vowel inventory

a. [+ATR] vowels
- front    central    back
  high    i         u
  mid     e         o

b. [-ATR] vowels
- front    central    back
  mid     ɛ         ɔ
  low     a

Languages that have the inventory in (4) are extremely common, while the Zaghawa inventory in (3) is much less common (Casali 2008). It is this uncommon vowel inventory that gives us a clue as to how [+ATR] mid vowels [ɛ, o] and [-ATR] high vowels [i, u] might be distinguished (which will henceforth be referred to as the cross-height pairs), in addition to potential acoustic-phonetic differences. The results of this study are of phonetic typological interest, giving us a more complete empirical picture of cross-linguistic variation in ATR harmony languages.
3. Methods

3.1. Speakers

Two male speakers of the Kube dialect were recruited for the study through the New York City based Endangered Language Alliance (http://elalliance.org/). Neither of the consultants reported any history of speech or hearing disorders. Because the consultants had very limited availability, the study was designed to elicit as much relevant acoustic data as possible to maximize the time period they were available. The consultants are referred to in this paper as speakers AM and NR.

3.2. Materials

In order to examine the Zaghawa [±ATR] vowels and cross-height pairs, a word list was culled from the French-Zaghawa word list of approximately 950 entries found in the glossary of Jakobi and Crass (2004). Monosyllabic, disyllabic and trisyllabic words were selected based on whether the vowel followed a coronal or labial consonant to minimize the influence of the consonant on the formant frequencies of the vowel. Vowels after velar consonants were excluded from the word list. The French words were translated into English using Google translate. The entire list totaled 94 words: 91 nouns, 2 imperatives and 1 adverb. The large corpus was chosen to maximize the sample size.

Zaghawa does not have a standardized orthography in wide use, as is common with languages in the Darfur region of Sudan. Since a reading task was not viable, the consultant was given the list of English words and asked to translate each word in isolation as naturally as possible with at least 4-5 repetitions. Occasionally, clarification had to be given regarding the meanings of words, e.g., 'pot' referred to a 'cooking pot' and not just any pot.

3.3. Data acquisition

The Zaghawa consultants were recorded on separate dates in a sound-treated booth at the Endangered Language Alliance offices using split-channel recording: channel 1 was recorded using a Countryman Isomax E61DP562 headset microphone, while channel 2 was interfaced with a Rode NTG-3-B shotgun microphone. A Sound Devices 722 digital recorder set to 44.1 kHz with 16-bit resolution saved the recordings to an internal hard drive in .wav format. The .wav files were then transferred to a laptop and channel 1 (Countryman Isomax headset microphone) was extracted for all acoustic analyses.

3.4. Acoustic measurements

F1, F2, f0, duration and intensity were examined as potential acoustic correlates of [ATR] in Zaghawa, while those measures plus phonation differences were considered for the cross-height pairs. For overviews of acoustic measures used in previous studies of ATR vowels, the interested reader should consult Fulop et al. (1998), Guion et al. (2004), Starwalt (2008), Kenstowicz (2009) and references cited therein. The spectral tilt measures H1-H2, which is the first harmonic amplitude minus the second harmonic amplitude of f0, and H1-A2, which is the first harmonic amplitude minus the amplitude of F2's strongest harmonic, were included for the cross-height pairs in this study because these measures have been found to reliably distinguish phonation differences (Kingston et al. 1997; Blankenship 2002; Przezdziecki 2005; Keating and Esposito 2006). An expanded study in the future will include more acoustic measures.

The acoustic analyses were conducted in Praat 5.1.25 (Boersma and Weenink 2010). Two TextGrid tiers were created to designate the English and Zaghawa transcriptions of the utterances. A third tier was created to mark the vowel boundaries using the waveform and spectrogram. The left edge of the vowel following stops was segmented at the closest zero-crossing of the first full glottal pulse corresponding with visible second formant energy. Following fricatives, the left edge was marked where the fricative turbulence ended and second formant energy was visible. The right edge was marked at the closest zero-crossing where second formant energy ended. A handful of lexical items were not marked when one consultant did not produce the same word as the other. Although the Zaghawa consultants were told to speak each word in isolation 4-5 times, there were instances when words were repeated 6-7 times. To keep the number of measurements consistent across and within
speakers, the vowels from the final 4 repetitions were chosen. Once the vowel boundaries had been marked, an automated Praat script was run which measured F1, F2, f0 and intensity at the vowel midpoints. The script was coded to detect 5 formants below 5000 Hz, window length .025 seconds and a dynamic range of 30 dB. Duration was calculated for each vowel by the same script. Although duration was measured for both long and short vowels, the long vowel measurements were excluded from the duration analysis, since Zaghawa has a phonological vowel length contrast. A second automated Praat script calculated the phonation measures H1-H2 and H1-A2 for the cross-height pairs. Statistical analysis was carried out in SPSS 19.

4. Results

4.1. [ATR] harmonic pairs

The acoustic properties of the [+ATR] pairs are considered first. Plots of the mean formant values for the data set are given in Figure 1 by speaker (n = 324 per speaker). As can be seen in the plots, [+ATR] vowels have a consistently lower F1 value than their [-ATR] harmonic counterparts.

These effects are consistent with the expansion of the pharyngeal cavity by advancement of the tongue root (Halle and Stevens 1969), and with what has been found in other languages (Hess 1992 for Akan; Fulop et al. 1998 for Degema; Guion et al. 2004 for Maa; Local and Lodge 2004 for Kalenjin; Przezdziecki 2005 for Yoruba; Gick et al. 2006 for Kinande). The [+ATR] mid vowels [e], [o] and the [-ATR] high vowels [ɪ], [ʊ] are in close proximity for both speakers, although [ʊ] has a higher mean F1 value than [o] for speaker NR. This vowel swapping will be investigated in future research and will not be addressed here.

A statistical analysis was conducted on the data set to determine whether the effect of [ATR] on F1 was statistically significant for the [+ATR] pairs and to check for interactions. A univariate analysis of variance (ANOVA) was run with F1 as the dependent variable and Vowel Pair (i, e, o, u) and ATR (+ or -) as independent factors. The low vowel [a] was excluded from the analysis since it did not appear to have a harmonic counterpart according to F1-F2 formant plots, which is consistent with the Zaghawa literature. For speaker AM, there are significant main effects of Vowel Pair, F(3, 280)=276.213, p<.001, ATR, F(1, 280)=411.652, p<.001, as well as a significant Vowel Pair * ATR interaction, F(3, 280)=5.303, p=.001. For speaker NR, there are significant main effects of Vowel Pair, F(3, 280)=131.291, p<.001, ATR, F(1, 280)=327.471, p<.001, as well as a significant Vowel Pair * ATR interaction, F(3, 280)=4.840, p=.003. The significant interactions are likely due to the amount of change in F1, which differs for each vowel even though the direction of change is the same.

The F1 and F2 data were analyzed by running t-tests for 6 planned comparisons: 4 comparisons of the [+ATR] pairs and 2 for the cross-height pairs (with adjusted alpha set to .0083). The results for the cross-height pairs will be discussed in the next section. For speaker AM, all [+ATR] pairs are significantly different with respect to F1, p<.001 in all cases: [i], [ɪ], t(35)=-13.571; [e], [ɛ], t(64)= -7.914; [u], [ʊ], t(70)=-9.091; [o], [ɔ], t(54)=-9.454. F2 is significantly different for all
[+ATR] vowels compared to their [-ATR] harmonic counterparts: [i], [ɪ], t(82) = 4.579, p < .001; [ɛ], [e], t(74) = 6.201, p < .001; [u], [ʊ], t(66) = -3.121, p = .003; [ɔ], [o], t(54) = 3.342, p = .002.

For speaker NR, t-tests reveal that F1 is significantly lower for [+ATR] vowels compared to their [-ATR] harmonic counterparts, p < .001 in all cases: [i], [ɪ], t(82) = -10.310; [ɛ], [e], t(74) = -10.228; [u], [ʊ], t(50) = -7.455; [o], [ɔ], t(54) = -14.625. Moreover, t-tests reveal that F2 is significantly higher for the [+ATR] front vowels compared to their [-ATR] harmonic counterparts: [i], [ɪ], t(82) = 4.135, p < .001; [ɛ], [e], t(74) = 7.338, p < .001. [u], [ʊ] and [o], [ɔ] are not significantly different.

Duration and intensity were checked for short vowels to see if these measures could potentially distinguish [+ATR] pairs. T-tests did not find a significant difference between the [+ATR] pairs with respect to duration for AM. For NR, the harmonic pair durations did not vary in any particular direction either. T-tests only found a significant difference between the front vowels [i] and [ɪ], t(52) = 3.335, p = .002. The rest of the harmonic pairs were not significant. Moreover, t-tests for both speakers did not find a significant difference between the mean intensities of [+ATR] pairs. To summarize, neither speaker’s short vowels have significant differences in average duration or intensity for the harmonic pairs, with the exception of [i] and [ɪ] for NR.

Fundamental frequency (f0) was analyzed to determine if the [+ATR] vowels differed in terms of this measure while taking into account Zaghawa’s three surface tones, Low (L), Mid (M) and High (H). Contour tones were excluded from the analysis. Figure 2 shows a plot of mean f0 values (Hz) for [+ATR] vowels by tone level for each speaker.

![Figure 2](image_url)

**Figure 2** Mean f0 for speaker AM (left) and NR (right) by tone level.

For speaker AM, a univariate ANOVA with Tone Level (L, M or H) and ATR (+ or -) as independent factors and f0 as the dependent variable shows a significant main effect of Tone Level, F(2, 285) = 45.294, p < .001. Tukey’s HSD post-hoc tests show a significant difference between f0 of the three tone levels L, M, and H for speaker AM. Nevertheless, the independent factor ATR and the Tone Level * ATR interaction are not significant. Thus, there is a trend for [+ATR] vowels to have higher f0 values than [-ATR] vowels at the three tone levels, but the difference is not significant. For NR, a univariate ANOVA shows a significant main effect of Tone Level, F(2, 297) = 61.229, p < .001. Tukey’s HSD post-hoc tests show that there is no significant difference in f0 between M and H, although M and H both significantly differ from L. Furthermore, there is a significant main effect of ATR, F(1, 297) = 16.391, p < .001, without a Tone Level * ATR interaction. While the results of a significantly higher f0 of [+ATR] vowels for NR and the trend of a higher f0 for [+ATR] vowels for AM are promising and consistent with Whalen and Gick (2001), it does not appear to be a robust acoustic correlate for these speakers.

### 4.2. Cross-height pairs

Although the cross-height pairs are in close proximity in the F1 versus F2 plane, the F1 means for the front vowel cross-height pair [ɛ], [ɪ] are significantly different for speaker AM, t(98) = -5.569, p < .001, and the F1 means for the back vowel cross-height pair [o], [ʊ] are significantly different,
$t(70)=-3.126$, $p=.003$. For NR, the front vowels $[e]$, $[i]$ are significantly different, $t(98)=-5.781$, $p<.001$, as well as the back vowels $[o]$, $[u]$, $t(59)=3.341$, $p=.001$. In terms of F2, the mean differences between the cross-height pairs $[e]$, $[i]$ and $[o]$, $[u]$ are not significant for AM. For NR, the F2 mean difference between $[e]$, $[i]$ is significant, $t(98)=4.539$, $p<.001$, but the F2 mean difference for $[o]$, $[u]$ is not.

With respect to the duration results, the cross-height pairs for AM have a pattern of the [-ATR] high vowels $[i]$, $[u]$ being shorter than the [+ATR] mid vowels $[e]$, $[o]$. The durations of $[e]$, $[i]$, $t(90)=-3.073$, $p=.003$, and $[o]$, $[u]$, $t(70)=-3.585$, $p=.001$, are significantly different. Speaker NR patterns in the same fashion but the differences are not significant. All that can be concluded is that there is a trend in speaker NR's cross-height vowel durations that is consistent with speaker AM's data.

Focusing on intensity, AM's $[i]$, $[o]$ have lower mean intensity values compared to $[e]$, $[o]$, respectively. There was significance between the front vowels $[e]$, $[i]$, $t(90)=-4.314$, $p<.001$, but not the back vowels $[o]$, $[u]$. For speaker NR, $[i]$ has a significantly lower mean intensity than $[e]$, $t(90)=-6.063$, $p<.001$, and $[o]$ is significantly lower than $[o]$, $t(70)=-4.642$, $p<.001$. Thus, speaker AM's cross-height pairs have a trend in the same direction as NR's, while NR's cross-height pairs are significantly different.

To summarize, speaker AM's [-ATR] high vowels $[i]$, $[u]$ were significantly shorter than his [+ATR] allophonic mid vowels $[e]$, $[o]$, while NR had a trend in the same direction. Speaker NR's high vowels $[i]$, $[u]$ were significantly lower in mean intensity compared to his [+ATR] mid vowels $[e]$, $[o]$, while AM had a trend in that direction. It may be that there is inter-speaker variation for the production of the cross-height pairs to help listeners discriminate these vowels, which can only be confirmed through future research with more speakers. It is worth noting that while F1 is significantly different for the cross-height pairs of both speakers, the mean differences are not large: $[e]$, $[i]=36$ Hz and $[o]$, $[u]=27$ Hz for AM; $[e]$, $[i]=37$ Hz and $[o]$, $[u]=57$ Hz for NR. These values are at the upper end or slightly higher than the Just Noticeable Difference (JND) range of 10-30 Hz for discriminating F1 (Stevens 1998).

The spectral tilt measures H1-H2 and H1-A2 were run for the cross-height pairs in this study to see if these vowels might be disambiguated by phonation differences. Planned comparison t-tests with H1-H2 as the test variable did not find significant differences for either speaker. For the H1-A2 measure, the only pair found to be significantly different was $[e]$, $[i]$ for speaker AM, $t(94)=3.348$, $p=.001$. The rest of the cross-height pairs were not significant.

5. Discussion

Tables 1 and 2 provide a description of which measures might be used to reliably distinguish the [ATR] and cross-height pairs for AM and NR, respectively. These measures were found to be statistically significant.

Table 1 Speaker AM summary.

<table>
<thead>
<tr>
<th>Vowels</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ATR] pairs</td>
<td>F1 and F2.</td>
</tr>
<tr>
<td>Cross-height pairs</td>
<td>F1; H1-A2 for front vowels [e], [i]; Duration.</td>
</tr>
</tbody>
</table>

Table 2 Speaker NR summary.

<table>
<thead>
<tr>
<th>Vowels</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ATR] pairs</td>
<td>F1; F2 for front vowels [i], [r]; [e], [e]; f0.</td>
</tr>
<tr>
<td>Cross-height pairs</td>
<td>F1; Intensity.</td>
</tr>
</tbody>
</table>
A review of Tables 1 and 2 demonstrates that the most stable acoustic correlate of the feature [ATR] in Zaghawa for these two speakers is F1, which is consistent with the literature on ATR harmony. F2 is not as stable, although AM has a significant difference for his [±ATR] pairs and NR shows a significant difference for his [±ATR] front vowels. Duration, intensity and f0 are not robust acoustic measures for the [±ATR] pairs for these speakers. In addition, while F1 is a stable acoustic correlate for the cross-height pairs, the phonation measures H1-H2 and H1-A2 are not. Speaker AM may be using F1 and duration while speaker NR uses F1 and intensity for the cross-height pairs since the mean difference is at the JND. I speculate that speakers use differing strategies to differentiate these vowels, which can only be confirmed through future research with more participants and additional acoustic measures. In any event, both speakers in this study produce their cross-height pairs significantly different with respect to F1.

Furthermore, since allophonic mid vowels [e], [o] only surface in the context of [±ATR] high vowels [i], [u], a striking effect can be observed in (5).

(5) a. tɪr 'name' b. ise 'food'
dɪm 'tail' dime 'wild cat'
ɪ: 'eye' tine 'thief'
ɑ: 'calf' eni 'riverbed'
u: 'giraffe' ori 'nice smell'
ʊdɑ 'fresh milk' sobu 'ashes'
ʊba 'baby' todu 'heart'
ʊdo 'back' oni 'behind'

The nouns in (5a) show that the [-ATR] high vowels can occur alone in monosyllables, with other vowels in disyllables, or in two successive syllables. On the other hand, [e] and [o] in (5b) are restricted to disyllabic words with either [i] or [u]. The restricted distribution of [e], [o] provides another cue for their identity: if the environment occurs, then the vowels are [e], [o]. Thus, the combination of F1 and distribution may help listeners distinguish these vowels.

References
